UNIVERSITY OF SOUTHAMPTON
FACULTY OF SOCIAL SCIENCES
School of Economic, Social and Political Sciences

Aspects of Estimation and Inference for Predictive Regression Models

by

Christis Katsouris
(B.Sc., M.Sc., MBA, M.Sc.)

ORCID ID 0000-0002-8869-6641

Thesis for the degree of Doctoral of Philosophy

December 2022
(September 2023)

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCES

SCHOOL OF ECONOMIC, SOCIAL AND POLITICAL SCIENCES

Doctor of Philosophy

Aspects of Estimation and Inference for Predictive Regression Models

by

Christis Katsouris

This PhD thesis[1] presents three essays on nonstationary time series econometrics which are grouped into three chapters. The chapters cover aspects of estimation and inference for predictive regression models through the lens of moderate deviation principles from the unit root boundary in a class of stable but nearly-unstable processes which exhibit high persistence.

The first chapter presents an overview of the research background which includes the persistence classes and the main asymptotic properties of estimators for nonstationary autoregressive processes. The persistence properties of time series is modeled via the local-to-unity parametrization which implies that the autoregressive coefficient is specified such that it approaches the unit boundary as the sample size increases. The second part of the chapter summarizes the structure of the thesis and the main contributions to the literature.

The second chapter, proposes an econometric framework for predictability testing in linear predictive regression models robust against parameter instability. In particular, the asymptotic theory for the proposed sup-Wald test statistics when regressors are assumed to be mildly integrated and persistent stochastic processes is established. The asymptotic theory of OLS and IVX based estimators and test statistics presented in this thesis is developed based on standard local-to-unity asymptotics and the limit theory of triangular arrays of martingales.

The third chapter, addresses the aspect of structural break detection for nonstationary time series when a conditional quantile specification form is used. In particular, the proposed econometric framework is suitable for testing for a structural break at unknown time in nonstationary quantile predictive regression models and can be further employed for investigating the aspect of quantile predictability against parameter instability.

The fourth chapter, proposes a novel estimation and inference methodology in systems of quantile predictive regressions with generated regressors. This econometric framework allows to address the issue of modelling systemic risk in financial networks when considering the interplay between network-type of dependence and time series nontationarity.

[1]Christis Katsouris (2022). "*Aspects of Estimation and Inference for Predictive Regression Models*", University of Southampton, Faculty of Social Sciences, Department of Economics, PhD Thesis.

Validation page

PhD Candidate: Christis Katsouris

PhD Thesis Title: Aspects of Estimation and Inference for Predictive Regression Models

The present PhD thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the Department of Economics in December 2022 for approval by the members of the Examination Committee.

Academic Advisors:

Academic Advisor: Dr. Jose Olmo, Professor, University of Southampton

Academic Advisor: Dr. Tassos Magdalinos, Professor, University of Southampton

Examination Committee:

Committee Member: Dr. Patrick Marsh, Associate Professor, University of Nottingham

Committee Member: Dr. Jan Podivinsky, Associate Professor, University of Southampton

Committee Member: Dr. Carmine Ornaghi, Professor, University of Southampton

Contents

List of Figures

List of Tables

Declaration of Authorship

Print Name: Christis Katsouris

Title of Thesis: Aspects of Estimation and Inference for Predictive Regression Models

I, Christis Katsouris, declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledge all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission;

Signed: C.Katsouris

Date: December 2022

Acknowledgements

I would like to express my gratitude to my academic advisors from the Department of Economics consisting of Jose Olmo (Professor of Financial Economics) and Tassos Magdalinos (Professor of Econometrics) as well as Jean-Yves Pitarakis (Professor of Economics) for guidance, support and continuous encouragement throughout the PhD programme. Jose Olmo trusted in my candidacy to the PhD programme and helped me to consolidate many ideas and improve the presentation of my work. Tassos Magdalinos provided expertise and insights that foster rigorous econometric thinking. Jean-Yves Pitarakis provided feedback that helped me improve my technical skills related to econometric theory aspects. I am grateful for feedback and extremely useful comments from all. Their approach to research and innovative contributions to the literature, inspired and motivated my passion to do research in the field of time series econometrics.

Moreover, I wish also to thank the academic staff at the Department of Economics, as well as the Faculty of Social Sciences, the Doctoral College and other PhD students for creating a stimulating environment for personal and professional growth as well as academic excellence. I also gratefully acknowledge financial support from the Vice Chancellor PhD studentship of the University of Southampton as well as the Department of Economics and the Department of Social Statistics for providing additional funding through opportunities to work as a Graduate Teaching Assistant.

I wish also to thank my family and friends for encouraging and supporting me during my studies.

Definitions and Abbreviations

$x \vee y$	$\max\{x, y\}$
$x \wedge y$	$\min\{x, y\}$
$:=$	equality by definition
$\equiv$	equivalent statement
$\lfloor\,.\,\rfloor$	integer part of the argument
$\lvert\,.\,\rvert$	absolute value of the argument
$I\{.\}$	indicator function
$\mathbb{E}[\,.\,]$	expectation operator
$\mathbb{P}(\,.\,)$	probability operator
$\mathrm{Var}(\,.\,)$	variance operator
$\top$	transpose operator
$\sup(\,.\,)$	supremum functional
$\inf(\,.\,)$	infimum functional
$\lVert\,.\,\rVert_2$	euclidean norm
$\mathcal{O}_p(\,.\,)$	Order of convergence
$\rightarrow_d$	Convergence in distribution
$\rightarrow_p$	Convergence in probability
$\Rightarrow$	Weak convergence argument
$\mathcal{D}([0,1])$	Skorokhod topology
$W(.)$	Brownian motion with variance σ^2
$B(.)$	Brownian motion with variance ω^2
$J_c(.)$	Ornestein-Uhlenbeck process

Research Thesis Title

Aspects of Estimation and Inference for Predictive Regression Models

To my family

Chapter 1

Introduction

1.1 Background

Consider the first-order autoregressive process $\{X_t\}_{t=1}^{+\infty}$, defined by the following recursive process

$$X_t = \vartheta_n X_{t-1} + \varepsilon_t, \qquad X_0 = 0, \tag{1.1}$$

where $\{\varepsilon_t\}$ are independent $\mathcal{N}(0,1)$ random sequences. The least squares estimator $\hat{\vartheta}_n$ of ϑ, based on a sample of n observations $\{X_1, ..., X_n\}$ is given by the following expression

$$\hat{\vartheta}_n = \left(\sum_{t=1}^{n} X_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^{n} X_{t-1} X_t \right). \tag{1.2}$$

It is well known that $\hat{\vartheta}_n$ is a consistent estimator for ϑ_n for all values of $\vartheta_n \in (-\infty, +\infty)$. More precisely, the asymptotic distribution of $\hat{\vartheta}_n$ depends on restrictions imposed on the admissible parameter space of ϑ. In particular, in the stable case, which implies that the parameter space of ϑ takes values within the unit circle, that is, $|\vartheta| < 1$, various seminal studies such as Mann and Wald (1943) and Anderson (1959) among others, have shown that

$$\sqrt{n} \left(\hat{\vartheta}_n - \vartheta_n \right) \xrightarrow{d} \mathcal{N}(0, 1 - \vartheta^2) \quad \text{as} \quad n \to \infty. \tag{1.3}$$

When the true parameter lies on the boundary of the parameter space, such that $|\vartheta| = 1$, it has been shown by White (1958) and Phillips (1987b) that the following limiting distribution holds

$$n \left(\hat{\vartheta}_n - \vartheta_n \right) \xrightarrow{d} \left(\int_0^1 W(r)^2 ds \right)^{-1} \left(\int_0^1 W(r) dW(r) \right) \quad \text{as} \quad n \to \infty, \tag{1.4}$$

where $W(r)$ for some $0 \leq r \leq 1$ is a standard Wiener process within the probability space $(\Omega, \mathbb{P}, \mathcal{F}_t)$. Furthermore, when the true parameter of the autoregressive process given by (1.1) is outside the unit circle (explosive parameter region), such that $|\vartheta| > 1$, it has been shown by White (1958) and Anderson (1959) that the following asymptotic result holds

$$\sqrt{n} \left(\hat{\vartheta}_n - \vartheta_n \right) \xrightarrow{d} \text{Cauchy} \left(0, \vartheta^2 - 1 \right) \quad \text{as} \quad n \to \infty. \tag{1.5}$$

The aforementioned statistical properties have been an important aspect of consideration in both the statistics and the time series econometrics literature (see, Rao (1978), Dickey and Fuller (1979), Chan and Wei (1987), Phillips (1989), Abadir (1993), Buchmann et al. (2007), and Phillips and Magdalinos (2007)). Without loss of generality, considering for each $n \in \mathbb{N}$, the statistical experiment $\mathcal{E}_n$ corresponding to the observations $\{X_0, ..., X_n\}$, the sequence $(\mathcal{E}_n)_{n \in \mathbb{N}}$, where $\mathcal{E}_n$ represents the experiment under consideration (i.e., recursive process), is locally asymptotically normal (LAN) if $|\vartheta| < 1$, it is locally asymptotically Brownian functional ($LABF$) if $|\vartheta| = 1$, and it is locally asymptotically mixed normal ($LAMN$) if $|\vartheta| > 1$ (Le Cam and Yang (2012). The local-to-unity asymptotic framework was introduced by Chan and Wei (1987), Phillips (1987a), Phillips (1987b), Phillips (1988a), Phillips (1988b), Phillips and Perron (1988).

Specifically, the class of nearly nonstationary time series models as proposed by the seminal studies of Phillips (1987a) and Phillips (1987b) who developed asymptotics for near-integrated time series provide a unified approach to inference. In practise, expressing the autocorrelation coefficient with the local-unit-root specification, such that, $\vartheta_n = (1 + c/n)$, where n the sample size and c denotes the nuisance parameter of persistence, provides a modelling methodology that encompasses such moderate deviations from the unit boundary (e.g., unit root, near-integrated or explosive). Furthermore, the nonstandard nature of the inference problem in predictive regressions has been previously discussed in the literature such as by the studies of Phillips and Hansen (1990), Cavanagh et al. (1995) and Jansson and Moreira (2006), with the main challenge being the robustness of inference methodologies to the nuisance parameter of persistence. In addition, Phillips and Magdalinos (2009) and Kostakis et al. (2015) have proposed a unified framework for robust estimation and inference regardless of the unknown degree of persistence.

Figure 1

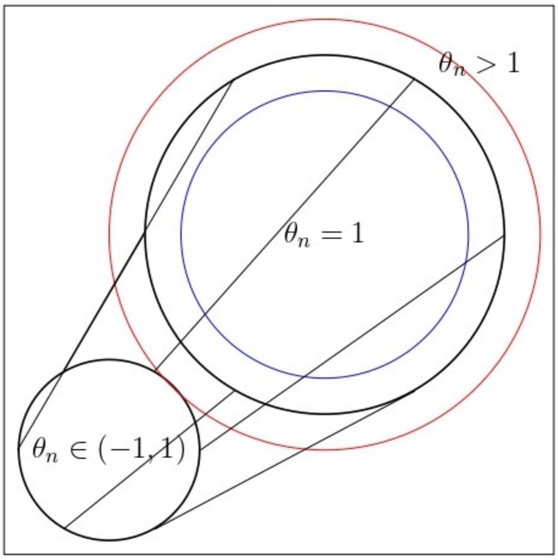

Figure 1.1: Spectrum of degrees of nonstationarity

All three essays in the thesis are developed around the idea that when considering nonstandard inference problems in time series models the asymptotic theory will be affected by the nuisance parameters of persistence which capture the nonstationary properties. In particular, the aspect of nonstationarity in time series has important implications for estimation and testing purposes and in simple terms describes the integration order. Although the conventional approach in stationary time series modelling environments is to consider the first difference approach to ensure covariance stationarity, it is found to result in loss of information especially with respect to long-memory properties. In contrast, within the predictive regression framework the persistence properties of regressors are captured by the nuisance parameter of persistence c_i and the exponent rate γ. Thus, in our modelling environment the degree of persistence allows to model moderate deviations from the unit boundary by employing the local-to-unity specification and the related asymptotics in order to establish weak convergence of sample moments to Brownian motion functionals.

Firstly, against this background testing the null hypothesis of no parameter stability in predictive regression models, implies that the testing hypothesis is formulated to capture unobserved time-variation in the coefficients of the model. Under the alternative hypothesis which implies the presence of a single structural break, we basically have the case where the model coefficients are non constant throughout the sample. However, allowing for an unknown break-point location, implies that we have the problem of an unidentified parameter under the null hypothesis, also called Davies problem, which requires to consider the supremum functional when testing for a structural break as well while deriving the asymptotic distribution of the proposed tests statistics (see, Davies (1977), Davies (1987) and Hansen (1996), Hansen (2000b)). Additionally, operating within the predictive regression framework implies that regressors are generated as nonstationary AR(1) autoregressive processes, which means expressing the autocorrelation coefficient with respect to the local unit root specification and therefore permits to capture abstract degree of persistence by considering different values of the parameter space of c_i and γ_x.

More precisely, we assume that the autocorrelation coefficient to take values near the unit boundary, but we consider two cases, such that it approaches unity from below but with different convergence rates (local unit root or mildly integrated). In other words, we restrict the permissible values of these nuisance parameters, such as $c_i > 0$ for all $i \in \{1, ..., p\}$ and the exponent rate γ, that controls the degree of persistence of the predictors, to take values such that $0 < \gamma \leq 1$. Consequently, the asymptotic behaviour of test statistics, especially when testing for structural breaks, relies on the implementation of suitable invariance principles. Specifically, for the development of the asymptotic theory in the thesis, we employ the fundamental building block when deriving asymptotics for nonstationary time series models such as the weakly convergence of invariance principles into their Brownian motion counterparts.

The concept of invariance principles was first introduced by the paper of Erdös and Kac (1946) as well as generalized by Prokhorov (1956) (see, discussion in Kuelbs (1968)). Furthermore, Kuelbs (1968) considers the weak convergence of these random elements into continuous functionals on the space of real-valued functions that includes Gaussian measures as analogues to Wiener measures. More precisely, the particular property of such topological spaces motivated the construction of a new metric such as the approach of Skorokhod (1956), who introduced a number of metrics on the space of cádlag functions on $[0, 1]$, with J_1 topology being the most widely used in statistics and econometric asymptotic theory. A first introduction of the particular weak convergence arguments when considering the asymptotics of autoregressive processes is presented by White (1958) and White (1959). Then, the asymptotic theory of model estimators for autoregressive processes was established by Chan and Wei (1987) and by the seminal papers of Phillips (1987a), Phillips (1987b) and Phillips and Durlauf (1986) who establish the concepts of invariance principles for time series regression models and the asymptotic theory for obtaining the limiting behaviour of estimators and test statistics (see, also Phillips and Solo (1992)).

Secondly, we consider the statistical estimation and inference problem in quantile time series models which occurs when the main purpose is to estimate the conditional quantile instead of the conditional mean for quantile time series models. Therefore, in Chapter 3 and 4 of the thesis we concentrate on the particular modelling methodology within the nonstationary econometric environment. The first application (Chapter 3) examines the structural break detection in nonstationary quantile time series models, while the second application (Chapter 4) proposes an estimation and inference framework for systemic risk in financial networks. In particular, a large literature that considers econometric methods for modelling systemic risk in financial markets (see, Härdle et al. (2016) and Adrian and Brunnermeier (2016)) illustrate the effect of increased tail dependence and the presence of systemic risk appears as higher level of connectedness. Therefore, the novelty in Chapter 4 of the thesis is that we consider the modelling of systemic risk by developing an econometric framework for estimation and inference in quantile predictive regression systems under the assumption of possibly nonstationary regressors.

1.2 The IVX Instrumentation

A key theoretical background to the thesis is the use of instrumental variable estimation methods in the context of nonstationary predictive regressions models, which is the case when the regressors are modelled using an autoregressive specification with an autocorrelation coefficient that has the local-to-unity form, such that, $\rho_n = (1 + c/n^{\gamma_x})$, where γ_x is the exponent rate of the localizing coefficient of persistence and c is the coefficient of persistence that can take either negative values ($c < 0$) when we assume that the underline stochastic processes that generate the regressors are near-stationary or takes a positive values ($c > 0$) when regressors exhibit mildly explosive or explosive asymptotic behaviour both cases in conjecture with what values we allow the exponent rate to take, such that $\gamma_x = 0, \gamma_x = 1$ or $\gamma_x \in (0, 1)$. All of the above cases can be generalized within the moderate deviations framework as proposed by Phillips and Magdalinos (2007), and require the use of IV techniques in order to construct the filtered instruments based on the IVX instrumentation proposed by Phillips and Magdalinos (2009). In this thesis we employ this robust inference procedure (see, Kostakis et al. (2015)) for predictive regressions, that does not require knowledge of the degree of persistence. In particular, the KMS test is a Wald-type statistic which is constructed based on endogenously generated variables and has a chi-squared limit distribution with degrees of freedom equal to the number of stochastic regressors. Usually the interest of standard procedures in the literature is in testing joint significance of all the stochastic regressors included in the regression. The specific testing hypothesis has an exact finite-sample distribution, although some presence of finite sample distortions could be unavoidable due to the bias in the residual-variance estimator which can be corrected using the fully-modified transformation of the corresponding quantities. These finite-sample size distortions disappear asymptotically.

1.2.1 Classification of Autoregressive regions

Next, we provide an assumption which summarizes the persistence properties one can consider with respect to the parameter space of the autocorrelation coefficient that appears in the autoregressive specification that models the possibly nonstationary stochastic regressors of the predictive regression models, the main econometric model of interest in the thesis.

Assumption 1.1. (Persistence class) Consider the specification

$$X_t = \theta_n X_{t-1} + u_t \tag{1.6}$$

Consider the following probability limit

$$\zeta_n := \lim_{n \to +\infty} n\left(\theta_n - 1\right) \to \zeta \tag{1.7}$$

- **P.1** ***nearly stable*** processes: if $(\theta_n)_{n \in \mathbb{N}}$ is such that $\zeta = -\infty$ and it holds that $\theta_n \to |\theta| < 1$.
- **P.2** ***nearly unstable*** processes: if $(\theta_n)_{n \in \mathbb{N}}$ is such that $\zeta \equiv c \in \mathbb{R}$ and it holds that $\theta_n \to \theta = 1$.
- **P.3** ***nearly explosive*** processes: if $(\theta_n)_{n \in \mathbb{N}}$ is such that $\zeta = +\infty$ and it holds that $\theta_n \to |\theta| > 1$.

Notice that various definitions exists in the literature. In particular, the class **P.1** corresponds to the near-stationary definition of Magdalinos (2020) and the class **P.2** corresponds to the near-nonstationary definition of Magdalinos (2020). Some further details regarding these definitions:

The class **P.1** implies that θ_n is near to the unit boundary but there is no local-to-unity specification, therefore the limit tends to $-\infty$. Furthermore, for the class **P.2** we have that $\theta_n \to 1$ but we specify $\theta_n = \left(1 + \frac{c}{k_n}\right)$, where $k_n = n^{\gamma_x}$ with $\gamma_x = 0$, $\gamma_x = 1$ or $\gamma_x \in (0, 1)$. Thus, when $\theta_n = \left(1 + \frac{c}{k_n}\right)$ we have that

$$\zeta_n := \lim_{n \to +\infty} n\left(\theta_n - 1\right) = \lim_{n \to +\infty} n\left[\left(1 + \frac{c}{k_n}\right) - 1\right] = \lim_{n \to +\infty} \left\{\frac{n}{k_n}\right\} c \tag{1.8}$$

In the thesis we mainly focus in the case when $c < 0$, which implies that

1. If $k_n = n^{\gamma_x}$ with $\gamma_x = 1$ $(k_n = n)$, then $\zeta \equiv c \in \mathbb{R},$ *whichfallsintheclassof* **nearly-unstable***processes(near-nonstationary)*.
2. If $k_n = n^{\gamma_x}$ with $\gamma_x \in (0, 1)$, then $\zeta = -\infty$. In this case, although we are in the perimeter of the circle given on Figure 1 but within the "blue zone" then we have a mildly integrated process which falls in the class of **nearly-stable** processes (near-stationary).
3. If $k_n = n^{\gamma_x}$ with $\gamma = 0$ $(k_n = 1)$, then $\zeta = -\infty$ which falls in the class of **nearly-stable** processes (near-stationary).

Based on the above definitions with respect to the stability properties of the autoregressive equation, various methodologies have been proposed for constructing valid statistical inference under the presence of the nuisance parameter of persistence and the exponent rate (e.g., in cases we assume that $\gamma_x \in (0,1)$). However, in our framework we employ the IVX instrumentation which is an approach that although is motivated by an instrumental variable regression estimation methodology, it only uses within the system information to construct the set of endogenously generated instruments without violating commonly used assumptions.

1.2.2 Instrument Construction

Specifically, the IVX filtration is constructed using endogenous instruments, which are based on information contained in the regressors of the predictive regression model. As a result, the degree of persistence of the instrumental variable has degree of persistence explicitly controlled so that the induced process is mildly integrated. To be precise, the IVX instrument is constructed with the first order difference of the corresponding autoregression model which is obtained by expanding the autoregression coefficient and rearranging as below

$$\Delta \boldsymbol{x}_t = -\frac{\boldsymbol{C}}{n^{\gamma_x}} \boldsymbol{x}_{t-1} + \boldsymbol{v}_t, \quad \gamma_x = 0, \gamma_x \in (0,1) \text{ or } \gamma_x = 1. \tag{1.9}$$

In practice, the above first difference sequence is not an innovation process unless the regressor belongs to the persistence class of integrated processes. However, it behaves asymptotically as an innovation after linear filtering by a matrix consisting of near-stationary roots[1]. Therefore, the procedure requires to choose an artificial coefficient matrix of the form

$$\boldsymbol{R}_{nz} = \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}_z}{n^{\delta_z}} \right), \ \delta_z \in (0,1), c_z > 0. \tag{1.10}$$

where $\boldsymbol{C}_z = \mathsf{diag}\{c_{z,1}, ..., c_{z,p}\}, c_z > 0$ for all $i \in \{1, ..., p\}$. Then, the instrumental regressor matrix $\tilde{\boldsymbol{z}}_t \in \mathbb{R}^{n \times p}$ can be constructed as below

$$\tilde{\boldsymbol{z}}_t = \sum_{j=1}^{t} \boldsymbol{R}_z^{t-j} \Delta \boldsymbol{x}_j, \quad \boldsymbol{R}_z = \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}_z}{n^{\delta_z}} \right), \delta_z \in (0,1), \boldsymbol{C}_z > 0. \tag{1.11}$$

The exponent rate δ_x of the generated mildly integrated instrument has a value, $\delta_x < 1$, which is below that of the exponent rate of the regressor. Using expression (1.9) we obtain

$$\tilde{\boldsymbol{z}}_t = \sum_{j=1}^{t} \boldsymbol{R}_z^{t-j} \left(\frac{\boldsymbol{C}}{n^{\gamma_x}} \boldsymbol{x}_{j-1} + \boldsymbol{u}_j \right) = \sum_{j=1}^{t} \boldsymbol{R}_z^{t-j} \boldsymbol{u}_j + \frac{\boldsymbol{C}}{n^{\gamma_x}} \sum_{j=1}^{t} \boldsymbol{R}_z^{t-j} \boldsymbol{x}_{j-1}. \tag{1.12}$$

which can be written via the following expression

$$\tilde{\boldsymbol{z}}_t = \boldsymbol{z}_t + \frac{\boldsymbol{C}}{n} \boldsymbol{\psi}_t, \quad \boldsymbol{z}_t = \sum_{j=1}^{t} \boldsymbol{R}_z^{t-j} \boldsymbol{v}_j \ \text{ and } \ \boldsymbol{\psi}_t = \sum_{j=1}^{t} \boldsymbol{R}_z^{t-j} \boldsymbol{x}_{j-1}. \tag{1.13}$$

[1] Note this assumption is a key idea in the development of the asymptotic theory in cointegrated systems for regressors with various types of nonstationarity (see, Phillips and Magdalinos (2009)).

The IVX filtration, employs the constructed $\widetilde{\boldsymbol{z}}_t$ instruments for the regressors $\boldsymbol{x}_t$ which are considered to behave asymptotically as mildly-integrated processes. More explicitly, by replacing $\boldsymbol{x}_t$ with the instrument $\boldsymbol{z}_t$ which has a controllable degree of persistence, result to a robust inference procedure which accounts for the effects of nonstationarity.

1.3 Summary of the chapters

This thesis comprises three self-contained chapters that cover some aspects of estimation and inference for predictive regression models. Specifically, this thesis considers three complementary nonstandard inference problems in predictive regression models with regressors being generated as local-to-unit root processes. In Chapter 2, the nonstandard inference problem of interest is predictability testing in predictive regressions with highly persistent regressors. In Chapter 3, the relevant nonstandard inference problem corresponds to econometric inference in quantile predictive regressions under the presence of parameter instability and predictors being parametrized as local-to-unity processes. In Chapter 4, we propose an econometric framework for estimation and inference in systems of quantile predictive regressions with generated covariates and near unit root regressors. In terms of empirical data applications, Chapter 2 employs the dataset of Welch and Goyal (2008) to investigate the stock return predictability puzzle robust against parameter instability. Chapter 3, utilizes the dataset of Yang et al. (2020) in order to examine the presence of quantile predictability in house price dynamics under the presence of a single structural break at an unknown location. Chapter 4, employs the dataset of Härdle et al. (2016) to test for the presence of systemic risk spillover effects in systems of quantile predictive regressions with possibly nonstationary regressors.

In all cases, we operate within a parametric setting which implies that the vector of innovation sequences generating the predictive regression has a well-defined density that follows a Gaussian random variable with prespecified moments. In terms of the nonstationarity regimes, in Chapter 2 and 3 of the thesis we develop structural break testing procedures for the linear and quantile predictive regression models, focusing only on the cases of near unit root and mildly integrated processes. In both cases, the Wald-type test statistics, which are the proposed structural break detectors, are assumed to have certain optimality properties such as a small probability of a false alarm and asymptotic power one, regardless of the abstract degree of persistence. In Chapter 4, we consider nonstationary processes using the local-to-unity parametrization but without extending to the case of mildly explosive processes. Specifically, due to the Seemingly Unrelated Regression representation for the system of quantile predictive regressions, these node-specific equations incorporate generated covariates that represent the presence of systemic risk, but under the assumption that these block of regressors have the same persistence properties. Overall, the asymptotic theory developed as well empirical and simulation results presented in this thesis, contribute to the literature of robust methods for estimation and inference in predictive regression models. These contributions are threefold and are summarized below.

1.3.1 Chapter 2: Predictability Testing Robust against Parameter Instability

In Chapter 2, we consider the problem of joint predictability and structural break testing which is a nonstandard econometric problem since the nuisance parameter of persistence c_i is present both under the null as well as the alternative while the innovations of the regressand and those of the regressors are correlated. The literature proposes various methodologies for handling the nuisance parameter of persistence in estimation and inference problems. In this chapter we investigate suitable testing methodologies that accommodate these features when identifying structural breaks in linear predictive regressions.

The first contribution to the literature is a structural break testing framework for predictive regression models with persistent regressors. We propose Wald-type test statistics based on the ordinary least squares estimator as well as the endogenous IVX instrumentation proposed by Phillips and Magdalinos (2009). We establish the limiting distributions of these test statistics and study suitable bootstrap-based inference methodologies for obtaining critical values. We find that the degree of persistence in regressors can affect the asymptotic theory of the tests. Specifically, under high persistence regardless of the chosen estimator the limiting distributions are nonstandard and nonpivotal, while under mildly integratedness both test statistics weakly converge to the conventional nuisance-parameter free limiting distribution. Our framework is extended to the case when testing for joint predictability and parameter instability in linear predictive regressions and we demonstrate that the asymptotic behaviour of test statistics is sensitivity to the stability of model intercepts.

1.3.2 Chapter 3: Detecting Structural Breaks in Quantile Predictive Regressions

In Chapter 3, we consider the problem of break detection in nonstationary quantile predictive regression models. We establish the limit distributions for a class of Wald and fluctuation type statistics based on both the ordinary least squares estimator and the endogenous instrumental regression estimator proposed by Phillips and Magdalinos (2009). Although the asymptotic distribution of these test statistics appear to depend on the chosen estimator, the IVX based tests are shown to be asymptotically nuisance parameter-free when regressors exhibit mildly integratedness. The finite-sample performance of both tests is evaluated via simulation experiments. An empirical application to house pricing index returns demonstrates the practicality of the proposed break tests for regression quantiles of nonstationary time series data.

The second contribution to the literature is the development of an econometric environment for break-point detection in nonstationary quantile time series models, that is, predictive regression models estimated using a conditional quantile specification functional form. Specifically, we focus on constructing Wald-type statistics for testing for the presence of parameter instability at an unknown break-point location, given some fixed quantile level within a compact set. Furthermore, by extending the scope of conditional quantile specifications to multiple quantile levels, we provide a novel structural break testing framework in quantile predictive regression models under high persistence. Similar to the test statistics for break detection in linear predictive regression models, we find that due to the presence of the unidentified break-point parameter under the null hypothesis, the limiting distributions have a discontinuity with respect to the parameter space of the nuisance parameter of persistence which captures the nonstationary properties of regressors.

1.3.3 Chapter 4: Estimation and Inference in Systems of Quantile Predictive Regressions with Generated Regressors

In Chapter 4, we propose a robust Wald test for Quantile Predictive Regression Systems with generated regressors. These generated regressors are estimated based on nodewise nonstationary quantile predictive regression models. Then, the proposed robust Wald test is constructed based on adding-up restrictions of the parameters across the system of quantile regressions. We demonstrate that the asymptotic behaviour of the test is asymptotically distribution-free, that is, no nuisance parameters are involved in the derived limit of the test statistic which weakly converges to a chi-squared limiting distribution. Moreover, we provide several examples and finite sample simulation experiments to demonstrate the relevance of the test for certain parameter restrictions of the system which are particularly useful when modelling systemic risk.

The third contribution to the literature is that we propose a framework for estimation and inference methods in quantile predictive regression systems when the set of regressors include both nonstationary regressors as well as a generated regressor to estimate the Value-at-Risk and the Conditional-Value-at-Risk respectively, based on the local-to-unity specification that we impose on the autoregressive equation that models predictors. Therefore, since our aim is to model a financial network the proposed modelling environment is a novel contribution to the literature of statistical inference methods for Seemingly Unrelated Regression (SUR) systems under the assumption of regressor's nonstationarity, specifically when individual equations are quantile predictive regression models under a more general dependence structure. Consequently, our estimation methodology has important applications when modelling and testing for systemic risk in financial networks using risk measures (see, Adrian and Brunnermeier (2016)).

Chapter 2

Predictability Testing Robust against Parameter Instability

Abstract

In this Chapter we establish the limiting distributions of Wald-based statistics when testing for a structural break in parameters of predictive regressions and illustrate that these have a discontinuity for certain degrees of regressors persistence regardless of the chosen model estimator. The test statistics based on both the OLS and the IVX estimators converge to a nonstandard and nonpivotal limiting distribution when the break-point is unknown and regressors exhibit nearly integrated or high persistent behaviour. A nuisance-parameter free distribution under the null hypothesis holds when regressors are mildly integrated or stationary. Furthermore, we consider the corresponding asymptotic theory results when testing for joint predictability and parameter instability in predictive regression models. We compare the finite-sample size performance of the proposed tests based on both estimators via simulation experiments. Critical values in the cases of nonstandard limiting distributions are obtained using bootstrap approximations.

2.1 Introduction

Predictive regression models are commonly used in financial econometrics and empirical finance when assessing time series predictability especially with regressors of unknown integration order. The development of methodologies for detecting the presence of parameter instability as well as stock return predictability is an important research question in the financial and time series econometrics literature. A related aspect is the predictability of indices such as equity premiums using predictive regression models with predictors macroeconomic and financial variables that has been found to be countercyclical. On one hand, a stream of literature focuses on identifying these periods of episodic predictability (see, Gonzalo and Pitarakis (2012, 2017), Chinco et al. (2019), Demetrescu et al. (2020)). On the other hand, a different stream of literature argues, that these periods of unstable predictability appear in time series models in the form of parameter instability (see, Rossi and Inoue (2012), Inoue et al. (2017), Pitarakis (2017) and Georgiev et al. (2018)). The predictive regression model operates under the strong assumption of parameter stability, which can be violated in certain regions of the sample. A related study to parameter instability in prediction models is presented by Paye and Timmermann (2006). However, the majority of the structural break literature propose testing methodologies under the assumption of stationary time series. Therefore, we focus on developing an econometric environment for predictability testing robust to parameter instability in predictive regressions under the presence of persistence regressors when the break-point is at an unknown location within the full sample. Consider the pair $(y_t, \boldsymbol{x}_{t-1})_{t\in\mathbb{Z}}$ with an underline martingale difference sequence $w_t = (u_t, v_t)'$ such that $(y_t, \boldsymbol{x}_{t-1}, w_t)_{t\in\mathbb{Z}}$ is generated by a predictive regression model while the vector of regressors $\boldsymbol{x}_t$ is assumed to be an autoregressive process with an autocorrelation coefficient as a local unit root, it allows us to examine the persistence properties of regressors in a unified way. However, in most of the aforementioned frameworks, estimation and inference in predictive regression models operates under the assumption of parameter constancy throughout the sample. In this paper, we consider structural break tests for predictive regression models in which the regressors are assumed to follow stable autoregressive processes. The stability of the autoregressive processes is determined by the parameters utilized in the local to unity specification of the autocorrelation coefficient. Specifically, we focus on autoregressive processes which are close to the unit boundary but have different order of convergence, namely high persistent regressors which are $\mathcal{O}(n^{-1/2})$ and mildly integrated regressors which are $\mathcal{O}(n^{-\gamma_x/2})$, where $\gamma_x \in (0,1)$ denotes the exponent rate of persistence and $c_i > 0$ is a positive persistence coefficient for all stochastic regressors. Specifically, to measure the degree of persistence in time series we follow the persistence classification in Kostakis et al. (2015). Furthermore, conventional structural break tests for the parameters of linear regression models employ the widely used sup Wald test proposed by Andrews (1993). However, the distributional theory of the Andrews's test depends on the strict stationarity assumption of regressors. On the other hand, the predictive regression model is usually fitted to economic datasets which contain time series that are highly persistent. Thus, within such econometric environment the traditional law of large numbers and central limit theorems can invalidate the standard econometric assumptions of linear regression models, which affect the large sample approximations. As a result distorted inferences can occur when testing for parameter instability in predictive regressions when these features are not accommodated in the asymptotic theory of the test statistics, especially under the presence of persistent regressors.

Our first objective is to theoretically demonstrate the impact of the nuisance parameter of persistence on the statistical validity of weakly convergence arguments to a Brownian Bridge process as in the case of Andrews sup Wald OLS based tests. Specifically, the particular limit result implies the use of the supremum functional on the Brownian Bridge process defined as $\sup_{s\in[0,1]}\left[W_n(s) - sW_n(1)\right]$. Under weak convergence, we have process convergence where B is a Brownian motion, that is, a pivot process, hence enabling the practitioner to use standard tabulated critical values. Our theoretical results show that the standard NBB asymptotic approximations of the sup OLS-Wald tests in the presence of nearly integrated regressors holds while under high persistence the same limit is no longer valid (see, also Katsouris (2022)).

Our second objective is to develop an instrumental variable based modification of the sup OLS-Wald statistic proposed by Andrews (1993), whose asymptotics are robust to the persistence properties of regressors. To do this, we employ the IVX filtration proposed by Phillips and Magdalinos (2009) and also extensively examined by Kostakis et al. (2015) in the context of predictability tests and by Gonzalo and Pitarakis (2012) in the context of predictability tests for threshold predictive regression models[1]. Due to the presence of the supremum functional when testing for structural break at an unknown break-point location both the persistence properties of regressors and the type of the Wald test can affect the distributional theory.

We study the statistical inference problem of structural break testing at an unknown break-point therefore the supremum functional is implemented and two test statistics are considered based on two different parameter estimation methods. The first estimation method considers the OLS estimator, while the second method considers an instrumental variable based estimator, namely the IVX estimator proposed by Phillips and Magdalinos (2009). These two estimators of the predictive regression model have different finite-sample and asymptotic properties, which allows us to compare the limiting distributions of the proposed tests for the two different types of persistence of the regressors. Furthermore, for both test statistics we assume that the regressors included in the model are permitted to be only one of the two persistence types which simplifies the asymptotic theory of the tests, however the presence of nuisance parameters under the null of parameter constancy, that is, the unknown break-fraction and the coefficient of persistence, requires careful examination of the asymptotic theory. An additional caveat is the inclusion of an intercept in the predictive regression model which induces different limiting distributions when is assumed to be stable vis-a-vis the case in which is permitted to shift. In summary, our main goal in this paper is the identification of the limit behaviour of Wald type tests when testing for structural breaks in predictive regression models based on the OLS vis-a-vis the IVX estimator.

Therefore, our asymptotic theory holds due to the invariance principle of the partial sum process of x_t, where $x_t = \left(1 - \frac{c}{n^{\gamma_x}}\right) x_{t-1}$, as proposed by Phillips (1987a). Furthermore, we denote with $\hat{\mathcal{U}}_T(s) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} x_t$, for some $s \in [0,1]$ and with $\hat{\mathcal{U}}_{n^{\gamma_x}}(s) := \frac{1}{n^{\gamma_x/2}} \sum_{t=1}^{\lfloor n^{\gamma_x} s \rfloor} x_t$, for some $s \in [0,1]$ and $0 < \gamma_x < 1$ for the invariance principle of the partial sum process of x_t in the case of mildly integrated processes, for the corresponding limit as proposed by Phillips and Magdalinos (2007).

[1] More specifically, the limit theory of Kostakis et al. (2015) provides a unified framework for robust inference and testing regardless the persistence properties of regressors. A simple example is the application of the IVX-Wald test for inferring the individual statistical significance of predictors under abstract degree of persistence. Further scenarios such as predictors of mixed integration order see Phillips and Lee (2013) and Phillips and Lee (2016) in which cases the mixed normality assumption still holds.

Similarly in this chapter we employ the invariance principles for the partial sums processes that correspond to the instrumental variable, IVX, proposed by Phillips and Magdalinos (2009), specifically within a structural break testing framework. Therefore, these results allow us to formally obtain the limiting distributions of the proposed tests with respect to the nuisance parameter of persistence along with the unknown break-point location, and observe in which cases we obtain nuisance-free inference that can simplify significantly the hypothesis testing procedure.

In summary, we propose an econometric framework for jointly testing against both predictability and structural break. The tests are constructed in a similar manner as the Wald type statistics proposed by Pitarakis (2014) and Gonzalo and Pitarakis (2012) in a threshold predictive regression model[2]. Our contributions are threefold: *(i)* We propose a test statistic which jointly tests against the alternative hypothesis of both predictability and structural break and show that the test is robust to the persistence properties of predictors in the model; *(ii)* We show that the test statistic has a nuisance parameter free limiting distribution under the assumption of stationary and mildly stationary regressors, while it converges to a nonpivotal asymptotic distribution under the assumption of local-to-unit root (LUR) regressors; *(iii)* We examine the finite-sample performance of the tests with Monte Carlo simulations where we obtain the empirical size the sup OLS-Wald and IVX-Wald tests. Lastly, we employ the proposed structural break testing framework to investigate the predictability of the equity premium based on US stock returns.

2.1.1 Literature Review

The ideas in this paper are related to research done in two different fields. From finance, it is related to the stock return predictability literature and from the econometrics and statistics perspective it is related to the literature of parameter instability and structural break testing methodologies. Therefore, our aim is to bridge the gap in the literature by developing an econometric environment for testing the presence of joint predictability and parameter instability.

Firstly, the implementation of the Student's t-test is commonly employed to detect statistical significance in stable relations of predictants such as equity index returns, on the lagged time series of predictors. From extensive empirical applications, this practise has been proved to cause distorted inference due to the presence of high persistent predictors, since nonstandard terms appear in the limiting distribution of the t-test. In particular, Stambaugh (1999) observed this finite sample bias[3] which occurs when the classical least squares estimator is employed for statistical inference. Furthermore, Amihud and Hurvich (2004) consider a second-order bias-correction and propose a reduced-bias OLS based estimator. Both aforementioned methods are considered to provide a post-estimation bias correction in finite samples.

Secondly, the predictability literature was extended to nonstandard inference problems to account for the presence of nonstationary predictors; some of the most notable contributions include the Bonferroni-type approach as in Cavanagh et al. (1995) and Campbell and Yogo (2006), the control function method proposed by Elliott (2011) as well as the conditional likelihood method using

[2] Gonzalo and Pitarakis (2017) propose tests which capture the effects of linearity and the presence of a threshold effect in predictive regressions with persistent predictors to test for episodic predictability.

[3] Specifically, the Stambaugh bias correction is based on the studies of Marriott and Pope (1954) and Kendall (1954) who proposed suitable bias correction in autocorrelations.

the framework of sufficient statistics[4] proposed by Jansson and Moreira (2006). However, the drawback of the aforementioned testing methodologies is that the asymptotic theory has some undesirable properties such as the non-correctable bias due to the presence of the nuisance parameter of persistence that appears in the limiting distribution of test statistics (see, Phillips and Lee (2013)). Furthermore, these testing approaches can be computational intense especially in multivariate predictive regression settings without ensuring that robust inference to the unknown parameter of persistence is guaranteed either in finite-samples or asymptotically.

Thirdly, a novel approach that recently has been attracting much attention is the instrumental variable-based test statistic proposed by Kostakis et al. (2015), (KMS, hereafter), which is build upon the theoretical framework developed by Phillips and Magdalinos (2009). This methodology, referred to as IVX-Wald test, provides a robust framework for predictive regression models which is valid for predictors with general persistence properties. Specifically, the asymptotic theory shows that the IVX estimator which converges to a mixed Gaussian distribution, successfully removes the long-run endogeneity, that appears due to the innovation structure of the model, and provides a pivotal statistic robust under different degrees of persistence or even regressors of mixed integration (e.g., see Phillips and Lee (2016)). Hence, a self-normalized Wald statistic can be constructed that converges to a nuisance parameter free χ^2 limiting distribution. More importantly, the IVX filtration implies a direct inference procedure via the various moment approximations (e.g., long-run covariance matrices) and can be easily extended to the multivariate predictive regression model under certain regulatory conditions.

An important assumption for the previously mentioned methodologies developed in the literature is the condition of parameter constancy[5] which implies a stable predictive relationship over the sample period. However, due to the nature of economic conditions, shifting between periods of market tranquillity and periods of market exuberance, the phenomenon of episodic predictability[6] has been proposed to capture these "pockets" of predictability across business cycles (see, Farmer et al. (2019)). This implies the existence of time-varying predictability which can be examined within an econometric framework which accommodates time-varying parameters. In this chapter, we approach this aspect by proposing a testing methodology for the existence of joint predictability and parameter stability, in the form of a single parameter shift at an unknown break-point location. Our approach is closer to the framework of Andrews (1993) who proposed Wald-type statistics for testing for a single structural break at an unknown break-point. The importance of having an inference methodology for predictability testing robust against parameter instability is shown both in empirical studies as well as in simulation experiments, where under the presence of a structural-break at an unknown location, model estimates and test statistics are affected.

[4] *Sufficient statistics* are developed as optimal tests invariant under transformation based on the curved exponential family. The framework proposed by Jansson and Moreira (2006) implies the use of conditional restrictions testing in the presence of nuisance parameters and is particularly appealing in the case of near integrated regressors.

[5] The literature of structural change goes back to 1940s and 1950s with the pioneering work of Wald on sequential hypothesis testing as well as the seminal work of Page (1954), who propose methods for detecting anomalies in control charting, an idea further developed by Chu et al. (1995) and Chu et al. (1996) who consider testing for structural break in the sense of contaminated and non-contaminated periods in linear regression models. Further seminal studies for structural break tests include Chow (1960), Hawkins (1987) and Ploberger and Krämer (1990).

[6] Testing methods for stock return predictability are proposed by Lanne (2002), Guo (2006), Gonzalo and Pitarakis (2012, 2017), Harvey et al. (2021a), Demetrescu et al. (2022b), Demetrescu et al. (2022a), Farmer et al. (2023) and Tu and Xie (2023), which includes the so-called episodic, pocket, subsample, or sporadic predictability.

The presence of parameter instability in predictive regression models implies variation in predictability across time. Consequently, standard testing methodologies for break detection, such as Andrew's Wald statistics for linear regression models as well as the family of tests proposed by Bai and Perron (1998) can no longer be valid for nonstationary[7] predictors as captured by the autoregressive specification of the predictive regression model. This invalidity is demonstrated by the simulation studies of Paye and Timmermann (2006). More specifically, the authors show that in the case of highly correlated innovations and (unfiltered) persistent predictors, an implementation of a sup-F test and UDMax test can cause severe size distortions when testing for structural breaks in time series. Moreover, the optimal test of Elliot and Muller has good empirical size performance, but distortions appear in the asymptotic properties of the power function of the test. The literature of structural break tests has recently focused in the construction of suitable testing frameworks for predictive regression models. Cai et al. (2015) propose modelling smooth structural breaks using an L_2- type statistic, in predictive regressions with nonstationary predictors. A different approach include the study of Pitarakis (2017) who consider a CUSUM-type statistic. Additionally, due to the occurrence of nonstandard limiting distributions in standard structural break test implementations, Georgiev et al. (2018) and Georgiev et al. (2019) propose the use of a fixed regressor wild bootstrap procedure[8] to approximate critical values. Another related study to the bootstrapping approach of structural break tests is presented by Boldea et al. (2019). However, the proposed framework is restricted to the case of exogenous regressors.

Despite the recent developments of the structural break literature for predictive regressions, these methodologies consider the detection of parameter instability in predictive regression models without simultaneously testing whether the tests are robust to the presence of predictability around the break-point location. The only existing study that jointly tests for the existence of both effects is the paper of Demetrescu et al. (2020) (see, also Andersen and Varneskov (2021) and Andersen and Varneskov (2022)), who propose to combine testing for predictability using appropriate subsampling techniques[9](see also similar techniques employed by Hansen (2000a)). The authors propose to use bootstrap-based inference due to the existence of a non-pivotal limiting distribution, which is a robust approach to provide statistical validity of the tests. Further aspects related to the IV based approach of KMS that we follow in this paper, can be also examined within the above parameter instability testing procedures. Recent applications include Magdalinos (2020) who consider predictability tests with GARCH-type effects (see also, Gungor and Luger (2020)) as well as Yang et al. (2020) who consider a modification of the KMS test that accounts for serial correlation in the error term of the linear predictive regression. Moreover, Pang et al. (2020) propose testing methodologies for multiple structural breaks under the presence of nonstationary predictors with an application to financial bubble detection. All these features are further refinements one can consider as future research related to our proposed framework.

[7]Specifically, Phillips and Magdalinos (2009) explicitly examine the various forms of nonstationarity attributed due to the structure of the predictive and cointegrated system, as captured by specific characteristics of the system. This is a major distinction in the literature which previously thought that the presence of nonstationarity in the regressors as a characteristic occurring due to the autoregressive equation of predictors solely.

[8]The fixed regressor wild bootstrap is an extension to the fixed regressor bootstrap employed by Hansen (2000b), to deal with the presence of a nonstandard limiting distribution due to nonstationary regressors.

[9]The subsampling technique is similarly employed by Davidson and Monticini (2010) in the context of detecting structural break due to break in cointegration of the relation under investigation.

The chapter is organized as follows. Section 2.2, presents the predictive regression model along with the background assumptions. This Section also includes a review of the IVX instrumentation procedure of KMS, which is employed for the construction of the proposed test statistic. Section 2.4, presents the asymptotic theory of the test statistic under different degrees of regressors persistence. Section 2.5 presents an extensive Monte Carlo simulation study. Section 2.6 illustrates an application to the US stock returns which provides evidence of the empirical relevance of the proposed tests for jointly testing predictability and structural break. Section 2.7 concludes.

2.2 Model Estimation and Assumptions

Consider the linear predictive regression model with a possible single structural break

$$y_t = \mu_t + \boldsymbol{\beta}_t' \boldsymbol{x}_{t-1} + u_t, \qquad \text{with} \;\; t \in \{1, ..., n\}\,, \tag{2.1}$$

$$\boldsymbol{x}_t = \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}_p}{n^{\gamma_x}} \right) \boldsymbol{x}_{t-1} + \boldsymbol{v}_t, \tag{2.2}$$

where $y_t \in \mathbb{R}$ is an 1−dimensional vector and $\boldsymbol{x}_t \in \mathbb{R}^{p \times n}$ is a p−dimensional vector of local unit root regressors, with an initial condition $\boldsymbol{x}_0 = 0$. Moreover, $\boldsymbol{C} = \mathsf{diag}\{c_1, ..., c_p\}$ is a $p \times p$ diagonal matrix which determines the degree of persistence of the regressors by the unknown persistence coefficients c_i's which are assumed to be positive constants. Furthermore, we consider that the exponent rate, γ_x, takes the following values:

- $\gamma_x = 1$, the case of local-unit-root (LUR) regressors.
- $\gamma_x \in (0, 1)$, the case of mildly integrated (MI) regressors.
- $\gamma_x = 0$, the case of stationary regressors.

The above definitions can be found in Phillips and Magdalinos (2009) and Kostakis et al. (2015). Notice that in the case of stationary predictors, i.e., $\gamma_x = 0$, this class of persistence holds for a suitably restricted coefficient matrix $(\boldsymbol{I}_p - \boldsymbol{C})$. Moreover, the predictive regression model (4.11)-(4.12) accommodates the existence of a single structural break at unknown location k. The model parameters take the following form indicating a time-varying parameter vector

$$\mu_t = \begin{cases} \mu_1, & t \leq k \\ \mu_1, & t > k \end{cases} \quad \text{and} \quad \boldsymbol{\beta}_t = \begin{cases} \boldsymbol{\beta}_1, & t \leq k \\ \boldsymbol{\beta}_2, & t > k \end{cases}, \quad s.t \;\; k = \lfloor n\pi \rfloor \tag{2.3}$$

where π denotes the first π fraction of the sample such that $\pi \in (0, 1)$ while the model parameters are denoted by $(\mu_1, \mu_2) \in \mathbb{R}$ and $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \in \mathbb{R}^{p \times 1}$. Notice that the introduction of the notation $(\mu_t, \boldsymbol{\beta}_t)$ implies that we consider predictive regression models with structural break type time-variation in the parameters of the predictive regression.

Let $\mathcal{F}_t$ denote the natural filtration, then for the error term of the predictive regression we assume that $\mathbb{E}\left(u_t | \mathcal{F}_{t-1}\right) = 0$ and $\mathbb{E}\left(u_t^2 | \mathcal{F}_{t-1}\right) = \sigma_{uu}^2$. Specifically, the innovation structure of the predictive regression model allows to impose a linear process dependence for $\boldsymbol{v}_t$, with a conditionally homoscedastic *martingale difference sequence* condition such that

$$\boldsymbol{v}_t = \sum_{j=0}^{\infty} \boldsymbol{\varphi}_j \boldsymbol{\varepsilon}_{t-j}, \quad \boldsymbol{\varepsilon}_t \sim mds\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \tag{2.4}$$

The particular error sequence representation for the autoregressive specification of the regressors that correspond to the predictive regression model is proposed by the seminal study of Phillips and Solo (1992). Specifically, the given representation is based on a simply polynomial decomposition such that $\boldsymbol{v}_t = \sum_{j=0}^{\infty} \boldsymbol{\varphi}_j \boldsymbol{\varepsilon}_{t-j}$ where $\boldsymbol{\varphi}_o(\mathsf{z}) = \sum_{j=0}^{\infty} \boldsymbol{\varphi}_j \mathsf{z}^j$.

Thus, the function $\boldsymbol{\varphi}_o(\mathsf{z})$ represents a power series such that $\sum_{j=0}^{\infty} \boldsymbol{\varphi}_j \mathsf{z}^j$ has a radius of convergence $\varrho \in [0, +\infty]$ which implies that $\boldsymbol{\varphi}_o(\mathsf{z})$ converges absolutely for $|z| < \varrho$ and does not converge for $|z| > \varrho$. Therefore, a uniform convergence result follows under regularity conditions on the radius of convergence which in practise can be accommodated by the summability condition of the positive coefficients $\boldsymbol{\varphi}_j$ such that $\sum_{j=0}^{\infty} \boldsymbol{\varphi}_j < \infty$.

Therefore, representing the innovation sequence of the autoregressive model as a linear process it permits to express the regression model $(\boldsymbol{x}_t - \boldsymbol{x}_{t-1}) = \sum_{j=0}^{\infty} \boldsymbol{\varphi}_j \boldsymbol{\varepsilon}_{t-j}$, where $\boldsymbol{\varepsilon}_t$ is a $m.d.s$ and $\sum_{j=0}^{\infty} j \left\| \boldsymbol{\varphi}_j \right\| < \infty$ such that $\left\| \boldsymbol{\varphi}_j \right\| = \mathsf{trace}\left[(\boldsymbol{\varphi}_j' \boldsymbol{\varphi}_j) \right]^{1/2}$ which results to a stationary process. The aforementioned conjecture was originated from the theory of cointegration; however when the autoregressive specification is expressed in the form of a near unit root process, such that $\boldsymbol{x}_t = \left(\boldsymbol{I}_p + \frac{C}{n^{\gamma_x}} \right) \boldsymbol{x}_{t-1}$ then a more sophisticated method is required to convert the nonstationarity property of regressors induced by the nuisance parameter of persistence. For this reason, the IVX filtration proposed by Phillips and Magdalinos (2009) can be employed to obtain mildly integrated processes. Our goal is to statistical infer on the stability of the model parameters regardless of the time series properties of the predictor vector x_t as captured by the unknown persistence coefficients c_i's and the exponent rate γ_x, such that $\gamma_x = 0$, $\gamma_x \in (0, 1)$ or $\gamma_x = 1$. More specifically, parameter instability is captured by the assumed time-variation in the vector (α_t, β_t), while the different types of persistence of predictors is modelled by the LUR specification. Therefore, for the development of the asymptotic theory of the structural break tests we impose the following regularity conditions given by Assumption 2.1 below.

Assumption 2.1. Let $\boldsymbol{e}_t = (u_t, \boldsymbol{\epsilon}_t')'$ be a $(p+1)-$dimensional vector. The innovation sequence e_t is a conditionally homoscedastic *martingale difference sequence* (m.d.s) such that the following conditions hold:

A1. $\mathbb{E}\left[\boldsymbol{e}_t | \mathcal{F}_{t-1} \right] = 0$, where $\mathcal{F}_t = \sigma\left(\boldsymbol{e}_t, \boldsymbol{e}_{t-1}, ... \right)$ is an increasing sequence of $\sigma-$fields.

A2. $\mathbb{E}\left[\boldsymbol{e}_t \boldsymbol{e}_t' | \mathcal{F}_{t-1} \right] = \boldsymbol{\Sigma}_{ee}$, where $\boldsymbol{\Sigma}_{ee} \in \mathbb{R}^{(p+1)\times(p+1)}$ is a positive-definite covariance matrix, which has the following form:

$$\boldsymbol{\Sigma}_{ee} = \begin{bmatrix} \sigma_u^2 & \sigma_{uv}' \\ \sigma_{vu} & \boldsymbol{\Sigma}_{vv} \end{bmatrix} > 0.$$

with $\sigma_u^2 \in \mathbb{R}$, $\sigma_{uv} \in \mathbb{R}^{p \times 1}$ and $\boldsymbol{\Sigma}_{vv} \in \mathbb{R}^{p \times p}$, where p is the number of predictors.

A3. The innovation to x_t is a linear process with the representation below

$$\boldsymbol{v}_t := \boldsymbol{\Phi}(L)\boldsymbol{\epsilon}_t \equiv \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j \boldsymbol{\epsilon}_{t-j}, \quad \boldsymbol{\epsilon}_t \sim^{i.i.d} (0, \boldsymbol{\Sigma}_{ee})$$

where $\{\boldsymbol{\Phi}\}_{j=0}^{\infty}$ is a sequence of absolute summable constant matrices such that $\sum_{j=0}^{\infty} \boldsymbol{\Phi}_j$ has full rank and $\boldsymbol{\Phi}_0 = \boldsymbol{I}_p$ with $\boldsymbol{\Phi}(1) \neq 0$, allowing for the presence of serial correlation.

Under conditions (A1)-(A3), the following Functional Central Limit Theorem ($FCLT$) applies

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} \boldsymbol{w}_t = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} \begin{bmatrix} u_t \\ \boldsymbol{v}_t' \end{bmatrix} \Rightarrow \text{ BM}\left(\boldsymbol{\Sigma}_w\right) = \begin{bmatrix} B_u(s) \\ B_v(s) \end{bmatrix} \equiv \boldsymbol{\Sigma}_w^{1/2} \boldsymbol{W}(s), \tag{2.5}$$

where $\boldsymbol{w}_t = (u_t, \boldsymbol{v}_t')'$ and $\{\boldsymbol{W}(s) : 0 \leq s \leq 1\}$ is the standard vector-valued Brownian motion, in the space of $\mathcal{D}([0,1])$, that is, the Skorokhod topology within which joint weak convergence arguments hold. Thus, expression (2.5) represents an invariance principle for the partial sum process of $\boldsymbol{w}_t$, which implies that the partial sum of $\boldsymbol{w}_t$, weakly converges to the stochastic quantity $\boldsymbol{\Sigma}_w^{1/2} \boldsymbol{W}(s)$, that is, a $(p+1)-$dimensional Brownian process with covariance matrix $\boldsymbol{\Sigma}_w$.

2.2.1 Preliminaries

Our main focus for the asymptotic theory are the following partial sum processes:

$$\hat{\mathcal{U}}_n(s) := \frac{1}{n^{\frac{1}{2}}} \sum_{t=1}^{\lfloor ns \rfloor} X_t \Rightarrow K_c(s), \quad s \in [0,1] \quad \text{when } \gamma_x = 1 \text{ in (4.12)}, \tag{2.6}$$

$$\hat{\mathcal{U}}_{n^{\gamma}}(s) := \frac{1}{n^{\frac{\gamma_x}{2}}} \sum_{t=1}^{\lfloor n^{\gamma_x} s \rfloor} X_t \Rightarrow J_c(s), \quad s \in [0,1] \quad \text{when } \gamma_x \in (0,1) \text{ in (4.12)}. \tag{2.7}$$

Corollary 2.1. Suppose Assumption 2.1 holds. If $\sup_{t \in \mathbb{Z}} \mathbb{E}\left[\epsilon_t^2\right] < \infty$, then for each $\gamma_x \in (0,1)$ and $c_i > 0$ for all $i \in \{1, ..., p\}$ it holds that

$$\sup_{s \in [0, n^{1-\gamma_x}]} \left| \, \hat{\mathcal{U}}_{n^{\gamma_x}}(s) - \mathcal{U}_{T^{\gamma_x}}(s) \, \right| = o_p(1) \quad \text{as } n \to \infty \tag{2.8}$$

where

$$\mathcal{U}_{n^{\gamma_x}}(s) := \int_0^s e^{(s-r)C} dW_{n^{\gamma_x}}(r) \tag{2.9}$$

In a shorter notation, we denote $x_{[nr]}/\sqrt{n} \Rightarrow \boldsymbol{K}_c(r)$ which represents the local-unit-root asymptotics proposed Phillips (1987a). Notice that when $\gamma_x \in (0,1)$ then suitable normalization are necessary to obtain an equivalent asymptotic limit as proposed by Phillips and Magdalinos (2007). Notice that, $\boldsymbol{K}_c(r)$ denotes a $p-$dimensional Gaussian process defined as below

$$\boldsymbol{K}_c(r) = \int_0^r e^{(r-s)C} d\boldsymbol{B}_v(s), \quad r \in (0,1). \tag{2.10}$$

that satisfies the Black-Scholes differential equation $d\boldsymbol{K}_c(r) \equiv c\boldsymbol{K}_c(r) + d\boldsymbol{B}_v(r)$, with $\boldsymbol{K}_c(r) = 0$, implying also that $\boldsymbol{K}_c(r) \equiv \sigma_v \boldsymbol{J}_c(r)$, where $\boldsymbol{J}_c(r) = \int_0^r e^{(r-s)C} d\boldsymbol{W}_v(s)$. Notice that $\boldsymbol{K}_c(r)$ represents the Ornstein-Uhlenbeck[10], (OU) process, which encompasses the unit root case such that $\boldsymbol{J}_c(r) \equiv \boldsymbol{B}_v(r)$, for $\boldsymbol{C} = 0$. Moreover, the assumption of a local-unit-root specification for the predictors of the model and more specifically by allowing the autocorrelation coefficient to be of the form $\boldsymbol{R}_n = \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}}{n^{\gamma}} \right)$, permits to consider the $\boldsymbol{K}_c(r)$ Gaussian process given by (2.10) as a stochastic integral approximation. The underline error structure introduced by Assumption 2.1, provides a realistic interpretation of macroeconomic shocks. More specifically, the shocks to y_t and $\boldsymbol{x}_{t-1}$, given by the sequences u_t and v_t respectively, for $t \in \{1, ..., n\}$, appear to be contemporaneously correlated, a commonly used assumption in predictive regression models. Related definitions can be found in the seminal papers of Phillips and Durlauf (1986) and Phillips (1987a). By imposing a MDS assumption on the disturbance sequence $(u_t, \boldsymbol{\epsilon}_t')'$ of the predictive regression, it follows that $\mathbb{E}\left[u_t \boldsymbol{\epsilon}_{t-k}\right] = 0$, for all $k \geq 1$ (see, Gonzalo and Pitarakis (2012)), which eliminates any covariance terms of the error term at time t with past innovations. Then, since the *FCLT*[11] given by (2.5) holds within our context, it allow us to derive the limiting distribution of Wald type statistics when testing for parameter instability in predictive regression models with persistent predictors, as functionals of the particular stochastic processes.

Assumption A3 gives the linear process representation of innovations proposed by Phillips and Solo (1992). The particular assumption regarding the innovation sequence of the predictive regression model, allows to impose further econometric conditions regarding how the shocks to the predictors of the model are generated, considering this way the case of shocks which can be serially correlated and heteroscedastic. This is an important result worth emphasizing, since although we permit these features for the shocks, the conditionally homoscedastic martingale difference assumption for the error term u_t that correspond to the predictive regression model, holds. This assumption can be further relaxed by introducing the conditionally heteroscedasticity for the variance of u_t (see, Kostakis et al. (2015)). Our modelling framework given by (4.11)-(4.12) and (2.3) encompasses the linear predictive regression model. Under the null hypothesis of parameter constancy, which is equivalent to the case when the time-varying parameters μ_t and β_t are both constant over the full sample, implies that $\mu_1 = \mu_2$ and $\beta_1 = \beta_2$, and the regression specification reduces to the standard predictive regression model. Moreover, our research interest in this paper is the detection of a parameter instability in the predictive regression model at a single break-point under the alternative hypothesis, where we denote with $k = \lfloor n\pi \rfloor$, the unknown break-point. We consider structural breaks in the coefficients of the predictive regression, assuming that the variance terms of the error sequence, (σ_u, σ_v) remains constant over the sample which also implies the FCLT to hold with no additional refinements. Furthermore, we consider related econometric aspects which arise within our context, such as the use of an appropriate bias correction to tackle the endogeneity bias under the presence of these two regimes[12].

[10] The OU process is a stationary Gaussian process with an autocorrelation function that decays exponentially over time. Moreover, the continuous time OU diffusion process has a unique solution and this property allows to approximate asymptotic terms which appear in the various expressions of the estimators and the corresponding test statistics as a function of the LUR parameters.

[11] A standard *FCLT* for linear regression models is introduced by Theorem 7.17 in White (2001).

[12] Notice that the underline regimes of our framework are motivated by the hypothesis of parameter instability in the predictive regression model rather than the presence of an independent threshold variable which induces the presence of the two regimes, as in Gonzalo and Pitarakis (2012, 2017) (see, also Zhu et al. (2022)).

2.2.2 Robust Inference with the IVX instrumentation

Our main research objectives in this chapter is using the IVX instrumentation proposed by Phillips and Magdalinos (2009) in order to robustify structural break tests to the unknown persistence coefficient, as this is captured via the local-unit-root specification of stochastic regressors employed to the predictive regression model. To do this, we begin by reviewing the econometric implementation of the particular methodology for the linear predictive regression model, which is equivalent to the case when $\mu_1 = \mu_2$ and $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ as mentioned earlier. Notice that the IVX methodology can robustify inferences in terms of the unknown persistence properties of predictors.

Then, the tuning parameters, which are the exponent rate δ_z and the diagonal matrix $\boldsymbol{C}_z$ are selected to ensure that $\boldsymbol{z}_t$ is mildly integrated; less persistence than a unit root or a regressor assumed to be generated via a local-unit-root process. Then, considering the IVX estimator for the standard linear predictive regression model with no structural breaks and $\gamma_x = 1$ we obtain

$$\widetilde{\boldsymbol{\beta}}^{IVX} = \left[\sum_{t=1}^{n} (\boldsymbol{x}_t - \bar{\boldsymbol{x}}_{n-1}) \widetilde{\boldsymbol{z}}_{t-1}' \right]^{-1} \sum_{t=1}^{n} (y_t - \bar{y}_n) \widetilde{\boldsymbol{z}}_{t-1}', \tag{2.11}$$

where $\bar{\boldsymbol{x}}_{n-1} = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{x}_{t-1}$ and $\bar{y}_n = \frac{1}{n} \sum_{t=1}^{n} y_t$, are the corresponding sample means.

As shown by Theorem A in the Appendix of Kostakis et al. (2015), the IVX estimator converges to a mixed Gaussian[13] limiting distribution, which holds regardless of the degree of persistence of the regressors in the model. In turn, this property allows to construct a self-normalized Wald-type statistic which is shown to converge to a standard χ^2-distribution. The classical testing hypothesis implies that, $\mathbb{H}_0 : \boldsymbol{R\beta} = \boldsymbol{0}$ vs $\mathbb{H}_1 : \boldsymbol{R\beta} \neq \boldsymbol{0}$, where $\boldsymbol{R}$ the full rank $(q \times p)$ restriction matrix with rank q. Then, the IVX-Wald statistic can be used to test the null hypothesis of no predictability in predictive regression models. The IVX-Wald statistic is expressed as below

$$\mathcal{W}_n^{IVX} = \widetilde{\boldsymbol{\beta}}^{IVX\prime} \boldsymbol{Q}_{\mathcal{R}}^{-1} \widetilde{\boldsymbol{\beta}}^{IVX} \tag{2.12}$$

Notice that $\boldsymbol{Q}_{\mathcal{R}}$ is a consistent estimator of the asymptotic variance-covariance matrix of $\widetilde{\beta}^{IVX}$ that accommodates both long-run endogeneity caused by the correlation between the error terms of the system, that is, u_t and v_t, and the finite-sample distortion which results from removing the model intercept. The covariance matrix $\boldsymbol{Q}_{\mathcal{R}}$, is derived via the following fully modified (FM) estimation of the system covariance terms

$$\boldsymbol{Q}_{\mathcal{R}} = \left(\sum_{t=1}^{n} \widetilde{z}_{t-1} x_{t-1}' \right)^{-1} \boldsymbol{M} \left(\sum_{t=1}^{T} x_{t-1}' \widetilde{z}_{t-1} \right)^{-1}$$

$$\boldsymbol{M} = \widehat{\sigma}_u^2 \left(\sum_{t=1}^{n} \widetilde{z}_{t-1} \widetilde{z}_{t-1}' \right) - T \bar{z}_{n-1} \bar{z}_{n-1}' \widehat{\boldsymbol{\Omega}}_{FM}$$

where the fully modified covariance matrix is defined as

$$\widehat{\boldsymbol{\Omega}}_{FM} = \widehat{\boldsymbol{\Sigma}}_{uu} - \widehat{\boldsymbol{\Omega}}_{uv} \widehat{\boldsymbol{\Omega}}_{uu} \widehat{\boldsymbol{\Omega}}_{uv}', \quad \text{and} \quad \bar{z}_{n-1} = \frac{1}{n} \sum_{t=1}^{n} \widetilde{z}_{t-1}.$$

[13]The mixed Gaussianity property of the IVX estimator is also extensively examined in the papers of Phillips and Lee (2013) and Phillips and Lee (2016) under various integration orders.

The corresponding population covariance matrices are such that

$$\boldsymbol{\Sigma}_{uu} = \mathbb{E}\left(u_t^2\right), \quad \boldsymbol{\Sigma}_{uv} = \mathbb{E}\left(u_t v_t'\right), \quad \boldsymbol{\Sigma}_{vv} = \mathbb{E}\left(v_t v_t'\right). \tag{2.13}$$

$$\boldsymbol{\Omega}_{uu} = \sum_{h=-\infty}^{+\infty} \mathbb{E}\left(v_t v_{t-h}'\right), \quad \boldsymbol{\Omega}_{uu} = \boldsymbol{\Sigma}_{uv} + \boldsymbol{\Lambda}_{uv} + \boldsymbol{\Lambda}_{uv}', \quad \boldsymbol{\Lambda}_{uv} = \sum_{h=1}^{+\infty} \mathbb{E}\left(v_t v_{t-h}'\right). \tag{2.14}$$

where $\boldsymbol{\Omega}_{uu}$ the long-run covariance matrix. Notice that the particular decomposition of the long-run covariance matrix holds under the assumption that the innovation process of the predictive regression system is serially uncorrelated and stationary. For instance, under the presence of serial correlation, additional terms for the matrix $\boldsymbol{\Lambda}_{uv}$ needs to be incorporated. However, we do not consider such an econometric feature within our framework.

Next, for the case of the univariate predictive regression model we define $\widehat{\boldsymbol{\Omega}}_{FM} \equiv \sigma_{FM}^2$ and $\widehat{\boldsymbol{\Sigma}}_{uu} \equiv \widehat{\sigma}_u^2$, where $\widehat{\sigma}_u^2$ is a consistent estimator of σ_u^2. Therefore, we can use the bias correction of KMS for the univariate model which is given by

$$M = \left[\sum_{t=1}^{n} \tilde{z}_{t-1}^2 - n\bar{z}_{t-1}^2 \left(1 - \widehat{\rho}_{uv}^2\right)\right] \widehat{\sigma}_u^2, \qquad \text{with} \quad \widehat{\rho}_{uv}^2 = \frac{\widehat{\boldsymbol{\Omega}}_{uv}}{\widehat{\boldsymbol{\Sigma}}_{uu} \widehat{\boldsymbol{\Omega}}_{vv}}. \tag{2.15}$$

A key aspect for the computation of the IVX-Wald statistic is the estimation procedure for the matrices $\widehat{\boldsymbol{\Omega}}_{uv}$ and $\widehat{\boldsymbol{\Omega}}_{uu}$, which represent the estimated long-run covariance between u_t and v_t and the long-run variance of u_t respectively. These covariance matrices can be constructed using nonparametric kernels with preselected bandwidth parameters, such as the Newey-West type estimators (see, Kostakis et al. (2015), Newey and West (1987) and Andrews (1991)).

Furthermore, it can be proved that under the null hypothesis which imposes a set of linear restrictions in the parameter vector β, we obtain that $\mathcal{W}_n^{IVX} \implies \chi^2(q)$ as $n \to \infty$ (see, Theorem 1 in KMS). This important limit theory result provides a unified framework for robust inference and testing regardless the persistence properties of regressors. A simple example is the application of the IVX-Wald test for inferring the individual statistical significance of predictors under abstract degree of persistence. For additional scenarios such as predictors of mixed integration order see Phillips and Lee (2013) and Phillips and Lee (2016) in which cases the mixed normality assumption still holds. The particular bias correction allows to control the empirical size when using the IVX Wald formulation for testing hypothesis.

In summary, our main objective in this paper is to propose a unified framework for structural break testing in predictive regression models with persistent predictors. Our framework operates under the assumption of a single structural break under the alternative hypothesis, even though the case of multiple structural breaks can be considered in a subsequent study. In terms of the proposed tests we implement the IVX-Wald test proposed by Phillips and Magdalinos (2009) and use the supremum functional which is suitable to detect structural breaks in a potential unknown break-point location within the full sample. Another interesting application, is to extend the structural break testing framework, to test for both the presence of predictability and structural break, in a similar manner as in the frameworks of Gonzalo and Pitarakis (2012, 2017).

2.3 Structural Break Testing Framework

2.3.1 Econometric Environment

Consider testing for a single structural break at an unknown break point location. Then, the predictive regression model is expressed as below

$$y_t = (\mu_1 + \boldsymbol{\beta}_1' \boldsymbol{x}_{t-1}) \, I\{t \leq k\} + (\mu_2 + \boldsymbol{\beta}_2' \boldsymbol{x}_{t-1}) \, I\{t > k\} + u_t \tag{2.16}$$

where $k = \lfloor n\pi \rfloor$ with $\pi \in (0,1)$ such that the regressors are LUR processes

$$\boldsymbol{x}_t = \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}}{n^{\gamma_x}} \right) \boldsymbol{x}_{t-1} + v_t \tag{2.17}$$

where $\boldsymbol{C} = diag\{c_1, ..., c_p\} > 0$ for all $i \in \{1, ..., p\}$ and $\gamma = 1$ (LUR) or $\gamma_x \in (0,1)$ (MI).

The presence of a nuisance parameter, such as the unknown break-point $k = \lfloor n\pi \rfloor$, under the null of parameter constancy is well-known as the unidentified parameters problem under the null hypothesis, introduced by the seminal paper of Davies (1987) and references therein. Our proposed econometric framework allows to formulate statistical tests for parameter instability, as time-varying parameters in the predictive regression model, investigating this way the presence of potential unstable coefficients. Furthermore, similar to the case of predictability testing, which implies finding statistical evidence against the null hypothesis of no predictability, we are interested in detecting such cases under the presence of parameter instability. However, we do focus on a linear parametric conditional mean specification of the model in this chapter. For instance, the nonparametric predictive hypothesis of Kasparis et al. (2015) is constructed by comparing the nonparametric functional form estimator with its parametric counterpart estimator.

The current settings here although are based on less involved stochastic approximations to the nonparametric approach, it still requires to determine the stochastic approximations that appear to the expression of the limiting distribution of Wald-type statistics under abstract degree of persistence in regressors and the presence of a single structural break. In particular, we develop an econometric framework for testing the null hypothesis of linearity, implying that, $\mu_1 = \mu_2, \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, specifically for the predictive regression model which permits to model the time series properties of regressors via the LUR specification. In this paper we consider only the univariate predictive regression model which has a single predictant, while we consider both single and multiple predictors to illustrate the related derivations for developing the asymptotic theory of the proposed tests under different econometric scenarios.

Remark 2.1. For the correct identification of a possible structural break in the predictive regression model, since $k = \lfloor n\pi \rfloor$, is the break position such that $0 < \pi < 1$, we assume that the fraction of the structural break defined as $\pi_0 = k_0/n$ is within the interior of $(0,1)$ for some fixed π_0 parameter. The particular condition ensures that no nuisance parameters appear at the boundary of the parameter space. Thus, we allow structural breaks to occur at locations which are considered to be fractions of the sample size, such as $\pi_0 = k_0/n$. Under the assumption of an unknown break-point, to avoid unidentified parameters under the null the supremum functional is employed for the construction of the Wald tests.

2.3.2 Classical Least Squares Estimation

We are interested about the parameter stability of the model intercept, μ, of the predictive regression model when one is included and that of the model coefficients with $\boldsymbol{\beta} := (\beta_1, ..., \beta_p)$. We denote with $\hat{\mu}$ the estimator of the model intercept for the predictive regression model and with $\hat{\boldsymbol{\beta}} := \left(\hat{\beta}_1, ..., \hat{\beta}_p\right)$. In this Section, we develop the asymptotic theory which corresponds to the sup OLS-Wald test, that is based on the ordinary least squares estimator, when testing for a single structural break in predictive regressions with multiple predictors. Moreover, the asymptotic theory based on the OLS estimator allow us to obtain an analytic form for the limiting distribution of the test which indicates how the assumption of nonstationarity and the presence of an unknown coefficient of persistence in the underline stochastic process, affects statistical inference for the sup OLS-Wald statistic.

We denote with $\boldsymbol{X}_1 := \left[I\{t \leq k\} \;\; \boldsymbol{x}_t I\{t \leq k\}\right]$ and $\boldsymbol{X}_2 := \left[I\{t > k\} \;\; \boldsymbol{x}_t I\{t > k\}\right]$ and define $\boldsymbol{X} = [\boldsymbol{X}_1 \; \boldsymbol{X}_2] \in \mathbb{R}^{n \times 2(p+1)}$ the corresponding partitioned matrix. Moreover, the linear restriction matrix is denoted with $\boldsymbol{\mathcal{R}} = [\boldsymbol{I}_{p+1} \;\; -\boldsymbol{I}_{p+1}]$ with $\boldsymbol{I}_{p+1}$ an identity matrix. Let $\theta_j = (\alpha_j, \beta_j)$ for $j = 1, 2$ and $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$, the parameter vector which implies that the predictive regression can be expressed as $y = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{u}$. Then, the OLS Wald statistic for testing the null hypothesis is

$$\mathbb{H}_0 : \boldsymbol{\theta}_1 := \left(\mu^{(1)}, \beta_1^{(1)}, ..., \beta_p^{(1)}\right) \equiv \boldsymbol{\theta}_2 := \left(\mu^{(2)}, \beta_1^{(2)}, ..., \beta_p^{(2)}\right) \tag{2.18}$$

against $\mathbb{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, for $k = \lfloor \pi n \rfloor$, is given by the following expression

$$\mathcal{W}_n^{OLS}(\pi) = \frac{1}{\hat{\sigma}_u^2} \left(\widehat{\boldsymbol{\theta}}_1(\pi) - \widehat{\boldsymbol{\theta}}_2(\pi)\right)' \left[\boldsymbol{\mathcal{R}} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{\mathcal{R}}'\right]^{-1} \left(\widehat{\boldsymbol{\theta}}_1(\pi) - \widehat{\boldsymbol{\theta}}_2(\pi)\right) \tag{2.19}$$

where $\hat{\sigma}_u^2 \overset{p}{\to} \sigma_u^2$ is a consistent (empirical) estimator of the residuals of the predictive regression model. Then, statistical inference under the null hypothesis of no structural break in the predictive regression are conducted using the supremum functional. Thus, taking into consideration the unknown nature of the break-point, the sup OLS-Wald statistic is expressed as below

$$\widetilde{\mathcal{W}}^{OLS} := \sup_{\pi \in [\pi_1, \pi_2]} \mathcal{W}_n^{OLS}(\pi), \tag{2.20}$$

where $0 < \pi_1 < 1$ and $0 < \pi_2 < 1$ with $\pi_2 = 1 - \pi_1$.

Under the null hypothesis the break-point k is unidentified and so the supremum functional selects the maximum Wald statistic corresponding to a sequence of Wald statistics evaluated at values within the interval $[\pi_1, \pi_2]$. Specifically, in order to construct the corresponding supremum Wald tests we split the full sample $\{y_t, x_{t-1}\}_{t=1}^{n}$ into two sub-samples; the first sub-sample corresponds to the time period before time τ, $\{y_t, x_{t-1}\}_{t=1}^{\tau}$ and the second sub-sample corresponds to the time period after time t, $\{y_t, x_{t-1}\}_{t=\tau+1}^{n}$. Due to the fact that we operate within the Skorokhod topology the sample moments that correspond to these sub-samples weakly converge to the corresponding asymptotic result which are based on the $\mathcal{D}\left([0,1]\right)$ topology. For instance, when deriving the sample moments we can use the notation $\pi \in [\pi_1, \pi_2]$, where π_1 and π_2 are the lower and upper bounds for the possible break-point location.

Theorem 2.1. Under conditions of Assumption 2.1, then the sup OLS-Wald statistic for the null hypothesis $\mathbb{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ against $\mathbb{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, where $\theta_j = (\mu_j, \beta_j)'$ with $j = 1, 2$ based on the predictive regression model (4.11)-(4.12) with $\gamma_x = 1$ (LUR predictors) weakly converges to the following limiting distribution

$$\widetilde{\mathcal{W}}^{OLS}(\pi) \equiv \sup_{\pi \in [\pi_1, \pi_2]} \mathcal{W}_T^{OLS}(\pi) \Rightarrow \sup_{\pi \in [\pi_1, \pi_2]} \left\{ \widetilde{\boldsymbol{N}}_c'(\pi) \widetilde{\boldsymbol{M}}_c(\pi)^{-1} \widetilde{\boldsymbol{N}}_c(\pi) \right\} \tag{2.21}$$

where

$$\widetilde{\boldsymbol{M}}_c(\pi)^{-1} = \widetilde{\boldsymbol{G}}_c(\pi) - \widetilde{\boldsymbol{G}}_c(\pi) \widetilde{\boldsymbol{G}}_c(1)^{-1} \widetilde{\boldsymbol{G}}_c(\pi) \tag{2.22}$$

and

$$\widetilde{\boldsymbol{N}}_c(\pi) = \left\{ \widetilde{\boldsymbol{G}}_c(\pi)^{-1} \widetilde{\boldsymbol{H}}_c(\pi) - \left[\widetilde{\boldsymbol{G}}_c(1) - \widetilde{\boldsymbol{G}}_c(\pi) \right]^{-1} \left[\widetilde{\boldsymbol{H}}_c(1) - \widetilde{\boldsymbol{H}}_c(\pi) \right] \right\}' \tag{2.23}$$

such that $\widetilde{\boldsymbol{K}}_c(r) := (\mathbf{1}, \boldsymbol{K}_c(r))'$, we define $\widetilde{\boldsymbol{G}}_c(\pi)$ and $\widetilde{\boldsymbol{H}}_c(\pi)$ as below

$$\widetilde{\boldsymbol{G}}_c(\pi) := \int_0^{\pi} \widetilde{\boldsymbol{K}}_c(r) \widetilde{\boldsymbol{K}}_c'(r) dr \quad \text{and} \quad \widetilde{\boldsymbol{H}}_c(\pi) := \int_0^{\pi} \widetilde{\boldsymbol{K}}_c(r) d\boldsymbol{B}_u(r) \tag{2.24}$$

with $\widetilde{\boldsymbol{G}}_c(\pi) \in \mathbb{R}^{(p+1)\times(p+1)}$ a positive-definite stochastic matrix and $\widetilde{\boldsymbol{H}}_c(\pi) \in \mathbb{R}^{(p+1)\times 1}$.

Theorem 2.1 is our first theoretical contribution to the literature, demonstrating that the sup OLS-Wald statistic for structural break testing with an unknown break-point in the predictive regression model with persistent predictors, does not converge to the NBB result as shown by Andrews (1993) for linear regression models. More specifically, as seen by the limiting distribution of the sup OLS-Wald test, the asymptotic theory of the OLS based test depends on the nuisance parameter of persistence c_i.

Remark 2.2. Notice that due to the presence of the unknown persistence parameter, c_i, the OLS-Wald test does not converge to the NBB asymptotic limit, when we have LUR predictors (i.e., $\gamma_x = 1$ and $c_i > 0 \ \forall \ i$), even under the assumption of a known break-point, that is, $\pi \equiv \pi_0$. The particular aspect, motivates us to investigate whether the implementation of an IVX based Wald test which is robust to the persistence properties, would induce an asymptotic distribution free of any nuisance parameters.

Remark 2.3. Notice that $\widehat{\sigma}_u^2$ refers to the residual variance under $\mathbb{H}_0$. Under $\mathbb{H}_0$, $\dfrac{\widehat{\sigma}_0^2}{\widehat{\sigma}_1^2} \overset{p}{\to} 1$, thus the residual variance does not affect the limiting distribution. However, when the finite sample properties of the test are examined, using the residual variance under the null instead the one under the alternative, results to incorrect power function dynamics. Since $\widehat{\sigma}_1^2$ takes the shape of the $\mathbb{H}_1$ into account then, this improves the ability of the test in detecting departures from the null hypothesis, towards specific local alternatives.

2.3.3 IVX based Estimation

In this section, we propose the IVX-Wald type statistic for testing for the presence parameter instability in the predictive regression model. The supremum IVX-Wald test is constructed by splitting the sample into two sub-samples. Then, the model estimates of the two sub-samples are used to construct the proposed test statistics. We denote with $\widetilde{\beta}_1^{IVX}(\pi)$ and $\widetilde{\beta}_2^{IVX}(\pi)$ the IVX estimates of the two sub-samples; and with $\widetilde{Q}_1(\pi)$ and $\widetilde{Q}_2(\pi)$ the corresponding covariance matrices. The sample estimates are given by

$$\widetilde{\boldsymbol{\beta}}_1^{IVX}(\pi) = \left(\frac{1}{\pi} \sum_{t=1}^{\pi} \widetilde{z}_{1,t-1} x'_{1,t-1} \right)^{-1} \left(\frac{1}{\pi} \sum_{t=1}^{\pi} \widetilde{z}_{1,t-1} y_t \right), \tag{2.25}$$

$$\widetilde{\boldsymbol{\beta}}_2^{IVX}(\tau) = \left(\frac{1}{n-\pi} \sum_{t=\pi+1}^{n} \widetilde{z}_{2,t-1} x'_{2,t-1} \right)^{-1} \left(\frac{1}{n-\pi} \sum_{t=\pi+1}^{T} \widetilde{z}_{2,t-1} y_t \right). \tag{2.26}$$

We begin our asymptotic theory analysis by considering the case in which under the null hypothesis both the model intercept and the model slopes have no structural break. Then, the testing hypothesis of interest is given by

$$\mathbb{H}_0 : \alpha_1 = \alpha_2 \quad \text{and} \quad \beta_1 = \beta_2 \tag{2.27}$$

For evaluating our hypothesis we use the IVX-Wald test which has the following form[14]

$$\mathcal{W}_{\beta}^{IVX}(\pi) = \left(\widetilde{\boldsymbol{\beta}}_1^{IVX}(\pi) - \widetilde{\boldsymbol{\beta}}_2^{IVX}(\pi) \right)' \left[\widetilde{\boldsymbol{Q}}_1(\pi) + \widetilde{\boldsymbol{Q}}_2(\pi) \right]^{-1} \left(\widetilde{\boldsymbol{\beta}}_1^{IVX}(\pi) - \widetilde{\boldsymbol{\beta}}_2^{IVX}(\pi) \right). \tag{2.28}$$

We denote with $\widetilde{\mathcal{W}}_{\beta}^{IVX}(\pi) \equiv \sup_{\pi \in [\pi_1, \pi_2]} \mathcal{W}_{\beta}^{IVX}(\pi)$, the corresponding sup IVX-Wald statistic, since we consider the case of the unknown break-point. The two IVX estimates of β and their corresponding asymptotic variances are computed, using the data from each sub-sample separately, defined as below

$$\widetilde{\boldsymbol{Q}}_1(\pi) = \left(\boldsymbol{Z}'_1 \boldsymbol{X}_1 \right)^{-1} \left(\boldsymbol{Z}'_1 \boldsymbol{Z}_1 \right) \left(\boldsymbol{X}'_1 \boldsymbol{Z}_1 \right)^{-1} \quad \text{and} \quad \widetilde{\boldsymbol{Q}}_2(\pi) = \left(\boldsymbol{Z}'_2 \boldsymbol{X}_2 \right)^{-1} \left(\boldsymbol{Z}'_2 \boldsymbol{Z}_2 \right) \left(\boldsymbol{X}'_2 \boldsymbol{Z}_2 \right)^{-1} \tag{2.29}$$

Thus, the $\mathcal{W}_{\beta}^{IVX}(\pi)$ statistic can be thought as a Chow-type statistic for detecting break at time π, such that $\pi_1 \leq \pi \leq \pi_2$. Then, due to the unknown nature of the break-point the sup Wald test is simply a sequence of Chow-type statistics within the same probability space. Theorem 3.2 presents the limiting distribution of $\widetilde{\mathcal{W}}_{\beta}^{IVX}(\pi)$, under the null hypothesis of no parameter instability in the predictive regression model.

[14]Note that the covariance matrices are computed based on the long-covariance matrices and corresponding FM corrections given by KMS, see also Kasparis et al. (2015). These definitions are employed since we do not rule out the strong assumption of a weakly covariance dependence in the model.

Theorem 2.2. Under Assumptions 2.1 and 4.3 hold and π denotes the unknown break-point, then the sup IVX-Wald statistic under the null hypothesis (with α known to be stable a priori) $\mathbb{H}_0 : \alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, based on the predictive regression model (4.11)-(4.12) and no restriction on the exponent rate γ_x, has the following asymptotic behaviour

$$\widetilde{\mathcal{W}}_{\beta}^{IVX}(\tau) \Rightarrow \sup_{\pi \in [\pi_1, \pi_2]} \left\{ \boldsymbol{N}(\pi)' \boldsymbol{M}(\pi)^{-1} \boldsymbol{N}(\pi) \right\} \tag{2.30}$$

with $\pi \in [\pi_1, \pi_2]$[15], where

$$\boldsymbol{N}(\pi) = \boldsymbol{B}_p(\pi) - \boldsymbol{R}(\pi)\boldsymbol{B}_p(1) \tag{2.31}$$

$$\boldsymbol{M}(\pi) = \pi \big(\boldsymbol{I}_p - \boldsymbol{R}(\pi)\big)\big(\boldsymbol{I}_p - \boldsymbol{R}(\pi)\big)' + (1-\pi)\boldsymbol{R}(\pi)\boldsymbol{R}(\pi)' \tag{2.32}$$

such that

$$\boldsymbol{R}(\pi) = \begin{cases} \left(\pi \boldsymbol{\Omega}_{xx} + \int_0^{\pi} \underline{\boldsymbol{B}} dB'\right)\left(\boldsymbol{\Omega}_{xx} + \int_0^{1} \underline{\boldsymbol{B}} dB'\right)^{-1} & , \text{if } \gamma_x > 1 \\ \left(\pi \boldsymbol{\Omega}_{xx} + \int_0^{\pi} \underline{\boldsymbol{J}}_C dJ_C'\right)\left(\boldsymbol{\Omega}_{xx} + \int_0^{1} \underline{\boldsymbol{J}}_C dJ_C'\right)^{-1} & , \text{if } \gamma_x = 1 \\ \pi \boldsymbol{I}_p & , \text{if } \gamma_x < 1 \end{cases} \tag{2.33}$$

where $B(.)$ is a $p-$dimensional standard Brownian motion, $J_C(\pi) = \int_0^{\pi} e^{C(\pi - s)} dB(r)$ is an *Ornstein-Uhkenbeck* (OU) process and we denote with $\underline{J}_C(\pi) = J_C(\pi) - \int_0^1 J_C(s)ds$ and $\underline{B}(\pi) = B(\pi) - \int_0^1 B(s)ds$ the demeaned processes of $J(\pi)$ and $B(\pi)$ respectively.

Theorem 2.2 demonstrates that the supremum functional of the IVX-Wald test weakly convergence to a stochastic quadratic functional and does not have the same asymptotic behaviour as the corresponding IVX-Wald statistic which is proved to follow a χ_p^2 distribution as in the framework proposed by Kostakis et al. (2015). Specifically, based on our asymptotic theory analysis we show that the limit distribution of the test statistic weakly converges to a non deterministic quadratic due to the dependence to Brownian motion functionals that contain two nuisance parameters, the unknown break-point $\pi \in [\pi_1, \pi_2]$ and the unknown localizing coefficient of persistence c_i.

An important implication of Theorem 2.2, which consists our second contribution to the limit theory of structural break tests in predictive regression models with nonstationary predictors, is that we show that the dependence of the limiting distribution to the unknown parameter of persistence takes different forms by restricting further the parameter space of the exponent rate. Specifically, Theorem 2.2 demonstrates that when testing for a structural break in predictive regression models, under the assumption that the predictors are integrated or nearly integrated, then the limiting distribution of the test, weakly converge to a nonstandard process, which verifies the corresponding result already mentioned by Hansen (2000b)). When the predictors are stationary or mildly stationary, then the test statistic weakly converges to the familiar squared tied-down Bessel process as proved by Andrews (1993) in the case of linear regression models.

[15]Note that, since we operate within the Skorokhod topology $\mathcal{D}(0,1)$ we can apply standard weakly convergence arguments.

The latter implies that when we consider separately the special case for which $\gamma_x < 1$, which covers both the cases of mildly integrated regressors, i.e., $\gamma_x \in (0,1)$, and stationary regressors, i.e., $\gamma_x = 0$, then it can be proved that the limiting distribution of the sup IVX-Wald test converges to the standard NBB result[16]. Corollary 3.1 below is a direct implication of Theorem 2.2 and summarizes this finding.

Corollary 2.2. Under the assumptions and definitions given by Theorem 2.2, when $\gamma_x < 1$ then the following asymptotic distribution holds

$$\widetilde{\mathcal{W}}_{\beta}^{IVX}(\tau) \Rightarrow \sup_{\pi \in [\pi_1, \pi_2]} \frac{\mathcal{BB}_p(\pi)' \mathcal{BB}_p(\pi)}{\pi(1-\pi)}, \tag{2.34}$$

where $\mathcal{BB}_p(.)$ is a $p-$dimensional standard Brownian bridge.

Remark 2.4. The weakly convergence of the sup IVX-Wald test into a normalized squared Brownian bridge, when the exponent rate is restricted such that $\gamma_x < 1$, implies standard statistical inference due to the known distribution, and this occurs in the case when we have a mildly integrated predictor. One rejects $\mathbb{H}_0$ for large values of the sup IVX-Wald test based on a significance level α such that $0 \leq \alpha \leq 1$ and thus the limit distribution can be used to derive associated critical values, denoted with c_α such that $\mathbb{P}(\widetilde{\mathcal{W}}_{\beta}^{IVX}(\pi) > c_\alpha) > 0$ with $\lim_{T \to \infty} \mathbb{P}(\widetilde{\mathcal{W}}_{\beta}^{IVX}(\pi) > c_\alpha) = 1$.

Our findings presented by Theorem 2.2 verify the conjuncture of Hansen (2000b) who argues that when testing for a structural break based on a sup-functional induces an asymptotic non-pivotal distribution under the assumption of nonstationarity. However, this seminal study has not examined in details certain forms of nonstationarity which can occur and how these are manifested in the limit theory of the tests. In particular, within our framework we demonstrate using local-to-unit root asymptotic arguments that the nonstandard limiting distribution occurs when x_t is properly modelled as a nonstationary stochastic process, and specifically the NBB result no longer holds when x_t is either a nearly integrated or an integrated process. The proofs of both Theorem 2.2 and Corollary 2.2 can be found in Appendix A.1. However, notice that in the case when the structural break is known, say $\pi \equiv \pi_0$ then the limiting distribution simplifies to the standard NBB result in all cases of Theorem 2.2, since $\int_0^{\pi_0} f(x)dx = \pi_0 \int_0^1 f(x)dx$.

[16]Note that in the case of a known break-point we can easily prove a convergence to a χ^2_p distribution which is free of nuisance parameters and so conventional inference methods apply

2.4 Joint Predictability and Structural Break Testing

For evaluating our hypothesis we use the IVX-Wald test which has the following form.

$$\mathcal{W}_{\beta}^{IVX}(\pi) = \left(\widetilde{\beta}_1^{IVX}(\pi) - \widetilde{\beta}_2^{IVX}(\pi) \right)' \left[\widetilde{\boldsymbol{Q}}_1(\pi) + \widetilde{\boldsymbol{Q}}_2(\pi) \right]^{-1} \left(\widetilde{\beta}_1^{IVX}(\pi) - \widetilde{\beta}_2^{IVX}(\pi) \right). \tag{2.35}$$

We denote with $\widetilde{\mathcal{W}}_{\beta}^{IVX}(\pi) \equiv \sup_{\pi \in [\pi_1, \pi_2]} \mathcal{W}_{\beta}^{IVX}(\pi)$, the corresponding sup IVX-Wald statistic, since we consider the case of the unknown break-point. The two IVX estimates of β and their corresponding asymptotic variances are computed, using the data from each sub-sample separately. Thus, the $\mathcal{W}_{\beta}^{IVX}(\pi)$ statistic can be thought as a Chow-type statistic for detecting break at time t, such that $\pi_1 \leq \pi \leq \pi_2$. Then, due to the unknown nature of the break-point the sup Wald test is simply a sequence of Chow-type statistics within the same probability space. Theorem 2.2 presents the limiting distribution of $\widetilde{\mathcal{W}}_{\beta}^{IVX}(\pi)$, under the null hypothesis of no parameter instability in the predictive regression model. The proofs of both Theorem 2.2 and Corollary 3.1 can be found in Appendix A.1. Next, we focus on designing test statistics for jointly testing against predictability and structural break in predictive regression models. We define a joint Wald test based on the IVX estimator under the assumption of an unknown break-point as below

$$\widetilde{\mathcal{W}}_{\beta}^{joint} := \sup_{\pi \in [\pi_1, \pi_2]} \left\{ \mathcal{W}_n^{IVX} + \mathcal{W}_{\beta}^{IVX}(\pi) \right\}, \quad \text{for} \ \ \pi_1 \leq \pi \leq \pi_2. \tag{2.36}$$

where the supremum functional applies only on the second component of the test above. Then, the Corollary 3.2 below gives the related limit theory result.

Corollary 2.3. *(i)* If the conditions of Theorem 2.2 hold, then under the null hypothesis, $\mathbb{H}_0 : \alpha_1 = \alpha_2$ and $\beta_1 = \beta_2 = 0$ and no restriction on the exponent rate γ_x, the large sample theory of the test statistic specified by (2.36) has the following form

$$\widetilde{\mathcal{W}}_{\beta}^{joint} \Rightarrow B(1)'B(1) + \sup_{\pi \in [\pi_1, \pi_2]} \left\{ \boldsymbol{N}(\pi)' \boldsymbol{M}(\pi)^{-1} \boldsymbol{N}(\pi) \right\}, \tag{2.37}$$

where $B(.)$, $\boldsymbol{N}(\pi)$ and $\boldsymbol{N}(\pi)$ are defined in Theorem 2.2.
(ii) As a special case, when $\gamma_x < 1$, it follows that

$$\widetilde{\mathcal{W}}_{\beta}^{joint} \Rightarrow \chi_p^2 + \sup_{\pi \in [\pi_1, \pi_2]} \frac{\mathcal{BB}_p(\pi)' \mathcal{BB}_p(\pi)}{\pi(1-\pi)}, \tag{2.38}$$

where $\mathcal{BB}_p(.)$ is a $p-$dimensional standard Brownian bridge and χ_p^2 denotes the χ^2 random variable with p degrees of freedom. Furthermore, the two stochastic quantities of the limiting distribution are assumed to be independent.

Remark 2.5. Notice that Corollary 2.3 demonstrates that when $\gamma_x < 1$, the large sample theory of the test statistic $\mathcal{W}_{\beta}$ is pivotal. Therefore, in this special case by having a limiting distribution being free of nuisance parameters, asymptotic critical values for testing the null hypothesis can be easily obtained. The particular test statistic provides a methodology for testing for both predictability and structural break which is robust to the persistence properties of regressors after replacing the OLS with an IVX estimator.

The main arguments we use for the proof of Corollary 3.2 is to consider the joint testing hypothesis as a composite hypothesis based on the mutually exclusive parameter space, and thus we construct the limit theory of this test based on the joint formulation of the two separate testing hypotheses. By employing the asymptotic matrix moments based on the IVX method we prove that the limiting distribution of the joint test can be decomposed into two components. Since these components are independent random variables and thus practically we could use the critical values of the limiting distributions that corresponds to each of these stochastic quantities.

2.4.1 Joint Wald Tests

Proposition 2.1. Consider the predictive regression model given by expressions (4.11)-(4.12). If Assumption 2.1-4.3 hold and α is known to be unstable a priori, then under the null hypothesis $\mathbb{H}_0 : \alpha_1 = \alpha_2$ and $\beta_1 = \beta_2 = \beta$, as $n \to \infty$ the following limit result holds

$$\widetilde{\mathcal{W}}_{\alpha}^{IVX}(\tau) \Rightarrow \sup_{\pi \in [\pi_1, \pi_2]} \frac{\mathcal{BB}_1(\pi)' \mathcal{BB}_1(\pi)}{\pi(1-\pi)} \tag{2.39}$$

where $\mathcal{BB}_1(.)$ is a one-dimensional standard Brownian bridge.

Remark 2.6. Notice that Proposition 2.1 provides an asymptotic result for a composite hypothesis since we consider jointly testing for a structural break in the model intercept and the slope coefficients while we test, that under the null hypothesis the slope coefficient has a fixed parameter value β. Additionally, we can investigate the limiting distribution of the joint Wald test when we assume that under the null hypothesis there is no predictability.

Proposition 2.2. Consider the predictive regression model . If Assumption 2.1-4.3 hold and α is known to be unstable a priori, then under the null hypothesis $\mathbb{H}_0 : \alpha_1 = \alpha_2$ and $\beta_1 = \beta_2 = 0$, as $n \to \infty$ the following limit results hold: (i)

$$\widetilde{\mathcal{W}}_{\alpha\beta}^{joint}(\pi) \Rightarrow B(1)'B(1) + \sup_{\pi \in [\pi_1, \pi_2]} \left\{ \widetilde{\boldsymbol{N}}(\pi)' \widetilde{\boldsymbol{M}}(\pi)^{-1} \widetilde{\boldsymbol{N}}(\pi) \right\} \tag{2.40}$$

where $\widetilde{\boldsymbol{N}}(\pi) = \big(\mathcal{BB}_1(\pi), \boldsymbol{N}(\pi)\big)'$ and $\widetilde{\boldsymbol{M}}(\pi) = \begin{pmatrix} \pi(1-\pi) & 0 \\ 0 & \boldsymbol{M}(\pi) \end{pmatrix}$. The terms, $\boldsymbol{N}(\pi)$ and $\boldsymbol{M}(\pi)$ are defined in Theorem 2.2.

(ii) As a special case, when $\gamma_x \in (0,1)$, it holds that

$$\widetilde{\mathcal{W}}_{\alpha\beta}^{joint}(\tau) \Rightarrow \chi_p^2 + \sup_{\pi \in [\pi_1, \pi_2]} \frac{\mathcal{BB}_{p+1}(\pi)' \mathcal{BB}_{p+1}(\pi)}{\pi(1-\pi)} \tag{2.41}$$

where $\mathcal{BB}_{p+1}(.)$ is a $(p+1)-$dimensional standard Brownian bridge, and χ_p^2 is a random variable following a χ^2 distribution with p degrees of freedom.

Remark 2.7. Notice that, Proposition 2.2 shows that the limiting distribution of the joint test for both predictability and structural break, when we consider simultaneously testing whether there is a structural break to the model intercept and no predictability using the set of regressors of the model, has an asymptotic distribution which takes a different form when we consider different values of the parameter space of the exponent rate.

2.5 Monte Carlo Simulation Study

Monte Carlo simulations were performed to examine the quality of the asymptotic approximations to the finite-sample distributions. In this section, we present extensive Monte Carlo simulation experiments to examine the finite size properties of the proposed Wald-type tests in terms of their empirical size and power performance, under the null hypothesis of parameter constancy in the predictive regression with persistent predictors. In practise, the degree of persistence in the time series of the regressors is unknown. More specifically, both the coefficient of persistence c_i as well as the exponent rate γ_x are both unknown parameters to the researcher. Moreover, we have proved that the limiting distribution of the Wald-type statistics for detecting structural break in predictive regression models depend on these unknown properties of the regressors for certain parameter value restrictions on the coefficient of persistence.

Therefore, to demonstrate the above theoretical result, under the null hypothesis of no structural break, we can generate a DGP with no breaks in the coefficients of the predictive regression. Then, constructing the sup-Wald test and using the Andrews' critical values we can observe whether size distortions indeed occur in this scenario[17]. Secondly, we propose an alternative approach to overcome this problem. In particular, using an IV based sup-Wald test, which is constructed using the IVX instrumentation, we prove that the limiting distribution of the statistic indeed weakly converges to a nonstandard limiting distribution, therefore a bootstrap methodology is necessary to control the empirical size. Furthermore, under certain parameter restrictions on the nuisance parameter we prove that the NBB limit holds, which allow us to use the Andrews' critical values, avoiding this way to simulate critical values which can be computational complex. Next, we focus on the data generating process, the test statistics as well as the size and power comparisons for the proposed Wald type statistics.

2.5.1 Data Generating Process

We use the following data generating process (DGP) where y_t is a scalar and x_t is a vector of LUR predictors (with the property of being highly persistent). We begin with the case of a predictive regression with two predictors.

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-1} + u_t \tag{2.42}$$

$$x_t = \begin{pmatrix} 1 - \frac{c_1}{n^{\gamma_x}} & 0 \\ 0 & 1 - \frac{c_2}{n^{\gamma_x}} \end{pmatrix} x_{t-1} + v_t, \ x_0 = 0. \tag{2.43}$$

with $t \in \{1, ..., n\}$ and $n = \{100, 250, 500, 1000\}$ for $B = 5,000$ replications, such that $\gamma_x = 1$.

[17] Recall that under the assumption of persistent predictors, we prove via Theorem 2.1 that when testing for parameter instability in the predictive regression the limiting distribution of the sup OLS-Wald statistic no longer follows a normalized squared brownian bridge, (NBB) for an unknown structural break. In particular, when contacting inference or assessing the statistical validity of the sup-Wald statistic via a MC experiment, using the corresponding critical values of the sup-Wald test proposed by Andrews (1993) we can observe that leads to size distortions due to the non-NBB limiting distribution.

Furthermore, we consider the effect of different localizing coefficients of persistence across the predictors. We use $c_i \in \{1, 5, 10, 20\}$ for $i = 1, 2$, which cover various cases of LUR regressors, with smaller values implying that we impose the assumption of higher persistence and lower values implying that existence of mild persistence in the predictor. The covariance matrix of the innovations $e_t = (u_t, v_t)' \sim \mathcal{N}(0_{3\times 1}, \Sigma_{ee})$ we assume that is parametrised with the following dependence structure

$$\Sigma_{ee} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{or} \quad \Sigma_{ee} = \begin{bmatrix} 1 & 0.10 & -0.29 \\ 0.10 & 1 & -0.03 \\ -0.29 & -0.03 & 1 \end{bmatrix}$$

2.5.2 Test statistics

We assess the statistical validity in finite and large samples of the Wald based statistics within our structural break framework, that is, the sup OLS-Wald test and the sup IVX-Wald test. In particular, we are interested to verify any size distortions under the null hypothesis of no structural change, when using the sup OLS-Wald test and we expect to observe improvements in the empirical size of the sup IVX-Wald test. Moreover, we examine the finite sample size of the two statistics for different degree of persistence as well as the rate at which we allow the IVX instrumentation procedure to induce a more mildly integrated regressor (controlled by the parameters c_z and the exponent rate γ_z). For the large sample properties of the test statistics a standard convergence result apply, that is, $\mathcal{W}_n(\pi; \gamma_z) \overset{d}{\to} \mathcal{W}(\pi; \gamma_z)$ as $n \to \infty$.

Then, the testing hypothesis of interest is a two-sided type hypothesis which is expressed as

$$\mathbb{H}_0 : \beta_1 = \beta_2 \quad \text{vesus} \quad \mathbb{H}_1 : \beta_1 \neq \beta_2 \tag{2.44}$$

Therefore, the test statistics are computed via expressions (2.45) and (2.46), while standard regularity and invariance principles holds.

Wald-OLS statistic

$$\mathcal{W}_n^{\text{OLS}}(k) = \frac{1}{\hat{\sigma}_u^2} \left(\hat{\beta}_1 - \hat{\beta}_2 \right)' \left[\left(X_1' X_1 \right)^{-1} + \left(X_2' X_2 \right)^{-1} \right]^{-1} \left(\hat{\beta}_1 - \hat{\beta}_2 \right) \tag{2.45}$$

Wald-IVX statistic

$$\mathcal{W}_n^{\text{IVX}}(\pi) = \frac{1}{\hat{\sigma}_u^2} \left(\hat{\beta}_1^{\text{IVX}} - \hat{\beta}_2^{\text{IVX}} \right)' \mathcal{Q}_{\mathcal{R}}^{-1} \left(\hat{\beta}_1^{\text{IVX}} - \hat{\beta}_2^{\text{IVX}} \right) \tag{2.46}$$

The covariance matrix, $\mathcal{Q}_{\mathcal{R}}(k)$, is estimated using the decomposition, $\mathcal{Q}_{\mathcal{R}}(k) := \widetilde{\boldsymbol{Q}}_1(k) + \widetilde{\boldsymbol{Q}}_2(k)$

as well as by adjusting for the second-degree bias for each sub-sample, as below

$$\widetilde{\boldsymbol{Q}}_j(k) = \left(\sum_{t=1}^{T} \widetilde{z}_{jt-1} x'_{jt-1} \right)^{-1} \boldsymbol{M}_j(k) \left(\sum_{t=1}^{T} x'_{jt-1} \widetilde{z}_{jt-1} \right)^{-1} \tag{2.47}$$

$$\boldsymbol{M}_j(k) = \widehat{\sigma}_u^2 \left(\sum_{t=1}^{T} \widetilde{z}_{jt-1} \widetilde{z}'_{jt-1} \right) - T_j \bar{z}_{jT-1} \bar{z}'_{jT-1} \widehat{\boldsymbol{\Omega}}_{FM} \tag{2.48}$$

where $\widehat{\boldsymbol{\Omega}}_{FM} = \widehat{\boldsymbol{\Sigma}}_{uu} - \widehat{\boldsymbol{\Omega}}_{uv} \widehat{\boldsymbol{\Omega}}_{uu} \widehat{\boldsymbol{\Omega}}'_{uv}$, is estimated based on the full sample, for $j \in \{1, 2\}$ and T_j denotes the sample size for each sub-sample, i.e., $T_j = k$ for $j = 1$ and $T_j = (T - k)$ for $j = 2$.

To evaluate the performance of the tests in relation to the asymptotic approximations in the theory section of the paper, we focus on two cases, that is, (i) Design 1: the case of a fixed break-point; (ii) Design 2: the case of a random (unknown) break-point within the permissible parameter space of the fraction parameter π such that $\pi \in \Pi = [\pi_1, \pi_2]$. Furthermore, the data that we generate from the DGP given by (4.106) and (4.107) only differ over the values of c_1 and c_2 and are the same for different values of the sample size T. Thus, as the sample size increases the degree of persistence of the endogenous regressors remain the same.

Design (Random break-point) $\pi \in \Pi = [\pi_1, \pi_2]$. In the case of an unknown break-point, we obtain the critical values proposed by Andrews (1993) with an appropriate significance size such as $\alpha = 5\%$ only under mildly integradeness. Following the common practise in the literature (see, Andrews (1993) and (Caner and Hansen, 2001, p. 1563)), we use the trimming parameters $\pi_1 = 0.15$ and $\pi_2 = 0.85$. On the other hand, a bootstrap based methodology can be implemented for both the OLS-Wald and IVX-Wald statistics in the case of persistent regressors (LUR) while for the case of mildly integrated regressors (MI) the critical values of Andrews can be utilized. Specifically, we use the residual wild bootstrap (RWB) procedure as given by Algorithm 1.

Our asymptotic theory analysis demonstrates that when the break is known a prior, then the limiting distributions of the test statistics can be simplified, while in the case we allow for a structural break at an an unknown location within the sample, to obtain critical values bootstrap-based inference is required. Since we focus on the case of an unknown break-point location, then the sup functional is employed to obtain the maximum test statistic value after estimating a sequence of statistics over the interval $\pi \in [\pi_1, 1 - \pi_2]$. Thus, critical values can be obtained from the corresponding limiting distribution of the bootstrapped test statistic due to the nonstandard and nonpivotal limiting distribution under the null hypothesis when regressors have high persistence[18]. In practise, this makes it difficult to control the empirical size in those cases in which the limiting distribution of the tests is nonstandard, due to two factors: (i) the presence of the unknown coefficient of persistence which cannot be consistently estimated, and (ii) the unavailability of asymptotic critical values for the limiting distribution given by Theorem 2.2 (e.g., $\gamma_x = 1$ and $\gamma_x \in (0, 1)$). Therefore, bootstrap implementations of the structural break tests such as via a fixed regressor bootstrap, in the same spirit as in Georgiev et al. (2018) are necessary to control size distortions given a prespecified significance level.

[18] As we have seen for a known break-point the IVX-Wald test gives an asymptotically pivotal limit distribution (a standard χ^2-distribution). A pivotal statistic is one whose distribution is independent of the true parameter. In the case of the sup IVX-Wald test this property doesn't hold and this is the reason we need to utilize bootstrap based methodologies.

Remark 2.8. Notice that the FM covariance correction provides a second-degree bias correction for the IVX-Wald test due to the presence of persistent predictors. For instance, since in finite sample the mixed Gaussanity assumption of the IVX estimator can be violated due to the fact that a component of the limit expression might not vanish as it depends on three factors: the degree of regressor persistence as captured by the exponent rate γ_x, the correlation between the innovation sequence u_t and v_t, and the choice of the exponent rate of the instrumentation procedure which is responsible to control the degree of persistence in the endogenous generated instruments. Therefore, the FM covariance correction employs a weighted demeaning of the IVX instruments by a matrix that depends on $\widehat{\Omega}_{uv}$ which balances the presence of bias in finite samples occurring for various combinations of persistence and correlations across the persistence classes that predictors are allowed to belong to. In other words, all finite-sample effects are simultaneously removed by the finite-sample correction on the self-normalizing component of the IVX-Wald statistic (see, Remark A (2) in Appendix of Kostakis et al. (2015)).

Remark 2.9. The particular finite-sample correction proposed by KMS is a weighted demeaning of the dominating term $\widetilde{Z}'\widetilde{Z} \otimes \widehat{\Sigma}_{uu}$ by the term $n\bar{z}_{n-1} \otimes \bar{z}_{n-1} \otimes \widehat{\Omega}_{FM}$. In simple words, this correction removes the finite-sample effects of estimating a model intercept in the predictive regression. Then, the FM covariance matrix correction given by $\widehat{\Omega}_{FM}$ is a fully modified correction that controls the effect of correlation between u_t and v_t on the remainder term of the Gaussian first-order approximation by the degree of demeaning of the instrument moment matrix $\widetilde{Z}'\widetilde{Z}$. Notice that these finite-sample effects are more prominent for highly persistent regressors that are strongly correlated with the predictive model's innovations. We find that the bias-corrected IVX estimator indeed leads to better size control (see, Kostakis et al. (2015)).

2.5.3 Size Comparison

To conduct a size comparison of the test statistics, we examine the empirical rejection rates of the sup-Wald OLS and sup-Wald IVX tests for detecting single structural change for LUR and MI predictors, under the null hypothesis of no structural change, that is, $\mathbb{H}_0 : \beta_1 = \beta_2$ using a significance level $\alpha = 5\%$. To do this, we generate 5,000 datasets from DGP (4.106) and (4.107) for various values of the model coefficients and compute the frequency of rejecting the null hypothesis. In particular, the increase in the sample size aims to reflect the properties of finite versus large sample asymptotics for the Wald type statistics in testing for a single unknown break-point in predictive regression models with highly persistent regressors.

All tests are run at the 5% nominal (asymptotic) significance level. The simulations were performed in MATLAB[19], version R2020a, using $M = 1000$ Monte Carlo replications and $B = 1000$ Bootstrap replications for both the empirical size and empirical power experiments. Table A1 (for sup Wald-OLS) and Table A2 (for sup Wald-IVX), present the probabilities of rejection of the two Wald-type statistics at the 5% nominal rate, under the null hypothesis. In particular, we consider different values for the exponent rate of the IVX parameter, such as $\delta_z \in \{0.75, 0.95\}$ in order to investigate the varying effect of the degree of persistent of the instrumental variable for detecting structural change in predictive regressions with persistent regressors, as well as different

[19]I grateful acknowledge Michalis P. Stamatogiannis for making the Matlab code of the paper of Kostakis et al. (2015), available on his website. Moreover the R implementation of the IVX procedure in the `IVX package` of Kostas Vasilopoulos has been particularly helpful. The MC simulations were run on Iridis 4, using 16 cores.

localizing coefficient of persistence, $c_i \in \{1, 5, 10, 20\}$. Comparing the empirical size results for Table A1 versus Table A2, we can see that the sup Wald IVX produces values for the empirical size quite close to the nominal size for critical value $c_\alpha = 13.42$, $\delta_z = 0.95$ and $-0.5 \leq \rho \leq 0.5$. The particular critical value is a closer representation to the $\alpha-$quantile from the corresponding limiting distribution.

In Table A1 and Table A2 we present the empirical size under the null hypothesis for the model with a single predictor and no model intercept. We can observe that size distortions appear for larger values of correlation between the u_t and v_t and this is more severe for high persistence regressors (i.e., low values of the coefficient of persistence). All main conclusions are in line with similar findings in the literature of predictive regression models, such as larger size distortions appear as the correlation between u_t and v_t increases, and this effect is more apparent in the case of persistent predictors (i.e., lower values of the persistence coefficient c). Moreover, we see that the sup OLS-Wald statistic with the standard asymptotics as derived by Andrews, is clearly immune to persistence when there is no correlation between the error terms u_t and v_t of the predictive regression. However, the predictive regression model is particularly useful for examining the case of non-zero correlation between the error sequence (u_t, v_t). Specifically, when $\text{Cov}(u_t, v_t) \neq 0$, then under the assumption of nonstationary predictors (LUR), we expect to have size distortions. This is in accordance with the asymptotic theory we proved in the paper, that is, the limiting distribution of the test no longer follows the standard NBB result of Andrews, as it depends on the nuisance coefficient of persistence c_i.

The initial simulation experiments we obtain indicate that the extent of size distortions in finite samples can be considerable if the empirical size is not controlled with a bootstrap methodology, for cases we have a non-standard limiting distribution (e.g, LUR). For instance, we observe that even though the simulated empirical size for the sup IVX-Wald test for detecting a single unknown structural break in the predictive regression model with no model intercept is not quite close to the 8.85 critical value that corresponds to the NBB result of Andrews. Furthermore, from the empirical size experiments across different values of the contemporaneous correlation coefficient, ρ, we can see that the simulated size are quite close when observing at a specific value of c. This is not surprising since the fully modified covariance estimator incorporated in the construction of the covariance matrix for the IVX takes into account the dependence structure of the regressors by applying a common long-run covariance structure. This holds across different values of c_i.

In summary, under mild persistence, such that $\gamma_x \in (0, 1)$, we expect that the sup OLS-Wald test to be properly sized based on the cut-off point from Table 1 of Andrews (1993). Similarly, this findings should also hold in the case of the sup IVX-Wald test as we have proved that the limiting distribution of both tests under the null converges to a NBB. Furthermore, under the assumption of LUR, such that $\gamma_x \in (0, 1)$, using the cut-off that correspond to the NBB, when testing for structural break based on the sup OLS-Wald we expect to have size distortions and this is also the case for the sup IVX-Wald test.

2.5.4 Bootstrap procedure

For robustness of the procedure we use a bootstrap algorithm in order to obtain corresponding critical values. The bootstrapped critical values are estimated based on the Residual Wild Bootstrap (RWB) within the structural break framework. Other approaches include the Fixed Regressor Wild bootstrap (FRWB) which can be found in the predictability literature such as in Georgiev et al. (2021) and Xu (2020). Further frameworks which describe suitable procedures for obtaining bootstrapped critical values are described by Hansen (2000b) and Georgiev et al. (2018) specifically for structural break tests.

Algorithm 1. Two-sided $100 \times (1-\alpha)\%$ Bootstrap sup IVX-Wald test for testing the null hypothesis $\mathbb{H}_0 : \theta_1 = \theta_2$, where $\theta_j = \left(\mu_j, \beta_j'\right)'$.

Consider the predictive regression with multiple predictors. Simulate the following DGP

$$y_t = \mu + x_{t-1}'\beta + u_t, \tag{2.49}$$

$$x_t = R x_{t-1} + v_t, \tag{2.50}$$

where R is a $k \times k$ dimensional coefficient matrix and $u_t = (u_t, v_t')'$ is $(k \times 1)-$dimensional martingale difference sequence with variance-covariance matrix

$$\text{Var}(u_t) = \Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv}' \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix} \tag{2.51}$$

Step 1. Estimation Step

1.1. Fit the predictive regression given by (2.49) to the sample data $(y_t, x_{t-1})'$ to obtain the OLS residuals $\hat{u}_t$, for $t \in \{1, ..., n\}$. Obtain the vector of coefficients $\left\{\hat{\mu}_1, \hat{\mu}_2, \hat{\beta}_1, \hat{\beta}_2\right\}$.

1.2. Fit by OLS the autoregression of order 1, AR(1), given by (2.50) to the vector of predictors x_t, to obtain the OLS residuals $\hat{v}_t$, for $t \in \{1, ..., n\}$. Set $\hat{v}_t = 0$. Obtain the vector of coefficients $\left\{\hat{R}\right\}$.

Step 2. Generate the bootstrap sample $\{y_t^*, x_t^*, t = 1, ..., n\}$.

2.1. Generate the bootstrap innovations $w_t^* = (u_t^*, v_t^{*\prime})' := (\kappa_t \otimes \hat{u}_t, \kappa_t \otimes \hat{v}_t)$, where $\otimes$ denotes element-wise vector multiplication, such that κ_t for $t \in \{1, ..., n\}$, is a scalar sequence independent from the data such that $\kappa_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$.

2.2. Generate x_t^* such that

$$x_{t_b}^* = \hat{R} x_{t_b-1}^* + u_{t_b}^*, \qquad \text{for } t_b \in \{1, ..., T\}, \tag{2.52}$$

with initial conditions $x_0^* = 0$ and $b \in \{1, ..., B\}$.

2.3. Generate the associated bootstrap IVX instrument z_t^* as below

$$z_0^* \quad \text{and} \quad z_t^* = \sum_{j=0}^{t_b-1} \left(I - \frac{C_z}{n}\right)^j \Delta x_{t_b-j}^*, \quad \text{for} \quad t \in \{1, ..., T\}, \tag{2.53}$$

where C_z is the coefficient matrix for the persistence of the bootstrap IVX instruments that contains the same values as the original IVX instruments.

2.4. Generate $y_{t_b}^*$ such that $y_{t_b}^* = \hat{\mu} + x_{t_b-1}^{*\prime}\hat{\beta} + u_{t_b}^*$. Set $y_1^* = y_1$.

2.5. Repeat Steps 2.1 to 2.4 B times (e.g., $B = 1000$).

Step 3. For each bootstrap sample,

$$y_{t_b}^* = \left(\hat{\mu}_1 + x_{t_b-1}^{*\prime}\hat{\beta}_1\right) \mathbf{1}\{t_b \leq k\} + \left(\hat{\mu}_2 + x_{t-1}^{*\prime}\hat{\beta}_2\right) \mathbf{1}\{t_b > k\} + u_{t_b}^* \tag{2.54}$$

calculate the bootstrap IVX-Wald statistic

$$\mathcal{W}_{IVX}^* = \left(\hat{\theta}_j^* - \hat{\theta}_j\right)' \left[\widehat{\mathcal{Q}}_1^* + \widehat{\mathcal{Q}}_2^*\right]^{-1} \left(\hat{\theta}_j^* - \hat{\theta}_j\right),$$

where $\widehat{\mathcal{Q}}_1^*$ and $\widehat{\mathcal{Q}}_2^*$ are computed like $\widehat{\mathcal{Q}}_1$ and $\widehat{\mathcal{Q}}_2$ except that the bootstrap sample $\{y_t^*, x_t^*, z_t^*\}$ is replaced by the original sample $\{y_t, x_t, z_t\}$ such that

$$\mathcal{Q}_1 := \left\{(Z_1'X_1)^{-1}\left(\widetilde{Z}_1'\widetilde{Z}_1\right)_{\text{adj}}(X_1'Z_1)^{-1}\right\} \quad \mathcal{Q}_2 := \left\{\left(\widetilde{Z}_2'X_2\right)^{-1}\left(\widetilde{Z}_2'\widetilde{Z}_2\right)_{\text{adj}}\left(X_2'\widetilde{Z}_2\right)^{-1}\right\}$$

where $X_1 = [1, x_{1t-1}']'$ and $X_2 = [1, x_{2t-1}']'$. Similarly, $\widetilde{Z}_1 = [1, \tilde{z}_{1t-1}']'$ and $\widetilde{Z}_2 = [1, \tilde{z}_{2t-1}']'$. The FM covariance matrix correction is implemented as below:

$$\left(\widetilde{Z}_1'\widetilde{Z}_1\right)_{\text{adj}} = \left(\widetilde{Z}_1'\widetilde{Z}_1\right) \otimes \widehat{\Sigma}_{uu} - k\bar{z}_{1n-1}\bar{z}_{1n-1}' \otimes \widehat{\Omega}_{FM} \tag{2.55}$$

$$\left(\widetilde{Z}_2'\widetilde{Z}_2\right)_{\text{adj}} = \left(\widetilde{Z}_2'\widetilde{Z}_2\right) \otimes \widehat{\Sigma}_{uu} - (T-k)\bar{z}_{2n-1}\bar{z}_{2n-1}' \otimes \widehat{\Omega}_{FM} \tag{2.56}$$

where $\widehat{\Omega}_{FM} = \widehat{\Sigma}_{uu} - \widehat{\Omega}_{uv}\widehat{\Omega}_{uu}'\widehat{\Omega}_{uv}'$.

Step 4. Use the $100(1-\alpha)\%$ quantile $\tau_{\mathcal{W}_{IVX}^*}(1-\alpha)$ of $\mathcal{W}_{IVX}^*$ over B bootstrap replications as the critical value, that is, the hypothesis $\mathbb{H}_0$ is rejected at the significance level α if it holds that $\mathcal{W}_{IVX} > \mathcal{W}_{IVX}^*$. We can also reject the null hypothesis based on the bootstrap p-value, as given by expression (2.57) below.

Remark 2.10. Notice that large sample theory and asymptotic validity for the IVX bootstrap Wald statistic via a bootstrap functional central limit theory is presented in Georgiev et al. (2018) and Georgiev et al. (2021) for the interested reader. The particular asymptotic results allow us to employ the Wild bootstrap procedure to generate critical values for the sup IVX-Wald test for cases in which the limiting distribution of the test does not converge to the standard NBB limit. In those cases (e.g., $\gamma_x \in (0,1)$ and $\gamma_x = 1$) the bootstrapped empirical distribution sufficiently converges to the true limiting distribution of the test.

Therefore, we compute the bootstrap p-value as below

$$p^*\left(\widehat{\tau}_{\mathcal{W}_{IVX}}\right) = \frac{1}{B}\sum_{j=1}^{B} I\left(\widehat{\tau}^*_{\mathcal{W}_{IVX}} > \widehat{\tau}_{\mathcal{W}_{IVX}}\right) \tag{2.57}$$

Under the null hypothesis H_0, as $T \to \infty$, $p-value^* := \mathbb{P}^*\left(\widetilde{\mathcal{W}}^*_{IVX} \leq \widetilde{\mathcal{W}}_{IVX}\right) \xrightarrow{w} U[0,1]$, where $\mathbb{P}^*$ denotes the probability space conditional on the data x and y.

Generally, if we denote with W_n a Wald type statistic, then the bootstrapped counterpart estimates the asymptotic distribution of the original statistic which implies that the convergence $G^*_n(x) := \mathbb{P}^*(W_n \leq x) \to_p G_\infty(x)$ holds. Moreover, the bootstrap consistency result holds

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}^*\left(W^*_n \leq x\right) - \mathbb{P}\left(W^*_n \leq x\right)\right| \to_p 0 \tag{2.58}$$

Then, the bootstrap p-value satisfies $p^*_n := \mathbb{P}^*\left(W^*_n \leq x\right)|_{x=W_n} = G^*_n(W_n) \Rightarrow U[0,1]$. In practise, the bootstrapped p-value given by expression (2.57), $p^*\left(\widehat{\tau}_{\mathcal{W}_{IVX}}\right)$ is the fraction of the bootstrap samples for which $\widehat{\tau}^*_{\mathcal{W}_{IVX}}$ is larger than $\widehat{\tau}_{\mathcal{W}_{IVX}}$. Then, the null hypothesis is rejected when the empirical p-value p^* of the test statistic is smaller than the significance level α.

An important aspect when simulating p-values for a certain test statistic is the assumption of the test being pivotal. However, it can be proved that the bootstrap p-values under H_0 are asymptotically uniformly distributed, and therefore the implementation of the Residual Wild Bootstrap is asymptotically valid in this sense. Since we utilize the above bootstrap procedure (modification of Algorithm 1) to obtain critical values for the sup Wald tests within a Monte Carlo setting then the bootstrap step is replicated in each Monte Carlo simulation. In other words, since asymptotic critical values are not available for the sup IVX-Wald test in the case of persistent or integrated regressors we employ the bootstrap critical values to obtain the empirical size under the null hypothesis of no parameter instability. Practically, since the asymptotic critical values for the sup IVX-Wald test (e.g., given by Theorem 2.2) are not available, then we use the simulated critical values. The regressor wild bootstrap is preferred to the residual wild based bootstrap for example since when the variance error $\text{Var}(\epsilon_i|X_i)$ depends on the value of the predictors X_i (i.e., heteroscedasticity), then the residual bootstrap will be unstable because the residual bootstrap will swap all the residuals regardless of the value of the predictor. Nevertheless, one could also examine the implementation of a residual wild bootstrap (RWB) to obtain the required critical values for the empirical size and power of the tests (e.g., see Cavaliere et al. (2013)).

The choice of the number of Bootstrap replications is another important aspect for consideration for bootstrap based inferences. A small number of artificial samples clearly affects the precision of estimation of p-values under the null hypothesis. In other words, we generate B bootstrap samples of size n where $u^*_{b1}, ..., u^*_{bn}$ is a random sample drawn, with replacement, from an asymptotically valid OLS residual-based EDF. Then, the bootstrap test statistics can be calculated and the p-values of the corresponding actual test statistics can be estimated.

Remark 2.11. Notice also that we consider the case of unconditional weakly convergence under the presence of correlation between the innovations of the predictive regression model. Furthermore, using for example conditional dependence arguments as in Georgiev et al. (2018), the limiting distribution under the null hypothesis for a random break-point $\pi \in [\pi_1, \pi_2]$ is still random. Obviously, when we impose the assumption that the break-point is fixed $\pi = \pi_0$, then even in the case of conditional convergence the limiting distribution of the tests can be shown to be χ^2 (see, the discussion in Georgiev et al. (2018) i.e., Remarks 14 to 17.). On the other hand, since we are particularly interested to evaluate the statistical performance of the IVX-Wald test under the null hypothesis and the assumption of an unknown break-point, then the best we can do is to rely on the corresponding bootstrap distribution. More specifically, Georgiev et al. (2018) discuss that the bootstrap statistics conditional on the data, and the original statistics, conditional on the data, share the same asymptotic distribution (due to weak convergence arguments) e.g., for example in the case of the wild-bootstrap procedure. The particular result has important implications. In other words, even though we can not exactly determine the asymptotic distribution of the sup Wald-IVX test for the case of persistent or integrated predictors, then bootstrapped counterpart statistics ensure both the the presence of asymptotically valid tests as well as they allow to obtain bootstrap p-values regardless of the random-nature of the original limiting distribution as well as the abstract degree of persistence.

The correct use of critical values close to the true asymptotic distribution of the test statistic under examination is crucial in Monte Carlo simulations. This allows to correctly identify the existence of size distortions. In order to decide whether to accept or reject the null hypothesis, one can use already tabulated critical values, only in those cases that the asymptotic distribution of the test statistic under consideration has a known Brownian functional (such as NBB). Therefore, implementing bootstrap inference ensures robust performance of the empirical size and power of the tests. In order to avoid the critical values of the tests to diverge to infinity for any significance level, we choose a trimming parameter which ensures that we do not select an unrestricted full range of values for the break fraction π.

2.6 Empirical Application

Our empirical implementation sheds light on the literature of stock return predictability. Identifying periods of predictability has important implications in various aspects of finance. Related modern reviews of these aspects are presented by Kostakis et al. (2018) and Chinco et al. (2019), among others. However, despite the extensive research of the field, the findings are still rather mixed (Kasparis et al., 2015) (see, Welch and Goyal (2008) for a full discussion). For instance, aspects such as the chosen sample period or the selected predictors can give different conclusions. Additionally, parameter instability due to certain economic events can also affect the reliability of predictability tests. Therefore, it is of paramount importance to develop robust testing methodologies for inferring predictability under conditions such as parameter instability or the presence of nonstationary regressors. Using the predictive regression model studied in this paper[20] our primary focus is to examine the robustness of the proposed tests.

Dataset Description The dataset we employ for the stock return predictability application of the paper, is constructed based on the set of variables considered in the study of Welch and Goyal (2008)[21] which capture economic and financial conditions for the US economy. Thus, we focus on examining the presence of predictability and parameter instability for monthly US stock market excess returns over the sampling period 1990Q1-2019Q4, which includes the economic shocks of the 2008 financial crisis.

Predictant The dependent variable is the monthly equity premium (excess return) of the US stock market based on the S&P500 index. We construct the excess return as in Kasparis et al. (2015), that is, the difference between the total rate of return and the risk-free rate for the same sample period. As a proxy of the US stock market return, we use the value-weighted S&P500 total stock market return including dividends (dividend adjusted returns) using the variables of the particular dataset. Moreover, to account for the risk-free rate we utilize the 3-month T-bill rate obtained from the FRED[22]database. After making these adjustments, we fit the univariate or multivariate predictive regression models with predictors as described below.

Predictors The predictor variables include a set of financial valuation variables, such as: dividend-payout ratio (d/e), long-term yield (lty), dividend yield spread (dy), dividend-price-ratio (d/p), T-bill rate (tbl), earnings-price-ratio (e/p), book-to-market ratio (b/m), default yield spread (dfs), net equity expansion (ntis), term spread (tms) and inflation rate (inf).

---

[20] Alternative model specifications can be considered; for instance a model which considers expected returns in relation to macroeconomic conditions and forecasting uncertainty. A first move towards this direction is presented by Atanasov et al. (2020) who examine consumption fluctuations and expected returns with respect to the predictability literature. More specifically, the authors indeed find statistical evidence of predictability at the one-quarter horizon using the IVX testing approach of KMS.

[21] The dataset can be retrieved from Amit Goyal's website at `http://www.hec.unil.ch/agoyal/`. Detailed descriptions of variables can be found in the Online Appendix of Welch and Goyal (2008).

[22] Time series of macroeconomic variables, such as the US inflation rate and the T-bill rate can be found at `https://fred.stlouisfed.org/`. Notice also that the proxy of equity premium and the other financial variables we consider in this paper, are commonly used in the predictability literature, see Gonzalo and Pitarakis (2012), Kasparis et al. (2015), Kostakis et al. (2015) and Kostakis et al. (2018).

2.6.1 Predictability Tests

We apply the predictability tests based on the the full sample 1946Q1-2019Q4 (Panel A), as commonly done in the literature, using as predictant the S&P500 Equity Premium. Moreover, in this paper we also focus on the sub-sample spanning the period 1990Q1-2019Q4 (Panel B). For both Panel A and B the time series observations correspond to monthly sampling frequency. Our first aim is to examine the stock return predictability of this subsample, that is, to identify the financial variables which are individually statistical significant as well as to identify for evidence of joint statistical significance. Furthermore, our second aim is to employ our proposed parameter instability tests for both the OLS and IVX estimators and compare the results we obtain. In particular Panel B, includes the period of the 2008 financial crisis, so it is natural to assume that certain predictors might exhibit structural break around that economic event. Therefore, this is a suitable sample to assess the statistical validity of our proposed framework for the case of a single structural break. Certain limitations of our approach are on sight, however these do not invalidate our findings. In particular, we do not consider the existence of multiple structural breaks neither we consider sample splitting techniques which can affect the power of the tests especially when using an out-of-sample forecasting scheme.

Table A3 summarizing our findings, which presents predictability tests based on both the OLS estimator and the IVX estimator. Notice that for these set of tests we fit the predictive regression models using the predictors without any detrending or differencing to preserve the unknown persistence properties and have comparability between the two estimators. As we can observe from the estimates on Table A3, using the HAC adjusted t-ratio based on the OLS estimator, indicates that for Panel A both the T-bill rate as well as the Inflation rate have predictive power for the US equity premium, while this is not the case for Panel B. This finding motivates us further to apply our novel Wald type statistics to detect whether any predictors exhibit parameter instability, causing any distortions to statistical inference on the parameters.

Table A4 presents standard structural break tests for the model parameters of the univariate predictive regressions. We implement a set of retrospective structural break tests, under the assumption of stationary time series, by taking the first difference of the predictors before fitting the AR(1) regressions, to avoid any violation of the related econometric assumptions of these test statistics. For instance, it is well-known that the events around the financial crisis of 2008, caused structural breaks in various time series that capture financial and economic conditions. As we can see from the estimates of Table A4, which is based on the dataset of Panel A, both the max-based and exp-based Likelihood Ratio tests indicate statistical evidence of structural breaks in the Dividend payout ratio as well as the earnings-price-ratio, based on the fitted AR(1) model. Therefore, we expect that these variables can affect the robustness of the predictability test, when are included in the model.

2.7 Conclusion

Our objective in this paper, is to provide a unified framework for parameter constancy tests in predictive regression models with nonstationary predictors. More specifically, we propose novel wild bootstap sup-Wald tests to detect parameter instability in predictive regression models with nonstationary predictors. Conducting inference, such as structural break testing, on the regression coefficient of the predictive regression model with multiple highly persistent regressors can lead to a nonstandard limiting distribution. In this paper, we propose a robust econometric framework for structural change testing in predictive regressions with highly persistent predictors as defined via the LUR specification.

In particular, we consider "pure" structural change as it is defined by Andrews (1993), in the sense that the entire parameter vector is subject to structural change under the alternative hypothesis. We have extensively examined the asymptotic theory of tests for parameter instability under the assumption of nonstationary regressors. We compare our results with previous seminal work in the fields of both structural break testing in linear regressions as well as predictability testing in predictive regression models. We find some interesting results not previously presented in the literature. Firstly, using the OLS estimator for the parameters of the predictive regression model we show that the limiting distribution of the Wald statistic for testing for a single structural break has a nonstandard limiting distribution which depends on the unknown coefficient of persistence. Secondly, by employing the IVX estimator proposed in the literature as a robust estimator which filters out the abstract degree of persistence in regressors, we have proved that the limiting distribution of the parameter constancy test is also nonstandard but in some special cases of the parameter space of γ_x, this simplifies to the NBB result. Finally, we prove that the sup OLS-Wald and sup IVX-Wald are equivalent in the case of mildly integrated predictors. However, for the case of highly persistent predictors, which is the case of interest in this paper, the two statistics have a different limiting distribution.

We evaluate Wald type test statistics that asymptotically converge to a nonstandard limiting distribution under the null hypothesis. The implementation of a fixed regressor bootstrap is computational feasible and since the econometrician is agnostic regarding the persistence properties of regressors, it allows to obtain asymptotic approximations of the distributions of the test statistics under the null hypothesis. Furthermore, the structure of the predictive regression model, can be proved that produces valid bootstrap inference as it is shown in the framework of Georgiev et al. (2021). The particular conjecture implies a weak convergence argument which allows to use the bootstrap asymptotic critical values even though the presence of the nuisance parameter of persistence in both the original limiting distribution and the corresponding bootstrap distribution under the null hypothesis.

Further aspects related to the IV based approach of KMS that we follow in this paper, can be also examined within the above parameter instability testing procedures. Recent applications include for instance, Gungor and Luger (2020) who consider predictability tests with GARCH-type effects (see also, Magdalinos (2020)) and Yang et al. (2020) who consider a modification of the KMS test that accounts for serial correlation in the error term of the linear predictive regression. Moreover, Pang et al. (2020) propose testing methodologies for multiple structural breaks under the presence of nonstationary predictors.

Chapter 3

Detecting Structural Breaks in Quantile Predictive Regressions

Abstract

We propose an econometric environment for structural break detection in nonstationary quantile predictive regressions. We establish the limit distributions for a class of Wald and fluctuation type statistics based on both the ordinary least squares estimator and the endogenous instrumental regression estimator proposed by Phillips and Magdalinos (2009). Although the asymptotic distribution of these test statistics appears to depend on the chosen estimator, the IVX based tests are shown to be asymptotically nuisance parameter-free regardless of the degree of persistence and consistent under local alternatives. The finite-sample performance of both tests is evaluated via simulation experiments. An empirical application to house pricing index returns demonstrates the practicality of the proposed break tests for regression quantiles of nonstationary time series.

3.1 Introduction

Time series predictability is a research question that sparked the development of various methodologies for estimation and inference in predictive regression models. Related studies include Jansson and Moreira (2006), Mikusheva (2007), Phillips (2014), Cai and Wang (2014), Kostakis et al. (2015), Kasparis et al. (2015), Gonzalo and Pitarakis (2012, 2017), Ren et al. (2019), Demetrescu et al. (2020), Yang et al. (2020), Georgiev et al. (2021), Andersen and Varneskov (2021), Harvey et al. (2021b) as well as Dou and Müller (2021) among others[1]. These frameworks operate under the assumption of a stable relation between the predictant and the predictors of the model. However, the possible presence of parameter instability under these conditions require a different treatment (e.g., see Pitarakis (2017) and Georgiev et al. (2018)). Thus, to derive limits for structural break tests suitable for predictive regressions, necessitate to employ certain regularity conditions and invariance principles of partial sum processes as in Phillips and Magdalinos (2007), Phillips and Magdalinos (2009), to obtain stochastic integral approximations.

Most of the current literature focuses on detecting structural break in the conditional mean of the distribution of $f(y_t|\boldsymbol{x}_t)$ or $f(y_t|\boldsymbol{x}_{t-1})$ where $\boldsymbol{x}_t$ is assumed to follow a near unit root process such as in Cai et al. (2015), Georgiev et al. (2018) and leaving the conditional quantile distribution vastly unexplored. Consequently, this paper addresses these issues by developing a framework suitable for structural break testing for quantile regressions under nonstationarity which has not seen much attention. Specifically, we build on related studies such as Lee (2016) and Fan and Lee (2019) who propose a framework for estimation and inference in quantile predictive regressions as well as the study of Qu (2008) who focus on break testing procedures for regression quantiles. The latter approach studies a linear quantile model with stationary covariates, while we consider the nonstationary quantile predictive regression model with persistent covariates as in Lee (2016). Another related study is presented by Aue et al. (2017) in the context of piecewise quantile autoregressions. Therefore, proposing tests which bridge the gap between these two approaches is a further development of the current toolkit while the local-to-unity theory proposed by Phillips (1987a) and Phillips and Perron (1988) can facilitate the asymptotic theory.

Consider the $\tau-$th conditional quantile of y_t which is defined as following

$$\mathsf{Q}_y\left(\tau|x\right) = F^{-1}_{y|x}(\tau|x) := \inf\{\mathsf{s} : F_{y|x}(\mathsf{s}|x) \geq \tau\} \tag{3.1}$$

Thus, using the conditional quantile function (3.1) proposed by Kiefer (1967), to be the specification form for the predictive regression model then the modelling environment permits to investigate the presence of quantile predictability. The parameter vector of the predictive regression model is quantile dependent for some fixed quantile level within a compact set and can be estimated as the unique solution of the following unconstrained optimization problem (see, Koenker and Portnoy (1987) and Portnoy (1991))

$$\underset{\boldsymbol{b}\in\mathbb{R}^p}{\arg\min} \sum_{t=1}^{n} \rho_\tau(y_t - \boldsymbol{x}'_{t-1}\boldsymbol{b}) \tag{3.2}$$

[1] The problem of distorted statistical inference in predictive regression models under nearly integrated predictors has been also reported in previous studies such as Elliott and Stock (1994), Elliott (2011), Campbell and Yogo (2006) and references therein.

where $\rho_\tau(u) \equiv u\,(\tau - \mathbb{1}\{u < 0\})$, is the check function as defined by Koenker and Bassett (1978), $\boldsymbol{b}$ is some parameter vector and $\boldsymbol{x}_{t-1}$ is the lagged regressor of the model. Our research objective is on the asymptotic behaviour of the parameter estimators of $\boldsymbol{b}$ as well as related test statistics about the parameter vector under the null hypothesis of no parameter instability. Under the alternative hypothesis there exists a structural break at an unknown break-point location.

The structural break literature has various applications for different modelling conditions, however there are currently limited studies related to break testing in quantile time series models even under the assumption of stationarity and ergodicity such as in the seminal paper of Andrews (1993) for linear models. Furthermore, when considering predictive regression models inference regarding structural break involves deriving nonstandard asymptotic theory due to the presence of the nuisance coefficient of persistence. More precisely, testing based on conventional estimation methods (such as OLS-based tests) has been proved to have size distortions for persistent regressors (e.g., see Georgiev et al. (2018) and as in Chapter 2 of the thesis). Our study is considered as a unified framework for structural break detection in quantile predictive regressions (see, Lee (2016) and Fan and Lee (2019)) which encompasses regressor properties such as high persistent, mildly integrated or stationary under certain parameter space restrictions on the persistence parameters. In particular, the proposed tests can statistically evaluate for breaks in the model coefficients at both fixed and multiple quantiles of the underline conditional quantile distribution. We investigate the asymptotic theory and implementation of both Wald type statistics as in Andrews (1993) as well as fluctuation type statistics as in Qu (2008) but within the setting of nonstationary quantile predictive regressions.

Our contributions are threefold. Firstly, we study the estimation problem of quantile predictive regressions with multiple predictors assumed to be generated as either near unit root or mildly integrated processes. We construct quantile predictability Wald statistics under the null of no predictability and derive the asymptotic distributions (as in Lee (2016)). Secondly, we propose a structural break testing procedure for quantile predictive regressions which permits testing for the presence of breaks at the tails. Thirdly, we examine the statistical performance of these tests for detecting parameter instability in the coefficients of quantile predictive regressions with extensive simulation experiments of empirical size. Our distribution theory is based on a double indexed empirical process (see, Caner and Hansen (2001)) with a weakly convergence to a two-parameter Brownian motion within the Skorokhod topology. Under the null hypothesis of no structural break the specific weakly convergence argument permits to establish convergence to stochastic integral approximations defined with respect to a two-parameter process; limit results employed to derive the asymptotic distributions of fluctuation type statistics. Similarly, for the Wald type statistics we employ the asymptotic theory developed by Phillips and Magdalinos (2009) and Lee (2016). Thus, we introduce invariance principles for partial sum processes of matrix moments based on these functionals to obtain asymptotic results for the proposed econometric environment.

We assume that all random elements are defined within a probability space denoted with the triple $(\Omega, \mathcal{F}, \mathbb{P})$. All limits are taken as $n \to \infty$, where n is the sample size. Denote with $\mathcal{D}([0,1])$ to be the set of functions on $[0,1]$ that are right continuous and have left limits, equipped with the Skorokhod metric. Then, the symbol " $\Rightarrow$ " is used to denote the weak convergence of the associated probability measures as $n \to \infty$. The symbol $\overset{\mathcal{D}}{\to}$ and $\overset{\mathbb{P}}{\to}$ are employed to denote convergence in distribution and convergence in probability respectively.

Outline Chapter 3 is organized as follows. Section 3.2 introduces the quantile predictive regression model along with main assumptions, the estimation methodology as well as the testing hypotheses of interest. Section 3.3 presents the testing procedure and asymptotic theory of structural break testing for a fixed quantile level. Section 3.4 presents the testing procedure and asymptotic theory of structural break testing across multiple quantile levels. Section 3.5 investigates the finite sample performance of the proposed tests via Monte Carlo experiments. Section 3.6 illustrates the implementation of of the proposed testing procedures with an empirical application. Section 3.7 concludes. Proofs of main limit results can be found in Appendix B.

3.1.1 Literature Review

Estimation and inference with quantile models has been proposed to the literature with the seminal work of Koenker and Bassett (1978) and Koenker and Bassett (1982). In particular, the uniform Bahadur-type representation established by Koenker and Portnoy (1987) as well as the representation of quantiles proposed by Knight (1989) (see also Koul and Saleh (1995)) are commonly used to derive limit distributions for estimators and test statistics. Several papers in the literature follow these methodologies that examine related aspects to the proposed estimation environment and testing procedures.

Firstly, the problem of structural change for regression quantiles include the studies of Su and Xiao (2008) and Qu (2008). Both frameworks develop diagnostic tools for break detection in quantile regression models under the assumption of stationary covariates. In the former case the alternative hypothesis of a single break is formulated with respect to the magnitude of the break-point. In the latter case, the author also propose a multiple break point testing procedure. Additionally, Furno (2014) implements these quantile regression based statistics using the methodology proposed by Chow (1960) which implies testing for break at a known location in the sample.

A different perspective is presented by Hoga (2018) who consider detecting for breaks using tail dependence measures (see also Hoga (2017)[2]) based on an empirical estimator of extremal dependence. An extension of the method to an alternative hypothesis with multiple breaks is also examined. All aforementioned procedures correspond to structural break tests suitable for quantile models[3] while in our study we focus on implementing testing procedures specifically for quantile predictive regression models in which the time series properties of predictors are modelled with the nuisance parameter of persistence. In particular, a separate autoregressive model with a local unit root coefficient matrix is used to model the unknown degree of persistence which implies that conventional approaches for deriving large sample theory are no longer valid. However, the fluctuation based tests implemented by Qu (2008) for break detection in quantile models with stationary covariates provides a suitable econometric environment for investigating the effect of nonstationarity to the asymptotic distribution of the tests.

[2] The framework proposed by Hoga (2017) considers a set of change point tests for the tail index of random variables and contributes to the literature of nonparametric modelling methods.

[3] Notice that the literature of break tests presented here differs from the literature of slope heterogeneity in quantile models. Related aspects to inference are Koenker and Hallock (2001), Chernozhukov (2005), Portnoy (2012) and Escanciano and Goh (2018) which are of independent interest.

Secondly, asymptotic theory for quantile time series regressions has been examined by Koenker and Xiao (2002, 2004, 2006) in the context of autoregressive model specifications and unit root testing[4]. Further applications in the time series literature include testing procedures for threshold effects under the assumption of a conditional quantile function as in the studies of Galvao et al. (2011, 2014) and the case of nonstationary (nonlinear) quantile regressions by Xiao (2009), Cho et al. (2015), Li et al. (2016) and Uematsu (2019). Moreover, Kato (2009) derives related limit results which are useful when considering the asymptotic behaviour of quantile estimators for a wide class of modelling approaches and econometric conditions. All these studies cover both stationary and nonstationary models however they operate under the null hypothesis of no parameter instability in model parameters. Therefore, our proposed testing procedure, is considered to be a novel contribution to the time series econometrics literature.

Thirdly, the proposed structural break testing methodology is closely related to the problem of unidentified parameters under the null hypothesis which is well-known in the econometrics and statistics literature as the Davies problem (see, Davies (1977)). The particular aspect which is relevant to parameter admissibility has been investigated in problems of estimation and testing such as in the studies of Hansen (1996), Andrews and Ploberger (1994), Pitarakis (2004) and Elliott et al. (2015). A more recent approach to the unidentified parameter problem under the null is presented by McCloskey (2017)[5]. Therefore, to overcome this challenge, we employ the supremum operator when constructing Wald type statistics. Furthermore, a nonparametric estimation approach for quantile regressions such in the study of Qu and Yoon (2015), may require different moment conditions for estimation and inference which is beyond the scope of this paper.

Lastly, in terms of the model structure of the proposed modelling approach of the paper, we impose standard econometric assumptions and conditions in the literature of predictive regressions. More precisely, by imposing a standard martingale difference condition on the equation innovations u_t, this implies an orthogonality condition between the innovations and the model regressors, such that $\mathsf{Cov}(u_t, \boldsymbol{x}_t) = \mathbb{E}\big[\boldsymbol{x}_t \mathbb{E}\left(u_t | \mathcal{F}_{t-1}\right)\big] = 0$. Furthermore, the model structure we follow does not allow for endogeneity even though the IVX instrumentation method implies the construction of endogenous instruments. Specifically, as explained by Phillips and Magdalinos (2009) and Kostakis et al. (2015), the IVX filtration implies the construction of instrumental covariates based on information obtained only from the regressors of the model. According to (Wang and Phillips, 2012, p. 731) the model structure would permit for the presence of endogeneity when the equation error could be serially dependent and cross-correlated under certain moment restrictions (see also Yang et al. (2020)[6]). However, the particular aspect can complicate the derivation of the limit distributions as additional considerations will be needed, such as to incorporate conditional heteroscedasticity, which is beyond the scope of our study and we leave as future research related to the proposed framework.

[4]A unified framework for econometric inference with nearly integrated regressors and unit roots is proposed by the seminal work of Phillips (1988a), Phillips (1987b). Further relevant literature includes the study of Stock (1994) who discuss aspects related to testing with unit roots and structural breaks in time series models as well as Chapter 14 in Davidson (2000) that has additional derivations and examples.

[5]Specifically, McCloskey (2017) proposes a framework for a set of flexible size-corrected critical values construction methods that lead to tests with correct asymptotic size and desirable power properties in testing problems with nuisance parameter under the null hypothesis.

[6]Specifically, Yang et al. (2020) propose a unified IVX-AR Wald statistic that accounts for serial correlation in the error terms of the linear predictive regression model. The IVX-AR estimator corrects the size distortions arising from serially correlated error under the presence of high persistence.

3.2 Econometric Environment and Testing Problem

3.2.1 Econometric Model and Assumptions

Consider the (linear) predictive regression model

$$y_t = \alpha + \boldsymbol{\beta}' \boldsymbol{x}_{t-1} + u_t, \qquad 1 \leq t \leq n, \tag{3.3}$$

$$\boldsymbol{x}_t = \boldsymbol{R}_n \boldsymbol{x}_{t-1} + \boldsymbol{v}_t \tag{3.4}$$

where $y_t \in \mathbb{R}^{n \times 1}$ is a scalar dependent variable and $\boldsymbol{x}_{t-1} \in \mathbb{R}^{n \times p}$ is a $p-$dimensional vector of predictors such that $\boldsymbol{x}_t$ is generated as a near unit root process (or local unit root process) with an autocorrelation coefficient matrix as defined by the studies of Phillips and Magdalinos (2009) and Kostakis et al. (2015) expressed as below

$$\boldsymbol{R}_n = \left(\boldsymbol{I}_p + \frac{\boldsymbol{C}_p}{n^{\gamma_x}} \right), \quad \text{for some } \gamma_x > 0. \tag{3.5}$$

where n is the sample size, and $\boldsymbol{C}_p = \mathsf{diag}\left\{ c_1, ..., c_p \right\}$ is a $p \times p$ diagonal matrix with the coefficients of persistence c_i for $i = 1, ..., p$. We consider that the predictors of the model are allowed to belong only to one of the two degree of persistence as specified below

- *Local Unit Root* (*LUR*): $\gamma_x = 1$ and $c_i \in (-\infty, 0), \ \forall \ i = 1, ..., p$.
- *Mildly Integrated* (*MI*): $\gamma_x \in (0, 1)$ and $c_i \in (-\infty, 0), \ \forall \ i = 1, ..., p$.

Let $\mathcal{F}_t$ denote the natural filtration, then for the error term of the predictive regression we assume that $\mathbb{E}\left(u_t | \mathcal{F}_{t-1} \right) = 0$ and $\mathbb{E}\left(u_t^2 | \mathcal{F}_{t-1} \right) = \sigma^2_{uu}$. Specifically, the innovation structure of the predictive regression model allows to impose a linear process dependence for $\boldsymbol{v}_t$, with a conditionally homoscedastic *martingale difference sequence* condition such that

$$\boldsymbol{v}_t = \sum_{j=1}^{\infty} \boldsymbol{\varphi}_j \boldsymbol{\varepsilon}_{t-j}, \quad \boldsymbol{\varepsilon}_t \sim mds\left(\boldsymbol{0}, \boldsymbol{\Sigma} \right),$$

with necessary conditions for the linear process representation to hold given as below

$$\boldsymbol{\Sigma} > 0, \qquad \sum_{j=0}^{\infty} j \left\| \boldsymbol{\varphi}_j \right\| < \infty \ \text{ such that } \ \boldsymbol{\varphi}_o(\mathsf{z}) = \sum_{j=0}^{\infty} \mathsf{z}^j \boldsymbol{\varphi}_j. \tag{3.6}$$

Denote with $\boldsymbol{e}_t = (u_t, \boldsymbol{v}_t')'$, then under regularity conditions, the following invariance principle (*FCLT*) holds (see Phillips and Solo (1992))

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{e}_t := \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} \begin{bmatrix} u_t \\ \boldsymbol{v}_t \end{bmatrix} \equiv \begin{bmatrix} B_{un}(r) \\ \boldsymbol{B}_{vn}(r) \end{bmatrix} \Rightarrow \begin{bmatrix} B_u(s) \\ \boldsymbol{B}_v(s) \end{bmatrix} := \mathcal{BM} \begin{bmatrix} \sigma^2_{uu} & \boldsymbol{\sigma}'_{uv} \\ \boldsymbol{\sigma}_{vu} & \boldsymbol{\Sigma}_{vv} \end{bmatrix}_{(p+1) \times (p+1)} \tag{3.7}$$

where $\boldsymbol{\Sigma}_{vv} \in \mathbb{R}^{p \times p}$ is a positive definite covariance matrix and $0 < r < 1$.

Specifically, the individual components of the vector sequence $\boldsymbol{e}_t = (u_t, \boldsymbol{v}_t')'$ have partial sums processes that weakly converge into their Brownian motion counterparts as below

$$B_{un}(r) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} u_t \Rightarrow B_u(r) := \mathcal{N}(0, r\sigma_{uu}^2) \tag{3.8}$$

$$\boldsymbol{B}_{vn}(r) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{v}_t \Rightarrow \boldsymbol{B}_v(r) := \mathcal{N}(\boldsymbol{0}, r\boldsymbol{\Sigma}_{vv}) \tag{3.9}$$

where $\boldsymbol{B}(r) = (B_u(r), \boldsymbol{B}_v(r)')'$ being a $(p \times 1)$ Brownian motion with long-run covariance matrix $\boldsymbol{\Sigma}_{ee}$, that is, a Gaussian vector process with almost surely continuous sample paths. More precisely, since $\boldsymbol{x}_t$ is an adapted process to $\mathcal{F}_t$ then in practise there exists a correlated vector Brownian motion $\boldsymbol{B}_n(r) = (B_{un}(r), \boldsymbol{B}_{vn}(r)')'$ such that

$$\left(\frac{1}{\sqrt{n}} \sigma_{uu}^{-1} \sum_{t=1}^{\lfloor nr \rfloor} u_t, \frac{1}{\sqrt{n}} \boldsymbol{\Sigma}_{vv}^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{v}_t \right)' \Rightarrow \boldsymbol{B}(r) = (B_u(r), \boldsymbol{B}_v(r)')', \quad 0 < r < 1 \tag{3.10}$$

on $\mathcal{D}\left([0,1]\right)^2$ as $n \to \infty$, with covariance matrix as in (4.25) implying joint convergence. Under the above conditions hold, then the following invariance principle holds (see, Phillips (1987a))

$$\frac{\boldsymbol{x}_{\lfloor nr \rfloor}}{\sqrt{n}} \Rightarrow \boldsymbol{J}_c(r), \quad \text{where} \quad \boldsymbol{J}_c(r) = \int_0^r e^{(r-s)\boldsymbol{C}_p} d\boldsymbol{B}_v(s). \tag{3.11}$$

The functional $\boldsymbol{J}_c(r)$ represents the *Ornstein-Uhlenbeck* process[7] which is employed to derive stochastic integral approximations for the nonstationary predictive regression model (4.11)-(4.12). Denote with $\boldsymbol{K}_c(r) := \boldsymbol{\Sigma}_{vv}\boldsymbol{J}_c(r)$, where $\boldsymbol{K}_c(r)$ is a $p-$dimensional Gaussian process defined as $\boldsymbol{K}_c(r) = \int_0^r e^{(r-s)\boldsymbol{C}_p} d\boldsymbol{B}_v(s)$ is the solution of *Black-Scholes* differential equation $d\boldsymbol{K}_c(r) \equiv c\boldsymbol{K}_c(r) + d\boldsymbol{B}_v(r)$, with $\boldsymbol{K}_c(r) = 0$ as the initial condition.

The specific autoregression matrix specification (3.5) allows to examine other persistence properties such as unit root processes, when $c_i = 0$ for all $i \in \{1, ..., p\}$, or explosive processes when $c_i > 0$. We consider two types of nonstationarity, that is, the near unit or high persistent regressors and the mildly integrated regressors. In both cases the coefficient of persistent, c_i, lies below the unit root boundary. However, the main difference between these two persistence classes is that the mildly integrated regressors have an exponent rate below the unit boundary, which implies that these regressors are less persistent than regressors generated from near unit root processes. One can consider extending our estimation and testing framework to the case of explosive regressors; we leave this aspect for future research. The proposed framework develops the testing procedure and relevant asymptotic theory for structural break testing in quantile predictive regressions while allowing for an endogenous predictive regressor with any degree of persistence. To evaluate the performance of the proposed sup-Wald tests we use Monte Carlo resampling techniques to control the familywise error rate in finite samples.

[7]The OU process is a stationary Gaussian process with an autocorrelation function that decays exponentially over time. Moreover, the continuous time OU diffusion process has a unique solution and this property allows to approximate asymptotic terms for estimators and corresponding test statistics as a function of the nuisance parameter of persistence (see, Perron (1991)).

Quantile Predictive Regression Model

Our main goal is to investigate the asymptotic theory and empirical implementation of structural break tests suitable for quantile predictive regression models. Therefore, we consider suitable modifications of the innovation structure that corresponds to the linear predictive regression model, following standard conditions and assumptions employed for quantile time series models, currently presented in literature. More precisely, the conditional quantile function of y_t denoted with $\mathsf{Q}_{y_t}(\tau|\mathcal{F}_{t-1})$, replaces the conditional mean function of the predictive regression which implies the following model specification

$$\mathsf{Q}_{y_t}(\tau|\mathcal{F}_{t-1}) := F^{-1}_{y_t|\boldsymbol{x}_{t-1}}(\tau) \equiv \alpha(\tau) + \boldsymbol{\beta}(\tau)'\boldsymbol{x}_{t-1}. \tag{3.12}$$

such that $F_{y_t|\boldsymbol{x}_{t-1}}(\tau) := \mathbb{P}(y_t \leq \mathsf{Q}_{y_t}(\tau|\mathcal{F}_{t-1})\,|\mathcal{F}_{t-1}) \equiv \tau$, where $\tau \in (0,1)$ is some quantile level in the compact set $(0,1)$. Therefore, in order to define the innovation structure that corresponds to the quantile predictive regression, we employ the piecewise derivative of the loss function such that $\psi_\tau(\mathsf{u}) = \left[\tau - \mathbb{1}\{\mathsf{u} < 0\}\right]$. Consequently, this implies that $u_t(\tau) := u_t - F_u^{-1}(\tau)$ where $F_u^{-1}(\tau)$ denotes the unconditional $\tau-$quantile of the error term u_t. Then, the corresponding invariance principle for the nonstationary quantile predictive regression model is formulated as below

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr \rfloor}\begin{bmatrix}\psi_\tau(u_t(\tau))\\ \boldsymbol{v}_t\end{bmatrix} \Rightarrow \begin{pmatrix}B_{\psi_\tau}(r)_{(1\times n)}\\ \boldsymbol{B}_v(r)_{(p\times n)}\end{pmatrix} \equiv \mathcal{BM}\begin{bmatrix}\tau(1-\tau) & \boldsymbol{\sigma}'_{\psi_\tau v}\\ \boldsymbol{\sigma}_{v\psi_\tau} & \boldsymbol{\Omega}_{vv}\end{bmatrix} \tag{3.13}$$

Assumption 3.1. The following conditions for the innovation sequence hold:

(i) The sequence of stationary conditional *probability distribution functions (pdf)* denoted with $\{f_{u_t(\tau),t-1}(.)\}$ evaluated at zero with a non-degenerate mean function such that $f_{u_t(\tau)}(0) := \mathbb{E}\left[f_{u_t(\tau),t-1}(0)\right] > 0$ satisfies a *FCLT* given as below

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr \rfloor}\left(f_{u_t(\tau),t-1}(0) - \mathbb{E}\left[f_{u_t(\tau),t-1}(0)\right]\right) \Rightarrow B_{f_{u_t(\tau)}}(r). \tag{3.14}$$

(ii) For each t and $\tau \in (0,1)$, $f_{u_t(\tau),t-1}(.)$ is uniformly bounded away from zero with a corresponding conditional distribution function $F_t(.)$ which is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$ (see, Neocleous and Portnoy (2008), Goh and Knight (2009), Lee (2016) and Fan and Lee (2019)).

Remark 3.1. Assumption 3.1 ***(i)*** provides a standard weak convergence argument to a Brownian motion process that corresponds to the underline distribution generating the innovation sequence of the quantile predictive regression model (see, Lee (2016) and Fan and Lee (2019)). Furthermore, Assumption 3.1 ***(ii)*** provides the weak convergence argument for the sparsity function of the model to its Brownian motion counterpart for some $0 < r < 1$.

3.2.2 Estimation Methodology

We investigate the statistical properties of the estimation methodology in relation to the handling of the nuisance parameter of persistence. We derive the asymptotic distributions of the associated test statistics based on two optimization methods which are known to have different convergence rates in the time series predictability literature.

Method A: OLS based estimation

We consider the OLS based estimation using the check function $\rho_\tau(.)$, which is common practise for optimization problems of quantile series. Specifically, the asymptotic behaviour of the quantity $\boldsymbol{\mathcal{E}}_n(\tau) \equiv \sqrt{n}(\boldsymbol{\beta}_n(\tau) - \boldsymbol{\beta}_0(\tau))$ is of interest. The traditional approach to asymptotics for $\hat{\boldsymbol{\beta}}(\tau)$ is to employ a Bahadur representation which allows to decompose the expression into a Brownian bridge component and an error term (see, Portnoy (2012) and Kato (2009)). Furthermore, various studies are concerned with the determination of sharp error bounds for the specific error term. However in our setting, $\sqrt{n}-$consistent asymptotics do not always apply due to the presence of nonstationarity. Additionally, the chosen estimator affects the stochastic rates of convergence.

Similar to the linear predictive regression (a model with a conditional mean functional form), under the assumption of persistent regressors, the OLS estimator has been proved to be biased due to the presence of nuisance parameters (e.g., see Campbell and Yogo (2006)), resulting to distorted statistical inference[8]. Nevertheless, focusing on the two persistence properties we introduced previously, we derive its limit distribution which is useful when considering structural break tests under nonstationarity.

Denote with $\boldsymbol{\theta}(\tau) = \left[\alpha(\tau), \boldsymbol{\beta}(\tau)'\right]' \in \mathbb{R}^{(p+1)\times 1}$ and $\boldsymbol{X}_{t-1} = \left(\mathbf{1}, \boldsymbol{x}_{t-1}'\right)' \in \mathbb{R}^{(p+1)\times n}$, then the OLS based estimator is obtained by solving the following optimization problem

$$\widehat{\boldsymbol{\theta}}_n^{qr}(\tau) := \underset{\boldsymbol{\theta}\in\mathbb{R}^{p+1}}{\text{arg min}} \sum_{t=1}^{n} \rho_\tau\left(y_t - \boldsymbol{X}_{t-1}'\boldsymbol{\theta}\right) \tag{3.15}$$

where $\rho_\tau(u) = u\left(\tau - \mathbb{1}\{u < 0\}\right)$ with $\tau \in (0,1)$, represents the asymmetric quantile regression function. Following Lee (2016), we use the normalization matrices below which are different according to the persistence properties of predictors such that

$$\boldsymbol{D}_n := \begin{cases} \text{diag}\left(\sqrt{n}, n\boldsymbol{I}_p\right) & \text{for } LUR, \\ \text{diag}\left(\sqrt{n}, n^{\frac{1+\gamma_x}{2}}\boldsymbol{I}_p\right) & \text{for } MI. \end{cases} \tag{3.16}$$

Then, Corollary 3.1 summarizes the asymptotic distribution of the OLS-QR estimator (see also Theorem 2.1 Lee (2016)) for mildly integrated and high persistent regressors.

[8] Relevant studies in the literature which examine the asymptotic behaviour of standard $t-$tests under these conditions include Phillips and Lee (2013, 2016), Lee (2016), Fan and Lee (2019), Kostakis et al. (2015) and Breitung and Demetrescu (2015) among others.

Corollary 3.1. Under Assumption 3.1 and *FCLT* (4.25)-(4.25) it follows that: $\boldsymbol{D}_n\left(\widehat{\boldsymbol{\theta}}_n^{qr}(\tau)-\boldsymbol{\theta}(\tau)\right)$

$$\Rightarrow \begin{cases} f_{u_t(\tau)}(0)^{-1}\begin{bmatrix} 1 & \int_0^1 \boldsymbol{J}_c(r)'dr \\ \int_0^1 \boldsymbol{J}_c(r)dr & \int_0^1 \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)'dr \end{bmatrix}^{-1}_{(p+1)\times(p+1)} \begin{bmatrix} B_{\psi_\tau}(1)_{(1\times n)} \\ \int_0^1 \boldsymbol{J}_c(r)dB_{\psi_\tau}dr_{(p\times n)} \end{bmatrix} & \textit{LUR}, \\ \mathcal{N}\left(0, \dfrac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2}\begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx}^{-1}\end{bmatrix}_{(p+1)\times(p+1)}\right) & \textit{MI}. \end{cases}$$

where the stochastic matrix $\boldsymbol{V}_{xx}$ is defined by the following expression

$$\boldsymbol{V}_{xx} := \int_0^\infty e^{r\boldsymbol{C}_p}\boldsymbol{\Omega}_{xx}e^{r\boldsymbol{C}_p}dr, \text{ where } \boldsymbol{\Omega}_{xx} := \sum_{m=-\infty}^{\infty}\mathbb{E}\left(\boldsymbol{v}_t\boldsymbol{v}_{t-m}'\right) = \boldsymbol{\varphi}_o(1)\boldsymbol{\Sigma}\boldsymbol{\varphi}_o(1)'.$$

Consequently, the limiting joint distribution of the model intercept and slopes for the nonstationary quantile predictive regression model under the assumption of mildly integrated regressors, is a mixed normal of the form $\mathcal{MN}(0,\boldsymbol{\Sigma}^\star)$, where

$$\boldsymbol{\Sigma}^\star := \frac{\tau(1-\tau)}{f_{u_t(\tau_0)}(0)^2}\begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx}^{-1}\end{bmatrix}, \quad \text{for some } \tau\in(0,1). \tag{3.17}$$

while under high persistence the asymptotic behaviour of $\boldsymbol{D}_n\left(\widehat{\boldsymbol{\theta}}_n^{qr}(\tau)-\boldsymbol{\theta}(\tau)\right)$ depends on functionals of OU processes which are more challenging to approximate, especially if one is interested to obtain sharp error bounds[9] as in Portnoy (2012).

Furthermore, we assume that the sparsity coefficient can be consistently estimated with a fixed unbiased estimator for all t. This a strong assumption which can be regarded as a trade-off relation between the complicated objective function (i.e., non-differentiable and nonstationary) and the tractable error term of the model (Uematsu, 2019). On the other hand, it permits to consider the limiting distributions when testing for a set of parameter restrictions. Thus, to overcome the problem of nonstandard statistical inference due to the presence of the nuisance coefficient of persistence, we employ the instrumental variable regression approach proposed by Phillips and Magdalinos (2009).

The estimator of Phillips and Magdalinos (2009) performs reasonably well in finite samples and even performs better that the OLS counterpart (see Georgiev et al. (2021)), demonstrating the robustness of the method in filtering out abstract degree of peristence when testing for linear restrictions in predictive regressions. In addition, the suggested instrumental variable approach is by definition neither spurious, since it is always correlated with the corresponding regressor, not a poor instrument, because the correlation of unit root processes tends to one asymptotically.

[9]In particular, the study of Portnoy (2012) obtains a near $\sqrt{n}-$consistent error bound by employing the "Hungarian construction" which requires to approximate the quantity $\boldsymbol{\mathcal{E}}_n(\tau)$ using a Brownian bridge limit which converges to this non-zero Gaussian process with an appropriate rate of convergence.

Method B: IVX based estimation

The endogenous instrumentation (IVX) procedure for predictive regression models[10] proposed by Phillips and Magdalinos (2009) implies the use of a mildly integrated instrumental variable. The instrumented variable is constructed as below

$$\tilde{\boldsymbol{z}}_t = \sum_{j=0}^{t-1} \left(\boldsymbol{I}_p + \frac{\boldsymbol{C}_z}{n^{\gamma_z}} \right) (\boldsymbol{x}_{t-j} - \boldsymbol{x}_{t-j-1}), \tag{3.18}$$

where $\boldsymbol{C}_z = \mathsf{diag}\{c_{z1}, ..., c_{zp}\}$ is a $p \times p$ diagonal matrix such that $c_{zj} < 0 \; \forall \; j \in \{1, ..., p\}$ with $0 < \gamma_z < 1$, where γ_z is the exponent rate of the persistence coefficient of the instrumental variable, such that $\gamma_z \neq \gamma_x$. The IVX filtering methodology transforms a nonstationary autoregressive process which is assumed to generate the regressors, $\boldsymbol{x}_t$, and can encompass both stable or unstable processes based on the behaviour of the local unit root coefficient, into a mildly integrated process which is less persistent than the endogenous variables. Another statistical property is the choice of the exponent rate for the coefficient of persistence that corresponds to the instrumental variable c_{zj}. Specifically, the econometric literature has documented a choice of γ_z close to 0.95 as a reasonable value with desirable finite-sample properties when constructing predictability tests (see, Lee (2016), Phillips and Lee (2016) and Kostakis et al. (2015)). Thus, to account for the different convergence rates due to nonstationarity and obtain the asymptotic distribution of the IVX-QR estimator we employ the following normalization matrices

$$\tilde{\boldsymbol{Z}}_{t-1,n} := \tilde{\boldsymbol{D}}_n^{-1} \tilde{\boldsymbol{z}}_{t-1} \quad \text{and} \quad \tilde{\boldsymbol{X}}_{t-1,n} := \tilde{\boldsymbol{D}}_n^{-1} \tilde{\boldsymbol{x}}_{t-1} \tag{3.19}$$

where $\tilde{\boldsymbol{D}}_n = n^{\frac{1+\gamma_x \wedge \gamma_z}{2}} \boldsymbol{I}_p$, such that $\gamma_x \wedge \gamma_z \equiv \mathsf{min}\,(\gamma_x, \gamma_z)$ which is identical for both the case of *local unit root* and *mildly integrated* regressors. Furthermore, we denote with $y_t(\tau) := y_t - \alpha(\tau) + \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ to be the zero-intercept *QR* dependent variable. The particular dequantiling procedure permits to reformulate the quantile model as $y_t(\tau) = \boldsymbol{x}'_{t-1}\beta(\tau) + u_t(\tau)$ that simplifies the derivations for the asymptotics of the model estimator which is known to have different convergence rates when an intercept is included (e.g., see Gonzalo and Pitarakis (2012, 2017)). Then, the IVX-QR estimator for the quantile regression, is defined by the following unconstrained optimization problem

$$\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) := \underset{\beta \in \mathbb{R}^p}{\mathsf{arg\ inf}} \; \frac{1}{2} \left\{ \left(\sum_{t=1}^{n} h_t(\boldsymbol{\beta}) \right)' \left(\sum_{t=1}^{n} h_t(\boldsymbol{\beta}) \right) \right\}, \tag{3.20}$$

where $h_t(.)$ is defined below such that $\psi_\tau(\mathsf{u}) := [\mathbb{1}\{\mathsf{u} \leq 0\} - \tau]$ and $\psi_\tau(\mathsf{u}) \overset{\mathsf{maps\ to}}{\mapsto} \rho_\tau^{-1}(\mathsf{u})$

$$h_t(\boldsymbol{\beta}) := \tilde{\boldsymbol{z}}_{t-1} \times \psi_\tau(u_t(\tau)) \equiv \tilde{\boldsymbol{z}}_{t-1} \times [\tau - \mathbb{1}\{y_t(\tau) < \boldsymbol{x}'_{t-1}\boldsymbol{\beta}\}] \tag{3.21}$$

[10]Notice that the related asymptotic theory which is robust to abstract degree of persistence and results to nuisance-parameter free inference was pioneered by Magdalinos and Phillips (2009a) in the context of cointegration models.

The minimization of expression (3.20) leads to the following first order condition:

$$\sum_{t=1}^{n} \tilde{\boldsymbol{Z}}_{t-1,n} \times \left[\tau - \mathbb{1}\{y_t(\tau) < \boldsymbol{x}'_{t-1}\boldsymbol{\beta}_n^{ivx-qr}(\tau)\}\right] = o_{\mathbb{P}}(1). \tag{3.22}$$

Furthermore, it can be proved that for both the cases of high persistent and mildly integrated predictors (*LUR* or *MI*) the asymptotic distribution of the IVX-QR estimator is identical as presented by Corollary 3.2. Notice that for the interested reader the limit distribution for other classes of persistence (e.g., such as predictors which exhibit near stationary or mildy explosive persistence) can be found in Theorem 3.1 of Lee (2016).

Corollary 3.2. (IVX-QR Limit Theory) Under Assumption 3.1 it follows that

$$\tilde{\boldsymbol{D}}_n \left(\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau) \right) \Rightarrow \mathcal{N}\left(0, \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2} \left(\boldsymbol{\Gamma}_{cxz} \boldsymbol{V}_{cxz}^{-1} \boldsymbol{\Gamma}'_{cxz} \right)^{-1} \right) \tag{3.23}$$

which is a mixed Gaussian distribution due to the stochastic covariance matrix.

Analytic definitions of the covariance matrices $\boldsymbol{\Gamma}_{cxz}$ and $\boldsymbol{V}_{cxz}$ can be found in expressions (3.4) and (3.5) in Lee (2016). Lastly, Lemma C.4 presents the asymptotic behaviour of the self-normalized Wald statistic based on the IVX-QR estimator (see, Proposition 3.1 in Lee (2016)) for both the cases of *LUR* and *MI* regressors.

Lemma 3.1. (Self-normalized IVX-QR) Under Assumption 3.1 it holds that,

$$\frac{\widehat{f_{u_t(\tau)}}(0)^2}{\tau(1-\tau)} \left(\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau) \right)' \left(\boldsymbol{X}'\boldsymbol{P}_{\tilde{\boldsymbol{Z}}}\boldsymbol{X} \right) \left(\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau) \right) \Rightarrow \chi_p^2 \tag{3.24}$$

where

$$\left(\boldsymbol{X}'\boldsymbol{P}_{\tilde{\boldsymbol{Z}}}\boldsymbol{X} \right) := \left(\boldsymbol{X}'\tilde{\boldsymbol{Z}} \right) \left(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}} \right)^{-1} \left(\tilde{\boldsymbol{Z}}'\boldsymbol{X} \right) \equiv \left(\sum_{t=1}^{n} \boldsymbol{x}_{t-1}\tilde{\boldsymbol{z}}'_{t-1} \right) \left(\sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}'_{t-1} \right)^{-1} \left(\sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1}\boldsymbol{x}'_{t-1} \right)$$

such that $\widehat{f_{u_t(\tau)}}(0)^2$ is a consistent estimator of $f_{u_t(\tau)}(0)^2$.

Furthermore, the above result can be generalized when testing for a set of linear restrictions under the null hypothesis, $\mathcal{H}_0 : \boldsymbol{R}\boldsymbol{\beta}(\tau) = \boldsymbol{q}(\tau)$ where $\boldsymbol{R}$ is a $r \times p$ known matrix and $\boldsymbol{q}(\tau)$ is a prespecified vector. The corresponding asymptotic distribution for the IVX-Wald statistic for the quantile predictive regression is given by the following expression

$$\frac{\widehat{f_{u_t(\tau)}}(0)^2}{\tau(1-\tau)} \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{q}(\tau) \right)' \left[\boldsymbol{R} \left(\boldsymbol{X}'\boldsymbol{P}_{\tilde{\boldsymbol{Z}}}\boldsymbol{X} \right)^{-1} \boldsymbol{R}' \right]^{-1} \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{q}(\tau) \right) \Rightarrow \chi_r^2$$

where χ_r^2 denotes the chi-square random variate with r degrees of freedom such that $\mathbb{P}\left(\chi^2 \geq \chi_{r;\alpha}^2\right) = \alpha$, where $0 < \alpha < 1$ denotes the fixed significance level.

3.2.3 Testing Hypotheses

In this section we present the testing problem of interest in this paper. More precisely, we consider two type of testing hypotheses, that is: *(i)* testing for structural break for a fixed quantile level $\tau \in (0,1)$ and *(ii)* testing for a structural break multiple quantile levels. For both testing hypotheses we operate under the assumption of a single structural break at an unknown location. Similar formulations can be found in Qu (2008); however the focus of our study is the nonstationary quantile predictive regression model with persistent covariates, commonly employed in the time series econometrics literature.

Testing Hypothesis A. The first testing hypothesis of interest is concerned with testing for structural break in a pre-specified quantile with the null and alternative hypothesis given as below

$$\mathcal{H}_0^{(A)} : \boldsymbol{\beta}_t(\tau) = \boldsymbol{\beta}_0(\tau) \quad \text{for all} \ \ 1 \leq t \leq n, \ textforafixed \ \tau \in (0,1),$$
$$\mathcal{H}_1^{(A)} : \boldsymbol{\beta}_t(\tau) = \begin{cases} \boldsymbol{\beta}_1(\tau) & \text{where } 1 \leq t \leq \kappa \\ \boldsymbol{\beta}_2(\tau) & \text{where } \kappa + 1 \leq t \leq n \end{cases}$$

for a fixed $\tau \in (0,1)$, where $\kappa = \lfloor \lambda n \rfloor$ the unknown break-point with $\lambda \in (0,1)$.

Testing Hypothesis B. The second testing hypothesis of interest is concerned with testing for structural break across multiple quantiles, that is, quantiles contained in a set $\mathcal{T}_\iota$, with the null and alternative hypothesis given as below

$$\mathcal{H}_0^{(B)} : \boldsymbol{\beta}_t(\tau) = \boldsymbol{\beta}_0(\tau) \quad \text{for all} \ \ 1 \leq t \leq n, \ \text{and for all} \ \ \tau \in \mathcal{T}_\iota,$$
$$\mathcal{H}_1^{(B)} : \boldsymbol{\beta}_t(\tau) = \begin{cases} \boldsymbol{\beta}_1(\tau) & \text{where } 1 \leq t \leq \kappa \\ \boldsymbol{\beta}_2(\tau) & \text{where } \kappa + 1 \leq t \leq n \end{cases} \tag{3.25}$$

for some $\tau \in \mathcal{T}_\iota$, where $\kappa = \lfloor \lambda n \rfloor$ the unknown break-point with $\lambda \in (0,1)$.

where $\boldsymbol{\beta}_0(\tau)$ is the value of the true population parameter under the null hypothesis of no parameter instability.

Remark 3.2. Notice that the statistical problem given by *Testing Hypothesis A* allow us to focus on a particular quantile of interest, e.g., any fixed quantile level $\tau \equiv \tau_0 \in (0,1)$. On the other hand, the inference problem given by *Testing Hypothesis B* permits testing for structural break in the coefficients of the quantile predictive regression model by investigating the presence of breaks in the conditional distribution, that is, at any possible quantile level within the compact set $\mathcal{T}_\iota := [\iota, 1-\iota]$ where $0 < \iota < 1/2$.

Both *Testing Hypothesis A* and *B* summarize the modelling environment under the null as well as under the alternative hypothesis. More precisely, under the null hypothesis the parameter vector is taken to be constant throughout the sample such that $\boldsymbol{\beta}_t(\tau) \equiv \boldsymbol{\beta}_0(\tau)$ for $t = 1, ..., n$ where $\boldsymbol{\beta}_0(\tau)$ is the unknown quantile dependent regression parameter. Therefore, in practise we are interested in testing the null hypothesis that $\boldsymbol{\beta}_t(\tau)$ remains constant, that is, $\boldsymbol{\beta}_t(\tau) = \boldsymbol{\beta}_0(\tau)$ for all t against the alternative that the quantile dependent parameter vector $\boldsymbol{\beta}_t(\tau)$ has a single structural break at an unknown location within the full sample, resulting to two regimes[11]. Under the alternative hypothesis:

$$\mathsf{Q}_{y_t}(\tau|\mathcal{F}_{t-1}) = \boldsymbol{\beta}_1(\tau)\boldsymbol{x}_{t-1}\mathbb{1}\{t \leq \kappa\} + \boldsymbol{\beta}_2(\tau)\boldsymbol{x}_{t-1}\mathbb{1}\{t > \kappa\} + u_t \tag{3.26}$$

where $\mathcal{F}_t$ denotes the $\sigma-$field generated by $\{\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, ..., \}$. Therefore, it is convenient to write the hypotheses with a different formulation. Denote with $\boldsymbol{\beta}_{(1)}(\tau) = \boldsymbol{\beta}_1(\tau)$ and $\boldsymbol{\beta}_{(2)}(\tau) = \boldsymbol{\beta}_2(\tau) - \boldsymbol{\beta}_1(\tau)$. Furthermore, to construct the model so that it can capture the magnitude of the structural break we denote with $\boldsymbol{\mathcal{X}}_{t-1} = \left(\boldsymbol{x}'_{t-1}, \boldsymbol{x}'_{t-1}\mathbb{1}\{t > \kappa\}\right)'$ and $\boldsymbol{\vartheta}(\tau) = \left(\boldsymbol{\beta}_{(1)}(\tau)', \boldsymbol{\beta}_{(2)}(\tau)'\right)'$. Thus, we express the null hypothesis as following

$$\mathcal{H}_0^{(A)} : \mathsf{Q}_{y_t}(\tau|\mathcal{F}_{t-1}) = \boldsymbol{\mathcal{X}}'_{t-1}\boldsymbol{\vartheta}(\tau), \text{ with } \boldsymbol{\beta}_{(2)}(\tau) = \mathbf{0} \text{ for some fixed } \tau_0 \in \mathcal{T}_\iota \tag{3.27}$$

Then, the alternative hypothesis can be formulated as below

$$\mathcal{H}_1^{(A)} : \mathsf{Q}_{y_t}(\tau|\mathcal{F}_{t-1}) = \boldsymbol{\mathcal{X}}'_{t-1}\boldsymbol{\vartheta}(\tau), \text{ with } \boldsymbol{\beta}_{(2)}(\tau) \neq \mathbf{0} \text{ for some fixed } \tau \in \mathcal{T}_\iota \tag{3.28}$$

Furthermore, since we consider the nonstationary quantile predictive regression model without an intercept, following the formulations given by expressions (3.27) and (3.28) for the null and alternative hypotheses respectively, then we define the quantile dependent estimator as the optimization problem below

$$\widehat{\boldsymbol{\vartheta}}_n(\lambda, \tau) := \underset{\boldsymbol{\vartheta} \in \mathbb{R}^{2p}}{\arg\min} \sum_{t=1}^{n} \rho_\tau(y_t - \boldsymbol{\mathcal{X}}'_{t-1}\boldsymbol{b}), \tag{3.29}$$

Therefore, with the above formulation of the estimator $\widehat{\boldsymbol{\vartheta}}_n(\tau, \lambda)$ is the quantile dependent regression estimator when we employ $\boldsymbol{\mathcal{X}}_{t-1}$ to be the model predictor variables. Specifically, when the $\mathcal{H}_0^{(A)}$ is true, under suitable regularity conditions, $\widehat{\boldsymbol{\vartheta}}_2(\lambda, \tau)$ converges in probability to $\mathbf{0}$ for each $(\lambda, \tau) \in \Lambda_\eta \times \mathcal{T}_\iota$. On the other hand, when $\mathcal{H}_1^{(A)}$ is true, $\widehat{\boldsymbol{\vartheta}}_2(\lambda; \tau_0)$ converges in probability to $\boldsymbol{\beta}_{(2)}(\tau_0) = (\boldsymbol{\beta}_2(\tau_0) - \boldsymbol{\beta}_1(\tau_0)) \neq \mathbf{0}$. In summary, since the quantile level τ_0 especially for *Testing Hypothesis B* is unknown a prior, then it is reasonable to reject $\mathcal{H}_0$ when the magnitude of $\widehat{\boldsymbol{\vartheta}}_2(\lambda, \tau)$ is suitable large for some $(\lambda, \tau) \in \Lambda_\eta \times \mathcal{T}_\iota$. Thus, an example of a suitable test statistic to test whether $\mathcal{H}_0$ against the alternative hypothesis $\mathcal{H}_1$ is to employ the supremum of the Wald process.

[11]Notice that the two regimes we refer to here, are not equivalent to testing methodologies proposed in studies such as Gonzalo and Pitarakis (2012, 2017) and Galvao et al. (2014) in which emphasis is given to testing the null hypothesis of linearity based on the presence of no threshold effect.

In general, a Wald type statistic has the following form

$$\mathcal{SW}_n(\tau, \lambda) := \sup_{(\lambda,\tau)\in\Lambda_\eta\times\mathcal{T}_\iota} n\hat{\boldsymbol{\beta}}_{(2)}(\tau_0)\Big[\boldsymbol{V}_n(\lambda;\tau_0)\Big]\hat{\boldsymbol{\beta}}_{(2)}(\tau_0) \tag{3.30}$$

where $\boldsymbol{V}_n(\lambda;\tau_0)$ is the asymptotic covariance matrix of the stochastic process $\sqrt{n}\hat{\boldsymbol{\beta}}_{(2)}(\tau_0)$, under the null hypothesis. However, since the covariance matrix that corresponds to the population regression parameters is in practise unknown, is replaced by a suitable consistent estimate that holds under the null hypothesis of no structural break in the quantile predictive regression model. Obviously within the aforementioned structural break setting which is our main research focus in the paper, the break-point location is not identified under the null hypothesis as we explained in the introduction.

Moreover, additionally to Remark 3.2, the Testing Hypotheses of interest A and B correspond to two different test functions such that Testing Hypothesis A requires to formulate test statistics by employing the supremum funcitonal while Testing Hypothesis B requires to construct test statistics with the use of the double supremum functional. Intuitively, we are interested to examine both hypotheses since it might be the case that A is not rejected, that is, there are no statistical evidence of the presence of a structural break for a given quantile level $\tau_0 \in (0,1)$, while when employing Testing Hypothesis B, it could be the case that the test statistic provides statistical evidence of rejecting the null hypothesis, implying that a structural break still exist at some other quantile level not the one which is kept fixed, within the set (0,1).

3.3 Testing for break for a fixed quantile

Next, we focus on the structural break testing procedures for the quantile regression model with regressors generated as near unit root processes based on two test statistics.

3.3.1 Preliminary Setting

We consider structural break tests for a fixed quantile level, say $\tau_0 \in (0,1)$. Consider the subgradient[12] $S_n(\lambda, \tau_0, \boldsymbol{b})$, based on the subsample $1 \leq t \leq \kappa$

$$S_n(\lambda, \tau_0, \boldsymbol{b}) = n^{-1/2}\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\psi_\tau(y_t - \boldsymbol{x}'_{t-1}\boldsymbol{b}), \tag{3.31}$$

where $\boldsymbol{b}$ corresponds to an estimator of the parameter vector $\boldsymbol{\beta}(\tau_0)$ which encompasses both the OLS and IVX estimators under suitable parametrizations.

The continuous function $\psi_\tau(.)$ is defined as $\psi_\tau(\mathsf{u}) = \big[\tau_0 - \mathbb{1}\{\mathsf{u} \leq 0\}\big]$ and $\kappa = \lfloor \lambda n \rfloor$ denotes the unknown break-point location implying a break fraction $\lambda \equiv \lim_{n\to\infty} \kappa/n$ such that $\lambda \in \Lambda_\eta := [\eta, 1-\eta]$ is a compact set.

[12]The required convexity arguments for obtaining estimators based on the conditional quantile function are based on the convexity lemma result presented by Pollard (1991). Furthermore, related results are presented by Koenker and Portnoy (1987).

Therefore, under the null hypothesis of no structural break with stationary and ergodic regressors, the quantity $\psi_\tau(y_t - \boldsymbol{x}'_{t-1}\boldsymbol{\beta}(\tau_0))$ is a pivotal statistic. In particular, since $\boldsymbol{x}'_{t-1}\boldsymbol{\beta}(\tau_0)$ is equal to the conditional $\tau-$quantile of y_t given $\boldsymbol{x}_{t-1}$, then the random variables

$$\mathbb{1}\left\{y_1 \leq \boldsymbol{x}'_{t-1}\boldsymbol{\beta}(\tau_0)\right\}, ..., \mathbb{1}\left\{y_n \leq \boldsymbol{x}'_{t-1}\boldsymbol{\beta}(\tau_0)\right\} \tag{3.32}$$

are independent Bernoulli trials with success probability τ (see, Galvao et al. (2014)), which implies a sequence of random variables with mean zero and variance $\tau_0(1-\tau_0)$. A similar result should hold in our modelling setting with nonstationary regressors.

Furthermore, denote with $\boldsymbol{X} = (x'_1, ..., x'_n)'$ and define the following auxiliary quantity

$$\mathcal{J}_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau)) := \left(n^{-1}\boldsymbol{X}'\boldsymbol{X}\right)^{-1/2} S_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau)). \tag{3.33}$$

Then, under stationarity and certain regularity conditions (e.g., see Assumptions 1 and 2 in Qu (2008)), $\mathcal{J}_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau_0))$ converges to a limit distribution that is nuisance parameter free. More specifically, it holds that (see, (Qu, 2008, p. 172))

$$\mathcal{J}_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau_0)) \overset{\mathcal{D}}{\to} \mathcal{N}\left(0, \lambda^2\tau_0(1-\tau_0)\right). \tag{3.34}$$

Thus, replacing the unknown parameter vector $\boldsymbol{\beta}_0(\tau_0)$ with the quantile dependent regression estimator based on the full sample, under the null hypothesis of no structural break in the model, the quantity given by (3.33) can be formulated as below

$$\hat{\mathcal{J}}_n(\lambda, \tau_0, \widehat{\boldsymbol{\beta}}_n(\tau_0)) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\psi_\tau\left(y_t - \boldsymbol{x}'_{t-1}\widehat{\boldsymbol{\beta}}_n(\tau_0)\right). \tag{3.35}$$

for some $0 < \lambda < 1$ and $\tau_0 \in (0, 1)$.

In particular, the use of fluctuation type statistics provide a way for statistical inference regarding the presence of structural breaks in model coefficients (Leisch et al. (2000)). Intuitively for these class of tests we consider the asymptotic behaviour of the corresponding empirical processes to decide whether to accept or reject the null. In particular, the fluctuation type test converges to a nondegenerate limiting distribution under the null hypothesis, since the random quantity given by expression (3.35) is essentially governed by the invariance principle under the null. In practise, when the quantile dependent parameters exhibit no structural break in the sample, then $\widehat{\boldsymbol{\beta}}_n(\tau_0)$ is a consistent estimator and as a result, $\hat{\mathcal{J}}_n(\lambda, \tau_0, \widehat{\boldsymbol{\beta}}_n(\tau_0))$ has the same stochastic order as its population counterpart. On the other hand, when the null hypothesis is false, the underline stochastic process exhibit excessive fluctuations. Specifically, under the alternative hypothesis, model parameters have a break at some unknown location in the sample, which implies that $\widehat{\boldsymbol{\beta}}_n(\tau_0)$ will differ significantly from the true value for some sub-sample and the estimated residuals will have high fluctuations (beyond the usual increments of a Wiener process) resulting to falsely rejecting the null due to a large value of the statistic (Qu, 2008, p. 172).

The focus of the proposed econometric environment in this paper is the structural break detection in the model parameters of the nonstationary quantile predictive regression, under the assumption that regressors are generated as near unit root processes. Intuitively, the two persistence classes (mildly integrated and near unit root) we consider encompasses moderate deviations from the unit boundary similar to the case of near integrated (see Phillips (1988a)). Thus, both the value (and sign) of the coefficient of persistence as well as its exponent rate[13] (a tuning parameter), determine the asymptotic behaviour of functionals based on these near unit root process. As a result, the chosen estimator can affect the asymptotic theory of the proposed test statistics as well as the corresponding functionals which we examine separately below.

OLS based functionals

Within the proposed econometric environment which corresponds to the modelling of nonstationary quantile time series models, regressors are assumed to follow a local unit root process. Thus, we expect that OLS based functionals will dependent on the nuisance coefficient of persistence[14]. Furthermore, due to the presence of both a model intercept and the set of nonstationary regressors, we also need to modify the functionals given by expressions (3.31)-(3.33) in order to account for the different convergence rates.

Therefore, to obtain equivalent representations to the quantity $S_n(\lambda, \tau_0, \boldsymbol{b})$, we consider the corresponding partial sum process of the functional $\boldsymbol{K}_{nx}(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0))$ as given by Definition 3.1. We obtain the limit result for these functionals based on the full sample and then focus on deriving invariance principles for the corresponding partial sum processes for the two estimators under examination (see, Section 3.3.2). More precisely, the functionals given by Definition 3.1 correspond to a quantile regression ordinary least squares estimator and are employed when the asymptotic behaviour of the OLS based test statistics is concerned (see also Lemma A1 in Lee (2016)).

Definition 3.1.

$$\boldsymbol{K}_{nx}(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0)) := \boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \boldsymbol{X}_{t-1} \psi_\tau(u_t(\tau_0)) \tag{3.36}$$

$$\boldsymbol{L}_{nx}(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0)) := \boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{n} f_{u_t(\tau), t-1}(0) \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}' \right] \boldsymbol{D}_n^{-1} \tag{3.37}$$

for some $\tau_0 \in (0,1)$ where $\psi_\tau(u_t(\tau_0)) = \left[\tau_0 - \mathbb{1}\{ y_t - \boldsymbol{X}_{t-1}' \boldsymbol{\theta}_n^{ols}(\tau_0) \leq 0 \} \right]$.

A key observation is that under the null hypothesis of no parameter instability these functional converge to a nondegenerate limit distribution. Corollary 3.3 demonstrates the asymptotic distributions of the functionals given by Definition 3.1.

[13]Practically, these are nuisance parameters however via Monte Carlo simulations we can choose suitable values for c_i and γ_x in order to simulate these experimental conditions and thus evaluate the finite-sample performance of the test statistics with high persistence or mildly integrated regressors.

[14]Notice that this is the standard inference problem in the predictability literature. Further details regarding the bias (nonstandard distortion) occurred in predictability tests (i.e., $t-$tests) in quantile predictive regression models can be found in the study of Lee (2016). In our study, we aim to compare both the OLS as well as the instrumental variable approach of Phillips and Magdalinos (2009).

Corollary 3.3. Under the assumption that the pair $\{y_t, \boldsymbol{x}_{t-1}\}_{t=1}^n$ is generated by the model (4.11)-(4.12) then for both *LUR* and *MI* regressors it holds that

(i) $\boldsymbol{K}_{nx}(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0)) \Rightarrow \boldsymbol{K}_x(\tau_0, \boldsymbol{\theta}_0(\tau_0))$, for some $\tau_0 \in (0,1)$ as $n \to \infty$,

(ii) $\boldsymbol{L}_{nx}(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0)) \Rightarrow \boldsymbol{L}_x(\tau_0, \boldsymbol{\theta}_0(\tau_0))$, for some $\tau_0 \in (0,1)$ as $n \to \infty$,

where

$$\boldsymbol{K}_x(\tau_0, \boldsymbol{\theta}_0(\tau_0)) \equiv \begin{cases} \begin{bmatrix} B_{\psi_\tau}(1)_{(1\times n)} \\ \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix}_{(p+1)\times n} & LUR, \\ \mathcal{N}\left(\boldsymbol{0}, \tau_0(1-\tau_0) \times \begin{bmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{V}_{xx} \end{bmatrix}_{(p+1)\times(p+1)}\right)_{(p+1)\times n} & MI. \end{cases} \tag{3.38}$$

$$\boldsymbol{L}_x(\tau_0, \boldsymbol{\theta}_0(\tau_0)) \equiv \begin{cases} f_{u_t(\tau)}(0) \times \begin{bmatrix} 1 & \int_0^1 \boldsymbol{J}_c(r)' \\ \int_0^1 \boldsymbol{J}_c(r) & \int_0^1 \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)' \end{bmatrix}_{(p+1)\times(p+1)} & LUR, \\ f_{u_t(\tau)}(0) \times \begin{bmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{V}_{xx} \end{bmatrix}_{(p+1)\times(p+1)} & MI. \end{cases} \tag{3.39}$$

where the stochastic matrix $\boldsymbol{V}_{xx}$ is defined by Phillips and Magdalinos (2009) as

$$\boldsymbol{V}_{xx} := \int_0^\infty e^{r\boldsymbol{C}_p} \boldsymbol{\Omega}_{xx} e^{r\boldsymbol{C}_p} dr, \text{ where } \boldsymbol{\Omega}_{xx} := \sum_{m=-\infty}^{\infty} \mathbb{E}\left(\boldsymbol{v}_t \boldsymbol{v}_{t-m}'\right) = \boldsymbol{\varphi}_o(1)\boldsymbol{\Sigma}\boldsymbol{\varphi}_o(1)'.$$

Remark 3.3. Notice that an important aspect for robust inference in quantile regressions[15] is the consistent estimation of the sparsity coefficient (see, discussion presented in Koenker and Machado (1999)) and also conditions proposed by Koltchinskii (1997), especially in finite samples. In our setting the self-normalized property of Wald type tests ensures that the sparsity coefficient does not affect the estimation accuracy.

[15]In some studies presented in the literature the use of the check function is defined to be the difference of the indicator function from the quantile level, as in Zhou and Portnoy (1998); however both expressions are equivalent due to the monotonicity property of the check function.

IVX based functionals

In this Section, we derive the asymptotic distribution of the IVX based functionals which are useful to obtain the asymptotic behaviour of the proposed structural break tests under the assumption of nonstationary regressors in the model. We employ the embedded normalization version of the instruments such that $\tilde{\boldsymbol{Z}}_{t-1,n} := \tilde{\boldsymbol{D}}_n^{-1} \tilde{\boldsymbol{z}}_{t-1}$.

Definition 3.2.

$$\boldsymbol{K}_{nz}(\tau_0, \boldsymbol{\beta}_n^{ivx}(\tau_0)) := \sum_{t=1}^{n} \tilde{\boldsymbol{Z}}_{t-1,n} \psi_\tau(u_t(\tau_0)) \tag{3.40}$$

$$\boldsymbol{L}_{nz}(\tau_0, \boldsymbol{\beta}_n^{ivz}(\tau_0)) := \left[\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \tilde{\boldsymbol{Z}}_{t-1,n} \tilde{\boldsymbol{Z}}'_{t-1,n} \right] \tag{3.41}$$

$$\boldsymbol{M}_{nz}(\tau_0, \boldsymbol{\beta}_n^{ivx}(\tau_0)) := \left[\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \tilde{\boldsymbol{Z}}_{t-1,n} \boldsymbol{X}'_{t-1,n} \right] \tag{3.42}$$

for some $\tau_0 \in (0,1)$ where $\psi_\tau(u_t(\tau_0)) = \left[\tau_0 - \mathbb{1}\{ y_t - \boldsymbol{\beta}_n^{ivx}(\tau_0)' \boldsymbol{x}_{t-1} \leq 0 \} \right]$.

Corollary 3.4. Under the assumption that the pair $\{ y_t, \boldsymbol{x}_{t-1} \}_{t=1}^{n}$ is generated by the model (4.11)-(4.12) then for both *LUR* and *MI* regressors it holds that

(i) $\boldsymbol{K}_{nz}(\tau_0, \boldsymbol{\beta}_n^{ivx}(\tau_0)) \Rightarrow \boldsymbol{K}_z(\tau_0, \boldsymbol{\beta}_0(\tau_0)) \equiv \mathcal{N}(\boldsymbol{0}, \tau_0(1-\tau_0)\boldsymbol{V}_{cxz})$,

(ii) $\boldsymbol{L}_{nz}(\tau_0, \boldsymbol{\beta}_n^{ivz}(\tau_0)) \Rightarrow \boldsymbol{L}_z(\tau_0, \boldsymbol{\beta}_0(\tau_0)) \equiv f_{u_t(\tau)}(0) \times \boldsymbol{V}_{cxz}$,

(iii) $\boldsymbol{M}_{nz}(\tau_0, \boldsymbol{\beta}_n^{ivx}(\tau_0)) \Rightarrow \boldsymbol{M}_z(\tau_0, \boldsymbol{\beta}_0(\tau_0)) \equiv f_{u_t(\tau)}(0) \times \boldsymbol{\Gamma}_{cxz}$,

where the definition of the asymptotic matrix $\boldsymbol{V}_{cxz}$ depends on the stochastic dominance of the two exponent rates (see, Phillips and Magdalinos (2009) and Lee (2016)) such as

$$\boldsymbol{V}_{cxz} \equiv \begin{cases} \boldsymbol{V}_{zz} = \displaystyle\int_0^\infty e^{r\boldsymbol{C}_z} \boldsymbol{\Omega}_{xx} e^{r\boldsymbol{C}_z} dr, & \text{when } 0 < \gamma_z < \gamma_x < 1, \\ \\ \boldsymbol{V}_{xx} = \displaystyle\int_0^\infty e^{r\boldsymbol{C}_p} \boldsymbol{\Omega}_{xx} e^{r\boldsymbol{C}_p} dr, & \text{when } 0 < \gamma_x < \gamma_z < 1. \end{cases} \tag{3.43}$$

Moreover, the definition of the moment matrix $\boldsymbol{\Gamma}_{cxz}$ is presented by Lee (2016) via expression (3.4) which is the corresponding asymptotic limit given by expression (20) in Phillips and Magdalinos (2009) as given below

$$\boldsymbol{\Gamma}_{cxz} := \begin{cases} -\boldsymbol{C}_z^{-1} \left(\boldsymbol{\Omega}_{xx} + \displaystyle\int_0^1 \boldsymbol{J}_c(r) d\boldsymbol{J}'_c \right), & \text{when } \gamma_x = 1, \\ -\boldsymbol{C}_z^{-1} \Big(\boldsymbol{\Omega}_{xx} + \boldsymbol{C}_p \boldsymbol{V}_{xx} \Big), & \text{when } 0 < \gamma_z < \gamma_x < 1, \\ \boldsymbol{V}_{xx}, & \text{when } 0 < \gamma_x < \gamma_z < 1. \end{cases} \tag{3.44}$$

The proofs of Corollary 3.3 and 3.4 can be found in the Appendix of the paper. Note that the stochastic convergence of these functional holds for large sample size, $n \to \infty$, and the existence of well-defined moment matrices with negligible higher-order terms. Therefore, to facilitate the development of the asymptotic theory we define the following empirical process for some parameter vector $\boldsymbol{b} \in \mathbb{R}^p$ such that

$$\boldsymbol{G}_n(\tau, \boldsymbol{b}) := n^{-(1+\gamma_x)/2} \sum_{t=1}^{n} \boldsymbol{z}_{t-1} \times \left\{ \psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\boldsymbol{b}) - \mathbb{E}_{\mathcal{F}_{t-1}} \left[\psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\boldsymbol{b}) \right] \right\}$$

where $\tau \in (0, 1)$ and $0 < \gamma_x < 1$. In particular, the empirical process $\boldsymbol{G}_n(\tau, \boldsymbol{b})$ is consider stochastically $\varrho-$equicontinuous over $\mathcal{T}_\iota \times B$, such that for any $\epsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{[\delta]} \left| G_n(\tau_1, \boldsymbol{b}_1) - G_n(\tau_2, \boldsymbol{b}_2) \right| > \epsilon \right) = 0, \tag{3.45}$$

where $[\delta] := \left\{ (\tau_1, \boldsymbol{b}_1), (\tau_2, \boldsymbol{b}_2) \in (\mathcal{T} \times B)^2 : \ \varrho\big((\tau_1, \boldsymbol{b}_1), (\tau_2, \boldsymbol{b}_2)\big) < \delta \right\}$.

Remark 3.4. The above expression is often employed to derive asymptotics for quantile regression models (with stationary regressors). Specifically, one can consider the validity of the stochastic equicontinuity proof of Bickel (1975) under nonstationarity. Practically, since the regressors employed when estimating the inverse of the quantile function $\psi_\tau(.)$, that is, $\tilde{\boldsymbol{z}}_{t-1}$ is mildly integrated, inducing a nearly stationary process, then the conditions given by Bickel (1975) are valid and the proof follows with modifications to accommodate the nonstationary quantile predictive regression (Lee (2016)).

An additional condition for convergence in probability for the empirical process is imposed by Lemma 3.2, which can be employed to derive the convergence rate of the IVX estimator for the nonstationary quantile predictive regression model.

Lemma 3.2. For a generic constant $\mathcal{C}_1 > 0$

$$\sup\left\{ \left\| \boldsymbol{G}_n(\tau, \boldsymbol{b}) - \boldsymbol{G}_n(\tau, \boldsymbol{0}) \right\| : \|\boldsymbol{b}\| \leq n^{(1+\delta)/2} \mathcal{C}_1 \right\} = o_{\mathbb{P}}(1). \tag{3.46}$$

where $\boldsymbol{b}$ is some estimator of the model parameter vector.

More precisely, Lemma 3.2 provides a simplified way to derive the convergence limit for the IVX-QR estimator (see, also Lee (2016)) that ensures consistent estimation of the model parameters for the quantile predictive regression model. A related study to our setting with detailed derivations for nonstandard inference problems, (Wald type statistics), for nonstationary quantile regressions is presented in the study of Goh and Knight (2009). Overall, the asymptotic theory of this paper aims to combine unit root asymptotics with empirical process methods. Specifically, we employ a two-parameter empirical process that converges weakly to a two-parameter Brownian motion. Therefore, our asymptotic distributions involve stochastic integrals with respect to this two-parameter process.

3.3.2 Invariance principles for partial sum processes

To obtain the asymptotic distributions of the test statistics, we consider the asymptotic behaviour of the partial sum processes of the functionals defined in the previous section. We focus in the case of nonstationary regressors which are either high persistent or mildly integrated (see, Section 4.2 for definitions and Kostakis et al. (2015)). Moreover, since we derive and compare the limit distributions of structural break tests based on the chose estimation methodology, we derive invariance principles that correspond to each of these two estimators. Therefore, we define with

$$S_{nx}^{ols}(\lambda, \tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0)) := \boldsymbol{D}_n^{-1} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1} \psi_\tau(u_t(\tau_0)), \quad \text{for some} \ \ 0 < \lambda < 1, \tag{3.47}$$

where $u_t(\tau_0) = (y_t - \boldsymbol{X}'_{t-1}\boldsymbol{\theta}_n^{ols}(\tau_0))$ for $\tau_0 \in (0,1)$, which can be determined uniquely, making the mapping $\psi_\tau(\mathsf{u}) \mapsto \rho_\tau^{-1}(\mathsf{u})$ one-to-one and well-defined. Moreover, we denote with $\boldsymbol{X}_{t-1} = (\boldsymbol{1}, \boldsymbol{x}'_{t-1})'$ the regressors and $\boldsymbol{\theta}(\tau_0) = (\alpha(\tau_0), \boldsymbol{\beta}'(\tau_0))'$ the parameters.

Recall that for mildly integrated regressors it holds that (see, Corollary 3.3)

$$\boldsymbol{K}_{nx}(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0)) := \boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \boldsymbol{X}_{t-1} \psi_\tau(u_t(\tau_0)) \Rightarrow \mathcal{N}\left(\boldsymbol{0}, \tau_0(1-\tau_0) \begin{bmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{V}_{xx} \end{bmatrix}\right) \tag{3.48}$$

Similarly, we can show that $S_{nx}^{ols}(\lambda, \tau, \boldsymbol{\theta}_n^{ols}(\tau_0)) \Rightarrow S_x(\lambda, \tau_0, \boldsymbol{\theta}_0(\tau_0))$ as $n \to \infty$, where

$$S_x(\lambda, \tau_0, \boldsymbol{\theta_0}(\tau_0)) \equiv \mathcal{N}\left(\boldsymbol{0}, \tau_0(1-\tau_0)\lambda \begin{bmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{V}_{xx} \end{bmatrix}\right) \tag{3.49}$$

for some $0 < \lambda < 1$. Then, for the corresponding IVX based functional it holds that

$$S_{nz}^{ivx}(\lambda, \tau_0, \boldsymbol{\beta}_n^{ivx}(\tau_0)) := \sum_{t=1}^{\lfloor \lambda n \rfloor} \tilde{\boldsymbol{Z}}_{t-1,n} \psi_\tau(u_t(\tau_0)) \Rightarrow \mathcal{N}(\boldsymbol{0}, \tau_0(1-\tau_0)\lambda \boldsymbol{V}_{cxz}) \tag{3.50}$$

where $\tilde{\boldsymbol{Z}}_{t-1,n} := \tilde{\boldsymbol{D}}_n^{-1} \tilde{\boldsymbol{z}}_{t-1}$ since we employ the corresponding dequantiled model.

Definition 3.3.

$$\hat{\mathcal{J}}_{nx}^{ols}(\lambda, \tau_0, \widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)) := (\boldsymbol{X}'\boldsymbol{X})^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1} \psi_\tau\left(y_t - \boldsymbol{X}'_{t-1}\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)\right), \tag{3.51}$$

$$\hat{\mathcal{J}}_{nx}^{ivx}(\lambda, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivx}(\tau_0)) := (\boldsymbol{X}'\tilde{\boldsymbol{Z}})^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \tilde{\boldsymbol{Z}}_{t-1,n} \psi_\tau\left(y_t - \boldsymbol{X}'_{t-1,n}\widehat{\boldsymbol{\beta}}_n^{ivx}(\tau_0)\right), \tag{3.52}$$

$$\hat{\mathcal{J}}_{nx}^{ivz}(\lambda, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0)) := (\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \tilde{\boldsymbol{Z}}_{t-1,n} \psi_\tau\left(y_t - \tilde{\boldsymbol{Z}}'_{t-1,n}\widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0)\right). \tag{3.53}$$

for some $0 < \lambda < 1$ and $\tau_0 \in (0,1)$.

Consider the functionals given by Definition 3.3, then when we employ the OLS estimator for a model with mildly integrated regressors, $\gamma_x \in (0,1)$, the following limit result holds

$$
\begin{aligned}
\hat{\mathcal{J}}^{ols}_{nx}(\lambda, \tau_0, \widehat{\boldsymbol{\theta}}^{ols}_n(\tau_0)) &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1}\psi_\tau\left(y_t - \boldsymbol{X}'_{t-1}\widehat{\boldsymbol{\theta}}^{ols}_n(\tau_0)\right) \\
&\equiv \left(\boldsymbol{D}_n^{-1}\left[\sum_{t=1}^{n} \boldsymbol{X}'_{t-1}\boldsymbol{X}_{t-1}\right]\boldsymbol{D}_n^{-1}\right)^{-1/2} \left\{ \boldsymbol{D}_n^{-1} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1}\psi_\tau(u_t(\tau_0)) \right\} \\
&\Rightarrow \left\{ \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right\}^{-1/2} \times \mathcal{N}\left(\mathbf{0}, \tau_0(1-\tau_0)\lambda \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix}\right) \\
&= \sqrt{\tau_0(1-\tau_0)} \times \mathcal{N}(\mathbf{0}, \lambda \boldsymbol{I}_p). \qquad (3.54)
\end{aligned}
$$

since the term $\frac{1}{n^{1+\gamma_x}} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1} \overset{\mathbb{P}}{\to} \lambda \boldsymbol{V}_{xx}$ converges in probability. A similar limit result holds for the IVZ based functional such that $\hat{\mathcal{J}}^{ivz}_{nx}(\lambda, \tau_0, \widehat{\boldsymbol{\theta}}^{ivz}_n(\tau_0)) \Rightarrow \sqrt{\tau_0(1-\tau_0)} \times \mathcal{N}(\mathbf{0}, \lambda \boldsymbol{I}_p)$, regardless of whether the regressors exhibit high persistence. On the other hand, the OLS based functional has a nonstandard limit distribution with regressors of high persistence. In the case of the IVX functional one needs to consider the limit result for these two classes of persistence separately. These conjectures are summarized and proved by Proposition 3.1 in the next section where we formalize the test statistics.

3.3.3 Test Statistics

We consider as detectors two types of test statistics commonly employed in the literature related to structural break testing methodologies. The first type of test corresponds to the fluctuation type statistic studied by Qu (2008) specifically for a quantile regression model, while the second type of test corresponds to the Wald statistic proposed by the seminal paper of Andrews (1993) for the linear regression model. Both test statistics utilize the supremum functional since the underline assumptions allow for a structural break for the coefficients of the nonstationary quantile predictive regression model at an unknown break-point location within the full sample.

Therefore, the null hypothesis of interest is formulated as below

$$
\mathcal{H}_0^{(A)} : \boldsymbol{\theta}^{(1)}_n(\lambda;\tau_0) = \boldsymbol{\theta}^{(2)}_n(\lambda;\tau_0) \quad \text{versus} \quad \mathcal{H}_1^{(A)} : \boldsymbol{\theta}^{(1)}_n(\lambda;\tau_0) \neq \boldsymbol{\theta}^{(2)}_n(\lambda;\tau_0) \qquad (3.55)
$$

where $\boldsymbol{\theta}^{(j)}_n(\lambda;\tau_0) = \left(\alpha^{(j)}_n(\lambda;\tau_0), \boldsymbol{\beta}^{(j)}_n(\lambda;\tau_0)'\right)'$, for $j \in \{1,2\}$ and the location of the break-point is denoted with $\kappa = \lfloor \lambda n \rfloor$ for some $0 < \lambda < 1$. Specifically, the implementation of structural break tests for the purpose of detecting parameter instability in nonstationary quantile predictive regressions is a novel aspect in the literature. To facilitate for the development of large sample theory, Assumption 3.2 presents necessary conditions relating the matrix moments to the quantile structure of the model.

Assumption 3.2. The regressors of the nonstationary quantile predictive regression model which follow a near unit process, are assumed to satisfy the following conditions:

(a) $\underset{n\to\infty}{\text{plim}} \dfrac{1}{n^{1+\gamma_x}} \sum_{t=1}^{\lfloor \lambda n \rfloor} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1} = \lambda f_{u_t(\tau)}(0)\boldsymbol{V}_{xx}$, uniformly for $0 < \lambda < 1$,

(b) $\underset{n\to\infty}{\text{plim}} \dfrac{1}{n^{1+\gamma_x}} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1} = \lambda \boldsymbol{V}_{xx}$, uniformly for some $0 < \lambda < 1$, where $\boldsymbol{V}_{xx}$ is a $p \times p$ non-random positive definite matrix and $\gamma_x \in (0,1)$,

(c) $\mathbb{E}\left(\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right)^{2+s} < L$ with $s > 0$ and $L < \infty$ for all $1 \leq t \leq n$,

(d) there exists a $\delta > 0$ and an $M < \infty$, such that $n^{-1}\sum_{t=1}^{n} \mathbb{E}\left\|\boldsymbol{x}_{t-1}\right\|^{3(1+\delta)} < M$ and $\mathbb{E}\left(n^{-1}\sum_{t=1}^{n}\left\|\boldsymbol{x}_{t-1}\right\|^{3}\right)^{(1+\delta)} < M$ hold for any n.

Assumption 3.2 (a) and (b) are standard convergence in probability limits for the nonstationary quantile predictive regression model. Assumption 3.2 (c) is employed for the convergence of the weighted empirical process $n^{-1/2}\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\left(\tau_0 - \mathbb{1}\left\{F_{y|x}(y_t) \leq \tau_0\right\}\right)$. Furthermore, Assumption 3.2 (d) ensures stochastic equicontinuity (see, Chapter 2 in Van Der Vaart and Wellner (1996)) of the sequential empirical process based on estimated quantile regression residuals, which is needed to establish weak convergence of the tests (see Bai (1996)). Moreover, when considering the case of nonstationary regressors, since standard quantile regression estimators follow a locally uniform weak convergence (see De Haan and Ferreira (2006)), then invariance principles hold uniformly for $\lambda \in (0,1)$.

Fluctuation type tests

Specifically, since we assume that the true break point is unknown, we need to search over all possible candidate subsets within the full sample. Furthermore, according to Qu (2008) recentering $\hat{\mathcal{J}}_n(\lambda,\tau_0,\widehat{\boldsymbol{\beta}}_n(\tau_0))$ by the quantity $\lambda\hat{\mathcal{J}}_n(1,\tau_0,\widehat{\boldsymbol{\beta}}_n(\tau_0))$ often yields better finite sample performance. Such considerations lead to the following test statistic:

$$\mathcal{SQ}_n(\lambda;\tau_0) = \sup_{\lambda\in[0,1]} \left\| \frac{1}{\sqrt{\tau_0(1-\tau_0)}} \left[\hat{\mathcal{J}}_n(\lambda,\tau_0,\widehat{\boldsymbol{\beta}}_n(\tau_0)) - \lambda\hat{\mathcal{J}}_n(1,\tau_0,\widehat{\boldsymbol{\beta}}_n(\tau_0)) \right] \right\|_{\infty} \tag{3.56}$$

where $\|.\|_{\infty}$ is the sup−norm such that for a generic vector $\mathsf{z} = (\mathsf{z}_1,...,\mathsf{z}_p)$ implies that $\|.\|_{\infty} := \max\left(|\mathsf{z}_1|,...,|\mathsf{z}_p|\right)$ (see, Koenker and Xiao (2002)).

We focus on the implementation of two different estimation methodologies. Therefore, to investigate the practical use of the proposed fluctuation type test for structural break detection in the nonstationary quantile predictive regression model, we consider the asymptotic distribution of the test statistics according to the estimator employed to construct the test function.

Therefore, Proposition 3.1 summarizes the formulations of the test according to the estimation methodology employed for a fixed quantile level $\tau_0 \in (0,1)$. Notice that the check function can be written as below:

$$\rho_\tau(\mathsf{u}) = |\mathsf{u}| \times \left[(1-\tau)\mathbb{1}\{\mathsf{u}<0\} + \tau\mathbb{1}\{\mathsf{u}>0\}\right] \tag{3.57}$$

The regression quantile dependent estimator $\boldsymbol{\beta}(\tau)$ for $\boldsymbol{\beta}$ with score function such that

$$\psi_\tau(y; \boldsymbol{\beta}) = \psi_\tau(y - \boldsymbol{x}'\boldsymbol{\beta}), \tag{3.58}$$

where $\psi_\tau(\mathsf{u}) = \tau \mathbb{1}\{\mathsf{u} > 0\} - (1-\tau)\tau\mathbb{1}\{\mathsf{u} < 0\} = (\tau - \mathbb{1}\{\mathsf{u} < 0\})$. Thus, we can generalize the particular family of uniformly bounded functions $\psi_\tau(\mathsf{u})$ from $\mathcal{T}_\iota$ to $\Theta \subset \mathbb{R}^p$ such that $\psi_{\tau,t}(\mathsf{u}) \equiv \psi_\tau(y_t - Q(\tau|\mathcal{F}_{t-1}))$ such that $Q(\tau|\mathcal{F}_{t-1})$ is identified as the $\tau-$th quantile of the conditional distribution of y_t given the information set $\mathcal{F}_{t-1}$. Furthermore, we can express the check function with respect to the score function such that $\rho_\tau(\mathsf{u}) = -\psi_\tau(\mathsf{u})\mathsf{u}$. Therefore, within the nonstationary quantile predictive regression model of our framework we are particularly interested for the asymptotic theory for $\sqrt{n}\left(\hat{\theta}(\tau) - \theta_0(\tau)\right)$, which is considered to be as a stochastic process dependent on the quantile parameter $\tau \in \mathcal{T}_\iota$. Specifically, in our study we consider two different types of estimators related to the quantile dependent parameter, that is, the OLS based QR estimator as well as the IVX based QR estimator. Both of these estimators are useful for demonstrating the effects of nonstationarity using the conditional distribution function when modelling the y_t variable based on the nonstationary regressors.

Similar asymptotic expressions are derived by Koenker and Machado (1999) who consider testing the null hypothesis, $\mathcal{H}_0 : \beta_2(\tau) = 0$ for some quantile $\tau \in \mathcal{T}_\eta$, against a sequence of local alternatives. Furthermore, to investigate the asymptotic behaviour of the processes some notation related to Bessel processes is employed. Let $\boldsymbol{W}_p(\lambda)$ denote $p-$dimensional vector of independent Brownian motions. Therefore, for each $\lambda \in [0,1]$, $\boldsymbol{\mathcal{BB}}_p(\lambda) = \boldsymbol{W}_p(\lambda) - \lambda \boldsymbol{W}_p(1)$ represents a $p-$vector of independent Brownian bridges. Notice that for any fixed $\lambda \in (0,1)$ it holds that

$$\boldsymbol{\mathcal{BB}}_p(\lambda) \sim \mathcal{N}(0, \lambda(1-\lambda)\boldsymbol{I}_p) \tag{3.59}$$

Moreover, the normalized Euclidean norm of $\boldsymbol{\mathcal{BB}}_p(\lambda)$, is defined as below

$$Q_p(\lambda) = \frac{\|\boldsymbol{\mathcal{BB}}_p(\lambda)\|}{\sqrt{\lambda(1-\lambda)}} \tag{3.60}$$

is referred to as a Bessel process of order p. In particular, critical values for the limiting distribution $\sup_{\lambda \in \Lambda_\eta} Q_p^2(\lambda)$ have been tabulated by Andrews (1993). Then, for any fixed $\lambda \in (0,1)$ it holds that $Q_p^2(\lambda) \sim \chi_p^2$. Furthermore, to characterize the behaviour of the test statistic under local alternatives, it is helpful to define a noncentral version of the squared Bessel process as an extension of the noncentral chi-squared distribution. Define with $\mu(\lambda)$ be a fixed, bounded function from $[0,1] \mapsto \mathbb{R}^p$. Then, we call the standardized squared norm below

$$Q^2_{p,\eta(\lambda)} := \frac{\|\boldsymbol{\mu}(\lambda) + \boldsymbol{\mathcal{BB}}_p(\lambda)\|^2}{\lambda(1-\lambda)} \quad \text{and} \quad \eta(\lambda) = \frac{\boldsymbol{\mu}(\lambda)'\boldsymbol{\mu}(\lambda)}{\lambda(1-\lambda)}, \tag{3.61}$$

as a squared noncentral Bessel process of order p with noncentrality function $\eta(\lambda)$. Then, for any fixed $\lambda \in (0,1)$, $Q^2_{p,\eta(\lambda)} \sim \chi^2_{p,\eta(\lambda)}$, a noncentral χ^2_p random variable with p degrees of freedom and noncentrality parameter $\eta(\lambda)$. An important quantity is the sparsity function and the regression estimates depend on this quantity, since it reflects the density of observations near the quantile. Thus, when the data are very sparse at the quantile τ, then it will be difficult to estimate while when the sparsity is low, (observations are dense), the quantile will be more precisely estimated.

Proposition 3.1. Under the null hypothesis $\mathcal{H}_0^{(A)}$ and given that Assumptions 3.1-3.2 hold, then the fluctuation type statistics weakly converge to the limit distributions below

$$\textbf{\textit{(i)}}\ \mathcal{SQ}_n^{ols}(\lambda;\tau_0) := \sup_{\lambda\in[0,1]} \left\| \frac{1}{\sqrt{\tau_0(1-\tau_0)}} \left[\hat{\mathcal{J}}_n\left(\lambda,\tau_0,\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)\right) - \lambda \hat{\mathcal{J}}_n\left(1,\tau_0,\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)\right)\right]\right\|_\infty$$

$$\Rightarrow \begin{cases} \sup_{\lambda\in[0,1]} \left\|\boldsymbol{\mathcal{BB}}_{p+1}(\lambda)\right\|_\infty, & \text{when } \gamma_x \in (0,1) \\ \sup_{\lambda\in[0,1]} \mathbb{S}_{xx}^{-1/2} \times \left\{ \begin{bmatrix} \boldsymbol{\mathcal{BB}}_{\psi_\tau}(\lambda)_{(1\times n)} \\ \boldsymbol{\mathcal{JB}}_{\psi_\tau}(\lambda)_{(p\times n)} \end{bmatrix}_{(p+1)\times n} \right\}, & \text{when } \gamma_x = 1 \end{cases}$$

$$\textbf{\textit{(ii)}}\ \mathcal{SQ}_n^{ivx}(\lambda;\tau_0) := \sup_{\lambda\in[0,1]} \left\| \frac{1}{\sqrt{\tau_0(1-\tau_0)}} \left[\hat{\mathcal{J}}_n\left(\lambda,\tau_0,\widehat{\boldsymbol{\beta}}_n^{ivx}(\tau_0)\right) - \lambda \hat{\mathcal{J}}_n\left(1,\tau_0,\widehat{\boldsymbol{\beta}}_n^{ivx}(\tau_0)\right)\right]\right\|_\infty$$

$$\Rightarrow \sup_{\lambda\in[0,1]} \left\|\boldsymbol{\mathcal{BB}}_{p}(\lambda)\right\|_\infty, \qquad \text{when } \gamma_x = (0,\gamma_z)$$

$$\textbf{\textit{(iii)}}\ \mathcal{SQ}_n^{ivz}(\lambda;\tau_0) := \sup_{\lambda\in[0,1]} \left\| \frac{1}{\sqrt{\tau_0(1-\tau_0)}} \left[\hat{\mathcal{J}}_n\left(\lambda,\tau_0,\widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0)\right) - \lambda \hat{\mathcal{J}}_n\left(1,\tau_0,\widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0)\right)\right]\right\|_\infty$$

$$\Rightarrow \sup_{\lambda\in[0,1]} \left\|\boldsymbol{\mathcal{BB}}_{p}(\lambda)\right\|_\infty, \qquad \text{when } \gamma_x = (0,1]$$

where $\boldsymbol{\mathcal{BB}}_p(\cdot)$ is a vector of p independent Brownian bridge processes[16] on $\mathcal{D}_{\mathbb{R}^p}([0,1])$,

$$\mathbb{S}_{xx} := \begin{bmatrix} 1 & \int_0^1 \boldsymbol{J}_c(r)'dr \\ \int_0^1 \boldsymbol{J}_c(r)dr & \int_0^1 \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)' \end{bmatrix}_{(p+1)\times(p+1)} \quad \text{with} \quad 0<r<1$$

where $\mathbb{S}_{xx}$ is a positive definite stochastic matrix, $\boldsymbol{\mathcal{BB}}_{\psi_\tau}(\lambda) := B_{\psi_\tau}(\lambda) - \lambda B_{\psi_\tau}(1)$.

A necessary condition to apply weak convergence arguments that yields invariance principles for partial sum processes in the Skorokhod space $\mathcal{D}([0,1])$ follows

$$\sup_{\lambda\in(0,1)} \boldsymbol{D}_n^{-1}\left| \left[\hat{\mathcal{J}}_n(\lambda,\tau_0,\widehat{\boldsymbol{\beta}}_n(\tau_0)) - \lambda\hat{\mathcal{J}}_n(1,\tau_0,\widehat{\boldsymbol{\beta}}_n(\tau_0))\right] - \left[\mathcal{J}_n(\lambda,\tau_0,\boldsymbol{\beta}_n(\tau_0)) - \lambda\mathcal{J}_n(1,\tau_0,\boldsymbol{\beta}_n(\tau_0))\right]\right| = o_{\mathbb{P}}(1).$$

Remark 3.5. The proposed test statistics extend the fluctuation type tests studied by Qu (2008), to the nonstationary quantile predictive regression model of our setting. More specifically, we compare the instrumental variable based method to the classical OLS approach for constructing the fluctuation test. Moreover, we employ the IVZ estimator given by Proposition 3.1 ***(iii)***, which replaces the original covariate vector with the constructed instruments as a post-estimation correction method that permits to obtain further simplifications of asymptotic terms. Our asymptotic theory analysis shows that the fluctuation type test weakly converges into a Brownian bridge limit when the IVZ estimator is employed and the same limit holds for both OLS and IVX based tests under mild integratedness.

[16] Note that $\boldsymbol{\mathcal{BB}}_p(\cdot)$ is known as the square of a standardized tied-down Bessel process of order p.

On the other hand, under high persistence the fluctuation type test based on the OLS estimator is proved to have a weak convergence into a nonstandard and nonpivotal asymptotic distribution. A similar asymptotic result holds for the corresponding IVX based test statistic when $\gamma_x = 1$ such that the following expression holds

$$\hat{\mathcal{J}}_n\left(\lambda, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivx}(\tau_0)\right) - \lambda\hat{\mathcal{J}}_n\left(1, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivx}(\tau_0)\right) \Rightarrow \sqrt{\tau_0(1-\tau_0)} \times \boldsymbol{\Gamma}_{cxz}^{-1/2} \times \left\{\mathcal{N}\Big(\mathbf{0}, \lambda\boldsymbol{V}_{cxz}\Big) - \lambda\mathcal{N}\Big(\mathbf{0}, \boldsymbol{V}_{cxz}\Big)\right\} \tag{3.62}$$

unless we assume that the exponent rate of persistence is such that $\gamma_x(0, \gamma_z)$.

Fluctuation type statistics have been previously examined as a detector for parameter instability in studies such as Kuan and Chen (1994), Chu et al. (1996) and Leisch et al. (2000). More precisely, the statistical advantage of fluctuation type statistics lies in the fact that they utilize properties of the maximum of Wiener processes (see Révész (1982)) and consequently weakly convergence arguments as defined by Billingsley (1968) can be employed to derive their asymptotic behaviour. Furthermore, these class of test statistics belong to the same class as CUSUM type tests although in the latter case the test function is constructed based on regression residuals (see Kulperger et al. (2005)).

Overall within the nonstationary quantile predictive regression model framework of our study, we observe some important conclusions for the implementation of fluctuation type tests as structural break detectors. Firstly, the asymptotic distribution of the test statistic $\mathcal{SQ}(\lambda; \tau_0)$ depends on the chosen estimator when constructing the test function as seen from the limiting distributions when constructing the test based on the two estimators, under high persistent regressors. On the other, under the assumption of mildly integrated regressors these test statistics weakly converge into a Brownian bridge type limit regardless of the chosen estimator when constructing the test function. Therefore, for the particular persistence class fluctuatuon type tests depends only on the number of parameters subject to structural break since the nuisance coefficient of persistence that captures the nonstationary properties of predictors is filtered out. Secondly, these test statistics do not require to estimate the sparsity coefficient $f_{y|x}(F_{y|x}^{-1}(\tau_0))$. According to Qu (2008) this occurs since the subgradient, when evaluated at the true parameter value $\boldsymbol{\beta}_0(\tau_0)$, does not depend on the distribution of the errors.

Thus, conducting statistical inference with some prior information regarding the presence of persistence regressors, using fluctuation type tests as detectors is preferable to construct the test function based on the IVZ estimator which can lead to conventional inference methods (e.g., using tabulated critical values). Next, we examine the self-normalized[17] property of Wald type tests by deriving the related asymptotic theory. Due to the assumptions and conditions under which we construct the proposed test statistics, to examine their limit distributions stochastic equicontinuity arguments are necessary in proofs (see, Newey (1991)).

[17]Specifically a relevant application of self normalized statistics is the construction of confidence intervals for model parameters as in the study of Shao (2010). We leave this aspect for future research.

Wald type tests

We now introduce the Wald type tests based on the two estimation methodologies which we focus on (OLS versus IVX based tests). The formulation of the model under the null and under the alternative hypothesis can change the interpretation in the notation we employ for model parameters. One approach is to employ the formulations given by expressions (3.27)-(3.28). In that case, the Wald type test is constructed for testing the null hypothesis that the parameter vector $\boldsymbol{\beta}_{(2)}(\tau_0)$ which implies that we are testing the null hypothesis that $\boldsymbol{\beta}_1(\tau_0) - \boldsymbol{\beta}_1(\tau_0) = \mathbf{0}$. However, one has to be careful when constructing the covariance matrix as the regressors need to be adjusted accordingly. Furthermore, a second approach is to construct the stacked regressors that correspond to the time series observations from each of the two subsamples.

To simplify the notation, we employ the second approach and denote with $\widehat{\boldsymbol{\beta}}_1(\lambda;\tau_0)$ the estimator of $\boldsymbol{\beta}_0(\tau_0)$, using observations up to $\kappa = \lfloor \lambda n \rfloor$ for some $0 < \lambda < 1$ and with $\widehat{\boldsymbol{\beta}}_2(\lambda;\tau_0)$ the corresponding parameter estimator based on the remaining observations in the sample. Moreover, denote with $\tilde{\boldsymbol{X}} \equiv [\boldsymbol{X}_1 \ \boldsymbol{X}_2]$ and $\boldsymbol{R} \equiv [\boldsymbol{I}_p \ \ -\boldsymbol{I}_p]$ the selection matrix, then

$$\Delta\widehat{\boldsymbol{\beta}}_n(\lambda;\tau_0) := \left(\widehat{\boldsymbol{\beta}}_2(\lambda;\tau_0) - \widehat{\boldsymbol{\beta}}_1(\lambda;\tau_0)\right) \tag{3.63}$$

Then, the Wald test for testing the null hypothesis that the two regimes have equivalent parameter vectors, based on the OLS estimator and some unknown break-point $\kappa = \lfloor \lambda n \rfloor$ is formulated as

$$\mathcal{W}_n(\lambda,\tau_0) = n\Delta\widehat{\boldsymbol{\beta}}_n(\lambda;\tau_0)' \left[\widehat{\boldsymbol{V}}_n(\lambda;\tau_0)\right]^{-1} \Delta\widehat{\boldsymbol{\beta}}_n(\lambda;\tau_0) \tag{3.64}$$

where $\widehat{\boldsymbol{V}}_n(\lambda;\tau_0)$ is a consistent estimate of the limiting variance of $\Delta\widehat{\boldsymbol{\beta}}_n(\lambda;\tau_0)$ under the null hypothesis, $\mathcal{H}_0^{(A)}$, of no parameter instability for a fixed quantile level $\tau_0 \in (0,1)$. The variance estimator is a key quantity which will affect the robustness of Wald type tests and takes different forms depending on the estimation method we employ when fitting the nonstationary quantile predictive regression.

Consider the following limiting variance estimate

$$\underset{n\to\infty}{\text{plim}} \left\{\widehat{\boldsymbol{V}}_n(\lambda;\tau_0)\right\} \equiv \left[\frac{\tau_0(1-\tau_0)}{\lambda(1-\lambda)}\right] \boldsymbol{\Omega}_0, \qquad (\lambda,\tau_0) \in (0,1)\times(0,1), \tag{3.65}$$

where $\boldsymbol{\Omega}_0 = \boldsymbol{H}_0^{-1}\boldsymbol{D}_0\boldsymbol{H}_0^{-1}$ is the unknown variance of the OLS-Wald test.

Furthermore, define with

$$\boldsymbol{H}_0 = \underset{n\to\infty}{\text{plim}} \ \frac{1}{n}\sum_{t=1}^{n} f_{y|x}\left(y_t|\boldsymbol{x}_{t-1}\right)\boldsymbol{x}_{t-1}'\boldsymbol{x}_{t-1} \tag{3.66}$$

$$\boldsymbol{D}_0 = \underset{n\to\infty}{\text{plim}} \ \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_{t-1}'\boldsymbol{x}_{t-1} \tag{3.67}$$

where $f_{y|x}(.|\boldsymbol{x}_{t-1})$ and $F_{y|x}(.|\boldsymbol{x}_{t-1})$ are the conditional density and conditional cumulative distribution function of y_t respectively (see Goh and Knight (2009) and Aue et al. (2017)).

However, the particular form of the asymptotic variance only holds under the assumption of stationarity in which case the variance estimator simplifies since it does not depend on any nuisance parameters (such as the coefficients of persistence) and thus equivalent matrix moments to the expressions in Qu (2008) (see also Andrews (1993)) hold. In our setting, we consider alternative variance estimators based on both the chosen estimator and the persistence class of regressors. For any of the aforementioned cases, the supremum Wald test statistic is defined as

$$\mathcal{SW}_n(\lambda;\tau_0) := \sup_{\lambda\in\Lambda_\eta}\left\{ n\Delta\widehat{\boldsymbol{\beta}}_n(\lambda;\tau_0)'\left[\widehat{\boldsymbol{V}}_n(\lambda;\tau_0)\right]^{-1}\Delta\widehat{\boldsymbol{\beta}}_n(\lambda;\tau_0)\right\} \tag{3.68}$$

In practise, a symmetric trimming coefficient is employed such that $0<\eta<1/2$ which lead to the admissible set $\Lambda_\eta := [\eta, 1-\eta]$, in order to ensure that the test statistics converge in distribution under the null hypothesis. Therefore, we investigate the asymptotic behaviour of the OLS based Wald test for the nonstationary quantile predictive regression model given by (4.11)-(4.12) which encompasses the case of stationary regressors. Under the assumption of stable regressors the asymptotic variance of the OLS-Wald test is equivalent to the case when regressors are stationary and ergodic.

Under the assumption of nonstationarity, the formulation of the OLS-Wald test statistic requires to determine the asymptotic behaviour of the following quantities

$$\widehat{\boldsymbol{\beta}}_1^{ols}(\lambda;\tau_0) = \left(\frac{1}{\kappa}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\right)^{-1}\left(\frac{1}{\kappa}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{x}_{t-1}y_t\right) \tag{3.69}$$

$$\widehat{\boldsymbol{\beta}}_2^{ols}(\lambda;\tau_0) = \left(\frac{1}{n-\kappa}\sum_{t=\lfloor\lambda n\rfloor+1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\right)^{-1}\left(\frac{1}{n-\kappa}\sum_{t=\lfloor\lambda n\rfloor+1}^{n}\boldsymbol{x}_{t-1}y_t\right) \tag{3.70}$$

and $\Delta\widehat{\boldsymbol{\beta}}_n^{ols}(\lambda;\tau_0) = \widehat{\boldsymbol{\beta}}_2^{ols}(\lambda;\tau_0) - \widehat{\boldsymbol{\beta}}_1^{ols}(\lambda;\tau_0)$, for some $0<\lambda<1$ and $\tau\in(0,1)$.

Then, due to orthogonality of the two set of regressors the covariance matrix simplifies into the following expression:

$$\widehat{\boldsymbol{V}}_n^{ols}(\lambda;\tau_0) := \left[\boldsymbol{R}\big(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\big)^{-1}\boldsymbol{R}'\right] \equiv \left[\big(\boldsymbol{X}_1'\boldsymbol{X}_1\big)^{-1} + \big(\boldsymbol{X}_2'\boldsymbol{X}_2\big)^{-1}\right] \tag{3.71}$$

3.3.4 Asymptotic Theory

As we discussed previously, the limit theory of the Wald type statistics for both the OLS and IVX estimators seems more difficult than the limit results for the fluctuation type tests, especially due to the dependence of regressors and parameter estimates to the nuisance parameter of persistence. Therefore, here we generalize the functionals introduced in Section 3.3.2 and 3.3.2 in order to study their asymptotic properties which can alleviate the difficulty in obtaining stochastic approximations under the presence of abstract degree of persistence; simplifying this way derivations for their limit distributions.

OLS-Wald test statistic

We focus on the asymptotic theory for the OLS-Wald test statistic, which is employed as a structural break detection for the nonstationary quantile predictive regression model. In particular, we investigate the asymptotic behaviour of the partial sum processes for the OLS based functionals we introduced previously. For a general parameter vector $\boldsymbol{b} \in \mathbb{R}^p$ we denote with $S_n(\lambda, \tau_0, \boldsymbol{b})$ the partial sum given by the following expression

$$S_n(\lambda, \tau_0, \boldsymbol{b}) = n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1} \psi_\tau (y_t - \boldsymbol{x}'_{t-1} \boldsymbol{b}) \tag{3.72}$$

where $\psi_\tau(\mathsf{u})$ is such that $\psi_\tau(\mathsf{u}) := \left[\tau - \mathbb{1}\left\{ \mathsf{u} \leq 0 \right\} \right]$. Therefore, $S_n(\lambda, \tau_0, \boldsymbol{b})$ is written as

$$S_n(\lambda, \tau_0, \boldsymbol{b}) = n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1} \left[\tau_0 - \mathbb{1}\{ y_t - \boldsymbol{x}'_{t-1} \boldsymbol{b} \leq 0 \} \right]. \tag{3.73}$$

Following conventional laws of invariance principles for $i.i.d$ partial sums the induced sequence of increments are tight within a suitable topological space[18] and in fact they converge weakly to Gaussian processes. Therefore, investigating the asymptotic behaviour and properties of these functionals is useful for the development of the asymptotic theory of the proposed test statistics as well as for other applications. To do this, we consider centering the quantity $\mathbb{1}\{ y_t - \boldsymbol{x}'_{t-1} \boldsymbol{b} \leq 0 \}$ at its expectation conditional on $\boldsymbol{x}_{t-1}$, instead around the quantile level τ_0. Since we assume that the nonparametric functional given by expression (3.73) can be employed as a stochastic process in $\mathcal{D}([0,1])$, which is the topological space of all right continuous functions with left limits then we can derive an invariance principle for this partial sum process.

To simplify derivations for the asymptotic theory, and following Qu (2008), we define the quantity $\widetilde{S}_n(\lambda, \tau_0, \boldsymbol{b})$ with the expression below

$$\widetilde{S}_n(\lambda, \tau_0, \boldsymbol{b}) = n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1} \left[F_{y|x}(\boldsymbol{x}'_{t-1} \boldsymbol{b}) - \mathbb{1}\{ y_t - \boldsymbol{x}'_{t-1} \boldsymbol{b} \leq 0 \} \right]. \tag{3.74}$$

where $F_{y|x}(\boldsymbol{x}'_{t-1} \boldsymbol{b})$ is assumed to be monotonic. A necessary and sufficient condition for the monotonicity property of the cumulative distribution function to hold is presented by Lemma B.2 (see, also Lemma A1 in Qu (2008)). Consequently, we obtain that

$$S_n(\lambda, \tau_0, \boldsymbol{b}) \equiv \widetilde{S}_n(\lambda, \tau_0, \boldsymbol{b}) + n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1} \left[\tau_0 - F_{y|x}(\boldsymbol{x}'_{t-1} \boldsymbol{b}) \right]. \tag{3.75}$$

[18] Related theory to weak convergence arguments of partial sum processes can be found in various studies. For instance, Wang and Phillips (2012) redefine the innovation sequence of their model, in the context of specification testing under nonstationarity, to a richer probability space which contains a standard Brownian motion. To do this, a triangular representation of the near unit process is employed in order to investigate the asymptotic behaviour of the transformed functional with respect to this triangular array. Although this would be an interesting way to represent our functionals we avoid the introduction of triangular arrays which could be more challenging to handle.

Our research objective here is to establish the weak convergence argument that holds for the random quantity $S_n(\lambda, \tau_0, \boldsymbol{b})$ on $(\mathcal{D}[0,1])^2$ by accommodating for the different convergence rates with appropriate matrix normalizations according to the estimator employed in each case. Moreover, the limit results of these functionals[19] can be utilized to show the following type of stochastic convergence

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{ols}(\lambda;\tau_0) - \boldsymbol{\beta}_0(\tau_0)\right) = \mathcal{O}_{\mathbb{P}}(1), \tag{3.76}$$

where $\widehat{\boldsymbol{\beta}}_1^{ols}(\lambda;\tau_0)$ is the quantile regression OLS based estimator that corresponds to the subsample $1 \leq t \leq \lfloor \lambda n \rfloor$ for some $0 < \lambda < 1$ and $\tau_0 \in (0,1)$, when the quantile regression has no model intercept. Intuitively, when the model structure incorporates both intercept and slopes then the different convergence rates of these coefficients due to the presence of nonstationarity is accommodated with the use of embedded normalization matrices. In the case of the IVX estimator obtained using observations from the full sample the following order of convergence holds: $n^{\frac{1+\gamma_x}{2}}(\widehat{\boldsymbol{\beta}}_n^{ivx}(\tau) - \boldsymbol{\beta}_0(\tau)) = \mathcal{O}_{\mathbb{P}}(1)$, which is proved by Corollary 3.2 in the Appendix of the paper (see also Theorem 3.1 in Lee (2016)).

For the remainder of this section, we suppose that the parameter vector is of the form $\boldsymbol{\theta}(\tau) = [\alpha(\tau), \boldsymbol{\beta}(\tau)']'$. Then, Lemma 3.3 below provides a useful decomposition for the functional given by expression (3.75) which we employ when deriving the limit distribution of the OLS-Wald test statistic under the null hypothesis of a single structural break at an unknown location denoted with $\kappa = \lfloor \lambda n \rfloor$, for some $0 < \lambda < 1$. Therefore, testing for a structural break in the quantile predictive regression model, using a Wald type statistic formulation based on the OLS estimator implies to use the parameter vector $\boldsymbol{\theta}(\tau)$ instead of $\boldsymbol{\beta}(\tau)$.

Remark 3.6. Notice that the Wald statistic of interest is a generalized version of the $t-$test statistic given by the following expression

$$\mathcal{T}_{n,i} := n \frac{\widehat{f}_\tau\left(\boldsymbol{R}_i \tilde{\boldsymbol{\theta}}_n(\tau) - r_i\right)}{\left\{\tau(1-\tau)\boldsymbol{R}_i \boldsymbol{\Sigma}_n^{-1} \boldsymbol{R}_i\right\}^{1/2}}, \quad \tau \in (0,1). \tag{3.77}$$

Therefore, it can been easily seen that the $t-$test statistic denoted with $\mathcal{T}_{n,i}^2$ equals the Wald statistic that tests the null hypothesis $\mathcal{H}_0 : \theta_i(\tau) = \theta_{0i}$ (e.g., if $r_i = 0$) against the alternative hypothesis $\mathcal{H}_1 : \theta_i(\tau) \neq \theta_{0i}$, where $\theta_i(\tau)$ is the $i-$th element of the parameter vector $\boldsymbol{\theta}_i$. We focus on the asymptotic behaviour of the Wald-based test statistics for testing parameter stability in quantile predictive regression models. We expect that as the sample size increases, the empirical level of these test statistics tend to converge to the nominal level.

[19]Notice that the proposed functionals in this paper similar to the framework of Qu (2008), clearly depend on the estimated parameter vector. Therefore, in our setting the assumption of a nonstationary quantile model contributes to some challenging asymptotic theory aspects, which we are motivated to tackle. Moreover, we shall note that a related large stream of literature considers functionals of estimated residuals with associated test statistics such as CUSUM and CUSUM-square commonly employed in the change-point literature. We avoid presenting the related literature here, as it beyond our scope.

Lemma 3.3. Under the null hypothesis, $\mathcal{H}_0^{(A)}$, and given that Assumptions 3.1-3.2 hold:

(i) uniformly in $\lambda \in \Lambda_\eta := [\eta, 1-\eta]$ as $n \to \infty$, then

$$\hat{S}_n^{ols}\big(\lambda, \tau_0, \widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0)\big) = S_n\big(\lambda, \tau_0, \boldsymbol{\theta}_0(\tau_0)\big) + \lambda \boldsymbol{L}_{nx}\left(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0)\right)\widehat{\boldsymbol{\mathcal{E}}}_1^{ols}(\lambda;\tau_0) + o_{\mathbb{P}}(1),$$

where $\widehat{\boldsymbol{\mathcal{E}}}_1^{ols}(\lambda;\tau_0) := \sqrt{n}\left(\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0) - \boldsymbol{\theta}_0(\tau_0)\right)$ and $\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0)$ denotes the estimate of $\boldsymbol{\theta}_0(\tau_0)$ using the sample up to $\lfloor \lambda n \rfloor$, and $\boldsymbol{L}_{nx}\left(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0)\right)$ as in Definition 3.1.

(ii) uniformly in $\lambda \in [0,1]$ as $n \to \infty$, then

$$\hat{S}_n^{ols}\big(\lambda, \tau_0, \widehat{\boldsymbol{\theta}}_n^{ols}(\lambda;\tau_0)\big) = S_n\big(\lambda, \tau_0, \boldsymbol{\theta}_0(\tau_0)\big) + \lambda \boldsymbol{L}_{nx}\left(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0)\right)\widehat{\boldsymbol{\mathcal{E}}}_n^{ols}(\tau_0) + o_{\mathbb{P}}(1),$$

where $\widehat{\boldsymbol{\mathcal{E}}}_n^{ols}(\tau_0) := \sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0) - \boldsymbol{\theta}_0(\tau_0)\right)$ and $\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)$ denotes the OLS estimate of $\boldsymbol{\theta}_0(\tau_0)$ using the full sample.

The proof of Lemma 3.3 can be found in the Appendix of the paper.

Proposition 3.2. Under the null hypothesis $\mathcal{H}_0^{(A)}$ and given that Assumptions 3.1-3.2 hold, then the Wald type statistics weakly converge to the limit distributions below

$$\boldsymbol{(i)} \quad \mathcal{SW}_n^{ols}(\lambda;\tau_0) \Rightarrow \sup_{\lambda\in[0,1]} \frac{\left\| \boldsymbol{\mathcal{BB}}_{p+1}(\lambda) \right\|^2}{\lambda(1-\lambda)}, \text{ for } \gamma_x \in (0,1)$$

$$\boldsymbol{(ii)} \quad \mathcal{SW}_n^{ols}(\lambda;\tau_0) \Rightarrow \sup_{\lambda\in\Lambda_\eta} \boldsymbol{\Delta}_0^{ols}(\lambda;\tau_0)'\left[\boldsymbol{\Sigma}_0^{-1}(\lambda;\tau_0)\right]\boldsymbol{\Delta}_0^{ols}(\lambda;\tau_0), \text{for } \gamma_x = 1$$

where $\boldsymbol{\mathcal{BB}}_{p+1}(.)$ is a vector of $(p+1)$ independent Brownian bridge processes on $\mathcal{D}_{\mathbb{R}^{p+1}}([0,1])$. Denote with

$$\boldsymbol{\Sigma}_0^{-1}(\lambda;\tau_0) := f_{u_t(\tau)}(0)^2\Big[\mathbb{S}_{xx}(\lambda) - \mathbb{S}_{xx}(\lambda)\mathbb{S}_{xx}^{-1}(1)\mathbb{S}_{xx}(\lambda)\Big]$$

$$\boldsymbol{\Delta}_0^{ols}(\lambda;\tau_0) := \mathbb{S}_{xx}^{-1}(\lambda)\begin{bmatrix} B_{\psi_\tau}(\lambda) \\ \int_0^\lambda \boldsymbol{J}_c(r)dB_{\psi_\tau} \end{bmatrix} - \left[\mathbb{S}_{xx}(1) - \mathbb{S}_{xx}(\lambda)\right]^{-1}\begin{bmatrix} B_{\psi_\tau}(1) - B_{\psi_\tau}(\lambda) \\ \int_0^1 \boldsymbol{J}_c(r)dB_{\psi_\tau} - \int_0^\lambda \boldsymbol{J}_c(r)dB_{\psi_\tau} \end{bmatrix}.$$

where $\boldsymbol{\Sigma}_0^{-1}(\lambda;\tau_0) \in \mathbb{R}^{(p+1)\times(p+1)}$ and $\boldsymbol{\Delta}_0^{ols}(\lambda;\tau_0) \in \mathbb{R}^{(p+1)\times n}$ since the model included both an intercept and slopes.

The proof of Proposition 3.2 can be found in the Appendix of the paper. Notice that $\boldsymbol{\Sigma}_0^{-1}(\lambda;\tau_0)$ represents the weakly convergence result of the inverse of the covariance matrix of the stochastic process $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(2)}(\tau_0) - \boldsymbol{\beta}_{(2)}(\tau_0))$, in which case $\mathcal{W}_n(\tau_0)$ is the Wald statistic for testing the null hypothesis $\mathcal{H}_0 : \hat{\boldsymbol{\beta}}_{(2)}(\tau_0) = \mathbf{0}$. Furthermore, notice that the event $\{y(t) \leq \boldsymbol{x}_{t-1}'\beta(\tau)\}$ is distributed exactly as a Bernoulli(τ) conditional on X regardless of the sample size. Therefore, any test statistic that depends only on this event, X, will have a distribution that does not depend on any unknown parameters in finite samples and thus can be used to construct valid finite sample inference statements.

IVX-Wald test statistic

The instrumentation methodology proposed by Phillips and Magdalinos (2009) has been proved to be robust in filtering abstract degree of persistence in predictive regression models (see also Phillips and Lee (2013, 2016)). Our research objective in this section is to study the asymptotic behaviour of the proposed structural break tests based on the endogenous instrumentation procedure in nonstationary quantile predictive regressions. The corresponding limit results for the self-normalized sup IVX-Wald test function as a break detector for the coefficients of linear predictive regressions is examined in Chapter 1. Specifically, the supremum IVX-Wald test corresponds to the maximum[20] of a sequence of test statistics constructed based on sequential sample splitting locations such that $\kappa = \lfloor \lambda n \rfloor$ where $\lambda \in \Lambda_\eta := [\eta, 1-\eta]$ with $0 < \eta < 1/2$. Furthermore, we employ the dequantiled model structure and denote with $\widehat{\boldsymbol{\beta}}_1^{ivx}(\lambda; \tau_0)$ and $\widehat{\boldsymbol{\beta}}_2^{ivx}(\lambda; \tau_0)$ the IVX based estimators for the two sub-samples occurred at each splitting step. Therefore, these estimators are computed via the following expressions

$$\widehat{\boldsymbol{\beta}}_1^{ivx}(\kappa; \tau_0) = \left(\frac{1}{\kappa} \sum_{t=1}^{\kappa+j} \tilde{\boldsymbol{z}}_{1,t-1} \boldsymbol{x}'_{1,t-1} \right)^{-1} \left(\frac{1}{\kappa} \sum_{t=1}^{\kappa+j} \tilde{\boldsymbol{z}}_{1,t-1} y_t \right), \tag{3.78}$$

$$\widehat{\boldsymbol{\beta}}_2^{ivx}(\kappa; \tau_0) = \left(\frac{1}{n-\kappa} \sum_{t=\kappa+1+j}^{n} \tilde{\boldsymbol{z}}_{2,t-1} \boldsymbol{x}'_{2,t-1} \right)^{-1} \left(\frac{1}{n-\kappa} \sum_{t=\kappa+1+j}^{n} \tilde{\boldsymbol{z}}_{2,t-1} y_t \right). \tag{3.79}$$

where $\kappa = \lfloor \lambda n \rfloor$ for some $0 < \lambda < 1$ and the indicator $j \in \{0, ..., (n-\kappa)\}$ shows that a sequence of parameter estimates is obtained by moving along all proportions within the compact set $\Lambda_\eta = [\eta, 1-\eta]$, to compute the maximum Wald statistic. However, for notation convenience we drop the index notation $(\kappa + j)$ and $(\kappa + 1 + j)$ which can be confused with notation used for time-varying parameter estimates. Also, $\boldsymbol{x}_{1,t-1} := \boldsymbol{x}_{t-1} \mathbb{1}\{t \leq \kappa\}$ and $\boldsymbol{x}_{2,t-1} := \boldsymbol{x}_{t-1} \mathbb{1}\{t > \kappa\}$. Furthermore, $\tilde{\boldsymbol{Q}}_1(\lambda; \tau_0)$ and $\tilde{\boldsymbol{Q}}_2(\lambda; \tau_0)$ denotes the covariance matrices which correspond to the two subsample parameter estimates and permits to decompose the covariance matrix[21] of the test with respect to each regime

$$\tilde{\boldsymbol{Q}}_1(\lambda; \tau_0) = \left(\tilde{\boldsymbol{Z}}'_1 \boldsymbol{X}_1 \right)^{-1} \left(\tilde{\boldsymbol{Z}}'_1 \tilde{\boldsymbol{Z}}_1 \right) \left(\boldsymbol{X}'_1 \tilde{\boldsymbol{Z}}_1 \right)^{-1} \quad \tilde{\boldsymbol{Q}}_2(\lambda; \tau_0) = \left(\tilde{\boldsymbol{Z}}'_2 \boldsymbol{X}_2 \right)^{-1} \left(\tilde{\boldsymbol{Z}}'_2 \tilde{\boldsymbol{Z}}_2 \right) \left(\boldsymbol{X}'_2 \tilde{\boldsymbol{Z}}_2 \right)^{-1}$$

Then, under the null hypothesis, $\mathcal{H}_0^{(A)}$, the sup IVX-Wald statistic is formulated as

$$\mathcal{SW}_n^{ivx}(\lambda; \tau_0) := \sup_{\lambda \in \Lambda_\eta} \left\{ \Delta \widehat{\boldsymbol{\beta}}_n^{ivx}(\lambda; \tau_0)' \left[\widehat{\boldsymbol{V}}_n^{ivx}(\lambda; \tau_0) \right]^{-1} \Delta \widehat{\boldsymbol{\beta}}_n^{ivx}(\lambda; \tau_0) \right\} \tag{3.80}$$

where $\Delta \widehat{\boldsymbol{\beta}}_n^{ivx}(\lambda; \tau_0) := \left(\widehat{\boldsymbol{\beta}}_1^{ivx}(\lambda; \tau_0) - \widehat{\boldsymbol{\beta}}_2^{ivx}(\lambda; \tau_0) \right)$ and $\widehat{\boldsymbol{V}}_n^{ivx}(\lambda; \tau_0) := \tilde{\boldsymbol{Q}}_1(\lambda; \tau_0) + \tilde{\boldsymbol{Q}}_2(\lambda; \tau_0)$.

---

[20] Further details regarding the formulation of Wald type tests and asymptotic theory is presented in the seminal study of Andrews (1993). The particular framework propose for structural change tests in linear regression models under the assumption of stationary and ergodic time series.

[21] The decomposition of the covariance matrix for the IVX-Wald statistic can be obtained using a formula for inverting partitioned matrices. In particular, since $\boldsymbol{Z}'_1 \boldsymbol{X}_2 = \boldsymbol{Z}'_2 \boldsymbol{X}_1 = 0$ then the matrix inversion formula simplifies further, allowing us to obtain an expression for the variance of the test.

Theorem 3.1. Under the null hypothesis and given that Assumptions 3.1-3.2 hold, then the sup IVX-Wald statistic weakly convergence to limit distribution below

$$\mathcal{SW}_n^{ivx}(\lambda; \tau_0) \Rightarrow \sup_{\lambda \in \Lambda_\eta} \left\{ \boldsymbol{\Delta}_0^{ivx}(\lambda; \tau_0)' \left[\boldsymbol{\Sigma}_0^{ivx}(\lambda; \tau_0) \right]^{-1} \boldsymbol{\Delta}_0^{ivx}(\lambda; \tau_0) \right\} \tag{3.81}$$

where $\Lambda_\eta := [\eta, 1 - \eta]$ with $0 < \eta < 1/2$ and

$$\boldsymbol{\Delta}_0^{ivx}(\lambda; \tau_0) := \boldsymbol{W}_p(\lambda) - \boldsymbol{\Psi}_c(\lambda) \boldsymbol{W}_p(1) \tag{3.82}$$

$$\boldsymbol{\Sigma}_0^{ivx}(\lambda; \tau_0) := \lambda \big(\boldsymbol{I}_p - \boldsymbol{\Psi}_c(\lambda) \big) \big(\boldsymbol{I}_p - \boldsymbol{\Psi}_c(\lambda) \big)' + (1 - \lambda) \boldsymbol{\Psi}_c(\lambda) \boldsymbol{\Psi}_c(\lambda)' \tag{3.83}$$

such that

$$\boldsymbol{\Psi}_c(\lambda) = \begin{cases} \left(\lambda \boldsymbol{\Omega}_{xx} + \int_0^\lambda \boldsymbol{J}_c^\mu(r) d\boldsymbol{J}_c' \right) \left(\boldsymbol{\Omega}_{xx} + \int_0^1 \boldsymbol{J}_c^\mu(r) d\boldsymbol{J}_c' \right)^{-1} & , \text{for } \gamma_x = 1 \\ \\ \lambda \boldsymbol{I}_p & , \text{for } \gamma_x \in (0, 1) \end{cases}$$

where $\boldsymbol{W}_p(.)$ is a $p-$dimensional standard Brownian motion, $\boldsymbol{J}_c(\lambda) = \int_0^\lambda e^{(\lambda - s)\boldsymbol{C}_p} d\boldsymbol{B}(s)$ is an *Ornstein-Uhkenbeck* process and we denote with $\boldsymbol{J}_c^\mu(\lambda) = \boldsymbol{J}_c(\lambda) - \int_0^1 \boldsymbol{J}_c(s) ds$ and $\boldsymbol{W}_p^\mu(\lambda) = \boldsymbol{W}_p(\lambda) - \int_0^1 \boldsymbol{W}(s) ds$ the demeaned processes of $\boldsymbol{J}_c(\lambda)$ and $\boldsymbol{W}_p(\lambda)$ respectively.

Remark 3.7. Notice that inference on $\boldsymbol{\beta}_n^{ivx}(\tau)$ critically depends on the estimator of the covariance matrix $\widehat{\boldsymbol{V}}_n^{ivx}(\lambda; \tau_0)$. Moreover, the estimation of the sparsity coefficient does not affect the estimation accuracy when constructing test statistics in quantile time series models due to the self-normalized property of Wald type tests. On the other hand, the robust estimation of the covariance matrix is ensured by employing fully modified type of transformations as in the linear model (see, Kostakis et al. (2015)).

Theorem 3.1 presents the asymptotic distribution of the sup IVX-Wald test under the null hypothesis of a single unknown break-point. Furthermore, it covers some practical considerations that arise in empirical work especially with respect to the persistence properties of regressors. As we can observe from the asymptotic behaviour of the test for local unit root regressors (high persistence), it converges to a nonstandard and nonpivotal distribution. On the other hand, for mildly integrated regressors the test behaves in large samples similar to the sup OLS-Wald test which weakly converge into a Brownian bridge type of limit. In the former case, a comparison of the limit distributions of the two tests does not necessarily indicate which test statistic might have better performance in detecting structural breaks to the coefficients of nonstationary quantile predictive regressions. To investigate the particular aspect, we use simulation experiments where allow us to use a suitable experimental design that accommodates these conditions. The proof of Theorem 3.1 can be found in the Appendix of the paper.

Some important implications follow from Theorem 3.1. More precisely, our asymptotic theory analysis confirms some of the conclusions drawn in similar studies. For instance, Hanson (2002) and Seo (1998) (see also Georgiev et al. (2018)) demonstrated that testing for structural breaks with integrated regressors based on the OLS estimation method converges to a nonstandard and nonpivotal limiting distribution. Furthermore, although the IVX-Wald statistic is proved to be robust to abstract degree of persistence when testing for parameter restrictions, within the structural break testing framework due to the presence of the unknown break-point location, the sup IVX-Wald test similar to the OLS counterpart is coverages to a nonpivotal limiting distribution, which is not Brownian bridge (as defined in the stationary case) even though it is still tied down.

Nevertheless, some interesting further simplifications occur; for instance under the assumption of mildly integrated regressors, it can be easily proved that the limiting distribution of the sup IVX-Wald test converges to a normalized Brownian bridge limit. This occurs due to the asymptotic matrix moments such that, for $0 < \gamma_x < 1$ it holds that $\sum_{t=1}^{n} \boldsymbol{x}_{t-1}\tilde{\boldsymbol{z}}_{t-1}' \Rightarrow -\boldsymbol{\Omega}_{xx}\boldsymbol{C}_z^{-1}$ by expression (20) in Phillips and Magdalinos (2009). Thus, it also holds that $\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\tilde{\boldsymbol{z}}_{t-1}' \Rightarrow -\lambda\boldsymbol{\Omega}_{xx}\boldsymbol{C}_z^{-1}$, which implies that $\boldsymbol{\Psi}_c(\lambda) = \lambda \boldsymbol{I}_p$. Furthermore, we also consider the limiting distribution of the sup IVZ-Wald test.

Corollary 3.5. Under the null hypothesis, $\mathcal{H}_0^{(A)}$ and suppose that Assumptions 3.1-3.2 hold, then the sup IVZ-Wald statistic weakly convergence to the asymptotic distribution below

$$\mathcal{SW}_n^{ivz}(\lambda; \tau_0) \Rightarrow \sup_{\lambda \in \Lambda_\eta} \frac{\left\| \boldsymbol{\mathcal{BB}}_p(\lambda) \right\|^2}{\lambda(1-\lambda)}, \quad \text{for} \quad 0 < \gamma_x \leq 1. \tag{3.84}$$

where $\boldsymbol{\mathcal{BB}}_p(.)$ is a vector of p independent Brownian bridge processes on $\mathcal{D}_{\mathbb{R}^p}([0,1])$.

In summary, in this section we show that the limiting distribution of the sup IVX-Wald test statistic under high persistence is nonstandard and nonpivotal. Furthermore, the particular limit result simplifies when regressors in the model are assumed to be mildly integrated resulting to weakly convergence into a Brownian bridge type limit. On the other hand, when we construct the test statistic based on the IVZ estimator then the sup IVZ-Wald test converges into a Brownian bridge type of limit regardless of the degree of persistence. Lastly, for a known break-point Wald type tests converge to a nuisance-parameter free limiting distribution, simplifying this way statistical inference. The proof of Corollary 3.5 can be found in the Appendix.

Another important aspect we consider for the development of the asymptotic theory of the paper, is the classical result of Huber for models with nonstandard conditions such as quantile regression models. More specifically, the first order condition (FOC) defined as the right derivative of the objective function plays a key role in deriving the asymptotic theory of estimators for the quantile model. In particular, we can show that the parameter vector estimator solves these FOC and then apply a Bahadur representation for the estimator. All these results hold with almost surely convergence in large samples.

3.4 Testing for break across multiple quantiles

3.4.1 Preliminary Setting

Testing for breaks at multiple quantiles implies that there is quantile-varying effects of these nonstationary regressors which implies that quantile predictability robust against parameter instability can only be detected if we test for structural breaks across multiple quantiles. In particular, when the quantile level is allowed to vary within the compact set $(0,1)$ this implies that the proposed test statistics[22] can detect for possible breaks in the coefficients of the nonstationary quantile predictive regression across multiple quantile levels[23]. The testing procedure under this scenario generalizes the search for structural breaks since parameter instability in the relation between predictant and predictors can occur across any quantile level within the admissible set.

Test statistics and functionals

Thus, we extend the test statistics $\mathcal{DQ}_n(\lambda,\tau)$ and $\mathcal{DW}_n(\lambda,\tau)$ as structural break detectors across multiple quantiles using the double supremum operator.

$$\mathcal{DQ}_n(\lambda,\tau) := \sup_{\tau\in\mathcal{T}_\iota} \sup_{\lambda\in[0,1]} \left\| \frac{1}{\sqrt{\tau(1-\tau)}} \left[\mathcal{J}_n(\lambda,\tau,\widehat{\boldsymbol{\beta}}_n(\tau)) - \lambda \mathcal{J}_n(1,\tau,\widehat{\boldsymbol{\beta}}_n(\tau)) \right] \right\|_\infty, \tag{3.85}$$

$$\mathcal{DW}_n(\lambda,\tau) := \sup_{\tau\in\mathcal{T}_\iota} \sup_{\lambda\in\Lambda_\eta} \left\{ n\, \Delta\widehat{\boldsymbol{\beta}}_n(\lambda,\tau)' \widehat{\boldsymbol{V}}_n(\lambda,\tau)^{-1} \Delta\widehat{\boldsymbol{\beta}}_n(\lambda,\tau) \right\}. \tag{3.86}$$

where $\mathcal{T}_\iota := [\iota, 1-\iota]$ for some $0<\iota<1/2$ and $(\lambda,\tau) \in (0,1)\times(0,1)$.

In particular, the proposed testing methodology for detecting breaks in quantile predictive regression models is based on the double supremum tests (see also Qu (2008) and Andrews (1993)) which are suitable test functions for evaluating *Testing Hypothesis B* (see Section 3.2.3). In particular, under the specific econometric environment, statistical inference encompasses cases such as structural breaks in the coefficients at different quantile levels such as the lower or the upper tails of the underline conditional quantile distribution function. Consequently, in order to strengthen the ability of these test statistics to detect structural-break which implies improving the monotonic performance of the asymptotic power function, the admissible quantile set $\mathcal{T}_\iota \subset (0,1)$ can be modified accordingly using for instance an asymmetric trimming factor.

[22]A relevant line of literature to structural break testing is concerned with the development of methodologies for specification testing of conditional distributions (as in Escanciano and Velasco (2010)). Although the null hypothesis for these procedures do not directly apply to the testing hypotheses we consider, the idea of testing for the correct specification of quantile models over multiple quantiles has similar intuition to detecting for structural breaks at random quantiles within the compact set $(0,1)$.

[23]Within our setting the estimation procedure implies that the quantile regression model is fitted on the pair of data $\{y_t, \boldsymbol{x}_t\}_{t=1}^n$ assuming a Gaussian error distribution. Moreover, based on the optimization function the model estimator corresponds to a parameter quantile dependent vector without imposing any assumptions regarding the presence of quantile effects. Therefore, structural break testing can only be interpreted as detecting for the presence of possible parameter instability for this quantile dependent vector either around a fixed quantile level or across multiple quantile levels within the set $(0,1)$.

Regularity Conditions for Weak Convergence

Therefore, the asymptotic theory of these tests statistics can be developed by examining their limit distributions in relation to the asymptotic behaviour of a two-parameter Gaussian process in τ and λ within a product of the compact set $(0,1)^2$. More precisely, following the framework of Qu (2008) further regularity assumptions can be imposed to ensure stochastic equicontinuity of the process $S_n(\lambda, \tau, \beta_0(\tau))$ on $(\mathcal{K}, d)$, where $\mathcal{K} = [0,1] \times [0,1]$ and d is some metric with well-defined properties on $\mathcal{K}$. Assumption 3.3 below imposes additional conditions that ensure the weakly convergence of the tests to two-parameter Gaussian processes.

Assumption 3.3 (Qu (2008))**.** For the vector-valued sequence $\left\{ \psi_\tau(u_t(\tau), \boldsymbol{X}_{t-1,n}, \tilde{\boldsymbol{Z}}_{t-1,n}) \right\}$, let $\mathcal{F}_t$ to denote the natural filtration. Then, for the functionals $\mathcal{DQ}_n(\lambda, \tau)$ and $\mathcal{DW}_n(\lambda, \tau)$:

(i) there exists a fixed scalar $\boldsymbol{\delta}_n$ such that the array $\mathcal{J}_{nt} = \boldsymbol{\delta}_n \mathcal{J}_t(\lambda, \tau, \boldsymbol{b}_n(\tau))$ satisfies $\mathcal{J}_{n,\lfloor \lambda n \rfloor}(., \tau) \Rightarrow \mathcal{J}^o(\lambda, \tau)$ for some $0 < \lambda < 1$ and $\tau \in (0,1)$, where $\mathcal{J}^o(\lambda, \tau)$ is continuous almost surely;

(ii) the functionals $\mathcal{DQ}_n(\lambda, \tau)$ and $\mathcal{DW}_n(\lambda, \tau)$ converge weakly in the Skorokhod space $\mathcal{D}\left([0,1]\right)^2$ equipped with the product J_1 topology, to a two-dimensional Gaussian process, such that $\{(B(\mathsf{u}), B(\mathsf{v})), 0 \leq \mathsf{u}, \mathsf{v} \leq 1\}$. Moreover, $\{B(\mathsf{u}), 0 \leq \mathsf{u} \leq 1\}$ and $\{B(\mathsf{v}), 0 \leq \mathsf{v} \leq 1\}$ are assumed to be two independent Gaussian processes.

Lemma 3.4 (Qu (2008))**.** Under the null hypothesis $\mathcal{H}_0^{(B)}$ and given that Assumptions 3.1-3.3 hold, then:

(i) uniformly in $(\lambda, \tau) \in \Lambda_\eta \times \mathcal{T}_\iota$ and for n large,

$$\hat{S}_n(\lambda, \tau, \widehat{\boldsymbol{\theta}}_1(\lambda, \tau)) = S_n(\lambda, \tau, \boldsymbol{\theta}_0(\tau)) + \lambda \boldsymbol{L}_{nx}\left(\tau, \boldsymbol{\theta}_n(\tau)\right) \widehat{\boldsymbol{\mathcal{E}}}_1(\lambda, \tau) + o_{\mathbb{P}}(1),$$

where $\widehat{\boldsymbol{\mathcal{E}}}_1(\lambda, \tau) := \sqrt{n}\left(\widehat{\boldsymbol{\theta}}_1(\lambda, \tau) - \boldsymbol{\theta}_0(\tau) \right)$ and $\widehat{\boldsymbol{\theta}}_1(\lambda, \tau)$ denotes the estimate of $\boldsymbol{\theta}_0(\tau)$ using the sample up to $\lfloor \lambda n \rfloor$, and $\boldsymbol{L}_{nx}(\tau, \boldsymbol{\theta}_n(\tau))$ as in Definition 3.1.

(ii) uniformly in $(\lambda, \tau) \in [0,1] \times \mathcal{T}_\iota$ and for n large,

$$\hat{S}_n(\lambda, \tau_0, \widehat{\boldsymbol{\theta}}_n(\lambda, \tau)) = S_n(\lambda, \tau, \boldsymbol{\theta}_0(\tau)) + \lambda \boldsymbol{L}_{nx}(\tau, \boldsymbol{\theta}_n(\tau)) \widehat{\boldsymbol{\mathcal{E}}}_n(\tau) + o_{\mathbb{P}}(1),$$

where $\widehat{\boldsymbol{\mathcal{E}}}_n(\tau) := \sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n(\tau) - \boldsymbol{\theta}_0(\tau) \right)$ and $\widehat{\boldsymbol{\theta}}_n(\tau)$ denotes the (OLS) estimate of $\boldsymbol{\theta}_0(\tau)$ using the full sample.

Denote with $\boldsymbol{\mathcal{B}}_p(\mathsf{u}, \mathsf{v}) = (B_1(\mathsf{u}, \mathsf{v}), ..., B_p(\mathsf{u}, \mathsf{v}))'$ be a $p-$vector of independent Gaussian processes with each component defined on $[0,1]^2$ having zero mean and a covariance function specified below

$$\mathsf{Cov}(B_i(\mathsf{r}, \mathsf{u}), B_i(\mathsf{s}, \mathsf{v})) := \mathbb{E}[B_i(\mathsf{r}, \mathsf{u}) B_i(\mathsf{s}, \mathsf{v})] = (\mathsf{r} \wedge \mathsf{s} - \mathsf{rs})(\mathsf{u} \wedge \mathsf{v} - \mathsf{uv}). \tag{3.87}$$

where $\{\mathsf{r}, \mathsf{u}, \mathsf{s}, \mathsf{v}\}$ are some random elements in a compact set such that $\mathsf{r} \neq \mathsf{s} \neq \mathsf{u} \neq \mathsf{v}$.

3.4.2 Asymptotic Theory

Fluctuation type tests

We begin with deriving the limiting distribution for the fluctuation type test statistic. Notice that the formulation of the fluctuation OLS based test given by (3.85) corresponds to the nonstationary quantile predictive regression model with both a model intercept and slopes. Furthermore, we consider testing for structural breaks across multiple quantiles, (i.e., random quantiles within the compact set $(0,1)$). Therefore, to derive the limiting distribution of the test we utilize a two-parameter Brownian motion and employ weak convergence arguments which apply in the Skorokhod topology $\mathcal{D}([0,1])$.

Proposition 3.3. Suppose that Assumptions 3.1-3.3 hold, then under the null hypothesis, $\mathcal{H}_0^{(B)}$, the fluctuation type statistics weakly converge to the limiting distributions below

$$\textbf{(i)}\ \mathcal{DQ}_n^{ols}(\lambda,\tau) := \sup_{\tau\in\mathcal{T}_\iota}\sup_{\lambda\in[0,1]}\left\|\frac{1}{\sqrt{\tau(1-\tau)}}\left[\hat{\mathcal{J}}_n\left(\lambda,\tau,\widehat{\boldsymbol{\theta}}_n^{ols}(\tau)\right)-\lambda\hat{\mathcal{J}}_n\left(1,\tau,\widehat{\boldsymbol{\theta}}_n^{ols}(\tau)\right)\right]\right\|_\infty$$

$$\Rightarrow\begin{cases}\sup\limits_{\tau\in\mathcal{T}_\iota}\sup\limits_{\lambda\in[0,1]}\left\|\boldsymbol{\mathcal{BB}}^\star_{p+1}(\lambda,\tau)\right\|_\infty, & \text{when } \gamma_x\in(0,1)\\ \sup\limits_{\tau\in\mathcal{T}_\iota}\sup\limits_{\lambda\in[0,1]}\mathbb{S}_{xx}^{-1/2}\times\left\{\begin{bmatrix}\boldsymbol{\mathcal{BB}}_{\psi_\tau}(\lambda)_{(1\times n)}\\ \boldsymbol{\mathcal{JB}}_{\psi_\tau}(\lambda)_{(p\times n)}\end{bmatrix}_{(p+1)\times n}\right\}, & \text{when } \gamma_x=1\end{cases}$$

$$\textbf{(ii)}\ \mathcal{DQ}_n^{ivx}(\lambda,\tau) := \sup_{\tau\in\mathcal{T}_\iota}\sup_{\lambda\in[0,1]}\left\|\frac{1}{\sqrt{\tau(1-\tau)}}\left[\hat{\mathcal{J}}_n\left(\lambda,\tau,\widehat{\boldsymbol{\beta}}_n^{ivx}(\tau)\right)-\lambda\hat{\mathcal{J}}_n\left(1,\tau,\widehat{\boldsymbol{\beta}}_n^{ivx}(\tau)\right)\right]\right\|_\infty$$

$$\Rightarrow\sup_{\tau\in\mathcal{T}_\iota}\sup_{\lambda\in[0,1]}\left\|\boldsymbol{\mathcal{BB}}^\star_{p}(\lambda,\tau)\right\|_\infty,\qquad \text{when } \gamma_x=(0,\gamma_z)$$

$$\textbf{(iii)}\ \mathcal{DQ}_n^{ivz}(\lambda,\tau) := \sup_{\tau\in\mathcal{T}_\iota}\sup_{\lambda\in[0,1]}\left\|\frac{1}{\sqrt{\tau(1-\tau)}}\left[\hat{\mathcal{J}}_n\left(\lambda,\tau,\widehat{\boldsymbol{\beta}}_n^{ivz}(\tau)\right)-\lambda\hat{\mathcal{J}}_n\left(1,\tau,\widehat{\boldsymbol{\beta}}_n^{ivz}(\tau)\right)\right]\right\|_\infty$$

$$\Rightarrow\sup_{\tau\in\mathcal{T}_\iota}\sup_{\lambda\in[0,1]}\left\|\boldsymbol{\mathcal{BB}}^\star_{p}(\lambda)\right\|_\infty,\qquad \text{when } \gamma_x=(0,1]$$

where $\boldsymbol{\mathcal{BB}}^\star_p(.,.)$ is a vector of p independent Brownian bridge processes on $\mathcal{D}_{\mathbb{R}^p}([0,1])$,

$$\mathbb{S}_{xx} := \begin{bmatrix}1 & \int_0^1 \boldsymbol{J}_c(r)'dr\\ \int_0^1\boldsymbol{J}_c(r)dr & \int_0^1\boldsymbol{J}_c(r)\boldsymbol{J}_c(r)'\end{bmatrix}_{(p+1)\times(p+1)},\quad \text{with } \ 0<r<1$$

such that $\mathbb{S}_{xx}$ is a positive definite stochastic matrix, $\boldsymbol{\mathcal{BB}}_{\psi_\tau}(\lambda) := B_{\psi_\tau}(\lambda)-\lambda B_{\psi_\tau}(1)$ and $\boldsymbol{\mathcal{JB}}_{\psi_\tau}(\lambda) := \int_0^\lambda \boldsymbol{J}_c(r)dB_{\psi_\tau}-\lambda\int_0^1\boldsymbol{J}_c(r)dB_{\psi_\tau}$.

Remark 3.8. The main ingredients in the proof of Proposition 3.3 are the three results:

(i) A Bahadur representation that holds uniformly in $(\lambda, \tau) \in \Lambda_\eta \times \mathcal{T}_\iota$ (Lemma 3.4),

(ii) A uniform approximation to the re-centered subgradient process

$$S_n(\lambda, \tau, \boldsymbol{\beta}_0(\tau) + n^{-1/2}\xi) \tag{3.88}$$

(iii) $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{1n}(\lambda, \tau) - \boldsymbol{\beta}_0(\tau)\right) = \mathcal{O}_{\mathbb{P}}(1)$ uniformly in $(\lambda, \tau) \in \Lambda_\eta \times \mathcal{T}_\iota$.

Wald type tests

The above functionals and test functions are employed for the purpose of conducting statistical inference with a fixed significance level $0 < \alpha < 1$ for *Testing Hypothesis B*; which corresponds to structural break detection in the quantile dependent parameter vector of the model for all $\tau \in \mathcal{T}_\iota$. Similarly, we consider the asymptotic theory for the Wald type test statistics that correspond to the double supremum functional. Therefore, the following formulation holds

$$\mathcal{DW}_n(\lambda, \tau) := \sup_{\tau \in \mathcal{T}_\iota} \sup_{\lambda \in \Lambda_\eta} \left\{ n \, \Delta\widehat{\boldsymbol{\beta}}_n(\lambda, \tau)' \widehat{\boldsymbol{V}}_n(\lambda, \tau)^{-1} \Delta\widehat{\boldsymbol{\beta}}_n(\lambda, \tau) \right\}.$$

Proposition 3.4. Suppose that Assumptions 3.1-3.3 hold, then under the null hypothesis, $\mathcal{H}_0^{(B)}$, the sup OLS-Wald test statistic weakly converge to the limiting distribution below

(*i*) $\mathcal{DW}_n^{ols}(\lambda, \tau) \Rightarrow \sup_{\tau \in \mathcal{T}_\iota} \sup_{\lambda \in \Lambda_\eta} \dfrac{\left\| \boldsymbol{\mathcal{BB}}_p^{\star}(\lambda, \tau) \right\|^2}{\lambda(1-\lambda)\tau(1-\tau)}$, for $\gamma_x \in (0, 1)$

(*ii*) $\mathcal{DW}_n^{ols}(\lambda, \tau) \Rightarrow$

$$\sup_{\tau \in \mathcal{T}_\iota} \sup_{\lambda \in \Lambda_\eta} \frac{f_{u_t(\tau)}(0)^2}{\lambda(1-\lambda)\tau(1-\tau)} \begin{bmatrix} B_{\psi_\tau}(1) \\ \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix}' \otimes \mathbb{S}_{xx}^{-1} \otimes \begin{bmatrix} B_{\psi_\tau}(1) \\ \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix}, \text{ for } \gamma_x = 1.$$

where $\boldsymbol{\mathcal{BB}}_p^{\star}(.,.)$ is a vector of p independent Brownian bridge processes on $\mathcal{D}_{\mathbb{R}^p}([0, 1])$.

Next, we focus on the formulation and asymptotic theory of the double supremum IVX-Wald test statistic under the null hypothesis, $\mathcal{H}_0^{(B)}$. Similar to the setting of testing for structural breaks for a fixed quantile level in the case of a random quantile level within the compact set $(0, 1)$, the crucial aspect to establish is whether the limit distribution of the test under the null hypothesis is nuisance-parameter free.

Theorem 3.2. Suppose that Assumptions 3.1-3.3 hold, then under the null hypothesis, $\mathcal{H}_0^{(B)}$, sup IVX-Wald statistic weakly convergence to the limiting distribution below

$$\textbf{\textit{(i)}} \qquad \mathcal{DW}_n^{ivx}(\lambda, \tau) \Rightarrow \sup_{\tau \in \mathcal{T}_\iota} \sup_{\lambda \in \Lambda_\eta} \left\{ \boldsymbol{\Delta}_0^{ivx}(\lambda, \tau)' \left[\boldsymbol{\Sigma}_0^{ivx}(\lambda, \tau) \right]^{-1} \boldsymbol{\Delta}_0^{ivx}(\lambda, \tau) \right\} \tag{3.89}$$

where $\Lambda_\eta := [\eta, 1-\eta]$ with $0 < \eta < 1/2$ and

$$\boldsymbol{\Delta}_0^{ivx}(\lambda, \tau) := \boldsymbol{W}_p(\lambda) - \boldsymbol{\Psi}_c(\lambda) \boldsymbol{W}_p(1) \tag{3.90}$$

$$\boldsymbol{\Sigma}_0^{ivx}(\lambda, \tau) := \lambda \big(\boldsymbol{I}_p - \boldsymbol{\Psi}_c(\lambda) \big) \big(\boldsymbol{I}_p - \boldsymbol{\Psi}_c(\lambda) \big)' + (1-\lambda) \boldsymbol{\Psi}_c(\lambda) \boldsymbol{\Psi}_c(\lambda)' \tag{3.91}$$

such that

$$\boldsymbol{\Psi}_c(\lambda) = \begin{cases} \left(\lambda \boldsymbol{\Omega}_{xx} + \int_0^\lambda \boldsymbol{J}_c^\mu(r) d\boldsymbol{J}_c' \right) \left(\boldsymbol{\Omega}_{xx} + \int_0^1 \boldsymbol{J}_c^\mu(r) d\boldsymbol{J}_c' \right)^{-1} & , \text{for } \gamma_x = 1 \\ \\ \lambda \boldsymbol{I}_p & , \text{for } \gamma_x \in (0,1) \end{cases}$$

Moreover, the corresponding sup IVZ-Wald statistic weakly converge to the limiting distribution given below

$$\textbf{\textit{(ii)}} \qquad \mathcal{DW}_n^{ivz}(\lambda, \tau) \Rightarrow \sup_{\tau \in \mathcal{T}_\iota} \sup_{\lambda \in \Lambda_\eta} \frac{\left\| \boldsymbol{\mathcal{BB}}_p^\star(\lambda, \tau) \right\|^2}{\lambda(1-\lambda)}, \quad \text{for} \quad 0 < \gamma_x \leq 1. \tag{3.92}$$

where $\boldsymbol{\mathcal{BB}}_p^\star(.,.)$ is a vector of p independent Brownian bridge processes on $\mathcal{D}_{\mathbb{R}^p}([0,1])$.

3.5 Monte Carlo Simulation Study

Practically, it is unclear how well the asymptotic theory can provide reliable reference and guidance in finite samples when applied to time series data since usually they can exhibit abstract degree of persistence. However, under the assumption that regressors incorporated in the quantile predictive regression model are generated by near unit root processes for which their asymptotic behaviour is well-understood, then our test statistics can provide an indication regarding the ability of the testing procedures in detecting structural breaks in coefficients of nonstationary quantile predictive regression models. Thus, to investigate the finite sample performance of the proposed tests for their adequacy in detecting parameter instability we focus on the empirical size simulation results as well as on asymptotic power analysis under relevant sequence of local alternatives.

3.5.1 Experimental Design

We conduct a number of Monte Carlo experiments to evaluate the performance of the limit distribution of the Wald and fluctuation type statistics against the conventional χ^2 asymptotic approximation. We simulate the following data generating process

$$y_t = \alpha + \sum_{j=1}^{3} \beta_j x_{j,t-1} + u_t, \qquad 1 \leq t \leq n \tag{3.93}$$

$$\boldsymbol{x}_t = \begin{bmatrix} \varrho_n(c_j, \gamma_x) & 0 & 0 \\ 0 & \varrho_n(c_j, \gamma_x) & 0 \\ 0 & 0 & \varrho_n(c_j, \gamma_x) \end{bmatrix} \boldsymbol{x}_{t-1} + \boldsymbol{v}_t \tag{3.94}$$

where $\boldsymbol{x}_t = (x_{1t}, x_{1t}, x_{1t})'$, $\boldsymbol{v}_t = (v_{1t}, v_{1t}, v_{1t})'$ and $\varrho_n(c_j, \gamma_x) = \left(1 + \frac{c_j}{n^{\gamma_x}}\right)$. Then, the innovation sequence of the model, denoted with $\boldsymbol{e}_t = (u_t, \boldsymbol{v}_t)'$ is generated such that $\boldsymbol{e}_t \sim \mathcal{N}\left(\boldsymbol{0}_{(p+1)\times 1}, \boldsymbol{\Sigma}_{ee}\right)$, where $\boldsymbol{\Sigma}_{ee}$ is an $(p+1) \times (p+1)$ positive-definite covariance matrix with a pre-specified variance-covariance structure given as below

$$\boldsymbol{\Sigma}_{ee} = \begin{bmatrix} \sigma^2_{uu} & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \boldsymbol{\Sigma}_{vv} \end{bmatrix} \tag{3.95}$$

where the matrix $\boldsymbol{\Sigma}_{vv}$ is of full rank p, resulting to a nonsingular matrix $\boldsymbol{\Sigma}_{ee}$.

The coefficients of persistence are such that $c_j \in \{-1, -2, -5\}$ and γ_x is defined to be $\gamma_x = 1$ to simulate near unit root predictors and $\gamma_x = 0.75$ to simulate mildly integrated predictors. Moreover, we consider different sample size such that $n \in \{250, 500, 750, 1000\}$. Under the null hypothesis of no parameter instability we use the following parameters $\alpha = 1$, $\beta_1 = 0.25, \beta_2 = 0.75, \beta_3 = -0.50$ and construct the proposed test statistics with significance level $\alpha = 5\%$. For the IVX instrumentation we use $c_z = 1$ and $\gamma_z = 0.95$.

In particular, to construct the test statistics the simulated pair $\{y_t, \boldsymbol{x}_t\}_{t=1}^{n}$ is formulated:

$$y_t = \left(\alpha^{(1)}(\tau) + \sum_{j=1}^{3} \beta_j^{(1)}(\tau) x_{j,t-1}\right) \mathbb{1}\{t \leq \kappa\} + \left(\alpha^{(2)}(\tau) + \sum_{j=1}^{3} \beta_j^{(2)}(\tau) x_{j,t-1}\right) \mathbb{1}\{t > \kappa\} + u_t$$

where $\kappa = \lfloor \lambda n \rfloor$ and the search over all possible subsets occurs for values of $\lambda \in \Lambda_\eta$. Denote with $\boldsymbol{\theta}^{(j)}(\tau) = \left(\alpha^{(j)}(\tau), \beta_1^{(j)}(\tau), \beta_2^{(j)}(\tau), \beta_3^{(j)}(\tau)\right)'$ with $j \in \{1, 2\}$ to be the quantile dependent parameter vector of each of the two regimes, for a fixed quantile τ_0 that belongs in the compact set[24] $\mathcal{T}_\eta$ such that $0 < \eta < \tau_0 < 1 - \eta < 1$.

Then, the testing hypothesis of interest is formulated as below

$$\mathcal{H}_0^{(A)} : \boldsymbol{\theta}_n^{(1)}(\tau) = \boldsymbol{\theta}_n^{(2)}(\tau) \quad \text{versus} \quad \mathcal{H}_1^{(A)} : \boldsymbol{\theta}_n^{(1)}(\tau) \neq \boldsymbol{\theta}_n^{(2)}(\tau) \tag{3.96}$$

with a fixed quantile $\tau_0 \in \mathcal{T}_\eta := [\eta, 1 - \eta]$, for an unknown break-point location $\kappa = \lfloor \lambda n \rfloor$ where $0 < \lambda < 1$ and a significance level $\alpha = 5\%$.

[24] Notice that the compact set $\mathcal{T}_\eta$ falls strictly within the unit interval to allow the conditional distribution to have an unbounded support.

Furthermore, the second set of simulation experiments we consider involves the consideration of the power performance for both the OLS and IVX based tests for detecting structural breaks in the coefficients of the quantile predictive regression model. More precisely, we consider a set of experiments against a sequence of alternatives given by

$$\mathcal{H}_{1n}^{(A)} : \boldsymbol{\beta}_t(\tau) = \boldsymbol{\beta}_0(\tau) + \frac{\delta(\tau)}{\sqrt{n}} \times g\left(\frac{t}{n}\right), \tag{3.97}$$

where $\delta(\tau)$ is a constant that is quantile dependent and $g(t/n)$ is a vector-valued function.

Notice that, we are particularly interested for the asymptotic behaviour of the test statistics under local alternatives. According to White (1996), the particular form of the parameter vector given above allows analysis of our test statistics under a form of local alternative, and therefore supports investigation of local power properties of the associated test procedures. Furthermore, in the literature of the behaviour of testing procedures under local alternatives, the standard approach is to subject the data generating mechanism to a drift, while holding the null hypothesis fixed. However, we do not consider a sequence of global alternatives, rather we focus on the case of a sequence of local alternative hypotheses. Therefore, we are particularly interested to investigate whether the proposed bootstrap procedure can consistently estimate the limiting distribution of the proposed structural break tests based on the OLS and IVX estimators in the context of the nonstationary quantile predictive regression model. For instance, it might be the case that the proposed bootstrap methodology cannot be used for correcting the induced endogeneity bias especially for the nonstationary quantile time series model.

3.5.2 Implementation Procedure

One important aspect for the correct implementation of the fluctuation based tests (which involves the estimation of a subgradient) is that no nuisance parameter is needed (without the LUR parametrization). Thus, for the Wald type statistics, we need a consistent estimate of the variance-covariance matrix $\boldsymbol{\Omega}_0$. In particular, it requires estimating the following matrix[25]

$$\boldsymbol{H}_0 = \underset{n\to\infty}{\text{plim}} \left(n^{-1} \sum_{t=1}^{n} f_{y|x}(F_{y|x}^{-1}(\tau))\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}' \right)^{-1} \tag{3.98}$$

A discussion regarding methodologies for estimating the matrix given by expression (3.98) can be found in Qu (2008). Our asymptotic theory analysis reveals that when using Wald type statistics as structural break detectors in nonstationary quantile predictive regression models, the chosen estimator can affect the limiting distributions of test statistics and therefore its finite-sample performance, especially under the presence of high persistence regressors. In particular, we have proved that when selecting the OLS estimator then the limit distribution is nonstandard and nonpivotal making inference challenging since critical values can be constructed only with the use of bootstrap-based methodologies. When the IVZ estimator is chosen then the limit distribution is proved to be nuisance-parameter free regardless of the persistence properties driving the behaviour of regressors employed when estimating the quantile predictive regression.

[25] Notice that the estimation of the sparsity function will affect the finite-sample performance of the test statistics if not consistently estimated.

Remark 3.9. Another relevant aspect to investigate is the sensitivity of the proposed test statistics to the break-point location within the full sample. Therefore, for instance we are interested to examone whether the proposed tests have roughly equal sensitivity to a break occurring early or late in the sample (see, also Leisch et al. (2000)). We expect that the break-point location will not affect the power performance of the proposed test statistics especially due to the fact that we do not operate within a sequential monitoring scheme in which case parameter estimates and functionals are updated continuously.

Remark 3.10. Furthermore, notice that we compare the performance of the sup OLS-Wald with the sup IVX-Wald for detecting structural break in nonstationary quantile predictive regression models, by comparing the empirical size results based on the bootstrapped test statistics when detecting structural breaks in these models.

Bootstrap-based inference

In this section, we discuss bootstrap based approximations to the asymptotic distribution of the fluctuation and Wald type statistics under the assumption of high persistence regressors. In particular, these bootstrap approximations can be employed to calculate critical values and p-values when evaluating the empirical size and power performance of the proposed test statistics. A bootstrap methodology which is robust to conditional heteroscedasticity of unknown form in the error term of a pure time series autoregressive model is the pairs bootstrap, where one resamples with replacement the vector that collects the dependent variable and its lagged values. Specifically, the asymptotic validity of the pairs bootstrap can be established following the same assumptions as those underlying the validity of the fixed-design wild bootstrap.

Theorem 3.3. Under regularity conditions, it follows that

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}^* \left\{ \sqrt{NT} \left(\hat{\theta}^*_{RD} - \hat{\theta} \right) \leq x \right\} - \mathbb{P}^* \left\{ \sqrt{NT} \left(\hat{\theta}^*_{RD} - \hat{\theta} \right) \leq x \right\} \right| \overset{p}{\to} 0. \tag{3.99}$$

Therefore, investigating the asymptotic validity of the recursive-design bootstrap is essential in establishing the practicality of the methodology in empirical and applied work. Furthermore, with simulation experiments we can evaluate the finite-sample performance of the proposed bootstrap.

Simulation Procedure

The simulation procedure is briefly described as follows.

(1) Under $\mathcal{H}_0$, estimate the coefficients in the model, $y_t = \boldsymbol{x}_{t-1}\boldsymbol{\beta} + u_t$ using the OLS method and obtain the corresponding OLS residuals $\hat{u}_t$. Run the autoregression such as $\boldsymbol{x}_t = \rho_n \boldsymbol{x}_{t-1} + \boldsymbol{v}_t$ where $t = 1, ..., n$ to obtain the vector of residuals $\hat{\boldsymbol{v}}_t$.

(2) Apply the proposed two-stage estimation procedure on the simulated sample in Step **(1)** to obtain estimates of the time-varying coefficients, calculate the estimated nonparametric residual $\tilde{u}_t$, and then estimate the conditional variance of J_n say $\tilde{\sigma}^2$.

(3) Replace with $\hat{u}_t$ and $\tilde{\sigma}^2$ and calculate the test statistic $\hat{H}_n$.

(4) Perform a large number of iterations, say 1,000, to find the empirical distribution of $\left\{\hat{H}_n\right\}$. The critical value at significance level α is given by the $(1-\alpha)-$th quantile.

Example 3.1. Consider the predictive regression model

$$y_t = \beta_0 + \beta_1 x_{t-1} + \epsilon_t \tag{3.100}$$

$$x_t = \rho x_{t-1} + u_t \tag{3.101}$$

Moreover, the innovations $\{\epsilon_t, u_t\}$ are assumed to be i.i.d bivariate normal $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Then, the AR(1) model is estimated via OLS and the residuals $\hat{\epsilon}_t$ are stored. Then, N different pseudo time series data $\{y_t^*, t = 1, ..., T\}$ are generated in a manner consistent with the estimated (no-break) model as below:

$$y_t^* = \hat{\mu} + \hat{\rho} y_{t-1}^* + u_t \tag{3.102}$$

where $\hat{\mu}$ and $\hat{\rho}$ are OLS estimates and the u_t the pseudo-disturbances are drawn randomly with replacement from the estimated residuals $\hat{\epsilon}_t$. Then, for each of these N simulated series, a sup Wald test statistic is calculated, and the $\alpha-$th percentile of the resulting empirical distribution is employed as the $(1-\alpha)-$percent critical value for the proposed testing procedure.

Notice that bootstrap methods can only be expected to work well when they provide a good approximation to the underlying data generating process. In light of this, one potential weakness of the sieve bootstrap approach is that the way that pseudo-disturbances u_t are generated (i.e., drawing randomly from the estimated residuals) might provide a poor description of DGPs that exhibit heteroscedasticity. When a model intercept is included the testing procedure represents a joint test for a simultaneous break in the intercept and the slope coefficients of the model. Furthermore, notice that for the case of high persistence regressors, the size distortions for asymptotic tests can be extremely large and this can be confirmed from empirical size results.

Therefore, our testing procedure aims to disentangle the problem of distinguishing between a persistent time series with no structural breaks for one that is less persistent but has a break in the intercept. Therefore, within our setting we aim to formalize the particular aspect using suitable formulations of the null hypothesis. In practise, most of the literature on testing for structural breaks formulates the hypotheses such that in the statistical model the stochastic process under the null hypothesis of no parameter instability is stationary. In our case, we consider that under the null hypothesis, the quantile predictive regression model with nonstationary regressors has in practise stable coefficients, meaning that the coefficients are constant throughout the sample. Furthermore, the formulation of the null hypothesis, implies that for the development of the asymptotic theory of the distributional properties of a test statistic we still operate under the assumption of nonstationary time series, although the model parameters are assumed to be stable over time. On the other hand, under the alternative hypothesis we operate under the assumption that there is a structural break in the model parameters at some unknown break-point location in the sample, operating this way under the assumption of both nonstationarity as well as unstable model parameters in the form of a sudden structural change.

Thus, it is clear that in our setting we are not testing for stationarity versus nonstationarity. The nonstationarity in the time series is present both under the null as well as under the alternative in the form of the model regressors being generated as near unit root processes. Therefore, generally speaking one can assume that a non-mean reversion is equivalent to nonstationarity, however in our framework specifically we restrict the properties of the underline stochastic processes to be such that both mean-reversion and nonstationarity can coexist, allowing this way to distinguish between different degrees of persistence that a regressor can exhibit regardless of the presence of structural break in the model parameters under the alternative hypothesis.

Furthermore, the particular aspect is discussed in the study of Diebold and Chen (1996). In particular, the asymptotic distribution theory of the supremum Wald test proposed by Andrews (1993), especially when testing for structural break in nonstationary time series models under persistence is an unreliable guide to finite-sample behaviour. Therefore, one of the main aims of this paper is to develop asymptotic theory suitable for structural break tests in nonstationary quantile predictive regression models. For instance, Diebold and Chen (1996) provide a detailed finite-sample evaluation of the size of supremum tests for structural change in a dynamic model, with attention focused on the comparative performance of asymptotic versus bootstrap procedures. Therefore, the particular results are of interest not only from the perspective of testing for structural change, but also from the broader perspective of compiling evidence on the adequacy of bootstrap approximations to finite-sample distributions in econometrics.

In practise, within our setting the algorithm we aim to construct is a bootstrap Monte Carlo. More precisely, a flow chart of the bootstrap Monte Carlo procedure is presented by Diebold and Chen (1996), which we summarize below for convenience. Let $i = 1, ..., M$ index Monte Carlo replications, and let $j = 1, ..., B$ index the bootstrap replications inside each Monte Carlo replication. Nominal size is denote with α. Then the procedure is as following:

Step [1] Draw the vector of innovations $\{\epsilon_1^{(i)}, ..., \epsilon_T^{(i)}\} \sim \mathcal{N}(0, 1)$. Then, generate a vector of true data such that $\{y_1^{(i)}, ..., y_T^{(i)}\}$ based on the following DGP:

$$y_t^{(i)} = \rho y_{t-1}^{(i)} + \epsilon_t^{(i)}, \quad t = 1, ..., T, \tag{3.103}$$

$$y_0 \sim \mathcal{N}\left(0, \frac{1}{(1-\rho^2)}\right). \tag{3.104}$$

and compute the OLS estimator for ρ and $\hat{\rho}^{(i)}$, and the associated de-meaned residuals $\{\hat{\epsilon}_1^{(i)}, ..., \hat{\epsilon}_T^{(i)}\}$. Finally compute the sup Wald test statistic $\mathcal{SW}_T^{(i)}$.

Step [2A] Draw $\{e_1^{(j)}, ..., e_T^{(j)}\}$ by sampling with replacement from the sequence $\{\hat{\epsilon}_1^{(i)}, ..., \hat{\epsilon}_T^{(i)}\}$. Then, generate the pseudo-data $\{y_1^{(ij)}, ..., y_T^{(ij)}\}$ via

$$y_t^{(ij)} = \hat{\rho}^{(i)} y_{t-1}^{(ij)} + \epsilon_t^{(j)}, \quad t = \{1, ..., T\}. \tag{3.105}$$

Choose the initial condition, $y_0^{(ij)}$, randomly from the stationary distribution as proxied by the vector of "true" data $\{y_1^{(i)}, ..., y_T^{(i)}\}$. Finally, compute the sup Wald test statistic $\mathcal{SW}_T^{(ij)}$.

Step [2B] Repeat the **Step [2A]** B times, yielding a $(B \times 1)$ vector of $\mathcal{SW}_T^{(ij)}$ values. Therefore, the particular vector constitutes the bootstrap distribution for Monte Carlo replication $i \in \{1, ..., M\}$. Then, to obtain the 5% critical value of the bootstrap distribution, for example, is estimated as the 950−th element in the vector, after sorting from smallest to largest.

Step [2C] Compare the $\mathcal{SW}_T^{(i)}$ value from **Step [1]** to the $\alpha\%$ bootstrap critical value from **Step [2B]**, and determine whether the critical value is exceeded.

Step [3] Repeat Steps **[1]**-**[2]** M times.

Step [4] Compute the percentage of times a rejection occurs in **Step [2C]**. Therefore, if nominal and empirical test size are equal, then rejection should occur at $\alpha\%$ of the time (up to Monte Carlo error).

Remark 3.11. Overall, for a Bootstrap Monte Carlo study one has to operate under the assumption that in each Monte Carlo step the simulated paired data $\{y_t, \boldsymbol{x}_{t-1}\}_{t=1}^n$ is the true data generating process from which we construct the bootstrapped empirical distribution. The particular, bootstrapped distribution is employed to calculate the critical value for determining the acceptance or rejection decision of the testing hypothesis for one iteration. Then, this procedure is repeated based on the fixed Monte Carlo replications. Therefore, the computational time can significantly increase when we increase the number of bootstrap replications within each Monte Carlo step. The advantages and the limitations of using the bootstrap when evaluating the performance of test statistics is discussed in Brownstone and Valletta (2001).

Remark 3.12. Another important point is that when constructing the Bootstrap Monte Carlo procedure, the data generating process is obtained under the null hypothesis of no structural break in the nonstationary quantile predictive regression model. Therefore, the two regimes induced due to the different model parameters should be constructed.

Specifically, when evaluating the performance of the proposed test statistics in detecting parameter instability the Residual Wild Bootstrap (RWB) procedure outperforms the Fixed Regressor Bootstrap (FRB) especially in the case when regressors are strongly persistent. The proposed Bootstrap Monte Carlo procedure aims to mimic the original model by retaining the nonstationary property of the modelling environment captured by the original time series in the sample pseudo-time series. In particular, these pseudo-time series maintain the same characteristics since are generated based on the same sample size. Therefore, the particular approach allow us to approximate the distribution of the proposed test statistics under the null hypothesis. Furthermore, although we do not consider the case of increasing the number of regressors in the model, it is clear that in that case the quality of the bootstrap approximation in relation the corresponding asymptotic critical values can deteriorate, as documented by Xu and Guo (2020). Therefore, we can conclude that indeed the RWB based Wald test, controls empirical size much better than the other bootstrap based tests with empirical rejection frequencies close to the nominal level.

Additionally, when testing for structural breaks under the assumption of high persistent regressors, either using conventional critical values or critical values obtained from suitable bootstrap based implementations, these tests can be found to deliver asymptotically nonpivotal inference regardless of the chosen estimator, unless we employ the IVZ estimator in which case the limiting distribution is found to be nuisance-parameter free both in finite-sample and large samples. Obviously this conclusion, is not equivalent to the asymptotically pivotal inference that holds in the case of standard predictability tests, which holds regardless of the degree of persistence or endogeneity of the regressor as found by Kostakis et al. (2015), Phillips and Lee (2016) and Georgiev et al. (2021) among others in the time series econometrics literature.

Thus, we can compare the performance of the proposed test statistics under high persistence when the OLS versus the IVX estimators are employed using bootstrap-based inference. Secondly, one can propose asymptotically valid RWB implementations of the proposed test statistics and to establish the conditions required for their asymptotic validity. In particular, simulation evidence can be obtained to demonstrate that structural break tests based on RWB resampling schemes perform well in finite samples, while correcting the finite sample size distortions seen with the corresponding asymptotic structural break tests. Furthermore, the proposed implementations which cover both single as well as multiple regressors in the quantile predictive regression model, hold provided that these satisfy the condition imposed by Kostakis et al. (2015) that the set of regressors belong to the same persistence class, that is, all of the included regressors in the model are generated as either strongly or weakly persistent (i.e., LUR or mildly integrated).

3.5.3 Simulation Results

The empirical sizes of the simulation experiments are calculated as the a proportion of the rejection number of the null hypothesis out of 1000 repetitions. To begin with, we evaluate the performance of the tests in the nonstationary setting. To achieve this task, we generate bivariate normal random variables and use a pre-specified covariance matrix. In particular, we evaluate the performance of the tests at different quantiles by employing the test statistics that correspond to the fixed quantile level. More precisely, this allow us to check for structural breaks at the median, or at the upper and lower quantiles for example, thus observing the presence of parameter instability at different levels of the predictant with respect to persistent predictors. Our simulation experiments verify the empirical and theoretical results that we illustrated for the case of the linear predictive regression model, that is, a trend of over-rejecting the null hypothesis when the OLS estimator is employed when constructing structural break tests[26].

Overall, the finite-sample results reflect the main conjectures presented in the asymptotic theory of the paper and appear to be reasonable for practical use in testing for structural breaks in the coefficients of nonstationary quantile predictive regression models, especially with persistent regressors and endogeneity[27]. In practise, in those cases in which the exact $\alpha-$level critical values for $0 < \alpha < 1$ depend on the unknown parameters of persistence, we employ bootstrap based resampling methods for inference purposes. Therefore, we can observe that these near unit root processes driving the regressors of the model, can affect the ability of the proposed structural break tests for detecting parameter instability in quantile predictive regression models. Specifically, in the simulation study of Wang and Phillips (2012), the authors mention that serial dependence can affect power. Furthermore, the lower long-run signal strength in the regressor tends to reduce discriminatory power.

[26] Notice that the concept of spurious break is discussed by Hansen (2000b) who emphasize that instability in the exogenous variables can cause over-rejection in the standard OLS-based tests.

[27] Notice that exogeneity plays an important role in dealing with non-stationary variables. More specifically, in Chapter of Banerjee et al. (1993) it is mentioned that dynamic regression equations in which the conditioning is on weakly or strongly exogenous variables (for the parameter of interest) provide asymptotically unbiased estimates.

3.6 Empirical Application

Our empirical application is concerned with the monitoring of the US housing price index returns (HPI). Using macroeconomic variables with predictive regression models has been demonstrated in various studies. For instance, the empirical study of Paye (2012) verifies the episodic predictability conclusions documented in the literature such as in Gonzalo and Pitarakis (2012, 2017) (see also Demetrescu et al. (2020)). The author finds statistical evidence of predictability in relation to countercyclical macroeconomic events when forecasting volatility using predictive regressions with macroeconomic covariates. Furthermore, Atanasov et al. (2020) investigate the impact of consumption fluctuations on predictability of expected returns using the IVX filter. Overall, our empirical study focuses in the implementation of the proposed test statistics[28]; thus our research goal here is to test for structural breaks in the relation between the predictant and the predictors at various quantile levels of the underline data specific distribution.

3.6.1 Data Description

For the empirical study, we utilize the dataset of Yang et al. (2020) that includes the US housing price index returns along with ten common macroeconomic variables. Specifically, the HPI covers more transactions and longer time interval, and thus can well represent the trend of the national-wide housing price such as the housing bubble collapsed during the 2007 subprime mortgage crisis. Furthermore, based on the HPI the authors obtain the quarterly growth rate of the housing price and use this rate as the dependent variable. More precisely, the ten macroeconomic variables are collected from FRED, and all data are quarterly between 1975:Q1 and 2018:Q2.

- CPI: Consumer price index with all items less shelter for all urban consumers (Index 1982 to 1984 = 100).
- DEF: The implicit price deflator of the gross domestic product (Index 2012 = 100).
- GDP: %−Change of the gross domestic product from the preceding period.
- INC: %−Change of the real disposable personal income from the quarter one year ago.
- IND: The industrial production index (Index 2012 = 100). An economic indicator that measures real output for all U.S. located facilities manufacturing, mining, and electric, and gas utilities (excluding those in U.S. territories).
- INT: The effective federal funds rate. The interest rate at which depository institutions trade federal funds (balances held at FRBs) with each other overnight.
- INV: The shares of the residential fixed investment in the gross domestic product. Gross private domestic investment is a critical component of gross domestic product as it provides an indicator of the future productive capacity of the economy. Residential investment represents expenditures on residential structures and residential equipment that is owned by landlords and rented to tenants.
- MOG: 30-year mortgage rate. It represents contract interest rates on commitments for fixed-rate first mortgages.
- RES: The total reserve balances maintained with the Federal Reserve banks.
- UNE: The civilian unemployment rate. It represents the number of unemployed as a percentage of the labor force.

[28] Notice, that a different stream of literature proposes testing procedures for detecting market exuberance and bubble effects. Our empirical application is concerned with the detection of structural breaks in the data based on the nonstationary quantile predictive regression.

3.6.2 Data Analysis

We begin our analysis by applying standard unit root tests to the predictors[29] employed for the quantile predictive regression model. Furthermore, we test each individual predictor separately for the presence of parameter instability using a toolkit of various structural break tests commonly employed in the literature. In particular, testing for breaks in housing price indices has been previously studied by Canarella et al. (2012). However, testing for quantile predictability as well as testing for breaks in nonstationary quantile regressions is a novel aspect not previously examined in the literature. Moreover, we are testing for breaks using a linear AR(1) model versus detection of breaks based on a quantile autoregression model for comparability purposes.

Secondly, we implement the IVX-Wald statistic under the null hypothesis that all slope coefficients simultaneously equal to zero[30]. More precisely, the particular hypothesis correspond to the null hypothesis of no quantile predictability. Thus, formulating the model in this manner allow us to investigate whether there is a stable relation between regressand and regressors at a specific quantile level[31] $\tau_0 \in (0, 1)$. Separately, we test the null hypothesis of no parameter instability by implementing the proposed tests for detecting structural breaks in the model coefficients of nonstationary quantile predictive regression models. As a third robustness check, we implement the joint IVX-Wald statistic under the null hypothesis that at least two of the slope coefficients have no structural break throughout the sample.

3.6.3 Main Findings

Our proposed structural break testing methodologies are able to identify structural breaks in the price index returns for the house market. In particular, there strong evidence of the presence of parameter instability which imply a potential price mismatch during periods of increased volatility and market uncertainty such as financial crises. During these period unidentified structural breaks can result to biased parameter estimates as well as inaccurate forecasts if not rigorously addressing these issues in model building and related statistical inference applications. In addition, by identifying the periods for which structural breaks indeed exist, we can utilized different modelling methodologies that can reduce model uncertainty. Furthermore, as robustness checks we can also apply our testing procedure to identify possible structural breaks in the quantiles of time series using different subsamples. In particular, we fit the quantile predictive regression model we introduce above.

[29] Notice that we have $t = 1, ..., 174$ time series observations which correspond to quarterly economic indicators and macroeconomic variables.

[30] We consider rejections of the null hypothesis at significance level 5% to match the rejection probabilities employed in the simulation study of the paper.

[31] In particular, the empirical study presented by Lee (2016) demonstrates statistical evidence of predictive ability using the nonstationary qunatile predictive regression model, at some specific quantiles of stock returns such as at lower or upper quantiles while on the other hand evidence of predictability disappear at the median of the conditional distribution of stock returns.

3.7 Conclusion

In Chapter 3, we develop a framework for structural break detection for nonstationary quantile predictive regression models, under the null hypothesis of no structural break[32]. A major challenge when deriving the asymptotic theory of these structural break tests is to obtain nuisance-parameter free limit distributions which is not a trivial task due to the stochastic approximation terms that depend on nuisance parameters, such as higher order covariance terms as functions of the coefficients of persistence that capture the time series properties of regressors. More precisely, we establish the asymptotic distributions for both Wald type (i.e., as in Andrews (1993)) and fluctuation type tests (i.e., as in Qu (2008)) with respect to two different estimation methods, that is, the OLS and IVX estimators[33]. Our test statistics show to have good finite-sample properties as shown by the Monte Carlo experiments in which we obtain the empirical size and power.

Firstly, we verify that indeed the self-normalization of Wald type statistics when testing the null hypothesis of no predictability in quantile predictive regressions results to a nuisance-free distribution, that is, ensuring their pivotal property (as also proved by Lee (2016) for abstract degree of persistence). Secondly, we demonstrate that the limit distribution of the proposed test statistics for structural break detection is not depending on the particular choice of the estimator of the quantile predictive regression model under mildly integrated; however under high persistence the choice of the estimator alters the limit theory due since different weakly convergence arguments apply. Furthermore, keeping the quantile level fixed versus testing for breaks across multiple quantile levels requires to consider extending the limit result into the two-parameter Gaussian process, for the latter case. Moreover, bootstrap-based methodologies can be applied when the limit distribution is nonstandard, allowing to infer regarding the presence of structural breaks under these conditions (e.g., high persistence). Further research aspects worth investigating that are worth mentioning include to extend the current framework in the case of alternative hypotheses with multiple structural breaks as in Qu (2008) as well as within a multivariate setting such that the framework proposed by the study of Qu and Perron (2007).

[32]Notice that we avoid to explicitly use the terminology of a null hypothesis of "stationarity" versus an alternative hypothesis of "non-stationarity" (see Pitarakis (2014) and Kwiatkowski et al. (1992)). Specifically, within our setting these two terms are interpreted in relation to the persistence properties of model predictors rather that with respect to the parameter constancy of coefficients in terms of temporal dependence. Related limit theory and conditions relevant to temporal dependence specifically for quantile and tail empirical processes can be found in Chapter 5 of De Haan and Ferreira (2006).

[33]Notice in the study of Phillips and Park (1988) the authors demonstrate the asymptotic equivalence of OLS and GLS based estimators in regression models with integrated regressors.

Chapter 4

Estimation and Inference in Seemingly Unrelated Systems of Nonstationary Quantile Predictive Regression Models

Abstract

This Chapter proposes a framework for estimation and inference in quantile predictive regression systems with nonstationary regressors and a generated regressor which is a proxy for systemic risk. In particular, the proposed modelling methodology employs a seemingly unrelated regression system, (SUR), with individual equations representing quantile predictive regression models with nonstationary regressors. The nonstationary properties of regressors is captured by the nuisance parameters of persistence which requires the local-to-unity asymptotic theory of Phillips (1987a) when deriving limiting distributions. Furthermore, inference is conducted using Wald-type statistics with linear restrictions as well as suitable subset restrictions for testing the null hypothesis of no systemic risk in our tail dependency driven system. Thus, to robustify estimation and inference against the nuisance parameters of persistence across these individual system equations the endogenous instrumentation methodology proposed by Phillips and Magdalinos (2009) is employed. We demonstrate that the asymptotic behaviour of these test statistics is asymptotically distribution-free as a weakly converge result to a χ^2_q-squared limiting distribution is established. The finite-sample properties of the proposed testing methodology is studied with simulation experiments that demonstrate the relevance of the tests when imposing parameter restrictions on the system and is particularly useful when modelling and testing for systemic risk.

4.1 Introduction

Modelling tail dependency with multivariate nonstationary time series is an aspect which has not seen much attention in the literature. Specifically, while several studies propose methodologies for estimation and inference using quantile regressions with stationary time series when modelling systemic risk, using nonstationary time series[1] for this purpose has received less attention. A key aspect of consideration of our proposed econometric framework is the use of quantile regression for modelling tail events. In particular, Adrian and Brunnermeier (2016), Härdle et al. (2016) and Chen et al. (2019) employ quantile-based regression models when estimating forecasts of tail risk measures[2] such as the Value at Risk (VaR) and Conditional Value at Risk (CoVaR), under the assumption of stationary time series. Moreover, there is currently a gap in the literature regarding suitable testing methodologies for the true presence of the coefficients of systemic risk appeared in these econometric specifications. Our research objective is to propose a framework for identification, estimation and inference in systems of nonstationary quantile predictive regressions as well as to establish the asymptotic theory of Wald-type statistics when testing the null hypothesis of no systemic risk across the system of equations of the cross-sectional units[3].

Regardless the complexity of the proposed interdependent system[4], this complexity does not preclude identification and consistent estimation of the parameters involved in these models (regardless of the presence of persistent data). Although variation of regression coefficients over time has been examined extensively in the single-equation context, there is less focus in the literature in a pooling context. These model specifications have the potential for describing both the cross-sectional and inter-temporal dynamics using a pooled data structure, although in this paper we exclude cases of time-varying coefficients when modelling systemic risk (e.g., see Bianchi et al. (2019)). Therefore, we introduce an identification and estimation approach in Seemingly Unrelated Systems of Nonstationary Quantile Predictive Regression Models, motivated by the aspects of modelling systemic risk in a tail-driven network under the presence of nonstationarity. In particular, the SUR estimator allows to gain estimation efficiency as well as to impose restrictions on the quantile predictive regressions in the system. Our identifying conditions relies on error covariance matrix restrictions by assuming no correlation between equation-specific error covariance matrices. We should emphasize that our techniques can also be used for estimating dynamic panel data structures based on suitable modifications, although we focus on our proposed system representation and estimation problem which aligns with our research motivation.

---

[1] A necessary condition for deriving the asymptotic behaviour of estimators in nonstationary time series models is for weakly convergence arguments of sample moments to stochastic integrals to hold.

[2] The risk measure pair $(\mathsf{VaR}, \mathsf{CoVaR})$ are commonly used in the literature for modelling systemic risk as well as in portfolio optimization problems in which case additional regularity conditions and constraints are required.

[3] Under the null hypothesis we assume that the presence of tail dependency across the cross-section is kept at minimal which has desirable properties in terms of portfolio risk as explained in the studies of Katsouris (2021) and Olmo (2021). In particular these studies derive conditions under which optimal portfolio allocation problems relate to the portfolio risk under network structure.

[4] Several studies in the literature highlight that modelling financial contagion and systemic risk by considering the economy requires the simultaneous estimation of equations in which case the SUR model provides a parsimonious representation. A simple example is the case of investment equations across financial institutions where the activities of a firm in a given year is expected to have a lagged effect within and across firms.

Our contributions to the literature are as below. First, the paper contributes to the modern literature on identification and estimation in seemingly unrelated systems of quantile predictive regression models with nonstationary regressors by considering the modelling of systemic risk, based on the pair of risk measures (VaR, CoVaR), as our focal point. Existing literature on network driven estimation techniques has focused on the use of models for stationary time series models (see, Henderson et al. (2015), Cho et al. (2015), Wang et al. (2018), Cai and Liu (2020), Tan et al. (2021), Xu et al. (2022) and Cai and Liu (2022)), while the literature which considers nonstationary time series models relies on different to ours representations either with respect to the cointegration properties of the multivariate time series (see, Poskitt (2006), **?** and Magdalinos (2021)) or with respect to the properties of the system equations (see, Mark et al. (2005), Chen et al. (2023)), while in our case our focus is the tail network-driven network approach for modelling systemic risk (see, Katsouris (2023b)). Second, the paper contributes to the literature of statistical inference on subvector of system parameters with our proposed Wald test that test for the presence of systemic risk effects in the identified and estimated network structure.

We consider a system of m nonstationary quantile predictive regressions where we assume that all system equations are estimated parametrically based on the same conditional quantile functional form. Moreover, similar to the Gaussian SUR model, the system equations are related through the underline correlation structure of the Gaussian errors such that

$$\boldsymbol{e} \overset{i.i.d}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n) \tag{4.1}$$

where the $\boldsymbol{e}_i$ error vector corresponds to the n time series observations of the $i-$th equation and $\boldsymbol{\Sigma}$ is a positive-definite $(m \times m)$ covariance matrix which also requires estimation. Moreover, we assume that the estimation of the observed time series based on independence conditions of the system errors, such as serial independence and conditional quantile independence although these two conditions does not preclude cross-sectional dependence dynamics based on further assumptions. The error terms are assumed to be independent and identically distributed across equations, while all equations are based on the same number of time series observations. Our modelling approach is also related to several recent developments in the literature of robust estimation methods in predictive regression models as we further explain below (see, Demetrescu and Rodrigues (2020) and Chen et al. (2023)).

A crucial assumption of our framework is that we assume the dimension of the multivariate time series vector $\boldsymbol{y}_t$ is the same as the number of system equations. Consider two elements of the $m-$dimensional vector $\boldsymbol{y}_t$, that is, $y_{(i)t}$ and $y_{(j)t}$ such that $(i,j) \in \{1,...,m\}$ with $i \neq j$. Then, the conditional expected mean function is defined as below

$$\mathbb{E}\left[y_{(j)t} | \boldsymbol{x}_{t-1}, y_{(i)t} \right] := \beta_0 + \boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} + \delta y_{(i)t}, \quad t = 1, ..., n, \tag{4.2}$$

where $y_{(.)t}$ is a scalar vector and $\boldsymbol{x}_{t-1}$ is $p-$dimensional vector of regressors. Then, the coefficients β_0, $\boldsymbol{\beta}_1$ and δ can be estimated consistently via OLS of $y_{(j)t}$ on $\boldsymbol{x}_{t-1}$ and $y_{(i)t}$. Furthermore, the predicted value of such an OLS regression would be the mean of $y_{(j)t}$ conditional on x_{t-1} and $y_{(i)t}$.

Similarly, when estimating the risk measures of VaR and CoVaR, we employ a conditional quantile function which is a common practise in the literature (see, Adrian and Brunnermeier (2016) and White et al. (2015)). Additionally, in the current paper we assume that the regressors follow a near unit root process such as $\boldsymbol{x}_t = \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}_p}{n} \right) \boldsymbol{x}_{t-1} + \boldsymbol{v}_t$, where $\boldsymbol{C}_p$ is a $(p \times p)$ matrix with the nuisance parameters of persistence. This assumption, allows to capture the nonstationary properties of regressors, however it requires a different modelling methodologies such as an instrumental based approach (see, Phillips and Magdalinos (2009)) which is robust to the nuisance parameter of persistence.

Furthermore, to provide a more realistic framework we consider that the conditional distribution $y_{(g)t} | \boldsymbol{x}_{t-1}$ for all $g \in \{1, ..., m\}$ is randomly generated although from the same underline stochastic process, which in practise allows to consider estimating a different quantile predictive regression model based on the aforementioned data structure. Thus, for the development of our asymptotic theory we consider a triangular system for each node by assuming that each innovation sequence corresponding to these system equations are generated by a set of triangular arrays which are independently generated. This property ensures that cross-section dependence and time series dependence are induced simultaneously without further assumptions regarding the correlation structure of the regressand and the regressors that correspond to each member of the cross-section[5]. Based on the aforementioned methodologies, we develop a framework for estimation and inference in such a complex tail dependency driven system under nonstationarity and a more general form of dependence.

Furthermore, to facilitate the econometric identification for these independently generated triangular arrays across the m nodes we employ the framework of Seemingly Unrelated Regressions, (SUR), in which each individual system equation corresponds to the quantile predictive regressions. Similarly, we want to investigate an implementation of the OLS and IVX estimators under the presence of possibly nonstationary regressors in the two econometric specifications that correspond to the risk measures of VaR and CoVaR. Then, the particular generated regressor is used as an additional regressor in the CoVaR specification. Specifically, we extend the current conditional probability definitions for VaR and CoVaR proposed by Adrian and Brunnermeier (2016) to incorporate the persistence properties of regressors when modelling these conditional risk measures. For example, Mark et al. (2005) propose a framework for estimation and inference in a SUR system of cointegrating regressions while separating the effects from each individual equation. This study considers the estimation and inference of a quantile predictive regression model with nonstationary regressors which is suitable when estimating the VaR risk measure while at the same time incorporating information regarding the nonstationary properties of regressors with the local-unit-root specification. Furthermore, we can estimate the CoVaR based on a similar modelling framework, by constructing a different quantile predictive regression model with the same set of regressors and an additional regressor which corresponds to the generated VaR risk measure.

[5] In particular, according to Phillips and Moon (1999), the simultaneous modelling of both cross section dependence and time series dependence remains a challenging problem and is a major area for future research in multi-index asymptotics.

4.1.1 Illustrative Examples

In order to motivate the modelling methodology we propose in the current study we demonstrate some examples of relevant econometric specifications which can provide some insights for the identification and estimation strategy we follow. In particular, we consider that the data generating process is based on the predictive regression models with nonstationary regressors as these were proposed in the studies of Phillips and Magdalinos (2009) and Kostakis et al. (2015). Specifically, the nonstationary specification of the autoregression equation permits to capture the persistence properties of regressors.

Example 4.1. Consider the lag-augmented version of the linear predictive regression

$$y_t = \beta x_{t-1} + \delta y_{t-1} + u_t \quad t = 1, ..., n, \tag{4.3}$$

$$x_t = \rho x_{t-1} + v_t, \qquad \rho = \left(1 - \frac{c}{n}\right), c > 0. \tag{4.4}$$

with $\delta \neq 0$ for estimation and inference. The lag-augmented modification proposed by Demetrescu (2014) is considered as a from of variable addition procedure (e.g., see Dolado and Lütkepohl (1996)). The model (4.106)-(4.107) implies the weak convergence $\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} (u_t, v_t) \Rightarrow (\sigma_u W(s), \sigma_v V(s))'$ with $(W(s), V(s))'$ a vector of correlated standard Wiener processes such that an invariance principle for x_t holds, that is, $\frac{1}{\sqrt{n}} x_{\lfloor ns \rfloor} \Rightarrow B_c(s)$, where $B_c(s)$ is the OU process driven by V such that $B_c(s) = V(s) - c \int_0^s e^{-c(s-r)} V(r)$. In practise, when the shocks u_t and v_t are correlated, then the regressor x_t is considered to be endogenous and given the presence of high persistence in the time series of the regressor various methods have been proposed in the literature to fix the finite-sample bias in the estimation of the standard OLS estimator of the predictive regression model. We do not impose any further assumptions that exclude such form of correlation structure between the error terms u_t and v_t, which allows for the presence of endogeneity. However, we consider inference methods that are robust to both endogeneity and persistence.

Remark 4.1. The model (4.106)-(4.107) presented by Example 4.1 has some parallelism with the proposed modelling framework (for the univariate case). Although, instead of using as a predictor the first lag of y_t, that is, y_{t-1}, for the econometric specification we introduce in Section 2, the additional predictor which does not belong in the set of nonstationary regressors (modelled via the LUR specificaton) is used as a proxy for the systemic risk. Specifically, following the literature econometric specifications for estimating systemic risk measures such as the framework of Adrian and Brunnermeier (2016), the particular generated regressor captures the VaR of each node. The challenging task of our framework is that to derive correctly asymptotic theory we need to take into account the fact that we additionally consider that these risk measures are estimated using nonstationary time series models, which provides additional information regarding the persistence properties of the regressors.

Example 4.2. Consider a pair of nodes (i, j) in the graph and we aim to obtain estimates for the CoVaR risk measures of node i and node j. Moreover, assume we have the observable vector sequence $\{y_{1t}, y_{2t}, x_t\}_{t=1}^n$. Then, the following bivariate system occurs, assuming a quantile objective function for some $\tau \in (0,1)$:

$$y_{1t} = \beta_1(\tau)x_{t-1} + \delta_1(\tau)y_{2t} + u_{1t}(\tau), \quad \text{with} \quad x_t = \rho_1 x_{t-1} + v_{1t} \tag{4.5}$$

$$y_{2t} = \beta_2(\tau)x_{t-1} + \delta_2(\tau)y_{1t} + u_{2t}(\tau), \quad \text{with} \quad x_t = \rho_2 x_{t-1} + v_{2t} \tag{4.6}$$

where $\rho_j = (1 - \frac{c_j}{n})$, for $j = \{1, 2\}$ with $c_1 \neq c_2 > 0$. Then, the null hypothesis is, $\mathbb{H}_0 : \delta_1(\tau) = \delta_2(\tau) = 0$ and the alternative, $\mathbb{H}_A : \delta_1(\tau) \neq \delta_2(\tau) \neq 0$.

Remark 4.2. Example 4.2 provides a simple illustration of the econometric complexity of modelling an observable vector sequence $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}_{t=1}^n$, where $\boldsymbol{y}_t = [y_{1t}, \dots, y_{mt}]$ in a similar manner as the pairwise quantile predictive regression specification given by the two systems in expressions (4.5)-(4.6), due to the presence of dependent covariate in each of those two separate predictive regressions. For instance, under the assumption that $\delta_1(\tau) = \delta_2(\tau) = 0$, then one can model the particular bivariate system as a multivariate predictive regression system in the spirit of Kostakis et al. (2015) and Lee (2016). The particular approach would hold for both the cases of either a single predictor x_t (scalar) or multiple predictors $\boldsymbol{x}_t$ (vector) while having a multivariate regressand $\boldsymbol{y}_t$.

Therefore, both Example 4.1 and 4.2 illustrate that for a large number of nodes, the estimation procedure can be done either in a pairwise manner equation-by-equation for all $m(m-1)$ pairs of nodes in the graph or using a parsimonious modelling approach which permit to group some of those pairs which have common indices. Thus, we propose the implementation of a SUR system for specific set pair of nodes in the graph. As a by-product, we also propose a novel estimator, so called, SUR-IVX estimator for a quantile predictive regression system of equations which permits estimation and inference in econometric models with graph structure, for the purpose of modelling systemic risk. More precisely, we consider the tail risk measures of VaR and CoVaR, focusing on the CoVaR as a systemic risk measure while the econometric identification is achieved via the pairwise quantile predictive regression models proposed by Adrian and Brunnermeier (2016). Thus, our research contributions are organized around two main pillars. First we contribute to the financial econometrics literature by proposing a unified approach for estimating systemic risk incorporating the persistence properties of the lagged regressors (see, the framework proposed by Lee (2016)). Second, our testing methodology for the presence of systemic risk contributes to the econometrics literature since we propose a Wald type statistic for SUR models with quantile predictive regression model specifications robust to the unknown persistence properties of regressors employed in these models.

4.1.2 Related Literature

Firstly, the econometric specifications we employ in this paper are motivated by current methodologies for modelling systemic risk as proposed by Adrian and Brunnermeier (2016) and extended by Härdle et al. (2016) to the case of nonlinar time series regressions to capture tail effects in graphs. Although these approaches consider an underline graph dependence, Shin et al. (2016) consider a SUR modelling approach for IV estimation of contagion models. An application to systemic risk modelling, is proposed by Bianchi et al. (2019) who implement a SUR system of equations using high dimensional graph modelling. Furthermore, in a slightly different context, Galvao et al. (2018b) propose an IV based estimation procedure for a system of investment equations. All aforementioned studies consider an identification strategy which is suitable for modelling financial spillover effects in markets, without imposing further assumptions regarding the exact form of graph dependence.

Secondly, in terms of the testing methodology for the presence of systemic risk, the idea of formulating the null hypothesis for testing linear restrictions on parameters across blocks has some similarities with the literature of slope homogeneity testing in panels (see, Galvao et al. (2018a) and Galvao et al. (2019)). A different approach for testing that these systemic coefficients are zero is proposed by Nkurunziza (2008, 2010). Additional features we consider with the testing procedure of our tail dependency driven system, are presented in the study of Ravikumar et al. (2000) who consider the use of adding-up restrictions when formulating the standard Wald test in a SUR system of equations. However, there is currently a gap in the literature regarding suitable testing methodologies for the true presence of these systemic risk effects in econometric models. In particular, the framework of White et al. (2015) allows the quantiles of the underline distributions to depend on lagged quantiles, past innovations and other covariates. In particular, this approach employs a Vector Autoregression representation for estimating the VaR for a set of firms as well as for testing the null hypothesis of no tail codependence based on the coefficients of the lagged quantiles. Our framework proposes the construction of a suitable testing procedure, which implies that under the null hypothesis the presence of tail dependency due to these systemic risk effects is zero.

Thirdly, the additional feature which we consider in this paper and has not given much attention in the literature before when modelling systemic risk measures (e.g., VaR and CoVaR), is the nonstationary properties of regressors employed within these models. Therefore, a novelty of our framework is that we incorporate information regarding the possible presence of nonstationarity with the use of the predictive regression framework (see, Phillips and Magdalinos (2009), Kostakis et al. (2015) as well as Magdalinos (2021)). Specifically, this approach requires to model the regressors as local to unit root processes that capture the unknown persistence in the time series of regressors and is based on the local-to-unity limit theory proposed by the seminal studies of Phillips (1987a,b) (see, also Phillips and Solo (1992)). The estimation of these risk measures requires to consider a conditional quantile specification form for modelling the quantiles[6] with nonstationary regressors as in the framework proposed by Lee (2016) and Fan and Lee (2019).

[6]Notice that related techniques to the estimation and inference in quantile regression models are examined by Koenker and Bassett (1978) and Koenker and Bassett (1982).

Our framework requires the estimation of the CoVaR risk measures for all cross-sectional units. Thus, for estimating the CoVaR risk measure, we follow the econometric specifications proposed by Adrian and Brunnermeier (2016) (AB, thereafter) which require a conditional quantile predictive regression based on an information set with both lagged endogenous regressors and the generated regressor for the VaR. Current methodologies for obtaining forecasts of the CoVaR risk measures use the conditional quantile regression estimation approach while these operate under the assumption of stationary regressors. In our study, we relax the particular assumption and allow for the inclusion of nonstationary regressors, which necessitates to combine both the AB specifications with the framework proposed by Lee (2016) who consider this feature. However, the particular econometric specification is only suitable for modelling the risk measure of VaR and therefore an extension for the modelling of the CoVaR based on the AB modelling approach has not been previously examined in the literature. Lastly, some further aspects we consider which require some refinements include the bias correction due to the presence of generated regressor (see, Chen et al. (2021)), although under nonstationarity a different approach is required which we discuss. Overall, due to the fact that each of the m nodes has an individual specific equation, a VAR representation for obtaining the CoVaR measures is not a suitable modelling approach, which motivates the use of a SUR type of system[7] for estimation and inference purposes.

Our procedure implies a novel lag-augmentation procedure due to the presence of systemic risk proxy (generated regressor) which has desirable properties since the score contributions of the particular methodology are serially uncorrelated. In other words, both due to the novel dependence structure of the node-specific predictive regression models as well as due to the fact that the generated regressor that corresponds to each of the predictive regression models is not correlated with the nonstationary regressors of the predictive regression model we are estimating, we contribute to the literature of uniform. Notice that the stationarity assumption for the system is ensured since we assume that all individual specific equations have zero initial conditions. Although the literature of seemingly unrelated regression system estimation requires the availability of $\sqrt{n}-$consistent first round estimator a natural choice is the quantile regression estimator. However, within our setting due to the nonstationarity effects we consider two estimators which exhibit different rate of convergence. An obvious problem with such a two-step procedure is that the first round estimation error, although asymptotically absent, can be such that correction is not worthwhile in small samples. The particular source of error can be amplified when the number of regressors is large since nonparametric covariance matrix estimators of high-dimensional functions are notoriously inaccurate. In terms of related assumptions that hold is that we consider that the joint conditional distribution of the regressors across equations is modelled at the same quantile level as when considering the conditional quantile function for each of these individual specific equations. The particular condition provides efficiency gains and is equivalent to an orthogonality condition between regressors and errors across equations in the conditional mean regression case. However, we do not consider the case of modelling different regression quantiles across equations.

[7]Discussion on aspects relevant to estimation and inference for SUR systems is presented in the studies of Zellner (1962), Hillier and Satchell (1986), Mark et al. (2005).

Since our identification strategy is based on a system of seemingly unrelated regressions of nonstationary quantile predictive regression model, then the SUR is expected to provide efficiency gains. Our modelling approach captures the possible presence of endogeneity between the error terms of regressors and predictive regressions when estimating risk measures, thereby under the assumption of no cross-sectional dependence through unobserved factor structures the joint estimation of these node specific equations provides efficiency gains in comparison to an equation-by-equation case. We also exclude misspecified regimes by assuming that the full information set consists only of the nonstationary regressors for each quantile predictive regression model as well as the generated regressor which is a proxy to systemic risk. Therefore, this homogeneity of conditions across equations ensures the consistent estimation of the system estimator. Investigating the effect of individual equation specific misspecification requires different assumptions and methodologies that we do not examine in this chapter. In addition the systemic risk proxy (generated variable) which is endogenously estimated for each of the system equations ensures that there is no cross-correlation of the particular variable across equations which would violate the modelling assumptions of the system estimator and thus our proposed estimation methodology has a great asymptotic efficiency in comparison to the OLS estimator for instance. In contrast, the classical SUR model assumes deterministic regressors and homoscedastic errors, which corresponds to independence of errors and regressors when regressors are random. Therefore, our modelling framework operates differently since we consider that each of the system equations corresponds to the quantile predictive regression model which has specific features such as endogeneity of regressors. On the other hand, this assumption does not violate the independence assumption of errors across system equations due to the proposed dependence structure we propose. Although the specification function of the multivariate system allows for quantile dependency we only consider fixed coefficients excluding time-varying effects from our framework.

Recently, Chen et al. (2023) investigate the use of a SUR system estimator for VAR models with explosive roots. In particular, due to the inconsistency that appears when testing for a common explosive roots (see, Phillips and Magdalinos (2013)) the approach proposed by Chen et al. (2023) which uses a SUR estimator, treating this way the individual specific equations as seemingly unrelated equations, is found to be consistent regardless of persistence properties and most importantly in the presence of both distinct explosive roots and common explosive roots. Furthermore, the authors demonstrate via simulations that the SUR estimate performs better than OLS and IV estimate in both of these explosive cases[8], which illustrate exactly the importance of our proposed framework as well. Specifically, it is found that there is a significant difference in the FEVD when using different estimators for explosive VAR models, which strongly suggests the use of the SUR estimator when the focus is on bubble periods due to its good statistical properties[9].

[8]Notice that Magdalinos and Phillips (2009b) developed limit theory for multivariate co-explosive processes. However, the common explosive root case yields a singular matrix for the sample variance matrix, which requires coordinates rotation in developing asymptotics. Moreover, Phillips and Lee (2016) apply the self-(endogenously)generated IVX instruments proposed by Kostakis et al. (2015), in the case of a co-explosive system.

[9]In particular, the SUR representation of a VAR system, is shown by Chen et al. (2023) that specifically for a moderately explosive system with a common explosive root, the SUR estimator is consistent and has better finite sample performance that the OLS and IV estimator proposed by Magdalinos and Phillips (2009b).

4.1.3 Two-Stage Estimation Algorithm

Our proposed two-stage estimation procedure is summarized as below:

Step 1. Let $\boldsymbol{\theta}_j$ be the parameter vector that corresponds to both the nonstationar regressors and the generated covariate. Estimate $\boldsymbol{\theta}_j$ and compute the fitted values such that $\widehat{x}_{tj} = g_j \left(\boldsymbol{w}_{tj}, \widehat{\boldsymbol{\theta}}_j \right)$ for $j \in \{1, ..., N\}$ and then obtain the generated regressors, where N denotes the number of nodes in the network or the number of firms in a cross-section. We denote with $g(\cdot)$ the optimization function which corresponds to the conditional quantile functional form.

Step 2. Compute the unknown parameter vector $\boldsymbol{\beta}$ from the following QR model

$$\widehat{\boldsymbol{\beta}}(\tau) = \underset{\beta \in \mathbb{R}^p}{\mathsf{arg\ min}} \sum_{i=1}^{n} \rho_\tau \Big(y_t - \boldsymbol{x}'_{t-1} \boldsymbol{\beta} \Big) \tag{4.7}$$

where the check function is defined as $\rho_\tau(\mathsf{u}) := \mathsf{u}\big[\tau - \mathbf{1}\{\mathsf{u} \geq 0\}\big]$ and $\boldsymbol{x}_t$ are LUR processes. One uses the estimates of $\boldsymbol{\theta}_j$, denoted by $\widehat{\boldsymbol{\theta}}_j$ for $j \in \{1, ..., N\}$, during the first-stage, to obtain the generated regressor $\widehat{x}_t^{\mathsf{VaR}}$.

Step 3. Compute the unknown parameter vector $\boldsymbol{\beta}^\star$ from the following QR model

$$\widehat{\boldsymbol{\beta}}^\star(\tau) = \underset{\beta^\star \in \mathbb{R}^{p+1}}{\mathsf{arg\ min}} \sum_{i=1}^{n} \rho_\tau \Big(y_t - \tilde{\boldsymbol{x}}'_{t-1} \boldsymbol{\beta}^\star \Big), \quad \tilde{\boldsymbol{x}}_{t-1} = \big[\boldsymbol{x}_{t-1} \quad \widehat{x}_t^{\mathsf{VaR}} \big] \tag{4.8}$$

where $\boldsymbol{\beta}^\star$ corresponds to the parameter vector of the nonstationary regressor of the second-stage as well as the unknown parameter that correspond to the fitted-values of the VaR[10] that have been estimated during the first-stage estimation procedure.

Recall that we can consistently estimate the conditional quantile based on the quantile-dependent model parameter $\widehat{\boldsymbol{\beta}}^\star(\tau)$ such that $\widehat{Q}_\tau(\tau | \boldsymbol{x}_{t-1}, \widehat{x}_t^{\mathsf{VaR}}) = \tilde{\boldsymbol{x}}'_{t-1} \widehat{\boldsymbol{\beta}}^\star(\tau)$.

Organization of the paper Section 2 introduces the econometric framework which consists of the identification strategy, the model, the main assumptions and the estimation methodology. Section 3 presents our theoretical results under general dependence conditions and our proposed statistical inference methodology. Section 4 discusses numerical studies using real and simulated data. Section 5 concludes. All technical proofs can be found in the Appendix of the paper.

[10] In the framework of Härdle et al. (2016), this coefficient corresponds to the unknown model parameter $\hat{\beta}_{j|i}$ in their expression (6). See also expressions (7) and (8) of Härdle et al. (2016) that correspond to the econometric specifications used to obtain the estimates of the risk measure pair under the assumption of stationary regressors.

Notation Throughout the paper, for any vector $\boldsymbol{a}$, we denote with $\|\boldsymbol{a}\|_1 = \sum_j |a_j|$ and with $\|\boldsymbol{a}\|_2 = \sqrt{\sum_j |a_j|^2}$. For any real arbitrary matrix $\boldsymbol{A}$, the norm is denoted by $\|\boldsymbol{A}\|$ and corresponds to the Frobenius norm defined by $\|\boldsymbol{A}\| = \sqrt{\mathsf{trace}(\boldsymbol{A}'\boldsymbol{A})}$. We use $\boldsymbol{a}_{(j)}$ to denote the $j-$th component of a vector $\boldsymbol{a}$ and $\boldsymbol{A}_{[i,j]}$ is defined similarly for any arbitrary matrix $\boldsymbol{A}$. Let $\lambda_i(\boldsymbol{M})$ denote the $i-$th largest eigenvalue of an $(n \times n)$ symmetric matrix $\boldsymbol{M}$ with its eigenvalues such that $\lambda_1(\boldsymbol{M}) \geq ... \geq \lambda_n(\boldsymbol{M})$. The spectral norm of $\boldsymbol{A}$ is denoted by $\|\boldsymbol{A}\|_2$, such that, $\|\boldsymbol{A}\|_2 = \sqrt{\lambda_1(\boldsymbol{A}'\boldsymbol{A})}$, is maximum column sum norm is denoted by $\|\boldsymbol{A}\|_1$, such that, $\|\boldsymbol{A}\|_1 = \mathsf{max}_{1 \leq j \leq n} \sum_{i=1}^{m} |A_{ij}|$ and its maximum row sum norm is denoted by $\|\boldsymbol{A}\|_\infty$, such that, $\|\boldsymbol{A}\|_\infty = \mathsf{max}_{1 \leq i \leq n} \sum_{i=1}^{m} |A_{ij}|$. Moreover, the operator $\overset{P}{\to}$ denotes convergence in probability, and $\overset{D}{\to}$ denotes convergence in distribution and $\Rightarrow$ denotes weak convergence in the $D[0,1]$ topological space. All limits are for $n \to \infty$ in all theories, and $O_p(1)$ is stochastically asymptotically bounded while $o_p(1)$ is asymptotically negligible.

4.2 Econometric Framework: Identification and Estimation

4.2.1 Identification Strategy

The modelling of systemic risk in the proposed complex tail dependency driven system is based on the econometric specifications of Adrian and Brunnermeier (2016), although we extend the particular framework in different directions as we explain below. To begin with, we assume that there exists an $m-$dimensional vector $\boldsymbol{y}_t$ for $t = 1, ..., n$ and consider two elements of this vector $(i,j) \in \{1, ..., m\}$ such that $i \neq j$, along with a $d-$dimensional vector of regressors $\boldsymbol{x}_{t-1}$ where $d < m$. Moreover, denote with $y_{(j)t}$ the $j-$th element of the $\boldsymbol{y}_t$ vector, such that $y_{(j)t}$ is a $(n \times 1)$ scalar vector. Then, the conditional quantile function $F^{-1}_{y_{(j)t}}\left(\tau | \boldsymbol{x}_{t-1}, y_{(i)t}\right)$ is the $\mathsf{VaR}_{(j)t}(\tau)$ conditional on $\boldsymbol{x}_{t-1}$ and $y_{(i)t}$, where $F(.)$ is the conditional distribution function and $F^{-1}(\cdot)$ is the conditional quantile function of the underline distribution.

Similarly, by conditioning on $y_{(i)t} = \mathsf{VaR}_{(i)t}(\tau)$, we obtain the $\mathsf{CoVaR}_{(j,i)t}(\tau)$ as defined below

$$\mathsf{CoVaR}_{(j,i)t}(\tau) := \inf_{\mathsf{VaR}_{(j)t}(\tau)} \left\{ \mathbb{P}\left(y_{(j)t} \middle| \left\{ \boldsymbol{x}_{t-1}, y_{(i)t} = \mathsf{VaR}_{(i)t}(\tau) \right\} \right) \geq \tau \right\} \tag{4.9}$$

Equivalently, it holds that

$$\mathsf{CoVaR}_{(j,i)t}(\tau) \equiv F^{-1}_{y_{(j)t}}\left(\tau | \boldsymbol{x}_{t-1}, y_{(i)t} = \mathsf{VaR}_{(i)t}(\tau)\right) \tag{4.10}$$

where $\tau \in (0,1)$ is a fixed quantile level.

The aforementioned definitions are instrumental in discussing the proposed econometric identification and estimation strategy in the present paper that considers further aspects to current modelling methodologies.

4.2.2 Predictive Regression Model

Consider the following seemingly unrelated regressions system with predictive regression models

$$y_{g,t} = \boldsymbol{\beta}_g \boldsymbol{x}'_{g,t-1} + u_{g,t}, \tag{4.11}$$

$$\boldsymbol{x}_{g,t} = \boldsymbol{R}_{g,n} \boldsymbol{x}_{g,t-1} + \boldsymbol{v}_{g,t}, \tag{4.12}$$

where $g \in \{1, ..., m\}$ and $t = 1, ..., n$ such that $y_{g,t} \in \mathbb{R}$ represents the regressand of the $g-$th predictive regression model and $\boldsymbol{x}_{g,t} = \left(x_{1t}^{(g)}, ..., x_{dt}^{(g)} \right)'$ is the $d-$dimensional vector of regressors and we assume that the number of regressors for the $g-$th equation is the same for all $g \in \{1, ..., m\}$. In addition, to model the persistence properties of regressors the autocorrelation coefficient matrix is defined as below

$$\boldsymbol{R}_{g,n} := \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d^{(g)}}{n^{\gamma}} \right) \quad \text{with} \quad \gamma = 1, \tag{4.13}$$

where $\boldsymbol{C}_d^{(g)} = \mathsf{diag}\{c_1^{(g)}, ..., c_{d_g}^{(g)}\}$ is a $(d \times d)$ diagonal matrix such that $c_j > 0 \ \forall \ j \in \{1, ..., d_g\}$ and n is the sample size. Then, the nonstationary properties of regressors are determined by the unknown coefficients of persistence c_j's which are positive constants. Thus, with the particular autocorrelation matrix specification we focus on near unit root processes (see, **?** and Buchmann et al. (2007)). For illustration purposes we show that each of the $g-$th system equations is represented by a predictive regression model as seen by the data generating mechanism below

$$y_{1,t} = \boldsymbol{\beta}'_1 \boldsymbol{x}_{1,t-1} + u_{1,t}, \quad \text{for} \quad t = 1, ..., n \tag{4.14}$$

$$\boldsymbol{x}_{1,t} = \boldsymbol{R}_{1,n} \boldsymbol{x}_{1,t-1} + \boldsymbol{v}_{1,t}, \quad \text{with} \quad \boldsymbol{R}_{1,n} = \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d^{(1)}}{n} \right), \tag{4.15}$$

where $\boldsymbol{w}_{1,t} = (u_{1,t}, \boldsymbol{v}'_{1,t})'$ such that $\boldsymbol{w}_{1,t} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{ww}^{(1)} \right)$. Similarly for $g = 2$ we have

$$y_{2,t} = \boldsymbol{\beta}'_2 \boldsymbol{x}_{2,t-1} + u_{2,t}, \quad \text{for} \quad t = 1, ..., n \tag{4.16}$$

$$\boldsymbol{x}_{2,t} = \boldsymbol{R}_{2,n} \boldsymbol{x}_{2,t-1} + \boldsymbol{v}_{2,t}, \quad \text{with} \quad \boldsymbol{R}_{2,n} = \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d^{(2)}}{n} \right), \tag{4.17}$$

where $\boldsymbol{w}_{2,t} = (u_{2,t}, \boldsymbol{v}'_{2,t})'$ such that $\boldsymbol{w}_{2,t} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{ww}^{(2)} \right)$ and so on. Our proposed modelling approach ensures that the independence assumption between the innovation sequences $\boldsymbol{w}_{1t}$ and $\boldsymbol{w}_{2t}$ is preserved. This assumption provides a statistical way of estimating the equations across the m units without imposing a cross-sectional dependence structure on the error terms using factor loadings while the nonstationarity of regressors across these cross-sectional units is independently determined.

We assume that all individual equations have equivalent number of regressors. Then, the first step in the estimation procedure is to obtain the IVX estimators for each of these individual quantile predictive regression models as well as the fitted values of $\hat{y}^{o}_{g,t}$ for all $g' \in \{1, ..., G\}$ based on a fixed quantile $\tau \in (0, 1)$. These quantities represent the estimated VaR under nonstationarity; which is itself a novel aspect in the current literature. Similarly, the estimated CoVaR is obtained based on the two-stage estimation procedure explained in Katsouris (2023a). Our econometric identification strategy aims to construct a tail dependency driven system while modelling the nonstationary properties of regressors. Although, is considered as a methodology for tail dependency modelling in multivariate nonstationary time series, our system representation is different than the conventional multivariate predictive regression models (see, Remark 4.8). Specifically the estimation procedure considers estimating a quantile predictive regression model for each of the m cross-sectional units using their set of nonstationary regressors and a generated regressor which is a proxy of systemic risk based on the nonstationary properties of regressors. Thus, as shown in Figure 4.1 the complex tail dependency system consists of $(m-1)$ quantile predictive regression models.

$$\begin{aligned}
y_{1,t} &= \boldsymbol{\theta}'_{(1)}(\tau)\widetilde{\boldsymbol{X}}_{1,t} + u_{1,t}(\tau) \equiv \boldsymbol{\beta}'_{(1)}(\tau)\boldsymbol{x}_{1,t-1} + \delta_{(1)}(\tau)\hat{y}^{o}_{2,t} + u_{1,t}(\tau) \\
y_{2,t} &= \boldsymbol{\theta}'_{(2)}(\tau)\widetilde{\boldsymbol{X}}_{2,t} + u_{2,t}(\tau) \equiv \boldsymbol{\beta}'_{(2)}(\tau)\boldsymbol{x}_{2,t-1} + \delta_{(2)}(\tau)\hat{y}^{o}_{3,t} + u_{2,t}(\tau) \\
y_{3,t} &= \boldsymbol{\theta}'_{(3)}(\tau)\widetilde{\boldsymbol{X}}_{3,t} + u_{3,t}(\tau) \equiv \boldsymbol{\beta}'_{(3)}(\tau)\boldsymbol{x}_{3,t-1} + \delta_{(3)}(\tau)\hat{y}^{o}_{4,t} + u_{3,t}(\tau) \\
\vdots &= \quad \vdots \\
y_{m-1,t} &= \boldsymbol{\theta}'_{(m-1)}(\tau)\widetilde{\boldsymbol{X}}_{(m-1)t} + u_{(m-1)t}(\tau) \equiv \boldsymbol{\beta}'_{(m-1)}(\tau)\boldsymbol{x}_{m-1,t-1} + \delta_{(m-1)}(\tau)\hat{y}^{o}_{m,t} + u_{m-1,t}(\tau)
\end{aligned}$$

Figure 4.1: A representation of the SUR system of equations

Figure 4.1 shows the node specific equations of the complex tail dependency system that represent quantile predictive regressions with nonstationary regressors. Our estimation methodology is based on a SUR system, where the parameter vector is given by $\boldsymbol{\theta}_{(g)}(\tau) = \left(\boldsymbol{\beta}'_{(g)}(\tau), \delta_{(g)}(\tau)\right)' \in \mathbb{R}^{d+1}$ for a fixed quantile level $\tau \in (0, 1)$. The set of regressors for each individual system equation are partitioned to include the d nonstationary regressors and the generated regressor $\hat{y}^{o}_{g't}$ which is employed as systemic risk proxies such that $\widetilde{\boldsymbol{X}}_{(g)t} = [\boldsymbol{x}_{g,t-1} \ \hat{y}^{o}_{st}]$ where $s \in \{1, ..., m\}$ such that $s \neq g$ and $t \in \{1, ..., n\}$. In terms of the estimation procedure, the set of generated regressors $\{\hat{y}^{o}_{2,t}, ..., \hat{y}^{o}_{m,t}\}$ of the system, as illustrated in Figure 4.1, are estimated individually from the corresponding quantile predictive regression models after appropriate indexing as defined by expressions (4.11)-(4.12). Our goal is to the robust estimation of a system of quantile predictive regressions under the specific network type of dependence induced by the VaR $-$ ΔCoVaR risk matrix proposed by Katsouris (2021). In particular, the unique structure of the $\boldsymbol{\Gamma}_{ij}$ risk matrix illustrates that the SUR modelling approach that we employ in our framework is suitable for this more general form of dependence. Our novel estimation approach implies that the equations of the SUR system correspond to node specific quantile predictive regression models with regressands $\boldsymbol{y}_g$ and nonstationary regressors $\boldsymbol{x}_{g,t-1}$.

These models are constructed using only the nonstastionry regressors which are node specific to the system. In other words, the quantities $\hat{y}^{o}_{s,t}$ represent the estimated VaR under nonstationarity and the particular notation is employed to distinguish them from the quantities $\hat{y}_{g,t}$ that correspond to the estimated CoVaR measures. The dependence structure of our tail dependency system implies that the computation procedure requires m of the above systems each of those with $(m-1)$ individual equations after using appropriate indexing. We initially focus on the estimation methodology and related testing methodology of the SUR system that corresponds separately to the columns, rows or diagonals of the $\boldsymbol{\Gamma}_{ij}$ matrix (see, Figure 4.2). A further generalization of our proposed tail dependency driven system can be also considered.

An important application of our estimation methodology is that it facilitates the construction of the VaR − ΔCoVaR matrix in quadratic forms which can be used in portfolio optimization problems with tail events. The exact functional form and the properties of the particular matrix that ensure its suitability for optimal choice problems is beyond the scope of the current study, we briefly discuss the use of our complex tail dependency system when estimating the particular risk matrix proposed by Katsouris (2021). Further regularity conditions regarding the elements of the tail-driven risk matrix with respect to time series properties can be found in Katsouris (2023b).

Consider that the risk matrix has dimensions $(m \times m)$, where m is the sample size of the cross-section of our system, then using the individual equations as displayed by Figure 4.1 we obtain the fitted values, $\{\hat{y}_{1,t}, ..., \hat{y}_{(m-1),t}\}$ which correspond to the estimated CoVaR measures under the assumption of individual-specific nonstationarity. We demonstrate below that the these sequence of values are employed to estimate the elements of the first diagonal in the upper diagonal region of the matrix $\boldsymbol{\Gamma}_{ij}$. Similarly, we can estimate all off-diagonal elements under appropriate indexing. An algorithm that estimates the VaR − ΔCoVaR risk matrix can be found in the study of Katsouris (2021).

$$\boldsymbol{\Gamma}_{ij} := \begin{bmatrix} \mathsf{VaR}_{(1)} & \bullet & & \\ & \mathsf{VaR}_{(2)} & \bullet & \\ & & \ddots & \bullet \\ & & & \mathsf{VaR}_{(m)} \end{bmatrix}_{m \times m}$$

Figure 4.2: A representation of the VaR − ΔCoVaR matrix

Remark 4.3. The proposed risk matrix is constructed by combining cross-sectional and time series information. A related study to the aspect of modelling cross section and time series data combinations is presented in the seminal study of Hussian and Wallace (1969) who provide a unifying theory for the use of error component models in combining cross section with time series data.

4.2.3 Main Assumptions

Our framework considers m deterministic equations which resemble a form induced network-dependence across these m nodes of the system which correspond to the nonstationary quantile predictive regressions. Thus, the corresponding m pairs of innovation sequences are independent and for each node $s \in \{1, ..., m\}$ their error vectors $\{u_t, \boldsymbol{v}_t\}_{t=1}^{n}$ satisfy an invariance principle (FCLT) that follows a correlated OU process (see, Phillips and Magdalinos (2009), Nkurunziza (2010) and Chen et al. (2023) among others). Suppose that we have a sequence of n observed time series such that

$$\{(\boldsymbol{X}_1^{(g)}, Y_1^{(g)}), (\boldsymbol{X}_2^{(g)}, Y_2^{(g)}), ..., (\boldsymbol{X}_n^{(g)}, Y_n^{(g)})\}, \quad \text{for} \quad g \in \{1, ..., m\} \tag{4.20}$$

are observed at discrete times $0 < t_1 < t_2 < ... < t_n$, where $\boldsymbol{X}_t^{(g)}$ and $Y_t^{(g)}$, represent respectively the sizes of the population of the two pairs observed at time t_i, $i = 1, ..., n$. From a methodological point of view, we construct a complex tail dependency driven system for modelling the pairs $(\boldsymbol{X}_{t_i}^{(g)}, Y_{t_i}^{(g)})$, for $i = 1, ..., n$ and $g = 1, ..., m$. Let s be a given node where $s \in \{1, ..., m\}$. Moreover, we assume that the dependent variable $\boldsymbol{y}_g$ of a node g, where $g \in \{1, ..., m\}$ is generated from a nonstationary predictive regression model with nonstationary regressors generated using the LUR parametrization. We impose the following assumptions on the variables and error sequences of the seemingly unrelated system of equations that facilitate the development of the asymptotic theory.

Assumption 4.1. (*innovation structure*) Define the filtration $\mathcal{F}_t = \sigma\{\boldsymbol{w}_t, \boldsymbol{w}_{t-1}, ...\}$, where $\boldsymbol{w}_t = (\boldsymbol{u}, \boldsymbol{v}_t)'$. Let $\{\boldsymbol{w}_t, \mathcal{F}_t\}$ be an independent *martingale difference sequence* with

(i) The variance-covariance matrix of the error vector $\boldsymbol{w}_{(g)t}$ for the $g-$th predictive regression model has the following form

$$\boldsymbol{\Sigma}_{ww}^{(g)} := \mathbb{E}_{\mathcal{F}_{t-1}}\left[\boldsymbol{w}_{(g)t}\boldsymbol{w}_{(g)t}'\right] \equiv \begin{bmatrix} \boldsymbol{\Sigma}_{uu}^{(g)} & \boldsymbol{\Sigma}_{uv}^{(g)} \\ \\ \boldsymbol{\Sigma}_{vu}^{(g)} & \boldsymbol{\Sigma}_{vv}^{(g)} \end{bmatrix} \text{ and } \mathbb{E}\left[\left\|\boldsymbol{w}_{(g)t}\right\|\right] > L \ a.s \ \forall \ t \leq n, \tag{4.21}$$

where $\boldsymbol{w}_{(g)t} = \left(u_{g,t}, \boldsymbol{v}_{g,t}'\right)'$ such that $\boldsymbol{\Sigma}_{ww}^{(g)} > 0$ is positive definite matrix $\forall \ t \leq n$.

(ii) Denote with $\boldsymbol{w}_t = \left(\boldsymbol{w}_{(1)t}, ..., \boldsymbol{w}_{(m)t}\right)'$, then $\boldsymbol{\Omega}$ represents the block covariance matrix of the system which has the following form[11]

$$\boldsymbol{\Omega} := \mathbb{E}_{\mathcal{F}_{t-1}}\left[\boldsymbol{w}_t\boldsymbol{w}_t'\right] \equiv \begin{bmatrix} \boldsymbol{\Sigma}_{ww}^{(1)} & & & \\ & \boldsymbol{\Sigma}_{ww}^{(2)} & & \\ & & \ddots & \\ & & & \boldsymbol{\Sigma}_{ww}^{(m)} \end{bmatrix}, \quad \boldsymbol{\Omega} \overset{i.i.d}{\sim} \mathsf{Wishart}_m(\boldsymbol{0}, \boldsymbol{V}). \tag{4.22}$$

[11]We assume a set of time invariant covariance matrices, but this condition can be further relaxed to capture conditional heteroscedasticity dynamics within each quantile predictive regression model.

Remark 4.4. Assumption 4.1 implies that $\left\{\boldsymbol{\Sigma}_{ww}^{(g)}\right\}$ where $g \in \{1,...,m\}$ represents an *i.i.d* sequence of covariance matrices generated randomly from a Wishart distribution. In particular, these generated covariance matrices obtained from the Wishart process are positive definite which implies that are well-defined as the covariance matrices for the Gaussian processes that generates the innovation sequences of the system equations, thereby preserving the corresponding linear processes representations and related properties of Phillips and Solo (1992).In other words, the data generating mechanism implies that each covariance matrix $\boldsymbol{\Sigma}_{ww}^{(g)}$ can be employed to generate the error vector of the predictive regression models for all $g \in \{1,...,m\}$. In addition, cross-sectional restrictions are imposed by assuming that $\boldsymbol{\Omega}$ is block-diagonal.

Moreover, following conventional assumptions commonly imposed in econometric environments of nonstationary time series models, we consider that the innovation vector for each individual equation of the system forms a martingale difference sequence, which is a condition commonly employed for estimation and inference in predictive regression models (see, Phillips and Magdalinos (2009), Kostakis et al. (2015), Kasparis et al. (2015), Phillips and Lee (2013), Phillips and Lee (2016), Lee (2016) Fan and Lee (2019) and Gonzalo and Pitarakis (2012, 2017)). Although, the proposed data generating mechanism of our framework implies randomly obtaining these covariance matrices when obtaining the innovation sequences from independent Gaussian processes that allow to obtain the pairs $\{Y_g, \boldsymbol{X}_g\}$ for all $g \in \{1,...,m\}$ the *martingale difference sequence* condition is not violated. Under suitable parametrizations conditional heteroscedasticity can be also incorporated, which is useful when modelling latent volatility dynamics (see, Magdalinos (2021)). We leave the implementation of the particular aspect, which requires extra refinements for future work.

Denote with $\mathcal{F}_t$ is a sequence of increasing $\sigma-$fields which for each $g \in \{1,...,m\}$ is independent of the innovation sequence of each system specific predictive regression model. Thus, the existence of a correlated vector Brownian motion (W, V), for each partial sum processes of the innovation vectors of these individual equations is applicable and facilitates the development of the asymptotic theory. Denote with $\left(u_{(g)m}, \boldsymbol{v}'_{(g)m}\right)'_{m\geq 1}$ to be a sequence of random vectors such that $\mathcal{F}_m = \sigma\left(u_{(g)m}, \boldsymbol{v}'_{(g)m}\right)$.

Assumption 4.2 (*triangular representation*)**.** Suppose that Assumptions 4.1 hold. Then, each individual system equation has the following triangular representation

$$\mathcal{M}_n := \left\{ \begin{array}{l} y_{(g)m,n} = \boldsymbol{\beta}'_{g,n}\boldsymbol{x}_{(g)m-1,n} + u_{(g)m,n} \\ \boldsymbol{x}_{(g)m,n} = \boldsymbol{R}_{g,n}\boldsymbol{x}_{(g)m-1,n} + \boldsymbol{v}_{(g)m,n} \end{array} \right\} \tag{4.23}$$

where the autocorrelation coefficient $\boldsymbol{R}_{g,n} = \left(\boldsymbol{I}_d - \dfrac{\boldsymbol{C}_d^{(g)}}{n}\right)$ where $m \geq 1$ and $g \in \{1,...,m\}$.

Remark 4.5. Assumption 4.3 provides a suitable triangular representation for the predictive regression model across the g system specific equations, for $g \in \{1,...,m\}$. Asymptotic theory relies in the long-run relation between the triangular arrays $y_{(g)m,n}$ and $\boldsymbol{x}_{(g)m,n}$ that are not cointegrated. In particular, limit theory corresponds to the case of $n \to \infty$ and $m \to \infty$ sequentially, which we denote with $(n,m)_{\textsf{seq}} \to \infty$. Detailed discussion for asymptotics when the limit behaviour of both dimensions are taken simultaneously can be found in the studies of Phillips and Moon (1999).

Assumption 4.3 (*conditional homoscedasticity*)**.** For each $n \geq 1$ $\left\{ u_{(g)t}, \mathcal{F}_{t,n} \right\}$ forms a martingale difference satisfying the following almost surely convergence

$$\lim_{t \to \infty} \sup_{n \geq t} \left| \mathbb{E}\left(u^2_{(g)t} | \mathcal{F}_{t,n} \right) - \sigma^2_{uu} \right| = 0, \quad \text{almost surely} \tag{4.24}$$

where $\mathcal{F}_{t,n} = \sigma(\boldsymbol{u}_1, ..., \boldsymbol{u}_n; \boldsymbol{x}_1, ..., \boldsymbol{x}_n)$ for $t = 1, ..., n$ with $n > 30$ for all $g \in \{1, ..., m\}$.

Assumption 4.4 (*FCLT*)**.** Suppose that Assumptions 4.1-4.5 hold. Denote with $\boldsymbol{w}_{(g)t} = \left(u_{(g)t}, \boldsymbol{v}'_{(g)t} \right)'$, then under regularity conditions, an invariance principle holds (see, Phillips and Solo (1992))

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{w}_{(g)t} := \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} \begin{bmatrix} u_{(g)t} \\ \boldsymbol{v}_{(g)t} \end{bmatrix} \equiv \begin{bmatrix} B^{(g)}_{un}(r) \\ \boldsymbol{B}^{(g)}_{vn}(r) \end{bmatrix} \Rightarrow \begin{bmatrix} B^{(g)}_{u}(s) \\ \boldsymbol{B}^{(g)}_{v}(s) \end{bmatrix} := \mathcal{BM}^{(g)} \begin{bmatrix} \sigma^2_{uu} & \boldsymbol{\sigma}'_{uv} \\ \boldsymbol{\sigma}_{vu} & \boldsymbol{\Sigma}_{vv} \end{bmatrix}_{(d+1)\times(d+1)} \tag{4.25}$$

where $\boldsymbol{\Sigma}^{(g)}_{vv} \in \mathbb{R}^{d \times d}$ is a positive definite covariance matrix and $0 < r < 1$ for all $g \in \{1, ..., m\}$.

To simplify the notation we exclude the index g for the various covariance terms of the above Brownian motions, then due to independently generated measures the following results hold for all $g \in \{1, ..., m\}$. In particular, the individual components of the vector sequence $\boldsymbol{w}_t = (u_t, \boldsymbol{v}'_t)'$ have partial sums processes that weakly converge into their Brownian motion counterparts

$$B_{un}(r) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} u_t \Rightarrow B_u(r) := \mathcal{N}(0, r\sigma^2_{uu}) \tag{4.26}$$

$$\boldsymbol{B}_{vn}(r) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{v}_t \Rightarrow \boldsymbol{B}_v(r) := \mathcal{N}(\boldsymbol{0}, r\boldsymbol{\Sigma}_{vv}) \tag{4.27}$$

Denote with $\boldsymbol{B}(r) = (B_u(r), \boldsymbol{B}_v(r)')'$ a $(d \times 1)$ Brownian motion with long-run covariance matrix $\boldsymbol{\Sigma}_{ee}$, that is, a Gaussian vector process with almost surely continuous sample paths. More precisely, since $\boldsymbol{x}_t$ is an adapted process to the filtration $\mathcal{F}_t$ then this implies that there exists a correlated vector Brownian motion $\boldsymbol{B}_n(r) = (B_{un}(r), \boldsymbol{B}_{vn}(r)')'$ such that

$$\left(\frac{1}{\sqrt{n}} \sigma^{-1}_{uu} \sum_{t=1}^{\lfloor nr \rfloor} u_t, \frac{1}{\sqrt{n}} \boldsymbol{\Sigma}^{-1/2}_{vv} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{v}_t \right)' \Rightarrow \boldsymbol{B}(r) = (B_u(r), \boldsymbol{B}_v(r)')', \quad 0 < r < 1 \tag{4.28}$$

on $\mathcal{D}_{\mathbb{R}^2}([0,1])^2$ as $n \to \infty$, with covariance matrix as in (4.25) implying joint convergence. Then the following local to unity principle applies (see, Phillips (1987a))

$$\frac{\boldsymbol{x}_{[nr]}}{\sqrt{n}} \Rightarrow \boldsymbol{J}_C(r), \quad \text{where} \;\; \boldsymbol{J}_c(r) = \int_0^r e^{(r-s)\boldsymbol{C}} d\boldsymbol{B}_v(s). \tag{4.29}$$

where $\boldsymbol{B}_v$ is a Brownian motion with a positive-definite covariance matrix $\boldsymbol{\Omega}_{xx}$. The functional $\boldsymbol{J}_C(r)$ represents the *Ornstein-Uhlenbeck* stochastic process, that is, the solution of *Black-Scholes* differential equation given by $d\boldsymbol{J}_C(r) = \boldsymbol{C}\boldsymbol{J}_C(r) + d\boldsymbol{B}_v(r)$, with an initial condition $\boldsymbol{J}_C(r) = 0$. We also assume that the initial values of the nonstationary series $\{\boldsymbol{x}_{\text{g},t}\}$ has initial values such that $\boldsymbol{x}_{\text{g},0} = \boldsymbol{0}$ for all $\text{g} \in \{1, ..., m\}$.

Assumption 4.3 provides the triangular representation for the system equations which is particularly useful for constructing equivalent Linderburg type of conditions that apply to the corresponding triangular arrays, such that $\lim_{n \to \infty} \sum_{j=1}^{n} \mathbb{E}|u_{(g)j}|^{2+s} = 0$ for some $s > 0$ for all $g \in \{1, ..., m\}$. This condition ensures that the error term that corresponds to the predictive regressions models that generate the particular general form of dependence for our complex system, excludes the presence of fat tails in their distribution functions. The array $\left\{ \boldsymbol{x}_{(g)m,n} \right\}_{m \geq 1}$ with $n \geq 30$ is correesponds to a near unit root process, which permits the use of the classical invariance principle. In terms of the estimation methodology, we employ the framework of quantile predictive regression models to obtain estimates for the risk measures of VaR and CoVaR.

Therefore, the LUR specification (with $\gamma = 1$) induces the presence of near unit root processes with localizing coefficients of persistence such that $c_j \neq 0$ where $j \in \{1, ..., d\}$ and d being the number of nonstationary regressors. The particular specification is a standard assumption in the near-integrated time series literature (see, Phillips (1987b), Chan and Wei (1987), Hansen (1992) and Buchmann et al. (2007) among others). As a result, the array $\left\{ y_{(g)m,n} \right\}_{m \geq 1}$ represents a predictive regression process with predictors generated as near unit root processes. Our research objective is to develop an estimation and inference framework for our complex tail dependency system under the assumption of stochastic processes with exactly these features; that is, high persistence regressors[12], which is a stylized feature of financial variables found in empirical studies.

Notice that the proposed data generating mechanism provided by Assumption 4.2 ensures that the dependence structure of the vector innovation across the individual equations of the system have desirable properties such as asymptotic independence. More precisely, these invariance principles hold for each of those predictive regressions which allows to employ conventional local-to-unity limit theory results. Moreover, Assumption 4.3 establishes a $m.d.s$ condition for the error term of the predictive regressions across all individual system equations holds; which ensures conditional homoscedasticity as well. Furthermore, Assumption 4.5 provides invariance principles laws for the innovation vector of each of the system specific equations which is essential for deriving the asymptotic behaviour of estimators and test statistics in our framework.

Remark 4.6. The proposed mechanism under the Assumption on the innovation structure of the system provides a more general form of dependence without imposing a factor structure. Current methodologies in the literature focus on estimation and inference when pooling cross-sectional and time-series data (see, Balestra and Nerlove (1966), Maddala (1971), Mundlak (1978)). We bridge the gap between *triangular-recursive* and *interdependent* (i.e., non-triangular) systems (Basmann (1963)). The classical causality hypothesis for a system of structural equations proposed by Wold-Strotz (Strotz and Wold (1960)) required the concept of triangular-recursivness (e.g., the matrix A to be triangular) as well as that each $i-$th equation to be interpreted as the OLS of y_{ti} on the remaining variables in the equation. Under triangular representation assumption imposed to each nonstationary quantile predictive regression model facilitates the development of the asymptotic theory for both the pooled estimator as well as the equation-specific estimator using conventional local-to-unity asymptotics. Therefore, the proposed modelling methodology preserves the conventional *structural causality interpretation* without destroying the notion of *isolation* as per Basmann (1963).

[12] Buchmann et al. (2007) proposes a framework for nearly unstable processes under strong dependence, which is related to high persistence predictors especially when a strong correlation between the error terms of these two processes exists.

Then, the conditional quantile function of y_t denoted with $\mathsf{Q}_{y_t}(\tau|\mathcal{F}_{t-1})$, replaces the conditional mean function of the predictive regression which implies the following model specification

$$\mathsf{Q}_{y_t}(\tau|\mathcal{F}_{t-1}) := F^{-1}_{y_t|\boldsymbol{x}_{t-1}}(\tau) \equiv \alpha(\tau) + \boldsymbol{\beta}(\tau)'\boldsymbol{x}_{t-1}. \tag{4.30}$$

such that $F_{y_t|\boldsymbol{x}_{t-1}}(\tau) := \mathbb{P}(y_t \leq \mathsf{Q}_{y_t}(\tau|\mathcal{F}_{t-1})\,|\mathcal{F}_{t-1}) \equiv \tau$, where $\tau \in (0,1)$ is some quantile level in the compact set $(0,1)$. Therefore, in order to define the innovation structure that corresponds to the quantile predictive regression, we employ the piecewise derivative of the loss function such that $\psi_\tau(\mathsf{u}) = \left[\tau - \mathbf{1}\{\mathsf{u} < 0\}\right]$. Consequently, this implies that $u_t(\tau) := u_t - F^{-1}_u(\tau)$ where $F^{-1}_u(\tau)$ denotes the unconditional $\tau-$quantile of the error term u_t. Then, the corresponding invariance principle for the nonstationary quantile predictive regression model is formulated as below

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \begin{bmatrix} \psi_\tau(u_t(\tau)) \\ \boldsymbol{v}_t \end{bmatrix} \Rightarrow \begin{pmatrix} B_{\psi_\tau}(r)_{(1\times n)} \\ \boldsymbol{B}_v(r)_{(p\times n)} \end{pmatrix} \equiv \mathcal{BM} \begin{bmatrix} \tau(1-\tau) & \boldsymbol{\sigma}'_{\psi_\tau v} \\ \boldsymbol{\sigma}_{v\psi_\tau} & \boldsymbol{\Omega}_{vv} \end{bmatrix} \tag{4.31}$$

Assumption 4.5. Under Assumption 4.2 and 4.3, then the following conditions hold:

(i) The sequence of stationary conditional *probability distribution functions (pdf)* denoted with $\{f_{u_t(\tau),t-1}(.)\}$ evaluated at zero with a non-degenerate mean function such that $f_{u_t(\tau)}(0) := \mathbb{E}\left[f_{u_t(\tau),t-1}(0)\right] > 0$ satisfies a *FCLT* given as below

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \left(f_{u_t(\tau),t-1}(0) - \mathbb{E}\left[f_{u_t(\tau),t-1}(0)\right] \right) \Rightarrow B_{f_{u_t(\tau)}}(r). \tag{4.32}$$

(ii) For each t and $\tau \in (0,1)$, $f_{u_t(\tau),t-1}(.)$ is uniformly bounded away from zero with a corresponding conditional distribution function $F_t(.)$ which is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$ (see, Goh and Knight (2009) and Lee (2016)).

$$\sup_{0\leq r\leq 1} \left| \frac{1}{n^{1-\delta}} \sum_{t=1}^{\lfloor nr \rfloor} \left[f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0) \right] \right| = o_p(1).$$

4.2.4 Equation Specific Estimation Methodology

As we mentioned earlier two important features which we need to take into consideration within our estimation environment is the quantile estimation procedure as well as the nonstationary properties of regressors when fitting the model (see related discussion in Dahlhaus (1997)). We begin by considering the estimation methodology for each individual system equation separately (i.e., equation-by-equation estimation) which has the same philosophy as the framework proposed by Lee (2016). Denote with $\mathcal{Q}_{y_j}(\tau|\boldsymbol{x}_j) = F^{-1}_{y_j}(\tau|\boldsymbol{x}_j)$, $\tau \in (0,1)$ the conditional quantile function. Then, the estimation methodology requires to fit a quantile predictive regression model using the conditional quantile specification function.

Therefore, the corresponding quantile regression optimization function for obtaining model estimates is expressed with the following form:

$$\mathcal{Q}_{\tau}\left(y_j|\boldsymbol{x}_j\right) = \underset{q(\boldsymbol{x})}{\text{arg min}}\ \mathbb{E}\Big[\rho_{\tau}\Big(y_j - q(\boldsymbol{x}_j)\Big)\Big] \tag{4.33}$$

where τ is a fixed quantile level in the compact set $(0,1)$ and $\rho_{\tau}(\mathsf{u}) = \mathsf{u}\left(\tau - \mathbf{1}\{\mathsf{u} \leq 0\}\right)$ is the check function which is convex but nondifferentiable. Then, the model estimator that correspond to one of these system equations is defined as below

$$y_{(g)t} = \boldsymbol{\theta}'_{(g)}(\tau)\widetilde{\boldsymbol{X}}_{(g)t} + u_{(g)t} \equiv \boldsymbol{\beta}'_{(g)}(\tau)\boldsymbol{x}_{g,t-1} + \delta_{(g)(\tau)}\hat{y}^{o}_{(g')t} + u_{(g)t}, \quad \text{for some } g \neq g', \tag{4.34}$$

$$\boldsymbol{x}_{g,t} = \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d^{(g)}}{n}\right)\boldsymbol{x}_{g,t-1} + \boldsymbol{v}_{g,t}, \quad \text{where} \quad \boldsymbol{C}_d^{(g)} > 0 \quad \text{for all} \quad g \in \{1,...,m\}\,. \tag{4.35}$$

The econometric specification given by expressions (4.34) and (4.35) encompasses the main idea of our proposed modelling methodology. In particular, the generated regressor $\hat{y}^{o}_{(g')t}$ represents the risk measure of VaR for a fixed quantile $\tau \in (0,1)$ which in our setting is obtained under the assumption of nonstationarity. On the other hand, $y_{(g)t}$ represents the risk measure of CoVaR which is obtained based on the generated regressor as well as the nonstationary regressors of the particular node. Therefore, our interest lies in the estimation and inference for the parameter $\boldsymbol{\theta}_{(g)}(\tau)$. Thus, the quantile dependent estimator $\boldsymbol{\theta}_{(g)}(\tau) \in \mathbb{R}^{d+1}$ for some fixed quantile $\tau \in (0,1)$ is estimated using the following optimization function

$$\widehat{\boldsymbol{\theta}}_{(g)}\left(\tau\right) = \underset{\boldsymbol{\theta}_{(g)} \in \mathbb{R}^{d+1}}{\text{arg min}} \sum_{t=1}^{n} \rho_{\tau}\left(y_{(g)t} - \boldsymbol{\theta}'_{(g)}(\tau)\widetilde{\boldsymbol{X}}_{(g)t}\right) \tag{4.36}$$

where $\widehat{\boldsymbol{\theta}}_{(g)}\left(\tau\right)$ is the quantile dependent estimator of $\boldsymbol{\theta}_{(g)}\left(\tau\right)$ based on the $d-$lagged regressors, $\boldsymbol{x}_{(g)t-1}$, plus the systemic risk proxy, $\hat{y}^{o}_{(s)t}$ which is a scalar vector, where $s \in \{1,...,m\}$ such that $s \neq g$ and is selected from the available vector of regressands $\boldsymbol{y}_t = \left[y_{1t},...,y_{mt}\right]$ to represent the tail dependence of the (g,s) pair of nodes (explanatory covariate for the $g-$th equation).

We model the regressors $\boldsymbol{x}_{(g)t-1}$ using an autoregressive model with local unit root coefficient, which captures the unknown peristence properties in the time series of these regressors. Further details of the corresponding modelling framework and asymptotic theory under the assumption of nonstationary regressors (i.e., regressors generated by the LUR specification) can be found in Lee (2016) and Fan and Lee (2019). Firstly, the implications of our proposed econometric environment can be better understood in the context of cross-sectional time series regression models, even though we assume that the cross-sectional dimension in our setting represents a graph structure. Therefore, some implications in the econometric literature include the consistency of estimators as well as the development of asymptotic theory[13].

[13] In particular, Balestra and Nerlove (1966) mention that: *"the reason that ordinary least squares estimates are inconsistent when lagged variables are included is that these variables are correlated with the current values of the residuals since they are determined to the same degree as the current value of the dependent variables"*. Furthermore, the authors mention: *"One solution to this difficulty is to use as instrumental variables a sufficient number of other exogenous or in the absence of serially correlated residuals, lagged endogenous variables appearing elsewhere in the system in the formation of the normal equations so that the current endogenous variables in the equation need not be used for this purpose."* These two sentences borrowed from Balestra and Nerlove (1966) describe the main idea behind the IVX instrumentation method proposed by Phillips and Magdalinos (2009)

Secondly, we consider the parametric estimation of our econometric specifications (e.g., quantile predictive regression models). In other words, the corresponding throughout the paper we operate under the assumption that the error processes follow a Gaussian distribution. Therefore, the appearance of different convergence rates due to the two estimators we employ and compare (OLS and IVX estimator) can only be contributed to the properties of these estimators rather than the tail behaviour of the underlying distribution. Moreover, when the design matrix includes nonstationary regressors, then we need to consider the near unit root properties of the underline processes, to avoid the presence of singularities (e.g., see Phillips and Magdalinos (2013), and Chen et al. (2023)). In addition, relevant frameworks from the quantile panel data regression literature are proposed in the studies of Chen (2022), Feng (2023) and Belloni et al. (2023).

4.2.5 IVX Instrumentation Methodology

The IVX instrumentation proposed by Phillips and Magdalinos (2009) Phillips and Magdalinos (2009) implies the use of a mildly integrated instrumental variable and is expressed via the following form

$$\widetilde{\boldsymbol{z}}_{tn} = \sum_{j=0}^{t-1} \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_z}{n^{\gamma_z}} \right) (\boldsymbol{x}_{t-j} - \boldsymbol{x}_{t-j-1}), \tag{4.37}$$

where $\boldsymbol{C}_z = \mathsf{diag}\{c_{z1}, ..., c_{zd}\}$ is a $(d \times d)$ diagonal matrix such that $c_{zj} > 0 \; \forall \; j \in \{1, ..., d\}$ with $0 < \gamma_z < 1$, where γ_z is the exponent rate of the persistence coefficient which corresponds to the instrumental variable. The IVX filtration transforms the autoregressive process generating the set of regressors, $\boldsymbol{x}_t$, which encompasses either stable or unstable processes depending on the behaviour of the local unit root coefficient, into a mildly integrated process which is less persistent than the endogenous variables of the system, denoted with $\boldsymbol{x}_t$. A relevant aspect for inference purposes is the choice of the exponent rate of persistence for the instrumental variables, where a choice of γ_z close to 0.95 is found to be a reasonable value with desirable finite-sample properties (see, Phillips and Lee (2016) and Lee (2016)) which we also employ in our simulation study. Furthermore, our framework is concerned with estimation and inference for quantile predictive regression systems. Thus, is crucial to establish the properties of the corresponding system IVX-based estimator.

We aim to demonstrate through our asymptotic theory analysis that the SUR-IVX estimator has the property of being robust against the unknown persistence properties of regressors as these are captured by the coefficients of persistence c_j. Therefore, the estimation procedure of the SUR system is implemented into two steps. Practically, our modelling methodology proposes a doubly IVX corrected estimator for Quantile predictive regression models with nonstationary and generated regressors. The proposed doubly IVX estimator can be shown to automatically provide robustness to abstract degree of persistence regardless of the presence of generated regressor in the the second stage of the estimation methodology. In other words, first for each equation of the system, that is, the quantile predictive regression model, we obtain the IVX estimator using the IVX instrumentation methodology. Specifically, we obtain the IVX residuals to estimate the overall covariance matrix using the IVX residuals such that

$$\hat{u}_j = y_j - \boldsymbol{X}_j' \boldsymbol{\beta}_j^{IVX}, \quad \text{with} \quad \sigma_{jj}^2 = \frac{1}{m} \sum_{j=1}^{m} \hat{\boldsymbol{u}}_j \hat{\boldsymbol{u}}_j' \tag{4.38}$$

which implies that the block diagonal matrix of the system has the following form

$$\widehat{\boldsymbol{\Omega}} := \frac{1}{n} \sum_{t=1}^{n} \left[\hat{\boldsymbol{w}}_t \hat{\boldsymbol{w}}_t' \right] = \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_{ww}^{(1)} & & & \\ & \widehat{\boldsymbol{\Sigma}}_{ww}^{(2)} & & \\ & & \ddots & \\ & & & \widehat{\boldsymbol{\Sigma}}_{ww}^{(m)} \end{bmatrix} \tag{4.39}$$

The existence of a feasible and consistent estimator of the overall covariance matrix $\widehat{\boldsymbol{\Omega}}$ of the SUR system is important for robust inference. In practise each covariance matrix estimator that corresponds to each of system's equations is constructed based on the residuals of the quantile predictive regression model. Notice also that in practise the particular distributional conditions given by Assumption 4.5 ensure that the m pairs of error processes are independent with the same distribution, while each pair follows correlated Ornstein-Uhlenbeck processes (see, also the framework of Nkurunziza (2010)).

4.3 Econometric Framework of Seemingly Unrelated Systems

Consider the linear predictive regression model formulated as below

$$y_t = \mu + \boldsymbol{\beta}' \boldsymbol{x}_{t-1} + u_t, \quad 1 \leq t \leq n \tag{4.40}$$

$$\boldsymbol{x}_t = \boldsymbol{R}_n \boldsymbol{x}_{t-1} + \boldsymbol{v}_t \tag{4.41}$$

where $\boldsymbol{x}_t \in \mathbb{R}^p$ is a $p-$dimensional vector and $\boldsymbol{R}_n = \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}_p}{n^\gamma} \right)$ with $\gamma = 1$.

Assumption 4.6. Let $\boldsymbol{\epsilon}_t = (u_t, \boldsymbol{e}_t')'$ be an $\mathbb{R}^{p+1}-$valued martingale difference sequence with respect to the filtration $\mathcal{F}_t = \sigma(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, ...)$ satisfying $\mathbb{E}\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \mid \mathcal{F}_{t-1}\right] = \boldsymbol{\Sigma}_\epsilon > 0$. Let $\boldsymbol{v}_t$ be an $\mathbb{R}^p-$valued stationary linear process such that $\boldsymbol{v}_t = \sum_{j=0}^{\infty} \boldsymbol{C}_j \boldsymbol{e}_{t-j}$, where $\boldsymbol{C}_j$ is a sequence of constant matrices such that $\boldsymbol{I}_0 = \boldsymbol{I}_p$, $\sum_{j=0}^{\infty} \boldsymbol{C}_j$ has full rank, and $\sum_{j=0}^{\infty} \| \boldsymbol{C}_j \| < \infty$.

Borrowing from the literature in the case when $\sigma_{uv} \neq 0$ (which is the case we consider in this paper), the exact OLS bias of $\hat{\beta}$ is computed from the predictive regression model is given by $\mathbb{E}\left[\hat{\beta} - \beta\right] = \delta \mathbb{E}\left[\hat{\rho} - \rho\right]$ where $\hat{\rho}$ is the OLS estimate of ρ and $\delta := \sigma_{uv} / \sigma_v^2$ is the slope coefficient of in a regression of u_t on v_t. Furthermore, since $\hat{\rho}$ is known to be downward biased in small-samples, and $(u_t, v_t)'$ are typically strongly negatively contemporaneously correlated, the autoregressive OLS bias feeds into the small-sample distribution of $\hat{\beta}$ causing over-rejections of the null hypothesis of no predictability, $H_0 : \beta = 0$. Moreover, notice even with the use of a possible finite-sample-bias correction on the OLS estimate, this reduces the noncentrality of the limiting distribution of the OLS t-statistic, but the distribution remains nonstandard in the near-integrated case. Therefore, a solution to the particular problem which provides robust statistical inference is the use of the IVX instrumentation, which ensures instrument relevance while controlling for persistence. In particular, the IVX instrument is constructed such that

$$z_t := \sum_{j=0}^{t-1} \rho_z^j \Delta x_{t-j}, \quad \rho_z = \left(1 - \frac{c_z}{n^{\gamma_z}} \right), c_z > 0 \text{ and } \gamma_z \in (0, 1). \tag{4.42}$$

In other words, since we operate within the framework of moderate deviations from unity, the constructed IVX instrument for x_t, is chosen so that z_t is by construction only mildly integrated when the predictor x_t is nearly integrated.

Therefore, the IVX estimator of β is found to have a slower convergence rate than the conventional OLS estimator under near integration, such that $n^{\frac{1+\gamma_z}{2}}$. In addition, the IVX estimator is mixed Gaussian in the limit irrespectively of the degree of endogeneity implied by δ, leading to standard inference in t and Wald tests. On the other hand, it can be proved that under low persistence, the IVX estimator is asymptotically equivalent to the OLS procedure. To obtain the IVX estimator we write the corresponding model with the demeaned variates such as

$$y_t^{\mu} = \boldsymbol{\beta}' \boldsymbol{x}_{t-1}^{\mu} + v_t^{\mu} \tag{4.43}$$

where the sample moments are estimated as $\bar{y}_t = y_t - \frac{1}{n-1} \sum_{t=1}^{n} y_t$. Then, the IVX estimator is constructed

$$\widehat{\boldsymbol{\beta}}^{ivx} = \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{t-1} \boldsymbol{z}_{t-1}' \right)^{-1} \left(\sum_{t=1}^{n} y_t \boldsymbol{z}_{t-1}' \right). \tag{4.44}$$

Furthermore, denote with $\widehat{\boldsymbol{\beta}}^{ols}$ the OLS estimator and define the residuals $\widehat{u}_t = \bar{y}_t - \widehat{\boldsymbol{\beta}}^{ols} \bar{\boldsymbol{x}}_{t-1}$ and the residual variance estimator $\widehat{\sigma}_u^2 = \frac{1}{n-1} \sum_{t=1}^{n} \widehat{u}_t^2$. Then, the IVX-Wald statistic under the null hypothesis is

$$\mathcal{W} = \widehat{\boldsymbol{\beta}}^{ivx\prime} \widehat{\boldsymbol{V}}^{-1} \widehat{\boldsymbol{\beta}}^{ivx} \tag{4.45}$$

where a feasible estimator for the covariance matrix is given by

$$\widehat{\boldsymbol{V}} = \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{t-1} \boldsymbol{z}_{t-1}' \right)^{-1} \left(\widehat{\sigma}_u^2 \sum_{t=1}^{n} \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}' \right) \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{t-1} \boldsymbol{z}_{t-1}' \right)^{-1} \tag{4.46}$$

Remark 4.7. The assumption of martingale difference sequence for the innovation term u_t allows to impose further assumptions regarding the modelling of conditional heteroscedasticity via volatility processes such as ARCH/GARCH. However, this aspect is beyond the scope of our study. Although, the assumption of KMS is that the innovations (u_t, v_t) are a correlated linear process, if we assume that u_t is a martingale difference sequence which is a special case of the more general case, then the asymptotic theory for the IVX estimator simplifies. The condition imposed by Assumption 4.2 does not violate conditional homoscedasticity assumption, especially since no Garch specifications are employed as in Kostakis et al. (2015). This condition can be further relaxed and a HC IVX-Wald test can be modified in our framework in a similar manner as in the study of Magdalinos (2021)). Therefore, our goal is to investigate the asymptotic theory and finite-sample performance of our proposed test statistic. In particular, we are interested in investigating the presence of systemic risk effects based on our complex tail dependency system.

Next, we establish the consistency and asymptotic mixed Gaussianity of the doubly IVX corrected estimator $\boldsymbol{\beta}^{ivx}(\tau)$. Recall that the QR model estimator is obtained as below

$$\widehat{\beta}(\tau) = \arg\min_{\beta} \sum_{i=1}^{n} \rho_{\tau}\Big(y_t - \boldsymbol{x}'_{t-1}\boldsymbol{\beta} \Big) \tag{4.47}$$

where the check function is defined as $\rho_{\tau}\{\tau - \mathbf{1}\,(u \geq 0)\,\}u$ is the check function. Therefore, one uses the estimates of $\boldsymbol{\theta}_j$, denoted by $\widehat{\boldsymbol{\theta}}_j$, in the first step, to obtain the generated regressor $\widehat{\boldsymbol{x}}_t$. Our aim is to demonstrate that our proposed methodology achieve rate and model double robustness simultaneously, provided that the parameter θ satisfies certain regularity conditions. Furthermore, in the current study the doubly robust estimation approach corresponds to conditional quantile specification forms specifically for nonstationary data.

4.3.1 Two-Stage Estimation Procedure

Assume the existence of any two pair of dependent random variables $\{y_{1t}, y_{2t}\}$ and a set of predictors $\boldsymbol{x}_t$ for $t = 1, ..., n$. Notice the steps below should not be confused with the first or second stage estimators in instrumental variable regressions. These procedures refer to the two different models under examination. Then, each of these two procedures indeed have a first and second stage estimation step due to the fact that the nonstationary regressors of the model have the local-to-unity specification. The two-stage estimation procedure for the risk measure pair using the nonstationary quantile predictive regression models is described in more details in the study of Katsouris (2023a).

First-Stage Procedure

During the first stage of our procedure we obtain a consistent estimator of the parameter vector. We focus on the IVX estimator which is found to be robust to the abstract degree of persistence. The econometric model is the linear predictive regression as below

$$y_{1,t} = \beta_{01}(\tau) + \boldsymbol{\beta}'_{11}(\tau)\boldsymbol{x}_{1,t-1} + u_{1t}(\tau), \quad \text{for} \quad t = 1, ..., n \tag{4.48}$$

$$\boldsymbol{x}_{1,t} = \boldsymbol{R}_{1,n}\boldsymbol{x}_{1,t-1} + \boldsymbol{v}_{1,t} \tag{4.49}$$

Denote with $\boldsymbol{\beta}(\tau) = \big(\beta_{01}(\tau), \boldsymbol{\beta}'_{11}(\tau)\big)'$ and the corresponding IVX estimator with $\boldsymbol{\beta}^{ivx}(\tau)$. The asymptotic behaviour of the IVX estimator that corresponds to the linear predictive regression model under abstract degree of regressors persistence is studied by Kostakis et al. (2015) while the asymptotic properties of the IVX estimator for the quantile predictive regression model is studied by Lee (2016). Since when estimating the risk measure pair $(\mathsf{VaR}, \mathsf{CoVaR})$, we use the conditional quantile distribution, i.e., to capture the effect of $\boldsymbol{x}_{t-1}$ on the conditional quantile of $\boldsymbol{y}_t$, but with the additional assumption of possibly nonstationary regressors, then estimation relies on the nonstationary quantile predictive regression.

Second-Stage Procedure

During the second stage of our procedure we consider a consistent estimator of the parameter vector for the set of regressors which includes the nonstationary regressors as well as the generated regressor from the first stage procedure. More precisely, the generated regressor in our study corresponds to the fitted values of the predictive regression model based on the IVX estimator obtained in the first stage procedure.

$$y_{2,t} = \beta_{02}(\tau) + \boldsymbol{\beta}'_{12}(\tau)\boldsymbol{x}_{2,t-1} + \delta(\tau)\widehat{y}_{1,t}(\widehat{\boldsymbol{\beta}}^{ivx}(\tau)) + u_{2t}(\tau), \quad \text{for} \quad t = 1, ..., n \tag{4.50}$$

$$\boldsymbol{x}_{2,t} = \boldsymbol{R}_{2,n}\boldsymbol{x}_{2,t-1} + \boldsymbol{v}_{2,t}, \quad \boldsymbol{R}_{jn} = \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}_{jp}}{n^\gamma}\right), \ \gamma = 1. \tag{4.51}$$

where $\boldsymbol{C}_{jp} = \mathsf{diag}\left\{c_{j1}, ..., c_{jp}\right\}$ and $j \in \{1, 2\}$.

Define with $\widetilde{\boldsymbol{\beta}}(\tau) = \left(\beta_{02}(\tau), \boldsymbol{\beta}'_{12}(\tau), \delta(\tau)\right)'$ the parameter vector of the second stage procedure and the corresponding IVX estimator with $\widetilde{\boldsymbol{\beta}}^{ivx}(\tau)$. Specifically, we observe that this extended parameter vector includes both the nearly integrated regressors as well as the generated regressor from the first stage estimation step. In this case, we need to develop the asymptotic distribution theory for both the OLS and IVX estimators when the generated regressor is included in the set of regressors in the second stage predictive regression model that corresponds to the econometric specification of the CoVaR risk measure. A modified IVX estimator is necessary to be developed in order to account for the presence of the particular effect in the setting of the quantile predictive regression models. The nuisance coefficient of persistence is defined such that $c_i > 0$ and is either $\gamma = 1$, which corresponds to near unit root regressors or $\gamma \in (0, 1)$ that corresponds to mildly integrated regressors.

Therefore, the focus of this paper is the implementation of the above econometric environment and the development of the corresponding asymptotic theory in the case of the quantile predictive regression model. Consider the piecewise derivative of the loss function which is defined as below $\psi_\tau(u) = \tau - \mathbf{1}(u < 0)$. Then the innovation sequence of the quantile predictive regression model, $u_t(\tau) = u_t - F_u^{-1}(\tau)$, and $F_u^{-1}(\tau)$ is the unconditional τ-quantile of u_t, where $\tau \in (0, 1)$ is a fixed quantile (see, Lee (2016) and Fan and Lee (2019)). Thus, the $\boldsymbol{\beta}^{ivx}$ estimator is obtained from the first stage procedure using the IVX estimation method[14] for the model coefficients based on the nonstationary regressors which implies that the generated regressor is defined

$$\widehat{y}_{1,t}\left(\widehat{\boldsymbol{\beta}}^{ivx}(\tau)\right) = \boldsymbol{x}'_{1,t-1}\widehat{\boldsymbol{\beta}}^{ivx}(\tau) \tag{4.52}$$

Thus to conduct inference we need to obtain a consistent estimator for the covariance matrix of $\tilde{\boldsymbol{\beta}}_2^{ivx}$ using the usual "sandwich" formula given below (see, Demetrescu and Rodrigues (2020)).

[14] The reason that we apply the IVX instrumentation procedure only to the estimate from the first stage quantile predictive regression model rather to the corresponding generated covariate is to ensure that those fitted values can preserve their definition as the estimated Value-at-Risk but in our case adjusted based on the presence of persistent predictors and corrected accordingly using the IVX methodology.

4.4 Statistical Inference and Asymptotic Theory

4.4.1 Preliminary Setting

Consider that the underline stochastic processes corresponding to the data generating mechanism of our complex tail network-driven system corresponds to a commonly defined probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a corresponding $\sigma-$algebra and $\mathcal{F}_t$ filtration. Due to the complex nature of our tail dependency driven system, we consider a statistical modelling methodology which is based on the "divide-and-conquer" principle. More precisely, the proposed hypothesis testing procedure is divided into several stages when testing for systemic risk effects to accommodate the complex estimation methodology. Therefore, based on the system displayed in Figure 4.1, the null hypothesis of interest is formulated as below

$$\mathbb{H}_0 : \left\{ \delta_{(1)}(\tau) = \delta_{(2)}(\tau) = \ \dots \ = \delta_{(j)}(\tau) \ \dots \ = \delta_{(m-1)}(\tau) = 0 \ | \ g \neq g' \right\}, \tag{4.53}$$

which implies that we are testing for the null hypothesis that the systemic risk coefficients of the risk matrix in the upper off-diagonal are all zero. Then, the alternative hypothesis is formulated as

$$\mathbb{H}_1 : \left\{ \delta_{(1)}(\tau) \neq \delta_{(2)}(\tau) \neq \ \dots \ \neq \delta_{(j)}(\tau) \ \dots \ \neq \delta_{(m-1)}(\tau) \neq 0 \ | \ g \neq g' \right\} \tag{4.54}$$

Under the alternative hypothesis at least one of the δ coefficients are different than zero. Due to the unique way the $\mathsf{VaR} - \mathsf{CoVaR}$ risk matrix is constructed further interesting statistical inference procedures can be implemented in order to accommodate testing for the corresponding lower off-diagonal elements of the risk matrix having all zero systemic risk coefficients. Nevertheless, we focus on the testing hypotheses as formulated above.

Our proposed testing methodology which focus on linear restrictions specifically for the systemic risk coefficients can be also modified with further adding up restrictions. A related study in the literature which considers such a testing formulation is proposed by Ravikumar et al. (2000). The use of adding-up restriction within our framework has specific interpretation in terms of systemic risk in the graph. Specifically, by employing a fixed weight vector, we can impose restrictions on the level of interconnectedness and systemic risk among the nodes of the graph. The proposed testing methodology is based on the IVX-Wald statistic which has been found to robustify inference in predictive regression models under abstract degree of persistence. Our strategy for the development of the asymptotic theory is based on first deriving the limiting distribution of the conventional IVX-Wald statistic for linear restrictions (see, Kostakis et al. (2015)). Furthermore, our complex system has some specific features we have to tackle such as the fact that we propose a double IVX instrumentation methodology. Therefore, to obtain the limiting distribution for testing the null hypothesis of no systemic risk we have to establish the asymptotic behaviour of the double IVX estimator, that is, a mixed Gaussian distribution regardless of the persistence properties of regressors.

Furthermore, deriving the limit theory for the corresponding system IVX-Wald test such that a weakly convergence to a Chi-square limiting distribution holds, permits to conduct robust inference on the parameters of interest. A reasonable research question is whether to consider all the cases of different degree of persistence for stage one nonstationary regressors versus stage two nonstationary regressors. We begin our analysis under the assumption that the nonstationary regressors of stage one are near unit root and similarly the nonstationary regressors of stage two (although not from the exact same process, due to the proposed data generating mechanism), to simplify the asymptotic theory. As a future research we can also consider the case in which the nonstationary regressors across the two procedures have also different degree of persistence. However, before doing that we concentrate on the correct construction of the statistical procedure for testing for systemic risk effects across the nodes of the graph. Therefore in order to construct the testing hypothesis we consider the following formulations of the individual specific equations that represent the quantile predictive regression models with nonstationary regressors. In particular, we have $g \in \{1, ..., m\}$ system equations. More specifically, Figure 4.3 below demonstrates the complexity in estimating the particular large parameter space using a representation of the different system equations across the different values of g, although we do not consider a high dimensional setting for which the number of parameters is larger than the time series observations.

$$
\begin{aligned}
y_{(1)t} &= \boldsymbol{\theta}'_{(1)}(\tau)\widetilde{\boldsymbol{X}}_{(1)t} + u_{(1)t} \equiv \boldsymbol{\beta}'_{(1)}(\tau)\boldsymbol{x}_{1,t-1} + \delta_{(1)}(\tau)\widehat{y}^{o}_{2,t} + u_{1,t} \\
\boldsymbol{x}_{1,t} &= \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d}{n^{\gamma}}\right)\boldsymbol{x}_{1,t-1} + \boldsymbol{v}_{1,t} \\
\\
y_{(2)t} &= \boldsymbol{\theta}'_{(2)}(\tau)\widetilde{\boldsymbol{X}}_{(2)t} + u_{2,t} \equiv \boldsymbol{\beta}'_{(2)}(\tau)\boldsymbol{x}_{2,t-1} + \delta_{(2)}(\tau)\widehat{y}^{o}_{3,t} + u_{2,t} \\
\boldsymbol{x}_{2,t} &= \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d}{n^{\gamma}}\right)\boldsymbol{x}_{2,t-1} + \boldsymbol{v}_{2,t} \\
\\
y_{(3)t} &= \boldsymbol{\theta}'_{(3)}(\tau)\widetilde{\boldsymbol{X}}_{(3)t} + u_{(3)t} \equiv \boldsymbol{\beta}'_{(3)}(\tau)\boldsymbol{x}_{3,t-1} + \delta_{(3)}(\tau)\widehat{y}^{o}_{4,t} + u_{3,t} \\
\boldsymbol{x}_{3,t} &= \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d}{n^{\gamma}}\right)\boldsymbol{x}_{3,t-1} + \boldsymbol{v}_{3,t} \\
\vdots &= \vdots \\
y_{(m-1)t} &= \boldsymbol{\theta}'_{(m-1)}(\tau)\widetilde{\boldsymbol{X}}_{(m-1)t} + u_{(m-1)t} \equiv \boldsymbol{\beta}'_{(m-1)}(\tau)\boldsymbol{x}_{m-1,t-1} + \delta_{(m-1)}(\tau)\widehat{y}^{o}_{m,t} + u_{m-1,t} \\
\boldsymbol{x}_{m-1,t} &= \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d}{n^{\gamma}}\right)\boldsymbol{x}_{m-1,t-1} + \boldsymbol{v}_{m-1,t}
\end{aligned}
$$

Figure 4.3: SUR system with quantile predictive regression models of upper diagonal of $\boldsymbol{S}_{ij}$

4.4.2 Wald tests on the Row and Column Spaces of the Risk matrix

In practice, based on the structure of the risk matrix, there are three different formulations of the testing hypotheses with respect to the systemic risk coefficients located in the diagonal above the main diagonal, the columns or the rows of the risk matrix $\boldsymbol{\Gamma}_{ij}$ proposed by Katsouris (2021). In addition, it is worth emphasizing that we operate under the assumption that the model size (number of regressors in each quantile predictive regression model) is fixed and relative small to the sample size. Extending our estimation and asymptotic analysis to allow for $q \to \infty$ is technically challenging, especially with persistent and endogenous regressors ($q \equiv d + 1$) and thus examining the performance of IVX-Wald test in a high-dimensional setting (e.g., see Gupta and Seo (2023)) is beyond the scope of our study. In that case one would be interested to compare the performance of the two test statistics under two different data structure, a multivariate predictive regression and a SUR representation. Nevertheless, we consider the derivation of Wald-type tests under quite general conditions and identifying restrictions that we explain in details below.

Hypothesis Testing Diagonally of the risk matrix

The following econometric specifications correspond to the demeaned versions of the original random variables and thus to avoid unnecessary complications with the notation we leave the current notation without further changes. We employ an alternative representation of the econometric specifications given by Figure 4.3 using matrix notation as displayed in Figure 4.4 below. Notice that due to the unique structure of the proposed risk matrix, with "diagonality-based" testing we main we consider the cases in which the joint estimation of the risk pair implies considering indices that lie on the diagonals of the risk matrix but not on the main diagonal which corresponds to all the estimated VaR risk measures of the cross-section. On the other hand, the flexibility of our system representation permits to consider various different statistical inference methodologies commonly used in the literature. Therefore, beyond the classical linear restrictions imposed by a Wald-test formulation which are examined in several studies in nonstationary time series econometrics (see, Kostakis et al. (2015)) we focus on constructing a suitable statistical testing approach for the significance of the systemic risk proxies, by leveraging the complexity of our risk matrix and the proposed SUR representation.

In this direction, a similar approach has been proposed by Xu and Guo (2022) who consider the special case when a multivariate regressand vector has all identical elements (see also Xu and Guo (2019)) and based on a relative large vector of predictors, a SUR system representation is employed for estimation and inference purposes (see, also). Therefore, the approach of Xu and Guo (2022) tackles the possible presence of high-dimensionality by assuming a large dimensional dependent variable based on the SUR representation which provides efficiency gains and convenient testing methods. Moreover, Chen et al. (2023) employ a SUR representation for a VAR(1) with explosive roots, thereby projecting the dimensions of the dependent vector into univariate nonstationary autoregressive processes. In contrast, our SUR representation implies projecting the m system equations into nonstationary quantile predictive regressions with univariate dependent variables but multivariate regressors.

Thus, we can compare the performance of the SUR-IVX-Wald test when estimating separately each individual equation of the system by fitting the quantile predictive regression model with nonstationary regressors with the performance of the IVX-Wald in a multivariate setting. In particular, based on our representation strategy, we expect the SUR-IVX-Wald test to perform better than the IVX-Wald counterpart in terms of empirical size. More precisely, to see this we can compare the performance of the two test statistics under two different data structures, a multivariate predictive regression and a SUR representation. Furthermore, when employing the IVX-Wald test in a high-dimensional setting, due to the large number of nonstationary regressors, the size distortions can be indeed severe. On the other hand, a SUR system representation provides a statistical mechanism for dimensionality reduction and therefore an IVX based Wald-test for the particular model structure is expected to have better performance under the null hypothesis. Similarly, one can consider the special case when such a multivariate regressand vector has all identical elements (as in Xu and Guo (2022)) and given a relative large vector of predictors, is formulated as a SUR system by splitting the large regressor vector into blocks equal to the number of estimating equations. In other words, the proposed Wald-based test statistics correspond to testing the quality of the columns of the non-linear risk matrix constructed using systems of predictive regressions.

In this case, since the presence of high-dimensionality is contributed due to the large dimension of the dependent variable, then consequently applying such a transformation to the data structure can clearly provide similar benefits as when considering that the $m-$dimensional vector $\boldsymbol{y}_t$ is decomposed such that each vector element corresponds to one of the system equations. Then the pooled estimator (i.e., SUR-IVX estimator) vis-a-vis the equation-specific estimators can be compared in terms of their asymptotic properties and finite-sample properties especially when these are used in the formulation of Wald-type test statistics with relevant restrictions. Thus, our proposed testing methodology that has been developed specifically for the accompanied framework of our complex tail dependency driven system, when reduced to an estimation and testing problem of lower dimensional space, then the resulting testing procedure can provide significant benefits in terms of empirical size and power performance under the null hypothesis. In matrix notation, the seemingly unrelated system of nonstationary quantile predictive regressions can be expressed as below

$$\boldsymbol{Y}_t = \boldsymbol{X}^{\star}_{t-1}\boldsymbol{\theta}^{\star} + \boldsymbol{U}_t, \quad \text{for} \ \ t = 1, ..., n \tag{4.56}$$

where the IVX instruments are given by $\boldsymbol{Z}_{t-1} = \mathsf{diag}\left\{ \boldsymbol{z}'_{1,t-1}, ..., \boldsymbol{z}'_{m-1,t-1} \right\}$ where $\boldsymbol{z}_{g,t-1}$ is of dimension $(d+1) \times 1$ for all $g \in \{1, ..., m-1\}$ since one of the system equations or one element of the $m-$dimensional dependent vector is not included. Therefore, in practise these individual equation specific instruments, $\boldsymbol{z}_{g,t-1}$, correspond to the IVX instruments for the d nonstationary regressors of the $m-$th equation and the self-instrumentation of the proxy systemic risk regressor, i.e., generated covariate (via the two-stage estimation procedure).

Using a matrix form representation we have that, for $d < m$:

$$\underbrace{\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \cdots \\ y_{m-1,t} \end{bmatrix}'}_{(m-1)\times 1} = \underbrace{\begin{bmatrix} [\boldsymbol{x}_{1,t-1} \ \widehat{y}^{o}_{2,t}]_{1\times(d+1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [\boldsymbol{x}_{2,t-1} \ \widehat{y}^{o}_{3,t}]_{1\times(d+1)} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & [\boldsymbol{x}_{m,t-1} \ \widehat{y}^{o}_{m,t}]_{1\times(d+1)} \end{bmatrix}}_{(m-1)\times(d+1)} \underbrace{\begin{bmatrix} [\boldsymbol{\beta}'_{(1)}(\tau) \ \delta_{(1)}(\tau)]' \\ [\boldsymbol{\beta}'_{(2)}(\tau) \ \delta_{(2)}(\tau)]' \\ \vdots \\ [\boldsymbol{\beta}'_{(m-1)}(\tau) \ \delta_{(m-1)}(\tau)]' \end{bmatrix}}_{(d+1)\times(m-1)} + \underbrace{\begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{m-1,t} \end{bmatrix}}_{(m-1)\times 1}$$

Figure 4.4: SUR system matrix formulation

Denote with

$$\underbrace{\boldsymbol{\theta}^{\star}}_{(d+1)\times(m-1)} := \Big[\boldsymbol{\theta}^{\star}_{1}(\tau), ..., \boldsymbol{\theta}^{\star}_{m-1}(\tau)\Big] = \Big[[\boldsymbol{\beta}'_{(1)}(\tau) \ \delta_{(1)}(\tau)], \dots, [\boldsymbol{\beta}'_{(m-1)}(\tau) \ \delta_{(m-1)}(\tau)]\Big]' \tag{4.58}$$

the $(m-1)-$dimensional parameter vector, which contains as elements the parameter vector of each separate econometric specification with $q = (d+1)$ model parameters. Furthermore, we denote with

$$\underbrace{\boldsymbol{X}^{\star}_{t-1}}_{(m-1)\times(d+1)} := \mathsf{diag}\Big\{[\boldsymbol{x}_{1,t-1} \ \widehat{y}^{o}_{2,t}]_{1\times(d+1)}, ..., [\boldsymbol{x}_{m-1,t-1} \ \widehat{y}^{o}_{m,t}]_{1\times(d+1)}\Big\} \tag{4.59}$$

and with $\boldsymbol{X}^{\star}$ the block diagonal matrix of regressors and with $\boldsymbol{u}^{\star} = \Big[\boldsymbol{u}_{1,t}, \dots, \boldsymbol{u}_{m-1,t}\Big]'$ the corresponding $(m-1)-$dimensional vector with the error terms corresponding to the $(m-1)$ predictive regressions of the system.

Linear Predictive Regression Model

We begin our analysis by considering the estimation methodology for the linear predictive regression model by considering the formulations of model estimators as well as their corresponding covariance matrices. We study the corresponding estimation and inference procedure when the system equations represent nonstationary quantile predictive regression models in the next section. Within our estimation framework we consider that $m < n$, that is, the number of cross-sectional units, is less than the number of time series observations (e.g., $m = 20$ and $n = 150$). The possible presence of high dimensionality[15] is not explicitly examined in this chapter, which might require modifications in our estimation methodology. Nevertheless, the parameter space is still quite large due to the presence of different nonstationary regressors for each system equation as well as the systemic risk proxy covariates. Moreover, to robustify against unknown persistence forms our system estimator is based on the IVX filtration of Phillips and Magdalinos (2009) which it has been proved to be robust to the abstract degree of persistence; in our case to filter out the persistence of the nonstationary regressors across the equations of the system.

Therefore, the SUR-IVX-diag estimator for the model parameter $\boldsymbol{\theta}^{\star}$ is estimated as below

$$\widehat{\boldsymbol{\theta}}^{\star}_{ivx} = \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \boldsymbol{X}^{\star}_{t-1} \right)^{-1} \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \boldsymbol{Y}_{t} \right), \tag{4.60}$$

where $\widehat{\boldsymbol{\theta}}^{\star}_{ivx} = \left(\widehat{\boldsymbol{\theta}}'_{ivx,1}, ..., \widehat{\boldsymbol{\theta}}'_{ivx,m-1} \right)'$ such that $\widehat{\boldsymbol{\theta}}_{g} = \left(\widehat{\boldsymbol{\beta}}'_{g}, \widehat{\delta}_{g} \right)'$ for all $g \in \{1, ..., m-1\}$. To construct the Wald-type test statistic we denote with $\widehat{\boldsymbol{U}}_{t} = \left(\widehat{u}_{1,t}, ..., \widehat{u}_{m-1,t} \right)'$ to be the OLS residual obtained from the multivariate model (4.56). Then, the $(m-1) \times (m-1)$ covariance matrix is expressed as

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} = \frac{1}{n-1} \sum_{t=1}^{n} \widehat{\boldsymbol{U}}_{t} \widehat{\boldsymbol{U}}'_{t}. \tag{4.61}$$

Proposition 4.1. Suppose that the conditions of Assumption **??** hold. Then, it follows that

$$n^{\frac{(1+\gamma_x)}{2}} \left(\widehat{\boldsymbol{\theta}}^{\star}_{ivx} - \boldsymbol{\theta}^{\star} \right) \overset{d}{\to} \mathcal{MN}\big(\mathbf{0}, \left(\boldsymbol{\Psi}_{C}^{-1} \right)' \boldsymbol{V}_{zz} \boldsymbol{\Psi}_{C}^{-1} \otimes \boldsymbol{\Sigma}\big), \quad \text{as} \quad n \to \infty. \tag{4.62}$$

where all system equations have near-unit root regressors not necessarily identical.

Notice that regardless of the fact that the nonstationary regressors across the equations not all belong to the same persistence regime we only consider nearly stationary and nearly nonstationary cases and exclude cases such as explosive unit roots either common or distinct (as in **?** and Chen et al. (2023)).

[15] The aspect of estimation for high-dimensional SUR models within a stationary time series environment is examined in the framework proposed by Tan et al. (2021).

Theorem 4.1. Suppose that the conditions of Assumption 1-4 hold, with $\boldsymbol{w}_{(g)t} = \left(u_{(g)t}, \boldsymbol{v}'_{(g)t}\right)'$. Define the following IVX-Wald statistic with an estimated covariance matrix

$$\mathcal{W}^{\star}_{sur-ivx} = \left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}^{\star}_{ivx} - \boldsymbol{r}\right)' \widehat{\boldsymbol{Q}}_{\boldsymbol{R}}^{-1} \left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}^{\star}_{ivx} - \boldsymbol{r}\right) \tag{4.63}$$

where $\widehat{\boldsymbol{Q}}_{\boldsymbol{R}}$ is the variance matrix estimator. Moreover, the test statistic can be expressed as below

$$\mathcal{W}^{\star}_{sur-ivx} \equiv \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1}\boldsymbol{U}_t\right)' \widehat{\mathbb{M}}^{-1} \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1}\boldsymbol{U}_t\right), \tag{4.64}$$

where $\widehat{\mathbb{M}}$ is defined as below

$$\widehat{\mathbb{M}} := \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} \odot \sum_{t=1}^{n} \boldsymbol{Z}_{g,t-1}\boldsymbol{Z}'_{g,t-1} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U},11} \sum_{t=1}^{n} \boldsymbol{z}_{1,t-1}\boldsymbol{z}'_{1,t-1} & \dots & \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U},1K} \sum_{t=1}^{n} \boldsymbol{z}_{1,t-1}\boldsymbol{z}'_{K,t-1} \\ \vdots & \ddots & \vdots \\ \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U},K1} \sum_{t=1}^{n} \boldsymbol{z}_{K,t-1}\boldsymbol{z}'_{1,t-1} & \dots & \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U},KK} \sum_{t=1}^{n} \boldsymbol{z}_{K,t-1}\boldsymbol{z}'_{K,t-1} \end{pmatrix}_{K\times K} \tag{4.65}$$

where $K \equiv m-1$. Then, under $\mathbb{H}_0 : \widehat{\boldsymbol{\theta}}^{\star}_{ivx} = \boldsymbol{0}$, it follows that $\mathcal{W}^{\star}_{ivx} \overset{d}{\to} \chi^2_{m-1}$ where the degrees of freedom equal to the number of system equations.

Furthermore, under the null hypothesis the $\mathcal{W}^{\star}_{ivx}$ statistic takes the following form

$$\mathcal{W}^{\star}_{ivx} = \left(\sum_{t=1}^{n} \widetilde{\boldsymbol{z}}_{t-1}\widetilde{u}_t\right)' \widehat{\mathbb{M}}^{-1} \left(\sum_{t=1}^{n} \widetilde{\boldsymbol{z}}_{t-1}\widetilde{u}_t\right) \tag{4.66}$$

The particular formulation can be compared with the standard IVX-Wald statistic such that

$$\mathcal{W}_{ivx} = \left(\sum_{t=1}^{n} \boldsymbol{z}_{t-1}u_t\right)' \left(\widehat{\sigma}^2_u \sum_{t=1}^{n} \boldsymbol{z}_{t-1}\boldsymbol{z}'_{t-1}\right)^{-1} \left(\sum_{t=1}^{n} \boldsymbol{z}_{t-1}u_t\right)', \tag{4.67}$$

An important conjecture is that in the special case when all regressors for the equations are the same and the system responses, which are constructed using the data generating mechanism provided by Assumption 4.2, are replaced by K replicates of the same vector (e.g., one of the response vector of the nodes), our procedure simplifies to the estimation procedure proposed by Xu and Guo (2019). In this special case, then it can be shown that $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} \overset{p}{\to} \sigma^2_u \boldsymbol{1}\boldsymbol{1}'$, where $\boldsymbol{1} = (1, ..., 1)'$ is a $(m \times 1)$ unit vector. Then, the test statistic $\mathcal{W}^{\star}_{sur-ivx}$ simply replaces $\widehat{\sigma}^2_u$ in the Wald statistic $\mathcal{W}_{ivx}$ (expression (19) in KMS) by estimators of σ^2_u with zero restrictions partially imposed, such that all the coefficients of $\boldsymbol{x}_{t-1}$ are zero expect the regressors that correspond to the $g-$equation, $\boldsymbol{x}_{g,t-1}$. Therefore, within our framework we will need to verify whether this convergence in probability, such that $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} \overset{p}{\to} \sigma^2_u \boldsymbol{1}\boldsymbol{1}'$ due to the fact that each system equation has a different Gaussian process.

Remark 4.8. Our procedure encompasses the testing methodology proposed by Xu and Guo (2019) since one can apply their dimensionality reduction method for each individual specific system equation. We leave the illustration of the particular application as future research. Our proposed estimation and testing methodology should be not viewed as an extension of the particular framework, when one employs the conditional functional form instead of the linear specification form. The motivation in the current study is driven from the perspective of systemic risk modelling and examines the SUR modelling methodology for the proposed complex tail dependency system of our framework. Furthermore, our framework considers a specific data generating mechanism for a complex system based on an underline graph structure, which in other words proposes a more general form of dependence.

However, in our case we need to evaluate whether the convergebce in probability result $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} \overset{p}{\to} \sigma_0^2 \mathbf{1}\mathbf{1}'$ holds. This is a crucial assumption especially due to the dependence structure we impose that allows to generate independent innovation sequences for each node specific predictive regression model. More specifically, due to the fact that each system equation is generated from a different Gaussian process then possibly a different variance for the residual terms applies. Therefore, in order to facilitate the development of the asymptotic theory within our proposed framework we can impose an assumption of conditional homoscedasticity across equations.

Suppose that we are interested in testing m_r linear restrictions such that $m_r \leq m$, where

$$\mathbb{H}_0^{\star} : \underbrace{\boldsymbol{R}}_{m_r \times m} \delta = \underbrace{\boldsymbol{r}}_{m_r \times 1} \tag{4.68}$$

where $\boldsymbol{R}$ and $\boldsymbol{r}$ contain known constants, and $\boldsymbol{R}$ has full rank. Then, the particular selector matrix can be employed to test linear restrictions as well as parameter-specific adding-up restrictions across the equations of the system.

Notice that in practise in the framework of Xu and Guo (2019) the authors keep fix a subset of nonstationary regressors with a small cardinality and then the remaining subset which includes a high dimensional vector of regressors, say K they divide this vector into equal smaller parts. Then, these K linear projections on the regressand with regressors the fix subset plus another set of regressors that corresponds to a partition from the large section of nonstationary regressors. The specific approach can be also found in the statisics literature as subset selection in high dimensional settings. However, the main difference in our framework is that we use a similar modelling methodology to estimate the SUR system with quantile predictive regressions.

4.4.3 System Specific Estimation Methodology

We focus on the development of the asymptotic theory for the case when we are testing for the equality of the systemic risk coefficients within each column of the risk matrix. To avoid confusion with the notation we assume that we have $(m-1)$ units and n time series observations such as $t = 1, ..., n$. Therefore, we are constructing the testing procedure for the $s-$th column of the matrix after removing the element of the $s-$th column which corresponds to the diagonal of the risk matrix, i.e., the VaR risk measure such that $j \neq i$ where the indices $(i, j, s) \in \{1, ..., m\}$.

A matrix representation for the first $s-$th column of the matrix is:

$$\begin{bmatrix} \boldsymbol{y}_{(1)t} \\ \boldsymbol{y}_{(2)t} \\ \vdots \\ \boldsymbol{y}_{(m-1)t} \end{bmatrix} = \begin{bmatrix} [\boldsymbol{x}_{t-1} \ \boldsymbol{y}_{(s)t}] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [\boldsymbol{x}_{t-1} \ \boldsymbol{y}_{(s)t}] & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & [\boldsymbol{x}_{t-1} \ \boldsymbol{y}_{(s)t}]' \end{bmatrix} \begin{bmatrix} [\boldsymbol{\beta}'_{(1)}(\tau) \ \delta_{(1)}(\tau)]' \\ [\boldsymbol{\beta}'_{(2)}(\tau) \ \delta_{(2)}(\tau)] \\ \vdots \\ [\boldsymbol{\beta}'_{(m-1)}(\tau) \ \delta_{(m-1)}(\tau)]' \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_{(1)t}(\tau) \\ \boldsymbol{u}_{(2)t}(\tau) \\ \vdots \\ \boldsymbol{u}_{(m-1)t}(\tau) \end{bmatrix}.$$

In practise, we have $(m-1)$ quantile predictive regressions with the same covariates, i.e., the set of lagged regressors $\boldsymbol{x}_{t-1}$ plus the same systemic risk covariate, $\boldsymbol{y}_{(s)t}$. Furthermore, the lagged regressors are assumed to follow an AR(1) model with a LUR coefficient and a common coefficient of persistence c_m for all $m \in \{1, ..., k\}$ where k is the number of regressors in each equation such that $k < m$. Equivalently, the $j-$th equation of the above SUR system corresponding to the $s-$th column of the risk matrix, such that $s = \{1, ..., m\}$, can be expressed using the following more compact form

$$\boldsymbol{y}_{(j\setminus\{s\})t} = \boldsymbol{\theta}'_{(j)}(\tau)\widetilde{\boldsymbol{X}}_j^{(s)} + \boldsymbol{u}_{(j)t}^{(s)}(\tau) \tag{4.69}$$

$$\boldsymbol{x}_t = \left(\boldsymbol{I}_k - \frac{\boldsymbol{C}_k}{n^{\gamma_x}}\right)\boldsymbol{x}_{t-1} + \boldsymbol{v}_t \tag{4.70}$$

$\boldsymbol{\theta}_{(j)}(\tau) := [\boldsymbol{\beta}'_{(j)}(\tau) \ \delta_{(j)}(\tau)]'$ and $\widetilde{\boldsymbol{X}}_j^{(s)} := [\boldsymbol{x}_{t-1} \ \boldsymbol{y}_{(s)t}]$ for all $j \in \{1, ..., m\} \setminus \{s\}$ with $s \in \{1, ..., m\}$.

In particular, due to the fact that each of the system equations has a specific generated innovation vector with first and second moments following the general principles of predictive regression models, this property ensures that we can work under the assumption of conditional homoscedasticity. Although extending to the conditional heteroscedasticity case with respect to imposing for example a GARCH structure on the innovation vector of each system's equation. Based on the dependence structure given by Assumption 4.1 which implies a block structure for the covariance matrices of the system, the zero conditional expectation (and second moments) across the cross-section is preserved (i.e., for $i \neq j$) since all those terms are equivalent with a zero $(m \times m)$ matrix.

Equivalently, formulating the Matrix From 4.4.3 in a vector form we obtain

$$\boldsymbol{Y}_{(j\setminus\{s\})t} = \mathbb{X}^{(s)}\boldsymbol{\beta}^{(s)}_{(j)}(\tau) + \left(\boldsymbol{I}_{m-1} \otimes \boldsymbol{y}_{(s)t}\right)\boldsymbol{\delta}^{(s)}(\tau) + \boldsymbol{U}^{(s)}_{(j)t}(\tau) \tag{4.71}$$

$$\boldsymbol{x}_t = \left(\boldsymbol{I}_k - \frac{\boldsymbol{C}_k}{n^{\gamma_x}}\right)\boldsymbol{x}_{t-1} + \boldsymbol{v}_t, \tag{4.72}$$

where

$$\boldsymbol{Y}_{(j\setminus\{s\})t} := \begin{bmatrix} \boldsymbol{y}_{(1)t} \\ \boldsymbol{y}_{(2)t} \\ \vdots \\ \boldsymbol{y}_{(m-1)t} \end{bmatrix}, \quad \mathbb{X}^{(s)} := \text{diag}\begin{bmatrix} \boldsymbol{X}^{(s)}_1 \\ \boldsymbol{X}^{(s)}_2 \\ \vdots \\ \boldsymbol{X}^{(s)}_{m-1} \end{bmatrix}', \quad \boldsymbol{U}^{(s)}_{(j)t}(\tau) := \begin{bmatrix} \boldsymbol{u}_{(1)t}(\tau) \\ \boldsymbol{u}_{(2)t}(\tau) \\ \vdots \\ \boldsymbol{u}_{(m-1)t}(\tau) \end{bmatrix}, \tag{4.73}$$

where $\widetilde{\boldsymbol{X}}^{(s)}_j = [\boldsymbol{x}_{t-1} \ y_{(s)t}]$ such that $\boldsymbol{X}^{(s)}_j \equiv \boldsymbol{x}_{t-1}$ for all $j \in \{1, ..., m\}$. Then, the model parameters are given by the vectors $\boldsymbol{\beta}^{(s)}_{(j)}(\tau) = \left(\beta^{(s)}_{(1)}(\tau), ..., \beta^{(s)}_{(m-1)}(\tau)\right)'$ and $\boldsymbol{\delta}^{(s)(\tau)} = \left(\delta^{(s)}_1(\tau), ..., \delta^{(s)}_{m-1}(\tau)\right)'$. Equivalently, we have the following SUR system:

$$\begin{bmatrix} \boldsymbol{y}_{(1)t} \\ \boldsymbol{y}_{(2)t} \\ \vdots \\ \boldsymbol{y}_{(m-1)t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_{t-1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{x}_{t-1} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{x}_{t-1} \end{bmatrix} \begin{bmatrix} \beta^{(s)\prime}_{(1)}(\tau) \\ \beta^{(s)\prime}_{(2)}(\tau) \\ \vdots \\ \beta^{(s)\prime}_{(m-1)}(\tau) \end{bmatrix} + \begin{bmatrix} \boldsymbol{y}_{(s)t} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{y}_{(s)t} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{y}_{(s)t} \end{bmatrix} \begin{bmatrix} \delta^{(s)}_1(\tau) \\ \delta^{(s)}_2(\tau) \\ \vdots \\ \delta^{(s)}_{m-1}(\tau) \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_{(1)t}(\tau) \\ \boldsymbol{u}_{(2)t}(\tau) \\ \vdots \\ \boldsymbol{u}_{(m-1)t}(\tau) \end{bmatrix}$$

Furthermore, all lagged regressors included in the individual equations, that is, the quantile predictive regressions, are identical and follow the LUR process as below:

$$\boldsymbol{x}_t = \left(\boldsymbol{I}_k - \frac{\boldsymbol{C}_k}{n^{\gamma_x}}\right)\boldsymbol{x}_{t-1} + \boldsymbol{v}_t, \tag{4.74}$$

Therefore, the testing hypothesis of interest is

$$\mathbb{H}_0 : \boldsymbol{\delta}^{(s)}(\tau) = \boldsymbol{0} \Rightarrow \delta^{(s)}_1(\tau) = ... = \delta^{(s)}_{m-1}(\tau) = 0. \tag{4.75}$$

i.e., the equality of the elements of the systemic risk coefficient of the $s-$th column of the risk matrix. Therefore, the above null hypothesis is tested using a Wald type statistic. In other words under the null hypothesis, the true population SUR system does not include the systemic risk covariate, i.e., $\boldsymbol{\delta}^{(s)}(\tau) = 0$, while under the alternative hypothesis we have the unrestricted model which implies that at least some of those systemic risk coefficients are present in the SUR system.

4.4.4 Parameter Estimation and Robust Wald Test

In order to estimate the econometric model within our setting we need to impose some regularity conditions and assumptions which permit to obtain the asymptotic behaviour of the system estimators as well as the proposed Wald statistic[16]. We begin by reviewing again the parameter estimation procedure. Denote with $\boldsymbol{\beta}^{\star}(\tau)$ the parameter vector of interest. Then, the estimator $\hat{\boldsymbol{\beta}}^{\star}(\tau)$ is obtained via the following expression:

$$\hat{\boldsymbol{\beta}}^{\star}(\tau) = \underset{\boldsymbol{\beta}^{\star} \in \mathbb{R}^{(m-1)\times(k+1)}}{\text{arg min}} \sum_{i=1}^{m} \sum_{t=1}^{n} \rho_{\tau} \left(y_{it} - \boldsymbol{X}^{\star\prime}_{(i)t-1} \boldsymbol{\beta}^{\star}(\tau) \right) \tag{4.76}$$

where $\boldsymbol{X}^{\star}_{(i)t-1} = \left[\boldsymbol{x}_{(i)t-1}, y_{(j)t} \right]'$ for all $j \neq i$.

The regressor matrix contains lagged regressors assumed to be generated as local-unit-root processes, representing the endogenous variables of the system plus the systemic risk covariate which is regression specific and time-invariant. There are several points to discuss here. First, the particular aspect when one of the endogenous covariates is chosen as one of the dependent while the rest are treated as if were independent is discussed by Balestra and Nerlove (1966) in the case of cross-sectional time series regressios models. However, out modelling approach considers certain features, such as the persistence properties of regressors as well as a possible graph dependence among pairs of nodes. Moreover, since the LUR specification incorporates an unknown abstract coefficient of persistence, we employ the IVX instrumentation methodology proposed by Phillips and Magdalinos (2009) that filters out and controls the degree of persistence. Next, in terms of graph dependence we impose an assumption that relates the standard covariance structure of the quantile predictive regression model with the autoregression specification in terms of the system.

Furthermore, in order to facilitate the development of the asymptotic theory for model estimators and associated test statistics we define the following matrices for some $\tau \in (0, 1)$.

$$\widehat{\boldsymbol{D}}_0 = \frac{1}{mn} \sum_{i=1}^{m} \sum_{t=1}^{n} \boldsymbol{X}^{\star}_{(i)t-1} \boldsymbol{X}^{\star\prime}_{(i)t-1}, \ \widehat{\boldsymbol{D}}_1(\tau) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{t=1}^{n} f_{it} \left\{ \boldsymbol{X}^{\star\prime}_{(i)t-1} \boldsymbol{\beta}^{\star}(\tau) \right\} \boldsymbol{X}^{\star}_{(i)t-1} \boldsymbol{X}^{\star\prime}_{(i)t-1}$$

where the sample moment matrices $\widehat{\boldsymbol{D}}_0$ and $\widehat{\boldsymbol{D}}_1(\tau)$ are well-defined and nonsingular.

The null hypothesis is now constructed using the selector matrix and considering linear restrictions to specific parameters from each block, such as the systemic risk covariate as well as adding-up restrictions on the systemic risk covariates as mentioned before.

$$\mathbb{H}_0 : \boldsymbol{R}\boldsymbol{\beta}^{\star}(\tau) = \boldsymbol{q} \quad \text{versus} \quad \boldsymbol{R}\boldsymbol{\beta}^{\star}(\tau) \neq \boldsymbol{q} \tag{4.77}$$

for any $\tau \in (0, 1)$, where $\boldsymbol{R} = (\mathbf{0}, ..., \mathbf{0}, \mathbf{1})$ and $q = (0, ..., 0, 1)'$ considers the adding-up restriction of interest.

[16] Notice that the propose testing methodology specifically for the proposed structural model is a novel aspect in the literature. For instance, in the framework of Zhu et al. (2019) the authors do not consider testing for the joint statistical significance of model estimates via a Wald type statistic but only consider a general measure of goodness-of-fit for the model.

Therefore, the particular null hypothesis considers the restriction on the systemic risk covariates which belong to the different blocks of the multivariate model specification. Next, we proceed on constructing the Wald test and study the asymptotic behaviour of the test statistic expressed as below:

$$\mathcal{W}_{SUR-IVX}(\tau) = \left(\boldsymbol{R}\hat{\beta}^{\star}_{SUR-IVX}(\tau) - \boldsymbol{q}\right)' \boldsymbol{\Omega}_{\boldsymbol{R}}^{-1} \left(\boldsymbol{R}\hat{\beta}^{\star}_{SUR-IVX}(\tau) - \boldsymbol{q}\right) \tag{4.78}$$

where $\boldsymbol{\Omega}_{\boldsymbol{R}}$ the covariance matrix of the Wald test based on the SUR-IVX estimator.

Theorem 4.2. Consider the predictive regression model given by expressions (4.11)-(4.12) and that the conditions of Assumption (3.2) hold. Then, the Wald test for testing the null hypothesis (4.77) has the following limiting distribution

$$\mathcal{W}_{SUR-IVX}(\tau) \Rightarrow \chi^2_q \quad \text{for some } \tau \in (0,1) \ \text{ as } n \to \infty. \tag{4.79}$$

4.4.5 Econometric Model and SUR system Representation

Consider the linear predictive regression model as below

$$\widetilde{y}_t = \mu + \underbrace{\widetilde{\boldsymbol{x}}'_{t-1}}_{1\times D} \boldsymbol{\vartheta} + \widetilde{u}_t, \quad 1 \leq t \leq n \tag{4.80}$$

$$\boldsymbol{x}_t = \boldsymbol{R}_n \boldsymbol{x}_{t-1} + \boldsymbol{v}_t \tag{4.81}$$

where $\widetilde{\boldsymbol{x}}_t \in \mathbb{R}^d$ is a $d-$dimensional vector and $\boldsymbol{R}_n = \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d}{n^{\gamma}}\right)$ with $\gamma = 1$.

Before proceeding with the estimation methodology the first step is to consider a standard demeaning transformation of the original predictive regression that yields exact invariance of estimation of the parameter vector $\boldsymbol{\beta}$ to the presence of a model intercept. Therefore, the demeaned random variables are denoted as

$$y_t = \widetilde{y}_t - \frac{1}{n}\sum_{t=1}^{n} \widetilde{y}_t, \qquad x_t = \widetilde{x}_t - \frac{1}{n}\sum_{t=1}^{n} \widetilde{x}_t, \qquad u_t = \widetilde{u}_t - \frac{1}{n}\sum_{t=1}^{n} \widetilde{u}_t, \tag{4.82}$$

Therefore, we obtain the transformed predictive regression as below

$$y_t = \underbrace{\boldsymbol{x}'_{t-1}}_{1\times d} \boldsymbol{\vartheta} + u_t, \quad 1 \leq t \leq n \tag{4.83}$$

The second step is to construct the IVX instruments based on the undemeand regressors of the model such that

$$\boldsymbol{z}_t = \sum_{j=0}^{t-1} \boldsymbol{R}_z^j \Delta \boldsymbol{x}_{t-j}, \quad \text{for } \ t = 1, ..., n \tag{4.84}$$

Then the IVX estimator is given by

$$\widehat{\boldsymbol{\vartheta}}^{ivx} = \left(\sum_{t=1}^{n} \boldsymbol{z}_{t-1}\boldsymbol{x}'_{t-1}\right)^{-1} \left(\sum_{t=1}^{n} \boldsymbol{z}_{t-1}\boldsymbol{y}_t\right), \tag{4.85}$$

Parsimonious System

Construct a system of K linear projections, each of which is onto only one group of regressors of $\boldsymbol{x}_{t-1}$ such that $\boldsymbol{x}_{g,t-1}$, where $K \equiv N-1$. In particular, we have that

$$\widetilde{y}_{g,t} = \alpha_g + \widetilde{\boldsymbol{x}}'_{g,t-1} \underbrace{\boldsymbol{\beta}_g}_{d_g \times 1} + \widetilde{u}_{g,t}, \quad 1 \leq t \leq n \tag{4.86}$$

where $g = 1, ..., K$ where K are the number of system equations and d_g is the number of nonstationary regressors for each system equation, however we assume that $d_g \equiv d$ for all $g = 1, ..., K$. Therefore, the parameter of interest is given by $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{m-1})'$ which is of dimension $d \times 1$ and note that the total number of regressors for the system is given by $D = \sum_{g=1}^{K} d_g = (N-1)d$ as all system regressors have the same number of nonstationary regressors. Therefore, we are interested in estimating an SUR system which provides a parsimonious representation for inference purposes. In practise, there is a duality between testing the zero linear restrictions on the predictive regression model with the large regressor vector and testing for zero linear restrictions across all system specific predictive regressions. In other words, instead of testing whether under the null hypothesis $\boldsymbol{\vartheta} = \boldsymbol{0}$, we construct the testing hypothesis such that $\boldsymbol{\beta} = \boldsymbol{0}$. Therefore, next step is to consider the demeaned version of (4.87) such that

$$\boldsymbol{y}_{g,t} = \boldsymbol{x}'_{g,t-1} \underbrace{\boldsymbol{\beta}_g}_{d_g \times 1} + \boldsymbol{u}_{g,t}, \quad 1 \leq t \leq n \tag{4.87}$$

Then by converting the above model with a matrix form we obtain that

$$\boldsymbol{Y}_t = \boldsymbol{X}_{t-1}\boldsymbol{\theta} + \boldsymbol{U}_t \tag{4.88}$$

where we define the following matrices as below which are thought as three-dimensional bars projected to an (x, y) plan thus "cutting-out" the time-dimension such that

$$\underbrace{\boldsymbol{Y}_t}_{K \times 1} = \begin{bmatrix} \boldsymbol{y}_{1,t} \\ \boldsymbol{y}_{2,t} \\ \vdots \\ \boldsymbol{y}_{K,t} \end{bmatrix} \quad \underbrace{\boldsymbol{X}_{t-1}}_{K \times D} = \begin{bmatrix} \boldsymbol{x}'_{1,t} & & \\ & \ddots & \\ & & \boldsymbol{x}'_{K,t} \end{bmatrix}, \quad \underbrace{\boldsymbol{\theta}}_{K \times 1} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_K \end{bmatrix}, \quad \underbrace{\boldsymbol{U}_t}_{K \times 1} = \begin{bmatrix} \boldsymbol{u}_{1,t} \\ \vdots \\ \boldsymbol{u}_{K,t} \end{bmatrix}. \tag{4.89}$$

Furthermore, we consider the block diagonal matrix with the IVX instruments that correspond to the nonstationary regressors of our complex system where $\boldsymbol{Z}_{g,t} = (\boldsymbol{z}^{o\prime}_{g,t}, \boldsymbol{z}'_{g,t})'$, and

$$\underbrace{\boldsymbol{Z}_t}_{K \times D} = \begin{bmatrix} \boldsymbol{Z}'_{1,t} & & \\ & \ddots & \\ & & \boldsymbol{Z}'_{K,t} \end{bmatrix}. \tag{4.90}$$

Therefore, we can obtain the IVX estimator for the system that corresponds to $\boldsymbol{\theta}$ such that

$$\widehat{\boldsymbol{\theta}}_{ivx} = \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1}\boldsymbol{X}_{t-1} \right)^{-1} \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1}\boldsymbol{Y}_t \right) \tag{4.91}$$

where $\widehat{\boldsymbol{\theta}}_{ivx} = (\widehat{\boldsymbol{\theta}}_1', ..., \widehat{\boldsymbol{\theta}}_K')'$ and $\widehat{\boldsymbol{\theta}}_g^{ivx} = \left(\widehat{\delta}_g^{ivx}, \widehat{\boldsymbol{\beta}}_g^{ivx\prime}\right)'$ with $g \in \{1, ..., K\}$.

Define the following IVX-J statistic

$$\mathcal{J}_n = \widehat{\boldsymbol{\theta}}_{ivx}' \widehat{\boldsymbol{V}}^{-1} \widehat{\boldsymbol{\theta}}_{ivx} = \left(\sum_{t=1}^{n} \boldsymbol{Z}_{t-1}' \boldsymbol{Y}_t \right)' \widehat{\boldsymbol{M}}^{-1} \left(\sum_{t=1}^{n} \boldsymbol{Z}_{t-1}' \boldsymbol{Y}_t \right), \tag{4.92}$$

where $\widehat{\boldsymbol{V}}^{-1}$ is the variance matrix estimator and $\widehat{\boldsymbol{M}}$ is defined as below

$$\widehat{\boldsymbol{M}} = \sum_{t=1}^{n} \boldsymbol{Z}_{t-1}' \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} \boldsymbol{Z}_{t-1}. \tag{4.93}$$

Theorem 4.3. Suppose that Assumption hold and $\boldsymbol{w}_t = (u_t, \boldsymbol{v}_t')'$. Then, under the null hypothesis $\boldsymbol{\theta} = \mathbf{0}$ it holds that

$$\mathcal{J}_n \overset{d}{\to} \chi_K^2, \qquad \text{as} \qquad n \to \infty. \tag{4.94}$$

where

$$\mathcal{J}_n = \widehat{\boldsymbol{\theta}}_{ivx}' \widehat{\boldsymbol{V}}^{-1} \widehat{\boldsymbol{\theta}}_{ivx} = \left(\sum_{t=1}^{n} \boldsymbol{Z}_{t-1}' \boldsymbol{Y}_t \right)' \widehat{\boldsymbol{M}}^{-1} \left(\sum_{t=1}^{n} \boldsymbol{Z}_{t-1}' \boldsymbol{Y}_t \right), \tag{4.95}$$

and

$$\widehat{\boldsymbol{M}} := \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} \odot \sum_{t=1}^{n} \boldsymbol{Z}_{g,t-1} \boldsymbol{Z}_{g,t-1}' = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U},11} \sum_{t=1}^{n} \boldsymbol{z}_{1,t-1} \boldsymbol{z}_{1,t-1}' & \dots & \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U},1K} \sum_{t=1}^{n} \boldsymbol{z}_{1,t-1} \boldsymbol{z}_{K,t-1}' \\ \vdots & \ddots & \vdots \\ \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U},K1} \sum_{t=1}^{n} \boldsymbol{z}_{K,t-1} \boldsymbol{z}_{1,t-1}' & \dots & \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U},KK} \sum_{t=1}^{n} \boldsymbol{z}_{K,t-1} \boldsymbol{z}_{K,t-1}' \end{pmatrix}_{K \times K} \tag{4.96}$$

such that $\odot$ denotes the blockwise Hadamard matrix product.

Furthermore, we aim to investigate whether for all $g \in \{1, ..., K\}$ the variance of the OLS residuals from each system specific equation converges to a fixed covariance term.

Wald-type Statistics Formulation with nonstationary and generated regressors

Consider the following formulation of the predictive regression model which incorporates both nonstationary and generated regressors

$$\widetilde{\boldsymbol{y}}_t = \mu + \underbrace{\widetilde{\boldsymbol{Y}}_t^{o\prime}}_{1\times K} \boldsymbol{\delta} + \underbrace{\widetilde{\boldsymbol{X}}_{t-1}^{\prime}}_{1\times D} \boldsymbol{\beta} + \widetilde{\boldsymbol{u}}_t, \quad \text{for} \quad t = 1, ..., n \tag{4.97}$$

$$\boldsymbol{X}_t = \boldsymbol{R}_n \boldsymbol{X}_{t-1} + \boldsymbol{v}_t \tag{4.98}$$

Therefore, in a more compact form we obtain the following expression

$$\widetilde{\boldsymbol{y}}_t = \mu + \underbrace{\widetilde{\boldsymbol{G}}_{t-1}^{\prime}}_{1\times(D+K)} \boldsymbol{\vartheta} + \widetilde{\boldsymbol{u}}_t, \quad \text{for} \quad t = 1, ..., n \tag{4.99}$$

where $\boldsymbol{\vartheta} = (\boldsymbol{\delta}', \boldsymbol{\beta}')'$ and $\widetilde{\boldsymbol{G}}_{t-1} = \left[\widetilde{\boldsymbol{Y}}_t^{o\prime} \; \widetilde{\boldsymbol{X}}_{t-1}^{\prime} \right]$ where $\widetilde{\boldsymbol{Y}}_t^{o}$ corresponds to the undemeneaned generated regressors and $\widetilde{\boldsymbol{X}}_{t-1}$ corresponds to the undemeneaned nonstationary regressors.

Similarly, consider the demeaned random variables are denoted as

$$\boldsymbol{y}_t = \widetilde{\boldsymbol{y}}_t - \frac{1}{n}\sum_{t=1}^{n} \widetilde{\boldsymbol{y}}_t, \quad \boldsymbol{X}_t = \widetilde{\boldsymbol{X}}_t - \frac{1}{n}\sum_{t=1}^{n} \widetilde{\boldsymbol{X}}_t, \quad \boldsymbol{u}_t = \widetilde{\boldsymbol{u}}_t - \frac{1}{n}\sum_{t=1}^{n} \widetilde{\boldsymbol{u}}_t, \quad \boldsymbol{Y}_t^{o} = \widetilde{\boldsymbol{Y}}_t^{o} - \frac{1}{n}\sum_{t=1}^{n} \widetilde{\boldsymbol{Y}}_t^{o} \tag{4.100}$$

Therefore, we obtain the transformed predictive regression and system form as below

$$\boldsymbol{Y}_t = \underbrace{\boldsymbol{G}_{t-1}^{\prime}}_{1\times(D+K)} \boldsymbol{\vartheta} + \boldsymbol{U}_t, \quad \text{for} \quad t = 1, ..., n \quad \text{(see figure on landscape page)} \tag{4.101}$$

Additionally, in order to simplify the estimation procedure due to the high dimensionality of the statistical problem we decide to partition the large vector of nonstationary regressors such that $\boldsymbol{X}_{t-1} = [\boldsymbol{x}_{1t-1}, ..., \boldsymbol{x}_{Kt-1}]$ which implies that we group the $p-$dimensional vector of nonstationary regressors into K groups with a small number of regressors in each group.

$$y_t = \mu_j + \boldsymbol{\gamma}_j^{\prime} \boldsymbol{f}_{t-1} + \boldsymbol{\beta}_j^{\prime} \boldsymbol{x}_{t-1} + w_{it} \tag{4.102}$$

where $j = 1, ..., d$ and denote with $\widetilde{\boldsymbol{x}}_{jt} = (\boldsymbol{f}_t^{\prime}, \boldsymbol{x}_t^{\prime})'$ and $\boldsymbol{\theta}_j = (\boldsymbol{\gamma}_j^{\prime}, \boldsymbol{\beta}_j^{\prime})'$ and w_t is the linear projection residual. Furthermore, we define with $\boldsymbol{\beta} = (\beta_1, ..., \beta_d)'$ which has the same dimension as $\boldsymbol{\delta}_x$.

Proof of Theorem 1 Consider the Wald test under the null hypothesis such that

$$\mathcal{W} = \left(\sum_{t=1}^{n} \boldsymbol{z}_{t-1} u_t \right)' \left(\widehat{\sigma}_0^2 \sum_{t=1}^{n} \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}^{\prime} \right)' \left(\sum_{t=1}^{n} \boldsymbol{z}_{t-1} u_t \right) \tag{4.103}$$

We consider the limiting distribution for the $\mathcal{J}$ under the null hypothesis.

First we consider the asymptotic behaviour of the following quantities

$$\frac{1}{n^{(1+\gamma_z)}} \widehat{\boldsymbol{M}} = \frac{1}{n^{(1+\gamma_z)}} \sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} \boldsymbol{Z}_{t-1} \tag{4.104}$$

Now, under the assumption that the covariance matrix of the OLS residuals for each system specific equation is the same, then we can simplify the asymptotics as below

$$\begin{aligned} \frac{1}{n^{(1+\gamma_z)}} \widehat{\boldsymbol{M}} = \frac{1}{n^{(1+\gamma_z)}} \sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} \boldsymbol{Z}_{t-1} &= \frac{1}{n^{(1+\gamma_z)}} \sigma^2 \sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \boldsymbol{1}\boldsymbol{1}' \boldsymbol{Z}_{t-1} + o_p(1) \\ &= \frac{1}{n^{(1+\gamma_z)}} \sigma^2 \sum_{t=1}^{n} \boldsymbol{z}'_{t-1} \boldsymbol{z}_{t-1} + o_p(1) \xrightarrow{p} \sigma^2 \boldsymbol{Q}_{zz} > 0, \end{aligned}$$

Remark 4.9. Notice that the estimation methodology proposed by Kostakis et al. (2015) corresponds to a system of equations where each element of the multivariate regressand can be formulated as an individual predictive regression with regressor the particular element and regressors the common nonstationary regressors of the system. Therefore, our proposed estimation and inference methodology is represented with a SUR system of equations where these fitted system specific equations correspond to the quantile predictive regressions with nonstationary and generated regressors. Furthermore, we allow for differently generated nonstationary regressors due to the underline data generating mechanism we propose.

Theorem 4.4. Suppose that the conditions of Theorem 1 hold. Under the alternative hypothesis, the SUR IVX-based Wald test, denoted with $\mathcal{W}^{\star}_{SUR-IVX}$ converges to the following asymptotic distribution as $n \to \infty$

$$\frac{1}{\kappa_n} \mathcal{W}^{\star}_{SUR-IVX} \xrightarrow{d} \boldsymbol{\vartheta}^{*\prime} \boldsymbol{Q}'_{zx} \left[\boldsymbol{\Sigma}^{*}_{\boldsymbol{U}} \odot \bar{\boldsymbol{Q}}_{zz} \right]^{-1} \boldsymbol{Q}_{zx} \boldsymbol{\vartheta}^{*} \tag{4.105}$$

for some slow varying function κ_n.

Remark 4.10. Note that the $\mathcal{W}^{\star}_{SUR-IVX}$ test statistic reduces to the Wald-IVX test when we have one estimating equation. On the other hand, with $K \geq 2$, the $\mathcal{W}^{\star}_{SUR-IVX}$ test statistic is expected to perform better than the IVX-Wald since it relies on multivariate predictive regression models with (univariate regressands) that have much smaller sets of regressors in comparison to the corresponding multivariate formulation. Moreover, the proposed test at the same time can be more powerful than the LM test since its construction does not impose the entire null hypothesis.

4.4.6 Large Sample Theory

Assumption 4.7. We impose the following conditions

A1. Each entry in the vector $\boldsymbol{\alpha}(\tau)$ is $k-$order differentiable in a neighbourhood of z_0 for any z_0.

A2. $f_z(z)$ is continuously marginal density of Z and $f_z(z_0) > 0$.

A3. The distribution of Y given X has an everywhere positive conditional density $f_{Y|X}(\cdot)$, which is bounded and satisfies the Lipschitz continuity condition.

Notice that large sample theory can crucially depend on the relative magnitudes of m and n. The advantage of the SURE procedure, is that it allows for contemporaneous error covariances to be freely estimated. On the other hand, when m (the number of system equations) is the same as the number of time series observations, then SURE is not feasible. Only when m is reasonably small relative to n the proposed procedure is feasible. Moreover, although usually the literature on the particular estimation procedure is concerned with linear cross-equation restrictions, in our case our concern with the common long-run coefficients implies nonlinear restrictions across the different equations of the system (e.g., see McElroy and Burmeister (1988) and Ravikumar et al. (2000)). Furthermore, notice that in pooled models decomposing the short-run from the long-run effects is of practical relevance (see, Pesaran et al. (1999)). Robust inference approaches for multivariate time series using M-estimators are proposed by Koenker and Portnoy (1990) and Bilodeau and Duchesne (2000).

Our limit theory is developed under the Gaussian errors assumption. Thus, we develop asymptotics based on the Gaussianity condition on the error terms of system equations. In addition, our proposed statistical testing methodology is relevant from both the economic as well as the econometric perspective. From the economic perspective it provides a testing mechanism for the presence of systemic risk in the constructed network, while from the econometric perspective it provides insights regarding the presence of strong and weak instruments. As noted by Phillips and Gao (2017), this is natural since with both strong and weak instruments the reduced-form estimates contain information about the structural parameter $\boldsymbol{\beta}$, while under irrelevant instruments these estimates carry no such information. Moreover, it can be proved that the asymptotic distribution of the unrestricted reduced-form (URRF) test statistic, first-order stochastically dominates the asymptotic distribution of the partially restricted reduced-form (PRRF) under all three different instrument strengths. In addition, under the null hypothesis that the vector of coefficients of the systemic risk proxy, $\delta = 0$, the asymptotic distribution of the PRRF test statistic also becomes invariant to the strength of instruments (see, also Andrews and Cheng (2012)), which in our case implies that the robustness to the unknown persistence properties of regressors is preserved. Note that the two cases of instrumentation strength which lead to different asymptotic theory is the case of strong instruments and the case of totally irrelevant instruments which implies that there is no information in the reduced form about the structural coefficients and so the structural parameter $\boldsymbol{\beta}$ is not identified.

4.5 Monte Carlo Simulation Study

This section conducts Monte Carlo experiments to investigate the finite sample performances of our proposed estimators and test statistics. More specifically, in this section, we provide a complete simulation study for the performance of the proposed estimators as well as the Wald type statistic based on these estimators. Consider the following Data Generating Process (DGP), assuming that for example that $\left\{ Y_{t_i}^{(g)}, X_{t_i}^{(g)} \right\}_{i=1}^{n}$ for all $g \in \{1, ..., m\}$ are drawn from P_0, a known "probability law" (as we describe below).

Step 1. Denote with $\boldsymbol{\Sigma}_{ww}^{(g)} \sim \mathcal{W}_m \left(\boldsymbol{V}_{p \times p}, m \right)$ be the covariance matrix of the $g-$th equation such that for all $g \in \{1, ..., m\}$, $\boldsymbol{\Sigma}_{ww}^{(g)}$ is an $i.i.d$ random variable that follows a Wishart distribution[17], where $\boldsymbol{V}$ is a $p \times p$ positive definite matrix and m are the degrees of freedom such that $m > p + 1$ with $p = d + 1$, the number of parameters in each quantile predictive regression equation of the SUR.

Step 2. For each $g \in \{1, ..., N\}$ we generate the following predictive regression model

$$y_{(g)t} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jt-1} + u_{(g)t} \tag{4.106}$$

$$\boldsymbol{x}_{(g)t} = \boldsymbol{R}_n \boldsymbol{x}_{(g)t-1} + \boldsymbol{v}_{(g)t}, \;\; \boldsymbol{x}_0 = \boldsymbol{0}. \tag{4.107}$$

where $\boldsymbol{R}_n = \left(\boldsymbol{I}_d - \frac{\boldsymbol{C}_d}{n^{\gamma}} \right)$ with $\boldsymbol{C}_d = \mathsf{diag}\left\{ c_1, ..., c_d \right\}$ and $\gamma = 1$ for $t = 1, ..., n$.

For the Monte Carlo design of this paper we use $n = \left\{ 250, 500, 750, 1000 \right\}$, where n is the sample size, $B = 5,000$, where B is the number of monte carlo replications and $c_i = \{1, 2, 3, 4, 5, 6, 7\}$ for $d = 7$ the number of regressors in each model.

To generate the predictive regression model given by expressions (4.106)-(4.107), we assume that each vector $\boldsymbol{w}_{(g)t} = \left(u_{(g)t}, \boldsymbol{v}'_{(g)t} \right)'$ is a multivariate Gaussian random variate with mean zero and covariance matrix given by $\boldsymbol{\Sigma}_{ww}^{(g)} := \mathbb{E}\left[\boldsymbol{w}_{(g)t} \boldsymbol{w}'_{(g)t} \right]$, that is $\boldsymbol{w}_{(g)t} \sim \mathcal{N}\left(\boldsymbol{0}_{p \times p}, \boldsymbol{\Sigma}_{ww}^{(g)} \right)$. The parametrization of the covariance matrix $\boldsymbol{\Sigma}_{ww}^{(g)}$ is shown in Assumption 4.2 *(i)*. Moreover, we assume that $\beta_j \sim \mathsf{Unif}[-1, 1]$.

Step 3. For each pair of simulated data $\left\{ y_{(g)t}, \boldsymbol{x}_{(g)t} \right\}$, $g \in \{1, ..., m\}$ and $t = 1, ..., n$, we fit the quantile predictive regression models and estimate the IVX-Wald statistic.

[17] Notice that the Wishart distribution corresponds to the random variate of (XX') where X is $p-$dimensional Gaussian random variable.

4.6 Empirical Application

As an empirical application of this chapter and motivated from our Monte Carlo simulation experiments where for each predictive regression we generate nonstationary regressors with different nuisance parameters of persistence across the node specific equations, we implement a complex tail-dependency system of quantile predictive regressions in order to model systemic risk based on a large dataset of financial institutions that includes both firm characteristic as well as macroeconomic variables that capture economic conditions and financial stability.

Specifically, we use the dataset of Härdle et al. (2016) which includes a panel of the top 100 publicly traded financial institutions by market capitalization. More precisely, these are categorized into four groups: (i) depositories, (ii) insurance companies, (iii) broker-dealers, and (iv) others; which allows us to construct a financial network and apply the proposed graph based optimal portfolio allocation methodology. The dataset contains the stock returns of these firms along with a set of macroeconomic variables both corresponding to the same period (between 5, January 2007 and 4, January 2013) based on weekly time series observations. Furthermore, the dataset includes a set of firm variables which we use as the observable firm factors. In particular, these include balance sheet information such as: (i) total assets/total equity to capture firm leverage, (ii) short term debt-cash/total liabilities to capture maturity mismatch, (iii) ratio of the market to the book value of the total equity to capture the market-to-book firm characteristics and (iv) log of the total book equity to capture the size of the firm (see, also Katsouris (2021)).

Furthermore, the presence of the generated regressor (systemic risk proxy) in the quantile predictive regression specification, although is a necessary covariate that captures the Value-at-Risk can be found to neutralize the appearance of these effects. More precisely, through both the empirical study as well as simulation experiments we can check whether the addition of the generated regressor in the model could provide a way of neutralizing these size distortions due to the presence of high persistence regressors in contrast to the usual increase of size distortions as the number of the nonstationary regressors increases. In other words, a possible trade off between the number of nonstationary regressors added in the model and the number of generated covariates can provide a mechanism for controlling the empirical size close to the nominal size, while keeping the dimensionality of the statistical problem not growing at a faster rate than the sample size.

4.7 Conclusion

The third chapter of the thesis proposes an econometric framework for robust inference in quantile predictive regression systems under a general form of dependence. The dependence structure we consider combines the cross section with time series data under the assumption that the underline stochastic processes represents nodes on a graph. Furthermore, such a joint time series and cross-section limit theory implies that the cross-sectional (network specific) parameters of interest depend solely on parameters governing the time-series processes, since the network specific parameters are constructed from *i.i.d* random covariance matrices (i.e., the innovation sequences) and so there is a form of separation between cross-section dependence and time series dependence. An application to systemic risk modelling using both simulated and real data provides evidence of the practicality of our novel methodology in testing for systemic risk effects in our complex tail dependency driven system of equations.

Furthermore, we consider the nonstationary properties of predictors in terms of a near unit root parametrization which depends on the nuisance parameter of persistence. Although, the proposed complex tail-driven system has a corresponding matrix representation which facilitates estimation and testing, a companion matrix representation is not applicable. In particular, the non applicability of the companion matrix, similar to when modelling vector autoregression processes makes it more difficult to discuss the stability of our system in terms if the underlying stochastic processes. In terms of the inference methodology, propose a Wald type statistic suitable for testing the null hypothesis of linear restrictions in quantile predictive regressions estimated using the SUR system representation. The individual equations include a set of lagged regressors which are assumed to be generated as local unit root processes to capture the persistence properties in the time series of regressors. Furthermore, the set of explanatory variables include the systemic risk covariate which based on the graph structure represents by definition an element of the vector of regressands not the same as the equation-specific regressand. More precisely, we aim to show that the limiting distribution of the proposed Wald test is nuisance-parameter free and is robust to the persistence properties of regressors, for cases such as mildly integrated or persistence regressors and thus can be generalised in further nonstationary processes. Despite that fact that our proposed modelling methodology has some computational complexity, this approach provides a suitable econometric identification and estimation strategy for tail dependency in multivariate time series under nonstationarity and weak dependence. Therefore, the specific identification strategy allows to incorporate the main features of nonstationary time series models especially when modelling tail dependency across regressands that represent nodes in a graph. Although the current study does not consider explicitly such cointegration dynamics, it is still a novel contribution to the literature as it helps to alleviate any concerns related to an inconsitent estimation and non-identification of the econometric environment under both nonstationarity and network dependence. Our testing framework corresponds to in-sample testing for systemic risk effects based on our complex tail dependency driven system. Through simulation experiments we demonstrate the usefulness of our testing methodology in detecting the presence of system risk when modelling quantile processes under nonstationarity and network dependence.

Appendix A

Supplement to Chapter 2

A.1 Asymptotic Theory

In this Appendix we present the main mathematical derivations and proofs related to the asymptotic results reported in the paper. We derive large sample approximations to the distribution of the parameter constancy tests based on both the OLS-Wald and IVX-Wald statistics, under the null hypothesis of no parameter instability. We begin by summarizing via Lemma A.1 below the limit theory results which can be found in Phillips and Magdalinos (2009) and Kostakis et al. (2015). We introduce the shorthand notation $\alpha \wedge \beta \equiv \min(\alpha, \beta)$ to denote the minimum operator, employed for the stochastic dominance of the convergence rates.

Lemma A.1. Let $\mathbb{V}_{xz} := \int_0^\infty e^{rC} \boldsymbol{V}_{xx} e^{rC_z} dr$, where $\boldsymbol{V}_{xx} := \int_0^\infty e^{sC} \boldsymbol{\Omega}_{xx} e^{sC} ds$, and $\boldsymbol{\Omega}_{xx}$ is the long-run covariance of u_t. Then, under the null hypothesis of no structural break in the predictive regression model, the following asymptotic results hold:

(i) the sample covariance satisfies that

$$\frac{1}{T^{\frac{1+\gamma_x \wedge \delta_z}{2}}} \sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{z}_{t-1}(u_t - \bar{u}_T) \Rightarrow U(\pi) \tag{A.1}$$

where $U(.)$ is a Brownian motion with variance $\sigma_u^2 \widetilde{\boldsymbol{V}}$, where $\widetilde{\boldsymbol{V}}$ is defined as

$$\widetilde{\boldsymbol{V}} = \begin{cases} \displaystyle\int_0^\infty e^{rC_z} \boldsymbol{\Omega}_{xx} e^{rC_z} dr & \text{, if } \gamma_x > \delta_z \\ \displaystyle\int_0^\infty e^{rC_z} \left(\boldsymbol{C}\mathbb{V}_{xz} + \boldsymbol{C}_z \mathbb{V}_{xz}' \boldsymbol{C} \right) e^{rC_z} dr & \text{, if } \gamma_x = \delta_z \\ \displaystyle\int_0^\infty e^{rC} \boldsymbol{\Omega}_{xx} e^{rC} dr & \text{, if } 0 < \gamma_x < \delta_z \\ \mathbb{E}\left(x_{0,1} x_{0,1}' \right) & \text{, if } \gamma_x = 0. \end{cases} \tag{A.2}$$

where $x_{0,t} = \sum_{j=0}^\infty (\boldsymbol{I}_p + \boldsymbol{C})^j u_{t-j}$ is the corresponding stationary sequence of the regressor vector x_t when $\gamma_x = 0$.

(ii) the sample second moment satisfies that

$$\frac{1}{T^{\left(1+\gamma_x \wedge \delta_z\right)}} \sum_{t=1}^{\lfloor T\pi \rfloor} \tilde{z}_{t-1} \left(x_{t-1} - \bar{x}_{T-1}\right)' \Rightarrow \mathbf{\Psi}\left(\pi\right) \tag{A.3}$$

where $\mathbf{\Psi}(\pi)$ has a different asymptotic convergence result as below, depending on the exponent rates γ_x and δ_z of the original regressor and instrumental regressor respectively.

$$\mathbf{\Psi}\left(\pi\right) = \begin{cases} -\boldsymbol{C}_z^{-1}\left(\pi \boldsymbol{\Omega}_{xx} + \int_0^{\pi} \underline{\boldsymbol{J}}_C dJ_C'\right) & \text{, if } \gamma_x = 1 \\ -\pi \boldsymbol{C}_z^{-1}\left(\boldsymbol{\Omega}_{xx} + \int_0^{\infty} e^{rC} \boldsymbol{\Omega}_{xx} e^{rC} dr \boldsymbol{C}\right) & \text{, if } \delta_x < \gamma_x < 1 \\ -\pi \boldsymbol{C} \mathbb{V}_{xz} & \text{, if } \gamma_x = \delta_x \\ \pi \int_0^{\infty} e^{rC} \boldsymbol{\Omega}_{xx} e^{rC} dr & \text{, if } 0 < \gamma_x < \delta_x \\ \pi \mathbb{E}\left(x_{0,1} x_{0,1}'\right) & \text{, if } \gamma_x = 0. \end{cases} \tag{A.4}$$

where $B(.)$ is a $p-$dimensional standard Brownian motion, $J_C(\pi) = \int_0^{\pi} e^{C(\pi - s)} dB(\pi)$ is an *Ornstein-Uhkenbeck* (OU) process and we denote with $\underline{J}_C(\pi) = J_C(\pi) - \int_0^1 J_C(s)ds$ and $\underline{B}(\pi) = B(\pi) - \int_0^1 B(s)ds$ the demeaned processes of $J(\pi)$ and $B(\pi)$ respectively.

(iii) The weakly joint convergence result applies and the asymptotic terms given by expressions in *(i)* and *(ii)* are stochastically independent.

Notice that for summarizing the above results we used that

$$\frac{1}{T^{1+\delta_z}} \sum_{t=1}^{T} z_{t-1} z_{t-1}' \overset{\text{plim}}{\to} \boldsymbol{V}_{zz} := \int_0^{\infty} e^{rC_z} \boldsymbol{\Omega}_{xx} e^{rC_z} dr \tag{A.5}$$

Moreover, we have the weakly convergence result from Phillips and Magdalinos (2009):

$$\frac{1}{T^{\frac{1+\delta_z}{2}}} \sum_{t=1}^{T} \left(z_{t-1} \otimes u_t\right) \Rightarrow \mathcal{N}(0, \boldsymbol{V}_{zz} \otimes \boldsymbol{\Sigma}_{uu}) \tag{A.6}$$

Expression (B.81) proves a mixed Gaussian limiting distribution. This, shows that the limit distribution of $T^{-(1+\delta_z)/2} \sum_{t=1}^{T} \left(z_{t-1} \otimes u_t\right)$ is Gaussian with mean zero and covariance matrix equal to the probability limit of $T^{-(1+\delta)/2} \sum_{t=1}^{T} \left(z_{t-1} \otimes u_t\right)$, which is equal to $\boldsymbol{V}_{zz} \otimes \boldsymbol{\Sigma}_{uu}$, where $\boldsymbol{V}_{zz} := \int_0^{\infty} e^{rC_z} \boldsymbol{\Omega}_{xx} e^{rC_z} dr$. Specifically, the above Mixed Gaussianity convergence, is a powerful result within the IVX framework and ensures the robustness of the methodology and the estimation procedure. The dependence of the covariance matrix on the degree of persistence of the IVX instrumentation methodology, induces exactly the Mixed Gaussianity. Similarly,

$$T^{-(1+\delta_z)/2} \sum_{t=1}^{T} \left(x_{t-1} \otimes u_t\right) \Rightarrow \mathcal{N}(0, \boldsymbol{V}_{xx} \otimes \boldsymbol{\Sigma}_{uu}), \text{ where } \boldsymbol{V}_{xx} := \int_0^{\infty} e^{rC} \boldsymbol{\Omega}_{xx} e^{rC} dr \tag{A.7}$$

is proved in Lemma 3.3 of PM.

Proof of Theorem 2.1.

Proof. We denote with $\widetilde{x}_t = (1, x_t')'$ and with $\theta_j = (\alpha_j, \beta_j)'$ for $j = 1, 2$ the parameter vector which is obtained via the OLS estimator. An expression for obtaining the OLS estimator is

$$\widehat{\theta}_j = \underset{\theta_j \in \mathbb{R}^{p+1}}{\arg\min} \sum_{t=1}^{T} \left(y_t - x_{t-1}' \theta_j \right)^2, \quad \text{for} \quad j \in \{1, 2\}. \tag{A.8}$$

Under the null hypothesis of no structural break, $\mathbb{H}_0 : \theta_1 = \theta_2$, against $\mathbb{H}_1 : \theta_1 \neq \theta_2$, we obtain

$$\left(\widehat{\theta}_1 - \theta^0 \right) = \left(\sum_{t=1}^{T} \widetilde{x}_{t-1} \widetilde{x}_{t-1}' I_{1t} \right)^{-1} \left(\sum_{t=1}^{T} \widetilde{x}_{t-1} u_t I_{1t} \right)$$
$$\left(\widehat{\theta}_2 - \theta^0 \right) = \left(\sum_{t=1}^{T} \widetilde{x}_{t-1} \widetilde{x}_{t-1}' I_{2t} \right)^{-1} \left(\sum_{t=1}^{T} \widetilde{x}_{t-1} u_t I_{2t} \right)$$

with I_{1t} and I_{2t} the dummy time variables and $\theta^0 = (\alpha_0, \beta_0)'$, the population value of the parameter vector θ. Therefore, we have that

$$T \left(\widehat{\theta}_1 - \theta^0 \right) = \left(\frac{1}{T^2} \sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{x}_{t-1} \widetilde{x}_{t-1}' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{x}_{t-1}' u_t \right)$$
$$T \left(\widehat{\theta}_2 - \theta^0 \right) = \left(\frac{1}{T^2} \sum_{t=\lfloor T\pi \rfloor + 1}^{T} \widetilde{x}_{t-1} \widetilde{x}_{t-1}' \right)^{-1} \left(\frac{1}{T} \sum_{t=\lfloor T\pi \rfloor + 1}^{T} \widetilde{x}_{t-1}' u_t \right)$$

Then, the weakly convergence result for the estimator of β_1 follows

$$T \left(\widehat{\theta}_1 - \theta^0 \right) \Rightarrow \left(\int_0^{\pi} \widetilde{K}_c(r) \widetilde{K}_c'(r) dr \right)^{-1} \left(\int_0^{\pi} \widetilde{K}_c(r) dB_u \right) \tag{A.9}$$

Similarly, for the estimator of β_2 we have the following weakly convergence result

$$T \left(\widehat{\theta}_2 - \theta^0 \right) \Rightarrow \left(\int_{\pi}^{1} \widetilde{K}_c(r) \widetilde{K}_c'(r) dr \right)^{-1} \left(\int_{\pi}^{1} \widetilde{K}_c(r) dB_u \right) \tag{A.10}$$

In order to simplify the expression of the Wald OLS statistic we denote with

$$\widetilde{G}_c(\pi) := \int_0^{\pi} \widetilde{K}_c(r) \widetilde{K}_c'(r) dr \quad \text{and} \quad H_c(\pi) := \int_0^{\pi} \widetilde{K}_c(r) dB_u(r) \tag{A.11}$$

which implies that due to the argument π in the expressions for $\widetilde{G}_c(\pi)$ and $\widetilde{H}_c(\pi)$

$$\widetilde{G}_c(1) := \int_0^{1} \widetilde{K}_c(r) \widetilde{K}_c'(r) dr \quad \text{and} \quad \widetilde{H}_c(1) := \int_0^{1} \widetilde{K}_c(r) dB_u(r) \tag{A.12}$$

Notice that, for example we can deduce that

$$\left(\int_{\pi}^{1} \widetilde{K}_c(r) \widetilde{K}_c'(r) dr \right) = \left(\int_0^{1} \widetilde{K}_c(r) \widetilde{K}_c'(r) dr \right) - \left(\int_0^{\pi} \widetilde{K}_c(r) \widetilde{K}_c'(r) dr \right) := \widetilde{\mathbf{G}}_c(1) - \widetilde{\mathbf{G}}_c(\pi)$$

Thus, the statistical distance component of the sup Wald-OLS statistic is given by

$$T\left(\widehat{\theta}_1-\widehat{\theta}_2\right)=\left\{\widetilde{\mathbf{G}}_c(\pi)^{-1}\widetilde{H}_c(\pi)-[\widetilde{\mathbf{G}}_c(1)-\widetilde{\mathbf{G}}_c(\pi)]^{-1}[\widetilde{H}_c(1)-\widetilde{H}_c(\pi)]\right\} \tag{A.13}$$

Denote with $X=[x_tI_{1t}\ \ x_tI_{2t}]\equiv[X_1\ X_2]$ then the convergence of the covariance matrix

$$\begin{aligned}\widetilde{M}_c(\pi):=\left[\mathcal{R}\left(X'X\right)^{-1}\mathcal{R}'\right]&=\left[\left(\frac{X_1'X_1}{T^2}\right)^{-1}+\left(\frac{X_2'X_2}{T^2}\right)^{-1}\right]\\&\Rightarrow\left\{\left(\int_0^\pi\widetilde{K}_c(r)\widetilde{K}_c'(r)dr\right)^{-1}+\left(\int_\pi^1\widetilde{K}_c(r)\widetilde{K}_c'(r)dr\right)^{-1}\right\}\\&\equiv\left\{\widetilde{\mathbf{G}}_c(\pi)^{-1}+\left[\widetilde{\mathbf{G}}_c(1)-\widetilde{\mathbf{G}}_c(\pi)\right]^{-1}\right\}\end{aligned}$$

Recall that the expression for the Wald statistic is as below

$$\mathcal{W}_T^{OLS}(\pi)=\frac{1}{\widehat{\sigma}_u^2}\left(\widehat{\theta}_1-\widehat{\theta}_2\right)'\left[\mathcal{R}\left(X'X\right)^{-1}\mathcal{R}'\right]^{-1}\left(\widehat{\theta}_1-\widehat{\theta}_2\right) \tag{A.14}$$

Therefore, we can now derive the limiting distribution of the sup OLS-Wald statistic in the case of the multiple predictive regression with persistent predictors.

$$\begin{aligned}\widetilde{\mathcal{W}}^{OLS}(\pi)\Rightarrow\sup_{\pi\in[\pi_1,\pi_2]}&\left\{\widetilde{\mathbf{G}}_c(\pi)^{-1}H_c(\pi)-[\widetilde{\mathbf{G}}_c(1)-\widetilde{\mathbf{G}}_c(\pi)]^{-1}[\widetilde{H}_c(1)-\widetilde{H}_c(\pi)]\right\}'\\&\times\left\{\widetilde{\mathbf{G}}_c(\pi)^{-1}+\left[\widetilde{\mathbf{G}}_c(1)-\widetilde{\mathbf{G}}_c(\pi)\right]^{-1}\right\}^{-1}\\&\times\left\{\widetilde{\mathbf{G}}_c(\pi)^{-1}\widetilde{H}_c(\pi)-[\widetilde{\mathbf{G}}_c(1)-\widetilde{\mathbf{G}}_c(\pi)]^{-1}[\widetilde{H}_c(1)-\widetilde{H}_c(\pi)]\right\}\end{aligned}$$

By applying the related inverse matrix formula to $\widetilde{\mathbf{M}}_c(\pi)^{-1}$ we obtain that

$$\begin{aligned}\widetilde{\mathbf{S}}_c(\pi)^{-1}&\equiv\left\{\widetilde{\mathbf{G}}_c(\pi)^{-1}+\left[\widetilde{\mathbf{G}}_c(1)-\widetilde{\mathbf{G}}_c(\pi)\right]^{-1}\right\}^{-1}\\&=\widetilde{\mathbf{G}}_c(\pi)-\widetilde{\mathbf{G}}_c(\pi)\left[\widetilde{\mathbf{G}}_c(\pi)+\widetilde{\mathbf{G}}_c(1)-\widetilde{\mathbf{G}}_c(\pi)\right]^{-1}\widetilde{\mathbf{G}}_c(\pi)\\&=\widetilde{\mathbf{G}}_c(\pi)-\widetilde{\mathbf{G}}_c(\pi)\widetilde{\mathbf{G}}_c(1)^{-1}\widetilde{\mathbf{G}}_c(\pi)\end{aligned}$$

Thus, we show that the limiting distribution of the sup OLS-Wald statistic is given by

$$\widetilde{\mathcal{W}}^{OLS}(\pi)\equiv\sup_{\pi\in[\pi_1,\pi_2]}\mathcal{W}_T^{OLS}(\pi)\Rightarrow\sup_{\pi\in[\pi_1,\pi_2]}\left\{\widetilde{\boldsymbol{N}}_c'(\pi)\widetilde{\boldsymbol{M}}_c(\pi)^{-1}\widetilde{\boldsymbol{N}}_c(\pi)\right\} \tag{A.15}$$

with quantities $\widetilde{\boldsymbol{M}}_c(\pi)$, $\widetilde{\boldsymbol{N}}_c(\pi)$,$\widetilde{\mathbf{G}}_c(\pi)$ and $\widetilde{H}_c(\pi)$ as defined by Theorem 2.1. □

Proof of limit result for Univariate Case.

Proof. Denote with $I_{1t} = \mathbf{1}\{t \leq k\}$ and $I_{2t} = \mathbf{1}\{t > k\}$. Moreover, we denote the standard OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ of the corresponding model coefficients β_1 and β_2. Assume that the structural break is at an unknown break point such as $k = \lfloor n\pi \rfloor$ for $\pi \in (0, 1)$.

Then, using the FCLT we obtain the following weakly convergence results under the null hypothesis, $\mathbb{H}_0 : \beta_1 = \beta_2$ given below

$$n\left(\hat{\beta}_1 - \beta^0\right) = \frac{\frac{1}{n}\sum_{t=1}^{\lfloor n\pi \rfloor} x_{t-1}u_t}{\frac{1}{n^2}\sum_{t=1}^{\lfloor n\pi \rfloor} x_{t-1}^2} \Rightarrow \frac{\int_0^\pi K_c(r)dB_u(r)}{\int_0^\pi K_c^2(r)dr} \tag{A.16}$$

$$n\left(\hat{\beta}_2 - \beta^0\right) = \frac{\frac{1}{n}\sum_{t=\lfloor n\pi \rfloor+1}^{n} x_{t-1}u_t}{\frac{1}{n^2}\sum_{t=\lfloor n\pi \rfloor+1}^{n} x_{t-1}^2} \Rightarrow \frac{\int_\pi^1 K_c(r)dB_u(r)}{\int_\pi^1 K_c^2(r)dr} \tag{A.17}$$

Thus, using (A.16) and (A.17) we obtain the following simplified expression

$$n\left(\hat{\beta}_1 - \hat{\beta}_2\right) = n\left(\left(\hat{\beta}_1 - \beta^0\right) - \left(\hat{\beta}_2 - \beta^0\right)\right) \Rightarrow \frac{\int_0^\pi K_c(r)dB_u(r)}{\int_0^\pi K_c^2(r)dr} - \frac{\int_\pi^1 K_c(r)dB_u(r)}{\int_\pi^1 K_c^2(r)dr}$$

Denoting with $X = [x_{t-1}I_{1t} \;\; x_{t-1}I_{2t}] \equiv [X_1 \;\; X_2]$ then the OLS-Wald statistic has an equivalent representation as below

$$\mathcal{W}_n(\pi) = \frac{1}{\hat{\sigma}_u^2}\left(\hat{\beta}_1 - \hat{\beta}_2\right)'\left[\mathcal{R}\left(X'X\right)^{-1}\mathcal{R}'\right]^{-1}\left(\hat{\beta}_1 - \hat{\beta}_2\right) \tag{A.18}$$

Using the orthogonality property of X_1 and X_2 and via a standard matrix inversion application, we obtain that

$$\begin{aligned}\left[\mathcal{R}\left(X'X\right)^{-1}\mathcal{R}'\right] &= \left[\left(\sum_{t=1}^{n} x_{t-1}^2 I_{1t}\right)^{-1} + \left(\sum_{t=1}^{n} x_{t-1}^2 I_{1t}\right)^{-1}\right] \\ &= \frac{\sum_{t=1}^{n} x_{t-1}^2 I_{1t} + \sum_{t=1}^{n} x_{t-1}^2 I_{2t}}{\left(\sum_{t=1}^{n} x_{t-1}^2 I_{1t}\right)\left(\sum_{t=1}^{n} x_{t-1} I_{2t}\right)} = \frac{\sum_{t=1}^{n} x_{t-1}^2}{\left(\sum_{t=1}^{n} x_{t-1}^2 I_{1t}\right)\left(\sum_{t=1}^{n} x_{t-1}^2 I_{2t}\right)}\end{aligned}$$

Therefore, the simplified expression of the Wald statistic is given by

$$\begin{aligned}\mathcal{W}_n(\pi) &= \frac{\left(\hat{\beta}_1 - \hat{\beta}_2\right)^2}{\hat{\sigma}_u^2}\left[\frac{\sum_{t=1}^{n} x_{t-1}^2}{\left(\sum_{t=1}^{n} x_{t-1}^2 I_{1t}\right)\left(\sum_{t=1}^{n} x_{t-1}^2 I_{2t}\right)}\right]^{-1} \\ &= \frac{\left(\hat{\beta}_1 - \hat{\beta}_2\right)^2}{\hat{\sigma}_u^2}\frac{\left(\sum_{t=1}^{n} x_{t-1}^2 I_{1t}\right)\left(\sum_{t=1}^{n} x_{t-1}^2 I_{2t}\right)}{\sum_{t=1}^{n} x_{t-1}^2} \\ &= n^2\frac{\left(\hat{\beta}_1 - \hat{\beta}_2\right)^2}{\hat{\sigma}_u^2}\frac{\left(\sum_{t=1}^{n} \frac{x_{t-1}^2 I_{1t}}{n^2}\right)\left(\sum_{t=1}^{n} \frac{x_{t-1}^2 I_{2t}}{n^2}\right)}{\sum_{t=1}^{n} \frac{x_{t-1}^2}{n^2}}\end{aligned}$$

Moreover, the following asymptotic convergence result also holds

$$\frac{\left(\sum_{t=1}^{n} \frac{x_{t-1}^2 I_{1t}}{n^2}\right)\left(\sum_{t=1}^{n} \frac{x_{t-1}^2 I_{2t}}{n^2}\right)}{\sum_{t=1}^{n} \frac{x_{t-1}^2}{n^2}} \Rightarrow \frac{\left(\int_0^{\pi} K_c^2(r)dr\right)\left(\int_{\pi}^{1} K_c^2(r)dr\right)}{\int_0^1 K_c^2(r)dr} \tag{A.19}$$

Assuming that the convergence in probability $\hat{\sigma}_u \xrightarrow{p} \sigma_u$ holds then the limiting distribution of the Wald test under persistent regressors is given by the following expression:

$$\mathcal{W}_n^*(\pi) \Rightarrow \sup_{\pi\in[\pi_1,\pi_2]} \frac{1}{\sigma_u^2}\left[\frac{\int_0^{\pi} K_c(r)dB_u(r)}{\int_0^{\pi} K_c^2(r)dr} - \frac{\int_{\pi}^{1} K_c(r)dB_u(r)}{\int_{\pi}^{1} K_c^2(r)dr}\right]^2 \frac{\left(\int_0^{\pi} K_c^2(r)dr\right)\left(\int_{\pi}^{1} K_c^2(r)dr\right)}{\int_0^1 K_c^2(r)dr} \tag{A.20}$$

Moreover, we can simplify further the above expression incorporating the covariance terms, σ_u and σ_v that corresponds to the innovation sequences u_t and v_t respectively. Recall also that we have: $B_u(r) = \sigma_u W_u(r)$ and $K_c(r) = \sigma_v J_c(r)$ and due to the orthogonality of the two predictors which implies that $\int_0^1 J_c^2(r)dr = \int_0^{\pi} J_c^2(r)dr + \int_{\pi}^1 J_c^2(r)dr$ and $\int_0^1 J_c(r)dW_u = \int_0^{\pi} J_c(r)dW_u + \int_{\pi}^1 J_c(r)dW_u$ we obtain the expression below:

$$\frac{\int_0^{\pi} K_c(r)dB_u(r)}{\int_0^{\pi} K_c^2(r)dr} = \frac{\sigma_u\sigma_v \int_0^{\pi} J_c(r)dW_u(r)}{\sigma_v^2 \int_0^{\pi} J_c^2(r)dr} = \frac{\sigma_u}{\sigma_v}\frac{\int_0^{\pi} J_c(r)dW_u(r)}{\int_0^{\pi} J_c^2(r)dr} \tag{A.21}$$

Similarly we have that,

$$\frac{\int_{\pi}^{1} K_c(r)dB_u(r)}{\int_{\pi}^{1} K_c^2(r)dr} = \frac{\sigma_u\sigma_v \int_{\pi}^{1} J_c(r)dW_u(r)}{\sigma_v^2 \int_{\pi}^{1} J_c^2(r)dr} = \frac{\sigma_u}{\sigma_v}\frac{\int_{\pi}^{1} J_c(r)dW_u(r)}{\int_{\pi}^{1} J_c^2(r)dr} \tag{A.22}$$

and also

$$\frac{\left(\int_0^{\pi} K_c^2(r)dr\right)\left(\int_{\pi}^{1} K_c^2(r)dr\right)}{\int_0^1 K_c^2(r)dr} = \frac{\sigma_v^2\sigma_v^2\left(\int_0^{\pi} J_c^2(r)dr\right)\left(\int_{\pi}^{1} J_c^2(r)dr\right)}{\sigma_v^2\int_0^1 J_c^2(r)dr} \tag{A.23}$$

which implies that the limiting distribution of the sup-Wald statistic can be simplified

$$\begin{aligned}
\mathcal{W}_n^*(\pi) &= \frac{1}{\sigma_u^2}\left[\frac{\int_0^{\pi} K_c(r)dB_u(r)}{\int_0^{\pi} K_c^2(r)dr} - \frac{\int_{\pi}^{1} K_c(r)dB_u(r)}{\int_{\pi}^{1} K_c^2(r)dr}\right]^2 \frac{\left(\int_0^{\pi} K_c^2(r)dr\right)\left(\int_{\pi}^{1} K_c^2(r)dr\right)}{\int_0^1 K_c^2(r)dr} \\
&= \frac{1}{\sigma_u^2}\left[\frac{\sigma_u}{\sigma_v}\frac{\int_0^{\pi} J_c(r)dW_u(r)}{\int_0^{\pi} J_c^2(r)dr} - \frac{\sigma_u}{\sigma_v}\frac{\int_{\pi}^{1} J_c(r)dW_u(r)}{\int_{\pi}^{1} J_c^2(r)dr}\right]^2 \frac{\sigma_v^2\left(\int_0^{\pi} J_c^2(r)dr\right)\left(\int_{\pi}^{1} J_c^2(r)dr\right)}{\int_0^1 J_c^2(r)dr}
\end{aligned}$$

Therefore, we obtain that

$$\mathcal{W}_n^*(\pi) \Rightarrow \sup_{\pi\in[\pi_1,\pi_2]}\left[\frac{\int_0^{\pi} J_c(r)dW_u}{\int_0^{\pi} J_c^2(r)dr} - \frac{\int_{\pi}^{1} J_c(r)dW_u}{\int_{\pi}^{1} J_c^2(r)dr}\right]^2 \frac{\left(\int_0^{\pi} J_c^2(r)dr\right)\left(\int_{\pi}^{1} J_c^2(r)dr\right)}{\int_0^1 J_c^2(r)dr} \tag{A.24}$$

which with simple algebra can be further simplified to the following expression

$$\mathcal{W}_n^*(\pi) \Rightarrow \sup_{\pi\in[\pi_1,\pi_2]}\frac{\left[\int_0^{\pi} J_c(r)dW_u - \frac{\int_0^{\pi} J_c^2(r)dr}{\int_0^1 J_c^2(r)dr}\int_0^1 J_c(r)dW_u\right]^2}{\left(\int_0^1 J_c^2(r)dr - \int_0^{\pi} J_c^2(r)dr\right)\left(\int_0^{\pi} J_c^2(r)dr\right)} \tag{A.25}$$

Expression (A.25) gives the main result of this paper, which is the limiting distribution of the sup-Wald statistic for testing for a single structural break when the regressor is highly persistent. Furthermore, letting $M_c(\pi) := \int_0^{\pi} J_c(r)dW_u$ and $Q_c(\pi) := \int_0^{\pi} J_c^2(r)dr$.

$$\mathcal{W}_n^*(\pi) \Rightarrow \sup_{\pi\in[\pi_1,\pi_2]}\frac{\left[M_c(\pi) - \frac{Q_c(\pi)}{Q_c(1)}M_c(1)\right]^2}{Q_c(\pi)\left[1 - \frac{Q_c(\pi)}{Q_c(1)}\right]} \tag{A.26}$$

□

Proof of Theorem 2.2.

Proof. Let Y denote the vector with all demeaned values of y_t and X be the matrix collecting all demeaned values of x_{t-1}, that is,

$$Y = (y_1 - \bar{y}_T, y_2 - \bar{y}_T, ..., y_T - \bar{y}_T)' \text{ and } X = (x_0 - \bar{x}_{T-1}, x_1 - \bar{x}_{T-1}, ..., x_{T-1} - \bar{x}_{T-1})'.$$

Similarly, we use $\mathcal{U}$ to denote the corresponding demeaned u_t vector. Furthermore, for any $1 \leq t \leq T$, we define X_t to be a $T \times p$ matrix, whose first t rows are the same as X while the rest are all zeros. Moreover, let $Z = (z_0, z_1, ...z_{T-1})'$ collect all the IVX instruments, and $Z_t = (z_0, ..., z_t, 0, ..., 0)'$ be the corresponding time$-t$ truncated matrix. Given these notations, we express the original predictive regression model as

$$Y = X\beta_2 + X_t \eta + \mathcal{U} \tag{A.27}$$

where $\eta := \beta_2 - \beta_1$ measures the magnitude of structural break. Moreover, we denote with ϕ_t to the corresponding estimator which captures the break size associated with the sample partition at time t. Therefore, given any particular t, testing for structural break in the parameter vector β is equivalent to testing the null hypothesis $\eta_t = 0$. Define with $\boldsymbol{M}_{xz} = \boldsymbol{I}_p - X\left(Z'X\right)^{-1} Z'$, which is idempotent and orthogonal to both X and Z and allows to rewrite (A.27) in its canonical form[1]. Multiplying $\boldsymbol{M}_{xz}$ on both sides of (A.27), we deduce that $\boldsymbol{M}_{xz}Y = \boldsymbol{M}_{xz}X_t\phi_t + \boldsymbol{M}_{xz}\mathcal{U}$. Now, using $\boldsymbol{M}_{xz}Z_t$ as the instrumental variables for $\boldsymbol{M}_{xz}X_t$, we obtain an estimator for the parameter η_t given by

$$\widetilde{\eta}_t = \left(Z_t'\boldsymbol{M}_{xz}X_t\right)^{-1} Z_t'\boldsymbol{M}_{xz}Y \tag{A.28}$$

Moreover, it holds that $\widetilde{\eta}_t = \widetilde{\beta}_2^{IVX} - \widetilde{\beta}_1^{IVX}$. Thus, the limiting distribution is given by

$$\begin{aligned}
\widetilde{\eta}_t - \eta_t &= \left(Z_t'\boldsymbol{M}_{xz}X_t\right)^{-1} Z_t'\boldsymbol{M}_{xz}\mathcal{U} \\
&= \left(Z_t'X_t - Z_t'X\left(Z'X\right)^{-1}Z'X_t\right)^{-1}\left(Z_t'\mathcal{U}_y - Z_t'X_t\left(Z'X\right)^{-1}Z'\mathcal{U}\right) \\
&= \left[\sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{z}_t\left(x_t - \bar{x}_{T-1}\right)' - \sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{z}_t\left(x_t - \bar{x}_{T-1}\right)'\left(\sum_{t=1}^{T} \widetilde{z}_t\left(x_t - \bar{x}_{T-1}\right)'\right)^{-1} \sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{z}_t\left(x_t - \bar{x}_{T-1}\right)'\right]^{-1} \\
&\times \left[\sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{z}_t\left(u_t - \bar{u}_T\right)' - \sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{z}_t\left(x_t - \bar{x}_{T-1}\right)'\left(\sum_{t=1}^{T} \widetilde{z}_t\left(x_t - \bar{x}_{T-1}\right)'\right)^{-1} \sum_{t=1}^{T} \widetilde{z}_t\left(u_t - \bar{u}_T\right)\right]
\end{aligned} \tag{A.29}$$

[1]Notice that the reparametrization of the model to its canonical form allows to shift the coordinates which transforms the model to its more general form within the exponential family.

We denote the weakly convergence of the above moments as below

$$\sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{z}_t \left(x_t - \bar{x}_{T-1} \right) \Rightarrow \boldsymbol{\Psi}(\pi) \quad \text{and} \quad \sum_{t=1}^{T} \widetilde{z}_t \left(x_t - \bar{x}_{T-1} \right) \Rightarrow \boldsymbol{\Psi}(1) \tag{A.30}$$

Applying the asymptotic results given by Lemma A.1 to (A.29) we obtain

$$T^{\frac{1+\gamma_x \wedge \delta_z}{2}} \left(\widetilde{\eta}_t - \eta_t \right) \Rightarrow \left[\boldsymbol{\Psi}(\pi) - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1}\boldsymbol{\Psi}(\pi)' \right]^{-1} \Big(U(\pi) - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} U(1) \Big)$$

Next, we focus on covariance estimators for $\widetilde{\boldsymbol{Q}}_1(t) = \left(Z_t' X_t \right)^{-1} \left(Z_t' Z_t \right) \left(X_t' Z_t \right)^{-1}$ and the corresponding one for $\widetilde{\boldsymbol{Q}}_2(t)$.

$$\widetilde{\boldsymbol{Q}}_1(t) = \left(\sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{z}_t \left(x_t - \bar{x}_{T-1} \right)' \right)^{-1} \left(\sum_{t=1}^{\lfloor T\pi \rfloor} \widetilde{z}_t \widetilde{z}_t' \right) \left(\sum_{t=1}^{\lfloor T\pi \rfloor} \left(x_t - \bar{x}_{T-1} \right) \widetilde{z}_t' \right)^{-1} \tag{A.31}$$

which implies that

$$T^{1+\gamma_x \wedge \delta_z} \widetilde{\boldsymbol{Q}}_1(t) \Rightarrow \boldsymbol{\Psi}(\pi)^{-1} \left(\pi \sigma_u^2 \widetilde{\boldsymbol{V}} \right) \boldsymbol{\Psi}(\pi)^{-1'} \tag{A.32}$$

Similarly,

$$\widetilde{\boldsymbol{Q}}_2(t) = \left(\sum_{t=\lfloor T\pi \rfloor+1}^{T} \widetilde{z}_t \left(x_t - \bar{x}_{T-1} \right)' \right)^{-1} \left(\sum_{t=\lfloor T\pi \rfloor+1}^{T} \widetilde{z}_t \widetilde{z}_t' \right) \left(\sum_{t=\lfloor T\pi \rfloor+1}^{T} \left(x_t - \bar{x}_{T-1} \right) \widetilde{z}_t' \right)^{-1} \tag{A.33}$$

which implies that

$$T^{1+\gamma_x \wedge \delta_z} \widetilde{\boldsymbol{Q}}_2(t) \Rightarrow \left(\boldsymbol{\Psi}(1) - \boldsymbol{\Psi}(\pi) \right)^{-1} \left((1-\pi) \sigma_u^2 \widetilde{\boldsymbol{V}} \right) \left(\boldsymbol{\Psi}(1) - \boldsymbol{\Psi}(\pi) \right)^{-1'} \tag{A.34}$$

Combining all the above we obtain the following result for the $\mathcal{W}_b(t)$ test statistic

$$\begin{aligned}
\mathcal{W}_b(t) &= \left(\widetilde{\beta}_2^{IVX}(t) - \widetilde{\beta}_1^{IVX}(t) \right)' \left[\widetilde{\boldsymbol{Q}}_1(t) + \widetilde{\boldsymbol{Q}}_2(t) \right]^{-1} \left(\widetilde{\beta}_2^{IVX}(t) - \widetilde{\beta}_1^{IVX}(t) \right) \\
&= \left\{ T^{\frac{1+\gamma_x \wedge \delta_z}{2}} \left(\widetilde{\beta}_2^{IVX}(t) - \widetilde{\beta}_1^{IVX}(t) \right) \right\}' \times \left[T^{1+\gamma_x \wedge \delta_z} \left(\widetilde{\boldsymbol{Q}}_1(t) + \widetilde{\boldsymbol{Q}}_2(t) \right) \right]^{-1} \\
&\quad \times \left\{ T^{\frac{1+\gamma_x \wedge \delta_z}{2}} \left(\widetilde{\beta}_2^{IVX}(t) - \widetilde{\beta}_1^{IVX}(t) \right) \right\}
\end{aligned} \tag{A.35}$$

which implies the following weakly convergence result

$$\begin{aligned}
\mathcal{W}_\beta^{IVX}(t) \Rightarrow & \Big(U(\pi) - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} U(1) \Big)' \left[\boldsymbol{\Psi}(\pi) - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1}\boldsymbol{\Psi}(\pi)' \right]^{-1'} \\
& \times \left[\boldsymbol{\Psi}(\pi)^{-1} \left(\pi \sigma_y^2 \widetilde{\boldsymbol{V}} \right) \boldsymbol{\Psi}(\pi)^{-1'} + \left(\boldsymbol{\Psi}(1) - \boldsymbol{\Psi}(\pi) \right)^{-1} \left((1-\pi) \sigma_y^2 \widetilde{\boldsymbol{V}} \right) \left(\boldsymbol{\Psi}(1) - \boldsymbol{\Psi}(\pi) \right)^{-1'} \right]^{-1} \\
& \times \left[\boldsymbol{\Psi}(\pi) - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1}\boldsymbol{\Psi}(\pi)' \right]^{-1} \Big(U(\pi) - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} U(1) \Big)
\end{aligned} \tag{A.36}$$

To simplify the notation of expression (A.36), we denote with $A = \mathbf{\Psi}(\pi)$, $C = \mathbf{\Psi}(1)$ and $\Sigma = \sigma_u^2 \widetilde{\boldsymbol{V}}$. Then, we have the following equivalent form of the IVX-Wald statistic

$$\begin{aligned} \mathcal{W}_\beta^{IVX}(t) \equiv & \Big(U(\pi) - AC^{-1}U(1)\Big)' \Big(A - AC^{-1}A'\Big)^{-1'} \\ & \times \Big[\pi A^{-1} \Sigma A^{-1'} + (1-\pi)(C-A)^{-1}\Sigma(C-A)^{-1'}\Big]^{-1} \\ & \times \Big(A - AC^{-1}A'\Big)^{-1} \Big(U(\pi) - AC^{-1}U(1)\Big) \end{aligned} \tag{A.37}$$

which can be written as below

$$\begin{aligned} \mathcal{W}_\beta^{IVX}(t) \equiv & \Big(U(\pi) - AC^{-1}U(1)\Big)' \\ & \times \Bigg(\pi \Big[\Big(A - AC^{-1}A'\Big) A^{-1} \Sigma A^{-1'} \Big(A - AC^{-1}A'\Big)'\Big] \\ & + (1-\pi) \Big[\Big(A - AC^{-1}A'\Big) (C-A)^{-1} \Sigma (C-A)^{-1'} \Big(A - AC^{-1}A'\Big)'\Big] \Bigg) \\ & \times \Big(U(\pi) - AC^{-1}U(1)\Big) \\ = & \Big(U(\pi) - AC^{-1}U(1)\Big)' \\ & \times \Big(\pi(I - AC^{-1})\Sigma(I - AC^{-1})' + (1-\pi)\Big(AC^{-1}\Big)\Sigma\Big(AC^{-1}\Big)'\Big)^{-1} \\ & \times \Big(U(\pi) - AC^{-1}U(1)\Big). \end{aligned} \tag{A.38}$$

Notice that since $U(.)$ is known to be a Brownian motion with variance Σ, the above expression can be further simplified as following

$$\begin{aligned} \mathcal{W}_\beta^{IVX}(t) \equiv & \Big(B(\pi) - AC^{-1}B(1)\Big)' \\ & \times \Big(\pi(I - AC^{-1})(I - AC^{-1})' + (1-\pi)\Big(AC^{-1}\Big)\Big(AC^{-1}\Big)'\Big)^{-1} \\ & \times \Big(B(\pi) - AC^{-1}B(1)\Big). \end{aligned} \tag{A.39}$$

Substituting back to expression (A.39) the notation for $A = \mathbf{\Psi}(\pi)$, $C = \mathbf{\Psi}(1)$ and $\Sigma = \sigma_u^2 \widetilde{\boldsymbol{V}}$,

$$\begin{aligned} \mathcal{W}_\beta^{IVX}(t) \Rightarrow & \Big(B(\pi) - \mathbf{\Psi}(\pi)\mathbf{\Psi}(1)^{-1}B(1)\Big)' \\ & \times \Big(\pi(\boldsymbol{I}_p - \mathbf{\Psi}(\pi)\mathbf{\Psi}(1)^{-1})(\boldsymbol{I}_p - \mathbf{\Psi}(\pi)\mathbf{\Psi}(1)^{-1})' + (1-\pi)\Big(\mathbf{\Psi}(\pi)\mathbf{\Psi}(1)^{-1}\Big)\Big(\mathbf{\Psi}(\pi)\mathbf{\Psi}(1)^{-1}\Big)'\Big)^{-1} \\ & \times \Big(B(\pi) - \mathbf{\Psi}(\pi)\mathbf{\Psi}(1)^{-1}B(1)\Big). \end{aligned}$$

where $B(.)$ is a standard Brownian motion.

Using the asymptotic results given by Lemma A.1, we simplify expression $\boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1}$ to

$$\boldsymbol{R}(\pi) := \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} = \begin{cases} \left(\pi\boldsymbol{\Omega}_{xx} + \int_0^{\pi} \underline{\boldsymbol{J}}_C dJ_C'\right)\left(\boldsymbol{\Omega}_{xx} + \int_0^{1} \underline{\boldsymbol{J}}_C dJ_C'\right)^{-1} & , \text{if } \gamma_x = 1 \\ \\ \pi\boldsymbol{I}_p & , 0 < \gamma_x < 1. \end{cases} \tag{A.40}$$

Notice that the last asymptotic result, that is, the case where $0 < \gamma_x < 1$ holds because under the assumption of mildly integrated (or near-stationary) predictors, then we have that $\boldsymbol{\Psi}(1) := \sum_{t=1}^{T} \tilde{z}_t \left(x_t - \bar{x}_{T-1}\right) \Rightarrow -\Omega_{xx}C_z^{-1}$ by expression (20) in PM (2009). It also holds that $\boldsymbol{\Psi}(\pi) := \sum_{t=1}^{\lfloor T\pi \rfloor} \tilde{z}_t \left(x_t - \bar{x}_{T-1}\right) \Rightarrow -\pi\Omega_{xx}C_z^{-1}$, which implies that $\boldsymbol{R}(\pi) = \pi I_p$

Therefore, by denoting the Brownian functional above with $\boldsymbol{N}(\pi) = \Big(B(\pi) - \boldsymbol{R}(\pi)B(1)\Big)$ and $\boldsymbol{M}(\pi) = \Big(\pi\left(\boldsymbol{I}_p - \boldsymbol{R}(\pi)\right)\left(\boldsymbol{I}_p - \boldsymbol{R}(\pi)\right)' + (1-\pi)\boldsymbol{R}(\pi)\boldsymbol{R}(\pi)'\Big)$, then we obtain

$$\widetilde{\mathcal{W}}_{\beta}^{IVX}(t) \Rightarrow \sup_{\pi \in [\pi_1, \pi_2]} \Big\{ \boldsymbol{N}(\pi)'\boldsymbol{M}(\pi)^{-1}\boldsymbol{N}(\pi) \Big\} \tag{A.41}$$

□

Proof of Corollary 2.2.

Proof. When $0 < \gamma_x < 1$, we have that $\boldsymbol{R}(\pi) = \pi\boldsymbol{I}_p$ which implies that

$$\boldsymbol{M}(\pi) = \pi\left(\boldsymbol{I}_p - \pi\boldsymbol{I}_p\right)\left(\boldsymbol{I}_p - \pi\boldsymbol{I}_p\right)' + (1-\pi)\pi\boldsymbol{I}_p\left(\pi\boldsymbol{I}_p\right)' = \left[\pi(1-\pi)^2 + (1-\pi)\pi^2\right]\boldsymbol{I}_p = \pi(1-\pi)\boldsymbol{I}_p \tag{A.42}$$

Hence, in this case, the limiting distribution in Theorem **??** will reduce to

$$\widetilde{\mathcal{W}}_{\beta}^{IVX}(t) \Rightarrow \sup_{\pi \in [\pi_1, \pi_2]} \frac{\left[B(\pi) - \pi B(1)\right]'\left[B(\pi) - \pi B(1)\right]}{\pi(1-\pi)}. \tag{A.43}$$

□

Thus, the above proof demonstrates that when we have predictors generated via the LUR specification with $0 < \gamma_x < 1$ (i.e., MI predictors) and we consider testing the null hypothesis of no parameter instability then the limiting distribution of the IVX Wald statistic converges to a NBB similar to the result of Andrews (1993) in the case of linear regression models, which allows us to use already tabulated critical values.

Proof of Corollary 2.3. *(i).* Using the above notations, it's straightforward to obtain that under the null hypothesis $\beta_1 = \beta_2 = 0$, we obtain that

$$\mathcal{W}_T^{IVX} = \widetilde{\beta}^{IVX\prime} \widetilde{\boldsymbol{Q}}_{\mathcal{R}}^{-1} \widetilde{\beta}^{IVX} \Rightarrow U(1)' \left[\boldsymbol{\Psi}(1)^{-1\prime} \boldsymbol{\Sigma} \boldsymbol{\Psi}(1)^{-1} \right]^{-1} U(1). \tag{A.44}$$

Combining this result with Theorem 2.2, we can deduce that

$$\begin{aligned} \mathcal{W}_\beta^{joint} &= \mathcal{W}_T^{IVX} + \mathcal{W}_\beta^{IVX}(t) \\ &\Rightarrow \begin{pmatrix} \boldsymbol{\Psi}(1)^{-1} U(1) \\ U(\tau) - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} U(1) \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Psi}(1)^{-1\prime} \boldsymbol{\Sigma} \boldsymbol{\Psi}(1)^{-1} & \boldsymbol{0}_{p\times p} \\ \boldsymbol{0}_{p\times p} & \boldsymbol{\Delta}_{p\times p} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\Psi}(1)^{-1} U(1) \\ U(\pi) - \boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}(1)^{-1} U(1) \end{pmatrix} \end{aligned}$$

$$\boldsymbol{\Delta} := \pi (\boldsymbol{I}_p - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1}) \boldsymbol{\Sigma} (\boldsymbol{I}_p - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1})' + (1-\pi) \left(\boldsymbol{I}_p - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} \right) \boldsymbol{\Sigma} \left(\boldsymbol{I}_p - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} \right)'$$

Therefore, we obtain that

$$\begin{aligned} \mathcal{W}_\beta^{joint} &= \mathcal{W}_T^{IVX} + \mathcal{W}_\beta^{IVX}(t) \\ &\Rightarrow \begin{pmatrix} \boldsymbol{\Psi}(1)^{-1} B(1) \\ B(\pi) - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} B(1) \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Psi}(1)^{-1\prime} \boldsymbol{\Psi}(1)^{-1} & \boldsymbol{0}_{p\times p} \\ \boldsymbol{0}_{p\times p} & \widetilde{\boldsymbol{\Delta}}_{p\times p} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\Psi}(1)^{-1} B(1) \\ B(\pi) - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} B(1) \end{pmatrix} \end{aligned}$$

$$\widetilde{\boldsymbol{\Delta}} := \pi (\boldsymbol{I}_p - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1}) (\boldsymbol{I}_p - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1})' + (1-\pi) \left(\boldsymbol{I}_p - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} \right) \left(\boldsymbol{I}_p - \boldsymbol{\Psi}(\pi)\boldsymbol{\Psi}(1)^{-1} \right)'$$

which shows that

$$\begin{aligned} \mathcal{W}_\beta^{joint} &\Rightarrow \begin{pmatrix} B(1) \\ B(\pi) - \boldsymbol{R}(\pi) B(1) \end{pmatrix}' \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{0}_{p\times p} \\ \boldsymbol{0}_{p\times p} & \boldsymbol{M}(\pi) \end{pmatrix}^{-1} \begin{pmatrix} B(1) \\ B(\pi) - \boldsymbol{R}(\pi) B(1) \end{pmatrix} \\ &\equiv B(1)B(1)' + \boldsymbol{N}(\pi)' \boldsymbol{M}(\pi)^{-1} \boldsymbol{N}(\pi) \end{aligned}$$

Since the first component of the above decomposition is independent of π, then by the Continuous Mapping Theorem, we conclude that

$$\widetilde{\mathcal{W}}_\beta^{joint} = \sup_{\pi \in [\pi_1, \pi_2]} \left\{ \mathcal{W}_T^{IVX} + \mathcal{W}_\beta^{IVX}(t) \right\} \Rightarrow B(1)B(1)' + \sup_{\pi \in [\pi_1, \pi_2]} \left\{ \boldsymbol{N}(\pi)' \boldsymbol{M}(\pi)^{-1} \boldsymbol{N}(\pi) \right\}.$$

(ii). In case that $\gamma_x < 1$, by Corollary 3.1 it holds that the second component of the $\mathcal{W}_\beta$ test statistic reduces to a functional of a standard Brownian bridge $B(\pi) - \pi B(1)$. Then, since both $B(1)$ and $B(\pi) - \pi B(1)$ are Gaussian processes which implies that $\text{Cov}(B(1), B(\pi) - \pi B(1)) = \pi - \pi = 0$. Therefore, these two stochastic quantities are independent of each other[2]. Hence, the proof of the statement follows.

[2]To validate the asymptotic independence of the two BM functionals, we can apply properties of the BM and prove that the covariance of the two terms is zero, which ensures independence.

Proof of Proposition 2.1. Under the null hypothesis, $\mathbb{H}_0 : \alpha_1 = \alpha_2$ and $\beta_1 = \beta_2 = \beta$, there is no break in the model intercept. In this section we consider in more details the estimator of the model intercept based on the IVX instrumentation before proving the asymptotic distribution given by Proposition 2.1. In particular, we propose to estimate the model intercept using the generated instrument instead of the predictor, we refer to this estimate as α^{IVZ} and the econometric intuition is explained below. Based on the IVX estimation procedure, the corresponding IVX estimate for the model intercept is given by $\hat{\alpha} = \bar{y}_T - \widetilde{\beta}^{IVX}\bar{x}_{T-1}$. However, we notice that due to the presence of the predictors, then the limit theory for the estimate of the model intercept will vary with the degree of persistence of predictors. We can see this below

$$\sqrt{T}\left(\hat{\alpha}^{IVX} - \alpha\right) = \sqrt{T}\bar{u}_T - \left[T^{\frac{1+\gamma_x \wedge \delta_z}{2}}\left(\widetilde{\beta}^{IVX} - \beta\right)\right]\left[T^{-\frac{\gamma_x \wedge \delta_z}{2}}\bar{x}_{T-1}\right], \tag{A.45}$$

where $\bar{u}_T = \frac{1}{T}\sum_{t=1}^{T} u_t$. Notice that both $\sqrt{T}\bar{u}_T$ and $T^{\frac{1+\gamma_x \wedge \delta_z}{2}}\left(\widetilde{\beta} - \beta\right)$ are both $\mathcal{O}_p(1)$, while the order of convergence of the last term depends on the persistence level of predictors. We have the following convergence rates

$$\sum_{t=1}^{T} x_{t-1} = \begin{cases} \mathcal{O}_p\left(T^{-1/2}\right) & \text{, if } \gamma_x = 1 \\ \\ \mathcal{O}_p\left(T^{1/2+\gamma_x}\right) & \text{, if } 0 < \gamma_x < 1 \\ \\ \mathcal{O}_p\left(T^{3/2}\right) & , \gamma_x > 1. \end{cases} \tag{A.46}$$

We can observe that the term $T^{-\frac{\gamma_x \wedge \delta_z}{2}}\bar{x}_{T-1}$ will dominate in the limit in the case that $\gamma_x > (\delta_z + 1)/2$ and vanish if the reverse holds. In the case that $\gamma_x = (\delta_z + 1)/2$, all three terms appear in the asymptotic distribution which will depend on an unknown parameter β. To simplify the asymptotic theory we need to derive we estimate the model intercept based on the generated endogenous instrument, that is, $\alpha^{IVZ} = \bar{y}_T - \widetilde{\beta}^{IVX}\bar{z}_{T-1}$, where $\bar{z}_{T-1} = \frac{1}{T-1}\sum_{j=2}^{T}\tilde{z}_{j-1}$. The advantage of the IVZ model estimate is that the persistence level of the instrument $\tilde{z}_t$ is controlled by the choice of the tuning parameters δ_z and c_z therefore the abstract degree of persistence is filtered out. Moreover, the particular choice of the estimate for the model intercept works as a power enhancement mechanism against nonzero β values, as seen below

$$\left(\hat{\alpha}^{IVZ} - \alpha\right) = \bar{u}_T - \left(\widetilde{\beta}^{IVX} - \beta\right)\bar{z}_{T-1} + \beta\left(\bar{x}_{T-1} - \bar{z}_{T-1}\right), \tag{A.47}$$

The order of the first term is $\mathcal{O}_p\left(T^{-1/2}\right)$. Moreover, we can show that the second term is asymptotically dominated by the first term, while the third term does not converge unless the parameter space for β is within the neighbourhood of zero. Therefore, power against the presence of predictability when a model intercept is included in the model is achieved when we control the convergence rate of the third term above. Next, we show that the second term of (A.47) is asymptotically dominated by the first term by expanding further the expression

$$\sqrt{T}\left(\widetilde{\beta}^{\text{IVX}} - \beta\right)\bar{z}_{T-1} \equiv \left[T^{\frac{1+\gamma_x \wedge \delta_z}{2}}\left(\widetilde{\beta}^{\text{IVX}} - \beta\right)\right] . \left[T^{-\left(1+\frac{\gamma_x \wedge \delta_z}{2}\right)}\sum_{t=1}^{T}\tilde{z}_{t-1}\right]. \tag{A.48}$$

The first term above is $\mathcal{O}_p(1)$ due to the convergence property of the IVX estimator to the mixed Gaussian distribution as established by Phillips and Magdalinos (2009). To establish the order of the second term we consider the convergence rate of the term $\sum_{t=1}^{T}\tilde{z}_{t-1}$. For instance, by Lemma A2 in the Online Appendix of KMS we have that

$$\sum_{t=1}^{T}\tilde{z}_{t-1} = \begin{cases} \mathcal{O}_p\left(T^{\frac{\gamma_x\wedge 1}{2}+\delta_z}\right) & \text{, if } \delta_z < \gamma_z \\ \\ \mathcal{O}_p\left(T^{\gamma_x+\frac{\delta_z}{2}}\right) & \text{, if } 0 < \gamma_x < \delta_z. \end{cases} \tag{A.49}$$

Hence, we have that the second term of (A.47) will be $\mathcal{O}_p\left(T^{\frac{\gamma_x\wedge 1+\delta_z}{2}+1}\right)$. Since $\gamma_x \wedge 1 + \delta_z < 2$, this implies that the order of the second term is $o_p(1)$. Therefore, we prove that the second term of $\left(\hat{\alpha}^{IVZ} - \alpha\right)$ is asymptotically dominated by the first term. For the third term of (A.47) we aim to show that is asymptotically dominated by the first term when β is small enough.

Thus, we rewrite with $\sqrt{T}\beta\left(\bar{x}_{T-1} - \bar{z}_{T-1}\right) = \beta\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(x_{t-1} - \tilde{z}_{t-1}\right)$. Using the representation formula of the instrument given by expression (23) of Phillips and Magdalinos (2009), we find an equivalent expression, that is, $\beta\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(x_{t-1} - \tilde{z}_{t-1}\right) = -\beta C_z T^{\left(\frac{1}{2}+\delta_z\right)}\sum_{t=1}^{T}\psi_{t-1}$, where $\psi_{t-1} = \sum_{j=1}^{t} R_T^{t-j} x_{j-1}$. Hence, if we set for simplicity $\boldsymbol{C}_z = \boldsymbol{I}_p$ indicating a common degree of persistence across the instruments, we then obtain the following probability bound for this expression

$$\begin{aligned} \left\|\sqrt{T}\beta\left(\bar{x}_{T-1} - \bar{z}_{T-1}\right)\right\| &\leq \|\beta\|\, T^{-\left(\frac{1}{2}+\delta_z\right)}\sum_{t=1}^{T}\|\psi_{t-1}\| \leq \|\beta\|\, T^{-\left(\frac{1}{2}+\delta_z\right)} T \sup_{2\leq t\leq T}\|\psi_{t-1}\| \\ &= \|\beta\|\, T^{\left(\frac{1}{2}-\delta_z\right)}\sup_{2\leq t\leq T}\|\psi_{t-1}\| \\ &\leq \|\beta\|\, T^{\left(\frac{1}{2}-\delta_z\right)}\mathcal{O}_p\left(T^{\frac{\gamma_x}{2}+\delta_z}\right) \\ &= \|\beta\|\,\mathcal{O}_p\left(T^{\frac{\gamma_x}{2}+\delta_z}\right). \end{aligned}$$

Notice that the above result is justified due to the uniform bound of $\|\psi_{t-1}\|$, which is shown to be $\mathcal{O}_p\left(T^{\frac{\gamma_x}{2}+\delta_z}\right)$ by PM. Thus, if $\beta = o_p\left(T^{-\frac{1+\gamma_x}{2}}\right)$, then $\left\|\sqrt{T}\beta\left(\bar{x}_{T-1} - \bar{z}_{T-1}\right)\right\| = o_p(1)$ and the third component of $\left(\tilde{\alpha}^{IVZ} - \alpha\right)$ will also be asymptotically dominated. Therefore, when β is a non-zero constant, this term will dominate, and this is the reason we obtain non-zero local power under the alternative hypothesis of predictability.

Proof of Proposition 2.2.

For the proof of Proposition 2.2, we have already proved most of the required results.The only step we need to additionally show is that the test statistics $\mathcal{W}_a(t)$ and $\mathcal{W}_b(t)$ are asymptotically independent of each other. Notice that since the test statistic $\mathcal{W}_a(t)$ is driven by $\sum_{t=1}^{\lfloor Tr\rfloor} u_t$, while the test statistic $\mathcal{W}_b(t)$ is driven by $\sum_{t=1}^{\lfloor Tr\rfloor}\tilde{z}_{t-1}u_t$. These two partial sums (invariance principles) have a joint weakly convergence to two independent Brownian motions, as shown by Proposition A1 in Phillips and Magdalinos (2009). Hence, the asymptotic independence is guaranteed. Then, convergence of the test statistic $\mathcal{W}_{\alpha\beta}(t)$ follows by an application of CMT.

Appendix B

Supplement to Chapter 3

B.1 Asymptotic Theory

We provide detailed mathematical derivations for the proofs of main asymptotic theory. Relevant references are Phillips and Magdalinos (2009) and Qu (2008).

B.1.1 Proofs of auxiliary theory results

For $\boldsymbol{b} \in \mathbb{R}^p$ we define (see, Lee (2016) and Galvao et al. (2014)) the empirical process:

$$\boldsymbol{G}_n(\tau, \boldsymbol{b}) := n^{-(1+\gamma_x)/2} \sum_{t=1}^{n} \boldsymbol{z}_{t-1} \times \left\{ \psi_\tau(u_t(\tau) - \boldsymbol{x}_{t-1}'\boldsymbol{b}) - \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau) - \boldsymbol{x}_{t-1}'\boldsymbol{b})\right] \right\}$$

such that $(\tau, \boldsymbol{b}) \in \mathcal{T}_\iota \times B \mapsto \boldsymbol{G}_n(\tau, \boldsymbol{b})$ is stochastically equicontinous for any $\epsilon > 0$.

Proof of Corollary 3.2

The consistency of the OLS quantile estimator is derived by Koenker and Xiao (2006). Here, we prove the convergence rate for the IVX-QR estimator by employing standard methods from the literature of extremum estimators and the regularity conditions of Bickel (1975) (see Lee (2016) and Angrist et al. (2006)). Thus, we aim to show that

$$\left(\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau)\right) = \mathcal{O}_{\mathbb{P}}\left(n^{-(1+\gamma_z)/2}\right). \tag{B.1}$$

Proof. For the remaining of the proof we denote with $\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) := \widehat{\boldsymbol{\beta}}_n^\star(\tau)$ to simplify notation. Then, consider the estimator distance such that $\widehat{\boldsymbol{\mathcal{E}}}_n^\star(\tau) = \left(\widehat{\boldsymbol{\beta}}_n^\star(\tau) - \boldsymbol{\beta}^\star(\tau)\right)$, which implies that by evaluating $\boldsymbol{G}_n(\tau, \boldsymbol{b})$ at $\boldsymbol{b} = \widehat{\boldsymbol{\mathcal{E}}}_n^\star(\tau)$ we obtain the following expression

$$\boldsymbol{G}_n(\tau, \boldsymbol{b})\Big|_{\boldsymbol{b}=\widehat{\boldsymbol{\mathcal{E}}}_n^\star(\tau)} = n^{-\frac{(1+\gamma_z)}{2}} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1}\left\{ \psi_\tau(u_t(\tau) - \boldsymbol{x}_{t-1}'\widehat{\boldsymbol{\mathcal{E}}}_n^\star(\tau)) - \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau) - \boldsymbol{x}_{t-1}'\widehat{\boldsymbol{\mathcal{E}}}_n^\star(\tau))\right] \right\}$$

Next, we apply the result[1] given by Lemma 3.2, which implies that for some constant $\mathcal{C}_1$

$$\sup\{ \| \boldsymbol{G}_n(\tau, \boldsymbol{b}) - \boldsymbol{G}_n(\tau, \boldsymbol{0}) \| : \|\boldsymbol{b}\| \leq n^{(1+\gamma_x)/2}\mathcal{C}_1 \} = o_{\mathbb{P}}(1). \tag{B.2}$$

Moreover, for $\boldsymbol{b} = \boldsymbol{0}$ the following empirical process holds

$$\boldsymbol{G}_n(\tau, \boldsymbol{0}) = n^{-\frac{(1+\gamma_z)}{2}} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \Big\{ \psi_\tau\big(u_t(\tau)\big) - \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau\big(u_t(\tau)\big) \right] \Big\} \tag{B.3}$$

Therefore, the argument of Lemma 3.2 is expanded as below

$$\begin{aligned}
\boldsymbol{G}_n(\tau, \boldsymbol{b}) - \boldsymbol{G}_n(\tau, \boldsymbol{0}) &= n^{-\frac{(1+\gamma_z)}{2}} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \{ \psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau)) - \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau)) \right] \} \\
&\quad - n^{-\frac{(1+\gamma_z)}{2}} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \{ \psi_\tau(u_t(\tau)) - \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau)) \right] \} \\
&= n^{-\frac{(1+\gamma_z)}{2}} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau)) \xrightarrow{\mathbb{P}} 0 \\
&\quad - n^{-\frac{(1+\gamma_z)}{2}} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau)) \right] \\
&\quad - n^{-\frac{(1+\gamma_z)}{2}} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \psi_\tau(u_t(\tau)) + n^{-\frac{(1+\gamma_z)}{2}} \sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau)) \right] = 0
\end{aligned}$$

Taking the absolute value since the result holds for the Euclidean norm we obtain

$$\Big| \boldsymbol{G}_n(\tau, \boldsymbol{b}) - \boldsymbol{G}_n(\tau, \boldsymbol{0}) \Big| = n^{-\frac{(1+\gamma_z)}{2}} \sum_{t=1}^{n} \Big\{ \tilde{\boldsymbol{z}}_{t-1} \psi_\tau(u_t(\tau)) + \tilde{\boldsymbol{z}}_{t-1} \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau)) \right] \Big\} + o_{\mathbb{P}}(1).$$

Similarly, with the embedded normalization matrices an equivalent expression holds

$$\Big| \boldsymbol{G}_n(\tau, \boldsymbol{b}) - \boldsymbol{G}_n(\tau, \boldsymbol{0}) \Big| = \sum_{t=1}^{n} \Big\{ \tilde{\boldsymbol{Z}}_{t-1,n} \psi_\tau(u_t(\tau)) + \tilde{\boldsymbol{Z}}_{t-1,n} \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau) - \boldsymbol{X}'_{t-1,n}\widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau)) \right] \Big\} + o_{\mathbb{P}}(1).$$

Then, the conditional expectation $\mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau\left(u_t(\tau) - \boldsymbol{x}'_{t-1}\widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau) \right) \right]$ can be expanded around the point $\boldsymbol{\mathcal{E}}(\tau) = \boldsymbol{0}$ using the first-order taylor expansion.

Since the term $\widehat{\boldsymbol{\beta}}^{\star}(\tau)\boldsymbol{x}_{t-1}$ is strictly monotonic, uniformly on $\{\|\boldsymbol{x}_{t-1}\| \leq \delta\}$ where $\iota \leq \tau \leq 1 - \iota$ for some $\delta > 0$ (see, Theorem 1 of Neocleous and Portnoy (2008)). Hence, we have that

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau\left(u_t(\tau) - \boldsymbol{x}'_{t-1}\widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau) \right) \right] &\equiv \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\boldsymbol{\mathcal{E}}^{\star}_n(\tau)) \right] \Big|_{\boldsymbol{\mathcal{E}}(\tau)=0} \\
&\quad + \frac{\partial \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\boldsymbol{\mathcal{E}}(\tau)) \right]}{\partial \boldsymbol{\mathcal{E}}(\tau)} \Bigg|_{\boldsymbol{\mathcal{E}}(\tau)=0} \times \widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau) + o_{\mathbb{P}}\left(\widehat{\boldsymbol{\mathcal{E}}}^{\star}_n(\tau) \right).
\end{aligned}$$

Note that $\boldsymbol{\mathcal{E}}(\tau) = 0$ implies that $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}^{\star}(\tau)$. Therefore, using the definition of $\psi_\tau(.)$ and by applying standard results for the conditional expectation operator we obtain

$$\mathbb{E}_{\mathcal{F}_{t-1}}\Big[\psi_\tau(u_t(\tau) - \boldsymbol{\mathcal{E}}(\tau)'\boldsymbol{x}_{t-1}) \Big] = \tau - \mathbb{E}_{\mathcal{F}_{t-1}}\Big[\mathbb{1}\{ u_t(\tau) < \boldsymbol{\mathcal{E}}(\tau)'\boldsymbol{x}_{t-1} \} \Big] = \tau - \int_{\infty}^{\boldsymbol{x}'_{t-1}\boldsymbol{\mathcal{E}}(\tau)} f_{u_t(\tau),t-1}(s)ds$$

[1]The norm $\|.\|$ represents the Euclidean norm, i.e., $\|x\| = \left(\sum_{i=1}^{p} x_i^2 \right)^{1/2}$ for $x \in \mathbb{R}^p$.

Hence, by differentiating expression (B.4) around the neighbourhood of $\boldsymbol{\mathcal{E}}(\tau)$ we get

$$\frac{\partial \mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau(u_t(\tau) - \boldsymbol{x}'_{t-1}\boldsymbol{\mathcal{E}}(\tau))\right]}{\partial \boldsymbol{\mathcal{E}}(\tau)}\bigg|_{\boldsymbol{\mathcal{E}}(\tau)=0} = -\boldsymbol{x}'_{t-1} f_{u_t(\tau),t-1}(0) \tag{B.4}$$

Therefore, it holds that

$$\mathbb{E}_{\mathcal{F}_{t-1}}\left[\psi_\tau\left(u_t(\tau) - \boldsymbol{x}'_{t-1}\right)\right] = -\boldsymbol{x}'_{t-1} f_{u_t(\tau),t-1}(0)\widehat{\boldsymbol{\mathcal{E}}}^\star_n(\tau) + o_{\mathbb{P}}(1) \tag{B.5}$$

Next substituting the limit given by (B.5) into the original expansion for the term $\big|\boldsymbol{G}_n(\tau,\boldsymbol{b}) - \boldsymbol{G}_n(\tau,\boldsymbol{0})\big|$ as well as by replacing $\boldsymbol{x}_{t-1}$ with the corresponding embedded normalization matrix $\boldsymbol{X}_{t-1,n}$ we obtain the expression

$$\begin{aligned}
&\bigg|\boldsymbol{G}_n(\tau,\boldsymbol{b}) - \boldsymbol{G}_n(\tau,\boldsymbol{0})\bigg| \\
&\equiv \boldsymbol{K}_{nz}(\tau,\boldsymbol{\beta}^\star(\tau)) - \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\tilde{\boldsymbol{Z}}_{t-1,n}\boldsymbol{X}'_{t-1} n^{-\frac{1+\gamma_z}{2}}\left(\widehat{\boldsymbol{\beta}}^\star_n(\tau) - \boldsymbol{\beta}^\star(\tau)\right) + o_{\mathbb{P}}(1)
\end{aligned} \tag{B.6}$$

where

$$\boldsymbol{K}_{nz}(\tau,\boldsymbol{\beta}^\star(\tau)) := \boldsymbol{D}_n^{-1}\sum_{t=1}^{n}\tilde{\boldsymbol{Z}}_{t-1}\psi_\tau(u_t(\tau)) \tag{B.7}$$

Moreover, we define with

$$\boldsymbol{M}_{nz}(\tau,\boldsymbol{\beta}^\star(\tau)) := \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\tilde{\boldsymbol{Z}}_{t-1,n}\tilde{\boldsymbol{X}}'_{t-1,n} \tag{B.8}$$

Next, by noting that from Lemma 3.2 it holds that $\mathsf{sup}\{\ \|\boldsymbol{G}_n(\tau,\boldsymbol{b}) - \boldsymbol{G}_n(\tau,\boldsymbol{0})\|\ \} = o_{\mathbb{P}}(1)$, then by rearranging expression (B.6) we obtain that

$$n^{\frac{1+\gamma_z}{2}}(\widehat{\boldsymbol{\beta}}^\star_n(\tau) - \boldsymbol{\beta}^\star(\tau)) = \Big[\boldsymbol{M}_{nz}(\tau,\boldsymbol{\beta}^\star(\tau))\Big]^{-1}\boldsymbol{K}_{nz}(\tau,\boldsymbol{\beta}^\star(\tau)) + o_{\mathbb{P}}(1) \tag{B.9}$$

which proves the result given by expression (B.1). In summary, we proved that $\widehat{\boldsymbol{\beta}}^\star_n(\tau)$ is a consistent estimator of $\boldsymbol{\beta}^\star(\tau)$ with convergence rate $\sqrt{n}\,\sqrt[\gamma_z]{n}$ where $\gamma_z \in (0,1)$. Furthermore, using the result given by expression (B.9) we can prove the limit results summarized in Corollary 3.2. Notice that for the proof of Corollary 3.3, we use the fact that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_\tau(u_t(\tau)) \overset{\mathbb{P}}{\to} \mathcal{N}(0,\tau(1-\tau)) \ \text{ for some } \tau \in (0,1), \tag{B.10}$$

since due to the structure of the model the quantile regression induced innovation term $\psi_\tau(u_t(\tau)) \sim \mathsf{mds}(0,\tau(1-\tau))$, i.e., it has a covariance which depends on the quantile τ.

□

Proof of Corollary 3.3

Part *(i)*

We begin by considering the limiting distribution of the functional,

$$\boldsymbol{K}_{nx}\left(\tau, \boldsymbol{\theta}_n^{ols}(\tau)\right) := \boldsymbol{D}_n^{-1}\sum_{t=1}^{n}\boldsymbol{X}_{t-1}\psi_\tau\big(u_t(\tau)\big) \Rightarrow \boldsymbol{K}_x\left(\tau, \boldsymbol{\theta}^{ols}(\tau)\right) \tag{B.11}$$

$$\boldsymbol{K}_x\left(\tau, \boldsymbol{\theta}^{ols}(\tau)\right) \equiv \begin{cases} \begin{bmatrix} B_{\psi_\tau}(1) \\ \displaystyle\int_0^1 \boldsymbol{J}_c(r)dB_{\psi_\tau} \end{bmatrix} & \textit{LUR}, \\ \mathcal{N}\left(\boldsymbol{0}, \tau(1-\tau) \times \begin{bmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{V}_{xx} \end{bmatrix}\right) & \textit{MI.} \end{cases}$$

Proof. **Mildly Integrated:** (MI)

By expanding the expression for $\boldsymbol{K}_{nx}(\tau, \boldsymbol{\theta}_n^{ols}(\tau))$, for the mildly integrated regressors case we obtain the following expression

$$\begin{aligned}
\boldsymbol{K}_{nx}\left(\tau, \boldsymbol{\theta}_n^{ols}(\tau)\right) &= \boldsymbol{D}_n^{-1}\sum_{t=1}^{n}\boldsymbol{X}_{t-1}\psi_\tau\big(u_t(\tau)\big) \\
&= \begin{bmatrix} \frac{1}{\sqrt{n}} & \boldsymbol{0}' \\ \boldsymbol{0} & n^{-\frac{1+\gamma_x}{2}}\boldsymbol{I}_p \end{bmatrix}_{(p+1)\times(p+1)} \times \sum_{t=1}^{n}\begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{x}_{t-1}' \end{bmatrix}_{(p+1)\times n} \times \psi_\tau\big(u_t(\tau)\big) \\
&= \begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_\tau\big(u_t(\tau)\big) \otimes \boldsymbol{1}_{(1\times n)}' \\ n^{-\frac{1+\gamma_x}{2}}\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\psi_\tau\big(u_t(\tau)\big) \otimes \boldsymbol{I}_p \end{bmatrix}_{(p+1)\times n} \Rightarrow \mathcal{N}_{p+1}\left(\boldsymbol{0}, \tau(1-\tau) \times \begin{bmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{V}_{xx} \end{bmatrix}\right).
\end{aligned}$$

Since the two scalar processes $\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_\tau\big(u_t(\tau)\big)\right\}$ and $\left\{n^{-\frac{1+\gamma_x}{2}}\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\psi_\tau\big(u_t(\tau)\big)\right\}$ are uncorrelated with negligible higher-order moment terms, thus mutually independent Gaussian with zero mean and variance 1 and and $\boldsymbol{V}_{xx}$ respectively. Therefore, a joint convergence to a Gaussian random variate holds due to their conditional independence. To see this, consider the following limit

$$n^{-\frac{1+\gamma_x}{2}}\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\psi_\tau\big(u_t(\tau)\big) \otimes \boldsymbol{I}_p \Rightarrow \mathcal{N}\big(\boldsymbol{0}, \tau(1-\tau)\boldsymbol{V}_{xx}\big) \tag{B.12}$$

when $0 < \gamma_x < 1$.

Remark B.1. Further details regarding these derivations can be found in the proof of Theorem 1 in Xiao (2009). Notice that although the proof of Theorem 1 in Xiao (2009) corresponds to the framework of quantile cointegrating regression model, after appropriate modifications we can obtain the limit given by expression (B.12). Furthermore, joint convergence of these two terms holds to a Gaussian random variable with mean vector zero and covariance matrix determined by the covariance of each individual term.

Additionally, the following invariance principle for the corresponding partial sum process holds for the exponent rate $0 < \gamma_x < 1$ such that

$$n^{-\frac{1+\gamma_x}{2}} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1} \psi_\tau(u_t(\tau)) \otimes \boldsymbol{I}_p \Rightarrow \mathcal{N}(\boldsymbol{0}, \tau(1-\tau)\lambda \boldsymbol{V}_{xx}). \tag{B.13}$$

which is useful when deriving the convergence limit for the OLS based functionals.

<u>**Local Unit Root:**</u> (LUR)

$$\begin{aligned}
\boldsymbol{K}_{nx}\left(\tau, \boldsymbol{\theta}_n^{ols}(\tau)\right) &= \boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \boldsymbol{X}_{t-1} \psi_\tau(u_t(\tau)) \\
&= \begin{bmatrix} \frac{1}{\sqrt{n}} & \boldsymbol{0}' \\ \boldsymbol{0} & \frac{1}{n}\boldsymbol{I}_p \end{bmatrix}_{(p+1)\times(p+1)} \times \sum_{t=1}^{n} \begin{bmatrix} \boldsymbol{1}'_{(1\times n)} \\ \boldsymbol{x}_{t-1} \end{bmatrix}_{(p+1)\times n} \times \psi_\tau(u_t(\tau)) \\
&= \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\tau(u_t(\tau)) \otimes \boldsymbol{1}' \\ \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{x}_{t-1} \psi_\tau(u_t(\tau)) \otimes \boldsymbol{I}_p \end{bmatrix}_{(p+1)\times n} \Rightarrow \begin{bmatrix} B_{\psi_\tau}(1) \\ \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix}_{(p+1)\times n}
\end{aligned}$$

where $\boldsymbol{x}_t = \left(\boldsymbol{I}_p + \frac{C_p}{n}\right)\boldsymbol{x}_{t-1} + \boldsymbol{v}_t$, for $1 \leq t \leq n$, since $\boldsymbol{X}_{t-1} = [\boldsymbol{1}\ \boldsymbol{x}'_{t-1}]'$.

Therefore, for deriving the limit result above the following weakly convergence arguments can be applied

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\tau(u_t(\tau)) \Rightarrow B_{\psi_\tau}(1) \quad \text{and} \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor \lambda n \rfloor} \psi_\tau(u_t(\tau)) \Rightarrow B_{\psi_\tau}(\lambda) \tag{B.14}$$

for some $0 < \lambda < 1$ and $\tau \in (0,1)$, where $B_{\psi_\tau}(1)$ is a standard Brownian motion that corresponds to the error function $\psi_\tau(u_t(\tau))$.

Remark B.2. Overall, in the case of LUR regressors (i.e., under high persistence) the convergence to Brownian motion functionals occurs due to the different normalization rates employed for these terms. Furthermore, since the local unit root coefficient is a general case that covers abstract degrees of persistence, then in practise we have convergence to correlated Brownian motions.

Part *(ii)*

Next, we consider the limiting distribution of the functional, $\boldsymbol{L}_{nx}\big(\tau, \boldsymbol{\theta}_n^{ols}(\tau)\big)$ for the two persistence classes separately as explained below.

Local Unit Root: (LUR)

We obtain the following expression

$$
\begin{aligned}
\boldsymbol{L}_{nx}\big(\tau, \boldsymbol{\theta}_n^{ols}(\tau)\big) &\overset{\text{def}}{=} \boldsymbol{D}_n^{-1}\left[\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right]\boldsymbol{D}_n^{-1} \\
&= \begin{bmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & \frac{1}{n}\boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) & \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1}' \\ \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1} & \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}' \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & \frac{1}{n}\boldsymbol{I}_p \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) & \frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1}' \\ \frac{1}{n}\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1}\otimes\boldsymbol{I}_p & \frac{1}{n}\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\otimes\boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & \frac{1}{n}\boldsymbol{I}_p \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) & \frac{1}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1}'\otimes\boldsymbol{I}_p \\ \frac{1}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1}\otimes\boldsymbol{I}_p & \frac{1}{n^2}\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\otimes\boldsymbol{I}_p \end{bmatrix} \\
&\Rightarrow f_{u_t(\tau)}(0) \times \begin{bmatrix} 1 & \int_0^1 \boldsymbol{J}_c(r)' \\ \int_0^1 \boldsymbol{J}_c(r) & \int_0^1 \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)' \end{bmatrix}
\end{aligned}
$$

Since, the following convergence in probability holds

$$
\frac{1}{n}\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \overset{\mathbb{P}}{\to} \mathbb{E}\big[f_{u_t(\tau),t-1}(0)\big] =: f_{u_t(\tau)}(0) \tag{B.15}
$$

Moreover, recall that for the case of LUR regressors it holds that

$$
\begin{aligned}
\boldsymbol{D}_n\big(\widehat{\boldsymbol{\theta}}_n^{qr}(\tau) - \boldsymbol{\theta}^{qr}(\tau)\big) &= \begin{bmatrix} \sqrt{n} & 0 \\ 0 & n\boldsymbol{I}_p \end{bmatrix} \begin{pmatrix} \widehat{\alpha}_n^{qr}(\tau) - \alpha(\tau) \\ \widehat{\boldsymbol{\beta}}_n^{qr}(\tau) - \boldsymbol{\beta}(\tau) \end{pmatrix} \\
&\equiv \left\{\boldsymbol{D}_n^{-1}\left[\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right]\boldsymbol{D}_n^{-1}\right\}^{-1} \times \left\{\boldsymbol{D}_n^{-1}\sum_{t=1}^{n}\boldsymbol{X}_{t-1}\psi_\tau\big(u_t(\tau)\big)\right\} \\
&\Rightarrow \left\{ f_{u_t(\tau)}(0) \times \begin{bmatrix} 1 & \int_0^1 \boldsymbol{J}_c(r)' \\ \int_0^1 \boldsymbol{J}_c(r) & \int_0^1 \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)' \end{bmatrix} \right\}^{-1} \times \begin{bmatrix} B_{\psi_\tau}(1) \\ \int_0^1 \boldsymbol{J}_c(r)dB_{\psi_\tau} \end{bmatrix}
\end{aligned}
$$

which follows by an application of the continuous mapping theorem to the first term of the expression above. Notice that in the case of persistent regressors, the limiting distribution of the normalized quantile OLS based estimator is nonstandard and nonpivotal.

Mildly Integrated: (MI)

We obtain the following expression

$$\begin{aligned}
\boldsymbol{L}_{nx}(\tau, \boldsymbol{\theta}_n^{ols}(\tau)) &\overset{\text{def}}{=} \boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}' \right] \boldsymbol{D}_n^{-1} \\
&= \begin{bmatrix} \frac{1}{n} \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) & \frac{1}{n^{1+\frac{\gamma_x}{2}}} \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \boldsymbol{x}_{t-1}' \otimes \boldsymbol{I}_p \\ \frac{1}{n^{1+\frac{\gamma_x}{2}}} \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \boldsymbol{x}_{t-1} \otimes \boldsymbol{I}_p & \frac{1}{n^{1+\gamma_x}} \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \boldsymbol{x}_{t-1} \boldsymbol{x}_{t-1}' \otimes \boldsymbol{I}_p \end{bmatrix} \\
&\Rightarrow f_{u_t(\tau)}(0) \times \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \qquad \text{(B.16)}
\end{aligned}$$

Since it holds that,

$$\frac{1}{n^{1+\frac{\gamma_x}{2}}} \sum_{t=1}^{n} \boldsymbol{x}_{t-1} \overset{\mathbb{P}}{\to} \mathbf{0} \quad \text{and} \quad \frac{1}{n^{1+\gamma_x}} \sum_{t=1}^{n} \boldsymbol{x}_{t-1} \boldsymbol{x}_{t-1}' \overset{\mathbb{P}}{\to} \boldsymbol{V}_{xx} \quad \text{when } \gamma_x \in (0,1),$$

where the second limit follows by Lemma B3 in the Appendix of Kostakis et al. (2015). Therefore, for the case of mildly integrated regressors in the model we obtain

$$\begin{aligned}
\boldsymbol{D}_n(\widehat{\boldsymbol{\theta}}_n^{qr}(\tau) - \boldsymbol{\theta}^{qr}(\tau)) &= \begin{bmatrix} \sqrt{n} & 0 \\ 0 & n^{\frac{1+\gamma_x}{2}} \boldsymbol{I}_p \end{bmatrix} \begin{pmatrix} \widehat{\alpha}_n^{qr}(\tau) - \alpha(\tau) \\ \widehat{\boldsymbol{\beta}}_n^{qr}(\tau) - \boldsymbol{\beta}(\tau) \end{pmatrix} \\
&\equiv \left\{ \boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}' \right] \boldsymbol{D}_n^{-1} \right\}^{-1} \times \left\{ \boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \boldsymbol{X}_{t-1} \psi_\tau(u_t(\tau)) \right\} \\
&\Rightarrow \left\{ f_{u_t(\tau)}(0) \times \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right\}^{-1} \times \mathcal{N}\left(0, \tau(1-\tau) \times \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right) \\
&= \frac{1}{f_{u_t(\tau)}(0)} \times \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx}^{-1} \end{bmatrix} \times \mathcal{N}\left(0, \tau(1-\tau) \times \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right) \\
&\equiv \mathcal{N}\left(0, \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2} \times \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx}^{-1} \end{bmatrix} \right) \qquad \text{(B.17)}
\end{aligned}$$

by an application of the continuous mapping theorem to the first term. □

Remark B.3. Notice also that the limit results given by Corollary 3.3 can be used to prove the limiting distribution provided by Corollary 3.1 of the paper. This can be done, by employing the trick presented in the proof of Theorem 1 in Xiao (2009). In particular, by linearizing the optimization function in terms of an arbitrary centered quantity $\boldsymbol{D}_n \left(\widehat{\boldsymbol{\theta}}_n(\tau) - \boldsymbol{\theta}(\tau) \right)$. Thus, using the convexity lemma we take the distributional limit of the linearized part and then minimize to get the desired expression as in (B.9).

Proof of Corollary 3.4

Proof. For the IVX based estimation of the quantile regression model, we use the dequantile procedure proposed by Lee (2016). Thus, $y_t(\tau) = y_t - \alpha(\tau) + \mathcal{O}_{\mathbb{P}}(n^{-1/2})$. Furthermore, we employ the following embedded normalization matrices

$$\tilde{\boldsymbol{Z}}_{t-1,n} := \tilde{\boldsymbol{D}}_n \tilde{\boldsymbol{z}}_{t-1} \quad \text{and} \quad \tilde{\boldsymbol{X}}_{t-1,n} := \tilde{\boldsymbol{D}}_n \boldsymbol{x}_{t-1}, \quad \text{where} \ \ \tilde{\boldsymbol{D}}_n := n^{\frac{1+(\gamma_x \wedge \gamma_z)}{2}} \boldsymbol{I}_p \tag{B.18}$$

<u>**Part *(i)***</u>

The limit holds regardless of the stochastic dominance of the exponent rates γ_x and γ_z

$$\begin{aligned}\boldsymbol{K}_{nz}(\tau, \boldsymbol{\beta}_n^{ivx}(\tau)) &:= \sum_{t=1}^{n} \tilde{\boldsymbol{Z}}_{t-1} \psi_\tau(u_t(\tau)) \equiv n^{\frac{1+(\gamma_x \wedge \gamma_z)}{2}} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \psi_\tau(u_t(\tau)) \otimes \boldsymbol{I}_p \\ &\Rightarrow \mathcal{N}(\boldsymbol{0}, \tau(1-\tau) \boldsymbol{V}_{cxz}).\end{aligned} \tag{B.19}$$

<u>**Part *(ii)***</u>

$$\boldsymbol{M}_{nz}(\tau, \boldsymbol{\beta}_n^{ivx}(\tau)) := \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \tilde{\boldsymbol{Z}}_{t-1,n} \tilde{\boldsymbol{Z}}'_{t-1,n} \Rightarrow f_{u_t(\tau)}(0) \times \boldsymbol{V}_{cxz} \tag{B.20}$$

which can be easily shown, since $\sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \tilde{\boldsymbol{z}}'_{t-1} \overset{\mathbb{P}}{\to} \boldsymbol{V}_{cxz}$.

<u>**Part *(iii)***</u>

Then, $\tilde{\boldsymbol{Z}}_{t-1,n} \tilde{\boldsymbol{X}}'_{t-1,n} \equiv \tilde{\boldsymbol{D}}_n \tilde{\boldsymbol{z}}_{t-1} \tilde{\boldsymbol{x}}'_{t-1} \tilde{\boldsymbol{D}}'_n$. Therefore, the limit result follows as below

$$\begin{aligned}\boldsymbol{M}_{nz}(\tau, \boldsymbol{\beta}_n^{ivx}(\tau)) &:= \left[\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \tilde{\boldsymbol{Z}}_{t-1,n} \tilde{\boldsymbol{X}}'_{t-1,n} \right] \equiv \tilde{\boldsymbol{D}}_n \left[\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \tilde{\boldsymbol{z}}_{t-1} \tilde{\boldsymbol{x}}'_{t-1} \right] \tilde{\boldsymbol{D}}'_n \\ &\Rightarrow f_{u_t(\tau)}(0) \times \boldsymbol{\Gamma}_{cxz}\end{aligned}$$

Since, a convergence in probability holds $\frac{1}{n} \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \overset{\mathbb{P}}{\to} \mathbb{E}[f_{u_t(\tau),t-1}(0)] =: f_{u_t(\tau)}(0)$. Moreover, we have that

$$\begin{aligned}&\tilde{\boldsymbol{D}}_n (\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau)) \\ &\equiv \left\{ \tilde{\boldsymbol{D}}_n \left[\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \tilde{\boldsymbol{z}}_{t-1} \boldsymbol{x}'_{t-1} \right] \tilde{\boldsymbol{D}}'_n \right\}^{-1} \times \left\{ \tilde{\boldsymbol{D}}_n \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \psi_\tau(u_t(\tau)) \right\} \\ &\Rightarrow \left\{ f_{u_t(\tau)}(0) \times \boldsymbol{\Gamma}_{cxz} \right\}^{-1} \times \mathcal{N}(\boldsymbol{0}, \tau(1-\tau) \boldsymbol{V}_{cxz}) \\ &\equiv \frac{1}{f_{u_t(\tau)}(0)} \times \boldsymbol{\Gamma}_{cxz}^{-1} \times \mathcal{N}(\boldsymbol{0}, \tau(1-\tau) \boldsymbol{V}_{cxz}) \\ &= \mathcal{N}\left(\boldsymbol{0}, \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2} \boldsymbol{\Gamma}_{cxz}^{-1} \boldsymbol{V}_{cxz} \left(\boldsymbol{\Gamma}_{cxz}^{-1} \right)' \right) \equiv \mathcal{N}\left(\boldsymbol{0}, \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2} \left(\boldsymbol{\Gamma}_{cxz} \boldsymbol{V}_{cxz}^{-1} \boldsymbol{\Gamma}'_{cxz} \right)^{-1} \right).\end{aligned}$$

The above result proves the Gaussian random variable limit given by Corollary 3.2 which holds for both the cases of local unit root and mildly integrated regressors in the quantile predictive regression model. Furthermore, in the case we employ the alternative IVX-QR estimator proposed by Lee (2016) (IVZ estimator); in which case the set of nonstationary regressors, $\boldsymbol{x}_{t-1}$, is replaced by the mildly integrated instruments, $\tilde{\boldsymbol{z}}_{t-1}$, we obtain the following limit result

$$\begin{aligned}\tilde{\boldsymbol{D}}_n(\widehat{\boldsymbol{\beta}}_n^{ivz-qr}(\tau) - \boldsymbol{\beta}(\tau)) &\Rightarrow \left\{ f_{u_t(\tau)}(0) \times \boldsymbol{V}_{cxz} \right\}^{-1} \times \mathcal{N}(\boldsymbol{0}, \tau(1-\tau)\boldsymbol{V}_{cxz}) \\ &= \frac{1}{f_{u_t(\tau)}(0)} \times \boldsymbol{V}_{cxz}^{-1} \times \mathcal{N}(\boldsymbol{0}, \tau(1-\tau)\boldsymbol{V}_{cxz}) \\ &= \mathcal{N}\left(\boldsymbol{0}, \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)} \boldsymbol{V}_{cxz}^{-1} \right) \end{aligned} \tag{B.21}$$

since, $\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \tilde{\boldsymbol{Z}}_{t-1,n} \tilde{\boldsymbol{Z}}_{t-1,n}' \Rightarrow f_{u_t(\tau)}(0) \times \boldsymbol{V}_{cxz}$, which is nuisance-parameter free for both the case of local unit root or mildly integrated regressors in the model. □

Remark B.4. Notice that in a standard time series quantile regression with stationary covariates it holds that the regression $\tau-$quantile is asymptotically normal, with

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}(\tau) \right) \overset{d}{\to} \mathcal{N}(\boldsymbol{0}, \tau(1-\tau)\boldsymbol{D}_1^{-1}(\tau)\boldsymbol{D}_0\boldsymbol{D}_1^{-1}(\tau)). \tag{B.22}$$

with an appropriate defined covariance matrix which is a function of the moments of the underline error distribution (see, Goh and Knight (2009)). Therefore, we can clearly see that under nonstationarity the covariance matrix of the Gaussian random variate is stochastic due to the presence of the nuisance parameter of persistence.

Remark B.5. Notice that the objective function for the setting of the nonstationary quantile predictive regression model, becomes globally convex in the parameter. Hence, the method based on the convexity lemma by Pollard (1991) is applicable. Therefore, our proofs for the asymptotic theory of test of parameter restrictions is based on the asymptotic theory framework proposed by Xiao (2009). Consequently, the limit results for the linear parameter restrictions can be employed when constructing the parameter specific restrictions that correspond to structural break tests.

Alternative IVZ-QR estimator

Following the framework proposed by Lee (2016) we also consider the limiting distribution of the IVX-QR estimator when the original persistent regressors are replaced by the instrumental variables in the optimization function. The particular approach is convenient as it significantly reduces the computational time by avoiding the nonconvex optimization procedure given by expression (3.20) which requires to use a grid search with several local optima. More specifically, we consider

$$\widehat{\gamma}_n^{ivx-qr}(\tau) := \underset{\gamma \in \mathbb{R}^p}{\arg\min} \sum_{t=1}^{n} \rho_\tau \left(y_t(\tau) - \tilde{z}_{t-1}' \gamma \right). \tag{B.23}$$

Corollary B.1. Under the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}(\tau) = 0$, it holds that

$$\tilde{\boldsymbol{D}}_n \left(\widehat{\gamma}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau) \right) \Rightarrow \mathcal{N} \left(\mathbf{0}, \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2} \boldsymbol{V}_{cxz}^{-1} \right) \tag{B.24}$$

both for near unit root and mildly integrated predictors, where $\tilde{\boldsymbol{D}}_n = n^{\frac{1+\gamma_x \wedge \gamma_z}{2}} \boldsymbol{I}_p$.

Lemma B.1. (Self-normalized IVX-QR) Under Assumption 3.1 it holds that,

$$\frac{\widehat{f_{u_t(\tau)}}(0)^2}{\tau(1-\tau)} \left(\widehat{\gamma}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau) \right)' \left(\tilde{\boldsymbol{Z}}' \tilde{\boldsymbol{Z}} \right)^{-1} \left(\widehat{\gamma}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau) \right) \Rightarrow \chi_p^2 \tag{B.25}$$

such that $\widehat{f_{u_t(\tau)}}(0)^2$ is a consistent estimator of $f_{u_t(\tau)}(0)^2$ and p degrees of freedom.

Therefore, Lemma B.1 provides a uniform inference limit result which allows to easily obtain critical values since is nuisance-parameter free. Furthermore, if we are interested to test for example the predictability of a specific subgroup among the predictors, say $\mathcal{H}_0 : \beta_1(\tau) = \beta_2(\tau) = 0$, then the formulation of the Wald statistic with the linear restrictions matrix can be employed. In the particular example, the restrictions matrix takes the form $\boldsymbol{R} = \left[\boldsymbol{I}_2, \mathbf{0}_{2\times(p-2)} \right]$. Then, generalizing the specific example for testing a set of linear restrictions, implies that the null hypothesis is formulated as $\mathcal{H}_0 : \boldsymbol{R}\boldsymbol{\beta}(\tau) = 0$ where $\boldsymbol{R}$ is a $r \times p$ known restriction matrix.

Then, the limiting distribution for the IVX-Wald statistic for the quantile predictive regression is given by the following expression

$$\frac{\widehat{f_{u_t(\tau)}}(0)^2}{\tau(1-\tau)} \left(\boldsymbol{R} \widehat{\gamma}_n^{ivx-qr}(\tau) \right)' \left[\boldsymbol{R} \left(\tilde{\boldsymbol{Z}}' \tilde{\boldsymbol{Z}} \right)^{-1} \boldsymbol{R}' \right]^{-1} \left(\boldsymbol{R} \widehat{\gamma}_n^{ivx-qr}(\tau) \right) \Rightarrow \chi_{p-2}^2$$

where χ_{p-2}^2 denotes the chi-square random variate with $(p-2)$ degrees of freedom such that $\mathbb{P}\left(\chi^2 \geq \chi_{p-2;\alpha}^2 \right) = \alpha$, where $0 < \alpha < 1$ denotes the significance level.

Proof of Lemma C.4

We have that

$$\mathcal{W}_n^{ivx-qr}(\tau) = \frac{\widehat{f_{u_t(\tau)}}(0)^2}{\tau(1-\tau)} \left(\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau)\right)' \left(\boldsymbol{X}'\boldsymbol{P}_{\tilde{\boldsymbol{Z}}}\boldsymbol{X}\right) \left(\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau)\right) \Rightarrow \chi_p^2 \tag{B.26}$$

where

$$\left(\boldsymbol{X}'\boldsymbol{P}_{\tilde{\boldsymbol{Z}}}\boldsymbol{X}\right) := (\boldsymbol{X}'\tilde{\boldsymbol{Z}})\,(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})^{-1}\,(\tilde{\boldsymbol{Z}}'\boldsymbol{X}) \equiv \left(\sum_{t=1}^{n} \boldsymbol{x}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\right)\left(\sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\right)^{-1}\left(\sum_{t=1}^{n} \boldsymbol{x}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\right)'$$

Moreover, we use the embedded normalization matrices such that

$$\left(\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\right) \times \left(\sum_{t=1}^{n} \tilde{\boldsymbol{Z}}_{t-1,n}\boldsymbol{X}'_{t-1,n}\right) \equiv \left(\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\tilde{\boldsymbol{Z}}_{t-1,n}\boldsymbol{X}'_{t-1,n}\right) \xrightarrow{\mathbb{P}} f_{u_t(\tau)}(0)\boldsymbol{\Gamma}_{cxz}$$

and the fact that $\left(\sum_{t=1}^{n} \boldsymbol{z}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\right) \xrightarrow{\mathbb{P}} \boldsymbol{V}_{cxz}$.

Proof. From Corollary 3.2 we have that

$$\begin{aligned}\tilde{\boldsymbol{D}}_n\left(\widehat{\boldsymbol{\beta}}_n^{ivx-qr}(\tau) - \boldsymbol{\beta}(\tau)\right) &= \left(\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\tilde{\boldsymbol{Z}}_{t-1,n}\boldsymbol{X}'_{t-1,n}\right)^{-1}\left(\sum_{t=1}^{n} f_{u_t(\tau),t-1}(0)\tilde{\boldsymbol{Z}}_{t-1,n}\psi_\tau(u_t(\tau))\right)\\ &\Rightarrow \mathcal{N}\left(0, \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2} \times (\boldsymbol{\Gamma}_{cxz}\boldsymbol{V}_{cxz}^{-1}\boldsymbol{\Gamma}'_{cxz})^{-1}\right)\end{aligned}$$

Denote the consistent sample estimator of $\widehat{f_{u_t(\tau)}}(0)$, with $f_{u_t(\tau)}(0)$ where

$$\widehat{f_{u_t(\tau)}}(0) = \sum_{t=1}^{n} f_{u_t(\tau),t-1}(0) \tag{B.27}$$

Therefore, the following asymptotic convergence result follows

$$\left(\boldsymbol{X}'\boldsymbol{P}_{\tilde{\boldsymbol{Z}}}\boldsymbol{X}\right) \Rightarrow [f_{u_t(\tau)}(0)\boldsymbol{\Gamma}_{cxz}] \times \boldsymbol{V}_{cxz}^{-1} \times [f_{u_t(\tau)}(0)\boldsymbol{\Gamma}_{cxz}]' \equiv f_{u_t(\tau)}(0)^2(\boldsymbol{\Gamma}_{cxz}\boldsymbol{V}_{cxz}^{-1}\boldsymbol{\Gamma}'_{cxz})$$

Thus, we obtain

$$\begin{aligned}\mathcal{W}_n^{ivx-qr}(\tau) \Rightarrow \frac{1}{\tau(1-\tau)} &\times \mathcal{N}\left(0, \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2} \times (\boldsymbol{\Gamma}_{cxz}\boldsymbol{V}_{cxz}^{-1}\boldsymbol{\Gamma}'_{cxz})^{-1}\right)\\ &\times \left\{f_{u_t(\tau)}(0)^2 \times (\boldsymbol{\Gamma}_{cxz}\boldsymbol{V}_{cxz}^{-1}\boldsymbol{\Gamma}'_{cxz})\right\}\\ &\times \mathcal{N}\left(0, \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2} \times (\boldsymbol{\Gamma}_{cxz}\boldsymbol{V}_{cxz}^{-1}\boldsymbol{\Gamma}'_{cxz})^{-1}\right)\end{aligned}$$

$$
\begin{aligned}
\mathcal{W}_n^{ivx-qr}(\tau) &\Rightarrow \frac{f_{u_t(\tau)}(0)^2}{\tau(1-\tau)} \frac{\tau(1-\tau)}{f_{u_t(\tau)}(0)^2} \times \left[\mathcal{N}\left(0, \left(\boldsymbol{\Gamma}_{cxz} \boldsymbol{V}_{cxz}^{-1} \boldsymbol{\Gamma}_{cxz}'\right)^{-1}\right) \right]' \\
&\quad \times \left(\boldsymbol{\Gamma}_{cxz} \boldsymbol{V}_{cxz}^{-1} \boldsymbol{\Gamma}_{cxz}'\right)^{-1} \times \left[\mathcal{N}\left(0, \left(\boldsymbol{\Gamma}_{cxz} \boldsymbol{V}_{cxz}^{-1} \boldsymbol{\Gamma}_{cxz}'\right)^{-1}\right) \right] \\
&= \mathcal{N}(0,1) \left[\left(\boldsymbol{\Gamma}_{cxz} \boldsymbol{V}_{cxz}^{-1} \boldsymbol{\Gamma}_{cxz}'\right)^{-1/2} \right]' \times \left(\boldsymbol{\Gamma}_{cxz} \boldsymbol{V}_{cxz}^{-1} \boldsymbol{\Gamma}_{cxz}'\right) \times \left[\left(\boldsymbol{\Gamma}_{cxz} \boldsymbol{V}_{cxz}^{-1} \boldsymbol{\Gamma}_{cxz}'\right)^{-1/2} \right] \mathcal{N}(0,1) \\
&= \left[\mathcal{N}(0,1) \right]^2 \equiv \chi_2^2.
\end{aligned}
$$

which is a standard χ^2-distribution with 2 degrees of freedom. □

Proof of Proposition 3.1

Proof.

Part (*i*)

First we consider the limit expression for the case of mildly integrated regressors

$$
\begin{aligned}
\mathcal{L} &:= \hat{\mathcal{J}}_n\left(\lambda, \tau_0, \widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)\right) - \lambda \hat{\mathcal{J}}_n\left(1, \tau_0, \widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)\right) \\
&= \left(\boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{n} \boldsymbol{X}_{t-1}' \boldsymbol{X}_{t-1} \right] \boldsymbol{D}_n^{-1} \right)^{-1/2} \left\{ \boldsymbol{D}_n^{-1} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1} \psi_\tau\big(u_t(\tau_0)\big) - \lambda \boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \boldsymbol{X}_{t-1} \psi_\tau\big(u_t(\tau_0)\big) \right\} \\
&\Rightarrow \left\{ \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right\}^{-1/2} \left\{ \mathcal{N}\left(\mathbf{0}, \tau_0(1-\tau_0)\lambda \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right) - \lambda \mathcal{N}\left(\mathbf{0}, \tau_0(1-\tau_0) \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right) \right\} \\
&= \sqrt{\tau_0(1-\tau_0)} \left\{ \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right\}^{-1/2} \left\{ \mathcal{N}\left(\mathbf{0}, \lambda \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right) - \lambda \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right) \right\} \\
&= \sqrt{\tau_0(1-\tau_0)} \left\{ \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right\}^{-1/2} \left\{ \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{V}_{xx} \end{bmatrix} \right\}^{1/2} \times \left\{ \mathcal{N}\left(\mathbf{0}, \lambda \boldsymbol{I}_p\right) - \lambda \mathcal{N}\left(\mathbf{0}, \boldsymbol{I}_p\right) \right\} \\
&\equiv \sqrt{\tau_0(1-\tau_0)} \Big[\boldsymbol{W}_p(\lambda) - \lambda \boldsymbol{W}_p(1) \Big].
\end{aligned}
$$

Therefore, for some $0 < \lambda < 1$ and $\tau_0 \in (0,1)$

$$
\mathcal{SQ}_n^{ols}(\lambda; \tau_0) \Rightarrow \sup_{\lambda \in [0,1]} \left\| \boldsymbol{\mathcal{BB}}_p(\lambda) \right\|_\infty. \tag{B.28}
$$

which is a nuisance-parameter free distribution that holds under the null hypothesis. In summary, suppose that the data are generated by the quantile predictive regression model and Assumptions 3.1-3.2 are satisfied. Then, under the null null hypothesis $\mathcal{H}_0^{(A)}$ the fluctuation type statistics weakly converge to the limiting distribution given by expression (B.28) for mildly integrated regressors for some unknown break-point location $0 < \lambda < 1$.

Second, for local unit root regressors (high persistence) then, following limit holds

$$
\begin{aligned}
\mathcal{L} &:= \hat{\mathcal{J}}_n\left(\lambda, \tau_0, \widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)\right) - \lambda \hat{\mathcal{J}}_n\left(1, \tau_0, \widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)\right) \\
&= \left(\boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{n} \boldsymbol{X}_{t-1}' \boldsymbol{X}_{t-1} \right] \boldsymbol{D}_n^{-1} \right)^{-1/2} \left\{ \boldsymbol{D}_n^{-1} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1} \psi_\tau\big(u_t(\tau_0)\big) - \lambda \boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \boldsymbol{X}_{t-1} \psi_\tau\big(u_t(\tau_0)\big) \right\} \\
&\Rightarrow \begin{bmatrix} 1 & \int_0^1 \boldsymbol{J}_c(r)' \\ \int_0^1 \boldsymbol{J}_c(r) & \int_0^1 \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)' \end{bmatrix}^{-1/2} \times \left\{ \begin{bmatrix} B_{\psi_\tau}(\lambda) \\ \int_0^\lambda \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix} - \lambda \begin{bmatrix} B_{\psi_\tau}(1) \\ \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix} \right\} \\
&\equiv \begin{bmatrix} 1 & \int_0^1 \boldsymbol{J}_c(r)' \\ \int_0^1 \boldsymbol{J}_c(r) & \int_0^1 \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)' \end{bmatrix}^{-1/2} \times \begin{bmatrix} B_{\psi_\tau}(\lambda) - \lambda B_{\psi_\tau}(1) \\ \int_0^\lambda \boldsymbol{J}_c(r) dB_{\psi_\tau} - \lambda \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix} \\
&\equiv \begin{bmatrix} 1 & \int_0^1 \boldsymbol{J}_c(r)'_{(1\times p)} \\ \int_0^1 \boldsymbol{J}_c(r)_{(p\times 1)} & \int_0^1 \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)'_{(p\times p)} \end{bmatrix}^{-1/2}_{(p+1)\times(p+1)} \times \begin{bmatrix} \boldsymbol{\mathcal{B}\mathcal{B}}_{\psi_\tau}(\lambda)_{(1\times n)} \\ \boldsymbol{\mathcal{J}\mathcal{B}}_{\psi_\tau}(\lambda)_{(p\times n)} \end{bmatrix}_{(p+1)\times n}
\end{aligned} \tag{B.29}
$$

where $\boldsymbol{\mathcal{B}\mathcal{B}}_{\psi_\tau}(\lambda) := B_{\psi_\tau}(\lambda) - \lambda B_{\psi_\tau}(1)$ and $\boldsymbol{\mathcal{J}\mathcal{B}}_{\psi_\tau}(\lambda) := \int_0^\lambda \boldsymbol{J}_c(r) dB_{\psi_\tau} - \lambda \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau}$. Thus, under the null hypothesis the OLS based functional with local unit root regressors converges into a nonstandard and nonpivotal limiting distribution.

Part (*ii*)

The limit expression of the fluctuation type test based on the IVX estimator is given as

$$
\begin{aligned}
\mathcal{L} &:= \hat{\mathcal{J}}_n\left(\lambda, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivx}(\tau_0)\right) - \lambda \hat{\mathcal{J}}_n\left(1, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivx}(\tau_0)\right) \\
&= \left(\sum_{t=1}^{n} \tilde{\boldsymbol{Z}}_{t-1,n} \boldsymbol{X}_{t-1,n}' \right)^{-1/2} \left\{ \sum_{t=1}^{\lfloor \lambda n \rfloor} \tilde{\boldsymbol{Z}}_{t-1,n} \psi_\tau\big(u_t(\tau_0)\big) - \lambda \sum_{t=1}^{n} \tilde{\boldsymbol{Z}}_{t-1,n} \psi_\tau\big(u_t(\tau_0)\big) \right\} \\
&= \left(\tilde{\boldsymbol{D}}_n^{-1} \left[\sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \boldsymbol{x}_{t-1}' \right] \tilde{\boldsymbol{D}}_n^{-1} \right)^{-1/2} \left\{ \tilde{\boldsymbol{D}}_n^{-1} \sum_{t=1}^{\lfloor \lambda n \rfloor} \tilde{\boldsymbol{z}}_{t-1} \psi_\tau\big(u_t(\tau_0)\big) - \lambda \tilde{\boldsymbol{D}}_n^{-1} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \psi_\tau\big(u_t(\tau_0)\big) \right\} \\
&\Rightarrow \boldsymbol{\Gamma}_{cxz}^{-1/2} \times \left\{ \mathcal{N}\Big(\boldsymbol{0}, \tau_0(1-\tau_0)\lambda \boldsymbol{V}_{cxz}\Big) - \lambda \mathcal{N}\Big(\boldsymbol{0}, \tau_0(1-\tau_0)\boldsymbol{V}_{cxz}\Big) \right\} \\
&= \sqrt{\tau_0(1-\tau_0)} \times \boldsymbol{\Gamma}_{cxz}^{-1/2} \times \left\{ \mathcal{N}\Big(\boldsymbol{0}, \lambda \boldsymbol{V}_{cxz}\Big) - \lambda \mathcal{N}\Big(\boldsymbol{0}, \boldsymbol{V}_{cxz}\Big) \right\} \equiv \sqrt{\tau_0(1-\tau_0)} \big[\boldsymbol{W}_p(\lambda) - \lambda \boldsymbol{W}_p(1) \big].
\end{aligned}
$$

provided that $\boldsymbol{\Gamma}_{cxz} \equiv \boldsymbol{V}_{xx}$, which applies when $\gamma_x \in (0, \gamma_z)$. Proposition 3.1 (***ii***) shows that the limiting distribution of the fluctuation type test is nonstandard in general when the IVX estimator is employed since we employ limit results which hold for both LUR and MI regressors. However, when the coefficient of persistence of regressors has an exponent rate with an absolute value less than the exponent rate of the mildly integrated instruments, then the asymptotic covariance matrix of the Gaussian variant has a simpified form due to the stochastic dominance property of these covariance matrices.

Part (*iii*)

We obtain the following limit result which holds for both LUR and MI regressors

$$
\begin{aligned}
\mathcal{L} &:= \hat{\mathcal{J}}_n^{ivz}\left(\lambda, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0)\right) - \lambda \hat{\mathcal{J}}_n^{ivz}\left(1, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0)\right) \\
&= \left(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}}\right)^{-1/2} \left\{ \tilde{\boldsymbol{D}}_n^{-1} \sum_{t=1}^{\lfloor \lambda n \rfloor} \tilde{\boldsymbol{z}}_{t-1} \psi_\tau \left(y_t - \tilde{\boldsymbol{z}}_{t-1}' \widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0) \right) - \lambda \tilde{\boldsymbol{D}}_n^{-1} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \psi_\tau \left(y_t - \tilde{\boldsymbol{z}}_{t-1}' \widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0) \right) \right\} \\
&= \left(\tilde{\boldsymbol{D}}_n^{-1} \left[\sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \tilde{\boldsymbol{z}}_{t-1}' \right] \tilde{\boldsymbol{D}}_n^{-1} \right)^{-1/2} \left\{ \tilde{\boldsymbol{D}}_n^{-1} \sum_{t=1}^{\lfloor \lambda n \rfloor} \tilde{\boldsymbol{z}}_{t-1} \psi_\tau \big(y_t - \tilde{\boldsymbol{z}}_{t-1}' \boldsymbol{\beta}_0(\tau_0) \big) - \lambda \tilde{\boldsymbol{D}}_n^{-1} \sum_{t=1}^{n} \boldsymbol{x}_{t-1} \psi_\tau \big(y_t - \tilde{\boldsymbol{z}}_{t-1}' \boldsymbol{\beta}_0(\tau_0) \big) \right\} \\
&\Rightarrow \boldsymbol{V}_{cxz}^{-1/2} \times \left\{ \mathcal{N}\Big(\boldsymbol{0}, \tau_0(1-\tau_0) \lambda \boldsymbol{V}_{cxz} \Big) - \lambda \mathcal{N}\Big(\boldsymbol{0}, \tau_0(1-\tau_0) \boldsymbol{V}_{cxz} \Big) \right\} \\
&= \boldsymbol{V}_{cxz}^{-1/2} \times \sqrt{\tau_0(1-\tau_0)} \times \left\{ \mathcal{N}\Big(\boldsymbol{0}, \lambda \boldsymbol{V}_{cxz} \Big) - \lambda \mathcal{N}\Big(\boldsymbol{0}, \boldsymbol{V}_{cxz} \Big) \right\} \\
&= \sqrt{\tau_0(1-\tau_0)}\, \boldsymbol{V}_{cxz}^{-1/2} \times \boldsymbol{V}_{cxz}^{1/2} \times \left\{ \mathcal{N}(\boldsymbol{0}, \lambda \boldsymbol{I}_p) - \lambda \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_p) \right\} \\
&= \sqrt{\tau_0(1-\tau_0)} \Big[\boldsymbol{W}_p(\lambda) - \lambda \boldsymbol{W}_p(1) \Big], \quad \text{since } \boldsymbol{V}_{cxz}^{-1/2} \times \boldsymbol{V}_{cxz}^{1/2} = \boldsymbol{I}_p.
\end{aligned}
$$

which implies that

$$
\frac{1}{\sqrt{\tau_0(1-\tau_0)}} \left[\mathcal{J}_n(\lambda, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0)) - \lambda \mathcal{J}_n(1, \tau_0, \widehat{\boldsymbol{\beta}}_n^{ivz}(\tau_0)) \right] \Rightarrow \frac{\sqrt{\tau_0(1-\tau_0)}}{\sqrt{\tau_0(1-\tau_0)}} \Big[\boldsymbol{W}_p(\lambda) - \lambda \boldsymbol{W}_p(1) \Big]
$$

Thus, for some $0 < \lambda < 1$ and $\tau_0 \in (0,1)$ it holds that

$$
\mathcal{SQ}_n^{ivz}(\lambda; \tau_0) \Rightarrow \sup_{\lambda \in [0,1]} \big\| \boldsymbol{\mathcal{BB}}_p(\lambda) \big\|_\infty \tag{B.30}
$$

Overall, it seems that the fluctuation type statistics are non-pivotal, at the first glance, for all estimators and across the two persistence classes, due to the dependence of their limiting distributions on the nuisance coefficient of persistence; appearing in the estimation of the covariance matrix $\boldsymbol{V}_{cxz}$ of the Gaussian variant that the corresponding partial sum processes converge to. However, the IVZ based statistic for both types of persistence induce an approximation which weakly converges into a Brownian bridge type limit. In practise, when we utilize the IVZ estimator (see also Theorem 3.2 and Proposition 3.2 in Lee (2016)) then the limiting distribution of the moment matrix between the covariates and the residuals of the model simplifies, and thus the overall limit is nuisance-parameter free. Furthermore, a similar limit for the IVX based test hold in the case of mildly integrated regressors. A nonstandard limit distribution appears in the case of high persistence (e.g., LUR) for both the OLS and IVX based test statistics. □

Proof of Proposition 3.2

We need to prove the following limit result

$$\mathcal{SW}_n^{ols}(\lambda;\tau_0) \Rightarrow \sup_{\lambda\in[0,1]} \frac{\left\| \mathcal{BB}_{p+1}(\lambda) \ \right\|^2}{\lambda(1-\lambda)} \tag{B.31}$$

where the exponent rate that captures the degree of persistence for the original regressors can be $\gamma_x = 1$ or $\gamma_x \in (0,1)$. In particular to prove the Brownian Bridge limit, we need to derive the limiting distribution of the term $\mathcal{J}_{nx}(\lambda,\tau_0,\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)) - \lambda\mathcal{J}_{nx}(1,\tau_0,\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0))$.

Using Definition 3.3, for the case of mildly integrated regressors, $\gamma_x \in (0,1)$, it holds that

$$\begin{aligned}\mathcal{J}_{nx}^{ols}(\lambda,\tau_0,\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)) &:= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1/2}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\psi_\tau\left(y_t - \boldsymbol{X}_{t-1}'\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)\right)\\ &\Rightarrow \sqrt{\tau_0(1-\tau_0)} \times \Big[\boldsymbol{W}_{p+1}(\lambda) - \lambda\boldsymbol{W}_{p+1}(1)\Big].\end{aligned} \tag{B.32}$$

where $\boldsymbol{W}_{p+1}(.)$ is a $p-$vector of independent Wiener processes and the convergence holds because $\{\boldsymbol{x}_{t-1}\psi_\tau(y_t - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_0(\tau_0))\}$ is a sequence of martingale differences under the null. The particular limit is derived in the proof of Proposition 3.1.

<u>Part *(i)*</u>

Proof. Let $\boldsymbol{X}_{t-1} = \left(\mathbf{1},\boldsymbol{x}_{t-1}'\right)'$ the design matrix and $\boldsymbol{\theta}(\tau_0) = \left(\alpha(\tau_0),\boldsymbol{\beta}'(\tau_0)\right)'$ to be the parameter vector. Then, to derive the asymptotic distribution of the OLS-Wald test we employ the following functional

$$\hat{S}_{nx}^{ols}(\lambda,\tau_0,\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)) := \boldsymbol{D}_n^{-1}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\psi_\tau\big(u_t(\tau_0)\big), \quad \text{for some} \ \ 0<\lambda<1, \tag{B.33}$$

where $u_t(\tau_0) = \big(y_t - \boldsymbol{X}_{t-1}'\widehat{\boldsymbol{\theta}}_n^{ols}(\tau_0)\big)$ with $\tau_0 \in (0,1)$. Then, the OLS based estimator for the subsample $1 \leq t \leq \lfloor\lambda n\rfloor$ denoted with $\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0)$ is given by

$$\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0) = \left(\frac{1}{k}\sum_{t=1}^{\lfloor\lambda n\rfloor} f_{u_t(\tau),t-1}(0)\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1}\left(\frac{1}{k}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}y_t\right) \tag{B.34}$$

Therefore, it holds that

$$\begin{aligned}\left\|\hat{S}_{nx}^{ols}\left(\lambda,\tau_0,\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0)\right)\right\| &\leq \left\|\ \boldsymbol{D}_n^{-1}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\Big[\mathbb{1}\{y_t = \boldsymbol{X}_{t-1}'\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0)\}\Big]\ \right\|\\ &\leq (p+1)D_n^{-1}\max_{1\leq i\leq n}\|\boldsymbol{X}_{t-1}\| \overset{\mathbb{P}}{\to} 0.\end{aligned} \tag{B.35}$$

which implies that $\hat{S}_{nx}^{ols}\left(\lambda,\tau_0,\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0)\right) = o_{\mathbb{P}}(1)$.

Moreover, for the estimator of the first subsample we obtain the following expression

$$
\begin{aligned}
\boldsymbol{D}_n\left(\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0)-\boldsymbol{\theta}_0(\tau_0)\right) &\overset{p}{\to} \left\{f_{u_t(\tau)}(0)\lambda\mathbb{V}_{xx}\right\}^{-1}\times S_{nx}^{ols}\big(\lambda,\tau_0,\boldsymbol{\theta}_0(\tau_0)\big)+o_{\mathbb{P}}(1)\\
&\equiv \frac{1}{f_{u_t(\tau)}(0)}\frac{1}{\lambda}\mathbb{V}_{xx}^{-1}\times\mathcal{N}\left(\mathbf{0},\tau_0(1-\tau_0)\lambda\begin{bmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{V}_{xx}\end{bmatrix}\right)\\
&= \frac{1}{f_{u_t(\tau)}(0)}\frac{1}{\lambda}\sqrt{\tau_0(1-\tau_0)}\mathbb{V}_{xx}^{-1}\times\begin{bmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{V}_{xx}\end{bmatrix}^{1/2}\times\mathcal{N}(\mathbf{0},\lambda)\\
&= \frac{1}{f_{u_t(\tau)}(0)}\frac{1}{\lambda}\sqrt{\tau_0(1-\tau_0)}\times\mathbb{V}_{xx}^{-1/2}\times\boldsymbol{W}_{p+1}(\lambda). \qquad \text{(B.36)}
\end{aligned}
$$

since it holds that $S_x^{ols}(\lambda,\tau_0,\boldsymbol{\theta}(\tau_0))\equiv\mathcal{N}\left(\mathbf{0},\tau_0(1-\tau_0)\lambda\begin{bmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{V}_{xx}\end{bmatrix}\right)$. Similarly,

$$
\begin{aligned}
\boldsymbol{D}_n\left(\widehat{\boldsymbol{\theta}}_2^{ols}(\lambda;\tau_0)-\boldsymbol{\theta}_0(\tau_0)\right) &\overset{p}{\to} \left\{f_{u_t(\tau)}(0)(1-\lambda)\mathbb{V}_{xx}\right\}^{-1}\times S_{nx}^{ols}\big(\lambda,\tau_0,\boldsymbol{\theta}_0(\tau_0)\big)+o_{\mathbb{P}}(1)\\
&\equiv \frac{1}{f_{u_t(\tau)}(0)}\frac{1}{1-\lambda}\sqrt{\tau_0(1-\tau_0)}\times\mathbb{V}_{xx}^{-1/2}\times\Big[\boldsymbol{W}_{p+1}(1)-\boldsymbol{W}_{p+1}(\lambda)\Big]. \qquad \text{(B.37)}
\end{aligned}
$$

Therefore, combining (B.36) and (B.37) we obtain the following expression

$$
\boldsymbol{D}_n\left[\Delta\widehat{\boldsymbol{\theta}}_n^{ols}(\lambda;\tau_0)\right]\Rightarrow-\frac{1}{f_{u_t(\tau)}(0)}\frac{1}{\lambda(1-\lambda)}\sqrt{\tau_0(1-\tau_0)}\times\mathbb{V}_{xx}^{-1/2}\times\big[\boldsymbol{W}_{p+1}(\lambda)-\lambda\boldsymbol{W}_{p+1}(1)\big].
$$

Moreover, the convergence of the covariance matrix follows as below

$$
\underset{n\to\infty}{\text{plim}}\,\widehat{\boldsymbol{V}}_n^{ols}(\lambda;\tau_0)\equiv\tau_0(1-\tau_0)\times\left\{\underset{n\to\infty}{\text{plim}}\,\widehat{\boldsymbol{V}}_{1n}^{ols}(\lambda;\tau_0)+\underset{n\to\infty}{\text{plim}}\,\widehat{\boldsymbol{V}}_{2n}^{ols}(\lambda;\tau_0)\right\} \qquad \text{(B.38)}
$$

Therefore, it holds that

$$
\begin{aligned}
&\underset{n\to\infty}{\text{plim}}\,\widehat{\boldsymbol{V}}_{1n}^{ols}(\lambda;\tau_0) \qquad \text{(B.39)}\\
&=\left\{\underset{n\to\infty}{\text{plim}}\,\tilde{\boldsymbol{L}}_{nx}\left(\tau_0,\boldsymbol{\theta}_n^{ols}(\tau_0)\right)\right\}^{-1}\left\{\underset{n\to\infty}{\text{plim}}\,\boldsymbol{D}_n^{-1}\left[\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right]\boldsymbol{D}_n^{-1}\right\}\left\{\underset{n\to\infty}{\text{plim}}\,\tilde{\boldsymbol{L}}_{nx}\left(\tau_0,\boldsymbol{\theta}_n^{ols}(\tau_0)\right)\right\}^{-1}\\
&=\left\{\underset{n\to\infty}{\text{plim}}\,\boldsymbol{D}_n^{-1}\left[\sum_{t=1}^{\lfloor\lambda n\rfloor}f_{u_t(\tau),t-1}(0)\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right]\boldsymbol{D}_n^{-1}\right\}^{-1}\left\{\underset{n\to\infty}{\text{plim}}\,\boldsymbol{D}_n^{-1}\left[\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right]\boldsymbol{D}_n^{-1}\right\}\\
&\times\left\{\underset{n\to\infty}{\text{plim}}\,\boldsymbol{D}_n^{-1}\left[\sum_{t=1}^{\lfloor\lambda n\rfloor}f_{u_t(\tau),t-1}(0)\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right]\boldsymbol{D}_n^{-1}\right\}^{-1}\\
&\equiv\left\{f_{u_t(\tau)}(0)\lambda\mathbb{V}_{xx}\right\}^{-1}\times\left\{\lambda\mathbb{V}_{xx}\right\}\times\left\{f_{u_t(\tau)}(0)\mathbb{V}_{xx}\right\}^{-1}\\
&=\frac{1}{f_{u_t(\tau)}(0)^2}\frac{1}{\lambda}\mathbb{V}_{xx}^{-1} \qquad \text{(B.40)}
\end{aligned}
$$

Similarly, we obtain that

$$\underset{n\to\infty}{\text{plim}}\,\widehat{\boldsymbol{V}}^{ols}_{2n}(\lambda;\tau_0) \Rightarrow \frac{1}{f_{u_t(\tau)}(0)^2}\frac{1}{1-\lambda}\mathbb{V}_{xx}^{-1} \tag{B.41}$$

Thus,

$$\underset{n\to\infty}{\text{plim}}\,\widehat{\boldsymbol{V}}^{ols}_{n}(\lambda;\tau_0) \equiv \frac{\tau_0(1-\tau_0)}{f_{u_t(\tau)}(0)^2}\mathbb{V}_{xx}^{-1}\left\{\frac{1}{\lambda}+\frac{1}{1-\lambda}\right\} = \frac{1}{f_{u_t(\tau)}(0)^2}\frac{\tau_0(1-\tau_0)}{\lambda(1-\lambda)}\mathbb{V}_{xx}^{-1} \tag{B.42}$$

which implies that

$$\begin{aligned}
\mathcal{W}^{ols}_n(\lambda;\tau_0) &:= \boldsymbol{D}_n\Big\{\Delta\widehat{\boldsymbol{\theta}}^{ols}_n(\lambda;\tau_0)\Big\}'\Big[\widehat{\boldsymbol{V}}_n(\lambda;\tau_0)\Big]^{-1}\Big\{\Delta\widehat{\boldsymbol{\theta}}^{ols}_n(\lambda;\tau_0)\Big\}\\
&\Rightarrow \frac{1}{f_{u_t(\tau)}(0)^2}\frac{\tau_0(1-\tau_0)}{[\lambda(1-\lambda)]^2}\big[\boldsymbol{W}_{p+1}(\lambda)-\lambda\boldsymbol{W}_{p+1}(1)\big]'\left(\mathbb{V}_{xx}^{-1/2}\right)'\times\left\{f_{u_t(\tau)}(0)^2\frac{\lambda(1-\lambda)}{\tau_0(1-\tau_0)}\mathbb{V}_{xx}\right\}\\
&\quad\times\mathbb{V}_{xx}^{-1/2}\big[\boldsymbol{W}_{p+1}(\lambda)-\lambda\boldsymbol{W}_{p+1}(1)\big]\\
&\equiv \frac{1}{\lambda(1-\lambda)}\big[\boldsymbol{W}_{p+1}(\lambda)-\lambda\boldsymbol{W}_{p+1}(1)\big]'\big[\boldsymbol{W}_{p+1}(\lambda)-\lambda\boldsymbol{W}_{p+1}(1)\big].
\end{aligned} \tag{B.43}$$

Hence, we have that

$$\begin{aligned}
\mathcal{SW}^{ols}_n(\lambda;\tau_0) &:= \sup_{\lambda\in\Lambda_\eta}\boldsymbol{D}_n\left\{\Delta\widehat{\boldsymbol{\theta}}^{ols}_n(\lambda;\tau_0)\right\}'\times\left[\widehat{\boldsymbol{V}}_n(\lambda;\tau_0)\right]^{-1}\times\left\{\Delta\widehat{\boldsymbol{\theta}}^{ols}_n(\lambda;\tau_0)\right\}\\
&\Rightarrow \sup_{\lambda\in\Lambda_\eta}\frac{\big[\boldsymbol{W}_{p+1}(\lambda)-\lambda\boldsymbol{W}_{p+1}(1)\big]'\big[\boldsymbol{W}_{p+1}(\lambda)-\lambda\boldsymbol{W}_{p+1}(1)\big]}{\lambda(1-\lambda)}\\
&\equiv \sup_{\lambda\in\Lambda_\eta}\frac{\big\|\boldsymbol{\mathcal{BB}}_{p+1}(\lambda)\big\|^2}{\lambda(1-\lambda)},\ \text{for } \gamma_x\in(0,1),
\end{aligned} \tag{B.44}$$

where $\boldsymbol{\mathcal{BB}}_{p+1}(\lambda)$ is a Brownian Bridge process, which holds for the case of mildly integrated regressors, that is, $\gamma_x \in (0,1)$ and holds even under the presence of model intercept. Next, we provide of some auxiliary derivations employed for Part ***(i)***. □

Proof.

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^{ols}_1(\lambda;\tau_0) &= \left(\frac{1}{\kappa}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}\right)^{-1}\left(\frac{1}{\kappa}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}y_t\right)\\
&= \left(\frac{1}{\kappa}\sum_{t=1}^{\lfloor\lambda n\rfloor}f_{u_t(\tau),t-1}(0)\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}\right)^{-1}\left(\frac{1}{\kappa}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\big[\boldsymbol{X}'_{t-1}\boldsymbol{\theta}_0(\tau_0)+\psi_\tau(u_t(\tau_0))\big]\right)\\
&= \boldsymbol{\theta}_0(\tau_0)+\left(\frac{1}{\kappa}\sum_{t=1}^{\lfloor\lambda n\rfloor}f_{u_t(\tau),t-1}(0)\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}\right)^{-1}\left(\frac{1}{\kappa}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\psi_\tau(u_t(\tau_0))\right)+o_{\mathbb{P}}(1).
\end{aligned}$$

Thus,

$$\boldsymbol{D}_n\left(\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0)-\boldsymbol{\theta}_0(\tau_0)\right)=\left(\boldsymbol{D}_n^{-1}\left[\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right]\boldsymbol{D}_n^{-1}\right)^{-1}\left(\boldsymbol{D}_n^{-1}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\psi_\tau(u_t(\tau_0))\right)+o_{\mathbb{P}}(1).$$

which implies that

$$\begin{aligned}\boldsymbol{D}_n\left(\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda;\tau_0)-\boldsymbol{\theta}_0(\tau_0)\right)&\Rightarrow\{\lambda\mathbb{V}_{xx}\}^{-1}\times\mathcal{N}(\boldsymbol{0},\tau_0(1-\tau_0)\lambda\mathbb{V}_{xx})\\&\equiv\frac{1}{\lambda}\sqrt{\tau_0(1-\tau_0)}\mathbb{V}_{xx}^{-1}\times\mathbb{V}_{xx}^{1/2}\times\boldsymbol{W}_{p+1}(\lambda)\\&=\frac{1}{\lambda}\sqrt{\tau_0(1-\tau_0)}\times\mathbb{V}_{xx}^{-1/2}\times\boldsymbol{W}_{p+1}(\lambda).\end{aligned}\tag{B.45}$$

Recall that

$$S_{nx}^{ols}(\lambda,\tau_0,\boldsymbol{\theta}_n^{ols}(\tau_0)):=\boldsymbol{D}_n^{-1}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{X}_{t-1}\psi_\tau(u_t(\tau_0)),\quad\text{for }(\lambda,\tau_0)\in(0,1),$$

where $u_t(\tau_0)=\left(y_t-\boldsymbol{X}_{t-1}'\boldsymbol{\theta}_n(\tau_0)\right)$. Thus, $S_{nx}^{ols}(\lambda,\tau_0,\boldsymbol{\theta}_n^{ols}(\tau_0))\Rightarrow\mathcal{N}\left(\boldsymbol{0},\tau_0(1-\tau_0)\lambda\begin{bmatrix}1&\boldsymbol{0}'\\\boldsymbol{0}&\boldsymbol{V}_{xx}\end{bmatrix}\right)$.
Denote with $\mathbb{V}_{xx}:=\begin{bmatrix}1&\boldsymbol{0}'\\\boldsymbol{0}&\boldsymbol{V}_{xx}\end{bmatrix}$, which implies that

$$S_{nx}^{ols}(\lambda,\tau_0,\boldsymbol{\theta}_n^{ols}(\tau_0))\Rightarrow\mathcal{N}(\boldsymbol{0},\tau_0(1-\tau_0)\lambda\mathbb{V}_{xx})\equiv\sqrt{\tau_0(1-\tau_0)}\times\mathbb{V}_{xx}^{-1/2}\times\boldsymbol{W}_{p+1}(\lambda).$$

Similarly, we can prove that

$$\boldsymbol{D}_n\left(\widehat{\boldsymbol{\theta}}_2^{ols}(\lambda;\tau_0)-\boldsymbol{\theta}_0(\tau_0)\right)\Rightarrow\frac{1}{1-\lambda}\sqrt{\tau_0(1-\tau_0)}\times\mathbb{V}_{xx}^{-1/2}\times\left[\boldsymbol{W}_{p+1}(1)-\boldsymbol{W}_{p+1}(\lambda)\right].$$

Practically, the above results can be deduced from Assumption 3.2 (b) such that

$$\sup_{r\in[0,\lambda]}\left|\frac{1}{n^{1+\gamma_x}}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'-r\mathbb{E}[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}']\right|=o_{\mathbb{P}}(1),\ \text{as }n\to\infty.\tag{B.46}$$

where $\mathbb{E}[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}']\equiv\boldsymbol{V}_{xx}$, which implies that $\frac{1}{n^{1+\gamma_x}}\sum_{t=1}^{\lfloor\lambda n\rfloor}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\xrightarrow{p}\lambda\boldsymbol{V}_{xx}$. Similarly,

$$\sup_{r\in(\lambda,1]}\left|\frac{1}{n^{1+\gamma_x}}\sum_{t=\lfloor\lambda n\rfloor+1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'-r\mathbb{E}[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}']\right|=o_{\mathbb{P}}(1),\ \text{as }n\to\infty.\tag{B.47}$$

which implies that $\frac{1}{n^{1+\gamma_x}}\sum_{t=\lfloor\lambda n\rfloor+1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\xrightarrow{p}(1-\lambda)\boldsymbol{V}_{xx}$.

□

Part *(ii)*

Next, we investigate the asymptotic behaviour of the OLS-Wald test statistic in the case of local unit root regressors (i.e., high persistent). To minimize complexity of notation for the derivations of this proof we denote with

$$\mathbb{S}_{xx} := \begin{bmatrix} 1 & \int_0^1 \boldsymbol{J}_c(r)' \\ \int_0^1 \boldsymbol{J}_c(r) & \int_0^1 \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)' \end{bmatrix} \text{ and } \mathbb{S}_{xx}(\lambda) := \begin{bmatrix} \lambda & \int_0^\lambda \boldsymbol{J}_c(r)' \\ \int_0^\lambda \boldsymbol{J}_c(r) & \int_0^\lambda \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)' \end{bmatrix} \tag{B.48}$$

Moreover, it holds that $\left(\boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{n} \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}' \right] \boldsymbol{D}_n^{-1} \right) \Rightarrow \mathbb{S}_{xx}$ and

$$\left(\boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}' \right] \boldsymbol{D}_n^{-1} \right) \Rightarrow \begin{bmatrix} \lambda & \int_0^\lambda \boldsymbol{J}_c(r)' \\ \int_0^\lambda \boldsymbol{J}_c(r) & \int_0^\lambda \boldsymbol{J}_c(r)\boldsymbol{J}_c(r)' \end{bmatrix}, \text{when } \gamma_x = 1. \tag{B.49}$$

Proof. Therefore, for LUR regressors, $\gamma_x = 1$, we have that

$$\boldsymbol{D}_n \left(\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda; \tau_0) - \boldsymbol{\theta}_0(\tau_0) \right) = \left(\boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}' \right] \boldsymbol{D}_n^{-1} \right)^{-1} \left(\boldsymbol{D}_n^{-1} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1}\psi_\tau\big(u_t(\tau_0)\big) \right) + o_{\mathbb{P}}(1)$$

which implies that

$$\boldsymbol{D}_n \left(\widehat{\boldsymbol{\theta}}_1^{ols}(\lambda; \tau_0) - \boldsymbol{\theta}_0(\tau_0) \right) \Rightarrow \mathbb{S}_{xx}^{-1}(\lambda) \times \begin{bmatrix} B_{\psi_\tau}(\lambda) \\ \int_0^\lambda \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix} \tag{B.50}$$

$$\boldsymbol{D}_n \left(\widehat{\boldsymbol{\theta}}_2^{ols}(\lambda; \tau_0) - \boldsymbol{\theta}_0(\tau_0) \right) \Rightarrow \big(\mathbb{S}_{xx}(1) - \mathbb{S}_{xx}(\lambda) \big)^{-1} \times \begin{bmatrix} B_{\psi_\tau}(1) - B_{\psi_\tau}(\lambda) \\ \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau} - \int_0^\lambda \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix} \tag{B.51}$$

Therefore, combining (B.50) and (B.51) we obtain the following expression

$$\begin{aligned}
&\boldsymbol{D}_n \left[\Delta \widehat{\boldsymbol{\theta}}_n^{ols} (\lambda; \tau_0) \right] \\
&\Rightarrow \begin{bmatrix} \left[\mathbb{S}_{xx}(1) - \mathbb{S}_{xx}(\lambda) \right]^{-1} \times \left\{ B_{\psi_\tau}(1) - B_{\psi_\tau}(\lambda) \right\} - \mathbb{S}_{xx}^{-1}(\lambda) \times B_{\psi_\tau}(\lambda) \\ \left[\mathbb{S}_{xx}(1) - \mathbb{S}_{xx}(\lambda) \right]^{-1} \times \left\{ \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau} - \int_0^\lambda \boldsymbol{J}_c(r) dB_{\psi_\tau} \right\} - \mathbb{S}_{xx}^{-1}(\lambda) \times \int_0^\lambda \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix} \\
&\equiv \left[\mathbb{S}_{xx}(1) - \mathbb{S}_{xx}(\lambda) \right]^{-1} \times \begin{bmatrix} B_{\psi_\tau}(1) - B_{\psi_\tau}(\lambda) \\ \int_0^1 \boldsymbol{J}_c(r) dB_{\psi_\tau} - \int_0^\lambda \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix} - \mathbb{S}_{xx}^{-1}(\lambda) \times \begin{bmatrix} B_{\psi_\tau}(\lambda) \\ \int_0^\lambda \boldsymbol{J}_c(r) dB_{\psi_\tau} \end{bmatrix}
\end{aligned} \tag{B.52}$$

Then, the OLS-Wald test for testing the null hypothesis of no parameter instability in the nonstationary quantile predictive regression model at an unknown break-point location $\kappa = \lfloor \lambda n \rfloor$ is given by the following expression

$$\mathcal{W}_n^{ols} (\lambda; \tau_0) := \tau_0 (1 - \tau_0) \boldsymbol{D}_n \left\{ \Delta \widehat{\boldsymbol{\theta}}_n^{ols} (\lambda; \tau_0) \right\}^\prime \left[\widehat{\boldsymbol{V}}_n (\lambda; \tau_0) \right]^{-1} \left\{ \Delta \widehat{\boldsymbol{\theta}}_n^{ols} (\lambda; \tau_0) \right\} \tag{B.53}$$

Furthermore, we study the limiting variance of the OLS-Wald test under the null hypothesis. In particular, the convergence of the covariance matrix follows as below

$$\begin{aligned}
&\underset{n \to \infty}{\text{plim}} \left\{ \widehat{\boldsymbol{V}}_{1n}^{ols} (\lambda; \tau_0) \right\} \\
&\equiv \left\{ \underset{n \to \infty}{\text{plim}} \ \boldsymbol{L}_{nx} \left(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0) \right) \right\}^{-1} \left\{ \underset{n \to \infty}{\text{plim}} \ \boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}^\prime \right] \boldsymbol{D}_n^{-1} \right\} \left\{ \underset{n \to \infty}{\text{plim}} \ \boldsymbol{L}_{nx} \left(\tau_0, \boldsymbol{\theta}_n^{ols}(\tau_0) \right) \right\}^{-1} \\
&= \left\{ \underset{n \to \infty}{\text{plim}} \ \boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{\lfloor \lambda n \rfloor} f_{u_t(\tau), t-1}(0) \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}^\prime \right] \boldsymbol{D}_n^{-1} \right\}^{-1} \times \left\{ \underset{n \to \infty}{\text{plim}} \ \boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}^\prime \right] \boldsymbol{D}_n^{-1} \right\} \\
&\times \left\{ \underset{n \to \infty}{\text{plim}} \ \boldsymbol{D}_n^{-1} \left[\sum_{t=1}^{\lfloor \lambda n \rfloor} f_{u_t(\tau), t-1}(0) \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}^\prime \right] \boldsymbol{D}_n^{-1} \right\}^{-1} \\
&\equiv \left\{ f_{u_t(\tau)}(0) \mathbb{S}_{xx}(\lambda) \right\}^{-1} \times \mathbb{S}_{xx}(\lambda) \times \left\{ f_{u_t(\tau)}(0) \mathbb{S}_{xx}(\lambda) \right\}^{-1} \\
&= \frac{1}{f_{u_t(\tau)}(0)^2} \mathbb{S}_{xx}^{-1}(\lambda).
\end{aligned} \tag{B.54}$$

and

$$\underset{n \to \infty}{\text{plim}} \left\{ \widehat{\boldsymbol{V}}_{2n}^{ols} (\lambda; \tau_0) \right\} \Rightarrow \frac{1}{f_{u_t(\tau)}(0)^2} \left[\mathbb{S}_{xx}(1) - \mathbb{S}_{xx}(\lambda) \right]^{-1}. \tag{B.55}$$

Therefore,

$$\underset{n\to\infty}{\text{plim}}\left\{\widehat{\boldsymbol{V}}_n^{ols}(\lambda;\tau_0)\right\} \Rightarrow \frac{1}{f_{u_t(\tau)}(0)^2}\left\{\mathbb{S}_{xx}^{-1}(\lambda) + \left[\mathbb{S}_{xx}(1) - \mathbb{S}_{xx}(\lambda)\right]^{-1}\right\}. \tag{B.56}$$

We can also simplify further the term $\left\{\underset{n\to\infty}{\text{plim}}\,\widehat{\boldsymbol{V}}_n^{ols}(\lambda;\tau_0)\right\}^{-1} = f_{u_t(\tau)}(0)^2\left[\mathbb{S}_{xx}(\lambda)-\mathbb{S}_{xx}(\lambda)\mathbb{S}_{xx}^{-1}(1)\mathbb{S}_{xx}(\lambda)\right]$. Denote with

$$\boldsymbol{\Delta}_0^{ols}(\lambda;\tau_0) := \left[\mathbb{S}_{xx}(1) - \mathbb{S}_{xx}(\lambda)\right]^{-1} \times \begin{bmatrix} B_{\psi_\tau}(1) - B_{\psi_\tau}(\lambda) \\ \int_0^1 \boldsymbol{J}_c(r)dB_{\psi_\tau} - \int_0^\lambda \boldsymbol{J}_c(r)dB_{\psi_\tau} \end{bmatrix} - \mathbb{S}_{xx}^{-1}(\lambda) \times \begin{bmatrix} B_{\psi_\tau}(\lambda) \\ \int_0^\lambda \boldsymbol{J}_c(r)dB_{\psi_\tau} \end{bmatrix}$$

Then, it follows that

$$\begin{aligned} \mathcal{W}_n^{ols}(\lambda;\tau_0) &:= \boldsymbol{D}_n\left\{\Delta\widehat{\boldsymbol{\theta}}_n^{ols}(\lambda;\tau_0)\right\}'\left[\widehat{\boldsymbol{V}}_n^{ols}(\lambda;\tau_0)\right]^{-1}\left\{\Delta\widehat{\boldsymbol{\theta}}_n^{ols}(\lambda;\tau_0)\right\} \\ &\Rightarrow \boldsymbol{\Delta}_0^{ols}(\lambda;\tau_0)'\left[\boldsymbol{\Sigma}_0^{-1}(\lambda;\tau_0)\right]\boldsymbol{\Delta}_0^{ols}(\lambda;\tau_0) \\ &\equiv f_{u_t(\tau)}(0)^2\left\{\boldsymbol{\Delta}_0^{ols}(\lambda;\tau_0)\right\}'\left[\mathbb{S}_{xx}(\lambda) - \mathbb{S}_{xx}(\lambda)\mathbb{S}_{xx}^{-1}(1)\mathbb{S}_{xx}(\lambda)\right]\left\{\boldsymbol{\Delta}_0^{ols}(\lambda;\tau_0)\right\}. \end{aligned}$$

□

Therefore, the asymptotic distribution of the sup OLS-Wald test statistic is nonpivotal and nonstandard and has the above analytical expression. Furthermore, we verify that a trivial aspect such as the inclusion of a model intercept can complicate the asymptotic theory of the structural break test since the model intercept and the slopes are known to have different rates of convergence.

B.1.2 Asymptotic results on stochastic equicontinuity

The limit results we present in this section are related to stochastic equicontinuity and finite dimensional convergence. In particular, the empirical process $\boldsymbol{G}_n(\tau, \boldsymbol{b})$ is consider stochastically $\varrho-$equicontinuous over $\mathcal{T}_\iota \times B$, such that for any $\epsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{[\delta]} |G_n(\tau_1, \boldsymbol{b}_1) - G_n(\tau_2, \boldsymbol{b}_2)| > \epsilon \right) = 0, \tag{B.57}$$

where $[\delta] := \left\{(\tau_1, \boldsymbol{b}_1), (\tau_2, \boldsymbol{b}_2) \in (\mathcal{T} \times B)^2 : \ \varrho((\tau_1, \boldsymbol{b}_1), (\tau_2, \boldsymbol{b}_2)) < \delta \right\}$.

Assumption B.1. (Assumption 6 in Qu (2008)) There exist constants $\nu > 1$ and $s > 1$ and $K < \infty$ such that for all $0 \leq \mathsf{u} \leq \mathsf{v} \leq 1$ and for all n,

$$n^{-1} \sum_{t=\lfloor \mathsf{u} n \rfloor}^{\lfloor \mathsf{v} n \rfloor} \mathbb{E}\left(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\right)^{\nu} \leq K(\mathsf{v} - \mathsf{u}) \tag{B.58}$$

$$n^{-1} \sum_{t=\lfloor \mathsf{u} n \rfloor}^{\lfloor \mathsf{v} n \rfloor} \mathbb{E}\left(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\right)^{\nu} \leq K(\mathsf{v} - \mathsf{u})^{s} \tag{B.59}$$

Lemma B.2. (Lemma A1 in Qu (2008)) Let $\mathcal{K} = [0,1] \times [0,1]$ be a parameter set with a metric $d(.,.)$ defined as

$$\varrho(\{\lambda_1, \tau_1\}, \{\lambda_2, \tau_2\}) = |\lambda_2 - \lambda_1| + |\tau_2 - \tau_1|. \tag{B.60}$$

Then, it can be proved that the process $S_n(\lambda, \tau)$ is stochastically equicontinuous on $(\mathcal{K}, d)$. That is, for any $\epsilon > 0$, $\zeta > 0$, there exists a $\delta > 0$ such that for any large n,

$$\mathbb{P}\left(\sup_{[\mathcal{A}]} \left\| S_n\left(\lambda_1, \tau_1, \widehat{\boldsymbol{\beta}}_0(\tau_1)\right) - S_n\left(\lambda_2, \tau_2, \widehat{\boldsymbol{\beta}}_0(\tau_2)\right) \right\| > \zeta \right) < \epsilon. \tag{B.61}$$

where $[\mathcal{A}] := \left\{ (\mathsf{s}_1, \mathsf{s}_2) \in \mathcal{K} \text{ such that } \mathsf{s}_1 = \{\lambda_1, \tau_1\}, \mathsf{s}_2 = \{\lambda_2, \tau_2\} \text{ and } \varrho(\mathsf{s}_1, \mathsf{s}_2) < \delta \right\}$.

Proof. For any given $\tau \in (0,1)$, say $\tau \equiv \tau_0$ we have that

$$\begin{aligned} S_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau_0)) &= n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\left[\tau_0 - \mathbb{1}\{y_t - \boldsymbol{\beta}_0(\tau_0)'\boldsymbol{x}_{t-1} \leq 0\}\right] \\ &= n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\left[\tau_0 - \mathbb{1}\{F_{y|x}(y_t) \leq \tau_0\}\right] \end{aligned}$$

where $F_{y|x}(.)$ is the conditional distribution function of y_t. Notice that the last equality follows since Assumption 3.2 implies $F_{y|x}(.)$ is absolute continuous and strictly increasing almost everywhere in the support of the function.

Define $w_t = F_{y|x}(y_t)$, then w_t follows a uniform distribution such that $w_t \sim \mathsf{Unif}[0,1]$. Hence,

$$S_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau_0)) = n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1} \big[\tau_0 - \mathbb{1}\{ w_t \leq \tau_0 \} \big]. \tag{B.62}$$

Moreover, since the term $\boldsymbol{x}_{t-1} \big[\tau_0 - \mathbb{1}\{ w_t \leq \tau_0 \} \big]$ is a sequence of vector martingale differences by construction, then standard martingale convergence results also apply. □

Next, we aim to prove that under the assumption of nonstationarity for the model regressors, then the corresponding matrix moments are well-defined.

Lemma B.3. (Lemma A2 in Qu (2008)) Let D be an arbitrary compact set in $\mathbb{R}^p$. Under Assumption B.1, it holds that

$$\sup_{\tau \in \mathcal{T}_{\iota}^*} \sup_{\lambda \in [0,1]} \sup_{\xi \in D} \big\| \widetilde{S}_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau_0) + n^{-1/2} \boldsymbol{\xi}) - \widetilde{S}_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau_0)) \big\| = o_{\mathbb{P}}(1), \tag{B.63}$$

where $\mathcal{T}_{\iota}^* = [\iota_1, \iota_2]$ with $0 < \iota_1 < \iota_2 < 1$.

Proof. The proof proceeds along similar lines as the proof of Theorem A3 in Bai (1996). Without loss of generality, we assume that the components of the near unit root process denoted with $\boldsymbol{x}_t$ are non-negative. However, since this a strong assumption and could be violated when considering that regressors are generated as near unit root processes, we proceed using a time series decomposition to account for possible nonnegative values (see, Qu (2008) and Bai (1996)).

Consider the $k-$th predictor such that $k \in \{1, ..., p\}$ and denote with $\boldsymbol{x}_{t,j}^{(k)}$ the $j-$th component of the $k-$th predictor, $\boldsymbol{x}_t^{(k)}$. Then, we can use the following decomposition

$$\boldsymbol{x}_{t,j}^{(k)} = \left(\boldsymbol{x}_{t,j}^{(k)} \right)^+ - \left(\boldsymbol{x}_{t,j}^{(k)} \right)^- \equiv \boldsymbol{x}_{t,j}^{(k)} \mathbb{1} \left\{ \boldsymbol{x}_{t,j}^{(k)} \geq 0 \right\} - \boldsymbol{x}_{t,j}^{(k)} \mathbb{1} \left\{ \boldsymbol{x}_{t,j}^{(k)} < 0 \right\}. \tag{B.64}$$

Then $\left(\boldsymbol{x}_{t,j}^{(k)} \right)^+$ and $\left(\boldsymbol{x}_{t,j}^{(k)} \right)^-$ are nonnegative and satisfy the required assumptions that ensure the monotonicity of the distribution function. Therefore, similarly to the above derivations, we assume that these results hold for the corresponding lagged predictor denoted with $\boldsymbol{x}_{t-1,j}^{(k)}$. Under the nonnegativity assumption it holds that

$$\boldsymbol{x}_{t-1} \mathbb{1} \big(y_t \leq \boldsymbol{\beta}_0(\tau)' \boldsymbol{x}_{t-1} + n^{-1/2} \boldsymbol{\xi}' \boldsymbol{x}_{t-1} \big) \ \text{ and } \ F_{y|x} \big(\boldsymbol{\beta}_0(\tau)' \boldsymbol{x}_{t-1} + n^{-1/2} \boldsymbol{\xi}' \boldsymbol{x}_{t-1} \big) \tag{B.65}$$

are nondecreasing in τ. We partition the compact set $\mathcal{T}_{\iota}$ into $N(\epsilon_n)$ intervals of equal length such that $\omega_1 = \tau_{(0)} < \tau_{(1)} < ... < \tau_{N(\epsilon_n)} = \omega_2$.

Suppose that $\tau \in [\tau_{j-1}, \tau_j]$ then

$$\begin{aligned}
&\widetilde{S}_n(\lambda, \tau, \boldsymbol{\beta}_0(\tau) + n^{-1/2}\boldsymbol{\xi}) - S_n(\lambda, \tau, \boldsymbol{\beta}_0(\tau)) \\
&\quad \leq \widetilde{S}_n(\lambda, \tau_j, \boldsymbol{\beta}_0(\tau_j) + n^{-1/2}\boldsymbol{\xi}) - \widetilde{S}_n(\lambda, \tau_{j-1}, \boldsymbol{\beta}_0(\tau_{j-1})) \\
&\quad + n^{-1/2}\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\{\tau_j - \tau_{j-1}\} \\
&\quad + n^{-1/2}\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\Big\{F_{y|x}(\boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1} + n^{-1/2}\boldsymbol{\xi}'\boldsymbol{x}_{t-1}) - F_{y|x}(\boldsymbol{\beta}_0(\tau_{j-1})'\boldsymbol{x}_{t-1} + n^{-1/2}\boldsymbol{\xi}'\boldsymbol{x}_{t-1})\Big\}
\end{aligned}$$

Similarly, we have that

$$\begin{aligned}
&\widetilde{S}_n(\lambda, \tau, \boldsymbol{\beta}_0(\tau) + n^{-1/2}\boldsymbol{\xi}) - S_n(\lambda, \tau, \boldsymbol{\beta}_0(\tau)) \\
&\quad \geq \widetilde{S}_n(\lambda, \tau_{j-1}, \boldsymbol{\beta}_0(\tau_{j-1}) + n^{-1/2}\boldsymbol{\xi}) - \widetilde{S}_n(\lambda, \tau_j, \boldsymbol{\beta}_0(\tau_j)) \\
&\quad + n^{-1/2}\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\{\tau_{j-1} - \tau_j\} \\
&\quad + n^{-1/2}\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\Big\{F_{y|x}(\boldsymbol{\beta}_0(\tau_{j-1})'\boldsymbol{x}_{t-1} + n^{-1/2}\boldsymbol{\xi}'\boldsymbol{x}_{t-1}) - F_{y|x}(\boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1} + n^{-1/2}\boldsymbol{\xi}'\boldsymbol{x}_{t-1})\Big\}.
\end{aligned}$$

Hence, we have that

$$\begin{aligned}
&\|\widetilde{S}_n(\lambda, \tau, \boldsymbol{\beta}_0(\tau) + n^{-1/2}\boldsymbol{\xi}) - S_n(\lambda, \tau, \boldsymbol{\beta}_0(\tau))\| \\
&\leq \|\widetilde{S}_n(\lambda, \tau_j, \boldsymbol{\beta}_0(\tau_j) + n^{-1/2}\boldsymbol{\xi}) - S_n(\lambda, \tau_{j-1}, \boldsymbol{\beta}_0(\tau_{j-1}))\| \\
&\quad + \|S_n^d(\lambda, \tau_{j-1}, \boldsymbol{\beta}_0(\tau_{j-1}) + n^{-1/2}\boldsymbol{\xi}) - S_n(\lambda, \tau_j, \boldsymbol{\beta}_0(\tau_j))\| \\
&\quad + \left\| n^{-1/2}\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\{\tau_j - \tau_{j-1}\} \right\| \\
&\quad + \left\| n^{-1/2}\sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1}\Big\{F_{y|x}(\boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1} + n^{-1/2}\boldsymbol{\xi}'\boldsymbol{x}_{t-1}) - F_{y|x}(\boldsymbol{\beta}_0(\tau_{j-1})'\boldsymbol{x}_{t-1} + n^{-1/2}\boldsymbol{\xi}'\boldsymbol{x}_{t-1})\Big\} \right\| \\
&\equiv (a) + (b) + (c) + (d).
\end{aligned}$$

Therefore, to complete the proof it is sufficient to show that (a), (b), (c), and (d) are $o_p(1)$ uniformly in $\tau \in \mathcal{T}_\iota^*, \lambda \in [0,1]$ and $\boldsymbol{\xi} \in D$. In particular for term (d), we follow the argument presented by Lemma 2.1 of Koul (1991) (see also Qu (2008)). Notice that similar arguments based on the monotonicity property of the check function are also employed in the proofs of Escanciano and Velasco (2010).

$$\max_{1\leq j\leq N(\epsilon_n)} \sup_{\lambda\in[0,1]} \sup_{\xi\in D} \left\|(c)\right\|$$

$$\max_{1\leq j\leq N(\epsilon_n)} \sup_{\lambda\in[0,1]} \sup_{\xi\in D} \left\| n^{-1} \sum_{t=1}^{\lfloor\lambda n\rfloor} \left[f_t(\boldsymbol{b}(\tau_j)'\boldsymbol{x}_{t-1}) - f_t(\boldsymbol{b}(\tau_{j-1})'\boldsymbol{x}_{t-1})\right]\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\xi\right\| + o_{\mathbb{P}}(1)$$

$$\begin{aligned}
&\leq 2 \max_{1\leq j\leq N(\epsilon_n)} \sup_{\lambda\in[0,1]} \sup_{\xi\in D} \left\| n^{-1} \sum_{t=1}^{\lfloor\lambda n\rfloor} \left[f_t(\boldsymbol{b}(\tau_j)'\boldsymbol{x}_{t-1}) - f_t(\boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1})\right]\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\xi\right\| \\
&+ \max_{1\leq j\leq N(\epsilon_n)} \sup_{\lambda\in[0,1]} \sup_{\xi\in D} \left\| n^{-1} \sum_{t=1}^{\lfloor\lambda n\rfloor} \left[f_t(\boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1}) - f_t(\boldsymbol{\beta}_0(\tau_{j-1})'\boldsymbol{x}_{t-1})\right]\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\xi\right\| \\
&+ o_{\mathbb{P}}(1),
\end{aligned} \tag{B.66}$$

where $\boldsymbol{b}(\tau_k)$ for $k = j-1$ and j is some vector between $\boldsymbol{\beta}_0(\tau_k)$ and $\boldsymbol{\beta}_0(\tau_k) + n^{-1/2}\xi$ and the first inequality follows from the mean value theorem and $\tau_j - \tau_{j-1} \leq n^{-1/2-d}$.

Furthermore, notice that expression (B.66) $= o_{\mathbb{P}}(1)$ if the following holds

$$\max_{1\leq j\leq N(\epsilon_n)} \max_{1\leq t\leq n} \left\| f_t(\boldsymbol{b}(\tau_j)'\boldsymbol{x}_{t-1}) - f_t(\boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1})\right\| = o_{\mathbb{P}}(1) \tag{B.67}$$

and

$$\max_{1\leq j\leq N(\epsilon_n)} \max_{1\leq t\leq n} \left\| f_t(\boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1}) - f_t(\boldsymbol{\beta}_0(\tau_{j-1})'\boldsymbol{x}_{t-1})\right\| = o_{\mathbb{P}}(1). \tag{B.68}$$

Notice that expression (B.67) holds because $f_t(s)$ is uniformly continuous in s for all t

$$\max_{1\leq j\leq N(\epsilon_n)} \max_{1\leq t\leq n} \left\| \boldsymbol{b}(\tau_j)'\boldsymbol{x}_{t-1} - \boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1}\right\| = \max_{1\leq t\leq n} \|\boldsymbol{x}_{t-1}'\|\mathcal{O}_{\mathbb{P}}\left(n^{-1/2}\right) = o_{\mathbb{P}}(1). \tag{B.69}$$

Furthermore, for expression (B.68) it holds that

$$\left[\boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1} - \boldsymbol{\beta}_0(\tau_{j-1})'\boldsymbol{x}_{t-1}\right] = \frac{\tau_j - \tau_{j-1}}{f_t(\mathsf{z}_t)} = \mathcal{O}_{\mathbb{P}}(\tau_j - \tau_{j-1}) = \mathcal{O}_{\mathbb{P}}\left(n^{-1/2-d}\right), \tag{B.70}$$

where the first equality follows from the mean value theorem with $\boldsymbol{\beta}_0(\tau_{j-1})'\boldsymbol{x}_{t-1} \leq \mathsf{z}_t \leq \boldsymbol{\beta}_0(\tau_j)'\boldsymbol{x}_{t-1}$, and the second equality follows because $f_t(.)$ is bounded away from 0 for all i. Furthermore, notice that D is a compact set, for any given $\delta > 0$, D can always be partitioned into a finite number of subsets, such that the diameter of each subset is less than or equal to δ.

Hence,

$$(f) := n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} f_t(\boldsymbol{\beta}_0(\tau)\boldsymbol{x}_{t-1})\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1} + o_{\mathbb{P}}(1), \tag{B.71}$$

uniformly in $\|\delta\| \leq K$ and $1 \leq t \leq n$.

Then, Assumption 3 (b) implies that $(f) = \lambda \boldsymbol{H}_0 \delta + o_{\mathbb{P}}(1)$. Combing the above results we have that

$$\widetilde{S}_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau) + n^{-1/2}\boldsymbol{\xi}) - S_n^d(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau_0)) \tag{B.72}$$

which holds uniformly in $\|\delta\| \leq K$ and $1 \leq t \leq n$. This completes the proof.

□

Proof of Lemma 2

Recall that in Lemma 2 of Qu (2008) we have that

$$S_n(\lambda, \tau_0, \boldsymbol{b}) = n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1} \times \left[\tau_0 - \mathbb{1}\{ y_t - \boldsymbol{b}'\boldsymbol{x}_{t-1} \leq 0 \} \right].$$

and specifically for $\boldsymbol{b} = \boldsymbol{\beta}_0(\tau_0)$ we have that

$$S_n(\lambda, \tau_0, \boldsymbol{\beta}_0(\tau_0)) = n^{-1/2} \sum_{t=1}^{\lfloor \lambda n \rfloor} \boldsymbol{x}_{t-1} \times \left[\mathbb{1}\{ y_t - \boldsymbol{\beta}_0(\tau_0)'\boldsymbol{x}_{t-1} \leq 0 \} - \tau_0 \right].$$

Notice that the crucial step for this proof is to show that

$$\sqrt{n} \left(\widehat{\boldsymbol{\beta}}_1(\lambda; \tau_0) - \boldsymbol{\beta}_0(\tau_0) \right) = \mathcal{O}_{\mathbb{P}}(1), \tag{B.73}$$

uniformly in $(\lambda, \tau) \in \Lambda_\eta \times \mathcal{T}_\iota$. In other words, we need to derive a similar convergence type result in the case we employ the IVX estimator

$$\widehat{\boldsymbol{\ell}}_n(\lambda; \tau_0) := n^{\frac{1+\gamma_x}{2}} \left(\widehat{\boldsymbol{\beta}}_1^{ivx-qr}(\lambda; \tau_0) - \boldsymbol{\beta}_0(\tau_0) \right) = \mathcal{O}_{\mathbb{P}}(1) \tag{B.74}$$

Notice that we have already proved the convergence in probability result given by expression (B.74) with Corollary 3.2 of the paper. Therefore we need to modify accordingly the remaining steps of the proof presented by Qu (2008) to account for the nonstationarity structure of the quantile regression model. Then, the following stochastic convergence result holds

$$\left\| S_n(\lambda, \tau_0, \widehat{\boldsymbol{\beta}}_1(\lambda, \tau)) \right\|_{L_1} = o_{\mathbb{P}}(1), \tag{B.75}$$

if for any $\epsilon > 0$ there exists a $K_0 > 0, N_0 > 0$ and $\eta > 0$, such that if $\|\sqrt{n}\,(\boldsymbol{\beta}^*(\tau) - \boldsymbol{\beta}_0(\tau))\| > K_0$,

$$\mathbb{P}\left(\inf_{\tau \in \mathcal{T}_\iota} \inf_{\lambda \in \Lambda_\eta} \| S_n(\lambda, \tau_0, \boldsymbol{\beta}^*(\tau) \| < \eta \right) < \epsilon \ \text{ for all } n > N_0. \tag{B.76}$$

B.1.3 Asymptotic results on stochastic integrals

In this section we summarize main invariance principles employed for deriving some of the theoretical results of the paper. Extensive details on these results can be found in the framework proposed by Phillips and Magdalinos (2009).

Lemma B.4. Let $\boldsymbol{V}_{xx} := \int_0^\infty e^{r\boldsymbol{C}_p}\boldsymbol{\Omega}_{xx}e^{r\boldsymbol{C}_p}dr$ where $\boldsymbol{\Omega}_{xx}$ is the long-run covariance of the error term $\boldsymbol{v}_t$. Then, under the null hypothesis of no structural break in the predictive regression model the following large sample theory holds:

(i) the sample covariance weakly convergence to the following limit (see Corollary 3.4)

$$\frac{1}{\sqrt{\tau_0(1-\tau_0)}}\tilde{\boldsymbol{D}}_n^{-1}\sum_{t=1}^{\lfloor \lambda n \rfloor}\tilde{\boldsymbol{z}}_{t-1}\psi_\tau(u_t(\tau_0)) \Rightarrow \boldsymbol{\mathcal{U}}_p(\lambda) \tag{B.77}$$

where $\boldsymbol{\mathcal{U}}_p(.)$ is a Brownian motion with variance $\boldsymbol{V}_{cxz}$ as defined below

$$\boldsymbol{V}_{cxz} \equiv \begin{cases} \boldsymbol{V}_{zz} = \displaystyle\int_0^\infty e^{r\boldsymbol{C}_z}\boldsymbol{\Omega}_{xx}e^{r\boldsymbol{C}_z}dr & , \text{ when } 0<\gamma_z<\gamma_x<1, \\ \boldsymbol{V}_{xx} = \displaystyle\int_0^\infty e^{r\boldsymbol{C}_p}\boldsymbol{\Omega}_{xx}e^{r\boldsymbol{C}_p}dr & , \text{ when } 0<\gamma_x<\gamma_z<1. \end{cases} \tag{B.78}$$

(ii) the sample covariance weakly convergence to the following limit

$$\tilde{\boldsymbol{D}}_n^{-1}\left[\sum_{t=1}^{\lfloor \lambda n \rfloor}\boldsymbol{x}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\right]\tilde{\boldsymbol{D}}_n^{-1} \Rightarrow \equiv \boldsymbol{\Gamma}_{cxz}(\lambda) \tag{B.79}$$

where the exact analytic form of $\boldsymbol{\Psi}(\lambda)$ depends on which of the two exponents rates of persistence stochastically dominates such as

$$\boldsymbol{\Gamma}_{cxz}(\lambda) := \begin{cases} -\boldsymbol{C}_z^{-1}\left(\lambda\boldsymbol{\Omega}_{xx} + \displaystyle\int_0^\lambda \boldsymbol{J}_c^\mu(r)d\boldsymbol{J}'_c\right) & , \text{ when } \gamma_x = 1 \\ -\lambda\boldsymbol{C}_z^{-1}\Big(\boldsymbol{\Omega}_{xx} + \boldsymbol{C}_p\boldsymbol{V}_{xx}\Big) & , \text{ when } 0<\gamma_z<\gamma_x<1 \\ \lambda\displaystyle\int_0^\infty e^{r\boldsymbol{C}_p}\boldsymbol{\Omega}_{xx}e^{r\boldsymbol{C}_p}dr \equiv \lambda\boldsymbol{V}_{xx} & , \text{ when } 0<\gamma_x<\gamma_z<1 \end{cases}$$

where $\boldsymbol{B}_p(.)$ is a $p-$dimensional standard Brownian motion, $\boldsymbol{J}_c(\lambda) = \int_0^\lambda e^{(\lambda-s)\boldsymbol{C}_p}d\boldsymbol{B}(s)$ is an *Ornstein-Uhkenbeck* process and we denote with $\boldsymbol{J}_c^\mu(\lambda) = \boldsymbol{J}_c(\lambda) - \int_0^1 \boldsymbol{J}_c(s)ds$ and $\boldsymbol{B}_p^\mu(\lambda) = \boldsymbol{B}(\lambda) - \int_0^1 \boldsymbol{B}(s)ds$ the demeaned processes of $\boldsymbol{J}_c(\lambda)$ and $\boldsymbol{B}_p(\lambda)$ respectively.

(iii) The weakly joint convergence result applies and the asymptotic terms given by expressions in *(i)* and *(ii)* are stochastically independent.

Proof. We present the main conjectures for deriving the invariance principles presented by Lemma B.4 (see, Phillips and Magdalinos (2009) and Kostakis et al. (2015) for details)

$$n^{-(1+\gamma_z)} \sum_{t=1}^{n} \tilde{\boldsymbol{z}}_{t-1} \tilde{\boldsymbol{z}}_{t-1}' \xrightarrow{\mathbb{P}} \boldsymbol{V}_{zz} := \int_0^{\infty} e^{r\boldsymbol{C}_z} \boldsymbol{\Omega}_{xx} e^{r\boldsymbol{C}_z} dr \tag{B.80}$$

Moreover, we have the weakly convergence result from Phillips and Magdalinos (2009):

$$n^{-\frac{1+\gamma_z}{2}} \sum_{t=1}^{n} \left(\tilde{\boldsymbol{z}}_{t-1} \otimes \boldsymbol{v}_t \right) \Rightarrow \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{V}_{zz} \otimes \boldsymbol{\Sigma}_{vv} \right) \tag{B.81}$$

Expression (B.81) shows weakly convergence into a mixed Gaussian limit distribution. In particular, this implies that the limit distribution of $n^{-(1+\gamma_z)/2} \sum_{t=1}^{n} \left(\tilde{\boldsymbol{z}}_{t-1} \otimes \boldsymbol{v}_t \right)$ is Gaussian with mean zero and covariance matrix equal to the probability limit of $n^{-(1+\gamma_z)/2} \sum_{t=1}^{n} \left(\tilde{\boldsymbol{z}}_{t-1} \otimes \boldsymbol{v}_t \right)$, which is equal to $\boldsymbol{V}_{zz} \otimes \boldsymbol{\Sigma}_{vv}$, where $\boldsymbol{V}_{zz}$ is defined in (B.80). Specifically, the above Mixed Gaussianity convergence argument, is a powerful property of the IVX filtration and ensures the robustness of the instrumental variable based procedure for abstract persistence. The dependence of the covariance matrix on the degree of persistence of the IVX instrument, induces exactly the Mixed Gaussianity. Similarly, the limit distribution below follows from Lemma 3.3 of PM.

$$n^{-(1+\gamma_z)/2} \sum_{t=1}^{n} \left(\boldsymbol{x}_{t-1} \otimes \boldsymbol{v}_t \right) \Rightarrow \mathcal{N}\left(0, \boldsymbol{V}_{xx} \otimes \boldsymbol{\Sigma}_{vv} \right), \text{ where } \boldsymbol{V}_{xx} := \int_0^{\infty} e^{r\boldsymbol{C}_p} \boldsymbol{\Omega}_{xx} e^{r\boldsymbol{C}_p} dr$$

□

Appendix C

Supplement to Chapter 4

C.1 Appendix A: Auxiliary Results

We introduce some useful limit results for deriving the asymptotic theory of our framework. These limit results are based on asymptotic theory established by Phillips and Magdalinos (2009) and extensions of some results of Chen et al. (2023). Assume that the $\mathrm{g}-$th equation of the seemingly unrelated system of nonstationary quantile predictive regression models has an autoregressive LUR parametrization for the nonstationary regressors of the particular system equation $\boldsymbol{x}_{\mathrm{g},t} = \boldsymbol{R}_{\mathrm{g}} \boldsymbol{x}_{\mathrm{g},t-1} + \boldsymbol{v}_{\mathrm{g},t}$ where $\mathrm{g} \in \{1, ..., m\}$ (system of m-equations).

Lemma C.1. Let $\boldsymbol{x}_{\mathrm{g},t} = \boldsymbol{R}_{\mathrm{g}} \boldsymbol{x}_{\mathrm{g},t-1} + \boldsymbol{v}_{\mathrm{g},t}$ where $\mathrm{g} \in \{1, ..., m\}$, be the data generating process for the nonstationary regressors for each of the system of m-equations (quantile predictive regressions).

(i). Denote with $\boldsymbol{\mathcal{Q}}_n(\boldsymbol{R}_{\mathrm{g}}) = \left[\mathcal{Q}_{1,n}(\boldsymbol{R}_1), ..., \mathcal{Q}_{m,n}(\boldsymbol{R}_m) \right]'$ where

$$\mathcal{Q}_{\mathrm{g},n}(\boldsymbol{R}_{\mathrm{g}}) := \sum_{t=1}^{n} \frac{1}{\sqrt{\sigma_{\mathrm{g},\mathrm{g}}}} \boldsymbol{R}_{\mathrm{g}}^{-t} \boldsymbol{v}_{\mathrm{g},t}, \quad \boldsymbol{R}_{\mathrm{g}} = \left(\boldsymbol{I} - \frac{\boldsymbol{C}_{\mathrm{g}}}{n} \right), \quad \text{for } \mathrm{g} \in \{1, ..., m\}. \tag{C.1}$$

We have that $\boldsymbol{\mathcal{Q}}_n(\boldsymbol{R}_{\mathrm{g}}) \Rightarrow \boldsymbol{\mathcal{Q}}(\boldsymbol{R}_{\mathrm{g}}) = \left[\mathcal{Q}_1(\boldsymbol{R}_1), ..., \mathcal{Q}_m(\boldsymbol{R}_m) \right]'$, where $\mathcal{Q}_{\mathrm{g}}(\boldsymbol{R}_{\mathrm{g}}) := \sum_{t=1}^{\infty} \frac{1}{\sqrt{\sigma_{\mathrm{g},\mathrm{g}}}} \boldsymbol{R}_{\mathrm{g}}^{-t} \boldsymbol{v}_{\mathrm{g},t}$.

(ii). Denote with $\widetilde{\boldsymbol{\mathcal{Q}}}_n(\boldsymbol{R}_{\mathrm{g}}) = \left[\widetilde{\mathcal{Q}}_{1,n}(\boldsymbol{R}_1), ..., \widetilde{\mathcal{Q}}_{m,n}(\boldsymbol{R}_m) \right]'$ where

$$\widetilde{\mathcal{Q}}_{j,n}(\boldsymbol{R}_{\mathrm{g}}) := \sum_{t=1}^{n} \frac{1}{\sqrt{\sigma_{j,j}}} \boldsymbol{R}_{\mathrm{g}}^{-(n-t)-1} \boldsymbol{v}_{j,t}, \quad \boldsymbol{R}_{\mathrm{g}} = \left(\boldsymbol{I} - \frac{\boldsymbol{C}_{\mathrm{g}}}{n} \right), \quad \text{for } (j, \mathrm{g}) \in \{1, ..., m\}. \tag{C.2}$$

We have that $\widetilde{\boldsymbol{\mathcal{Q}}}_n(\boldsymbol{R}_{\mathrm{g}}) \Rightarrow \widetilde{\boldsymbol{\mathcal{Q}}}(\boldsymbol{R}_{\mathrm{g}}) = \left[\widetilde{\mathcal{Q}}_1(\boldsymbol{R}_1), ..., \widetilde{\mathcal{Q}}_m(\boldsymbol{R}_m) \right]'$, where $\widetilde{\mathcal{Q}}_j(\boldsymbol{R}_{\mathrm{g}}) := \sum_{t=1}^{\infty} \frac{1}{\sqrt{\sigma_{\mathrm{g},\mathrm{g}}}} \boldsymbol{R}_{\mathrm{g}}^{-(n-t)-1} \boldsymbol{v}_{j,t}$. Notice that the subscript j of $\widetilde{\mathcal{Q}}_j(\cdot)$ corresponds to the error sequence $\boldsymbol{v}_{j,t}$.

(iii). We assume that $\boldsymbol{\mathcal{Q}}(\boldsymbol{R}_{\mathrm{g}})$ and $\widetilde{\boldsymbol{\mathcal{Q}}}(\boldsymbol{R}_{\mathrm{g}})$ are asymptotically independent.

Lemma C.2. Consider the martingale array given by

$$\boldsymbol{M}_n(s) := \left[\sum_{t=1}^{n} \boldsymbol{x}_{1,t-1}\boldsymbol{v}_{j,t}, ..., \sum_{t=1}^{n} \boldsymbol{x}_{\mathsf{g},t-1}\boldsymbol{v}_{j,t} \right] \tag{C.3}$$

for a fixed $j \in \{1, ..., m\}$ and some $\mathsf{g} \in \{1, ..., m\}$. Then, the following joint converges implies

$$\begin{bmatrix} \boldsymbol{M}_n(s) \\ \dfrac{1}{\sqrt{n}} \displaystyle\sum_{t=1}^{\lfloor ns \rfloor} \boldsymbol{u}_t \end{bmatrix} \Rightarrow \begin{bmatrix} \boldsymbol{M}(s) \\ \boldsymbol{B}(s) \end{bmatrix}, \quad \text{on the Skorokhod space } \mathbb{D}_{\mathbb{R}^{j+m}}[0,1],$$

where $\boldsymbol{M}$ and $\boldsymbol{B}$ are independent Brownian motions with well-defined covariance matrices.

C.2 Asymptotic Theory

Notation For two sequences of real numbers (a_n) and (b_n), we say that $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$. Furthermore, notice that our setting operates under the assumption of a fixed quantile level $\tau \in (0,1)$. Further regularity conditions we assume that are satisfied which include: *(i)* the spectral density matrix of the vector of equilibrium errors is bounded away from zero, *(ii)* the long-run covariance matrix exists, and *(iii)* the fourth-order cumulants are absolutely summable.

In terms of the asymptotic limit scheme we do not consider the sequential $(m,n) \to \infty$ case unless we otherwise specify so. In particular, the main limit results correspond to the case where m remains fixed but $n \to \infty$. This allows to establish the consistency of the pooled IVX estimator irrespective of whether m is large or not. Moreover, to obtain the asymptotic distribution of the common long-run coefficients, we will need to impose an additional condition such that m is a monotonic function of n, say $m(n)$, such that $m \to \infty$ only as $n \to \infty$. In that case, the correct rate of convergence for $\boldsymbol{\vartheta}$ towards its true value is given by $\sqrt{mn}$, in the case of stationary regressors, denoted with $I(0)$. All long-run coefficients are the same across groups, although we can allow only a subset of the long-run coefficients to be the same while others to differ.

Lemma C.3. *(Consistency of variance-covariance matrix)* The following result holds:

$$\left| \frac{1}{Nn} \sum_{i=1}^{N} \sum_{t=1}^{n} \boldsymbol{X}^{\star}_{(i)t-1} \boldsymbol{X}^{\star\top}_{(i)t-1} - \mathbb{E}\left[\boldsymbol{X}^{\star}_{(i)t-1} \boldsymbol{X}^{\star\top}_{(i)t-1} \right] \right| = o_p(1). \tag{C.4}$$

Lemma C.4. *(Consistency of variance-covariance matrix)* The following result holds:

$$\left| \frac{1}{mn} \sum_{i=1}^{m} \sum_{t=1}^{n} \boldsymbol{X}^{\star}_{(i)t-1} \boldsymbol{X}^{\star\prime}_{(i)t-1} - \mathbb{E}\left[\boldsymbol{X}^{\star}_{(i)t-1} \boldsymbol{X}^{\star\prime}_{(i)t-1} \right] \right| = o_p(1). \tag{C.5}$$

In addition, let $f_{it}(\cdot|x)$ denote the conditional density of u_{it} given $X_{it-1} = x$ and let $\xi_{i,n}$ denote the smallest eigenvalue of $n^{-1} \sum_{t=1}^{n} \mathbb{E}\big[f_{it}(\cdot|\boldsymbol{X}_{it-1}) \widetilde{\boldsymbol{X}}_{it-1} \widetilde{\boldsymbol{X}}'_{it-1} \big]$.

Lemma C.5. *(SUR Graph-based Bahadur Representation)*

The following representation holds uniformly over the compact set $\mathcal{B} \subset (0,1)$ such that $\tau \in \mathcal{B}$

$$\left(\hat{\boldsymbol{\beta}}^{\star}_{SUR-IVX}(\tau) - \boldsymbol{\beta}^{\star}(\tau)\right) = \hat{\boldsymbol{D}}_1(\tau) \left\{ \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n} \sum_{t=1}^{n} \psi_\tau \left(\boldsymbol{\mathcal{E}}_{(i)t}(\tau) \right) \boldsymbol{X}^{\star}_{(i)t-1} \right\} + \mathcal{R}_{mn}(\tau), \tag{C.6}$$

where $\psi_\tau(\mathsf{u}) \mapsto \rho_\tau^{-1}(\mathsf{u}) := (\tau - \mathbf{1}\{\mathsf{u} \leq 0\})$ and $\mathcal{R}_{mn}(\tau)$ is the remainder term. Then,

$$\hat{\boldsymbol{\beta}}^{\star}_{SUR-IVX}(\tau) \xrightarrow{p} \boldsymbol{\beta}^{\star}(\tau) \quad \text{uniformly for } \tau \in (0,1) \text{ as } \min\{m,n\} \to \infty. \tag{C.7}$$

Proof. Notice that the remainder term of the expression which satisfies $\sup_{\tau \in (0,1)} |\mathcal{R}_{mn}(\tau)| = \mathcal{O}_p\left(\frac{1}{\sqrt{mn}}\right)$.

Denote with

$$\boldsymbol{\mathcal{E}}_{(i)t}(\tau) := \left(y_{(i)t} - \boldsymbol{X}^{\star\prime}_{(i)t-1} \boldsymbol{\beta}^{\star}(\tau) \right) \quad \text{and} \quad \mathbb{S}_{in}(\boldsymbol{\beta}^{\star}(\tau)) := \frac{1}{n} \sum_{t=1}^{n} \psi_\tau \left(\boldsymbol{\mathcal{E}}_{(i)t}(\tau) \right) \boldsymbol{X}^{\star}_{(i)t-1}.$$

Moreover, let $\mathbb{S}_i(\boldsymbol{\beta}^{\star}(\tau)) \equiv \mathbb{E}_{\mathcal{F}_{t-1}}\left[\mathbb{S}_{in}(\boldsymbol{\beta}^{\star})\right] = \mathbb{E}\left[\tau - F_i(\boldsymbol{\beta}^{\star}(\tau) - \boldsymbol{\beta}^{\star}_{i0}(\tau) | \boldsymbol{X}^{\star}_{i,t-1} = \boldsymbol{x}^{\star}_{i,t-1})\right]$. Thus, we can establish the following probability bound

$$\sqrt{n}\left(\left[\mathbb{S}_{in}(\hat{\boldsymbol{\beta}}^{\star}_i(\tau)) - \mathbb{S}_{in}(\boldsymbol{\beta}^{\star}_{i0}(\tau)) \right] - \left[\mathbb{S}_i(\hat{\boldsymbol{\beta}}^{\star}_i(\tau)) - \mathbb{S}_i(\boldsymbol{\beta}^{\star}_{i0}(\tau)) \right] \right) = o_p(1).$$

for each cross-sectional unit due to the stochastic equicontinuity of the process and $\hat{\boldsymbol{\beta}}^{\star}_i(\tau) \xrightarrow{p} \boldsymbol{\beta}^{\star}_i(\tau)$.

□

Theorem C.1. Under the conditions of Assumption 3.2, we have that

$$\sqrt{Nn}\, \boldsymbol{\Sigma}_{\beta}^{-1/2}(\tau) \left(\hat{\boldsymbol{\beta}}^{\star}_{SUR-IVX}(\tau) - \boldsymbol{\beta}^{\star}(\tau) \right) \xrightarrow{\mathcal{D}} \boldsymbol{\mathcal{BB}}_{q+1}(\tau), \tag{C.8}$$

as $\min\{N,n\} \to \infty$, where $\boldsymbol{\Sigma}_{\beta}(\tau) = \boldsymbol{D}_1^{-1}(\tau) \boldsymbol{D}_0 \boldsymbol{D}_1^{-1}(\tau)$ with the matrix $\boldsymbol{D}_0$ defined as:

$$\boldsymbol{D}_0 \tag{C.9}$$

such that $\boldsymbol{\mathcal{BB}}_{q+1}(\tau)$ is a $(q+1)-$dimensional Brownian Bridge.

Corollary C.1. Under Assumption 3.2, for any fixed value of τ in the compact set $\mathcal{B}$, say $\tau_0 \in (0,1)$, the following limit result holds:

$$\sqrt{Nn}\left(\hat{\boldsymbol{\beta}}^{\star}_{SUR-IVX}(\tau_0) - \boldsymbol{\beta}^{\star}(\tau_0) \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \tau_0(1-\tau_0) \boldsymbol{\Sigma}_{\beta}(\tau_0)\right) \text{ as } \min\{N,n\} \to \infty. \tag{C.10}$$

Notice that Theorem C.1 provides the limiting distribution for the normalized estimators distance term. The particular normalization constant allows to construct a self-normalized Wald type statistic which accounts for the different estimation effects in the SUR system.

Remark C.1. A discussion of the pivotal property of self-normalized test statistics is presented by Shao (2010). More precisely, the pivotal property ensures the applicability of conventional inference methods without the need to use bootstrap based inference methods to obtain critical values from the bootstrapped distribution of the test statistic. Notice that Corollary C.1 is a direct implication of Theorem C.1. In particular, Corollary C.1 permits to conduct pointwise inference, for any fixed value τ_0, in the estimated SUR system parameters since we obtain weakly convergence to a mixed Gaussian random variant. Moreover, notice that the assumption of a compact set $\mathcal{B} \subset (0,1)$, which is an open subset of the boundaries of the parameter space for the quantile level $\tau \in (0,1)$ and the corresponding quantile function, ensures the uniform convergence of the model estimators.

Proof of Lemma:

Proof. We need to show that the quantities $\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n} \boldsymbol{X}^{\star}_{(i)t-1}\boldsymbol{X}^{\star\prime}_{(i)t-1}$ and $\mathbb{E}\left[\boldsymbol{X}^{\star}_{(i)t-1}\boldsymbol{X}^{\star\prime}_{(i)t-1}\right]$ are close to each other element by element, such that the following probability limit holds:

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n} \boldsymbol{X}^{\star}_{(i)t-1}\boldsymbol{X}^{\star\prime}_{(i)t-1} \overset{p}{\to} \begin{bmatrix} \mathbb{E}\left[\boldsymbol{x}_{(i)t-1}\boldsymbol{x}'_{(i)t-1}\right] & \mathbb{E}\left[\boldsymbol{x}_{(i)t-1}y'_{(j)t}\right] \\ \mathbb{E}\left[y_{(j)t}\boldsymbol{x}'_{(i)t-1}\right] & \mathbb{E}\left[y_{(j)t}y'_{(j)t}\right] \end{bmatrix}, \tag{C.11}$$

where $(i,j) \in \{1,...,m\}$ with $i \neq j$.

Moreover, since we assume that the set of regressors of the SUR system is the same, then it remains to obtain the asymptotic behaviour of the elements of the matrix above for the two types of persistence we consider, that is, $\gamma_x = 1$ for LUR regressors and $\gamma_x \in (0,1)$ for mildly integrated (MI) regressors. For LUR regressors, that is, $\gamma_x = 1$, Kostakis et al. (2015) show that as $n \to \infty$

$$\frac{1}{n^{1+\gamma_x}}\boldsymbol{x}'_{(i)t-1}\boldsymbol{x}_{(i)t-1} \overset{p}{\to} \boldsymbol{V}_c := \int_0^{\infty} e^{r\boldsymbol{C}}\boldsymbol{\Omega}_{vv}e^{r\boldsymbol{C}}dr. \tag{C.12}$$

Consider the asymptotic behaviour for the term $\frac{1}{mn}\sum_{i=1}^{N}\sum_{t=1}^{n} y_{(j)t}y'_{(j)t}$ where

$$y_{(j)t} = \boldsymbol{\beta}'_{(j)}(\tau)\boldsymbol{x}_{t-1} + \delta_{(j)}(\tau)y_{(i)t} + u_{(j)t} \tag{C.13}$$

$$\boldsymbol{x}_t = \left(\boldsymbol{I}_k - \frac{\boldsymbol{C}}{n^{\gamma_x}}\right)\boldsymbol{x}_{t-1} + \boldsymbol{v}_t \tag{C.14}$$

such that it expands to

$$\begin{aligned}
\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n} y'_{(j)t}y_{(j)t} &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\left[\boldsymbol{\beta}'_{(j)}(\tau)\boldsymbol{x}_{t-1} + \delta_{(j)}(\tau)y_{(i)t} + u_{(j)t}\right]\left[\boldsymbol{\beta}'_{(j)}(\tau)\boldsymbol{x}_{t-1} + \delta_{(j)}(\tau)y_{(i)t} + u_{(j)t}\right]' \\
&= \frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{\beta}'_{(j)}(\tau)\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\boldsymbol{\beta}_{(j)}(\tau) + \frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{\beta}'_{(j)}(\tau)\boldsymbol{x}_{t-1}y'_{(i)t}\delta_{(j)}(\tau) \\
&+ \frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{\beta}'_{(j)}(\tau)\boldsymbol{x}_{t-1}u'_{(j)t}
\end{aligned}$$

$$+\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\delta_{(j)}(\tau)y_{(i)t}\boldsymbol{x}'_{t-1}\boldsymbol{\beta}_{(j)}(\tau)+\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\delta_{(j)}(\tau)y_{(i)t}y'_{(i)t}\delta_{(j)}(\tau)+\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\delta_{(j)}(\tau)y_{(i)t}u'_{(j)t}$$

$$+\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}u_{(j)t}\boldsymbol{x}'_{t-1}\boldsymbol{\beta}_{(j)}(\tau)+\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}u_{(j)t}y'_{(i)t}\delta_{(j)}(\tau)+\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}u_{(j)t}u'_{(j)t}$$

Next, we consider the asymptotic behaviour of each of those terms separately such as

$$\boldsymbol{\beta}'_{(j)}(\tau)\mathbb{E}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right]\boldsymbol{\beta}_{(j)}(\tau) \tag{C.15}$$

$$\boldsymbol{\beta}'_{(j)}(\tau)\mathbb{E}\left[\boldsymbol{x}_{t-1}y'_{(i)t}\right]\delta_{(j)}(\tau) \tag{C.16}$$

Next, consider the asymptotic behaviour of the upper right term of the matrix given by expression (C.11), that is, $\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{x}_{(i)t-1}y'_{(j)t}$, which can be expanded further as

$$\begin{aligned}\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{x}_{(i)t-1}y'_{(j)t} &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{x}_{(i)t-1}\left[\boldsymbol{\beta}'_{(j)}(\tau)\boldsymbol{x}_{t-1}+\delta_{(j)}(\tau)y_{(i)t}+u_{(j)t}\right]'\\ &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{x}_{(i)t-1}\boldsymbol{x}'_{(i)t-1}\boldsymbol{\beta}_{(j)}(\tau)+\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{x}_{(i)t-1}y'_{(i)t}\delta_{(j)}(\tau)\\ &+\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{x}_{(i)t-1}u'_{(j)t}\end{aligned}$$

For example for the last term, we have that

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}\boldsymbol{x}_{(i)t-1}u'_{(j)t}=\frac{1}{m}\sum_{i=1}^{m}\left\{\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_{(i)t-1}u'_{(j)t}\right\} \tag{C.17}$$

Assume that $\forall\ i\in\{1,...,m\}$, we have a common set of regressors, that is, $\boldsymbol{x}_{(i)t-1}\equiv\boldsymbol{x}_{t-1}$ and that $u_{(j)t}$ represents the error term of the $j-$th specification such that $j\neq i$.

□

Proof of Theorem 4.2:

Consider the formulation for the SUR-IVX Wald statistic with respect to the proposed SUR-IVX estimator, $\hat{\boldsymbol{\beta}}^{\star}_{SUR-IVX}(\tau)$ for any quantile level $\tau\in(0,1)$ given by

$$\mathcal{W}_{SUR-IVX}(\tau)=\left(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{\star}_{SUR-IVX}(\tau)-\boldsymbol{q}\right)^{\top}\boldsymbol{\Omega}_{\boldsymbol{R}}^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{\star}_{SUR-IVX}(\tau)-\boldsymbol{q}\right) \tag{C.18}$$

where $\boldsymbol{\Omega}_{\boldsymbol{R}}$ the covariance matrix of the Wald test based on the SUR-IVX estimator. More precisely, the covariance matrix is defined as below

$$\boldsymbol{\Omega}_{\boldsymbol{R}}=\left\{\boldsymbol{R}\left[\left(\boldsymbol{X}^{\star\prime}\boldsymbol{P}_{\tilde{\boldsymbol{Z}}}\boldsymbol{X}^{\star}\right)^{-1}\otimes\hat{\boldsymbol{\Sigma}}_{vv}\right]\boldsymbol{R}^{\top}\right\} \tag{C.19}$$

where $\boldsymbol{P}_{\tilde{\boldsymbol{Z}}}$ represents the projection matrix with respect to the IVX instrumental variable $\tilde{\boldsymbol{Z}}$ given by expression (4.37) and $\boldsymbol{R}$ is the linear restrictions matrix.

Proof. We need to show that $\mathcal{W}_{SUR-IVX}(\tau) \Rightarrow \chi^2_q \quad \text{for some } \tau \in (0,1) \text{ as } T \to \infty.$ □

Proof of Theorem C.1:

Recall that a Brownian-Bridge process is defined as below:

$$\boldsymbol{\mathcal{BB}}_{q+1}(\tau) := \left[\boldsymbol{W}_{q+1}(\tau) - \tau \boldsymbol{W}_{q+1}(1)\right]^{\top} \left[\boldsymbol{W}_{q+1}(\tau) - \tau \boldsymbol{W}_{q+1}(1)\right] \tag{C.20}$$

where $\boldsymbol{W}_{q+1}(.)$ is a $(q+1)-$dimensional standard Wiener process on the same or equivalent probability space. Therefore, we need to show that the normalized distance measure for the SUR-IVX estimator weakly converges to the Brownian Bridge defined by (C.20).

Proof. We denote the following moment matrices as below

$$\widehat{\boldsymbol{D}}_0 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{X}^{\star}_{(i)t-1} \boldsymbol{X}^{\star\top}_{(i)t-1}, \tag{C.21}$$

and

$$\widehat{\boldsymbol{D}}_1(\tau) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f_{it} \left(\boldsymbol{X}^{\star\top}_{(i)t-1} \boldsymbol{\beta}^{\star}(\tau) \right) \boldsymbol{X}^{\star}_{(i)t-1} \boldsymbol{X}^{\star\top}_{(i)t-1}. \tag{C.22}$$

where $f_{it}(.)$ is a continuous function. Moreover, we have that $\boldsymbol{\Sigma}_{\beta}(\tau) = \boldsymbol{D}_1^{-1}(\tau) \boldsymbol{D}_0 \boldsymbol{D}_1^{-1}(\tau)$ with the matrix $\boldsymbol{D}_0$ defined as:

Notice that the convergence result requires that both N and T are required to diverge to infinity to obtain a $\sqrt{NT}-$consistent result.

□

Proof of Corollary 3.1:

Proof. For any fixed $\tau_0 \in (0,1)$ we need to show that

$$\sqrt{NT} \left(\hat{\boldsymbol{\beta}}^{\star}_{SUR-IVX}(\tau_0) - \boldsymbol{\beta}^{\star}(\tau_0) \right) \overset{\mathcal{D}}{\to} \mathcal{N}\big(\boldsymbol{0}, \tau_0(1-\tau_0) \boldsymbol{\Sigma}_{\beta}(\tau_0)\big) \text{ as } \mathsf{min}\{N,T\} \to \infty.$$

□

Proof of Theorem 3 ($\mathcal{J}$ test for subset predictability)

Denote with $\boldsymbol{1} = (1, ..., 1)'$ be a $K \times 1$. Furthermore, demeaning the original econometric specification we obtain the following expression

$$\boldsymbol{Y}_t = \widetilde{\boldsymbol{X}}_{t-1} \boldsymbol{\delta}_f + \boldsymbol{x}'_{t-1} \delta_x + u_t \boldsymbol{1} \tag{C.23}$$

Therefore, using the particular representation, we can express the IVX estimator as below

$$\widehat{\boldsymbol{\theta}} = \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \widetilde{\boldsymbol{X}}_{t-1} \right)^{-1} \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \widetilde{\boldsymbol{Y}}_{t} \right) = \bar{\boldsymbol{\delta}}_f + \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \widetilde{\boldsymbol{X}}_{t-1} \right)^{-1} \sum_{t=1}^{n} \left(\boldsymbol{Z}'_{t-1} \mathbf{1} \right) \left(\tilde{\boldsymbol{x}}_{t-1} \boldsymbol{\delta}_x + u_t \right). \tag{C.24}$$

Furthermore, we employ the restriction matrix such that $\boldsymbol{R}\bar{\boldsymbol{\delta}}_f = \mathbf{0}_{(N-1)\times 1}$, therefore by implementing these linear restrictions we obtain the following expression

$$\widehat{\beta} := \boldsymbol{R}\widehat{\boldsymbol{\theta}} = \boldsymbol{R} \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \widetilde{\boldsymbol{X}}_{t-1} \right)^{-1} \left(\boldsymbol{Z}'_{t-1} \mathbf{1} \right) \left(\tilde{\boldsymbol{x}}_{t-1} \boldsymbol{\delta}_x + u_t \right) \overset{\mathcal{H}_0}{=} \boldsymbol{R} \left(\sum_{t=1}^{n} \boldsymbol{Z}'_{t-1} \widetilde{\boldsymbol{X}}_{t-1} \right)^{-1} \left(\boldsymbol{Z}'_{t-1} \mathbf{1} \right) u_t \tag{C.25}$$

We also define the following matrices we derive the asymptotic behaviour of the above quantity

$$\boldsymbol{P} = \begin{bmatrix} P_1 \\ \vdots \\ P_{N-1} \end{bmatrix} \quad \text{and} \quad \boldsymbol{P} = \begin{bmatrix} P_1 & & \\ & \ddots & \\ & & P_{N-1} \end{bmatrix} \tag{C.26}$$

$$P_g = \begin{bmatrix} \boldsymbol{I}_{m_f} \\ \Big(\underbrace{\mathbf{0}}_{k_g \times k_1}, \dots, \underbrace{\mathbf{0}}_{k_g \times k_{g-1}}, \underbrace{\boldsymbol{I}}_{k_g \times k_g}, \underbrace{\mathbf{0}}_{k_g \times k_{g+1}}, \dots, \underbrace{\mathbf{0}}_{k_g \times k_{N-1}} \Big) \end{bmatrix}. \tag{C.27}$$

Then, we have that $\boldsymbol{Z}'_{t-1}\mathbf{1} = \boldsymbol{P}\boldsymbol{z}_{t-1}$. Furthermore, it holds that

$$\boldsymbol{Z}'_{t-1} = \boldsymbol{S}(\boldsymbol{I}_{N-1} \otimes \boldsymbol{G}_{t-1}) \quad \text{and} \quad \boldsymbol{Z}'_{t-1} = \boldsymbol{S}(\boldsymbol{I}_{N-1} \otimes \boldsymbol{G}_{t-1}) \tag{C.28}$$

which implies that

$$\boldsymbol{Z}'_{t-1} \widetilde{\boldsymbol{X}}_{t-1} = \boldsymbol{S}(\boldsymbol{I}_{N-1} \otimes \boldsymbol{z}_{t-1} \boldsymbol{g}'_{t-1}) \boldsymbol{S}' \tag{C.29}$$

Then, under the null hypothesis it holds that

$$n^{\frac{1+\gamma_z}{2}} \widehat{\boldsymbol{\beta}} = \boldsymbol{R} \left\{ \sum_{t=1}^{n} \boldsymbol{S} \Big(\boldsymbol{I}_{N-1} \otimes \boldsymbol{z}_{t-1} \tilde{\boldsymbol{g}}'_{t-1} \Big) \boldsymbol{S}' \right\}^{-1}. \tag{C.30}$$

which implies that

$$n^{-\left(\frac{1+\gamma_z}{2}\right)} \sum_{t=1}^{n} \boldsymbol{P} \boldsymbol{z}_{t-1} u_t \overset{d}{\to} \boldsymbol{R} \Big[\boldsymbol{S} \Big(\boldsymbol{I}_{N-1} \otimes \boldsymbol{Q}_{zg} \Big) \boldsymbol{S}' \Big]^{-1} \boldsymbol{P} \sigma_0 \boldsymbol{Q}_{zz}^{1/2} \mathcal{Z} \equiv \boldsymbol{R} \tag{C.31}$$

In particular, consider that $\boldsymbol{\beta}_i$ is the limit of the OLS estimator, then we have that

$$\boldsymbol{\beta}_i = \operatorname*{plim}_{n \to \infty} \left(\frac{1}{n^2} \sum_{t=1}^{n} \boldsymbol{x}_{g,t-1} \boldsymbol{x}_{g,t-1}^{\top} \right)^{-1} \left(\frac{1}{n^2} \sum_{t=1}^{n} \boldsymbol{x}_{g,t-1} \boldsymbol{X}_{t-1}^{\top} \right) \widetilde{\boldsymbol{\beta}}, \quad \text{for} \quad g \in \{1, \dots, K\} \tag{C.32}$$

where $\widetilde{\boldsymbol{\beta}}$ is the value of the parameter vector under the alternative hypothesis.

C.3 Doubly Corrected Estimation: Conditional Mean Case

C.3.1 Residual Augmented IVX Estimation Example

We begin our asymptotic theory analysis by considering the framework for the residual augmented IVX estimation methodology previously examined in the literature (see, Demetrescu and Rodrigues (2020)). The IVX instrumentation provides a methodology such that the asymptotic distribution of the IVX-Wald based test statistic converges to a nuisance-parameter free distribution. Consider the feasible predictive regression defined with the following form

$$y_t^{\mu} = \beta x_{t-1}^{\mu} + \delta \hat{v}_t + \varepsilon_t, \quad \varepsilon_t \sim_{i.i.d} \text{ and } \; t = 1, ..., n. \tag{C.33}$$

where y_t^{μ} and x_{t-1}^{μ} denote the demeaned variates, following the notation employed in several studies in the literature. Moreover, $\hat{v}_t$ represents a generated covariate which in our study corresponds to the estimated VaR under nonstationarity.

The parameter of interest for inference purposes is the slope coefficient of the quantile predictive regression of the second-stage procedure. Based on the specification above, the IVX estimator of β, has a slower convergence rate than the conventional OLS estimator under near integration. In addition, the IVX estimator is mixed Gaussian in the limit regardless of the degree of endogeneity which implies standard inference based on the Wald test. Under low persistence it is asymptotically equivalent to the OLS procedure which implies that the IVX estimator is easy to implement. The null hypothesis of no predictability implies the presence of no serial correlation. Furthermore, we conjecture that the generated regressor since it is stationary it will not affect the convergence of the IVX estimator to its asymptotic limit, given the lower convergence rate of IVX compared to the OLS estimator. Although our procedure has some similarities with residual-based augmentation procedures proposed in the literature, our objective is to demonstrate that in finite-samples and asymptotically has similar desirable properties. All directly observable variables and demeaned. Let $\bar{y}_t = y_t - \bar{y}_n = \left(y_t - \frac{1}{n} \sum_{t=1}^{n} y_t \right)$, $\bar{x}_t = x_t - \bar{x}_n = \left(x_t - \frac{1}{n} \sum_{t=1}^{n} x_t \right)$ and $\bar{x}_{t-1} = x_{t-1} - \bar{x}_{n-1} = \left(x_{t-1} - \frac{1}{n} \sum_{t=1}^{n} x_{t-1} \right)$. Then the IVX estimator is defined as below

$$\widehat{\beta}^{ivx} := \frac{\displaystyle\sum_{t=1}^{n} \bar{y}_t z_{t-1}}{\displaystyle\sum_{t=1}^{n} \bar{x}_{t-1} z_{t-1}}, \quad \text{with} \quad se\left(\widehat{\beta}^{ivx} \right) := \left(\sum_{t=1}^{n} \bar{x}_{t-1} z_{t-1} \right)^{-1} \widehat{\sigma}_u \left(\sum_{t=1}^{n} z_{t-1}^2 \right)^{1/2}.$$

where $se\left(\widehat{\beta}^{ivx} \right)$ is the standard error of the IVX estimator. To improve the finite-sample behaviour Kostakis et al. (2015) suggest the use of OLS residuals $\widehat{u}_t$ for the computation of $\widehat{\sigma}_u$, do not demean the instrument z_t and correct the standard errors by subtracting from $\widehat{\sigma}_u \left(\sum_{t=1}^{n} z_{t-1}^2 \right)^{1/2}$. In addition, we define with $\tilde{\varepsilon}_t := \bar{y}_t - \left(\widehat{\beta}^{ols} \bar{x}_{t-1} + \widehat{\delta} \widehat{v}_t \right)$, and use those instead of the IVX residuals due to the superconsistency properties of the OLS-based residuals in the near-integrated context. Recall that the above expressions correspond to the model estimates from a linear predictive regression model with conditional mean functional form. Based on the above illustrative example, we replace the regressor $\hat{v}_t$ with the generated dependent variable, $\widehat{\widehat{y}}_{1,t}$, which is used as an additional covariate. We do this first in the case of a linear predictive regression measure, so the estimation procedures does not correspond to the risk measure pair yet.

C.3.2 Generated Regressor in Nonstationary Linear Predictive Regression

In compact form we have that

$$\bar{y}_{2,t} = \boldsymbol{X}_{t-1}\boldsymbol{\Gamma}' + \varepsilon_t, \quad \varepsilon_t \sim_{i.i.d} \text{ and } \ t = 1, ..., n. \tag{C.34}$$

where $\boldsymbol{X}_{t-1} = (\bar{\boldsymbol{x}}_{t-1} \ \widehat{\widehat{y}}_{1,t})$ and $\boldsymbol{\Gamma} = (\boldsymbol{\beta} \ \delta)$. Moreover, define with $\boldsymbol{Z}_{t-1} = (\boldsymbol{z}_{t-1} \ \widehat{\widehat{y}}_{1,t})$ where $\boldsymbol{z}_{t-1}$ is the IVX instrument for $\boldsymbol{x}_{t-1}$ and the generated regressor is self-instrumented.

Notice that the generated covariate $\widehat{\widehat{y}}_{1,t}$ aims to resemble the estimated VaR under nonstationarity. In other words, the generated regressor, denoted with $\widehat{\widehat{y}}_{1,t}$ is obtained from a predictive regression model which includes only the nonstationary predictors. In practise, these nonstationary predictors can be different than the nonstationary predictors corresponding to the regressand $\widehat{y}_{2,t}$ due to the proposed dependence structure. However, for simplicity we can also consider the case in which we use the same nonstationary regressors for both predictive regression models. In this section, we obtain derivations for the linear predictive regression model. However, when estimating the risk measures of VaR and CoVaR, the nonstationary quantile predictive regression models are used which implies that all parameters are quantile-dependent for a fixed quantile $\tau \in (0, 1)$. Therefore, we are interested for the consistent estimation of the β_2 estimator (e.g., via IVX) as well as constructing hypothesis testing for: (i) only β parameter (e.g., no predictability), (ii) only δ parameter (e.g., no presence of generated regressor) and (iii) linear restrictions on all the parameters of the predictive regression model that includes both nonstationary regressors and the generated regressor (e.g., univariate generated regressor). The covariance estimator for $\tilde{\boldsymbol{\beta}}_2^{ivx}$ expressed as

$$\widehat{\mathsf{Cov}\left(\tilde{\boldsymbol{\beta}}_2^{ivx}\right)} := \left(\boldsymbol{B}_n^{-1}\right) \mathbb{V}_n \left(\boldsymbol{B}_n^{-1}\right)', \quad \boldsymbol{B}_n := \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{t-1}\boldsymbol{z}_{t-1}'\right)$$

$$\mathbb{M}_n := \left(\sum_{t=1}^{n} \boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\tilde{\varepsilon}_t^2\right) + \left[\boldsymbol{\gamma}^\top \otimes \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{z}_{t-1}\bar{\boldsymbol{x}}_{t-1}^\top\right)\left(\frac{1}{n}\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{t-1}\bar{\boldsymbol{x}}_{t-1}^\top\right)^{-1}\right]\left(\sum_{t=1}^{n} \boldsymbol{\nu}_t\boldsymbol{\nu}_t^\top \otimes \bar{\boldsymbol{x}}_{t-1}\bar{\boldsymbol{x}}_{t-1}^\top\right)$$

$$\times \left[\boldsymbol{\gamma} \otimes \left(\frac{1}{n}\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{t-1}\bar{\boldsymbol{x}}_{t-1}^\top\right)^{-1}\left(\frac{1}{n}\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{t-1}\boldsymbol{z}_{t-1}^\top\right)\right].$$

Therefore, the just-identified model for the IV estimator of $\boldsymbol{\Gamma}$ is given by

$$\hat{\boldsymbol{\Gamma}} = \left(\sum_{t=1}^{n} \boldsymbol{Z}_{t-1}'\boldsymbol{X}_{t-1}\right)^{-1}\left(\sum_{t=1}^{n} \boldsymbol{Z}_{t-1}'\bar{y}_{2,t}\right). \tag{C.35}$$

Furthermore, notice that the nonstationary regressors x_{t-1} and the generated regressor are orthogonal since the generated regressor can be considered as exogenous variable to the particular model, then we can compute the generated regressor augmented estimator and test statistic in two steps. Notice the constructed test statistic refers to the IVX estimator. Therefore, we focus on deriving the asymptotic distribution for the IVX estimator under the presence of the generated regressor in the quantile predictive regression model. The estimation procedure is implemented into two-stages, and our asymptotic theory analysis demonstrates that although the limit converges into a nonstandard distribution the particular limit is nuisance-parameter free.

Remark C.2. The key point in this estimation step, is that the variance-covariance needs to be adjusted to account for the first-step estimation error. In other words, when estimating the VaR using the quantile predictive regression of the first stage, the forecast at the current time period t, is estimated with some additional source of error. For this reason, an adjustement is necessary. Consider the estimated coefficient of the quantile predictive regression model of the first stage as

$$Q_{\tau}(y_{2t}|x_{t-1}, y_{1t}) = \theta X_{t-1}, \quad \tau \in (0,1). \tag{C.36}$$

In other words, the estimation step needs to account for two sources of errors, that is, the usaul estimation error in obtaining a consistent estimator that corresponds to the IVX instrumentation of the nearly integrated regressors, and the second source of error is the sampling error in generating the forecast for the VaR from the first stage of the process. More specifically, this implies that using a Bahadur representation of the QR-IVX estimator we need to determine the precise stochastic order of the remainder term when the generated regressor is included in the conditional quantile specification of the model. Thus the main focus of the estimation step should be how to correct for the effect to the overall variance due to the presence of this generated regressor. In particular, the parameter vector from the first stage has the usual convergence rate that the IVX estimator has (regardless of the degree of persistence). However, we need to check whether this convergence rate remains the same for the parameter vector of the second stage estimation under the presence of the generated regressor. Furthermore, the property of stochastic equicontinuity still holds in this case, since we have a plug-in estimator, regardless if the underline stochastic processes correspond to nonstationary time series data.

Estimation Procedure in Nonstationary Linear Predictive Regression Case

Step 1. OLS-regress $\bar{y}_{2,t}$ on the generated regressor $\widehat{\bar{y}}_{1,t}$ and obtain the OLS estimator below

$$\widehat{\delta}^{ols} = \left(\sum_{t=1}^{n} \widehat{\bar{y}}^{2}_{1,t} \right)^{-1} \left(\sum_{t=1}^{n} \widehat{\bar{y}}_{1,t} \bar{y}_{2,t} \right). \tag{C.37}$$

and compute the residual such that $\widehat{\eta}_t = \bar{y}_{2,t} - \widehat{\delta}^{ols} \widehat{\bar{y}}_{1,t}$. Here we assume that the elements of the dependent vector are not serially correlated, which implies that y_{1t} and y_{2t} are uncorrelated. This simplifies the development of the asymptotic theory, despite the presence of endogeneity and highly persistent regressors.

Step 2. IVX-regress $\bar{y}_{2,t}$ on $\bar{\boldsymbol{x}}_{t-1}$, leading to

$$\widehat{\boldsymbol{\beta}}^{ivx}_{2} = \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{t-1} \boldsymbol{z}'_{t-1} \right)^{-1} \left(\sum_{t=1}^{n} \tilde{y}_{t} \boldsymbol{z}'_{t-1} \right). \tag{C.38}$$

Recall the linear predictive regression model (for node 1) is given by

$$y_{1,t} = \boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} + u_{1,t} \tag{C.39}$$

$$\boldsymbol{x}_t = \boldsymbol{R}\boldsymbol{x}_{t-1} + \boldsymbol{v}_{1,t} \tag{C.40}$$

Then we estimate the following model

$$\bar{y}_{2,t} = \boldsymbol{\beta}_2' \bar{\boldsymbol{x}}_{t-1} + \delta \widehat{\widehat{y}}_{1,t} + u_{2,t} \tag{C.41}$$

Moreover, by OLS-regressing $\bar{y}_{2,t}$ on $\widehat{\widehat{y}}_{1,t}$ such that $\bar{y}_{2,t} = \theta \widehat{\widehat{y}}_{1,t} + \eta_t$, where η_t is a disturbance term, we construct the transformed dependent variable $\tilde{y}_{2,t}$ given by the following expression

$$\tilde{y}_{2,t} := \left(\bar{y}_{2,t} - \widehat{\theta}^{ols} \widehat{\widehat{y}}_{1,t} \right) \tag{C.42}$$

- In the second stage of the estimation procedure, we need to re-estimate the quantile kernel function when the generated regressor is included in the design matrix, especially when constructing the Wald test under the null hypothesis of no systemic risk in the network. To obtain a measure of the performance of the testing procedure, we need to compare the standard estimation procedure without the variance-covariance matrix correction against the estimation procedure that includes the adjustment, in both cases under the null hypothesis. Then, under the alternative hypothesis of deviations, that is, existence of non-zero systemic risk effect in the network (see, Katsouris (2021, 2023b)), then the larger this distance is, under the adjustement then the larger the power function of the test would be when all the individual quantile predictive regression models of the system correspond to a non-zero coefficient of node-specific systemic risk.

- When the nonstationary regressors for the two nodes are the same then second-order effects can contribute to the covariance of the estimator. However, due to the proposed dependence structure (see, Katsouris (2021, 2023b)), in practice we ensure that these nonstationary (or near unit roots) are not equivalent across the different predictive regression models. Notice that the quantity $\widehat{y}_{1,t}$ is the forecaster VaR which is obtained using time series observations from t to $(t-1)$. Therefore, it might have an effect on the convergence of the parameter estimate δ. In summary, we need to control for two effects, that is, a bias effect on the estimate of δ and a variance effect. In particular, the bias effect will disappear when $\boldsymbol{x}_{t-1}$ and the regressors used to estimate $\widehat{y}_{1,t}$ are uncorrelated. The variance effect appears due to the choice of the generated regressors. Standard OLS estimation methods do not correct for the particular variance effect. On the other hand, the variance effect does not influence the inference methodology for the estimator $\hat{\beta}_2$ in which case an important property to hold is to obtain a consistent estimator regardless of the presence of nonstationary regressors. In the derivations below, we use $\widehat{y}_{1,t}$ e.g., the estimated VaR under nonstationarity (when we employ the quantile predictive regression model).

Therefore, the IVX estimator for the second econometric specification (CoVaR quantile predictive regression) is obtained by the following expression:

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}_2^{ivx} &= \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1}\boldsymbol{z}'_{2,t-1}\right)^{-1}\left(\sum_{t=1}^{n} \tilde{y}_{2,t}\boldsymbol{z}'_{2,t-1}\right) \\
&= \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1}\boldsymbol{z}'_{2,t-1}\right)^{-1}\left(\sum_{t=1}^{n} \left[\boldsymbol{\beta}'_2\bar{\boldsymbol{x}}_{2,t-1} + \delta\widehat{y}_{1,t} + u_{2,t} - \widehat{\delta}^{ols}\widehat{y}_{1,t}\right]\boldsymbol{z}'_{2,t-1}\right) \\
&= \boldsymbol{\beta}_2 - \left(\widehat{\delta}^{ols} - \delta\right)\left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1}\boldsymbol{z}'_{2,t-1}\right)^{-1}\left(\sum_{t=1}^{n} \widehat{y}_{1,t}\boldsymbol{z}'_{2,t-1}\right) + \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1}\boldsymbol{z}'_{2,t-1}\right)^{-1}\left(\sum_{t=1}^{n} u_{2,t}\boldsymbol{z}'_{2,t-1}\right)
\end{aligned}$$

which gives

$$\left(\tilde{\boldsymbol{\beta}}_2^{ivx} - \boldsymbol{\beta}_2\right) = -\left(\widehat{\delta^{ols}} - \delta\right)\left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1}\boldsymbol{z}'_{2,t-1}\right)^{-1}\left(\sum_{t=1}^{n} \widehat{y}_{1,t}\boldsymbol{z}'_{2,t-1}\right) + \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1}\boldsymbol{z}'_{2,t-1}\right)^{-1}\left(\sum_{t=1}^{n} u_{2,t}\boldsymbol{z}'_{2,t-1}\right)$$

Notice that for the first term we have that

$$\left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1}\boldsymbol{z}'_{2,t-1}\right)^{-1}\left(\sum_{t=1}^{n} \widehat{\boldsymbol{\beta}}_1^{ivx}\bar{\boldsymbol{x}}_{1,t-1}\boldsymbol{z}'_{2,t-1}\right) \equiv \widehat{\boldsymbol{\beta}}_1^{ivx} \tag{C.43}$$

Notice that the above expression only holds when we have identical nonstationary regressors such that $\bar{\boldsymbol{x}}_{1,t-1} \equiv \bar{\boldsymbol{x}}_{2,t-1}$. Therefore, we obtain the following expression

$$\left(\tilde{\boldsymbol{\beta}}_2^{ivx} - \boldsymbol{\beta}_2\right) = -\left(\widehat{\delta}^{ols} - \delta\right) \times \widehat{\boldsymbol{\beta}}_1^{ivx} + \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1}\boldsymbol{z}'_{2,t-1}\right)^{-1}\left(\sum_{t=1}^{n} u_{2,t}\boldsymbol{z}'_{2,t-1}\right) \tag{C.44}$$

Remark C.3. Notice that when $\boldsymbol{x}_{t-1}$ and $\widehat{y}_{1,t}$ are not correlated, then $\widehat{\theta}^{ols}$ is a consistent estimator of δ. However, in the case we employ the same nonstationary predictors to estimate the value of $\widehat{y}_{1,t}$ then these two quantities are correlated and then it is not so clear to determine the correlation structure due to the presence of different effects. On the other hand, the dependence structure proposed in Katsouris (2021, 2023b) ensures that this case doesn't occur, so that we can consistently estimate these two parameters which are important for inference purposes.

Note that $\widehat{\boldsymbol{\beta}}_1^{ivx}$ corresponds to the IVX estimator of the quantile predictive regression model for the VaR risk measure. Additionally, note that from Lemma B4 of Kostakis et al. (2015) it holds,

$$\frac{1}{n^{\frac{1+\gamma_z}{2}}} \sum_{t=1}^{n} \boldsymbol{z}_{2,t-1} u_{2,t} \Rightarrow \mathcal{N}\left(0, \sigma^2_{u_2} \times \boldsymbol{V}_{c_z}\right). \tag{C.45}$$

Therefore, by incorporating the convergence rate we have that

$$n^{\frac{1+\gamma_z}{2}} \left(\tilde{\boldsymbol{\beta}}_2^{ivx} - \boldsymbol{\beta}_2 \right) = n^{\frac{1+\gamma_z}{2}} \left(\delta - \widehat{\delta}^{ols} \right) \times \widehat{\boldsymbol{\beta}}_1^{ivx} + \left(\frac{1}{n^{1+\gamma_z}} \sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1} \boldsymbol{z}'_{2,t-1} \right)^{-1} \left(\frac{1}{n^{\frac{1+\gamma_z}{2}}} \sum_{t=1}^{n} \boldsymbol{z}_{2,t-1} u_{2,t} \right).$$

Thus, for the LUR regressor case we have that the second term of expression (C.46) converges to

$$\left(\frac{1}{n^{1+\gamma_z}} \sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1} \boldsymbol{z}'_{2,t-1} \right)^{-1} \left(\frac{1}{n^{\frac{1+\gamma_z}{2}}} \sum_{t=1}^{n} \boldsymbol{z}_{2,t-1} u_{2,t} \right) \Rightarrow - \left(\boldsymbol{\Omega}_{uu} + \int_0^1 \boldsymbol{J}_c^{\mu} d J_c' \right) \boldsymbol{C}_z^{-1} \times \mathcal{N}\left(0, \sigma^2_{u_2} \times \boldsymbol{V}_{C_z}\right).$$

To see this, from the Appendix of KMS expression (27) gives that

$$\left(\frac{1}{n^{1+\gamma_z}} \sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1} \boldsymbol{z}'_{2,t-1} \right) = \left(\boldsymbol{\Omega}_{uu} + \boldsymbol{V}_C \boldsymbol{C} \right) \boldsymbol{C}_z^{-1} + o_p(1). \tag{C.46}$$

where

$$\boldsymbol{V}_C = \int_0^{\infty} e^{r\boldsymbol{C}} \boldsymbol{\Omega}_{uu} e^{r\boldsymbol{C}} dr \quad \text{and} \quad \boldsymbol{V}_{C_z} = \int_0^{\infty} e^{r\boldsymbol{C}_z} \boldsymbol{\Omega}_{uu} e^{r\boldsymbol{C}_z} dr. \tag{C.47}$$

Notice that by Kostakis et al. (2015) we have that $\boldsymbol{\Psi}_{uu} = \left(\boldsymbol{\Omega}_{uu} + \int_0^1 \boldsymbol{J}_c^{\mu} d J_c' \right)$ since we consider local unit root regressors. Therefore, similar to Theorem A (i) of KMS it holds that

$$\left(\frac{1}{n^{1+\gamma_z}} \sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1} \boldsymbol{z}'_{2,t-1} \right)^{-1} \left(\frac{1}{n^{\frac{1+\gamma_z}{2}}} \sum_{t=1}^{n} \boldsymbol{z}_{2,t-1} u_{2,t} \right) \Rightarrow \mathcal{MN}\left(0, \sigma^2_{u_2} \times \left(\boldsymbol{\Psi}^{-1}_{u_2 u_2} \right)' \boldsymbol{C}_z \boldsymbol{V}_{C_z} \boldsymbol{C}_z \boldsymbol{\Psi}^{-1}_{u_2 u_2} \right)$$

which is a mixed Gaussian random variate since the covariance matrix is a function of the OU process. Next we investigate the asymptotic behaviour of the first term of expression (C.46):

$$\mathcal{A} := n^{\frac{1+\gamma_z}{2}} \left(\delta - \widehat{\delta}^{ols} \right) \times \widehat{\boldsymbol{\beta}}_1^{ivx} \tag{C.48}$$

Specifically, assuming that node 1 and 2 have a set of nonstationary regressors that are generated from non-identical stochastic processes (i.e., not identical regressors), then it follows that

$$\begin{aligned} \mathcal{A} &= \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1} \boldsymbol{z}'_{2,t-1} \right)^{-1} \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{1,t-1} \boldsymbol{z}'_{2,t-1} \right) \times \widehat{\boldsymbol{\beta}_1^{ivx}} \\ &= \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1} \boldsymbol{z}'_{2,t-1} \right)^{-1} \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{1,t-1} \boldsymbol{z}'_{2,t-1} \right) \times \left[\boldsymbol{\beta}_1 + \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{1,t-1} \boldsymbol{z}'_{1,t-1} \right)^{-1} \left(\sum_{t=1}^{n} \bar{u}_{1,t} \boldsymbol{z}'_{1,t-1} \right) \right]. \end{aligned}$$

Regarding the asymptotic behaviour of the first term of the expression we have that the IVX estimator from the quantile predictive regression that corresponds to the VaR risk measure is

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_1^{ivx} &= \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{1,t-1}\boldsymbol{z}'_{1,t-1}\right)^{-1}\left(\sum_{t=1}^{n} y_{1,t}\boldsymbol{z}'_{1,t-1}\right), \quad \bar{y}_{1,t} = \boldsymbol{\beta}'_1\bar{\boldsymbol{x}}_{1,t-1} + \bar{u}_{1,t} \\ &= \boldsymbol{\beta}_1 + \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{1,t-1}\boldsymbol{z}'_{1,t-1}\right)^{-1}\left(\sum_{t=1}^{n} \bar{u}_{1,t}\boldsymbol{z}'_{1,t-1}\right)\end{aligned}$$

Moreover, the number of nonstationary regressors for both econometric specifications is the same although these two sets of regressors possibly have different nuisance parameters of persistence. The limiting distributions for their IVX estimators have the same dimensions such that

$$\begin{aligned}n^{\frac{1+\gamma_z}{2}}\left(\widehat{\boldsymbol{\beta}}_1^{ivx} - \boldsymbol{\beta}_1\right) &= \left(\frac{1}{n^{1+\gamma_z}}\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{1,t-1}\boldsymbol{z}'_{1,t-1}\right)^{-1}\left(\frac{1}{n^{\frac{1+\gamma_z}{2}}}\sum_{t=1}^{n} \bar{u}_{1,t}\boldsymbol{z}'_{1,t-1}\right) \\ &\Rightarrow \mathcal{MN}\left(0, \sigma^2_{u_1} \times \left(\boldsymbol{\Psi}^{-1}_{u_1u_1}\right)' \boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}^{-1}_{u_1u_1}\right).\end{aligned}$$

Rearranging the above expressions, we obtain the following limit result

$$\begin{aligned}n^{\frac{1+\gamma_z}{2}}\left[\left(\tilde{\boldsymbol{\beta}}_2^{ivx} - \boldsymbol{\beta}_2\right) - \boldsymbol{\beta}_1\right] &= \left(\delta - \widehat{\delta}^{ols}\right) \times \mathcal{MN}\left(0, \sigma^2_{u_1} \times \left(\boldsymbol{\Psi}^{-1}_{u_1u_1}\right)' \boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}^{-1}_{u_1u_1}\right) \\ &\quad + \mathcal{MN}\left(0, \sigma^2_{u_2} \times \left(\boldsymbol{\Psi}^{-1}_{u_2u_2}\right)' \boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}^{-1}_{u_2u_2}\right). \qquad \text{(C.49)}\end{aligned}$$

Remark C.4. In this paper we are not specifically attempting to solve the endogeneity issue in quantile predictive regressions models. We employ a well-investigated method in the literature that tackles the endogeneity problem, meaning it produces weak convergence results into standard asymptotic distributions (see, Lee (2016), Fan and Lee (2019)). Alternative approaches that tackle the endogeneity problem and produce uniform valid inference regardless of the unknown persistence properties are proposed in the frameworks of Cai et al. (2023) and Liu et al. (2023). In contrast, our study provides a general framework and establish the asymptotic properties for quantile predictive regression models with a generated regressor, which is particularly useful when jointly estimating the risk measure pair of (VaR, CoVaR) under the presence of time series nonstationarity. We conjecture that regardless of the estimation method employed to robustify the quantile-based model parameters to the unknown persistence, when estimating the CoVaR based on the specifications we follow (Härdle et al. (2016)) and using the LUR parametrization, similar challenges are needed to be tackled to ensure statistical properties hold.

Asymptotic Properties

Next, consider for a moment the limiting distribution of $\left(\widehat{\delta}^{ols} - \delta\right)$ where $\bar{y}_{2,t} = \delta \widehat{y}_{1,t} + \eta_t$. Therefore, the OLS estimator of δ is given by $\widehat{\delta}^{ols} = \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \left(\sum_{t=1}^{n} \bar{y}_{2,t} \widehat{y}_{1,t} \right)$, where $\widehat{y}_{1,t} := \widehat{\boldsymbol{\beta}}_1^{\prime ivx} \boldsymbol{x}_{1,t-1}$ which implies that it can be expressed as below

$$\widehat{\delta}^{ols} = \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \left(\sum_{t=1}^{n} \bar{y}_{2,t} \widehat{\boldsymbol{\beta}_1^{\prime ivx} \boldsymbol{x}_{1,t-1}} \right) = \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \left(\sum_{t=1}^{n} \bar{y}_{2,t} \boldsymbol{x}_{1,t-1} \right) \widehat{\boldsymbol{\beta}_1^{\prime ivx}}. \tag{C.50}$$

However, $\bar{y}_{2,t} = \boldsymbol{\beta}_2' \bar{\boldsymbol{x}}_{2,t-1} + \delta \widehat{y}_{1,t} + u_{2,t}$ which implies that

$$\begin{aligned} \widehat{\delta}^{ols} &= \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \left(\sum_{t=1}^{n} \left[\boldsymbol{\beta}_2' \bar{\boldsymbol{x}}_{2,t-1} + \delta \widehat{y}_{1,t} + u_{2,t} \right] \boldsymbol{x}_{1,t-1}' \right) \widehat{\boldsymbol{\beta}_1^{ivx}} \\ &= \left(\sum_{t=1}^{n} \widehat{y}_{1,t} \widehat{y}_{1,t}' \right)^{-1} \times \left\{ \boldsymbol{\beta}_2' \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1} \boldsymbol{x}_{1,t-1}' \right) + \delta \left(\sum_{t=1}^{n} \widehat{y}_{1,t} \boldsymbol{x}_{1,t-1}' \right) \left(\sum_{t=1}^{n} u_{2,t} \boldsymbol{x}_{1,t-1}' \right) \right\} \widehat{\boldsymbol{\beta}_1^{ivx}} \end{aligned}$$

where

$$\left(\sum_{t=1}^{n} \widehat{y}_{1,t} \widehat{y}_{1,t}' \right)^{-1} = \left(\sum_{t=1}^{n} \boldsymbol{\beta}_1^{\prime ivx} \widehat{\boldsymbol{x}_{1,t-1} \boldsymbol{x}_{1,t-1}'} \widehat{\boldsymbol{\beta}}_1^{ivx} \right)^{-1} = \left(\widehat{\boldsymbol{\beta}}_1^{ivx} \right)^{-1} \left(\sum_{t=1}^{n} \boldsymbol{x}_{1,t-1} \boldsymbol{x}_{1,t-1}' \right)^{-1} \left(\widehat{\boldsymbol{\beta}}_1^{\prime ivx} \right)^{-1}$$

Therefore, we obtain that

$$\widehat{\delta}^{ols} = \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \boldsymbol{\beta}_2' \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1} \boldsymbol{x}_{1,t-1}' \right) \widehat{\boldsymbol{\beta}_1^{ivx}} + \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \delta \left(\sum_{t=1}^{n} \widehat{y}_{1,t} \boldsymbol{x}_{1,t-1}' \right) \widehat{\boldsymbol{\beta}_1^{ivx}} + \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \left(\sum_{t=1}^{n} u_{2,t} \boldsymbol{x}_{1,t}' \right.$$

Notice that the second term above becomes:

$$\left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \delta \left(\sum_{t=1}^{n} \boldsymbol{\beta}_1^{\prime ivx} \widehat{\boldsymbol{x}_{1,t-1} \boldsymbol{x}_{1,t-1}'} \widehat{\boldsymbol{\beta}}_1^{ivx} \right) = \delta \left(\sum_{t=1}^{n} \boldsymbol{\beta}_1^{\prime ivx} \widehat{\boldsymbol{x}_{1,t-1} \boldsymbol{x}_{1,t-1}'} \widehat{\boldsymbol{\beta}}_1^{ivx} \right)^{-1} \left(\sum_{t=1}^{n} \boldsymbol{\beta}_1^{\prime ivx} \widehat{\boldsymbol{x}_{1,t-1} \boldsymbol{x}_{1,t-1}'} \widehat{\boldsymbol{\beta}}_1^{ivx} \right) = \delta$$

Thus, we have that

$$\left(\widehat{\delta}^{ols} - \delta \right) = \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \boldsymbol{\beta}_2' \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1} \boldsymbol{x}_{1,t-1}' \right) \widehat{\boldsymbol{\beta}}_1^{ivx} + \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^2 \right)^{-1} \left(\sum_{t=1}^{n} u_{2,t} \boldsymbol{x}_{1,t-1}' \right) \widehat{\boldsymbol{\beta}_1^{ivx}}.$$

Furthermore, since the nonstationary regressors of node 1 are not correlated with the nonstationary regressors of node 2, then the term $\underset{n\to\infty}{\text{plim}}\left(\sum_{t=1}^{n} u_{2,t}\boldsymbol{x}_{1,t-1}'\right) = 0$, which implies that

$$\begin{aligned}\left(\widehat{\delta}^{ols} - \delta\right) &= \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^{2}\right)^{-1} \boldsymbol{\beta}_2' \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{2,t-1}\boldsymbol{x}_{1,t-1}'\right) \times \widehat{\boldsymbol{\beta}_1^{ivx}} \\ &= \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^{2}\right)^{-1} \boldsymbol{\beta}_2' \left(\mathbb{E}[\boldsymbol{x}_{2,t-1}\boldsymbol{x}_{1,t-1}']\right) \times \left[\boldsymbol{\beta}_1 + \left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{1,t-1}\boldsymbol{z}_{1,t-1}'\right)^{-1}\left(\sum_{t=1}^{n} \bar{u}_{1,t}\boldsymbol{z}_{1,t-1}'\right)\right] \\ &= \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^{2}\right)^{-1} \boldsymbol{\beta}_2' \left(\mathbb{E}[\boldsymbol{x}_{2,t-1}\boldsymbol{x}_{1,t-1}']\right)\boldsymbol{\beta}_1 + \left(\sum_{t=1}^{n} \widehat{y}_{1,t}^{2}\right)^{-1} \boldsymbol{\beta}_2' \left(\mathbb{E}[\boldsymbol{x}_{2,t-1}\boldsymbol{x}_{1,t-1}']\right) \\ &\quad\times \left[\left(\sum_{t=1}^{n} \bar{\boldsymbol{x}}_{1,t-1}\boldsymbol{z}_{1,t-1}'\right)^{-1}\left(\sum_{t=1}^{n} \bar{u}_{1,t}\boldsymbol{z}_{1,t-1}'\right)\right] + o_p(1).\end{aligned}$$

Since the term $\left(\sum_{t=1}^{n} \widehat{y}_{1,t}^{2}\right)$ is always positive then we assume that $\underset{n\to\infty}{\text{plim}}\left(\sum_{t=1}^{n} \widehat{y}_{1,t}^{2}\right)^{-1} = \mathcal{K}$ where $\mathcal{K}$ is some positive constant. Therefore, putting all related expressions together we obtain

$$\begin{aligned}n^{\frac{1+\gamma_z}{2}}\left[\left(\tilde{\boldsymbol{\beta}}_2^{ivx} - \boldsymbol{\beta}_2\right) - \boldsymbol{\beta}_1\right] &= \left\{\mathcal{K}\boldsymbol{\beta}_2'\left(\mathbb{E}[\boldsymbol{x}_{2,t-1}\boldsymbol{x}_{1,t-1}']\right)\boldsymbol{\beta}_1 + \mathcal{K}\boldsymbol{\beta}_2'\left(\mathbb{E}[\boldsymbol{x}_{2,t-1}\boldsymbol{x}_{1,t-1}']\right)\right. \\ &\quad\left.\times \mathcal{MN}\left(0, \sigma_{u_1}^2 \times \left(\boldsymbol{\Psi}_{u_1u_1}^{-1}\right)'\boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}_{u_1u_1}^{-1}\right)\right\} \\ &\quad\times \mathcal{MN}\left(0, \sigma_{u_1}^2 \times \left(\boldsymbol{\Psi}_{u_1u_1}^{-1}\right)'\boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}_{u_1u_1}^{-1}\right) \\ &\quad+ \mathcal{MN}\left(0, \sigma_{u_2}^2 \times \left(\boldsymbol{\Psi}_{u_2u_2}^{-1}\right)'\boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}_{u_2u_2}^{-1}\right) \\ &= \mathcal{K}\boldsymbol{\beta}_2'\left(\mathbb{E}[\boldsymbol{x}_{2,t-1}\boldsymbol{x}_{1,t-1}']\right)\boldsymbol{\beta}_1 \times \mathcal{MN}\left(0, \sigma_{u_1}^2 \times \left(\boldsymbol{\Psi}_{u_1u_1}^{-1}\right)'\boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}_{u_1u_1}^{-1}\right) \\ &\quad+ \mathcal{K}\boldsymbol{\beta}_2'\left(\mathbb{E}[\boldsymbol{x}_{2,t-1}\boldsymbol{x}_{1,t-1}']\right) \times \left\{\mathcal{MN}\left(0, \sigma_{u_1}^2 \times \left(\boldsymbol{\Psi}_{u_1u_1}^{-1}\right)'\boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}_{u_1u_1}^{-1}\right)\right\}^2 \\ &\quad+ \mathcal{MN}\left(0, \sigma_{u_2}^2 \times \left(\boldsymbol{\Psi}_{u_2u_2}^{-1}\right)'\boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}_{u_2u_2}^{-1}\right).\end{aligned}$$

The second term above $\left\{\mathcal{MN}\left(0, \sigma_{u_1}^2 \times \left(\boldsymbol{\Psi}_{u_1u_1}^{-1}\right)'\boldsymbol{C}_z\boldsymbol{V}_{C_z}\boldsymbol{C}_z\boldsymbol{\Psi}_{u_1u_1}^{-1}\right)\right\}^2$, corresponds to a squared mixed Gaussian distribution which gives a form of a χ^2 distribution. Thus, the limiting distribution of (C.51) is a linear combination of a χ^2 distribution and a mixed Gaussian distribution, which is a Generalized χ^2 distribution. Thus the y_t inherits the properties of x_t, through a cointegrating relation, especially our goal is to have a stationary innovation sequence such that $u_t \sim I(0)$. Under the assumption of a consistent estimator for $\hat{\delta}^{ols}$ then we have convergence in probability to zero such that $\left(\hat{\delta}^{ols} - \delta\right) \overset{p}{\to} 0$. Thus, the first term of the expression asymptotically tends to zero (negligible) which implies that the second term converges into a mixed Gaussian distribution.

C.4 Doubly Corrected Estimation: Conditional Quantile Case

In this section we consider the double corrected conditional quantile estimation methodology which is the main focus of our research study. The derivations presented in the previous sections were useful to shed light on some challenges we have to overcome to develop a robust framework in the proposed setting of quantile predictive regression models. In terms of estimation approach we focus on the QR estimators under nonstationarity which are adjusted using the IVX filtration. During the first stage of the estimation procedure in the same spirit as several studies in the literature related to the joint estimation of the risk measures of $(\mathsf{VaR}, \mathsf{CoVaR})$ (see, Härdle et al. (2016) and Patton et al. (2019) among others), the VaR is estimated given a fixed quantile $\tau \in (0,1)$ level but using information on the nonstationary properties of regressors as given by the quantile predictive regression system below

Stage 1:

$$y_t^{(1)} = \alpha_1(\tau) + \boldsymbol{\beta}_1'(\tau)\boldsymbol{x}_{t-1}^{(1)} + u_t^{(1)}(\tau), \quad \text{for} \quad t = 1, ..., n \tag{C.51}$$

$$\boldsymbol{x}_t^{(1)} = \boldsymbol{R}_n^{(1)}\boldsymbol{x}_{t-1}^{(1)} + \boldsymbol{v}_t^{(1)} \tag{C.52}$$

such that $\alpha_1(\tau) + \boldsymbol{\beta}'(\tau)\boldsymbol{x}_{t-1}^{(1)}$ is the $\tau-$conditional quantile of y_t given $\boldsymbol{x}_{t-1}$, where the unknown parameters $(\alpha_1(\tau), \boldsymbol{\beta}_1(\tau))$ are estimated using the QR objective function.

Stage 2:

$$y_t^{(2)} = \alpha_2(\tau) + \boldsymbol{\beta}_2'(\tau)\boldsymbol{x}_{t-1}^{(2)} + \delta(\tau)\widehat{y}_{1,t}\left(\hat{\boldsymbol{\theta}}_1^{ivx}(\tau)\right) + u_t^{(2)}(\tau), \quad \text{for} \quad t = 1, ..., n \tag{C.53}$$

$$\boldsymbol{x}_t^{(2)} = \boldsymbol{R}_n^{(2)}\boldsymbol{x}_{t-1}^{(2)} + \boldsymbol{v}_t^{(2)} \tag{C.54}$$

where the parameter of interest corresponds to the estimator $\boldsymbol{\vartheta}_n^{ivx} = (\alpha_2(\tau), \boldsymbol{\beta}_2'(\tau), \delta(\tau))$.

Remark C.5. Similar to the remark found in Wang and Zhao (2016), we point out that it is up to the practitioner to determine the specific model and parameter estimation method, which are the starting point to carry out any subsequent CoVaR estimation and inference. In our study, we use the nonstationary quantile predictive regression when obtaining an estimate for CoVaR. Thus, our approach gives a $\sqrt{n^{1+\delta}}-$consistent estimator for the unknown coefficient of systemic risk based on the IVX filter. Recall that $\widehat{\boldsymbol{\beta}}^{\star}(\tau) := \underset{\boldsymbol{\beta}^{\star} \in \mathbb{R}^{p+1}}{\mathsf{arg\ min}} \sum_{i=1}^{n} \rho_{\tau}\left(y_t - \tilde{\boldsymbol{x}}_{t-1}'\boldsymbol{\beta}^{\star}\right)$ where the set of regressors $\tilde{\boldsymbol{x}}_{t-1} = \left[\boldsymbol{x}_{t-1} \quad \widehat{x}_t^{\mathsf{VaR}}\right]$. The particular optimization function corresponds to the second-stage procedure which gives the model estimates to construct the one-period ahead forecasted CoVaR risk measure. However, our approach differs from the estimation method of **?** and Härdle et al. (2016) since nonstationary time series data are modelled via the LUR parametrization and thus we employ the QR-IVX estimator proposed by Lee (2016) but with the additional generated regressor from the first-stage procedure.

Appendix D

Tables

D.1 Tables of Chapter 2

Table A1: Empirical size with nominal level $\alpha = 5\%$.

sup OLS-Wald test $(\beta_1 = \beta_2)$									
				$c = 1$					
T\ρ	**-0.9**	**-0.7**	**-0.5**	**-0.3**	**0**	**0.3**	**0.5**	**0.7**	**0.9**
100	0.0804	0.0736	0.0704	0.0672	0.0556	0.0540	0.0512	0.0664	0.0700
250	0.0968	0.0812	0.0700	0.0628	0.0564	0.0608	0.0696	0.0764	0.0904
500	0.0924	0.0812	0.0720	0.0644	0.0608	0.0668	0.0744	0.0916	0.1044
1000	0.1060	0.0824	0.0788	0.0716	0.0684	0.0668	0.0760	0.0860	0.1032
				$c = 5$					
T\ρ	**-0.9**	**-0.7**	**-0.5**	**-0.3**	**0**	**0.3**	**0.5**	**0.7**	**0.9**
100	0.0588	0.0564	0.0592	0.0548	0.0492	0.0480	0.0500	0.0552	0.0560
250	0.0740	0.0608	0.0596	0.0548	0.0516	0.0504	0.0568	0.0692	0.0744
500	0.0860	0.0784	0.0696	0.0648	0.0564	0.0572	0.0632	0.0680	0.0848
1000	0.0908	0.0816	0.0732	0.0708	0.0636	0.0596	0.0620	0.0724	0.0864
				$c = 10$					
T\ρ	**-0.9**	**-0.7**	**-0.5**	**-0.3**	**0**	**0.3**	**0.5**	**0.7**	**0.9**
100	0.0448	0.0524	0.0508	0.0488	0.0444	0.0448	0.0480	0.0516	0.0492
250	0.0572	0.0540	0.0528	0.0488	0.0472	0.0472	0.0548	0.0608	0.0604
500	0.0788	0.0768	0.0676	0.0604	0.0544	0.0592	0.0592	0.0648	0.0736
1000	0.0796	0.0768	0.0704	0.0620	0.0568	0.0528	0.0552	0.0652	0.0704
				$c = 20$					
T\ρ	**-0.9**	**-0.7**	**-0.5**	**-0.3**	**0**	**0.3**	**0.5**	**0.7**	**0.9**
100	0.0380	0.0416	0.0440	0.0432	0.0476	0.0416	0.0440	0.0440	0.0388
250	0.0560	0.0476	0.0480	0.0408	0.0436	0.0460	0.0496	0.0480	0.0484
500	0.0620	0.0592	0.0600	0.0560	0.0512	0.0504	0.0564	0.0596	0.0648
1000	0.0676	0.0656	0.0608	0.0548	0.0496	0.0520	0.0548	0.0576	0.0572

Table A1 presents finite-sample empirical sizes for the sup Wald-OLS test, with nominal level $\alpha = 5\%$ for $B = 5,000$ replications. The predictive regression model under the null hypothesis, $H_0 : \beta_1 = \beta_2$, is given by, $y_t = 0.25x_{t-1} + u_t$, $x_t = (1 - \frac{c}{T})x_{t-1} + v_t$, with $\Sigma_{ee} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, where $\rho = \dfrac{\sigma_{uv}}{\sigma_u \sigma_v}$, is the contemporaneous correlation coefficient.

Table A2: Empirical size with nominal level $\alpha = 5\%$.

sup IVX-Wald test $(\alpha_1 = \alpha_2, \beta_1 = \beta_2)$									
c = 1 $(c_z = 1,\ \delta_z = 0.75,\ c_\alpha = 13.42)$									
T\\ρ	**-0.9**	**-0.7**	**-0.5**	**-0.3**	**0**	**0.3**	**0.5**	**0.7**	**0.9**
100	0.0821	0.0781	0.0821	0.0680	0.0668	0.0596	0.0610	0.0516	0.0526
250	0.0636	0.0878	0.0586	0.0624	0.0532	0.0522	0.0506	0.0538	0.0489
500	0.0616	0.0508	0.0604	0.0606	0.0586	0.0582	0.0562	0.0572	0.0526
1000	0.0576	0.0644	0.0412	0.0564	0.0400	0.0460	0.0598	0.0668	0.0544
c = 5 $(c_z = 1,\ \delta = 0.75,\ c_\alpha = 13.42)$									
T\\ρ	**-0.9**	**-0.7**	**-0.5**	**-0.3**	**0**	**0.3**	**0.5**	**0.7**	**0.9**
100	0.0634	0.0604	0.0580	0.0526	0.0450	0.0546	0.0596	0.0550	0.0610
250	0.0598	0.0662	0.0456	0.0548	0.0448	0.0526	0.0452	0.0461	0.0686
500	0.0544	0.0692	0.0510	0.0650	0.0580	0.0582	0.0548	0.0518	0.0516
1000	0.0492	0.0488	0.0618	0.0436	0.0590	0.0450	0.0564	0.0522	0.0488

c = 1 $(c_z = 1,\ \delta_z = 0.75,\ c^*_\alpha)$									
T\\ρ	**-0.9**	**-0.7**	**-0.5**	**-0.3**	**0**	**0.3**	**0.5**	**0.7**	**0.9**
100	0.0686	0.0710	0.0708	0.0568	0.0632	0.0514	0.0638	0.0492	0.0582
250	0.0604	0.0736	0.0816	0.0631	0.0790	0.0558	0.0622	0.0438	0.0542
500	0.0538	0.0780	0.0828	0.0638	0.0760	0.0536	0.0692	0.0532	0.0501
1000	0.0534	0.0548	0.0672	0.0448	0.0356	0.0454	0.0686	0.0548	0.0564
c = 5 $(c_z = 1,\ \delta_z = 0.75,\ c^*_\alpha)$									
T\\ρ	**-0.9**	**-0.7**	**-0.5**	**-0.3**	**0**	**0.3**	**0.5**	**0.7**	**0.9**
100	0.0722	0.0734	0.0736	0.0604	0.0668	0.0552	0.0566	0.0532	0.0518
250	0.0588	0.0694	0.0714	0.0548	0.0658	0.0544	0.0766	0.0780	0.0648
500	0.0576	0.0784	0.0716	0.0606	0.0656	0.0594	0.0508	0.0508	0.0522
1000	0.0490	0.0828	0.0578	0.0628	0.0658	0.0556	0.0630	0.0498	0.0526

Table A2 presents finite-sample sizes for the sup Wald-IVX test, with nominal size $\alpha = 5\%$. The predictive regression model under the null hypothesis is given by $y_t = 0.25 + 0.5x_{t-1} + u_t$, $x_t = (1 - \frac{c_1}{n})x_{t-1} + v_t$, with $\Sigma_{ee} = \begin{bmatrix} 0.25 & \sigma_{uv} \\ \sigma_{uv} & 0.75 \end{bmatrix}$, $\rho = \dfrac{\sigma_{uv}}{\sigma_u \sigma_v}$, is the contemporaneous correlation coefficient. Furthermore, for the IVX estimation step, we use an IVX persistence parameter[1] $\delta_z \in \{0.75, 0.95\}$ and the localizing coefficient is set to $c_z = 1$. The number of replications is $B = 5,000$ and the critical values are based on the bootstrap.

Remark D.1. As we can observe from the empirical size results presented on Table A2, when using the sup IVX-Wald statistic for testing parameter constancy in the predictive regression models with LUR predictors are based on conventional asymptotic critical values, then we obtain excessive size distortions. This is especially the case for values of the contemporaneous correlation coefficient, ρ, close to the vicinity of unity. Therefore, to obtain correctly controlled empirical size within this context, bootstrap based inference is necessary.

[1]Notice that the IVX persistent parameter δ_z is chosen to ensure that the instrument z_t is mildly integrated. A value of $\delta_z = 0.95$ it has been documented in the literature to work well (see, Phillips and Lee (2016)). For comparability purposes we also use an exponent rate with value $\delta_z = 0.75$.

Table A3: Predictability Tests for the Equity Premium

Predictor	$\widehat{\beta}^{\text{OLS}}$	t^{HAC}	R^2	$\widehat{\beta}^{\text{IVX}}$	$\mathcal{W}^{\text{IVX}}$	$sup\mathcal{W}^{\text{OLS}}$	$sup\mathcal{W}^{\text{IVX}}$
Panel A:	1946Q1 - 2019Q4						
Dividend payout ratio	0.0056	0.3898	0.0003				
Long-term yield	-0.0824	-1.5958	0.0032				
Dividend yield	0.0150	1.9880	0.0047				
Dividend-price-ratio	0.0142	1.8741	0.0042				
T-bill rate	-0.1043	-2.1461*	0.0060				
Earnings-price-ratio	0.0115	1.2885	0.0028				
Book-to-market ratio	0.0042	0.6590	0.0006				
Default yield spread	0.0150	1.9880*	0.0047				
Net equity expansion	-0.0501	-0.6149	0.0005				
Term spread	0.1813	1.6834	0.0035				
Inflation rate	-0.9075	-2.6520**	0.0096				
Panel B:	1990Q1 - 2019Q4						
Dividend payout ratio	0.0028	0.1597	0.0001				
Long-term yield	-0.0883	-0.8409	0.0017				
Dividend yield	0.0354	1.6497	0.0102				
Dividend-price-ratio	0.0351	1.6389	0.0101				
T-bill rate	-0.0393	-0.3885	0.0005				
Earnings-price-ratio	0.0173	0.8403	0.0041				
Book-to-market ratio	0.0317	1.0439	0.0041				
Default yield spread	0.5275	0.5506	0.0025				
Net equity expansion	0.1188	0.9655	0.0037				
Term spread	-0.0727	-0.4442	0.0005				
Inflation rate	0.2005	0.2831	0.0003				

Table A3 presents simple predictability tests of the null hypothesis $\beta = 0$ and model estimates based on the predictive regression with a single predictor given by $y_t = \alpha + \beta x_{t-1} + u_t$, applied to all individual predictors. For the t-tests we consider a t-ratio based on the HAC (Newey-West) covariance estimator. For all test statistics, we denote the rejection probabilities under the null hypothesis of no predictability at significance levels 1%(***), 5%(**), and 10%(*), respectively.

Table A4: Structural Break Tests for predictors

Panel A:	1990Q1 - 2019Q4					
Predictor		$\widehat{\rho}$	R^2	BP Seq F-test	Max LR F-test	Exp LR F-test
Dividend payout ratio		0.7477	0.5592	4.0061	49.4450***	19.1970***
Long-term yield		0.0429	0.0046	3.4002	2.3062	0.3676
Dividend yield		0.0424	0.0001	3.5761	7.2497	1.4443
Dividend-price-ratio		0.0511	0.0008	3.6305	7.5262	1.5400
T-bill rate		0.4734	0.2169	2.6279	5.4363	0.6033
Earnings-price-ratio		0.4385	0.1922	1.6033	14.6679***	4.6243***
Book-to-market ratio		-0.0439	0.0012	5.5639	3.1215	0.1470
Default yield spread		0.4616	0.2131	3.1040	7.4651	1.9635**
Net equity expansion		0.2440	0.0596	0.9001	2.5452	0.2197
Term spread		0.1035	0.0107	1.6129	2.1243	0.2136
Inflation rate		-0.0979	0.0096	3.2371	5.1032	0.9540
Predictor	$\widehat{\mu}$	$\widehat{\rho}$	R^2	BP Seq F-test	Max LR F-test	Exp LR F-test
Dividend payout ratio	-0.0001	0.7476	0.5592	26.5211***	26.5211***	7.7360***
Long-term yield	-0.0002	0.0368	0.0014	1.7122	1.1633	0.1840
Dividend yield	-0.0007	0.0408	0.0017	3.4744	5.0447	1.1532
Dividend-price-ratio	-0.0007	0.0495	0.0024	3.3367	5.0907	1.1876
T-bill rate	-0.0001	0.4684	0.2197	1.3134	2.6698	0.4553
Earnings-price-ratio	-0.0003	0.4384	0.1922	0.7989	7.5124***	2.2289
Book-to-market ratio	-0.0006	-0.0447	0.0020	2.7760	2.0446	0.2170
Default yield spread	0.0000	0.4616	0.2131	1.6369	3.7191	0.7621
Net equity expansion	0.0000	0.2440	0.0596	0.8209	1.6173	0.1941
Term spread	0.0000	0.1034	0.0107	1.3025	1.7316	0.2593
Inflation rate	0.0000	-0.0979	0.0097	1.6209	2.5489	0.4311

Table A4 presents structural break tests and model estimates based on an AR(1) model applied to all individual stationary predictors (based on first differences[2], i.e., $\Delta x = x_t - x_{t-1}$) for Panel A with sampling period: 1990Q1 - 2019Q4. The covariance matrix for the AR(1) model is constructed using the HAC (Newey-West) estimator. We consider the following structural break tests[3] (i) Bai-Perron F-test based on the sequential break detection algorithm of Bai and Perron (2003) with $\epsilon = 0.15$; (ii) Maximum LR F-statistic with $\epsilon = 0.15$; and (iii) Exp LR F-statistic with $\epsilon = 0.15$. The last two structural break tests represent the Andrews unknown breakpoint tests (see, Andrews (1993)). Notice also for the BP statistic we assume a common data distribution across the blocks to ensure consistent estimation of the variance. For all test statistics, we denote the rejection probabilities under the null hypothesis of no structural break at significance levels 1%(***), 5%(**), and 10%(*), respectively.

[2]Notice that considering the first differences of the predictors ensures that the structural break tests are asymptotically valid, regardless of whether these have unit roots.

[3]Further details regarding the specification and testing algorithms can be found in the User's Guide of Eviews under the section titled "Stability Diagnostics", see `https://www.eviews.com/help/`

Appendix E

Coding Procedures

E.1 Matlab Code

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Procedure 1: Generating the DGP under the null hypothesis
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [ysim, xsim] = simulate_null( n , beta1, beta2, c1, c2, gamma, rho );

outputVector = [];
x1 = zeros(n,1);
x2 = zeros(n,1);

sigma_v1v2 = (rho)*sqrt(1)*sqrt(1);
mu         = [0 0 0];
Sigma      = [1 0.10 -0.29 ; 0.10 1 sigma_v1v2 ; -0.29 sigma_v1v2 1];
innov_e    = mvnrnd(mu,Sigma,n);
u          = innov_e(:,1);
v1         = innov_e(:,2);
v2         = innov_e(:,3);

for i = 2:n;

  x1(i,1) = ( 1 - c1 / ( n^gamma) )*x1(i-1,1) + v1(i,1);
  x2(i,1) = ( 1 - c2 / ( n^gamma) )*x2(i-1,1) + v2(i,1);
  y(i,1)  = beta1*x1(i-1,1) + beta2*x2(i-1,1) + u(i,1);

end

ysim = y;
xsim = [x1(:,1) x2(:,1)];
[ysim , xsim ];

end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Procedure 2: Wald IVX test for structrural break testing under the null hypothesis
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [sup_Wald_IVX] = estimate_sup_Wald_IVX( ysim, xsim );

outputVector = [];
pi0 = 0.15;
p = 2;
n = length(ysim);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Obtain the simulated data
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xlag = xsim(1:n-1,:);
xt   = xsim(2:n,:);
y    = ysim(2:n,:);
nn   = length(y);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Fit the LUR specification
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
rn = zeros(p,p);
for i = 1:p
   rn(i,i) = regress( xt(:,i),xlag(:,i) );
end

% autoregressive residual estimation
u = xt - xlag*rn;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Construct the IVX instrument
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
K = 1;
n  = nn-K+1;
Rz = ( 1 -1/(nn^0.95) )*eye(p);
diffx = xt-xlag;
z = zeros(nn,p);
z(1,:) = diffx(1,:);

for i=2:nn
  z(i,:) = z(i-1,:)*Rz+diffx(i,:);
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Loop for the supremum functional
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
t = length( y );
lower = t*pi0;
upper = t*(1-pi0);
```

```
lower_bound = round(lower);
upper_bound = round(upper);
sequence    = ( lower_bound : upper_bound );
dim_seq     = ( upper_bound - lower_bound );

% Wald IVX vector that stores the sequence of statistics
Wald_IVX_vector = zeros(dim_seq ,1);

s = 1;
while ( s <= dim_seq  );

    k = sequence( s );

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % Step 1: Obtain the estimate of the variance of OLS regression
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    xlag1 = xlag;
    xlag1( (k+1):nn, : ) = 0;

    xlag2 = xlag;
    xlag2(  1:k, : ) = 0;

    X_matrix = [ xlag1 xlag2 ];

    [Aols,bhat,epshat] = regress( y, X_matrix );
    covepshat = ( epshat'*epshat ) / nn;

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % Step 2: Obtain the estimates of the IVX estimators
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % We define with Z the IVX instrument
    Z = [ zeros(1,p) ; z(1:n-1,:) ];

    % Estimate the Z1 (before break) and Z2 (after break) matrices
    Z1 = Z;
    Z1( (k+1):nn, : ) = 0;

    Z2 = Z;
    Z2( 1:k, : ) = 0;

    beta1_ivx = y'*Z1*pinv(xlag1'*Z1);
    beta2_ivx = y'*Z2*pinv(xlag2'*Z2);
    beta_ivx_distance = ( beta1_ivx - beta2_ivx );

    % covariance matrix estimation (predictive regression)
    covu = zeros(p,p);
    for t=1:nn
       covu=covu+u(t,:)'*u(t,:);
    end
```

```
% covariance matrix estimation (autoregression)
covu    = covu/nn;
covuhat = zeros(1,p);

for i=1:p
   covuhat(1,i) = sum( epshat'*u(:,i));
end

% covariance matrix between 'epshat' and 'u'
covuhat = covuhat'/nn;
m  = floor(nn^(1/3));
uu = zeros(p,p);
for h = 1:m
   a = zeros(p,p);
   for t = (h+1):nn
      a = a+u(t,:)'*u(t-h,:);
   end
uu = uu+(1-h/(m+1))*a;
end
uu = uu/nn;
Omegauu = covu+uu+uu';

q = zeros(m,p);
for h = 1:m
   pa = zeros(nn-h,p);
   for t = (h+1):nn
      pa(t-h,:) = u(t,:)*epshat(t-h)';
   end
q(h,:) = (1-h/(1+m))*sum(pa);
end

residue = sum(q)/nn;
Omegaeu = covuhat+residue';

% FM is estimated based on the regression model with xlag1 and xlag2 as predictors
FM = covepshat - Omegaeu'*Omegauu^(-1)*Omegaeu;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Estimation of Q1 and Q2 matrices
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Z1_mean   = mean( Z1( 1:k, : ) );
M1_matrix = Z1'*Z1*covepshat - k*Z1_mean'*Z1_mean*FM;
Q1_matrix = pinv(Z1'*xlag1)*( M1_matrix )*pinv(xlag1'*Z1);

Z2_mean   = mean( Z2( (k+1):nn, : ) );
M2_matrix = Z2'*Z2*covepshat - ( n - k )*Z2_mean'*Z2_mean*FM;
Q2_matrix = pinv(Z2'*xlag2)*( M2_matrix )*pinv(xlag2'*Z2);

Q_matrix       = ( Q1_matrix + Q2_matrix );
```

```
    Wald_IVX_break = beta_ivx_distance*pinv( Q_matrix )*(beta_ivx_distance');

    Wald_IVX_vector( s, 1 ) = Wald_IVX_break;

% end-of estimation of sequence of Wald-IVX statistics
s = s + 1;
end

% Obtain the sup-Wald statistic
sup_Wald_IVX = max( Wald_IVX_vector );

end
%END-OF-FUNCTION


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Procedure 3: Monte Carlo Simulations under the null hypothesis
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [empirical_size, sup_Wald_IVX_matrix, sup_Wald_IVX_matrix_bootstrap]
                       = MC_simulations_size( M, B, n, beta1, beta2, c1, c2, gamma, rho );

outputVector = [];

%  Step 1: Simulate a DGP using Function 1
sup_Wald_IVX_matrix           = zeros(M,1);
sup_Wald_IVX_matrix_bootstrap = zeros(M,1);
empirical_size = 0;
count = 0;
ysim  = 0;
xsim  = 0;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MC simulation step
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i = 1:M
%begin-of-MC-loop

   [ ysim , xsim ]  = simulate_null( n , beta1, beta2, c1, c2, gamma, rho );
   xlag = xsim(1:n-1,:);
   xt   = xsim(2:n,:);
   y    = ysim(2:n,:);
   nn   = length(y);

   Wald_IVX = estimate_sup_Wald_IVX( ysim, xsim );
   sup_Wald_IVX_matrix( i, 1 ) = Wald_IVX;

   %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
   % Bootstrap step (Fixed regressor bootstrap)
   %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
   yt_star = zeros( nn,B );

   for b = 1:B
      [Aols,bhat,epshat] = regress( y, xlag );
      kappa      = normrnd(0,1,[nn,1]);
      u_hat_star = ( epshat.*kappa);
      yt_star( : , b ) = Aols(1,1)*xlag( :,1) + Aols(2,1)*xlag( :,2) + u_hat_star;
   end

   bootstrap_Wald_IVX_matrix = zeros( B , 1 );

   R = 40;
   parfor (j = 1:B, R)
       ysim = yt_star( : , j );
       estimation_sup_Wald_IVX_step     = estimate_sup_Wald_IVX( ysim, xsim );
       bootstrap_Wald_IVX_matrix( j, 1) = estimation_sup_Wald_IVX_step;
   end

   the_boot_statistic = sort( bootstrap_Wald_IVX_matrix );
   quantile           = round(0.95*B);
   boot_IVX_statistic = the_boot_statistic( quantile,  1 );

   %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
   sup_Wald_IVX_matrix_bootstrap( i, 1 ) = boot_IVX_statistic;
   ysim = 0;
   xsim = 0;

%end-of-MC-loop
end

   % Obtain the empirical size
   count = 0;
   for j = 1:M

       if ( sup_Wald_IVX_matrix( j, 1 ) <  sup_Wald_IVX_matrix_bootstrap( j, 1 ) );
           count = count + 1;
        end
    end

    empirical_size =  count / M;

%%%%%%%%%%%%%%%%%%%%%%%%%
%END-OF-FUNCTION
end
```

E.2 R Code

```
################################################################################
### Function 1: Simulate the i.i.d covariance matrices from Wishart distribution
################################################################################
simulate_Wishart_matrices <- function( n_Nodes = n_Nodes, p = p )
{
  n_Nodes = n_Nodes; p = p
  S        <- toeplitz((p:1)/p)
  sim_data <- rWishart( n_Nodes, (p+2), S )
  mylist  <- 0; mylist  <- list(); sigma   <- 0
  for ( j in 1:n_Nodes )
  {
    sim_data <- rWishart( n_Nodes, (p+2), S )
    Sigma    <- sim_data[ , , j ]
    mylist[[j]] <- Sigma
    Sigma <- 0
  }
  return( mylist )
}
################################################################################
### Function 2: Simulate data pair under the null hypothesis
################################################################################
simulate_data_null_function <- function( N = N_size, gamma.x = gamma.x, sim_Wishart = sim_Wishart )
{# begin of function
  N  <- N_size; gamma.x <- gamma.x; sim_Wishart <- sim_Wishart
  # We simulate 7 different regressors for each model
  p   <- 7
  mu.vector <- matrix(0, nrow = p+1, ncol = 1 )
  Sigma     <- as.matrix( simulate_Wishart )
  # generate random error sequence from Multivariate Normal Distribution
  innov.e  <- rmvnorm( n = N, mean = mu.vector, sigma = Sigma )
  innov.u  <- as.matrix( innov.e[ ,1] )
  innov.v  <- as.matrix( innov.e[ ,2:8] )
  # generates beta from random Uniform distribution in the interval [-1,1]
  beta1 <- as.numeric( runif(1, min = -1, max = 1) )
  beta2 <- as.numeric( runif(1, min = -1, max = 1) )
  beta3 <- as.numeric( runif(1, min = -1, max = 1) )
  beta4 <- as.numeric( runif(1, min = -1, max = 1) )
  beta5 <- as.numeric( runif(1, min = -1, max = 1) )
  beta6 <- as.numeric( runif(1, min = -1, max = 1) )
  beta7 <- as.numeric( runif(1, min = -1, max = 1) )
  C      <- diag(p)
  C[1,1] <- 1; C[2,2] <- 2; C[3,3] <- 3; C[4,4] <- 4; C[5,5] <- 5; C[6,6] <- 6; C[7,7] <- 7
  Rn <- diag(p) - C/( N^gamma.x )
  x  <- matrix( 0,N,p)
  for (j in 1:p){
    for(t in 2:N){
      x[t,j] <- Rn[j,j]*x[t-1,j] + innov.v[t,j]
```

```
    }
  }
  y <- matrix(0, nrow = N, ncol = 1)
  for (t in 2:N)  {
    y[t,1] <- beta1*x[t-1,1] + beta2*x[t-1,2] + beta3*x[t-1,3] + beta4*x[t-1,4]
              + beta5*x[t-1,5] + beta6*x[t-1,6] + beta7*x[t-1,7] + innov.u[t,1]
  }
  simulated.data <- structure( list( y = y, x  = x ) )
  return( simulated.data )
}# end of function


################################################################################
### Function 3: IVX-QR Estimation Methodology (see, Lee (JoE, 2016))
################################################################################
IVXQR_estimation_function <- function( y_data = y_data, x_data = x_data, tau = tau )
{# begin of function
  # obtain the simulated data from Function 1 which are in the form of lists
  y_data <- y_data; x_data <- x_data
  y   <- as.matrix( y_data )
  x   <- as.matrix( x_data )
  tau <- tau
  # Lag Adjustment
  n  <- nrow( x ); p  <- ncol( x )
  y.t    <- as.matrix( y[2:n,1] )
  x.t    <- as.matrix( x[2:n,  ] )
  x.lag  <- as.matrix( x[1:(n-1), ] )
  ####################################
  # Autoregression Estimation
  # Rn contains the estimated coefficients of the autoregression matrix Rn
  ####################################
  Rn <- matrix(0, p, p)
  for ( i in 1:p ){
    Rn[i, i] <- lm( x.t[, i] ~ 0 + x.lag[, i] )$coefficients
  }
  OLS.model    <- lm( x.t ~  x.lag - 1)
  matrix.coef  <- OLS.model$coefficients
  coef <- matrix( 0, 7, 7)
  for (i in 1:7){
    coef[i,i] <- matrix.coef[i,i]
  }
  # autoregressive residual estimation: ( n x p ) matrix with estimated residuals
  # Note that these are the individual residuals from each AR(1) separately in each column
  u.hat <- as.matrix( x.t - x.lag %*% Rn )
  ####################################
  # Ordinary QR with intercept
  ####################################
  model.QR <- rq( y.t  ~ x.lag, tau = tau )
  model.QR.summary <- summary( model.QR, se = "boot", bsmethod= "xy" )
  model.QR.coef    <- as.data.frame( model.QR.summary$coefficients )
```

```
  model.QR.coef    <- as.matrix(  model.QR.coef$Value )
  residuals.QR.model <- as.matrix( residuals( model.QR ) )
  #  estimating lambda
  e.tau <- matrix( 0, nrow = nrow(x.lag) , ncol = 1)
  for ( i in 1:(n-1) ){
    if ( residuals.QR.model[i,1] < 0 ){ e.tau[i,1] <- 1}
  }
  # lambda <- cor( e.tau, u.hat )
  lambda <- as.matrix( ( ( t(e.tau) %*% u.hat  ) / (n-1) ) / sqrt( tau*(1-tau) ) )
  c <- -6; cz <- -5
  delta <- ( 1-(log(-c)-log(-cz))/ log(n) )
  # IVX construction
  n_ivx <-  n^(delta)
  z     <-  matrix( 0 , nrow = n, ncol = ncol(x) )
  z[1, ] <- x[1, ]
  rho_z  <- ( 1 + cz / n_ivx )
  for ( t in 2:n ){
    z[t, ] <- rho_z*z[t-1, ] + ( x[t, ] - x[t-1, ] )
  }
  # IVX-QR with dequantiling (see, Lee (JoE, 2016))
  intercept.ones <- matrix( 1, nrow = nrow(y.t) , ncol = 1)
  z.lag.deq <- z[1:(n-1), ]
  beta.tau  <- model.QR.coef[1,1]
  y.t.tau   <- y.t - intercept.ones*beta.tau
  # Next we obtain the IVX-QR estimates after dequantiling
  model.IVX.QR         <- rq( y.t.tau  ~ z.lag.deq - 1, tau = tau )
  model.IVX.QR.summary <- summary( model.IVX.QR, se = "boot", bsmethod= "xy" )
  model.IVX.QR.coef    <- as.data.frame( model.IVX.QR.summary$coefficients )
  IVX.QR.coef          <- as.matrix( model.IVX.QR.coef$Value )
  residuals.QR.model <- as.matrix( residuals( model.IVX.QR ) )

  # Estimate the kernel density function
  kernel.density.residuals.QR.model <- kdensity( residuals.QR.model,
                                 start = "gumbel", kernel = "gaussian", normalized = TRUE)
  kde <- kernel.density.residuals.QR.model
  f_u <- kde(0)
  x.values <- as.matrix( d.residuals$x )
  y.values <- as.matrix(  d.residuals$y  )
  position.first.positive <- min( which( x.values > -0.5 ) )
  # IVX QR t stats for simulation
  ttau             <- tau*(1-tau)
  zzinv            <- inv( t(z.lag.deq)%*%(z.lag.deq) )
  sigma.IVX.QR.coef <- ( ttau/ ( f_u^2 ) )*(zzinv)
  chi.square.test   <- t(IVX.QR.coef)%*%( inv(sigma.IVX.QR.coef) )%*%(IVX.QR.coef)
  return( list( IVX.QR.coef, sigma.IVX.QR.coef, chi.square.test ) )
}# end of function
```

```
################################################################################
### Function 4: Obtain Pairwise Data and IVX-QR Estimation for each equation
################################################################################
model_coef_list <- list(); wald_IVX_test <- list()
for ( j in 1:n_Nodes ){
  y_data <- as.matrix( y_matrix[ ,j] )
  x_data <- as.matrix( x_matrix[[j]] )
  estimation_IVXQR_model <- IVXQR_estimation_function( y_data = y_data, x_data = x_data, tau = tau )
  IVX_model_coefficients <- as.matrix( estimation_IVXQR_model[[1]] )
  model_coef_list[[j]]   <- IVX_model_coefficients
  Wald_IVX_test[[j]]     <- as.matrix( estimation_IVXQR_model[[3]] )
}
# Now in order to ensure that we estimate the QR-SUR estimator we use the de-quantile method
# de-quantiling the errors of the predictive regression model
for (i in 1:NB){
   uy[i] = uy[i] - qnorm( tau, 0, sqrt(sigma2[i]) )
}
################################################################################
### Function 5A: Estimate the VaR-CoVaR matrix (stationary time series models)
################################################################################
Risk_Matrix_forecast <- function( Nr_C = Nr_C, nhist = nhist, returns = returns_historical, macro = m
  # Initialize inputs
  Nr_C  <- Nr_C; nhist <- nhist
  currenttime <- currenttime
  returns     <- as.matrix(returns_historical)
  macro       <- as.matrix(macro_historical)
  #We match the lag of the two series since the macro in the model is M_{t-1}
  nr <- NROW(returns)
  #We pick the last observation from the historical period to get the predicted value of the VaR in t
  current.macro <- macro[nhist-1, ]
  current.macro <- as.matrix(as.vector(current.macro))
  returns.lag <- returns[1:(nr - 1), , drop = FALSE]
  macro.lag   <- macro[1:(nr - 1), , drop = FALSE]
  returns.t   <- returns[2:nr, , drop = FALSE]
  macro.t     <- macro[2:nr, , drop = FALSE]
  returns.t   <- as.matrix( returns.t )
  macro.lag   <- as.matrix( macro.lag )
  return.t <- 0
  other.return.t <-0; CoVaR.est <- 0; var_est <- 0
  forecast_risk_matrix <- matrix( 0, nrow = Nr_C  , ncol = Nr_C )
  beta.estimate        <- matrix( 0, nrow = Nr_C  , ncol = Nr_C )
  sigma.beta.estimate  <- matrix( 0, nrow = Nr_C  , ncol = Nr_C )
  pvalue.beta.estimate <- matrix( 0, nrow = Nr_C  , ncol = Nr_C )
  for (i in 1:Nr_C){#begin of outer loop
    # Each i iteration first estimates the CoVaR_i|i = VaR_i
    var_est <- 0; var_forecast <- 0; return.t <- returns.t[ , i]; return.t <- as.matrix( return.t )
    model.var <- rq( return.t  ~ macro.lag, tau = 0.05 )
    model.var.summary <- summary( model.var, se = "boot", bsmethod= "xy" )
    model.coef <- as.data.frame( model.var.summary$coefficients )
    coef.macro <-  model.coef$Value
```

```
    coef.macro <- as.matrix( coef.macro )
    # Extract the coefficients of the CoVaR_1|1 model
    coef_const_var  <- coef.macro[1,1]
    var_forecast    <- coef_const_var + t(current.macro) %*% coef.macro[2:8,1 ]
    forecast_risk_matrix[i,i] <- var_forecast
    #get the fitted values of the model
    var_est <- fitted.values(model.var)
    var_est <- as.vector(var_est)
    var_est <- as.matrix(var_est)
    ##################################
    for (j in 1:Nr_C){#begin of inner loop
      CoVaR.est <- 0
      if (j!=i)
      {# condition to estimate
        other.return.t <- returns.t[ , j]
        other.return.t <- as.matrix( other.return.t )
        ###Estimation of CoVaR_j|i
        model.covar <- rq( other.return.t ~ macro.lag + return.t , tau = 0.05 )
        model.covar.summary <- summary( model.covar, se = "boot", bsmethod= "xy" )
        model.estimates <- as.data.frame( model.covar.summary$coefficients )
        coef.covar  <-  model.estimates$Value
        coef.covar  <-  as.matrix( coef.covar )
        coef.st.error <- model.estimates$‘Std. Error‘
        coef.st.error <- as.matrix( coef.st.error )
        coef.pvalue   <- model.estimates$‘Pr(>|t|)‘
        coef.pvalue   <- as.matrix( coef.pvalue )
        # We need 3 matrices to collect the coefficients of the returns
        # Matrix 1: coefficients
        # Matrix 2: standard errors
        # Matrix 3: p-values
        # Define the 3 matrices to keep the above
        beta.estimate[j,i] <- coef.covar[9,1]
        sigma.beta.estimate[j,i]  <- coef.st.error[9,1]
        pvalue.beta.estimate[j,i] <- coef.pvalue[9,1]
        ### Extract the coefficients of the CoVaR_12 model
        coef_const_covar  <- coef.covar[1,1]
        coef_const_covar  <- as.numeric(coef_const_covar)
        covar_forecast <- coef_const_covar + t(current.macro) %*% coef.covar[2:8,1] + coef.covar[9,1
        forecast_risk_matrix[j,i] <- covar_forecast
        covar_forecast<-0
      }#end of estimate condition
    }#end of inner loop
  }#end of outer loop
  return(forecast_risk_matrix)
}#end of function
```

```
############################################################################
# FUNCTION 5B: Constuction of symmetrized time-varying CoVaR Matrices #####
############################################################################
covar_symmetric_function <- function(sigma_tilda = sigma_tilda )
{#begin of function
  sigma_tilda <- sigma_tilda
  sigma_tilda <- abs( sigma_tilda )
  sigma_tilda_sym <- 0.5*( sigma_tilda + t(sigma_tilda) )
  return( sigma_tilda_sym )
}#end of function

################################################################################
# FUNCTION 6: Bootstrapping the predictive regression model ##########
################################################################################
bootstrap_data_function <- function( Nr_C = Nr_C, time = time, returns=returns, macro=macro  )
{# begin of function
  Nr_C <- Nr_C; time <- time
  returns <- as.matrix(returns); macro <- as.matrix(macro)
  ### Step 1: Run OLS predictive regressions to obtain the model coefficients
  nr <- NROW(returns); p  <- ncol(macro)
  returns.t <- as.matrix( returns[2:nr, ] )
  macro.t   <- as.matrix( macro[2:nr, ] )
  macro.lag <- as.matrix( macro[1:(nr-1), ] )
  ##############################################
  ### Step 1A: Fit a VAR(1) model for the autoregressive model
  ##############################################
  coefficients.xt.model.matrix <- matrix( 0, nrow = ncol(macro) , ncol = ncol(macro) )
  intercepts.xt.model.matrix   <- matrix( 0, nrow = ncol(macro) , ncol = 1)
  VAR_model <- VAR( y = macro, p = 1 )
  B.estimates <- as.matrix( Bcoef(VAR_model) )
  B.matrix    <- as.matrix( B.estimates[ ,1:p ] )
  mu.x.vector <- as.matrix( B.estimates[ ,(p+1)]  )
  eigenvalues.B.matrix <- as.matrix( roots(VAR_model) )
  ux.t.hat             <- as.matrix( residuals( VAR_model ) )
  ux.t.hat.centered    <- matrix(0, nrow = nrow(ux.t.hat), ncol = ncol(ux.t.hat) )
  for (j in 1: ncol(ux.t.hat) ){
    ux.t.hat.centered[ ,j] <-  ( as.matrix( ux.t.hat[ ,j] ) - mean( as.matrix(ux.t.hat[ ,j])) )
  }
  ### Next I construct the bootstrap residuals
  ux.t.hat.star <- matrix(0, nrow =nrow(ux.t.hat), ncol = ncol(ux.t.hat) )
  data <- 0; x <- 0
  for (j in 1: ncol(ux.t.hat) ) {
    x     <- as.matrix( ux.t.hat.centered[ ,j] )
    nboot <- 10000
    data  <- matrix(sample(x, size = length(x) * nboot, replace = T), nrow = nboot)
    for (i in 1:nrow(ux.t.hat) ) {
      ux.t.hat.star[i,j] <- mean( data[ ,i] )
    }
    data <- 0; x <- 0
  }
```

```
###############################################
### Bias Correction for the coefficient matrix
### (Reference: Testing for Multiple Horizon Predictability)
####################### #######################
Omega.x.hat <- as.matrix( cov( ux.t.hat ) )
unit.matrix <- diag( ncol(ux.t.hat) )
term1 <- inv( unit.matrix - t(B.matrix)  )
term2 <- t(B.matrix)%*%( inv( unit.matrix - t(B.matrix)%*%B.matrix  ) )
term3 <- 0
for (j in 1:ncol(ux.t.hat) ){
  term3 <- term3
  + eigenvalues.B.matrix[j,1]*( inv( unit.matrix - eigenvalues.B.matrix[j,1]*t(B.matrix) ) )
}
term3 <- as.matrix(term3)
macro.lag.tilde <- matrix( 0, nrow = nrow(macro.lag), ncol= ncol(macro.lag) )
for (j in 1:ncol(macro.lag) ){
  macro.lag.tilde[ ,j] <- ( as.matrix( macro.lag[ ,j] ) -  mean( macro.lag[ ,j] ) )
}
term4 <- inv( t( macro.lag.tilde )%*%macro.lag.tilde )
B.matrix.hat.bc <- as.matrix( B.matrix + Omega.x.hat%*%( term1 + term2 + term3  )%*%term4 )
###############################################
### Step 1B: Fit the predictive regression model
###############################################
model.yt   <- lm( returns.t ~ macro.lag )
model.coef <- as.matrix( model.yt$coefficients )
beta.coefficients <- as.matrix( model.coef[2:8, ] )
mu.y.vector       <- as.matrix( model.coef[1, ] )
uy.t.hat          <- as.matrix( residuals(model.yt) )
uy.t.hat.centered <- matrix(0, nrow = nrow(uy.t.hat), ncol = ncol(uy.t.hat) )
for (j in 1: ncol(uy.t.hat) ){
  uy.t.hat.centered[ ,j] <-  ( as.matrix( uy.t.hat[ ,j] ) - mean( as.matrix(uy.t.hat[ ,j])) )
}
### Next we construct the bootstrap residuals
uy.t.hat.star <- matrix(0, nrow =nrow(uy.t.hat), ncol = ncol(uy.t.hat) )
data <- 0; x    <- 0
for (j in 1: ncol(uy.t.hat) ){
  x     <- as.matrix( uy.t.hat.centered[ ,j] )
  nboot <- 10000
  data  <- matrix(sample(x, size = length(x) * nboot, replace = T), nrow = nboot)

  for (i in 1:nrow(uy.t.hat) ){
    uy.t.hat.star[i,j] <- mean( data[ ,i] )
  }
  data <- 0
  x <- 0
}
```

```
  ##############################################
  ### Step 2: Generate the bootstrap sample {yt*, xt*, t =1,...,n =313}
  ##############################################
  #### Generate the xt_star series ####
  x.t.star      <- matrix( 0, nrow = nrow(macro.lag), ncol = ncol(macro.lag)  )
  x.t.star[1, ] <- ux.t.hat.star[1,]
  ones.matrix   <- matrix( 1, nrow = nrow(macro.lag), ncol = ncol(macro.lag) )
  intrecepts.x  <- matrix( 0, nrow = nrow(macro.lag), ncol = ncol(macro.lag) )
  # Define the matrix of intercepts
  for (j in 1:ncol(macro.lag) ){
    intrecepts.x[, j]  <- as.matrix( ones.matrix[ ,j] )*as.numeric( mu.x.vector[j,1] )
  }
  # Define the x.t.star series
  for ( j in 1: ncol(macro.lag)) {
    for ( t in 2: nrow(macro.lag) ){
      x.t.star[t,j] <- intrecepts.x[t,j]
       +  ( B.matrix.hat.bc[ ,j] ) %*% ( as.matrix(x.t.star[t-1, ]) )
       +  as.matrix( ux.t.hat.star[t,j] )
    }
  }
  #### Generate the yt_star series ####
  y.t.star      <- matrix(0, nrow = (nrow(returns.t)), ncol = ncol(returns.t) )
  y.t.star[1, ] <- uy.t.hat.star[1,]
  ones.matrix   <- matrix( 1, nrow = nrow(returns.t), ncol = ncol(returns.t) )
  intrecepts.y  <- matrix( 0, nrow = nrow(returns.t), ncol = ncol(returns.t) )
  # Define the matrix of intercepts
  for (j in 1:ncol(returns.t) ){
    intrecepts.y[, j]  <- as.matrix( ones.matrix[ ,j] )*as.numeric( mu.y.vector[j,1] )
  }
  for ( j in 1: ncol(returns.t)  ){
    for ( t in 2: nrow(returns.t) ){
      y.t.star[t,j] <- intrecepts.y[t,j]
      + ( beta.coefficients[ ,j] ) %*% ( as.matrix(x.t.star[t-1, ]) )
      +  as.matrix( uy.t.hat.star[t,j] )
    }
  }
  return( list( y.t.star = y.t.star, x.t.star = x.t.star) )
}# end of function
################################################################################
```

```
################################################################################
# FUNCTION 7: Subset Testing for Quantile Regressions ##########
################################################################################

p < - nrow(X); nc <- ncol(X); m <- nc; ns <- n; nr <- 1000; h <- 1
for (ii in 1:nr)
{
  Xs[, 1, ii] <- 0
  Xs1[, 1, ii] <- colMeans(XX[, 2:ncol(XX)])
  Us <- t(mvrnorm(ns, mu = rep(0, kk+1), Sigma = Cov))
  Ys[, 1, ii] <- Beta1[1] + Us[, 1]
  Us <- t(Us)
  for (jj in 1:ns) {
    Xs[, jj+1, ii] <- Beta %*% Xs[, jj, ii] + Us[2:(kk+1), jj]
    Xs1[, jj+1, ii] <- Xs[, jj+1, ii] + mean_X[1:kk, 1]
  }
  Ys[, 1, ii] <- Ys[, 1, ii] + coef * Xs1[1, 1:(ns-1), ii]
}
Xs1 <- Xs1[, 2:ns, ]
Xs1 <- aperm(Xs1, c(2, 1, 3))
dataset$Y <- Ys
dataset$X <- Xs1

rej_KMS <- matrix(0, nrow = length(kk), ncol = length(hh))
for (kk in 1:length(kk)) {
  for (hh in 1:length(hh)) {
    result <- compute_size_0608(dataset, nr, h, epsilon)
    rej_KMS[kk, hh] <- result[[1]]
  }
}

Wald_test <- function(Y, X)
{
  n <- nrow(X)
  m_x <- ncol(X)
  X <- cbind(1, X) # add a constant
  b <- solve(t(X) %*% X) %*% t(X) %*% Y
  sig2 <- sum((Y - X %*% b)^2) / (n - 1 - (m_x + 1))
  V <- solve(t(X) %*% X) * sig2
  W <- t(b[-1]) %*% solve(V[-1, -1]) %*% b[-1]
  pv <- 1 - pchisq(W, m_x)
  return(list(pv = pv, W = W))
}

IV_max_J_IVX_statistic <- function(R, X, h, epsilon)
{
  T <- nrow(X); m <- ncol(X); mf <- 0; B <- 1000;
  YY <- y[2:T]
  XX <- X
  YY <- R[2:T]
```

```
  XX <- X[1:(T-1), ]
  nn <- length(YY)
  YY <- YY - mean(YY)
  XX <- XX - matrix(rep(mean(XX), nn), nrow = nn, ncol = m, byrow = TRUE)
  Yvec <- rep(YY, each = m)
  Xmat <- matrix(0, nrow = nn * m, ncol = m * (1 + mf))
  for (i in 1:nn) {
    for (j in 1:m) {
      Xmat[(i - 1) * m + j, 1 + (j - 1) * (1 + mf)] <- XX[i, j]
    }
  }
  theta <- solve(t(Xmat) %*% Xmat) %*% t(Xmat) %*% Yvec
  xlag <- X[1:(T-1), ]; xt   <- X[2:T, ]
  n <- nn - h + 1; Rz <- (1 - 1 / (nn ^ 0.95)) * diag(m + mf)
  diffx <- xt - xlag
  z <- matrix(0, nrow = nn, ncol = m + mf); z[1, ] <- diffx[1, ]
  for (i in 2:nn) {
    z[i, ] <- z[i - 1, ] %*% Rz + diffx[i, ]
  }
  Z <- rbind(matrix(0, nrow = 1, ncol = m + mf), z[1:(n - 1), ])
  zz <- rbind(matrix(0, nrow = 1, ncol = m + mf), z[1:(nn - 1), ])
  ZK <- matrix(0, nrow = n, ncol = m + mf)
  for (i in 1:n) {
    ZK[i, ] <- colSums(zz[i:(i + h - 1), ])
  }
  Zmat <- matrix(0, nrow = nn * m, ncol = m * (1 + mf))
  for (i in 1:nn) {
    for (j in 1:m) {
      Zmat[(i - 1) * m + j, 1 + (j - 1) * (1 + mf)] <- Z[i, j]
    }
  }
  theta_IV <- solve(t(Zmat) %*% Xmat) %*% t(Zmat) %*% Yvec
  u <- matrix(0, nrow = nn, ncol = m + mf)
for (i in 1:(m+mf)) {
  rn <- lm(xt[,i] ~ xlag[,i])
  u[,i] <- xt[,i] - xlag[,i] %*% rn$coefficients[-1]
}
covu <- matrix(0, nrow = m+mf, ncol = m+mf)
for (t in 1:nn) {
  covu <- covu + u[t,] %*% t(u[t,])
}
covu <- covu/nn; mm <- floor(nn^(1/3))
uu <- matrix(0, nrow = m+mf, ncol = m+mf)
for (h in 1:mm) {
  a <- matrix(0, nrow = m+mf, ncol = m+mf)
  for (t in (h+1):nn) {
    a <- a + u[t,] %*% t(u[t-h,])
  }
  uu <- uu + (1-h/(mm+1)) * a
}
```

```
uu <- uu/nn; Omegauu <- covu + uu + t(uu)

UU <- matrix(0, nrow = m, ncol = m)
ZX <- matrix(0, nrow = m*(1+mf), ncol = m*(1+mf))
XZ <- matrix(0, nrow = m*(1+mf), ncol = m*(1+mf))
ZZ <- matrix(0, nrow = m*(1+mf), ncol = m*(1+mf))
ZS <- matrix(0, nrow = m, ncol = m*(1+mf))
XX <- matrix(0, nrow = m*(1+mf), ncol = m*(1+mf))
XS <- matrix(0, nrow = m*(1+mf), ncol = m*(1+mf))

for (ii in 1:nn) {
  UU <- UU + W0_mat[,ii] %*% t(W0_mat[,ii])
}
UU <- UU/nn; covepshat <- UU
covuhat <- matrix(0, nrow = m, ncol = m+mf)
for (i in 1:(m+mf)) {
  for (j in 1:m) {
    covuhat[j,i] <- sum(W0_mat[j,] * u[,i])
  }
}
covuhat <- t(covuhat)/nn; m <- floor(nn^(1/3)); uu <- matrix(0, l, l)
for (h in 1:m) {
  a <- matrix(0, l, l)
  for (t in (h+1):nn) {
    a <- a + t(t-h) %*% t(t-h)
  }
  uu <- uu + (1 - h/(m+1)) * a
}
uu <- uu/nn
Omegauu <- covu + uu + t(uu)
q <- matrix(0, m, l)
for (h in 1:m) {
  p <- matrix(0, nn-h, l)
  for (t in (h+1):nn) {
    p(t-h, ) <- u(t, ) %*% epshat(t-h)
  }
  q(h, ) <- (1 - h/(1+m)) * colSums(p)
}
residue <- colSums(q)/nn
Omegaeu <- covuhat + t(residue)
n <- nn - K + 1
Rz <- (1 - 1/(nn^0.95)) * diag(l)
diffx <- xt - xlag
z <- matrix(0, nn, l)
z[1, ] <- diffx[1, ]
for (i in 2:nn) {
  z[i, ] <- z[i-1, ] %*% Rz + diffx[i, ]
}
Z <- rbind(matrix(0, 1, l), z[1:(n-1), ])
zz <- rbind(matrix(0, 1, l), z[1:(nn-1), ])
```

```
ZK <- matrix(0, n, l)
for (i in 1:n) {
  ZK[i, ] <- colSums(zz[i:(i+K-1), ])
}
yy <- matrix(0, n, 1)
for (i in 1:n) {
  yy[i] <- sum(y[i:(i+K-1)])
}
xK <- matrix(0, n, l)
for (i in 1:n) {
  xK[i, ] <- colSums(xlag[i:(i+K-1), ])
}
meanxK <- colMeans(xK)
Yt <- yy - mean(yy)
Xt <- matrix(0, n, l)
for (i in 1:l) {
  Xt[, i] <- xK[, i] - meanxK[i] * rep(1, n)
}
Aivx <- t(Yt) %*% Z %*% solve(t(Xt) %*% Z)
meanzK <- colMeans(ZK)
FM <- covepshat - t(Omegaeu) %*% solve(Omegauu) %*% Omegaeu
M <- t(ZK) %*% ZK %*% covepshat - n * t(meanzK) %*% meanzK %*% FM
H <- diag(l)
Q <- H %*% solve(t(Z) %*% Xt) %*% M %*% solve(t(Xt) %*% Z) %*% H
Wivx <- matrix(0, 2, 1)
Wivx[1, 1] <- t(H %*% Aivx) %*% solve(Q) %*% (H %*% Aivx)
Wivx[2, 1] <- 1 - pchisq(Wivx[1, 1], l)
WivxInd <- matrix(0, 2, l)
WivxInd[1, ] <- (Aivx/((diag(Q))^(1/2)))^2
WivxInd[2, ] <- 1 - pchisq(WivxInd[1, ], 1)
}


################################################################################
# FUNCTION 8: Testing for Slope Homogeneity ##########
################################################################################

y <- data[, 1]
x <- data[, 2:ncol(data)]
p <- ncol(x)
for (i in 1:N) {
  X[[i]] <- x[((1-T)+(T*i)):(T*i), ]
  Y[[i]] <- y[((1-T)+(T*i)):(T*i)]
}
I <- diag(n); tau <- rep(1, n)
M <- I - (tau %*% t(tau) / T)
betahat_ols <- vector("list", n)
for (i in 1:N) {
  betahat_ols[[i]] <- solve(t(X[[i]]) %*% M %*% X[[i]], t(X[[i]]) %*% M %*% Y[[i]])
}
```

```
sigsq_hat <- rep(0, n)
for (i in 1:n) {
  sigsq_hat[i] <- t(Y[[i]] - X[[i]] %*% betahat_ols[[i]]) %*% M %*% (Y[[i]] - X[[i]] %*% betahat_ols
}
L1 <- vector("list", n); L2 <- vector("list", n)
for (i in 1:n) {
  L1[[i]] <- t(X[[i]]) %*% M %*% X[[i]] / sigsq_hat[i]
  L2[[i]] <- t(X[[i]]) %*% M %*% Y[[i]] / sigsq_hat[i]
}
sum1 <- matrix(0, nrow = p, ncol = p)
sum2 <- matrix(0, nrow = p, ncol = 1)
for (i in 1:n) {
  sum1 <- sum1 + L1[[i]]
  sum2 <- sum2 + L2[[i]]
}
betatilde_wfe <- solve(sum1) %*% sum2
swamytilde <- 0
for (i in 1:n) {
  swamytilde <- swamytilde + t(betahat_ols[[i]] - betatilde_wfe) %*% ((X[[i]] %*% M %*% X[[i]]) / si
}
delta <- (N^(1/2)) * (((N^-1) * swamytilde - p) / ((2 * p)^(1/2)))
EZ <- p
VZ <- (2 * p) * (n - p - 1) / (n + 1)
deltaadj <- (N^(1/2)) * (((N^-1) * swamytilde - EZ) / (VZ^(1/2)))
if (delta > 0) {
  pvalue_delta <- 2 * (1 - pnorm(delta, 0, 1))
} else {
  pvalue_delta <- 2 * pnorm(delta, 0, 1)
}
UCV_delta <- qnorm((1 - (alpha/2)), 0, 1)
LCV_delta <- qnorm((alpha/2), 0, 1)
if (deltaadj > 0) {
  pvalue_deltaadj <- 2 * (1 - pnorm(deltaadj, 0, 1))
} else {
  pvalue_deltaadj <- 2 * pnorm(deltaadj, 0, 1)
}
UCV_deltaadj <- qnorm((1 - (alpha/2)), 0, 1)
LCV_deltaadj <- qnorm((alpha/2), 0, 1)
```

Bibliography

Abadir, K. M. (1993). The limiting distribution of the autocorrelation coefficient under a unit root. *The Annals of Statistics*, pages 1058–1070.

Adrian, T. and Brunnermeier, M. K. (2016). Covar. *The American Economic Review*, 106(7):1705.

Amihud, Y. and Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, pages 813–841.

Andersen, T. G. and Varneskov, R. T. (2021). Consistent inference for predictive regressions in persistent economic systems. *Journal of Econometrics*, 224(1):215–244.

Andersen, T. G. and Varneskov, R. T. (2022). Testing for parameter instability and structural change in persistent predictive regressions. *Journal of Econometrics*, 231(2):361–386.

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics*, pages 676–687.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, pages 817–858.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856.

Andrews, D. W. and Cheng, X. (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80(5):2153–2211.

Andrews, D. W. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414.

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the us wage structure. *Econometrica*, 74(2):539–563.

Atanasov, V., Møller, S. V., and Priestley, R. (2020). Consumption fluctuations and expected returns. *The Journal of Finance*, 75(3):1677–1713.

Aue, A., Cheung, R. C., Lee, T. C., and Zhong, M. (2017). Piecewise quantile autoregressive modeling for nonstationary time series. *Bernoulli*, 23(1):1–22.

Bai, J. (1996). Testing for parameter constancy in linear regressions: an empirical distribution function approach. *Econometrica: Journal of the Econometric Society*, pages 597–622.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1):1–22.

Balestra, P. and Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica: Journal of the econometric society*, pages 585–612.

Banerjee, A., Dolado, J. J., Galbraith, J. W., Hendry, D., et al. (1993). Co-integration, error correction, and the econometric analysis of non-stationary data. *OUP Catalogue*.

Basmann, R. L. (1963). The causal interpretation of non-triangular systems of economic relations. *Econometrica: Journal of the Econometric Society*, pages 439–448.

Belloni, A., Chen, M., Madrid Padilla, O. H., and Wang, Z. (2023). High-dimensional latent panel quantile regression with an application to asset pricing. *The Annals of Statistics*, 51(1):96–121.

Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2019). Modeling systemic risk with markov switching graphical sur models. *Journal of Econometrics*.

Bickel, P. J. (1975). One-step huber estimates in the linear model. *Journal of the American Statistical Association*, 70(350):428–434.

Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons.

Bilodeau, M. and Duchesne, P. (2000). Robust estimation of the sur model. *Canadian Journal of Statistics*, 28(2):277–288.

Boldea, O., Cornea-Madeira, A., and Hall, A. R. (2019). Bootstrapping structural change tests. *Journal of Econometrics*, 213(2):359–397.

Breitung, J. and Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics*, 187(1):358–375.

Brownstone, D. and Valletta, R. (2001). The bootstrap and multiple imputations: harnessing increased computing power for improved statistical tests. *Journal of Economic Perspectives*, 15(4):129–141.

Buchmann, B., Chan, N. H., et al. (2007). Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *The Annals of statistics*, 35(5):2001–2017.

Cai, Z., Chen, H., and Liao, X. (2023). A new robust inference for predictive quantile regression. *Journal of Econometrics*, 234(1):227–250.

Cai, Z. and Liu, X. (2020). A functional coefficient var model for dynamic quantiles with an application for constructing financial network. Technical report, Working Paper, Department of Economics, University of Kansas.

Cai, Z. and Liu, X. (2022). A nonparametric dynamic network via multivariate quantile autoregressions. Technical report, University of Kansas, Department of Economics.

Cai, Z. and Wang, Y. (2014). Testing predictive regression models with nonstationary regressors. *Journal of Econometrics*, 178:4–14.

Cai, Z., Wang, Y., and Wang, Y. (2015). Testing instability in a predictive regression model with nonstationary regressors. *Econometric Theory*, 31(5):953.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of financial economics*, 81(1):27–60.

Canarella, G., Miller, S., and Pollard, S. (2012). Unit roots and structural change: an application to us house price indices. *Urban Studies*, 49(4):757–776.

Caner, M. and Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica*, 69(6):1555–1596.

Cavaliere, G., Georgiev, I., and Robert Taylor, A. (2013). Wild bootstrap of the sample mean in the infinite variance case. *Econometric Reviews*, 32(2):204–219.

Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric theory*, 11(5):1131–1147.

Chan, N. H. and Wei, C.-Z. (1987). Asymptotic inference for nearly nonstationary ar (1) processes. *The Annals of Statistics*, pages 1050–1063.

Chen, C. Y.-H., Härdle, W. K., and Okhrin, Y. (2019). Tail event driven networks of sifis. *Journal of Econometrics*, 208(1):282–298.

Chen, L. (2022). Two-step estimation of quantile panel data models with interactive fixed effects. *Econometric Theory*, pages 1–28.

Chen, L., Galvao, A. F., and Song, S. (2021). Quantile regression with generated regressors. *Econometrics*, 9(2):16.

Chen, Y., Li, J., and Li, Q. (2023). Seemingly unrelated regression estimation for var models with explosive roots. *Oxford Bulletin of Economics and Statistics*.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics*, 33(2):806–839.

Chinco, A. M., Clark-Joseph, A. D., and Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, 74.1(1):449–491.

Cho, J. S., Kim, T.-h., and Shin, Y. (2015). Quantile cointegration in the autoregressive distributed-lag modeling framework. *Journal of econometrics*, 188(1):281–300.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605.

Chu, C.-S. J., Hornik, K., and Kuan, C.-M. (1995). The moving-estimates test for parameter stability. *Econometric Theory*, 11(4):699–720.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica: Journal of the Econometric Society*, pages 1045–1065.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The annals of Statistics*, 25(1):1–37.

Davidson, J. (2000). *Econometric theory*. Wiley-Blackwell.

Davidson, J. and Monticini, A. (2010). Tests for cointegration with structural breaks based on subsamples. *Computational statistics & data analysis*, 54(11):2498–2511.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2):247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1):33–43.

De Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*, volume 21. Springer.

Demetrescu, M. (2014). Enhancing the local power of ivx-based tests in predictive regressions. *Economics Letters*, 124(2):269–273.

Demetrescu, M., Georgiev, I., Rodrigues, P. M., and Taylor, A. R. (2020). Testing for episodic predictability in stock returns. *Journal of Econometrics*.

Demetrescu, M., Georgiev, I., Rodrigues, P. M., and Taylor, A. R. (2022a). Extensions to ivx methods of inference for return predictability. *Journal of Econometrics*.

Demetrescu, M., Georgiev, I., Rodrigues, P. M., and Taylor, A. R. (2022b). Testing for episodic predictability in stock returns. *Journal of Econometrics*, 227(1):85–113.

Demetrescu, M. and Rodrigues, P. M. (2020). Residual-augmented ivx predictive regression. *Journal of Econometrics*.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.

Diebold, F. X. and Chen, C. (1996). Testing structural stability with endogenous breakpoint a size comparison of analytic and bootstrap procedures. *Journal of Econometrics*, 70(1):221–241.

Dolado, J. J. and Lütkepohl, H. (1996). Making wald tests work for cointegrated var systems. *Econometric reviews*, 15(4):369–386.

Dou, L. and Müller, U. K. (2021). Generalized local-to-unity models. *Econometrica*, 89(4):1825–1854.

Elliott, G. (2011). A control function approach for testing the usefulness of trending variables in forecast models and linear regression. *Journal of econometrics*, 164(1):79–91.

Elliott, G., Müller, U. K., and Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, 83(2):771–811.

Elliott, G. and Stock, J. H. (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric theory*, 10(3-4):672–700.

Erdös, P. and Kac, M. (1946). On certain limit theorems of the theory of probability. *Bulletin of the American Mathematical Society*, 52(4):292–302.

Escanciano, J. C. and Goh, S. (2018). Quantile-regression inference with adaptive control of size. *Journal of the American Statistical Association*.

Escanciano, J. C. and Velasco, C. (2010). Specification tests of parametric dynamic conditional quantiles. *Journal of Econometrics*, 159(1):209–221.

Fan, R. and Lee, J. H. (2019). Predictive quantile regressions under persistence and conditional heteroskedasticity. *Journal of Econometrics*, 213(1):261–280.

Farmer, L., Schmidt, L., and Timmermann, A. (2019). Pockets of predictability. *Available at SSRN 3152386*.

Farmer, L. E., Schmidt, L., and Timmermann, A. (2023). Pockets of predictability. *The Journal of Finance*, 78(3):1279–1341.

Feng, J. (2023). Nuclear norm regularized quantile regression with interactive fixed effects. *Econometric Theory*, pages 1–31.

Furno, M. (2014). Quantile regression estimates and the analysis of structural breaks. *Quantitative Finance*, 14(12):2185–2192.

Galvao, A., Montes-Rojas, G., and Olmo, J. (2011). Threshold quantile autoregressive models. *Journal of Time Series Analysis*, 32(3):253–267.

Galvao, A. F., Juhl, T., Montes-Rojas, G., and Olmo, J. (2018a). Testing slope homogeneity in quantile regression panel data with an application to the cross-section of stock returns. *Journal of Financial Econometrics*, 16(2):211–243.

Galvao, A. F., Kato, K., Montes-Rojas, G., and Olmo, J. (2014). Testing linearity against threshold effects: uniform inference in quantile regression. *Annals of the Institute of Statistical Mathematics*, 66(2):413–439.

Galvao, A. F., Montes-Rojas, G., and Olmo, J. (2019). Tests of asset pricing with time-varying factor loads. *Journal of Applied Econometrics*, 34(5):762–778.

Galvao, A. F., Montes-Rojas, G., Olmo, J., and Song, S. (2018b). On solving endogeneity with invalid instruments: an application to investment equations. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(3):689–716.

Georgiev, I., Demetrescu, M., Rodrigues, P. M., and Taylor, A. (2021). Extensions to ivx methods of inference for return predictability.

Georgiev, I., Harvey, D. I., Leybourne, S. J., and Taylor, A. R. (2018). Testing for parameter instability in predictive regression models. *Journal of Econometrics*, 204(1):101–118.

Georgiev, I., Harvey, D. I., Leybourne, S. J., and Taylor, A. R. (2019). A bootstrap stationarity test for predictive regression invalidity. *Journal of Business & Economic Statistics*, 37(3):528–541.

Goh, S. C. and Knight, K. (2009). Nonstandard quantile-regression inference. *Econometric Theory*, 25(5):1415–1432.

Gonzalo, J. and Pitarakis, J.-Y. (2012). Regime-specific predictability in predictive regressions. *Journal of Business & Economic Statistics*, 30(2):229–241.

Gonzalo, J. and Pitarakis, J.-Y. (2017). Inferring the predictability induced by a persistent regressor in a predictive threshold model. *Journal of Business & Economic Statistics*, 35(2):202–217.

Gungor, S. and Luger, R. (2020). Small-sample tests for stock return predictability with possibly non-stationary regressors and garch-type effects. *Journal of Econometrics*.

Guo, H. (2006). On the out-of-sample predictability of stock market returns. *The Journal of Business*, 79(2):645–670.

Gupta, A. and Seo, M. H. (2023). Robust inference on infinite and growing dimensional time-series regression. *Econometrica*, 91(4):1333–1361.

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, pages 489–500.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, pages 413–430.

Hansen, B. E. (2000a). Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603.

Hansen, B. E. (2000b). Testing for structural change in conditional models. *Journal of Econometrics*, 97(1):93–115.

Hanson, B. E. (2002). Tests for parameter instability in regressions with i (1) processes. *Journal of Business & Economic Statistics*, 20(1):45–59.

Härdle, W. K., Wang, W., and Yu, L. (2016). Tenet: Tail-event driven network risk. *Journal of Econometrics*, 192(2):499–513.

Harvey, D. I., Leybourne, S. J., Sollis, R., and Taylor, A. R. (2021a). Real-time detection of regimes of predictability in the us equity premium. *Journal of Applied Econometrics*, 36(1):45–70.

Harvey, D. I., Leybourne, S. J., and Taylor, A. R. (2021b). Simple tests for stock return predictability with good size and power properties. *Journal of Econometrics*.

Hawkins, D. (1987). A test for a change point in a parametric model based on a maximal wald-type statistic. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 368–376.

Henderson, D. J., Kumbhakar, S. C., Li, Q., and Parmeter, C. F. (2015). Smooth coefficient estimation of a seemingly unrelated regression. *Journal of Econometrics*, 189(1):148–162.

Hillier, G. and Satchell, S. (1986). Finite-sample properties of a two-stage single equation estimator in the sur model. *Econometric Theory*, 2(1):66–74.

Hoga, Y. (2017). Change point tests for the tail index of β-mixing random variables. *Econometric Theory*, 33(4):915–954.

Hoga, Y. (2018). A structural break test for extremal dependence in β-mixing random vectors. *Biometrika*, 105(3):627–643.

Hussian, A. and Wallace, T. (1969). The use of error components in combining cross section with time series data. *Econometrica*, 37:52–72.

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, 196(1):55–67.

Jansson, M. and Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, 74(3):681–714.

Kasparis, I., Andreou, E., and Phillips, P. C. B. (2015). Nonparametric predictive regression. *Journal of Econometrics*, 185(2):468–494.

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis*, 100(8):1816–1829.

Katsouris, C. (2021). Optimal portfolio choice and stock centrality for tail risk events. *arXiv preprint arXiv:2112.12031*.

Katsouris, C. (2022). Partial sum processes of residual-based and wald-type break-point statistics in time series regression models. *arXiv preprint arXiv:2202.00141*.

Katsouris, C. (2023a). Estimating conditional value-at-risk with nonstationary quantile predictive regression models. *arXiv preprint arXiv:2311.08218*.

Katsouris, C. (2023b). Statistical estimation for covariance structures with tail estimates using nodewise quantile predictive regression models. *arXiv preprint arXiv:2305.11282*.

Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika*, 41(3-4):403–404.

Kiefer, J. (1967). On bahadur's representation of sample quantiles. *The Annals of Mathematical Statistics*, 38(5):1323–1342.

Knight, K. (1989). Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pages 261–278.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica: Journal of the Econometric Society*, pages 43–61.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, 15(4):143–156.

Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448):1296–1310.

Koenker, R. and Portnoy, S. (1987). L-estimation for linear models. *Journal of the American statistical Association*, 82(399):851–857.

Koenker, R. and Portnoy, S. (1990). M estimation of multivariate regressions. *Journal of the American Statistical Association*, 85(412):1060–1068.

Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612.

Koenker, R. and Xiao, Z. (2004). Unit root quantile autoregression inference. *Journal of the American Statistical Association*, 99(467):775–787.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *Journal of the American statistical association*, 101(475):980–990.

Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. *The annals of Statistics*, pages 435–477.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5):1506–1553.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2018). Taking stock of long-horizon predictability tests: Are factor returns predictable? *Available at SSRN 3284149*.

Koul, H. L. (1991). A weak convergence result useful in robust autoregression. *Journal of Statistical Planning and Inference*, 29(3):291–308.

Koul, H. L. and Saleh, A. M. E. (1995). Autoregression quantiles and related rank-scores processes. *The Annals of Statistics*, pages 670–689.

Kuan, C.-M. and Chen, M.-Y. (1994). Implementing the fluctuation and moving-estimates tests in dynamic econometric models. *Economics Letters*, 44(3):235–239.

Kuelbs, J. (1968). The invariance principle for a lattice of random variables. *The Annals of Mathematical Statistics*, 39(2):382–389.

Kulperger, R., Yu, H., et al. (2005). High moment partial sum processes of residuals in garch models and their applications. *The Annals of Statistics*, 33(5):2395–2422.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178.

Lanne, M. (2002). Testing the predictability of stock returns. *Review of Economics and Statistics*, 84(3):407–415.

Le Cam, L. and Yang, G. L. (2012). *Asymptotics in statistics: some basic concepts*. Springer Science & Business Media.

Lee, J. H. (2016). Predictive quantile regression with persistent covariates: Ivx-qr approach. *Journal of Econometrics*, 192(1):105–118.

Leisch, F., Hornik, K., and Kuan, C.-M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory*, 16(6):835–854.

Li, H., Zheng, C., and Guo, Y. (2016). Estimation and test for quantile nonlinear cointegrating regression. *Economics Letters*, 148:27–32.

Liu, X., Long, W., Peng, L., and Yang, B. (2023). A unified inference for predictive quantile regression. *Journal of the American Statistical Association*, pages 1–15.

Maddala, G. S. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica: Journal of the Econometric Society*, pages 341–358.

Magdalinos, T. (2020). Least squares and ivx limit theory in systems of predictive regressions with garch innovations. *Econometric Theory (forthcoming)*.

Magdalinos, T. (2021). Least squares and ivx limit theory in systems of predictive regressions with garch innovations. *Econometric Theory*, pages 1–38.

Magdalinos, T. and Phillips, P. C. B. (2009a). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory*, 25(2):482–526.

Magdalinos, T. and Phillips, P. C. B. (2009b). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory*, 25(2):482–526.

Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica, Journal of the Econometric Society*, pages 173–220.

Mark, N. C., Ogaki, M., and Sul, D. (2005). Dynamic seemingly unrelated cointegrating regressions. *The Review of Economic Studies*, 72(3):797–820.

Marriott, F. and Pope, J. (1954). Bias in the estimation of autocorrelations. *Biometrika*, 41(3/4):390–402.

McCloskey, A. (2017). Bonferroni-based size-correction for nonstandard testing problems. *Journal of Econometrics*, 200(1):17–35.

McElroy, M. B. and Burmeister, E. (1988). Arbitrage pricing theory as a restricted nonlinear multivariate regression model iterated nonlinear seemingly unrelated regression estimates. *Journal of Business & Economic Statistics*, 6(1):29–42.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, 75(5):1411–1452.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, pages 69–85.

Neocleous, T. and Portnoy, S. (2008). On monotonicity of regression quantile functions. *Statistics & probability letters*, 78(10):1226–1229.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pages 1161–1167.

Nkurunziza, S. (2008). Likelihood ratio test for a special predator-prey system. *Statistics*, 42(2):149–166.

Nkurunziza, S. (2010). Testing concerning the homogeneity of some predator–prey populations. *Journal of statistical planning and inference*, 140(1):323–333.

Olmo, J. (2021). Optimal portfolio allocation and asset centrality revisited. *Quantitative Finance*, pages 1–16.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

Pang, T., Du, L., and Chong, T. T.-L. (2020). Estimating multiple breaks in nonstationary autoregressive models. *Journal of Econometrics*.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2):388–413.

Paye, B. S. (2012). Deja vol: Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, 106(3):527–546.

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.

Perron, P. (1991). A continuous time approximation to the unstable first-order autoregressive process: the case without an intercept. *Econometrica: Journal of the Econometric Society*, pages 211–236.

Pesaran, M. H., Shin, Y., and Smith, R. P. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American statistical Association*, 94(446):621–634.

Phillips, P. C. and Gao, W. Y. (2017). Structural inference from reduced forms with many instruments. *Journal of Econometrics*, 199(2):96–116.

Phillips, P. C. and Magdalinos, T. (2013). Inconsistent var regression with common explosive roots. *Econometric Theory*, 29(4):808–837.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica: Journal of the Econometric Society*, pages 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547.

Phillips, P. C. B. (1988a). Regression theory for near-integrated time series. *Econometrica: Journal of the Econometric Society*, pages 1021–1043.

Phillips, P. C. B. (1988b). Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. *Econometric Theory*, 4(3):528–533.

Phillips, P. C. B. (1989). Partially identified econometric models. *Econometric Theory*, 5(2):181–240.

Phillips, P. C. B. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195.

Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495.

Phillips, P. C. B. and Hansen, B. E. (1990). Statistical inference in instrumental variables regression with i (1) processes. *The Review of Economic Studies*, 57(1):99–125.

Phillips, P. C. B. and Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2):250–264.

Phillips, P. C. B. and Lee, J. H. (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics*, 192(2):433–450.

Phillips, P. C. B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1):115–130.

Phillips, P. C. B. and Magdalinos, T. (2009). Econometric inference in the vicinity of unity. *Singapore Management University, CoFie Working Paper*, 7.

Phillips, P. C. B. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111.

Phillips, P. C. B. and Park, J. Y. (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association*, 83(401):111–115.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, pages 971–1001.

Pitarakis, J.-Y. (2004). Least squares estimation and tests of breaks in mean and variance under misspecification. *The Econometrics Journal*, 7(1):32–54.

Pitarakis, J.-Y. (2014). A joint test for structural stability and a unit root in autoregressions. *Computational Statistics & Data Analysis*, 76:577–587.

Pitarakis, J.-Y. (2017). A simple approach for diagnosing instabilities in predictive regressions. *Oxford Bulletin of Economics and Statistics*, 79(5):851–874.

Ploberger, W. and Krämer, W. (1990). The local power of the cusum and cusum of squares tests. *Econometric Theory*, 6(3):335–347.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(2):186–199.

Portnoy, S. (1991). Asymptotic behavior of regression quantiles in non-stationary, dependent cases. *Journal of Multivariate analysis*, 38(1):100–113.

Portnoy, S. (2012). Nearly root-n approximation for regression quantile processes. *The Annals of Statistics*, 40(3):1714–1736.

Poskitt, D. S. (2006). On the identification and estimation of nonstationary and cointegrated armax systems. *Econometric Theory*, 22(6):1138–1175.

Prokhorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, 1(2):157–214.

Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics*, 146(1):170–184.

Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502.

Qu, Z. and Yoon, J. (2015). Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics*, 185(1):1–19.

Rao, M. (1978). Asymptotic distribution of an estimator of the boundary parameter of an unstable process. *The Annals of Statistics*, pages 185–190.

Ravikumar, B., Ray, S., and Savin, N. E. (2000). Robust wald tests in sur systems with adding-up restrictions. *Econometrica*, 68(3):715–719.

Ren, Y., Tu, Y., and Yi, Y. (2019). Balanced predictive regressions. *Journal of Empirical Finance*, 54:118–142.

Révész, P. (1982). On the increments of wiener and related processes. *The Annals of Probability*, pages 613–622.

Rossi, B. and Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, 30(3):432–453.

Seo, B. (1998). Tests for structural change in cointegrated systems. *Econometric Theory*, 14(2):222–259.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):343–366.

Shin, D. W., Kim, H. J., and Seo, J. (2016). Sur approach for iv estimation of canonical contagion models. *Communications in Statistics-Simulation and Computation*, 45(1):378–387.

Skorokhod, A. V. (1956). Limit theorems for stochastic processes. *Theory of Probability & Its Applications*, 1(3):261–290.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Stock, J. H. (1994). Unit roots, structural breaks and trends. *Handbook of econometrics*, 4:2739–2841.

Strotz, R. H. and Wold, H. O. (1960). Recursive vs. nonrecursive systems: An attempt at synthesis (part i of a triptych on causal chain systems). *Econometrica: Journal of the Econometric Society*, pages 417–427.

Su, L. and Xiao, Z. (2008). Testing for parameter stability in quantile regression models. *Statistics & Probability Letters*, 78(16):2768–2775.

Tan, L., Chiong, K. X., and Moon, H. R. (2021). Estimation of high-dimensional seemingly unrelated regression models. *Econometric Reviews*, 40(9):830–851.

Tu, Y. and Xie, X. (2023). Penetrating sporadic return predictability. *Journal of Econometrics*, 237(1):105509.

Uematsu, Y. (2019). Nonstationary nonlinear quantile regression. *Econometric Reviews*, 38(4):386–416.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer Series in Statistics.

Wang, C.-S. and Zhao, Z. (2016). Conditional value-at-risk: Semiparametric estimation and inference. *Journal of Econometrics*, 195(1):86–103.

Wang, Q. and Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *The Annals of Statistics*, 40(2):727–758.

Wang, W., Xuening, Z., Wang, H., and Haerdle, W. (2018). Network quantile autoregression. *Journal of Econometrics*.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

White, H. (1996). *Estimation, inference and specification analysis*. Number 22. Cambridge university press.

White, H. (2001). *Asymptotic theory for econometricians*. Academic press.

White, H., Kim, T.-H., and Manganelli, S. (2015). Var for var: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics*, 187(1):169–188.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics*, pages 1188–1197.

White, J. S. (1959). The limiting distribution of the serial correlation coefficient in the explosive case ii. *The Annals of Mathematical Statistics*, pages 831–834.

Xiao, Z. (2009). Quantile cointegrating regression. *Journal of econometrics*, 150(2):248–260.

Xu, K.-L. (2020). Testing for multiple-horizon predictability: Direct regression based versus implication based. *The Review of Financial Studies*, 33(9):4403–4443.

Xu, K.-L. and Guo, J. (2019). A dimensionality-robust test in multiple predictive regression. *Available at SSRN 3458074*.

Xu, K.-L. and Guo, J. (2022). A new test for multiple predictive regression. *Journal of Financial Econometrics*, pages 1–38.

Xu, X., Wang, W., Shin, Y., and Zheng, C. (2022). Dynamic network quantile regression model. *Journal of Business & Economic Statistics*, pages 1–15.

Yang, B., Long, W., Peng, L., and Cai, Z. (2020). Testing the predictability of us housing price index returns based on an ivx-ar model. *Journal of the American Statistical Association*, 115(532):1598–1619.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, 57(298):348–368.

Zhou, K. Q. and Portnoy, S. L. (1998). Statistical inference on heteroscedastic models based on regression quantiles. *Journal of Nonparametric Statistics*, 9(3):239–260.

Zhu, F., Liu, M., Ling, S., and Cai, Z. (2022). Testing for structural change of predictive regression model to threshold predictive regression model. *Journal of Business & Economic Statistics*, 41(1):228–240.

Zhu, X., Wang, W., Wang, H., and Härdle, W. K. (2019). Network quantile autoregression. *Journal of econometrics*, 212(1):345–358.