Robust Inference with Stochastic Local Unit Root Regressors in Predictive Regressions*

Yanbo Liu
School of Economics, Shandong University

Peter C. B. Phillips
Yale University, University of Auckland,
University of Southampton, Singapore Management University

June 13, 2022

Abstract

This paper explores predictive regression models with stochastic unit root (STUR) components and robust inference procedures that encompass a wide class of persistent and time-varying stochastically nonstationary regressors. The paper extends the mechanism of endogenously generated instrumentation known as IVX, showing that these methods remain valid for short and long-horizon predictive regressions in which the predictors have STUR and local STUR (LSTUR) generating mechanisms. Both mean regression and quantile regression methods are considered. The asymptotic distributions of the IVX estimators are new and require some new methods in their derivation. The distributions are compared to previous results and, as in earlier work, lead to pivotal limit distributions for Wald testing procedures that remain robust for both single and multiple regressors with various degrees of persistence and stochastic and fixed local departures from unit roots. Numerical experiments corroborate the asymptotic theory, and IVX testing shows good power and size control. The IVX methods are illustrated in an empirical application to evaluate the predictive capability of economic fundamentals in forecasting S&P 500 excess returns.

JEL classification: C12, C22, G01

Keywords: IVX, Long horizon, LSTUR, Predictability, Quantile regression, Robustness, Short horizon, STUR.

*Thanks are due to the CoEditor, Torben Andersen, and two referees for helpful comments on the paper and guidance for revision. We also thank Jia Li, Nan Liu, Tassos Magdalinos, Liangjun Su, Katsuto Tanaka, Qiying Wang, Jun Yu, Yichong Zhang and the participants of the 2019 SETA, the 2019 AMES, and the 2019 CMES conferences for suggestions and discussions. Phillips acknowledges support from a Lee Kong Chian Fellowship at SMU, the Kelly Fund at the University of Auckland, and the NSF under Grant No. SES 18-50860. Yanbo Liu, School of Economics, Shandong University, 27 Shanda South Road, Jinan, Shandong, 250100. Email: yanbo.liu.2015@phdecons.smu.edu.sg. Corresponding author: Peter C.B. Phillips, Cowles Foundation for Research in Economics, Yale University, Box 208281, Yale Station, New Haven, Connecticut 06520-8281. Email: peter.phillips@yale.edu.

1 Introduction

Testing the predictability of financial asset returns has generated a vast literature in empirical finance and remains a primary focus of ongoing research. Continuing concerns in the use of testing procedures are the validity, accuracy, and robustness of the econometric methods to the properties of the regressors that are used as economic fundamentals in the regressions. Several methods have been suggested to address these concerns and are now used in empirical work. Much attention in the development of these methods has been given to achieving robustness and size control in testing to the (typically unknown) degree of persistence in the regressors.

One well-established approach employs a local to unity (LUR) formulation to model potential persistence and simulation-based Bonferroni bounds for inference, as developed in Campbell and Yogo (2006). This approach is successful in the scalar LUR case but fails for stationary regressors (Phillips, 2014), is challenging to extend to multiple regressors and is not designed for regressors with varying degrees of persistence or potentially stochastic departures from unit-roots (Phillips and Lee, 2013). Since these characteristics commonly arise in economic fundamentals, and their precise nature remains unknown, methods that are capable of robust inference under these conditions and which apply in multiple regressions are needed to support the growing body of empirical research on financial predictability.

Another approach that seeks to address these needs involves the use of endogenous instrumentation for the predictive regressors. This method, which is known as IVX instrumentation and developed in Phillips and Magdalinos (2009), was extended and utilized in an extensive empirical application by Kostakis et al. (2015) to reveal its potential in applied research. The method has several significant advantages compared with Bonferroni-type simulation-based approaches. First, standard significance testing procedures such as Wald tests are applicable with convenient pivotal chi-square limit distributions that hold for a wide range of persistence characteristics in the regressors, including LUR and mildly integrated root (MIR) regressors (Phillips and Magdalinos, 2007). Other new methods can also achieve pivotal limit theory test but rely on the use of additional regressors (Lin and Tu, 2020; Cai et al., 2020) and have only been applied with non-stochastic regression coefficients. No simulation methods are needed for implementation of these techniques.

Second, no prior knowledge regarding the degree of persistence or the presence of stochastic departures from unit root conditions is required for IVX instrumentation. In contrast, the Bonferroni bound approach applies only to LUR regressors, and supplemental methods (like switching) are needed to support testing in the presence of MIR or stationary regressors (Elliott et al., 2015). Third, in contrast to most other methods, IVX conveniently accommodates multiple regressors, as illustrated in Kostakis et al. (2015), and allows an extension to locally explosive and mildly explosive regressor cases as well as mixed-root cases (Phillips and Lee, 2016). Fourth, the method applies in both short-horizon and long-horizon predictive regressions, again with a pivotal chi-square limit theory for Wald tests. Fifth, the IVX methodology may be used in quantile regressions

(QR), as well as mean predictive regressions, as shown by Lee (2016) and subsequently, Fan and Lee (2019) in models with conditional heteroskedasticity errors. This QR-IVX-Wald test is particularly useful in checking for predictability under tail conditions where more extreme return behavior occurs.

A feature of the empirical finance literature that is particularly important for the present study is the recognition that parameter instability can be critical in both asset price determination as well as economic fundamentals. Coefficients can vary over time for many reasons, such as changes in regulatory conditions, shifts in market sentiment, or adjustments in monetary policy and targeting, as well as evolution in financial institutions and the impact of technology on transmission mechanisms and information dissemination. These influences can, in turn, affect the generating mechanism of predictive regressors, including the degree of persistence, thereby supporting formulations such as stochastic departures from unit roots and time-varying coefficient formulations. In the stock price application discussed later in the paper, a detection method for stochastic deviations from unity in the autoregressive coefficient is used to show empirical evidence of STUR behavior in the predictive regressor. In past research, there is ample support for time variation of the parameters in much financial econometric and macroeconomic modeling work. Bossaerts and Hillion (1999), for instance, cited poor performance in many prediction models and indications that the parameters of even the best models change over time. Amongst many other studies, Bekaert et al. (2007) showed patterns of time variation in model coefficients across sub-periods, and random walk specifications are frequently employed to capture parameter randomness and time variation in dynamic macroeconomic models (e.g. Cogley and Sargent, 2001, 2005).

The econometric literature has a long history of modeling with time-varying parameters, including stochastic process formulations. Granger and Swanson (1997) introduced a stochastic unit root model in which a stationary process is embedded into the autoregression coefficient, proposing a unit root test that accommodates such departures. Other early contributions of stochastic deviations from unit-roots appear in McCabe and Smith (1998), and Yoon (2006). Parameter instability is also discussed in the context of cointegration and predictive regression models (Gonzalo and Pitarakis, 2012; Georgiev et al., 2018; Demetrescu et al., 2020). More recently, Lieberman and Phillips (2017) developed asymptotic theory and inferential procedures for nonlinear least-squares estimation of the STUR model and extended the Black-Scholes asset pricing formula to this more general model setup. Lieberman and Phillips (2020a) further extended the analysis to a stochastic version of the local unit root model called the LSTUR model. Tao et al. (2019) studied a continuous-time variant of the same LSTUR model, which has been used to model derivative pricing in mathematical finance, employed infill asymptotics to establish asymptotic properties, and analyzed evidence of instability and bubble behavior in financial data.

An essential feature of LSTUR models that is relevant for the use of persistent regressors in predictive regressions is that LSTUR processes have means and variances the same as an elementary random walk process but with kurtosis exceeding 3. This property is consistent with the heavy-tail behavior of much financial data, thereby offering the prospect of improved modeling representations for both asset prices and economic fundamentals in predictive regressions. Whereas these features of STUR, LSTUR, and more general time-varying stochastic models are recognized as useful in capturing relevant financial data characteristics, there is as yet no treatment of the properties of predictive regression in the presence of such regressors.

The present paper responds to this need by studying short, and long-horizon mean predictive regressions and quantile predictive regressions with LSTUR regressors. The analysis reveals size distortions in predictability testing for both short-horizon and long-horizon regressions with standard methods. A primary contribution of the paper is to develop a version of the IVX endogenous instrumentation technique to address these difficulties in conventional predictive regression test procedures, together with the asymptotic properties of the IVX estimators and associated asymptotically pivotal tests. Both mean predictive regression (IVX and LHIVX) and quantile predictive regression (QR-IVX) approaches are considered. The IVX methods are shown to have excellent finite sample performance in simulations, not only with mixed (stationary and explosive) roots but also with random departures. In sum, the attractive features of IVX methodology are the availability of standard asymptotic chi-square inference in models with multiple predictive regressors and, as shown here, the allowance for random departures from unity in the autoregressive roots of the predictors. A second contribution of the paper is technical. The presence of random departures from unity brings new complications to the development of the IVX limit theory and, in particular, the establishment of mixed normal asymptotics for IVX estimators. New methods of proof of mixed normality are developed that use Réyni mixing techniques and take advantage of some of the special properties of the mildly integrated processes used as instruments in IVX estimation.

The paper is structured as follows. Section 2 gives the model setup and briefly revisits the size distortion problems of standard methods of inference. Section 3 presents the self-generated endogenous instrument and establishes the pivotal limit theory for IVX tests in both short and long horizons. Section 4 proves the validity of the IVX methods with mixed roots and random departures from unity. Section 5 reports the results of simulations that explore the finite sample performance of the IVX methods. Section 6 applies the short-horizon and long-horizon IVX tests to the S&P 500 financial market. Section 7 concludes.

Notational conventions are as follows. We use $:=$ and $=:$ for definitional identity, and for an $m \times r$ matrix A with $m \geq r$, $\|A\| \left(:= \max_i \left\{ \lambda_i^{1/2} : \lambda_i = \text{an eigenvalue of } A'A \right\}\right)$ denotes the spectral norm, L_2 (Frobenius) and L_1 norms are specified as $\|\cdot\|_F$ and $\|\cdot\|_1$, $P_A = A\left(A'A\right)^{+} A'$, where A^{+} is the Moore Penrose generalized inverse of A, $M_A = I_m - P_A$, and I_m is the identity matrix of order m. The symbol $\mathbb{E}_{t-1}(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_{t-1})$ denotes conditional expectation with respect to the filtration $\mathcal{F}_{t-1}$. The symbol $=_d$ denotes equivalence in distribution, $a_T \preceq b_T$ signifies that a_T/b_T is either $O_p(1)$ or $o_p(1)$, $\mathbf{0}_{m\times r}$ denotes an $m\times r$ matrix of zeros, and $\rightsquigarrow$ signifies weak convergence in both Euclidean space and function space according to context. The notation $\to_p$ is convergence in probability, for two sequences a_T and b_T the notation $a_T \sim b_T$ signifies $\lim_{T\to\infty} a_T/b_T = 1$, and $a_T \sim_a b_T$ denotes $\Pr\left(|a_T/b_T| \neq 1\right) \to 0$ as $T \to \infty$.

2 Model Setup and Size Distortions

This section presents three models that are commonly used in predictive regression: the short-horizon mean predictive regression model, the short-horizon quantile predictive regression model, and the long-horizon mean prediction model. Problems of size control in such predictive regressions when the regressors have persistent characteristics are well known. The present section discusses these issues in the context of conventional estimation and inferential methods. The framework provides a unified setting for inference that allows for both stochastic and deterministic local departures from unity in the generating mechanisms of the persistent regressors.

2.1 Model and assumptions

The standard predictive mean regression model has the form

$$y_t = \beta_0 + \beta_1' x_{t-1} + u_{0t} = \beta' \check{X}_{t-1} + u_{0t}, \text{ with } \mathbb{E}_{t-1}(u_{0t}) = 0, \tag{1}$$

where $\check{X}_{t-1} := \left(1, x_{t-1}'\right)'$ includes both the intercept and regressor x_{t-1}, $\beta' = (\beta_0, \beta_1')$, β_1 is an n-vector of regression coefficients and $\mathcal{F}_t$ is a suitable natural filtration, defined later, for which u_{0t} is a martingale difference sequence (mds). It is convenient to write the model (1) in observation format as

$$y = \check{X}_{-1}\beta + u_0,$$

where $y' = (y_1, ..., y_T)$, $\check{X}_{-1}' = \left(\check{X}_0, ..., \check{X}_{T-1}\right)$, $X_{-1}' = (x_0, ..., x_{T-1})$ and $u_0' = (u_{01}, ..., u_{0T})$ for sample size T .

The n-vector of predictors x_{t-1} in (1) is assumed throughout this paper to have a stochastic unit root generating mechanism that belongs to the STUR or LSTUR family, implying the following locally persistent autoregressive form[1]

$$x_t = R_{Tt} x_{t-1} + u_{xt}, \;\; R_{Tt} = \begin{cases} I_n + \frac{C}{T} + \frac{a' u_{at}}{\sqrt{T}} I_n + \frac{(a' u_{at})^2}{2T} I_n & \text{under LSTUR} \\ I_n + \frac{a' u_{at}}{\sqrt{T}} I_n + \frac{(a' u_{at})^2}{2T} I_n & \text{under STUR} \end{cases}. \tag{2}$$

Under this scheme only a finite number of regressors is considered, so n is fixed and $T \to \infty$. In Eqn (2), $C := \text{diag}\left\{c_1, c_2, ..., c_n\right\}$ with $\{c_i\}_{i=1}^n$ being a set of scalar localizing coefficients,

[1] As in Lieberman and Phillips (2020a), it is often convenient to employ asymptotically equivalent matrix exponential formulations of the autoregressive coefficient in (2) without affecting the limit theory, viz., $R_{Tt} = \exp\left(\frac{C}{T} + \frac{a' u_{at}}{\sqrt{T}} I_n\right)$ for LSTUR. Such formulations are often used in the proofs in Appendix A and B. The errors induced by the differences between such formulations and (2) can be neglected in the development of the asymptotic theory, as shown in Lieberman and Phillips (2020a).

and u_{at} being a p-vector of random variables that influence stochastic departures from unit roots in the coefficient matrix R_{Tt}, which has a time varying character in contrast to the more usual LUR specification in which the stochastic component u_{at} does not appear. The parameter vector $a = (a_1, ..., a_p)'$ is a p-dimensional coefficient that appears in both STUR and LSTUR formulations and is, in the formulation (2), common across equations. The latter homogeneity is a matter of notational convenience and the results of the paper extend to the case where the coefficient a may differ across equations.[2] This formulation induces further notational burdens and these are avoided for convenience in the main text by assuming commonality in the coefficients. A discussion of the heterogeneous coefficient case is given in Section C of the Online Supplement. The initialization of x_t is assumed for convenience to be $x_0 = \mathbf{0}_{n\times 1}$ but this can be relaxed to the weaker requirement $x_0 = o_p(\sqrt{T})$ without affecting the main results.

In practical work neither the degree of persistence nor the presence of any stochastic departures from unit roots in the observed time series is known. In the case where there is a correlation between u_{at} and u_{xt} in the regression model (1), a further complication is that the STUR coefficient a is not consistently estimable, as shown in Lieberman and Phillips (2017). Moreover, standard unit root tests have little discriminatory power in distinguishing between the unit root and STUR (or LSTUR) processes in models such as (2). It is therefore desirable to have methods of estimation and inference that are robust to various formulations within the class of near-unity regressors (Phillips, 1987; Phillips and Magdalinos, 2007) and allow for general forms of dependence on the innovations. The following assumptions are used throughout the paper.

Assumption 1

(i) Let $\epsilon_t = (\epsilon_{0t}, \epsilon_{xt}', \epsilon_{at}')'$, with ϵ_{0t}, ϵ_{xt} and ϵ_{at} as in (4) and (5), denote an $\mathbb{R}^{1+n+p}$-valued martingale difference sequence with respect to the natural filtration $\mathcal{F}_t = \sigma(\epsilon_t, \epsilon_{t-1}, ...)$ satisfying

$$\mathbb{E}_{t-1}\left[\epsilon_t\epsilon_t'\right] = \Sigma_\epsilon \text{ a.s. and } \sup_{1\le t\le T} \mathbb{E}\left\|\epsilon_t\right\|^{2s} < \infty, \tag{3}$$

for some $s > 1$, where Σ_ϵ is a positive definite matrix.

(ii) The process $\{u_{0t}\}_{1\le t\le T}$ admits the following GARCH(q_1,q_2) representation,

$$u_{0t} = H_t^{1/2}\epsilon_{0t}, \;\; H_t = \varphi_0 + \sum_{l=1}^{q_1} \varphi_{1l} u_{0,t-l}^2 + \sum_{k=1}^{q_2} \varphi_{2k} H_{t-k}, \tag{4}$$

where $\{\epsilon_{0t}\}_{1\le t\le T}$ is the above mentioned martingale difference sequence with a unit variance, φ_0 is a positive constant value, φ_{1l} and φ_{2k} are nonnegative for all l, k, and $\sum_{l=1}^{q_1}\varphi_{1l} + \sum_{k=1}^{q_2}\varphi_{2k} < 1$.

[2] In this case, it is necessary to use a matrix formulation of multiple inner products $\{a_i' u_{at}\}$ in the LSTUR and STUR specifications, where the $\{a_i\}_{i=1}^n$ are each p-dimensional coefficient vectors. This extension to the heterogeneous coefficient case can be achieved by introducing the matrix $\breve{D}_{at} := \text{diag}\{a_1'u_{at}, ..., a_n'u_{at}\}$, as shown in Section C of the Online Supplement.

(ii') The process $\{u_{0t}\}_{1\le t\le T}$ *admits the martingale difference sequence structure* $u_{0t} = \sqrt{\varphi_0}\cdot\epsilon_{0t}$*, where* $\{\epsilon_{0t}\}_{1\le t\le T}$ *is the above mentioned martingale difference sequence with a unit variance and* φ_0 *is a positive constant.*

(iii) The processes $\{u_{xt}\}_{1\le t\le T}$ *and* $\{u_{at}\}_{1\le t\le T}$ *follow the stationary linear processes*

$$u_{xt} = \sum_{j=0}^{\infty} F_{xj}\epsilon_{x,t-j},\ u_{at} = \sum_{j=0}^{\infty} F_{aj}\epsilon_{a,t-j}, \tag{5}$$

where $\{F_{xj}\}_{j\ge 0}$ *and* $\{F_{aj}\}_{j\ge 0}$ *are matrix sequences such that* $F_{x0} = I_n$*,* $F_{a0} = I_p$*, both* $\sum_{j=0}^{\infty} F_{xj}$ *and* $\sum_{j=0}^{\infty} F_{aj}$ *have full rank,* $\sum_{j=0}^{\infty} j\,\|F_{xj}\| < \infty$*, and* $\sum_{j=0}^{\infty} j\,\|F_{aj}\| < \infty$*.*

Assumption 1 (iii) imposes homoskedasticity on the errors u_{xt} and u_{at}, whereas Assumption 1 (ii) allows for both conditional heteroskedasticity and second-order serial correlation taking the form of a stationary GARCH(q_1,q_2) process for the prediction error u_{0t}. In particular, as in Assumption 1 (i), u_{0t} is a martingale difference sequence with respect to the natural filtration $\mathcal{F}_t$. Assumption 1 (i), (ii), and (iii) apply to the discussion of the mean predictive regressions, in both short and long horizons. Assumption 1 (ii') simplifies Assumption 1 (ii) by allowing conditional homoskedasticity. For the discussion of the quantile predictive regression model and its estimation we impose Assumption 1 (i), (ii'), and (iii).

Remark 2.1 *The GARCH process is a typical example of a strong mixing process with exponential tails (Fan and Lee, 2019). As usual, the* α*-mixing coefficients* $\alpha(j)$ *are defined as*

$$\alpha(j) = \sup_m \sup_{A\in\mathcal{F}_{(-\infty,m)},B\in\mathcal{F}_{(m+j,\infty)}} |\Pr(A)\Pr(B) - \Pr(AB)|,$$

in which $\mathcal{F}_{(-\infty,m)} = \sigma(\ldots,\check{v}_m)$ *and* $\mathcal{F}_{(m+j,\infty)} = \sigma(\check{v}_{m+j},\check{v}_{m+j+1},\ldots)$*. When* $\alpha(j)\to 0$ *as* $j\to\infty$*, then* $\{\check{v}_t\}_{1\le t\le T}$ *is an* α*-mixing sequence and* $\check{v}_t\in\mathcal{F}_t$*. We assume that the prediction error* $u_{0t} = H_t\epsilon_{0t}$ *and the process* $\check{v}_t = (\epsilon_{0t}, H_{t+1})'$ *are* α*-mixing of size* $\frac{-r}{r-2}$ *for* $r > 2$*, i.e., the mixing coefficients* $\alpha(j) = O(j^{\lambda})$ *for* $\lambda < \frac{-r}{r-2} < 0$*. Then the functional central limit theorem in (8) still holds for* u_{0t} *and the derivations of the short-horizon and long-horizon mean predictive regression models have no material changes. So without losing generality and following Kostakis et al. (2015), the properties of IVX instrumentation may be based on a* $GARCH(q_1,q_2)$ *error process.*

The instantaneous and long-run variance matrices of the innovations $(u_{0t},u'_{xt},u'_{at})'$ are defined as follows.

$$\sigma_{00} = \mathbb{E}\left(u_{0t}^2\right),\ \Sigma_{xx} = \mathbb{E}\left(u_{xt}u'_{xt}\right),\ \Sigma_{aa} = \mathbb{E}\left(u_{at}u'_{at}\right),\ \Sigma_{a0} = \mathbb{E}\left(u_{at}u_{0t}\right),$$
$$\Omega_{xx} = \sum_{h=-\infty}^{\infty}\mathbb{E}\left(u_{xt}u'_{x,t-h}\right),\ \Omega_{aa} = \sum_{h=-\infty}^{\infty}\mathbb{E}\left(u_{at}u'_{a,t-h}\right),\ \Sigma_{x0} = \mathbb{E}\left(u_{xt}u_{0t}\right),$$

$$\Lambda_{xx} = \sum_{h=1}^{\infty} \mathbb{E}\left(u_{xt}u'_{x,t-h}\right),\ \Lambda_{aa} = \sum_{h=1}^{\infty} \mathbb{E}\left(u_{at}u'_{a,t-h}\right),\ \Lambda_{ax} = \sum_{h=1}^{\infty} \mathbb{E}\left(u_{at}u'_{x,t-h}\right),$$
$$\Omega_{ax} = \sum_{h=-\infty}^{\infty} \mathbb{E}u_{at}u'_{x,t-h},\ \Sigma_{ax} = \mathbb{E}\left(u_{at}u'_{xt}\right),\ \Delta_{ax} = \sum_{h=0}^{\infty} \mathbb{E}\left(u_{at}u'_{x,t-h}\right),$$
$$\Lambda_{a0} = \sum_{h=1}^{\infty} \mathbb{E}\left(u_{at}u_{0,t-h}\right),\ \Lambda_{x0} = \sum_{h=1}^{\infty} \mathbb{E}\left(u_{xt}u_{0,t-h}\right),\ \Delta_{x0} = \Lambda_{x0} + \Sigma_{x0},$$
$$\Omega_{xx} = \Lambda_{xx} + \Lambda'_{xx} + \Sigma_{xx},\ \Omega_{aa} = \Lambda_{aa} + \Lambda'_{aa} + \Sigma_{aa},\ \Delta_{a0} = \Lambda_{a0} + \Sigma_{a0}. \tag{6}$$

Since u_{0t} is uncorrelated in first order, the two-sided long-run variance Ω_{00} equals the instantaneous variance σ_{00}. Letting $\widehat{u}_{0t} = y_t - \widehat{\beta}_0 - \widehat{\beta}'_1 x_{t-1}$ be the residuals from consistent regression estimation of (1) by least squares or IVX, the instantaneous prediction error variance σ_{00} can be estimated in the usual way by

$$\widehat{\sigma}_{00} = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_{0t}^2. \tag{7}$$

Under Assumption 1, the usual Beveridge-Nelson decomposition applies and by standard functional limit theory (Phillips and Solo, 1992)

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{\lfloor Ts \rfloor} \begin{bmatrix} u_{0j} \\ u_{xj} \\ u_{aj} \end{bmatrix} \rightsquigarrow \begin{bmatrix} B_0(s) \\ B_x(s) \\ B_a(s) \end{bmatrix} = BM \begin{bmatrix} \sigma_{00} & \Delta'_{x0} & \Delta'_{a0} \\ \Delta_{x0} & \Omega_{xx} & \Omega'_{ax} \\ \Delta_{a0} & \Omega_{ax} & \Omega_{aa} \end{bmatrix}, \tag{8}$$

where $BM(\cdot)$ denotes the vector Brownian motion. Using the limit theory (8), the corresponding limit processes for STUR and LSTUR cases are given by

$$\frac{x_{\lfloor Tr \rfloor}}{\sqrt{T}} \rightsquigarrow G_a(r) := e^{a'B_a(r)}\left\{\int_0^r e^{-a'B_a(p)}dB_x(p) - \left(\int_0^r e^{-a'B_a(p)}dp\right)\Delta_{xa}a\right\}, \tag{9}$$

and

$$\frac{x_{\lfloor Tr \rfloor}}{\sqrt{T}} \rightsquigarrow G_{a,c}(r) := e^{rC+a'B_a(r)I_n}\left\{\int_0^r e^{-pC-a'B_a(p)I_n}dB_x(p) - \left(\int_0^r e^{-pC-a'B_a(p)I_n}dp\right)\Delta_{xa}a\right\}, \tag{10}$$

where $\Delta_{xa} = \Sigma_{xa} + \Lambda_{xa}$. The limits in (9) and (10) were obtained by Lieberman and Phillips (2017, 2020a) in the scalar case and similar derivations (not repeated) lead to (9) and (10) here in the general case.

Remark 2.2 *Lieberman and Phillips (2020a) derived the stochastic differential equation satisfied by these stochastic processes and showed that they have instantaneous means and variances that resemble those of Brownian motion but with instantaneous kurtosis (Lieberman and Phillips, 2020b) exceeding 3, a feature that is coherent with the properties of much financial data for which heavy tail behavior is commonly present. To illustrate, consider the scalar LSTUR case in which* $dim\left(u_{at}\right) = dim\left(u_{xt}\right) = 1$, u_{at} *and* u_{xt} *are independently* *and identically distributed (i.i.d.), and* $\mathbb{E}(u_{at}u_{xt}) = 0$*. Let* $b = a^2\sigma_{aa}$*. For the variance and kurtosis of* $G_{a,c}(r)$*, the following properties hold*

$$\lim_{b+c\to 0} \mathbb{E}\left(G_{a,c}^2(r)\right) = \sigma_{xx}r, \quad \lim_{b+c\to 0} \mathbb{E}\left(G_{a,c}^4(r)\right) = \frac{3\sigma_{xx}^2\left(\exp\left(-4cr\right) + 4cr - 1\right)}{8c^2}. \tag{11}$$

Therefore, when $c + b = 0$ *and* $c < 0$*, the variance of* $G_{a,c}(r)$ *matches that of Brownian motion but the kurtosis of* $G_{a,c}(r)$ *exceeds* 3*. As* $|c|$ *increases the kurtosis of the LSTUR process satisfies*

$$\lim_{b+c\to 0} \frac{\mathbb{E}\left(G_{a,c}^4(r)\right)}{\left(\mathbb{E}\left(G_{a,c}^2(r)\right)\right)^2} = \frac{3\left(\exp\left(-4cr\right) + 4cr - 1\right)}{8\left(cr\right)^2},$$

which diverges as $c \to -\infty$ *and has its minimum of* 3 *at* $c = 0$*. The above analysis confirms the heavy-tailed property of the LSTUR process.*

In what follows we explore the potential of using IVX methods to deal with this broad class of persistent regressors with parameter instability and heavy tails in the context of short-horizon predictive regressions (with mean and quantile regressions) and long-horizon mean predictive regression.

2.1.1 Short-horizon mean and quantile predictive regressions

For the short-horizon mean predictive regression case, the model setup and innovation structures are given in (1), (2), and Assumption 1. In addition to the mean predictive regression, the following quantile predictive regression model is considered. Following Xiao (2009), our model is based on the linear quantile representation

$$Q_{y_t}\left(\tau|\mathcal{F}_{t-1}\right) = \beta_{0,\tau} + \beta_{1,\tau}' x_{t-1}, \tag{12}$$

where $\mathcal{F}_t$ is the natural filtration defined in Assumption 1 and $Q_{y_t}(\tau|\mathcal{F}_{t-1})$ is the conditional quantile of y_t, so that

$$\Pr\left(y_t \leq Q_{y_t}\left(\tau|\mathcal{F}_{t-1}\right)|\mathcal{F}_{t-1}\right) = \tau \in (0, 1).$$

The loss function of quantile regression is defined by $\rho_\tau(u) := u \cdot \Psi_\tau(u)$ where $\Psi_\tau := \tau - \mathbf{1}\left(u < 0\right)$ and $\mathbf{1}\left(\cdot\right)$ denotes an indicator function. With Assumption 1 (ii'), the quantile regression innovation satisfies

$$\Psi_\tau\left(u_{0t\tau}\right) \sim mds\left(0, \tau\left(1 - \tau\right)\right),$$

where $u_{0t\tau} := u_{0t} - P_{u_0}^{-1}(\tau)$ and $P_{u_0}^{-1}(\tau)$ is the unconditional $\tau-$quantile of u_{0t}. Further, we define $\Delta_{x\Psi_\tau} := \sum_{h=0}^{\infty} \mathbb{E}\left(u_{xt}\Psi_\tau\left(u_{0,t-h,\tau}\right)\right)$ and $\Delta_{a\Psi_\tau} := \sum_{h=0}^{\infty} \mathbb{E}\left(u_{at}\Psi_\tau\left(u_{0,t-h,\tau}\right)\right)$. In addition, to facilitate the asymptotic development, several regularity conditions on the conditional densities $\{p_{u0t\tau,t-1}\}$ of $u_{0t\tau}$ given $\mathcal{F}_{t-1}$ are imposed (c.f., Xiao, 2009; Lee, 2016).

Assumption 2 *(i) A sequence of stationary conditional probability densities* $\{p_{u0t\tau,t-1}(\cdot)\}$ *evaluated at zero satisfies a functional central limit theorem with a nondegenerate mean* $p_{u0\tau}(0) = \mathbb{E}[p_{u0t\tau,t-1}(0)]$*, where* $p_{u0t\tau,t-1}(\cdot)$ *is the conditional density of* $u_{0t\tau}$ *given* $\mathcal{F}_{t-1}$*, so that*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}(p_{u0t\tau,t-1}(0) - p_{u0\tau}(0)) \rightsquigarrow B_{pu0\tau}(r).$$

(ii) For each t *and* $\tau \in (0,1)$*, the conditional density* $p_{u0t\tau,t-1}(0)$ *is bounded above with probability one, i.e.* $p_{u0t\tau,t-1}(x) < \infty$ *in probability for all* $|x| < M$ *with some* $M > 0$*.*

2.1.2 Long-horizon mean predictive regression

In short-horizon predictive regressions, the time horizon of prediction only covers a single period. However, it is common to use a longer horizon in empirical research in which the time horizon is mildly increasing with the sample size, T. In particular, we set the horizon as $k = T^{\nu}$ for some $\nu \in (0,1)$ and use the implied rate restriction, $\frac{1}{k} + \frac{k}{T} \to 0$. As discussed in Phillips and Lee (2013), the long-horizon prediction model with a mildly increasing time window is formulated as

$$y_{t+k} = B_0 + B_1 x_t^k + u_{0,t+k}, \text{ and } x_t^k = \sum_{j=1}^{k} x_{t+j-1}, \tag{13}$$

with the null hypothesis $\mathcal{H}_0(k)$: $B_1 = \mathbf{0}_{1\times n}$ and alternative $\mathcal{H}_1(k)$: $B_1 \neq \mathbf{0}_{1\times n}$, where B_1 is a $1 \times n$ row vector. Inference concerning long-horizon predictability under STUR and LSTUR regressors can be conducted by empirically fitting equation (13) and allowing for stationary disturbances in the long-horizon model.

2.2 Size distortions in traditional regressions

Allowing for STUR and LSTUR regressors, we provide below the limit theory for standard short-horizon quantile predictive regressions. These results facilitate the study of the size distortions that arise in using traditional testing methods.

Standard quantile regression (QR) estimation optimizes the following objective function

$$\widehat{\beta}_{\tau}^{QR} := \arg\min_{\beta^{\star}} \sum_{t=1}^{T} u_{0t} \cdot \Psi_{\tau}(y_t - \beta^{\star\prime}\check{X}_{t-1}), \tag{14}$$

where $\Psi_{\tau}(u) := (\tau - \mathbf{1}(u < 0))$ for the given $\tau \in (0,1)$. In the scalar regressor case in which $n = p = 1$, $u_{xt} \sim mds(0, \Sigma_{xx})$, $u_{at} \sim mds(0, \Sigma_{aa})$, and $C \neq 0$, then the multivariate LSTUR model reduces to a scalar case. We define

$$\Sigma_{\Psi_{\tau}|(x,a)} := \tau(1-\tau) - \Sigma'_{\Psi_{\tau}(x,a)}\Sigma^{-1}_{(x,a)}\Sigma_{\Psi_{\tau}(x,a)}, \tag{15}$$

where $\Sigma_{\Psi_\tau(x,a)} = \left[\begin{array}{cc} \Sigma_{\Psi_\tau x} & \Sigma_{\Psi_\tau a} \end{array}\right]'$, $\Sigma_{(x,a)} = \left[\begin{array}{cc} \Sigma_{xx} & \Sigma_{xa} \\ \Sigma_{ax} & \Sigma_{aa} \end{array}\right]$, $\Sigma_{\Psi_\tau x} = \Sigma_{x\Psi_\tau} = \mathbb{E}\left(\Psi_\tau\left(u_{0t\tau}\right)u_{xt}\right)$, $\Sigma_{\Psi_\tau a} = \Sigma_{a\Psi_\tau} = \mathbb{E}\left(\Psi_\tau\left(u_{0t\tau}\right)u_{at}\right)$, and $\Sigma_{ax} = \Sigma_{xa} = \mathbb{E}\left(u_{xt}u_{at}\right)$. We employ the orthogonal decomposition of a Brownian motion (Phillips, 1989)

$$dB_{\Psi_\tau}(r) = dB_{\Psi_\tau|(x,a)}(r) + \Sigma'_{\Psi_\tau(x,a)}\Sigma^{-1}_{(x,a)}\left[\begin{array}{c} dB_x(r) \\ dB_a(r) \end{array}\right], \tag{16}$$

and set $\overline{G}_{a,c} := G_{a,c}(r) - \int_0^1 G_{a,c}(s)ds$. Using the estimated standard error $s.e.\left(\widehat{\beta}^{QR}_{1,\tau}\right) := [\tau(1-\tau)]^{\frac{1}{2}}\,\widehat{p}_{u0\tau}(0)^{-1}\left\{\sum_{t=1}^T (x^\mu_{t-1})^2\right\}^{-\frac{1}{2}}$ where $x^\mu_{t-1} := x_{t-1} - \frac{1}{T}\sum_{t=1}^T x_{t-1}$, and a consistent nonparametric kernel density estimate $\widehat{p}_{u0\tau}(0)$ for $p_{u0\tau}(0)$, the t-ratio test for a given τ under the null hypothesis $\mathcal{H}_0 : \beta_{1,\tau} = 0$ satisfies

$$\begin{aligned} t_{\widehat{\beta}^{QR}_{1,\tau}} = \frac{(\widehat{\beta}^{QR}_{1,\tau} - \beta_{1,\tau})}{s.e.\left(\widehat{\beta}^{QR}_{1,\tau}\right)} &\rightsquigarrow \left[1 - \frac{\Sigma'_{\Psi_\tau(x,a)}\Sigma^{-1}_{(x,a)}\Sigma_{\Psi_\tau(x,a)}}{\tau(1-\tau)}\right]^{\frac{1}{2}}\mathcal{N}(0,1) \\ &\quad + \left[\frac{\Sigma'_{\Psi_\tau(x,a)}\Sigma^{-1}_{(x,a)}}{\sqrt{\tau(1-\tau)}}\right]\frac{\int_0^1 \overline{G}_{a,c}(r)\left[\begin{array}{c} dB_x(r) \\ dB_a(r) \end{array}\right]}{\left[\int_0^1 \overline{G}^2_{a,c}(r)dr\right]^{\frac{1}{2}}}, \\ &= \left[1 - \lambda_1(\tau)\right]^{\frac{1}{2}} Z_{\mathcal{N}} + \left[\lambda_2(\tau)\right]\cdot\eta_{LP}(a,c), \end{aligned} \tag{17}$$

where $Z_{\mathcal{N}} := \mathcal{N}(0,1)$ and

$$\lambda_1(\tau) := \frac{\Sigma'_{\Psi_\tau(x,a)}\Sigma^{-1}_{(x,a)}\Sigma_{\Psi_\tau(x,a)}}{\tau(1-\tau)},\;\; \lambda_2(\tau) := \frac{\Sigma'_{\Psi_\tau(x,a)}\Sigma^{-1}_{(x,a)}}{\sqrt{\tau(1-\tau)}},$$

$$\eta_{LP}(a,c) := \left[\frac{\int_0^1 \overline{G}_{a,c}(r)dB_x(r)}{\left(\int_0^1 \overline{G}^2_{a,c}(r)dr\right)^{1/2}},\; \frac{\int_0^1 \overline{G}_{a,c}(r)dB_a(r)}{\left(\int_0^1 \overline{G}^2_{a,c}(r)dr\right)^{1/2}}\right]'.$$

The non-zero factors $\lambda_1(\tau)$ and $\lambda_2(\tau)$ in the limit expression (17) reveal the presence of size distortion in the usual t-ratio statistic and determine its magnitude.

3 Limit Theory for IVX Filtering

The intuition behind IVX instrumentation is to filter persistent data on endogenous regressors x_t to generate an instrument $\widetilde{z}_t$ that has the appealing property of asymptotic orthogonality to structural equation errors while retaining asymptotic relevance for the regressors. The generation process involves a simple reproductive mechanism

$$\widetilde{z}_t = F\widetilde{z}_{t-1} + \Delta x_t,$$

so that the time series innovations $\{\Delta x_t\}$ in an endogenous regressor are passed through an autoregressive filter to produce an instrument $\widetilde{z}_t$ using some suitable autoregressive coefficient matrix F. When $F = \mathbf{0}_{n\times n}$, then $\widetilde{z}_t = \Delta x_t$, and the first-order difference operator is used to remove distortions at the cost of substantial information loss. When $F = I_n$, then $\widetilde{z}_t = x_t$ and IVX reduces to OLS and standard quantile regression estimations with bias and size distortions as discussed above. The key idea in IVX is the use of a coefficient matrix F to generate instruments $\widetilde{z}_t$ with properties that are intermediate in persistence between the first-differenced data and the level data.

A simple way to achieve the advantageous intermediate form suggested in Phillips and Magdalinos (2009) is to select F so that $\widetilde{z}_t$ is a mildly integrated (MIR) process. Setting $F = R_{Tz}$ where $R_{Tz} = I_n + C_z/T^\gamma$, we have

$$\widetilde{z}_t = R_{Tz}\widetilde{z}_{t-1} + \Delta x_t,\ R_{Tz} = I_n + \frac{C_z}{T^\gamma}, \tag{18}$$

in which $\gamma \in (0,1)$, $C_z = c_z I_n$, $c_z < 0$, $\widetilde{z}_0 = \mathbf{0}_{n\times 1}$. With this construction, the coefficient matrix R_{Tz} is diagonal with entries that lie between zero and unity but which move close to unity for large sample sizes. The central advantage of IVX estimation, including both QR-IVX and LHIVX, is that robust pivotal inference is achieved for unit root and local unit root regressors, as well as regressors with mixed properties, thereby covering a broad class of cointegrating regression and predictive regression models and avoiding the size distortion of standard methods of inference.

These robustness properties are now shown to extend to cases of STUR and LSTUR regressors, as well as short-horizon and long-horizon predictive regression models, validating inference and providing a convenient tool for empirical work in predictive regression with a more extensive class of endogenous regressors. The presence of stochastic departures from unit roots in STUR and LSTUR systems introduces nonlinearities and additional sources of endogeneity in these models that complicate the limit theory. But as in simpler LUR regressor cases the asymptotics continue to have the advantage of mixed normality and pivotal test criteria, which facilitate inference.

3.1 IVX in short-horizon predictive regression

In short-horizon predictive regressions, IVX instruments are constructed using the observed data $\{x_t\}_{t=1}^T$ by means of the transform

$$\widetilde{z}_t = \sum_{j=1}^{t} R_{Tz}^{t-j}\Delta x_j, \tag{19}$$

where $R_{Tz} = I_n + \frac{C_z}{T^\gamma}$, $\gamma \in (0,1)$, $C_z = c_z I_n$ and $c_z < 0$. Some common settings for the parameters in (19) are $c_z = -1, \gamma = 0.95$ (Kostakis et al., 2015)[3], and $c_z = -5$ (Phillips and Lee, 2016). The transform (19) corresponds to the vector autoregression $\widetilde{z}_t = R_z\widetilde{z}_{t-1} + \Delta x_t$

[3]There are several reasons for selecting $c_z = -1$. This choice ensures that the properties of the mildly integrated instrument are determined by the localizing rate parameter γ alone, which is parsimonious and takes account of the fact that the dual parameters (c_z, γ) are not jointly identified in observations of the IVX instrument. Further, attempts to estimate c_z using observations lead to the pseudo-true value limit $c_z = -1$ in mildly integrated data (Phillips, 2021).

calculated with the observed differences $\{\Delta x_t\}$ as innovations from a zero initialization $\widetilde{z}_0 = \mathbf{0}_{n\times 1}$. This construction is designed as a feasible version of the latent instrument given by the transform $z_t = \sum_{j=1}^{t} R_{Tz}^{t-j} u_{xj}$ that depends on the innovations $\{u_{xt}\}$ that are unobserved except for the special case of a pure unit root process regressor x_t for which $\Delta x_t = u_{xt}$. The precise relationship between $\widetilde{z}_t$ and z_t depends on the generating mechanism of x_t.

The short-horizon IVX estimate of β_1 is obtained by the standard IV regression giving

$$\widehat{\beta}_1^{IVX} = (\widetilde{Z}'_{-1}\underline{X}_{-1})^{-1}(\widetilde{Z}'_{-1}\underline{y}), \tag{20}$$

where the instrument matrix $\widetilde{Z}_{-1} = [\widetilde{z}_0, ..., \widetilde{z}_{T-1}]'$. Above and in what follows, it is convenient to use notations to signify mean deviations and means. Demeaned variables are represented as $\underline{X}_{-1} := [\underline{x}_0, ..., \underline{x}_{T-1}]'$, $\underline{y} := \left[\underline{y}_1, ..., \underline{y}_T\right]'$, $\underline{u}_0 := [\underline{u}_{01}, ..., \underline{u}_{0T}]'$. The 'underline' signifies deviations from means, e.g. $\underline{y}_t := y_t - \overline{y}$, $\underline{x}_{t-1} := x_{t-1} - \overline{x}_{-1}$, $\underline{u}_{0t} := u_{0t} - \overline{u}_0$ and the 'overline' denotes sample means, e.g. $\overline{y} := \frac{1}{T}\sum_{t=1}^{T} y_t$, $\overline{x}_{-1} := \frac{1}{T}\sum_{t=1}^{T} x_{t-1}$, and $\overline{u}_0 := \frac{1}{T}\sum_{t=1}^{T} u_{0t}$. The short-horizon QR-IVX estimate is obtained by extremum estimation

$$\widehat{\beta}_{1,\tau}^{QRIVX} := \arg\min_{\beta_{1,\tau}^{\star}} \frac{1}{2}\left(\sum_{t=1}^{T} m_t\left(\beta_{1,\tau}^{\star}\right)\right)'\left(\sum_{t=1}^{T} m_t\left(\beta_{1,\tau}^{\star}\right)\right), \tag{21}$$

with $m_t\left(\beta_{1,\tau}^{\star}\right) = \widetilde{z}_{t-1}\left(\tau - \mathbf{1}\left(y_{t\tau} \leq \widehat{\beta}_{0,\tau} + \beta_{1,\tau}^{\star\,\prime} x_{t-1}\right)\right)$ where $\mathbf{1}(\cdot)$ is the indicator function and $\widehat{\beta}_{0,\tau}$ is a preliminary QR estimator for $\beta_{0,\tau}$ that optimizes (14).

The asymptotic properties of the normalized short-horizon IVX and QR-IVX estimates are provided in this Section 3. The asymptotic development shows that the effect of estimating the intercept (β_0 and $\beta_{0,\tau}$) matters only in the limit distribution of the normalized denominator $\widetilde{Z}'_{-1}\underline{X}_{-1}$ in the normalized IVX estimator. The normalized numerators in both estimates (IVX and QR-IVX) converge weakly to Gaussian processes centered at the origin and are independent of their corresponding signal matrices. Independence between the denominator (stochastic variance) and numerator successfully removes the size distortion induced by endogeneity and assists in generating a mixed Gaussian limit distribution.

Asymptotic analysis depends on the relationship between the IVX instrument $\widetilde{z}_t$ and the latent instrument z_t that involves the unobserved innovations u_{xt}. The relationship differs between the two generating mechanisms for the regressors x_t. In the LSTUR case we have the decomposition $\Delta x_j = u_{xj} + \frac{a' u_{aj}}{\sqrt{T}} x_{j-1} + \frac{C}{T} x_{j-1} + \frac{(a' u_{aj})^2}{2T} x_{j-1}$, which specializes in the STUR case where $C = \mathbf{0}_{n\times n}$ to $\Delta x_j = u_{xj} + \frac{a' u_{aj}}{\sqrt{T}} x_{j-1} + \frac{(a' u_{aj})^2}{2T} x_{j-1}$. The dynamics of the IVX instrument $\widetilde{z}_{t-1}$ in the LSTUR case then follow from the moving average representation

$$\begin{aligned}\widetilde{z}_{t-1} =& \sum_{j=1}^{t-1} R_{Tz}^{t-j-1}\left(u_{xj} + \frac{C}{T} x_{j-1} + \frac{a' u_{aj}}{\sqrt{T}} x_{j-1} + \frac{(a' u_{aj})^2}{2T} x_{j-1}\right) \\ =& z_{t-1} + \frac{C}{T}\eta_{T,t-1}^{(1)} + \frac{1}{\sqrt{T}}\eta_{T,t-1}^{(2)} + \frac{1}{2T}\eta_{T,t-1}^{(3)},\end{aligned} \tag{22}$$

where

$$\eta_{T,t-1}^{(1)} := \sum_{j=1}^{t-1} R_{Tz}^{t-j-1} x_{j-1},$$

$$\eta_{T,t-1}^{(2)} := \sum_{j=1}^{t-1} R_{Tz}^{t-j-1} \left(a' u_{aj}\right) x_{j-1},$$

$$\eta_{T,t-1}^{(3)} := \sum_{j=1}^{t-1} R_{Tz}^{t-j-1} \left(a' u_{aj}\right)^2 x_{j-1},$$

while in the STUR case $\widetilde{z}_{t-1} = z_{t-1} + \frac{1}{\sqrt{T}}\eta_{T,t-1}^{(2)} + \frac{1}{2T}\eta_{T,t-1}^{(3)}$.

As is apparent in (22), the IVX remainder terms measuring the difference between the observed and latent IVX instruments take a much more complex form than in LUR and unit root (UR) cases. Correspondingly, these remainders have a considerable impact on the limit theory. Under the case of LUR and MIR regressors, Phillips and Magdalinos (2009) gave the following decomposition for the IVX instrument $\widetilde{z}_{t-1}$

$$\widetilde{z}_{t-1} = \sum_{j=1}^{t-1} R_{Tz}^{t-1-j} \left(u_{xj} + \frac{C}{T^{\alpha}} x_{j-1}\right) = z_{t-1} + \frac{C}{T^{\alpha}} \eta_{T,t-1}^{(1)},$$

where $\alpha \in (0,1]$ and $\eta_{T,t-1}^{(1)}$ is defined in (22). It turns out that in the numerator of the IVX estimator $\widehat{\beta}_1^{IVX}$, the latent instrument z_{t-1} itself asymptotically dominates the remainder term $\frac{C}{T^{\alpha}}\eta_{T,t-1}^{(1)}$, whereas in the denominator both terms have the same order (Kostakis et al., 2015; Phillips and Magdalinos, 2009). When mildly explosive root (MER hereafter) regressors (Phillips and Lee, 2016) are considered, the IVX procedure still successfully removes the effects of endogeneity, but the asymptotic theory for MER regressors is different: the IVX remainder term $\frac{C}{T^{\alpha}}\eta_{T,t-1}^{(1)}$ dominates all other terms in the limit theory. These cases already reveal how the precise form of local persistence in the regressors plays a role in determining the asymptotics.

Compared to the LUR, MER and, MIR cases, the asymptotic behavior of IVX with STUR and LSTUR regressors is novel. Notably, in the numerator of the IVX (and the asymptotic form of the QR-IVX) estimates, terms containing both the components $\frac{1}{\sqrt{T}}\eta_{T,t-1}^{(2)}$ and z_{t-1} of (22) dominate. In the denominator of the IVX estimator all component moment matrices have the same stochastic order, while in the QR-IVX denominator the demeaned term asymptotically diminishes due to the $\sqrt{T}$-consistency of the preliminary QR estimate $\widehat{\beta}_{0,\tau}$. These features produce additional terms that figure in the asymptotics and considerable new complexities in the derivations, leading to very different limit theory from existing results when stochastic unit root regressors are present in the model. In spite of these complexities, the process of IVX instrumentation is shown to be remarkably robust. In particular, the approach still succeeds in producing mixed normal limit theory, which provides a mechanism for establishing pivotal limit theory for inference. We collect the relevant results for the estimators in the following theorem.

Theorem 3.1 *Suppose Assumptions 1 and 2 hold for (1) and (2). As $T \to \infty$, both $\widehat{\beta}_1^{IVX}$ and $\widehat{\beta}_{1,\tau}^{QRIVX}$ are asymptotically mixed normal with distributions*

$$T^{\frac{1+\gamma}{2}}(\widehat{\beta}_1^{IVX} - \beta_1) \sim_a \mathcal{MN}\left(\mathbf{0}_{n\times 1},\ \mathbb{V}^{IVX}\right),$$
$$T^{\frac{1+\gamma}{2}}(\widehat{\beta}_{1,\tau}^{QRIVX} - \beta_{1,\tau}) \sim_a \mathcal{MN}\left(\mathbf{0}_{n\times 1},\ \mathbb{V}^{QRIVX}\right),$$

where the covariance matrices $\mathbb{V}^{IVX}$ and $\mathbb{V}^{QRIVX}$ are defined in (58) and (59).

Remark 3.1 *In Phillips and Magdalinos (2009), IVX instrumentation was developed for models with LUR regressors and use of IVX was shown to produce mixed normal limit theory, thereby eliminating endogeneity effects arising from the joint dependence of the innovations u_{xt} and u_{0t}. In the present paper, the regressors are STUR/LSTUR processes with dual sources of endogeneity driven by u_{xt} and u_{at}. The additional source of endogeneity from u_{at} does not change the convergence rate of IVX estimation as partial sums of u_{at} still enjoy the same $\sqrt{T}$-rate. But the existence of u_{at} dramatically impacts the limit distribution of IVX and the asymptotic variance matrix. The presence of u_{at} also produces added complexity in the proofs and new methods for obtaining the limit theory are required. In particular, as will become clear, estimation involves weighted sums of both u_{xt} and u_{at} and the asymptotics rely on Réyni-mixing (stable convergence) techniques to deliver limiting mixed normality with an asymptotic variance that embodies the effects of both components in contrast to the IVX limit theory under LUR regressors. These asymptotics enable the establishment of pivotal limit theory for IVX based testing, as is now described.*

We consider both IVX-Wald and QR-IVX-Wald tests based on the corresponding IVX mean and QR regressions. The mixed normal estimation limit theory leads to pivotal asymptotic tests with chi-square limit distributions. For example, to test hypotheses of block restrictions on the cointegration coefficients such as $\mathcal{H}_0$: $\breve{H}\beta_1 = h^0$ or $\mathcal{H}_{0,\tau}$: $\breve{H}\beta_{1,\tau} = h_\tau^0$ (where $\breve{H}$ is a known $\ell \times n$ matrix of full rank ℓ and both h^0 and h_τ^0 are known ℓ-dimensional vectors), we simply use the following IVX-Wald and QR-IVX-Wald test statistics (W_{IVX} and $W_{QRIVX,\tau}$) constructed in the usual manner, viz.,

$$W_{IVX} := \frac{1}{\widehat{\sigma}_{00}}\left(\breve{H}\widehat{\beta}_1^{IVX} - h^0\right)'\left(\breve{H}\left(\underline{X}'_{-1}P_{\widetilde{Z}}\underline{X}_{-1}\right)^{-1}\breve{H}'\right)^{-1}\left(\breve{H}\widehat{\beta}_1^{IVX} - h^0\right), \tag{23}$$

where $\underline{X}'_{-1}P_{\widetilde{Z}}\underline{X}_{-1} := \left(\sum_{t=1}^T \underline{x}_{t-1}\widetilde{z}'_{t-1}\right)\left(\sum_{t=1}^T \widetilde{z}_{t-1}\widetilde{z}'_{t-1}\right)^{-1}\left(\sum_{t=1}^T \widetilde{z}_{t-1}\underline{x}'_{t-1}\right)$ and $\widehat{\sigma}_{00}$ is any consistent estimator for σ_{00}; and

$$W_{QRIVX,\tau} := \frac{\widehat{p}_{u0\tau}^2(0)}{\tau(1-\tau)}\left(\breve{H}\widehat{\beta}_{1,\tau}^{QRIVX} - h_\tau^0\right)'\left(\breve{H}\left(X'_{-1}P_{\widetilde{Z}}X_{-1}\right)^{-1}\breve{H}'\right)^{-1}\left(\breve{H}\widehat{\beta}_{1,\tau}^{QRIVX} - h_\tau^0\right), \tag{24}$$

where $X'_{-1}P_{\widetilde{Z}}X_{-1} := \left(\sum_{t=1}^T x_{t-1}\widetilde{z}'_{t-1}\right)\left(\sum_{t=1}^T \widetilde{z}_{t-1}\widetilde{z}'_{t-1}\right)^{-1}\left(\sum_{t=1}^T \widetilde{z}_{t-1}x'_{t-1}\right)$ and $\widehat{p}_{u0\tau}(0)$ is any consistent nonparametric estimate for $p_{u0\tau}(0)$. The limit distribution theory for both statistics is standard.

Theorem 3.2 *(i) Suppose Assumptions 1 and 2 hold. As $T \to \infty$,*

$$W_{IVX} \rightsquigarrow \chi^2(\ell),$$

under $\mathcal{H}_0$: $\breve{H}\beta_1 = h^0$, where $\breve{H}$ is an $\ell \times n$ matrix of full rank ℓ.

(ii) Suppose Assumptions 1 and 2 hold. As $T \to \infty$ for given τ,

$$W_{QRIVX,\tau} \rightsquigarrow \chi^2(\ell),$$

under $\mathcal{H}_{0,\tau}$: $\breve{H}\beta_{1,\tau} = h^0_\tau$, where $\breve{H}$ is an $\ell \times n$ matrix of full rank ℓ.

Beyond conditional heteroskedasticity, serial correlation can also be embedded into the IVX framework. A two-stage approach called IVX-AR (Yang et al., 2020) combines a Cochrane-Orcutt transformation with the construction of the IVX-Wald statistic to reduce the finite sample size distortion in testing predictability. Suppose the error term u_{0t} of (1) is serially correlated and the regressor in (2) has LSTUR/STUR persistence. In that case, the IVX-AR-Wald test has a robust pivotal distribution under the null hypothesis and discriminatory power under local alternatives. To proceed with the theory we introduce the following condition.

Assumption 3 *The error term $\{u_{0t}\}_{1\le t\le T}$ is governed by an $AR(m) + GARCH(q_1, q_2)$ process*

$$\begin{aligned} u_{0t} =& \rho_1 u_{0,t-1} + \rho_2 u_{0,t-2} + \ldots + \rho_m u_{0,t-m} + v_t, \\ v_t =& H_t^{\frac{1}{2}} \epsilon_{0t}, \quad H_t = \varphi_0 + \sum_{l=1}^{q_1} \varphi_{1l} v_{t-l}^2 + \sum_{k=1}^{q_2} \varphi_{2k} H_{t-k}, \end{aligned} \tag{25}$$

where $\{\epsilon_{0t}\}_{1\le t\le T}$ is the martingale difference sequence in Assumption 1 with unit variance, φ_0 is a positive constant value, φ_{1l} and φ_{2k} are nonnegative for all l, k, and $\sum_{l=1}^{q_1} \varphi_{1l} + \sum_{k=1}^{q_2} \varphi_{2k} < 1$.

The following discussion is conducted under Assumptions 1 (i) (iii), and 3. To perform IVX-AR estimation, the IVX method is first applied to (1) and (2) to derive the preliminary estimates $(\widehat{\rho}_j)_{1\le j\le m}$ of the AR coefficients. Feasibile AR transformations are then implemented to generate $\widehat{\underline{y}}_t := \underline{y}_t - \sum_{j=1}^m \widehat{\rho}_j \underline{y}_{t-j}$, $\widehat{\underline{x}}_{t-1} := \underline{x}_{t-1} - \sum_{j=1}^m \widehat{\rho}_j \underline{x}_{t-1-j}$, $\widehat{\tilde{z}}_{t-1} := \tilde{z}_{t-1} - \sum_{j=1}^m \widehat{\rho}_j \tilde{z}_{t-1-j}$, and the IVX-AR coefficient estimates are obtained as follows

$$\widehat{\beta}_1^{IVXAR} := \left(\sum_{t=1}^T \widehat{\tilde{z}}_{t-1} \widehat{\underline{x}}'_{t-1} \right)^{-1} \left(\sum_{t=1}^T \widehat{\tilde{z}}_{t-1} \widehat{\underline{y}}_t \right).$$

Using this estimate, an IVX-AR-Wald test of the hypothesis $\mathcal{H}_0$: $\breve{H}\beta_1 = h^0$ is constructed as

$$W_{IVXAR} := \frac{1}{\widehat{\sigma}_{vv}} \left(\breve{H}\widehat{\beta}_1^{IVXAR} - h^0 \right)' \left\{ \breve{H} \left(\widehat{\underline{X}}'_{-1} P_{\widehat{\tilde{Z}}} \widehat{\underline{X}}_{-1} \right)^{-1} \breve{H}' \right\}^{-1} \left(\breve{H}\widehat{\beta}_1^{IVXAR} - h^0 \right), \tag{26}$$

where $\widehat{\check{X}}'_{-1} P_{\widehat{\check{Z}}} \widehat{\check{X}}_{-1} := \left(\sum_{t=1}^{T} \widehat{\check{x}}_{t-1} \widehat{\check{z}}'_{t-1}\right) \left(\sum_{t=1}^{T} \widehat{\check{z}}_{t-1} \widehat{\check{z}}'_{t-1}\right)^{-1} \left(\sum_{t=1}^{T} \widehat{\check{z}}_{t-1} \widehat{\check{x}}'_{t-1}\right)$ and $\widehat{\sigma}_{vv}$ is any consistent estimate for σ_{vv}, the instantaneous variance of v_t. The limit distribution of the statistic (26) is pivotal with $W_{IVXAR} \rightsquigarrow \chi^2(\ell)$ under the null hypothesis $\mathcal{H}_0$: $\check{H}\beta_1 = h^0$ and provides robust inference in the case of serially correlated errors. So, the IVX-AR methodology retains the advantage of the usual IVX limit theory and is easy to implement with pivotal critical values in the nonstationary prediction model with STUR/LSTUR regressors.

3.2 IVX in long-horizon predictive regression

To simplify discussion it is convenient to exclude the intercept β_0 and use the data generating process (DGP)

$$y_t = \beta_1' x_{t-1} + u_{0t}, \tag{27}$$

analogous to the demeaned model in the short-horizion prediction model. Rather than using (1) or (27), we estimate the long horizon prediction model formulated in (13) and apply the long-run variant of the IVX approach, called LHIVX, to the model with LSTUR and STUR regressors. The LHIVX estimate $\widehat{B}_1^{LHIVX}$ is given by

$$\widehat{B}_1^{LHIVX} - B_1 = \left(\sum_{t=1}^{T-k} u_{0,t+k} \widetilde{z}_t^{k\,\prime}\right) \left(\sum_{t=1}^{T-k} x_t^k \widetilde{z}_t^{k\,\prime}\right)^{-1}. \tag{28}$$

The LHIVX instrument $\widetilde{z}_t^k$ in (28) is constructed as $\widetilde{z}_t^k := \sum_{j=1}^{k} \widetilde{z}_{t+j-1}$, and we employ the notations $z_t^k := \sum_{j=1}^{k} z_{t+j-1}$, $\eta_{T,t}^{1,k} := \sum_{j=1}^{k} \eta_{T,t+j-1}^{(1)}$, $\eta_{T,t}^{2,k} := \sum_{j=1}^{k} \eta_{T,t+j-1}^{(2)}$, and $\eta_{T,t}^{3,k} := \sum_{j=1}^{k} \eta_{T,t+j-1}^{(3)}$, in which $\widetilde{z}_{t+j-1}$, z_{t+j-1}, $\eta_{T,t+j-1}^{(1)}$, $\eta_{T,t+j-1}^{(2)}$, and $\eta_{T,t+j-1}^{(3)}$ are defined in (22). Analogous to (22), the following decomposition now holds

$$\widetilde{z}_t^k := \begin{cases} z_t^k + \frac{C}{T}\eta_{T,t}^{1,k} + \frac{1}{\sqrt{T}}\eta_{T,t}^{2,k} + \frac{1}{2T}\eta_{T,t}^{3,k} & \text{with} \quad \text{LSTUR}, \\ z_t^k + \frac{1}{\sqrt{T}}\eta_{T,t}^{2,k} + \frac{1}{2T}\eta_{T,t}^{3,k} & \text{with} \quad \text{STUR}. \end{cases} \tag{29}$$

In developing the limit theory we set the following condition on the time horizon k and MIR rate parameter γ as follows

$$\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0. \tag{30}$$

This restriction requires the time horizon k to diverge slower than T but faster than T^γ.

The limit theory for the STUR/LSTUR case differs from the LUR, MIR and MER cases. In the LUR and MIR cases, the behavior of the numerator in the matrix quotient (28) of the LHIVX estimator is determined by the term $\sum_{t=1}^{T-k} u_{0,t+k} z_t^{k\,\prime}$, whereas in the MER case the term $\sum_{t=1}^{T-k} u_{0,t+k} \eta_{T,t}^{1,k\,\prime}$ dominates the other term of the numerator. But in the case of STUR/LSTUR regressors, the sample covariances $\sum_{t=1}^{T-k} u_{0,t+k} z_t^{k\,\prime}$ and

$\sum_{t=1}^{T-k} u_{0,t+k}\eta_{T,t}^{2,k\,\prime}$ dominate other terms and upon normalizations generate a joint asymptotic mixed normality via the following approximation:

$$\frac{1}{T^{\frac{1}{2}+\gamma}\sqrt{k}}\sum_{t=1}^{T-k} u_{0,t+k}\tilde{z}_t^{k\,\prime} = \frac{1}{T^{\frac{1}{2}+\gamma}\sqrt{k}}\sum_{t=1}^{T-k} u_{0,t+k}z_t^{k\,\prime} + \frac{1}{T^{1+\gamma}\sqrt{k}}\sum_{t=1}^{T-k} u_{0,t+k}\eta_{T,t}^{2,k\,\prime} + o_p(1). \quad (31)$$

Similarly, the limit theory for the LHIVX denominator in (28) is obtained upon normalization using

$$\frac{1}{T^{1+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k\tilde{z}_t^{k\,\prime} \sim_a \frac{1}{T^{1+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k z_t^{k\,\prime} + \frac{1}{T^{\frac{3}{2}+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k\eta_{T,t}^{2,k\,\prime} + \frac{1}{T^{2+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k\eta_{T,t}^{1,k\,\prime}C$$
$$+ \frac{1}{2T^{2+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k\eta_{T,t}^{3,k\,\prime}, \quad (32)$$

where now all four terms contribute to the asymptotics. Combining the components in both (31) and (32) delivers the limit distribution of the LHIVX estimate given in the following result.

Theorem 3.3 *Suppose Assumptions 1 and 2 hold. Under the rate condition (30), as $T \to \infty$, the LHIVX estimator is asymptotically mixed normal*

$$\sqrt{T}k^{\frac{3}{2}}\left(\widehat{B}_1^{LHIVX} - B_1\right)' \sim_a \mathcal{MN}\left(\mathbf{0}_{n\times 1},\ \mathbb{V}^{LHIVX}\right),$$

where the variance matrix $\mathbb{V}^{LHIVX}$ is defined in (60).

The LHIVX estimator is consistent, asymptotically unbiased, and has a mixed normal limit distribution. The distribution is nonpivotal as both the unknown localizing coefficients a and C occur in the variance matrix $\mathbb{V}^{LHIVX}$. Nonetheless, given the asymptotic mixed normality of the estimates, the use of a consistent variance matrix estimate enables robust inference and the LHIVX-Wald test has the usual standard chi-square limit distribution. This pivotal limit theory feature, as in the short-horizon case considered earlier, demonstrates the significant advantage in practical applications of IVX procedures over simulation methods.

Theorem 3.4 *Suppose Assumptions 1 and 2 hold. Under $\mathcal{H}_0(k)$: $\breve{H}B_1' = h^0$, where $\breve{H}$ is a known $\ell \times n$ matrix of rank ℓ and h^0 is a known ℓ-vector,*

$$W_{LH} := \left(\breve{H}\widehat{B}_1^{LHIVX\,\prime} - h^0\right)'\left[\breve{H}\left\{\left(X^{k\prime}P_{\tilde{Z}}^k X^k\right)^{-1}\widehat{\sigma}_{00}\right\}\breve{H}'\right]^{-1}\left(\breve{H}\widehat{B}_1^{LHIVX\,\prime} - h^0\right) \rightsquigarrow \chi^2(\ell), \quad (33)$$

where

$$X^{k\,\prime}P_{\tilde{Z}}^k X^k = \left(\sum_{t=1}^{T-k} x_t^k\tilde{z}_t^{k\,\prime}\right)\left(\sum_{t=1}^{T-k}\tilde{z}_t^k\tilde{z}_t^{k\,\prime}\right)^{-1}\left(\sum_{t=1}^{T-k}\tilde{z}_t^k x_t^{k\,\prime}\right), \quad (34)$$

and $\widehat{\sigma}_{00}$ is any consistent estimator for σ_{00}.

Remark 3.2 *Long horizion predictability can also be tested by regressing the k-period return* $y_t^k \left(:= \sum_{i=1}^k y_{t+i-1}\right)$ *on a constant intercept* μ_f *and the STUR/LSTUR regressors* x_{t-1} *in the model*

$$\begin{aligned} y_t^k &= \mu_f + \Phi' x_{t-1} + u_{0t}^k, \\ y_t^k &:= \sum_{i=1}^k y_{t+i-1}, \ u_{0t}^k := \sum_{j=1}^k u_{0,t+j-1}, \end{aligned} \tag{35}$$

with null hypothesis $\mathcal{H}_0^\star(k)$: $\Phi = \Phi^0$ *and alternative* $\mathcal{H}_1^\star(k)$: $\Phi \neq \Phi^0$. *The IVX instrument can accommodate (35) and deliver a pivotal test for long-horizon predictability. As the intercept* μ_f *is nonzero, the IVX estimate is based on the demeaned equation:*

$$\left(y_t^k - \overline{y}^k\right) = \Phi' \underline{x}_{t-1} + \left(u_{0t}^k - \overline{u}_0^k\right),$$

where $\overline{y}^k := \frac{1}{T-k+1} \sum_{t=1}^{T-k+1} y_t^k$, *and* $\overline{u}_0^k := \frac{1}{T-k+1} \sum_{t=1}^{T-k+1} u_{0t}^k$. *The corresponding IVX estimator of* Φ *is given by*

$$\widehat{\Phi}_{IVX}^k = \left(\sum_{t=1}^{T-k+1} \widetilde{z}_{t-1} \underline{x}_{t-1}^{k\,\prime}\right)^{-1} \left(\sum_{t=1}^{T-k+1} \widetilde{z}_{t-1} \underline{y}_t^k\right), \tag{36}$$

where $\underline{x}_{t-1}^k := x_{t-1}^k - \overline{x}_{-1}^k$, $x_{t-1}^k := \sum_{j=1}^k x_{t-2+j}$, *and* $\overline{x}_{-1}^k := \frac{1}{T-k+1} \sum_{t=1}^{T-k+1} x_{t-1}^k$.

Upon appropriate normalization and centering, the numerator matrix of (36) has the following decomposition:

$$\sum_{t=1}^{T-k+1} \left(u_{0t}^k - \overline{u}_0^k\right) \widetilde{z}_{t-1}' \sim_a \sum_{j=1}^{k-1} u_{0j} \sum_{t=1}^{j} \widetilde{z}_{t-1}' + \sum_{t=0}^{T-2k} u_{0,t+k} \widetilde{z}_t^{k\,\prime} + \sum_{j=0}^{k-1} u_{0,T-j} \sum_{t=0}^{j} \widetilde{z}_{T-k-t-2}' \tag{37}$$

$$\sim_a \sum_{t=0}^{T-2k} u_{0,t+k} \widetilde{z}_t^{k\,\prime}, \tag{38}$$

where (37) follows Kostakis et al. (2020, equation (72)) or Kostakis et al. (2018, equation (19)) and (38) follows Kostakis et al. (2020, equation (26)) or Kostakis et al. (2018, equation (19)). By a similar derivation of the LHIVX estimator, the numerator of (36) still has the mixed normal distribution and is asymptotically independent from the limit Gaussian variates of (8). Therefore, as $T \to \infty$, *the asymptotic unbiasedness and the mixed normality of* $\widehat{\Phi}_{IVX}^k$ *are retained.*

In consequence, when the rate condition $\frac{1}{k} + \frac{k}{T^\gamma} \to 0$ *holds, the corresponding long-horizon IVX-Wald test has the following pivotal chi-square limit theory*

$$\begin{aligned} W_{IVX}^k :=& \frac{1}{\widehat{\sigma}_{00}} \left[\widehat{\Phi}_{IVX}^k - \Phi^0\right]' \left[\left(\sum_{t=1}^{T-k+1} \underline{x}_{t-1}^k \widetilde{z}_{t-1}'\right) \left(\sum_{t=1}^{T-k+1} \widetilde{z}_{t-1}^k \widetilde{z}_{t-1}^{k\,\prime}\right)^{-1} \left(\sum_{t=1}^{T-k+1} \widetilde{z}_{t-1} \underline{x}_{t-1}^{k\,\prime}\right)\right]^{-1} \\ & \times \left[\widehat{\Phi}_{IVX}^k - \Phi^0\right] \rightsquigarrow \chi^2(n), \end{aligned} \tag{39}$$

under the null $\mathcal{H}_0^\star(k)$: $\Phi = \Phi^0$, and $\widetilde{z}_t^k$ follows the definition in (28). The limit theory of (39) shows that the IVX methodology retains validity in a long-horizon predictive regression model with STUR/LSTUR regressors, thereby justifying its usage in more general and empirically realistic nonstationary models than LUR settings.

4 IVX Regression with Mixed Roots

This section applies IVX procedures to a predictive regression model containing mixed roots and random departures from unity. Again, suitably defined Wald statistics have pivotal chi-square limit distributions under the null hypothesis of no predictability.

4.1 IVX regression with mixed roots in the short-horizon case

Let $n = 2$ and the predictive regressor x_{t-1} be a bivariate AR(1) process with mixed roots and stochastic departures from unity. For convenience, but with some abuse of notation, the model is written in the form

$$\begin{aligned} y_t =& \beta_0 + \beta_1' x_{t-1} + u_{0t},\ \beta_1' = [\beta_{11}, \beta_{12}],\ x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}, \\ x_t =& R_{Tt} x_{t-1} + u_{xt},\ R_{Tt} = \begin{bmatrix} \rho_{Tt} & 0 \\ 0 & \theta_T \end{bmatrix},\ u_{xt} = \begin{bmatrix} u_{x1t} \\ u_{x2t} \end{bmatrix}. \end{aligned} \tag{40}$$

This system allows for mixed roots and stochastic departures from a unit root with AR coefficients[4] defined as follows

$$\rho_{Tt} = 1 + \frac{c_1}{T^{\alpha_1}} + \frac{a_1 u_{at}}{\sqrt{T}} + \frac{(a_1 u_{at})^2}{2T}, \tag{41}$$

where

$$\begin{cases} c_1 \in (-\infty, \infty) & \text{if} \quad \alpha_1 = 1 \\ c_1 \in (-\infty, 0) & \text{if} \quad \alpha_1 \in [0,1) \end{cases}, \text{ and } a_1 \in (-\infty, +\infty),$$

and

$$\theta_T = 1 + \frac{c_2}{T^{\alpha_2}},\ \text{where } c_2 \in (0, \infty) \text{ and } \alpha_2 \in (0,1). \tag{42}$$

Accordingly, x_{1t} falls within one of the following specifications including unit root (UR), LUR, STUR, LSTUR, MIR or purely stationary regressors, whereas x_{2t} is a regressor with a mildly explosive root (MER). The error process $u_t = (u_{0t}, u_{xt}', u_{at}')'$, long-run variance and asymptotics are the same except that the subscripts $0, 1, 2$, and a here represent u_{0t}, u_{x1t}, u_{x2t} and u_{at}. The IVX instrument $\widetilde{z}_t$ is constructed in the same form as before with localizing coefficient matrix $C_z := \operatorname{diag}\{c_z, c_z\} < 0$, $\gamma \in (0,1)$, and the intercept-augmented instrument as $\widetilde{Z}_t := (1, \widetilde{z}_t')'$.

[4] When $\alpha_1 = 0$, we assume $a_1 = 0$.

Irrespective of the stochastic deviations from unity, the short-horizon IVX methods continue to provide pivotal tests for the multivariate predictive regression model under the null hypothesis. Suppose Assumptions 1 and 2 hold and assume that $\alpha_1 \in \left(\frac{1}{3}, 1\right)$, and $\gamma \in \left(\left(\alpha_2 \vee \frac{2}{3}\right),\ 1\right)$. Under the null hypothesis $\mathcal{H}_0 : \beta_1 = \beta_1^0$

$$\left(\widehat{\beta}_1^{IVX} - \beta_1^0\right)' \left(\underline{X}'_{-1} \check{P}_{\widetilde{Z}} \underline{X}_{-1}\right) \left(\widehat{\beta}_1^{IVX} - \beta_1^0\right) \rightsquigarrow \chi^2(2), \tag{43}$$

where the weight matrix is

$$\underline{X}'_{-1} \check{P}_{\widetilde{Z}} \underline{X}_{-1} := \left(\sum_{t=1}^{T} \underline{x}_{t-1} \widetilde{z}'_{t-1}\right) \left(\sum_{t=1}^{T} \widetilde{z}_{t-1} \widetilde{z}'_{t-1} \widehat{u}_{0t}^2\right)^{-1} \left(\sum_{t=1}^{T} \widetilde{z}_{t-1} \underline{x}'_{t-1}\right), \tag{44}$$

in which $\widehat{u}_{0t}$ is the least squares or IVX estimated equation residual. Similarly, under the null hypothesis $\mathcal{H}_{0,\tau} : \beta_{1,\tau} = \beta_{1,\tau}^0$ for any quantile $\tau \in (0,1)$ and imposing Assumptions 1 and 2 and the rate restrictions that $\alpha_1 \in \left(\frac{1}{3}, 1\right)$, and $\gamma \in \left(\left(\alpha_2 \vee \frac{2}{3}\right),\ 1\right)$, we have

$$\frac{\widehat{p}_{u0\tau(0)}^2}{\tau(1-\tau)} \left(\widehat{\beta}_{1,\tau}^{QRIVX} - \beta_{1,\tau}^0\right)' \left(X'_{-1} P_{\widetilde{Z}} X_{-1}\right) \left(\widehat{\beta}_{1,\tau}^{QRIVX} - \beta_{1,\tau}^0\right) \rightsquigarrow \chi^2(2), \tag{45}$$

where $\widehat{p}_{u0\tau}(0)$ is any consistent estimator for $p_{u0\tau}(0)$, and $X'_{-1} P_{\widetilde{Z}} X_{-1}$ is defined in (24).

The proofs of (43) and (45) rely on orthogonality conditions as in Lemmas 3.1 and 3.2 of Phillips and Lee (2016) that establish the mixed Gaussian distributions for the short-horizon IVX and QR-IVX estimates. In general, the short-horizon IVX approaches provide a unified theory of delivering pivotal chi-square limit theory in the presence of mixed roots and stochastic deviations from unity. In addition, the heteroskedasticity-robust estimator in (44) for the asymptotic variance matrix is embedded in the IVX-Wald test (43) for the short-horizon mean predictive regression model to address the effect of heteroskedasticity and retain an asymptotically pivotal test under stationary regressors.[5]

4.2 IVX regression with mixed roots in the long-horizon case

Similar to short-horizon predictive regressions with mixed roots and random deviations in (40), (41) and (42)[6], a corresponding long-horizon model can be estimated as

$$y_{t+k} = b_1' x_t^k + u_{0,t+k}, \text{ and } x_t^k = \sum_{j=1}^{k} x_{t+j-1}, \tag{46}$$

in which the zero-intercept case (i.e., $\beta_0 = b_0 = 0$) is considered. Correspondingly, the LHIVX estimation follows precisely as in (28). For the bivariate regressor case of (40) and using subscripts $0, 1, 2$, and a to signify u_{0t}, u_{x1t}, u_{x2t} and u_{at}, the LHIVX estimate can be further given in the partitioned form:

$$\widehat{b}_1^{LHIVX} - b_1 = \begin{bmatrix} \sum_{t=1}^{T-k} \widetilde{z}_{1t}^k x_{1t}^k & \sum_{t=1}^{T-k} \widetilde{z}_{1t}^k x_{2t}^k \\ \sum_{t=1}^{T-k} \widetilde{z}_{2t}^k x_{1t}^k & \sum_{t=1}^{T-k} \widetilde{z}_{2t}^k x_{2t}^k \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T-k} \widetilde{z}_{1t}^k u_{0,t+k} \\ \sum_{t=1}^{T-k} \widetilde{z}_{2t}^k u_{0,t+k} \end{bmatrix}.$$

[5] We thank a referee for indicating the need to deal with heteroskedasticity in the case where stationary time series regressors are included.

[6] We assume that $\alpha_1 \neq 0$ in the long-horizon predictive regression.

Suppose Assumptions 1 and 2 hold. Assume the rate condition $\frac{\sqrt{T}}{T^{\gamma}} + \frac{T^{\gamma}}{k} + \frac{k}{T} \to 0$, as before in (30), and $\alpha_1 \in (\nu, 1)$, $\alpha_2 \in (\gamma, 1-\gamma+\nu)$. Under the null hypothesis $\mathcal{H}_0^b(k)$: $b_1 = b_1^0$, the following pivotal χ^2 limit theory holds for the LHIVX-Wald statistic:

$$\frac{1}{\widehat{\sigma}_{00}}(\widehat{b}_1^{LHIVX} - b_1^0)'(X^{k\prime} P_{\widetilde{Z}}^k X^k)(\widehat{b}_1^{LHIVX} - b_1^0) \rightsquigarrow \chi^2(2), \tag{47}$$

for any consistent estimate $\widehat{\sigma}_{00}$ of σ_{00}. The notation $X^{k\prime} P_{\widetilde{Z}}^k X^k$ is defined in (34).

Extension of the theory to the case of long-horizon prediction with multiple regressors (i.e., $n \geq 3$) is possible. In such cases, persistent regressors x_t of dimension $n = n_1 + n_2$ can be modeled with autoregressive roots in the partitioned form

$$R_{Tt} = \begin{bmatrix} R_{1,Tt} & \mathbf{0}_{n_1 \times n_2} \\ \mathbf{0}_{n_2 \times n_1} & R_{2,Tt} \end{bmatrix}, \tag{48}$$

where the $n_1 \times n_1$ matrix $R_{1,Tt}$ has the roots of the UR, STUR, LUR, LSTUR, or MIR type, while the $n_2 \times n_2$ matrix $R_{2,Tt}$ has the roots of the MER type. In this multiple regressor case in which $n \geq 3$, the LHIVX procedure leads to an asymptotically pivotal $\chi^2(n)$ test analogous to the bivariate case of (47) under the corresponding null hypothesis.

5 Simulation Findings

This section reports size and power performances of the IVX-Wald, QR-IVX-Wald[7] , and LHIVX-Wald tests in both univariate and multivariate prediction models. The generated DGP follows (40) and (41).

5.1 IVX-Wald and QR-IVX-Wald testing

The IVX instruments are generated using the localizing coefficient matrix $C_z = -10I_n$ for the QR-IVX-Wald test, $C_z = -5I_n$ for the IVX-Wald test and the power coefficients $\gamma \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$. To explore local power performance we use two sets of alternatives

$$\mathcal{H}_{\beta_T}^{(1)} : \beta_T = \frac{\delta}{T} \quad \text{or} \quad \mathcal{H}_{\beta_T,\tau}^{(1)} : \beta_{T,\tau} = \frac{\delta_\tau}{T}, \tag{49}$$

or

$$\mathcal{H}_{\beta_T,\gamma}^{(2)} : \beta_{T,\gamma} = \frac{\delta}{T^{\frac{1+\gamma}{2}}} \quad \text{or} \quad \mathcal{H}_{\beta_T,\tau,\gamma}^{(2)} : \beta_{T,\tau,\gamma} = \frac{\delta_\tau}{T^{\frac{1+\gamma}{2}}}, \tag{50}$$

over a grid of integer values, $1 \leq \delta, \delta_\tau \leq 20$, where T is the sample size. For different values of IVX scaling parameters γ, the corresponding alternative hypotheses $\mathcal{H}_{\beta_T,\gamma}^{(2)}$ (or $\mathcal{H}_{\beta_T,\tau,\gamma}^{(2)}$) are employed. The number of replications is 1,000.

[7] We thank Ji-Hyung Lee for sharing the codes of the QR-IVX method. Numerical simulations of the QR-IVX-Wald test are conducted using his codes. Validity of the method follows from the same procedures as Proposition 3.2 in Lee (2016).

Test performance was investigated in the multiple regressor case with three types of regressors: (i) LSTUR and MIR; (ii) LSTUR and LUR; and (iii) LSTUR and MER. Case (ii) corresponds to the non-bubble STUR case with near-integrated components. Case (iii) captures potential market bubble conditions. Case (i) allows for conditions of market collapse. The model set up follows (40) with localizing scale and rate parameters

$$(c_1, a_1, c_2, \alpha_2) = \begin{cases} (-1, 1, -1, 0.75) & \text{for} \quad \text{Case (i)} \\ (-1, 1, 0, 0.75) & \text{for} \quad \text{Case (ii)} \\ (-1, 1, 0.2, 0.75) & \text{for} \quad \text{Case (iii)} \end{cases} . \tag{51}$$

The prediction error is given by $u_{0t} = \sqrt{2}\epsilon_{0t}$, the regressor errors by $u_{xt} = 0.5u_{x,t-1} + \epsilon_{xt}$, and the stochastic STUR variates by $u_{at} = 0.5u_{a,t-1} + \epsilon_{at}$, all driven by multivariate normal innovations $\epsilon_t = (\epsilon_{0t}, \epsilon_{x_1t}, \epsilon_{x_2t}, \epsilon_{at})'$ with variance matrix

$$\Sigma_\epsilon = \begin{bmatrix} 1 & -0.75 & -0.75 & -0.75 \\ -0.75 & 1 & 0.5 & 0.5 \\ -0.75 & 0.5 & 1 & 0.5 \\ -0.75 & 0.5 & 0.5 & 1 \end{bmatrix}. \tag{52}$$

[Insert Tables 1 and 2 here]

Tables 1 and 2 report size performance of the IVX-Wald and QR-IVX-Wald tests for cases (i)–(iii). The joint null hypothesis is $\mathcal{H}_0 : \beta_{11} = \beta_{12} = 0$ (or $\mathcal{H}_{0,\tau} : \beta_{11,\tau} = \beta_{12,\tau} = 0$) and sample size is $T = 200$ and 500. The results show that empirical size is well controlled around the nominal level when a small value of γ is used in (i) or (ii) and size distortion occurs only at extreme quantiles. Mild size distortion is evident for large values of γ around 0.8 or 0.9 in all quantiles of case (iii) and these are not entirely eliminated even with larger sample sizes[8]. The choice of γ values close to unity can therefore be troublesome in controlling size in STUR and LSTUR predictive regressions, an outcome that differs from the recommended setting $\gamma = 0.95$ suggested in Kostakis et al. (2015), although we note that the impact of an explosive root as in (iii) was not considered in their work.

[Insert Figure 1 here]

Local alternatives are generated according to (49) and (50). The sample size used for the local power analysis is 200, in which the power functions of Case (ii) are presented. Powers of the QR-IVX-Wald tests are calculated at the median ($\tau = 0.5$) and these curves with those of IVX-Wald tests are plotted in Figure 1. When the alternative hypotheses given in (50) are employed, the power curves are steep and show fast convergence to unity

[8]Irregular behavior at extreme quantiles is an intrinsic property of quantile regression models (Chernozhukov, 2005; Chernozhukov et al., 2016) and is not particular to instrumental variable estimation or IVX. Larger sample sizes increase the number of observations around the extreme quantiles and therefore alleviate somewhat the phenomenon of over-rejection. But it is still difficult for inference at extreme quantiles to behave like inference at moderate quantiles even in very large sample sizes. To assess relative performance as the sample size increases, some further simulation results are given in Section E of the Online Supplement.

in all cases. But under the $O(1/T)$ alternatives in (49) the power curves for small values of γ show slower rates of increase toward unity. These curves corroborate the asymptotic theory where the convergence rate of the short-horizon IVX estimators is $O(T^{\frac{1+\gamma}{2}})$ and slower than $O(T)$.

5.2 LHIVX-Wald testing

Results for the LHIVX procedure are only reported for the univariate case. The LHIVX instruments are obtained using $C_z = -5$ and $\gamma \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$. To explore local power performance, alternatives are defined by $\mathcal{H}_{\beta_T}^{LH} : \beta_T = \frac{\delta}{T}$ for integer values of $\delta \in [0, 20]$. The sample sizes are $T = 100, 200$ and the number of replications is $1,000$. The localizing parameter settings are $c_1 \in \{-1, 0, 1\}$ and $a_1 \in \{0.5, 1\}$. The innovations are generated as follows: a GARCH prediction error $u_{0t} = \sqrt{h_t}\epsilon_{0t}$, where $h_t = 1 + 0.25u_{0,t-1}^2 + 0.25h_{t-1}$; autoregressive error processes $u_{xt} = 0.5u_{x,t-1} + \epsilon_{xt}$, $u_{at} = 0.5u_{a,t-1} + \epsilon_{at}$; and ϵ_t is multivariate normal with the covariance matrix (53):

$$\Sigma_\epsilon = \begin{bmatrix} 1 & -0.75 & -0.75 \\ -0.75 & 1 & 0.50 \\ -0.75 & 0.50 & 1 \end{bmatrix}. \tag{53}$$

The sample size $T = 100$ or 200. The prediction horizon is set as $k = T^\nu$ with $\nu = 0.75$.

[Insert Table 3 here]

Table 3 reports the size performance of the LHIVX-Wald test under the null hypothesis $\mathcal{H}_0 : \beta_1 = 0$ for the sample size $T = 100$, 200. Evidently, size is well controlled below the nominal level 5% in all cases.

[Insert Figure 2 here]

To investigate power in long horizon testing we employ the local alternatives $\mathcal{H}_{\beta_T}^{(1)} : \beta_T = \frac{\delta}{T}$ where $\delta \in [0, 20]$. Power curves are shown in Figure 2. Evidently, the power functions for all choices of γ approach unity at an almost indistinguishable rate under these settings, illustrating the consistency of the LHIVX-Wald test. This uniformly strong power performance is unsurprising in this case because the convergence rate of the LHIVX estimator with the horizon setting $k = T^{3/4}$ is $\sqrt{T}k^{\frac{3}{2}} = O(T^{13/8})$, which exceeds the $O(T)$ rate on which the local alternatives are based.

Overall the results show that the LHIVX-Wald test controls size well and gives good power performance with STUR/LSTUR regressors in the long-horizon prediction model.

5.3 Choice of IVX rate parameter γ

Consider the case in which $\dim(u_{at}) = \dim(u_{xt}) = 1$. As in Phillips and Lee (2016), the asymptotic mean squared error (AMSE) criterion for predictive regression with an LSTUR regressor is found to have the form

$$\text{AMSE}\left(\widehat{\beta}_{IVX}\right) = \frac{1}{T^{1+\gamma}}\,\text{Avar}\left(\widehat{\beta}_{IVX} - \beta\right) + \frac{1}{T^{2\gamma}}\,\text{B}^2, \tag{54}$$

in which the expression of the asymptotic bias term B is given in Section F of the Online Supplement[9]. From (54) observe that the AMSE strictly decreases as γ increases. The maximum value of γ is 1, but that setting does not eliminate the size distortion induced by LSTUR regressors. It turns out therefore that use of the AMSE criterion or cross-validation fails to set a suitable upper bound for the IVX rate parameter γ.

Instead, Phillips and Lee (2016) employed numerical simulations to corroborate the use of the setting $\gamma = 0.95$ proposed in Kostakis et al. (2015) in predictive regressions with LUR regressors. Lee (2016) suggested a simulation rule for choosing γ in the quantile predictive regression case. Based on our simulation findings in regressions with LSTUR regressors, a unified upper bound of γ for IVX and QR-IVX instrumentation was set to $\gamma = 0.7$ to maintain both satisfactory power performance and good size controls[10].

6 Empirical Application

The IVX procedures, including the QR-IVX-Wald, IVX-Wald, and LHIVX-Wald tests, are employed to assess the predictability of S&P 500 US stock returns over short and long horizons. In the short-horizon prediction case, our empirical results show evidence of return predictability over various quantiles and conditional expectations. They also reveal the existence of stochastic departures from unity in the generating mechanisms of the persistent predictors. Long-term predictability is also evident in the long-horizon prediction tests.

6.1 Models and data

The DGP used in the application is the linear predictive regression model (55). Using y_t for S&P 500 excess returns and x_{t-1} as a vector of posited fundamentals, the empirical model takes the following conventional form

$$y_t = \beta_0 + \beta_1' x_{t-1} + u_{0t}. \tag{55}$$

Following Rapach et al. (2016), the excess stock return is calculated by

$$y_t = \log(1 + P_t) - \log(RF_{t-1} + 1), \tag{56}$$

where P_t and RF_t indicate the S&P 500 index and the risk-free rate. For discussions of long-horizon predictability, we estimate the following fitted model

$$y_{t+k} = b_1 x_t^k + u_{0,t+k}, \text{ and } x_t^k = \sum_{j=1}^{k} x_{t+j-1}, \tag{57}$$

in which y_{t+k} and x_{t+j-1} follow the definitions in (55) and (56). To measure fundamentals, we employ monthly, quarterly and annual data ranging from January 1952 to December

[9] Detailed discussion of the AMSE in predictive regression with LSTUR regressors is provided in Section F of the Online Supplement.

[10] We thank a referee for suggesting further discussion of the choice of the IVX rate parameter γ.

2019. Following Kostakis et al. (2015), we use a similar set of financial variables for potential predictors: Book-to-market value ratio (b/m), dividend payout ratio (d/e), default yield spread (dfy), dividend-price ratio (d/p), dividend yield (d/y), earnings-price ratio (e/p), inflation rate ($infl$), long-term yield (lty), net equity expansion ($ntis$), T-bill rate (tbl), and term spread (tms). The definitions of these variables as well as references to the many prior studies of return predictability are given in Welch and Goyal (2008) and Rapach et al. (2016).

To detect potential stochastic departures from unity in the predictors, the following approach is used: an exogenous IV-assisted test of the hypothesis $\mathcal{H}_0 : a = 0$ (Lieberman and Phillips, 2018) is conducted to assess the relevance of potential stochastic regressors influencing the autoregressive coefficient. The method is discussed and a computational algorithm is provided in Section D of the Online Supplement. The data employed in the stochastic unit root test are based on the updated factor library (i.e., SMB, HML, RMW, CMA, MOM) maintained by Kenneth French[11], covering the period from July 1963 to December 2019. Details of the factor construction are given in Fama and French (2015). Statistical significance in the reported outcomes is denoted by starred affixes according to the usual convention.[12]

6.2 Stochastic unit root tests

We report test results for the presence of stochastic deviations from unity in the persistence characteristics of the predictive regressor. A persistent regressor of interest that has empirical support from past research as a predictor is the Treasury Bill Rate defined as $x_{t-1}(:= 100 \times tbl_{t-1})$[13] with a potential stochastic unit root component $u_{at} = \Delta\,(d/p)_t$, where $(d/p)_t$ is the dividend-price ratio. The exogenous instrument employed in the IV test is $Z_t = RMW_t$, the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios. The observed short-run sample correlations in the data are $\widehat{\rho}_{a,\Delta x} = 0.1397$, $\widehat{\rho}_{a,Z} = 0.1519$, and $\widehat{\rho}_{Z,\Delta x} = 0.0612$, and the observed long-run sample correlations are $\widehat{\rho}^{lr}_{a,\Delta x} = -0.0897$, $\widehat{\rho}^{lr}_{a,Z} = 0.1495$, and $\widehat{\rho}^{lr}_{Z,\Delta x} = -0.0794$ supporting the suitability of IV estimation in terms of both orthogonality and relevance.

We apply the test statistic S_T for randomness detection that is discussed in Section D of the Online Supplement. The statistic is asymptotically pivotal and respective critical values and confidence intervals (CIs) are computed from the asymptotic standard Cauchy distribution under the null hypothesis of no random departure from unity ($\mathcal{H}_0^a : a_1 = 0$). Two-sided confidence intervals generated by the exogenous IV-assisted method (Lieberman and Phillips, 2018) for 90% and 95% significance levels are found to be $(-6.3138, 6.3138)$ and $(-12.7062, 12.7062)$. Correspondingly, the test statistic S_T for randomness detection generated by the IV-assisted method is 14.8169 with a two-sided p-value 0.0429, showing

[11] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[12] * for $p < 0.1$, ** for $p < 0.05$, and *** for $p < 0.01$.

[13] Wald tests are self-normalized and unit-free, so that tbl_{t-1} and $100tbl_{t-1}$ are equivalent predictors.

that the stochastic departure from unity caused by the $\Delta(d/p)_t$ factor is statistically significant at the 5% significance level over the whole sample. These results provide strong empirical support for the presence of stochastic unit root regressors in predictive regression of S&P 500 US stock returns.

To check robustness of these findings of the STUR effects of $\Delta(d/p)_t$ in the generation of the regressor, the momentum factor MOM_t[14] is used as another potential instrument. The observed short-run sample correlations in the data are $\widehat{\rho}_{a,\Delta x} = 0.1397$, $\widehat{\rho}_{a,Z} = 0.1661$, and $\widehat{\rho}_{Z,\Delta x} = 0.0756$, and the observed long-run sample correlations are $\widehat{\rho}^{lr}_{a,\Delta x} = -0.0897$, $\widehat{\rho}^{lr}_{a,Z} = 0.1372$, and $\widehat{\rho}^{lr}_{Z,\Delta x} = 0.0385$. These results provide evidence for the suitability of the momentum factor MOM_t as an instrument in terms of the orthogonality and relevance conditions. Using this instrument the computed value of the statistic S_T is 30.8293, with the p-value 0.0206 under the null hypothesis $\mathcal{H}^a_0 : a_1 = 0$, giving further evidence for the presence of stochastic unit root regressors.

6.3 Results of predictability inferences

[Insert Tables 4–6 and Figure 3 here]

Tables 4–6 report the short-horizon IVX-Wald and QR-IVX-Wald tests in the univariate regression model. For short-horizon testing, Tables 4–7 and Figure 3 are generated with the setting $\gamma = 0.7$. At the 40%, 50% and 60% quantiles of excess returns there is evidence of predictability in (d/p), (lty) and (tbl) with monthly data at the 5% significance level. Similarly, at 40%, 50% and 60% quantiles of excess returns, there is evidence of predictability in (b/m), (d/p), (d/y), (e/p), and (tbl) for quarterly data at the 5% significance level. At these quantiles of excess return, only (d/p) and (d/y) show potential for return predictability with annual data at the 5% significance level.

Figure 3 plots the IVX and QR-IVX slope estimates and the corresponding 95% confidence intervals for various market fundamentals, corresponding to the results of Table 4 for monthly data. These curves show evidence of return predictability for tbl_t at many quantiles, for $(d/p)_t$ at intermediate quantiles 0.4 and 0.7, and for lty_t at most quantiles. Testing for stochastic departures from unity, tbl_t belongs to the class of STUR/LSTUR regressors for which the IVX-Wald and QR-IVX-Wald tests detect short-horizon predictive ability. Overall, monthly predictability is not significant for many of the other variables at most quantiles from the QR-IVX-Wald test and for the mean from the IVX-Wald test. Predictability is rarely found at the median except for (lty) and (tbl). Comparatively, predictive ability is evident at extreme quantiles for several persistent predictors, as found in Cenesizoglu and Timmermann (2008). But this result is tempered by the knowledge from the simulations that mild over-rejection at the extreme quantiles can lead to spurious predictability and cast doubt on the findings. For testing the forecasting capacity at the mean, only (lty) and (tbl) again demonstrate significant results, corroborating doubts found by Kostakis et al. (2015) about predictability results from least squares methods.

[14] MOM_t is the average return on the two highest prior return portfolios minus the average return on the two lowest prior return portfolios

[Insert Table 7 here]

Table 7 shows the multiple regression results for tests involving various combinations of fundamentals. First, we examine combinations of fundamentals that are well-used in empirical finance, including (a) (d/p), (tbl) (Ang and Bekaert, 2007); (b) (e/p), (tbl) (Kostakis et al., 2015); (c) (d/p), (b/m) (Kothari and Shanken, 1997); (d) (d/p), (d/e) (Lamont, 1998). The findings reveal that return predictability is more manifest in monthly than quarterly or annual data with these regressor combinations. After eliminating over-rejections in the extreme quantiles of monthly excess returns, the QR-IVX-Wald tests show predictive ability in (a) (d/p), (tbl) and (b) (e/p), (tbl) combinations at the 5% level. In comparison only the (b) (e/p), (tbl) combination shows strong evidence of predicting S&P 500 excess returns in quarterly data and no combination shows predictability in annual data. Only (a) (d/p), (tbl) is detected by the IVX-Wald test to demonstrate predictive ability for the conditional expectation of excess returns. Since (tbl) demonstrates stochastic departures from unity and evidences widespread predictability in these combination regressions the results show the usefulness of IVX robustness to LSTUR and STUR instruments in predictive regression.

[Insert Table 8 here]

Table 8 reports the LHIVX-Wald tests for long-horizon predictability over the univariate regression models, in which we set $\gamma = 0.5$. For monthly and quarterly data, we select the prediction horizons $k = 30$, 60, and 90. For annual data, we set the prediction horizons to $k = 10$, 20, and 30. Compared to the potential predictors detected in the short-horizon IVX tests, the only fundamental variable to preserve some of its predictive ability in these long horizon regressions is (tbl), confirming the usefulness of IVX robustness to STUR/LSTUR regressors. The predictors $(infl)$ and (tms) also reject the null hypothesis of no long-horizon predictability at the 5% significance level.

7 Conclusion

The IVX and QR-IVX methodology provides a convenient approach to robust estimation and inference in a broad class of predictive models, either in standard linear regression form or quantile regression format. Moreover, the approach is available in both short-horizon and long-horizon regressions.

The present work shows that this methodology allows for a much broader class of predictive regressors that includes stochastic as well as deterministic local departures from unit root regressors. Notably, the process of IVX instrument construction continues to use the same simple autoregressive formulation that utilizes potentially endogenous regressors as the vehicle of instrumentation so that external instruments are not needed. The resulting inferential framework involves a pivotal chi-square distribution and requires no simulations to compute CIs. As Kostakis et al. (2015) demonstrated in simulations and empirical work, a key advantage in practical work is the capability of IVX methodology to accommodate multiple regressor cases with no fundamental changes in the approach. The present paper shows that this advantage carries over to a very wide class of regressors that encompass stochastic, deterministic, and mixed near-stationary and near-explosive departures from a simple unit root or local unit root regressor processes, thereby extending the range of practical application of this methodology in empirical work.

Appendices

Throughout the following proofs we use the same notation as in the paper. The technical lemmas listed in the next section are proved in the Online Supplement and play central roles in the proofs of the main theorems in the paper.

Further, we define the following asymptotic variance matrices of the IVX and QR-IVX estimators in terms of component matrices as

$$\mathbb{V}^{IVX} := \sigma_{00} V_{zx}^{-1} \left[V_{zz} + V_{\eta\eta}^{(2)} + V_{z\eta}^{(2)} + V_{\eta z}^{(2)} \right] V_{xz}^{-1}, \tag{58}$$

and

$$\mathbb{V}^{QRIVX} := \frac{\tau(1-\tau)}{p_{u0\tau}^2(0)} \left(V_{zx}^{QRIVX} \right)^{-1} \left[V_{zz} + V_{\eta\eta}^{(2)} + V_{z\eta}^{(2)} + V_{\eta z}^{(2)} \right] \left(V_{xz}^{QRIVX} \right)^{-1}, \tag{59}$$

where V_{zz}, $V_{z\eta}^{(2)}$ and $V_{\eta\eta}^{(2)}$ are defined in Lemma A.4, V_{xz} and V_{xz}^{QRIVX} are defined in Lemma A.5, $V_{\eta z}^{(2)} = V_{z\eta}^{(2)\,\prime}$, $V_{zx} = V_{xz}'$, and $V_{zx}^{QRIVX} = V_{xz}^{QRIVX\,\prime}$.

The asymptotic variance matrix of the LHIVX estimator is similarly defined as

$$\mathbb{V}^{LHIVX} := \left(\Upsilon^{-1}\right)' \left(V_{zz}^{LH} + V_{\eta\eta}^{(2),LH} + V_{z\eta}^{(2),LH} + V_{z\eta}^{(2),LH\,\prime} \right) \left(\Upsilon^{-1}\right) \sigma_{00}, \tag{60}$$

in which the explicit expressions for V_{zz}^{LH}, $V_{\eta\eta}^{(2),LH}$, $V_{z\eta}^{(2),LH}$, and Υ are given below in (85)-(88) in this Appendix.

A Technical lemmas in short-horizon predictive regression

Lemma A.1 *Let Assumptions 1 and 2 hold. As $T \to \infty$,*

$$\sup_{r\in[0,1]} \left\| \eta_{T,\lfloor Tr \rfloor -1}^{(1)} \right\| = O_p\left(T^{\gamma+\frac{1}{2}}\right),$$
$$\sup_{r\in[0,1]} \left\| \eta_{T,\lfloor Tr \rfloor -1}^{(2)} \right\| = O_p\left(T^{\frac{1+\gamma}{2}}\right),$$
$$\sup_{r\in[0,1]} \left\| \eta_{T,\lfloor Tr \rfloor -1}^{(3)} \right\| = O_p\left(T^{\gamma+\frac{1}{2}}\right).$$

Lemma A.2 *Let Assumptions 1 and 2 hold. Let all $t = \lfloor Tr \rfloor$ with $r \in [0,1]$ as $T \to \infty$.*

(i) For the IVX residual term $\eta_{T,t-1}^{(2)}$, then

$$\frac{1}{T^{\frac{1+\gamma}{2}}} \eta_{T,t-1=\lfloor Tr \rfloor -1}^{(2)} \sim_a Z_a G_{a,c}(r) + O_p\left(\frac{1}{T^{\frac{1-\gamma}{2}}}\right)$$

$$\rightsquigarrow \mathcal{MN}\left(\mathbf{0}_{n\times 1},\ \frac{a'\Omega_{aa}a}{-2c_z}G_{a,c}(r)G_{a,c}(r)'\right), \tag{61}$$

where $Z_a =_d \mathcal{N}\left(0,\ \frac{a'\Omega_{aa}a}{-2c_z}\right)$. *Replace* $G_{a,c}(r)$ *by* $G_a(r)$ *when* $C = \mathbf{0}_{n\times n}$.

(ii) For the IVX instrument $\widetilde{z}_{t-1}$*, then*

$$\frac{1}{T^{\frac{1}{2}+\gamma}}\sum_{t=1}^{T}\widetilde{z}_{t-1=\lfloor Tr\rfloor-1} = -\frac{C_z^{-1}}{\sqrt{T}}x_T + O_p\left(\frac{1}{T^{\frac{1-\gamma}{2}}}\right) \sim_a -C_z^{-1}G_{a,c}(1), \tag{62}$$

for the LSTUR case. Replace $G_{a,c}(r)$ *by* $G_a(r)$ *when* $C = \mathbf{0}_{n\times n}$.

Lemma A.3 *Let Assumptions 1 and 2 hold. As* $T\to\infty$,

(i) For the numerator of the IVX estimator, the following asymptotic approximation holds

$$\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\underline{u}_{0t} = \frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}z_{t-1}u_{0t} + \frac{1}{T^{1+\frac{\gamma}{2}}}\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}u_{0t} + o_p(1).$$

(ii) For the asymptotic variance of the IVX numerator, the following approximation holds

$$\begin{aligned}\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\widetilde{z}_{t-1}' =& \frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}z_{t-1}z_{t-1}' + \frac{1}{T^{\frac{3}{2}+\gamma}}\left(\sum_{t=1}^{T}z_{t-1}\eta_{T,t-1}^{(2)\,\prime} + \sum_{t=1}^{T}\eta_{T,t-1}^{(2)}z_{t-1}'\right)\\ &+ \frac{1}{T^{2+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}\eta_{T,t-1}^{(2)\,\prime} + o_p(1).\end{aligned}$$

(iii) For the denominator of the IVX estimator, the following approximation holds

$$\begin{aligned}\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\underline{x}_{t-1}' =& \frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}z_{t-1}x_{t-1}' + \frac{1}{T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}x_{t-1}' + \frac{C}{T^{2+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}x_{t-1}'\\ &+ \frac{1}{2T^{2+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(3)}x_{t-1}' + \left(\frac{1}{\sqrt{T}}x_T\right)\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}x_{t-1}\right)'C_z^{-1} + o_p(1).\end{aligned}$$

Lemma A.4 *Let Assumptions 1 and 2 hold. As* $T\to\infty$,

(i) For the terms of the IVX numerator,

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}z_{t-1}z_{t-1}' \to_p V_{zz} := -\frac{1}{2c_z}\Omega_{xx},$$

$$\frac{1}{T^{2+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}\eta_{T,t-1}^{(2)\,\prime} \rightsquigarrow V_{\eta\eta}^{(2)} := \begin{cases}-\frac{1}{2c_z}(a'\Omega_{aa}a)\int_0^1 G_{a,c}(r)G_{a,c}'(r)dr, & \text{under LSTUR},\\ -\frac{1}{2c_z}(a'\Omega_{aa}a)\int_0^1 G_a(r)G_a'(r)dr, & \text{under STUR},\end{cases}$$

$$\frac{1}{T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T}z_{t-1}\eta_{T,t-1}^{(2)\,\prime} \rightsquigarrow V_{z\eta}^{(2)} := \begin{cases}-\frac{1}{2c_z}(\Omega_{xa}a)\int_0^1 G_{a,c}'(r)dr, & \text{under LSTUR},\\ -\frac{1}{2c_z}(\Omega_{xa}a)\int_0^1 G_a'(r)dr, & \text{under STUR}.\end{cases}$$

(ii) *For the terms of the IVX denominator,*

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}x_{t-1}z_{t-1}' \rightsquigarrow V_{xz}^{l} := \begin{cases} \frac{-1}{c_z}\left(\int_0^1 G_{a,c}(r)dB_x'(r)+\Omega_{xx}+\int_0^1 G_{a,c}(r)\left(a'\Omega_{ax}\right)dr\right), & \text{LSTUR}, \\ \frac{-1}{c_z}\left(\int_0^1 G_{a}(r)dB_x'(r)+\Omega_{xx}+\int_0^1 G_{a}(r)\left(a'\Omega_{ax}\right)dr\right), & \text{STUR}, \end{cases}$$

$$\frac{1}{T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T}x_{t-1}\eta_{T,t-1}^{(2)\,\prime} \rightsquigarrow V_{x\eta}^{(2)},$$

$$\frac{1}{T^{2+\gamma}}\sum_{t=1}^{T}x_{t-1}\eta_{T,t-1}^{(1)\,\prime}C \rightsquigarrow V_{x\eta}^{(1)} := \begin{cases} -\frac{1}{c_z}\int_0^1 G_{a,c}(r)G_{a,c}'(r)drC, & \text{under LSTUR}, \\ -\frac{1}{c_z}\int_0^1 G_{a}(r)G_{a}'(r)drC, & \text{under STUR}, \end{cases}$$

$$\frac{1}{T^{2+\gamma}}\sum_{t=1}^{T}x_{t-1}\eta_{T,t-1}^{(3)\,\prime} \rightsquigarrow V_{x\eta}^{(3)} := \begin{cases} -\frac{1}{c_z}\int_0^1 G_{a,c}(r)G_{a,c}'(r)\left(a'\Sigma_{aa}a\right)dr, & \text{under LSTUR}, \\ -\frac{1}{c_z}\int_0^1 G_{a}(r)G_{a}'(r)\left(a'\Sigma_{aa}a\right)dr, & \text{under STUR}. \end{cases}$$

where

$$\begin{aligned} V_{x\eta}^{(2)} &:= -\frac{1}{c_z}\left[\int_0^1 G_{a,c}(r)G_{a,c}'(r)\left(a'dB_a(r)\right)+\int_0^1 G_{a,c}(r)G_{a,c}'(r)\left(a'\Omega_{aa}a+a'\Lambda_{aa}a\right)dr\right. \\ &\quad \left.+\int_0^1\left(\Omega_{xa}a\right)G_{a,c}'(r)dr+\int_0^1 G_{a,c}(r)\left(a'\Lambda_{ax}\right)dr\right], \end{aligned} \tag{63}$$

where $C \neq \mathbf{0}_{n\times n}$ in the case of the LSTUR regressors;

$$\begin{aligned} V_{x\eta}^{(2)} &:= -\frac{1}{c_z}\left[\int_0^1 G_{a}(r)G_{a}'(r)\left(a'dB_a(r)\right)+\int_0^1 G_{a}(r)G_{a}'(r)\left(a'\Omega_{aa}a+a'\Lambda_{aa}a\right)dr\right. \\ &\quad \left.+\int_0^1\left(\Omega_{xa}a\right)G_{a}'(r)dr+\int_0^1 G_{a}(r)\left(a'\Lambda_{ax}\right)dr\right], \end{aligned} \tag{64}$$

where $C = \mathbf{0}_{n\times n}$ in the case of the STUR regressors.

Lemma A.5 *Suppose Assumptions 1 and 2 hold. As $T\to\infty$,*

(i) *For the IVX numerator, then*

$$\begin{pmatrix} \frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}z_{t-1}u_{0t} \\ \frac{1}{T^{1+\frac{\gamma}{2}}}\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}u_{0t} \end{pmatrix} \rightsquigarrow \mathcal{MN}\left(\mathbf{0}_{2n\times 1},\ \sigma_{00}\begin{bmatrix} V_{zz} & V_{z\eta}^{(2)} \\ V_{z\eta}^{(2)\,\prime} & V_{\eta\eta}^{(2)} \end{bmatrix}\right), \tag{65}$$

where V_{zz}, $V_{z\eta}^{(2)}$, and $V_{\eta\eta}^{(2)}$ are defined in Lemma A.4.

(ii) *For the QR-IVX numerator, then*

$$\begin{pmatrix} \frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}z_{t-1}\Psi_\tau\left(u_{0t\tau}\right) \\ \frac{1}{T^{1+\frac{\gamma}{2}}}\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}\Psi_\tau\left(u_{0t\tau}\right) \end{pmatrix} \rightsquigarrow \mathcal{MN}\left(\mathbf{0}_{2n\times 1},\ \tau\left(1-\tau\right)\begin{bmatrix} V_{zz} & V_{z\eta}^{(2)} \\ V_{z\eta}^{(2)\,\prime} & V_{\eta\eta}^{(2)} \end{bmatrix}\right), \tag{66}$$

where V_{zz}, $V_{z\eta}^{(2)}$, and $V_{\eta\eta}^{(2)}$ are defined in Lemma A.4.

(iii) For the IVX denominator, then

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\underline{x}_{t-1}\widetilde{z}'_{t-1}\rightsquigarrow V_{xz}:=V_{xz}^{l}+V_{x\eta}^{(1)}+V_{x\eta}^{(2)}+\frac{1}{2}V_{x\eta}^{(3)}+V_{demean}, \tag{67}$$

where V_{xz}^{l}, $V_{x\eta}^{(1)}$, $V_{x\eta}^{(2)}$ and $V_{x\eta}^{(3)}$ are defined in Lemma A.4; and

$$V_{demean}:=\begin{cases}\frac{1}{c_z}\left(\int_0^1 G_{a,c}(r)dr\right)G'_{a,c}(1) & \text{under} \quad \text{LSTUR}\\ \frac{1}{c_z}\left(\int_0^1 G_{a}(r)dr\right)G'_{a}(1) & \text{under} \quad \text{STUR}\end{cases}.$$

(iv) For the QR-IVX denominator, then

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}p_{u0t\tau,t-1}(0)\,x_{t-1}\widetilde{z}'_{t-1}\rightsquigarrow p_{u0\tau}(0)\;V_{xz}^{QRIVX}, \tag{68}$$

where $V_{xz}^{QRIVX}:=V_{xz}^{l}+V_{x\eta}^{(1)}+V_{x\eta}^{(2)}+\frac{1}{2}V_{x\eta}^{(3)}$. The limiting covariance matrices V_{xz}^{l}, $V_{x\eta}^{(1)}$, $V_{x\eta}^{(2)}$ and $V_{x\eta}^{(3)}$ are defined in Lemma A.4.

Lemma A.6 *Suppose Assumptions 1 and 2 hold. Then we have the following results.*

(i) For the numerator of short-horizon QR-IVX and IVX: as $T\to\infty$,

$$\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\underline{u}_{0t}\rightsquigarrow\mathcal{MN}\left(\mathbf{0}_{n\times1},\;\sigma_{00}\left(V_{zz}+V_{\eta\eta}^{(2)}+V_{z\eta}^{(2)}+V_{z\eta}^{(2)\,\prime}\right)\right), \tag{69}$$

for the short-horizon IVX estimator; as $T\to\infty$,

$$\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\Psi_{\tau}(u_{0t\tau})\rightsquigarrow\mathcal{MN}\left(\mathbf{0}_{n\times1},\;\tau\left(1-\tau\right)\left(V_{zz}+V_{\eta\eta}^{(2)}+V_{z\eta}^{(2)}+V_{z\eta}^{(2)\,\prime}\right)\right),$$

for the short-horizon QR-IVX estimator. The limiting covariance matrices V_{zz}, $V_{\eta\eta}^{(2)}$, and $V_{z\eta}^{(2)}$ are defined in Lemma A.4.

(ii) For the sample variance of the short-horizon QR-IVX and IVX numerators: as $T\to\infty$, then

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\widetilde{z}'_{t-1}\rightsquigarrow\left(V_{zz}+V_{\eta\eta}^{(2)}+V_{z\eta}^{(2)}+V_{z\eta}^{(2)\,\prime}\right),$$

where V_{zz}, $V_{z\eta}^{(2)}$ and $V_{\eta\eta}^{(2)}$ are defined in Lemma A.4.

(iii) For the denominator of short-horizon QR-IVX and IVX: as $T\to\infty$,

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\underline{x}_{t-1}\widetilde{z}'_{t-1}\rightsquigarrow V_{xz},$$

for the short-horizon IVX estimator, where V_{xz} is defined in Lemma A.5. Moreover, as $T \to \infty$,

$$\frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} p_{u0t\tau,t-1}(0) x_{t-1} \widetilde{z}_{t-1}' \rightsquigarrow p_{u0\tau}(0)\, V_{xz}^{QRIVX},$$

for the short-horizon QR-IVX estimator, where $V_{xz}^{QRIVX} := V_{xz}^{l} + V_{x\eta}^{(1)} + V_{x\eta}^{(2)} + \frac{1}{2} V_{x\eta}^{(3)}$.

Lemma A.7 *For some constant $M > 0$,*

$$\sup \left\{ \| H_T(\vartheta) - H_T(0) \| : \|\vartheta\| \leq T^{\frac{1+\gamma}{2}} M \right\} = o_p(1),$$

where $H_T(\vartheta) := T^{-\frac{1+\gamma}{2}} \sum_{t=1}^{T} \widetilde{z}_{t-1} \left\{ \Psi_\tau \left(u_{0t\tau} - \vartheta' \check{X}_{t-1} \right) - \mathbb{E}_{t-1} \left[\Psi_\tau \left(u_{0t\tau} - \vartheta' \check{X}_{t-1} \right) \right] \right\}$.

Proof of Theorem 3.1

Proof: Since uniform convergence is confirmed in Lemma A.7, the standard result for extremum estimation with a non-smooth criterion function holds following Lee (2016). Let $\widehat{\beta}_{1,\tau} := \widehat{\beta}_{1,\tau}^{QRIVX}$ and $\widehat{\beta}_{0,\tau}$ be a preliminary QR estimator in the following argument. Define $\widehat{\vartheta}_{1,\tau} := \left(\widehat{\beta}_{1,\tau} - \beta_{1,\tau} \right)$, $\widehat{\vartheta}_{0,\tau} := \left(\widehat{\beta}_{0,\tau} - \beta_{0,\tau} \right)$ and then

$$\widehat{\beta}_{1,\tau} \sim_a \arg\min_{\beta_{1,\tau}^{\star}} \left(\sum_{t=1}^{T} m_t \left(\beta_{1,\tau}^{\star} \right) \right)' \left(\sum_{t=1}^{T} m_t \left(\beta_{1,\tau}^{\star} \right) \right),$$

where $m_t \left(\beta_{1,\tau}^{\star} \right) = \widetilde{z}_{t-1} \Psi_\tau \left(u_{0t\tau} \left(\beta_{1,\tau}^{\star} \right) \right) = \widetilde{z}_{t-1} \left(\tau - \mathbf{1} \left(y_t \leq \widehat{\beta}_{0,\tau} + \beta_{1,\tau}^{\star\prime} x_{t-1} \right) \right)$, and $\mathbf{1}(\cdot)$ is an indicator function. By Theorem 1 of Xiao (2009), the QR estimator $\widehat{\beta}_{0,\tau}$ is $\sqrt{T}$-consistent. Therefore, in the QR-IVX procedure

$$\begin{aligned}
&\frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} \widetilde{z}_{t-1} \Psi_\tau \left(y_t - \beta_{0,\tau} - \beta_{1,\tau}^{\star\prime} x_{t-1} \right) - \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} \widetilde{z}_{t-1} \Psi_\tau \left(y_t - \widehat{\beta}_{0,\tau} - \beta_{1,\tau}^{\star\prime} x_{t-1} \right) \\
=& \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} \widetilde{z}_{t-1} \left\{ \mathbf{1} \left(u_{0t\tau} \leq \left(\beta_{1,\tau}^{\star\prime} - \beta_{1,\tau}' \right) x_{t-1} \right) - \mathbf{1} \left(u_{0t\tau} \leq \left(\widehat{\beta}_{0,\tau} - \beta_{0,\tau} \right) + \left(\beta_{1,\tau}^{\star\prime} - \beta_{1,\tau}' \right) x_{t-1} \right) \right\} \\
=& \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} \widetilde{z}_{t-1} \xi_t, \qquad (70)
\end{aligned}$$

where

$$\xi_t := \left\{ \mathbf{1} \left(u_{0t\tau} \leq \left(\beta_{1,\tau}^{\star\prime} - \beta_{1,\tau}' \right) x_{t-1} \right) - \mathbf{1} \left(u_{0t\tau} \leq \left(\widehat{\beta}_{0,\tau} - \beta_{0,\tau} \right) + \left(\beta_{1,\tau}^{\star\prime} - \beta_{1,\tau}' \right) x_{t-1} \right) \right\}.$$

By taking conditional expectations $\mathbb{E}_{t-1}(\cdot)$ and for any possible value of $\beta_{1,\tau}^{\star}$ we have

$$\left\| \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} \widetilde{z}_{t-1} \cdot \mathbb{E}_{t-1} \xi_t \right\| \sim_a \left\| p_{u0\tau}(0) \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} \widetilde{z}_{t-1} \left(\widehat{\beta}_{0,\tau} - \beta_{0,\tau} \right) \right\|$$

$$= O_p\left(\frac{1}{T^{1+\frac{\gamma}{2}}}\right)\cdot\left\|\sum_{t=1}^{T}\widetilde{z}_{t-1}\right\| = O_p\left(T^{\frac{-1+\gamma}{2}}\right), \tag{71}$$

where the last equality is due to Lemma A.2 and the penultimate equality results from Theorem 1 of Xiao (2009). In addition, for any possible value of $\beta^{\star}_{1,\tau}$, the conditional quadratic variation of the objective function is asymptotically diminishing as

$$\left\|\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\mathbb{E}_{t-1}\left(\widetilde{z}_{t-1}\widetilde{z}'_{t-1}\cdot\xi_t^2\right)\right\| \leq \left\|\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\widetilde{z}'_{t-1}\right\|\cdot\left(\sup_{1\leq t\leq T}\mathbb{E}_{t-1}\xi_t^2\right)$$
$$= O_p\left(\sup_{1\leq t\leq T}\mathbb{E}_{t-1}\xi_t^2\right) = o_p(1), \tag{72}$$

under the $\sqrt{T}$-consistency of the preliminary estimator $\widehat{\beta}_{0,\tau}$. Combining (71) and (72) shows that (70) is $o_p(1)$. So $\widehat{\beta}_{0,\tau}$ does not asymptotically affect QR-IVX estimation. Following Lee (2016) and Lemma A.7, the first order condition follows

$$o_p(1) = \sum_{t=1}^{T}\widetilde{z}_{t-1}\Psi_\tau(u_{0t\tau}) + \sum_{t=1}^{T}p_{u0t\tau,t-1}(0)\,\widetilde{z}_{t-1}x'_{t-1}\left(\widehat{\beta}_{1,\tau}-\beta_{1,\tau}\right), \tag{73}$$

and for any $\tau\in(0,1)$, then the limiting distributions of $\widehat{\beta}_{1,\tau}$ follow equation (73) and Lemmas A.5 and A.6.

Similarly, the derivations for the short-horizon IVX estimate are straightforward and follow directly from the representation of the IVX estimation error. In particular,

$$T^{\frac{1+\gamma}{2}}\left(\widehat{\beta}_1^{IVX}-\beta_1\right) = \left(\frac{1}{T^{1+\gamma}}\widetilde{Z}'_{-1}\underline{X}_{-1}\right)^{-1}\left(\frac{1}{T^{\frac{1+\gamma}{2}}}\widetilde{Z}'_{-1}\underline{u}_0\right), \tag{74}$$

and, using Lemma A.6, we have

$$\frac{1}{T^{\frac{1+\gamma}{2}}}\widetilde{Z}'_{-1}\underline{u}_0 = \frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}z_{t-1}u_{0t} + \frac{1}{T^{\frac{2+\gamma}{2}}}\sum_{t=1}^{T}\eta^{(2)}_{T,t-1}u_{0t} + o_p(1)$$
$$\rightsquigarrow \mathcal{MN}\left(\mathbf{0}_{n\times 1},\ \sigma_{00}\left(V_{zz}+V^{(2)}_{\eta\eta}+V^{(2)}_{z\eta}+V^{(2)\,\prime}_{z\eta}\right)\right), \tag{75}$$

and $\frac{1}{T^{1+\gamma}}\widetilde{Z}'_{-1}\underline{X}_{-1}\rightsquigarrow V_{zx}$. Combining these results, we have

$$T^{\frac{1+\gamma}{2}}\left(\widehat{\beta}_1^{IVX}-\beta_1\right)\rightsquigarrow\mathcal{MN}\left(\mathbf{0}_{n\times 1},\ \sigma_{00}V_{zx}^{-1}\left(V_{zz}+V^{(2)}_{\eta\eta}+V^{(2)}_{z\eta}+V^{(2)\,\prime}_{z\eta}\right)\left(V_{zx}^{-1}\right)'\right), \tag{76}$$

giving the required limit. This completes the proof of Theorem 3.1. ■

Proof of Theorem 3.2

Proof: Based on the limit theory in Theorem 3.1 and the asymptotic independence between the numerator and the denominator, the Wald test asymptotics follow directly. ■

B Technical lemmas in long-horizon predictive regression

We introduce several additional random components to facilitate the discussion of long-horizon predictive regression. We define the Gaussian variates $(Z_x^{\star\prime}, Z_a^{\star\prime})'$ driven jointly by taking averages around the time point $t^\star$ across the whole prediction horizon k as

$$\left(\frac{1}{\sqrt{k}}\sum_{j=1}^{k} u'_{x,t^\star+j}, \frac{1}{\sqrt{k}}\sum_{j=1}^{k} u'_{a,t^\star+j}\right)' \rightsquigarrow (Z_x^{\star\prime}, Z_a^{\star\prime})' =_d \mathcal{N}\left(\mathbf{0}_{2n\times 1}, \begin{pmatrix} \Omega_{xx} & \Omega_{xa} \\ \Omega_{ax} & \Omega_{aa} \end{pmatrix}\right), \tag{77}$$

for any given $t^\star$ with $t^\star = \lfloor Tr^\star \rfloor$ and $r^\star \in [0,1]$ and $1 \le j \le k$. Comparatively, the Brownian motions commonly used in the long-horizon predictive regression models, $(B_0(r), B'_x(r), B'_a(r))'$, are driven by partial summations over the time horizon T as

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor - k} u_{0,t+j}, \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor - k} u'_{x,t+j-1}, \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor - k} u'_{a,t+j-1}\right)' \rightsquigarrow \left(B_0(r), B'_x(r), B'_a(r)\right)', \tag{78}$$

where $1 \le j \le k$. The FCLT results of (77) and (78) are derived by assuming the rate restriction $\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0$.

By (77) and (78), it follows that the two groups of Gaussian variates are uncorrelated. So for any $t^\star = \lfloor Tr^\star \rfloor$ with $r^\star \in [0,1]$, then

$$\left(Z_x^{\star\prime}, Z_a^{\star\prime}\right)' \perp \left(B_0(r), B'_x(r), B'_a(r)\right)', \tag{79}$$

as $\left(\frac{1}{\sqrt{k}}\sum_{j=1}^{k} u'_{x,t^\star+j}, \frac{1}{\sqrt{k}}\sum_{j=1}^{k} u'_{a,t^\star+j}\right)'$ escapes asymptotically from its probability space and demonstrates the stable mixing property. In addition, for any $t^\star$ with $t^\star = \lfloor Tr^\star \rfloor$ and $r^\star \in [0,1]$, we have joint convergence of the above two groups of Gaussian variates as

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor - k} u_{0,t+j}, \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor - k} u'_{x,t+j-1}, \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor - k} u'_{a,t+j-1}, \frac{1}{\sqrt{k}}\sum_{j=1}^{k} u'_{x,t^\star+j}, \frac{1}{\sqrt{k}}\sum_{j=1}^{k} u'_{a,t^\star+j}\right)'$$
$$\rightsquigarrow \left(B_0(r), B'_x(r), B'_a(r), Z_x^{\star\prime}, Z_a^{\star\prime}\right)', \tag{80}$$

and asymptotic independence between $(Z_x^{\star\prime}, Z_a^{\star\prime})'$ and $(B_0(r), B'_x(r), B'_a(r))'$. Based on the joint convergence in (80) and asymptotic independence, the mixed normality of the LHIVX estimator and the pivotal chi-squared test statistics can be established, as collected in the following lemmas and theorems.

Lemma B.1 *Suppose Assumptions 1 and 2 hold. Under the rate condition* $\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0$, *then*

(i) For any $t = \lfloor Tr \rfloor$ *and* $r \in [0,1]$, $-\frac{C_z}{kT^{\frac{1}{2}+\gamma}}\eta_{Tt}^{1,k} \rightsquigarrow \begin{cases} G_{a,c}(r), & \text{under LSTUR} \\ G_a(r), & \text{under STUR} \end{cases}.$

(ii) $\sum_{t=1}^{T-k} u_{0,t+k}\eta_{T,t}^{1,k\,\prime} = O_p\left(kT^{1+\gamma}\right)$, *and* $\frac{1}{\sqrt{k}T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T-k} u_{0,t+k}\eta_{T,t}^{1,k\,\prime} = o_p(1)$.

(iii) $-\frac{C_z}{\sqrt{k}T^{\gamma}} z_t^k \rightsquigarrow Z_x^\star$ *where* $Z_x^\star$ *is defined in* (77), *and*

$$\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{C_z}{\sqrt{k}T^{\gamma}} z_t^k\right)\left(\frac{C_z}{\sqrt{k}T^{\gamma}} z_t^k\right)' \to_p \Omega_{xx}, \tag{81}$$

in which $t = \lfloor Tr \rfloor$ *with* $r \in [0,1]$.

(iv) $\frac{1}{\sqrt{k}T^{\frac{1}{2}+\gamma}}\sum_{t=1}^{T-k} u_{0,t+k} z_t^{k\,\prime} \rightsquigarrow -\int_0^1 Z_x^{\star\prime} dB_0(r)\cdot C_z^{-1} =_d \mathcal{N}(0,\ \sigma_{00}\Omega_{xx}C_z^{-2})$, .

(v) *For any* $t = \lfloor Tr \rfloor$ *and* $r \in [0,1]$, *then*

$$-\frac{C_z}{\sqrt{k}T^{\frac{1}{2}+\gamma}}\eta_{T,t}^{2,k} \rightsquigarrow \begin{cases} (a'Z_a^\star)\, G_{a,c}(r), & \text{under} \quad \text{LSTUR} \\ (a'Z_a^\star)\, G_a(r), & \text{under} \quad \text{STUR} \end{cases}, \tag{82}$$

where $Z_a^\star$ *is defined in* (77).

(vi) $-\frac{C_z}{\sqrt{k}T^{1+\gamma}}\sum_{t=1}^{T-k} u_{0,t+k}\eta_{T,t}^{2,k\,\prime} \rightsquigarrow \begin{cases} \int_0^1 G'_{a,c}(r)\,(a'Z_a^\star)\, dB_0(r), & \text{under} \quad \text{LSTUR} \\ \int_0^1 G'_a(r)\,(a'Z_a^\star)\, dB_0(r), & \text{under} \quad \text{STUR} \end{cases}.$

(vii) $\frac{1}{kT^{\frac{3}{2}+2\gamma}}\sum_{t=1}^{T-k} z_t^k \eta_{T,t}^{2,k\,\prime} \rightsquigarrow \begin{cases} \int_0^1 (\Omega_{xa}a)\, G'_{a,c}(r) dr C_z^{-2}, & \text{under} \quad \text{LSTUR} \\ \int_0^1 (\Omega_{xa}a)\, G'_a(r) dr C_z^{-2}, & \text{under} \quad \text{STUR} \end{cases}.$

(viii) $\frac{1}{kT^{2+2\gamma}}\sum_{t=1}^{T-k} \eta_{T,t}^{2,k}\eta_{T,t}^{2,k\,\prime} \rightsquigarrow \begin{cases} \int_0^1 G_{a,c}(r) G'_{a,c}(r)\,(a'\Omega_{aa}a)\, dr C_z^{-2}, & \text{under} \quad \text{LSTUR} \\ \int_0^1 G_a(r) G'_a(r)\,(a'\Omega_{aa}a)\, dr C_z^{-2}, & \text{under} \quad \text{STUR} \end{cases}.$

(ix) *For any* $t = \lfloor Tr \rfloor$ *and* $r \in [0,1]$, *then*

$$-\frac{C_z}{kT^{\frac{1}{2}+\gamma}}\eta_{Tt}^{3,k} \rightsquigarrow \begin{cases} (a'\Sigma_{aa}a)\, G_{a,c}(r), & \text{under} \quad \text{LSTUR} \\ (a'\Sigma_{aa}a)\, G_a(r), & \text{under} \quad \text{STUR} \end{cases}. \tag{83}$$

(x) $\sum_{t=1}^{T-k} u_{0,t+k}\eta_{T,t}^{3,k\,\prime} = O_p\left(kT^{1+\gamma}\right)$, *therefore* $\frac{1}{\sqrt{k}T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T-k} u_{0,t+k}\eta_{T,t}^{3,k\,\prime} = o_p(1)$.

Lemma B.2 *Suppose Assumptions 1 and 2 hold. Under the rate condition* $\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0$,

$$\frac{1}{\sqrt{k}T^{\frac{1}{2}+\gamma}}\sum_{t=1}^{T-k} u_{0,t+k}\tilde{z}_t^{k\,\prime} \rightsquigarrow \mathcal{MN}\left(\mathbf{0}_{1\times n},\ \left(V_{zz}^{LH} + V_{\eta\eta}^{(2),LH} + V_{z\eta}^{(2),LH} + V_{z\eta}^{(2),LH\,\prime}\right)\sigma_{00}\right), \tag{84}$$

where

$$V_{zz}^{LH} := \Omega_{xx}C_z^{-2}, \tag{85}$$

$$V_{\eta\eta}^{(2),LH} := \begin{cases} (a'\Omega_{aa}a)\int_0^1 G_{a,c}(r)G_{a,c}'(r)drC_z^{-2}, & \text{under} \quad \text{LSTUR} \\ (a'\Omega_{aa}a)\int_0^1 G_a(r)G_a'(r)drC_z^{-2}, & \text{under} \quad \text{STUR} \end{cases}, \tag{86}$$

and

$$V_{z\eta}^{(2),LH} := \begin{cases} (\Omega_{xa}a)\int_0^1 G_{a,c}'(r)drC_z^{-2}, & \text{under} \quad \text{LSTUR} \\ (\Omega_{xa}a)\int_0^1 G_a'(r)drC_z^{-2}, & \text{under} \quad \text{STUR} \end{cases}. \tag{87}$$

Lemma B.3 *Suppose Assumptions 1 and 2 hold. Under the rate condition,* $\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0$,

(i) $\frac{1}{T^{1+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k z_t^{k\prime} \to_p -\frac{1}{2}\Omega_{xx}C_z^{-1}$.

(ii) $\frac{1}{T^{2+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k \eta_{T,t}^{1,k\prime} \rightsquigarrow \begin{cases} -\int_0^1 G_{a,c}(r)G_{a,c}'(r)drC_z^{-1}, & \text{under} \quad \text{LSTUR} \\ -\int_0^1 G_a(r)G_a'(r)drC_z^{-1}, & \text{under} \quad \text{STUR} \end{cases}.$

(iii) $\frac{1}{T^{\frac{3}{2}+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k \eta_{T,t}^{2,k\prime} = \begin{cases} -\int_0^1 G_{a,c}(r)G_{a,c}'(r)(a'dB_a(r))C_z^{-1} & \text{under} \quad \text{LSTUR} \\ -\int_0^1 G_a(r)G_a'(r)(a'dB_a(r))C_z^{-1} & \text{under} \quad \text{STUR} \end{cases}.$

(iv) $\frac{1}{T^{2+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k \eta_{T,t}^{3,k\prime} \rightsquigarrow \begin{cases} -\int_0^1 G_{a,c}(r)G_{a,c}'(r)(a'\Sigma_{aa}a)drC_z^{-1}, & \text{under} \quad \text{LSTUR} \\ -\int_0^1 G_a(r)G_a'(r)(a'\Sigma_{aa}a)drC_z^{-1}, & \text{under} \quad \text{STUR} \end{cases}.$

Lemma B.4 *Suppose Assumptions 1 and 2 hold. Under the rate condition that* $\frac{1}{T} + \frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0$

$$\frac{1}{T^{1+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k \tilde{z}_t^{k\prime} \rightsquigarrow \Upsilon, \tag{88}$$

where

$$\begin{aligned} \Upsilon := &\left(-\frac{1}{2}\Omega_{xx} - \int_0^1 G_{a,c}(r)G_{a,c}'(r)dr \cdot C - \frac{1}{2}\int_0^1 G_{a,c}(r)G_{a,c}'(r)\left(a'\Sigma_{aa}a\right)dr\right)\cdot C_z^{-1} \\ &- \int_0^1 G_{a,c}(r)G_{a,c}'(r)\left(a'dB_a(r)\right)C_z^{-1} \end{aligned} \tag{89}$$

in the LSTUR case, and

$$\Upsilon := \left(-\frac{1}{2}\Omega_{xx} - \frac{1}{2}\int_0^1 G_a(r)G_a'(r)\left(a'\Sigma_{aa}a\right)dr\right)\cdot C_z^{-1} - \int_0^1 G_a(r)G_a'(r)\left(a'dB_a(r)\right)C_z^{-1}, \tag{90}$$

in the STUR case.

Proofs of Theorems 3.3 and 3.4

Proof: These two proofs follow directly from the continuous mapping theorem, the mixed normality and asymptotic independence in Lemma B.2, and the weak convergence in Lemma B.4. ■

C Simulation & Empirical Results

Table 1: Empirical size of QR-IVX-Wald and IVX-Wald tests with two mixed regressors ($T = 200, c_1 = -1, a_1 = 1, a_2 = 0$)

	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	IVX-Wald
$c_2 = -1$										
$\gamma = 0.5$	0.087	0.054	0.046	0.033	0.031	0.039	0.038	0.028	0.072	0.022
$\gamma = 0.6$	0.092	0.055	0.050	0.044	0.041	0.041	0.045	0.038	0.076	0.034
$\gamma = 0.7$	0.105	0.051	0.052	0.056	0.045	0.044	0.042	0.046	0.085	0.049
$\gamma = 0.8$	0.105	0.060	0.066	0.053	0.049	0.046	0.053	0.053	0.076	0.060
$\gamma = 0.9$	0.097	0.073	0.067	0.068	0.055	0.064	0.059	0.071	0.077	0.061
$c_2 = 0$										
$\gamma = 0.5$	0.093	0.060	0.050	0.031	0.038	0.039	0.046	0.038	0.069	0.020
$\gamma = 0.6$	0.092	0.058	0.056	0.047	0.047	0.043	0.055	0.046	0.083	0.029
$\gamma = 0.7$	0.115	0.060	0.063	0.053	0.059	0.042	0.048	0.058	0.080	0.036
$\gamma = 0.8$	0.108	0.073	0.071	0.067	0.059	0.058	0.060	0.067	0.082	0.051
$\gamma = 0.9$	0.109	0.082	0.097	0.079	0.072	0.084	0.091	0.091	0.106	0.055
$c_2 = 0.2$										
$\gamma = 0.5$	0.103	0.066	0.049	0.044	0.036	0.043	0.055	0.041	0.075	0.015
$\gamma = 0.6$	0.100	0.064	0.057	0.051	0.054	0.048	0.062	0.056	0.088	0.027
$\gamma = 0.7$	0.120	0.064	0.072	0.061	0.076	0.050	0.058	0.064	0.087	0.033
$\gamma = 0.8$	0.122	0.089	0.084	0.081	0.071	0.076	0.077	0.084	0.102	0.043
$\gamma = 0.9$	0.129	0.100	0.115	0.109	0.102	0.104	0.115	0.122	0.135	0.051

[a] τ refers to the quantile index
[b] The number of replications is $1,000$.

Table 2: Empirical size of QR-IVX-Wald and IVX-Wald tests with two mixed regressors ($T = 500, c_1 = -1, a_1 = 1, a_2 = 0$)

	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	IVX-Wald
$c_2 = -1$										
$\gamma = 0.5$	0.072	0.054	0.042	0.038	0.042	0.037	0.034	0.044	0.059	0.024
$\gamma = 0.6$	0.075	0.053	0.042	0.048	0.045	0.034	0.037	0.046	0.059	0.032
$\gamma = 0.7$	0.079	0.048	0.049	0.047	0.045	0.034	0.037	0.042	0.056	0.040
$\gamma = 0.8$	0.087	0.056	0.052	0.048	0.050	0.034	0.053	0.048	0.069	0.058
$\gamma = 0.9$	0.073	0.064	0.054	0.057	0.056	0.051	0.058	0.065	0.081	0.062
$c_2 = 0$										
$\gamma = 0.5$	0.069	0.058	0.046	0.044	0.045	0.044	0.040	0.048	0.059	0.013
$\gamma = 0.6$	0.082	0.057	0.046	0.053	0.048	0.048	0.044	0.048	0.055	0.028
$\gamma = 0.7$	0.084	0.050	0.046	0.058	0.048	0.046	0.049	0.051	0.062	0.036
$\gamma = 0.8$	0.089	0.072	0.064	0.065	0.056	0.053	0.062	0.065	0.078	0.049
$\gamma = 0.9$	0.092	0.086	0.078	0.072	0.066	0.070	0.076	0.080	0.092	0.053
$c_2 = 0.2$										
$\gamma = 0.5$	0.074	0.057	0.053	0.051	0.048	0.045	0.043	0.053	0.057	0.015
$\gamma = 0.6$	0.084	0.064	0.058	0.055	0.049	0.048	0.047	0.056	0.067	0.029
$\gamma = 0.7$	0.095	0.063	0.066	0.072	0.062	0.055	0.055	0.063	0.072	0.040
$\gamma = 0.8$	0.116	0.091	0.076	0.083	0.076	0.068	0.069	0.080	0.092	0.048
$\gamma = 0.9$	0.142	0.124	0.114	0.098	0.103	0.107	0.105	0.114	0.123	0.048

[a] τ refers to the quantile index;
[b] The number of replications is $1,000$.

Table 3: Empirical size of LHIVX-Wald tests with a single regressor

		$c_1 = -1$ $a_1 = 0.5$	$c_1 = -1$ $a_1 = 1$	$c_1 = 0$ $a_1 = 0.5$	$c_1 = 0$ $a_1 = 1$	$c_1 = 1$ $a_1 = 0.5$	$c_1 = 1$ $a_1 = 1$
$T = 100$	$\gamma = 0.3$	0.022	0.023	0.026	0.021	0.037	0.028
	$\gamma = 0.4$	0.034	0.029	0.031	0.029	0.039	0.042
	$\gamma = 0.5$	0.034	0.037	0.031	0.038	0.044	0.045
	$\gamma = 0.6$	0.033	0.036	0.032	0.041	0.045	0.049
	$\gamma = 0.7$	0.035	0.035	0.042	0.045	0.046	0.054
$T = 200$	$\gamma = 0.3$	0.029	0.030	0.026	0.032	0.030	0.029
	$\gamma = 0.4$	0.040	0.034	0.029	0.035	0.033	0.028
	$\gamma = 0.5$	0.041	0.035	0.032	0.037	0.035	0.036
	$\gamma = 0.6$	0.040	0.037	0.033	0.039	0.040	0.037
	$\gamma = 0.7$	0.040	0.039	0.042	0.038	0.046	0.037

[a] The prediction horizon $k = T^{\nu}$ and $\nu = 0.75$;
[b] The number of replications is $1,000$.

Table 4: IVX-Wald and QR-IVX-Wald tests for univariate predictive regressions with monthly data (1952-2019)

	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	IVX-Wald
b/m	10.019***	1.818	2.378	1.547	0.320	1.814	4.810**	1.989	5.652**	0.031
d/e	3.316*	1.903	1.335	0.005	0.349	1.134	1.695	2.518	3.363*	0.470
dfy	5.635**	1.626	0.974	0.015	0.154	1.256	5.557**	9.956***	6.458**	0.003
d/p	1.196	0.018	0.776	7.355***	2.131	11.436***	22.133***	16.119***	0.913	0.247
d/y	0.001	0.154	0.350	7.531***	1.009	0.713	8.934***	2.960*	0.512	0.433
e/p	2.503	0.117	0.000	3.058*	0.241	0.552	0.552	0.072	0.918	0.003
infl	0.743	0.000	2.157	2.204	1.554	2.075	0.251	1.383	2.932*	1.200
lty	4.784**	17.160***	15.212***	8.560***	7.540***	8.080***	6.983***	2.799*	5.314**	7.712***
ntis	0.603	0.502	1.001	0.067	0.040	0.003	0.142	0.023	0.341	0.422
tbl	1.382	3.629*	4.303**	4.034**	4.439**	3.409*	3.437*	4.036**	7.633***	5.834**
tms	2.195	0.867	0.339	0.012	0.152	1.033	2.969*	1.059	2.474	0.173

[a] The sample period is January 1952 – December 2019.
[b] *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level, respectively.

Table 5: IVX-Wald and QR-IVX-Wald tests for univariate predictive regressions with quarterly data (1952-2019)

	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	IVX-Wald
b/m	1.313	0.106	0.000	0.207	0.233	6.429**	7.515***	3.185*	7.659***	0.496
d/e	0.770	2.951*	4.956**	0.133	1.305	0.849	3.145*	1.704	0.353	0.055
dfy	1.390	2.922*	1.660	0.208	0.376	0.241	4.499**	4.402**	6.824***	0.075
d/p	14.945***	0.091	0.452	1.278	22.798***	42.095***	25.023***	10.213***	4.260**	0.107
d/y	23.327***	0.728	1.251	5.471**	35.279***	34.225***	11.748***	14.280***	0.825	0.016
e/p	0.649	0.156	0.009	0.000	4.865**	11.720***	0.773	0.011	0.619	0.017
infl	0.402	0.856	0.265	0.330	0.123	0.809	1.140	1.451	0.605	0.146
lty	2.398	4.251**	3.730*	0.227	0.271	0.939	1.674	0.789	1.594	3.433*
ntis	0.001	0.029	0.431	0.484	0.802	0.666	0.222	0.210	0.223	0.187
tbl	0.023	2.084	2.596	2.932*	4.128**	4.119**	5.444**	11.708***	11.393***	2.500
tms	2.344	0.002	0.263	0.209	2.014	0.675	1.562	4.041**	3.653*	0.051

[a] The sample period is January 1952 – December 2019.
[b] *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level, respectively.

Table 6: IVX-Wald and QR-IVX-Wald tests for univariate predictive regressions with annual data (1952-2019)

	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	IVX-Wald
b/m	0.248	0.575	0.048	1.371	3.427*	5.859**	6.173**	8.267***	6.819***	2.053
d/e	0.433	0.382	0.425	0.029	0.003	0.424	0.041	0.008	0.012	0.154
dfy	0.178	0.150	1.548	0.287	0.954	1.425	0.891	0.802	4.627**	0.920
d/p	100.460***	20.636***	27.645***	4.037**	2.067	2.644	2.474	9.717***	0.412	0.000
d/y	12.029***	3.370*	1.808	1.951	7.779***	8.352***	2.643	5.013**	0.848	0.505
e/p	2.853*	2.779*	2.754*	1.523	1.780	1.602	0.634	0.523	1.298	0.070
infl	1.615	1.229	0.423	0.046	0.034	0.204	0.401	0.135	0.068	0.426
lty	0.088	0.073	0.496	0.246	0.682	0.409	0.055	0.108	6.784***	0.374
ntis	4.067**	0.953	1.217	0.952	0.881	0.581	1.535	0.295	0.000	0.067
tbl	1.381	0.570	0.569	0.006	0.033	0.164	0.038	0.000	1.697	0.235
tms	0.065	0.184	1.708	0.484	0.176	0.242	0.253	0.256	0.431	0.014

[a] The sample period is January 1952 – December 2019.
[b] *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level, respectively.

Table 7: IVX-Wald and QR-IVX-Wald tests for multivariate predictive regressions with monthly, quarterly and annual data (1952-2019)

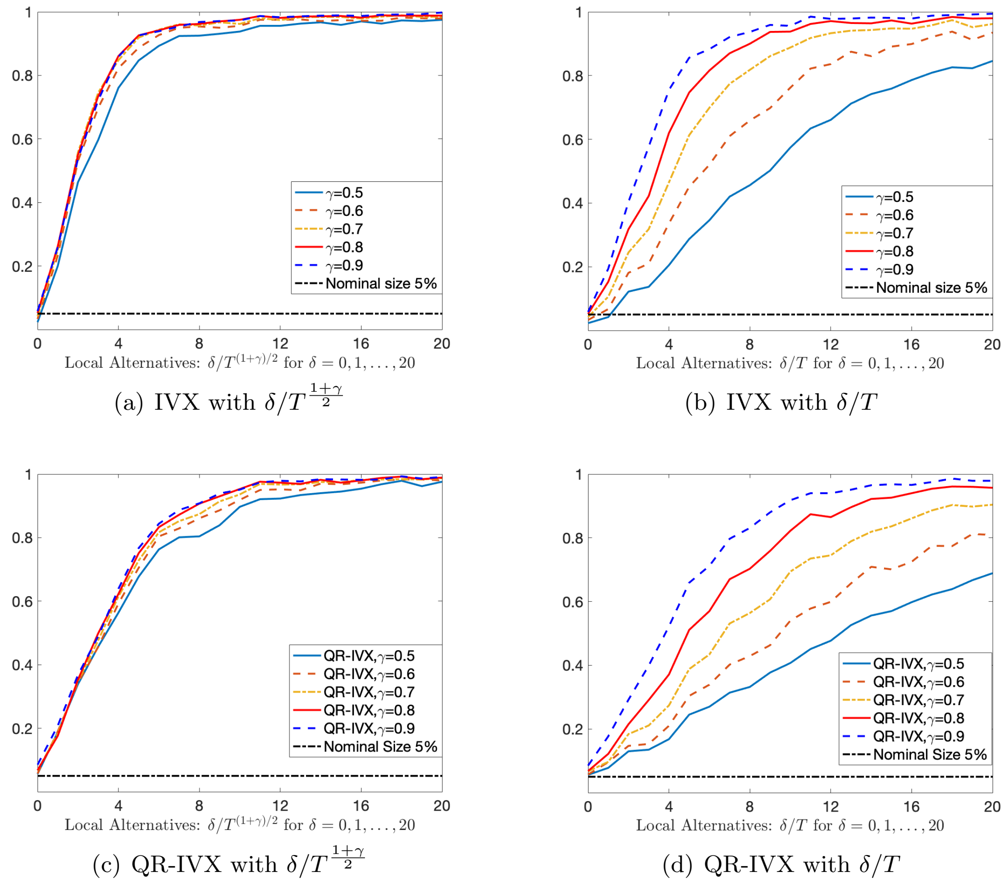
Monthly	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	IVX-Wald
d/p, tbl	3.490	4.697*	5.913*	3.987	5.153*	2.903	5.787*	4.264	6.852**	6.347**
e/p, tbl	0.791	3.724	3.213	4.724*	6.453**	3.819	5.844*	4.150	6.622**	2.919
d/p, b/m	13.769***	7.534**	5.114*	4.065	1.542	2.840	8.621**	9.639***	7.635**	0.611
d/p, d/e	6.273**	5.035*	3.899	0.353	1.537	0.022	1.151	1.842	3.947	0.673
Quarterly	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	IVX-Wald
d/p, tbl	1.660	1.690	2.951	2.582	4.202	6.337**	4.419	14.492***	15.522***	3.385
e/p, tbl	1.321	0.465	2.473	1.530	3.860	6.757**	4.875*	13.830***	12.167***	2.154
d/p, b/m	3.654	1.196	0.414	2.526	3.014	3.371	3.707	10.071***	5.455*	1.395
d/p, d/e	1.331	2.859	6.253**	0.796	1.476	0.449	0.439	1.876	0.579	0.373
Annual	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	IVX-Wald
d/p, tbl	0.277	1.577	2.182	1.497	0.868	1.464	1.925	2.256	0.363	0.966
e/p, tbl	0.562	1.515	2.518	1.157	0.900	1.677	0.426	0.642	2.169	0.677
d/p, b/m	0.355	2.946	1.967	2.182	2.420	2.795	13.178***	10.072***	6.026**	2.263
d/p, d/e	3.161	1.981	0.062	0.351	0.666	1.059	1.663	0.327	0.159	0.223

[a] The sample period is January 1952 – December 2019.
[b] *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level, respectively.

Table 8: LHIVX-Wald tests for univariate predictive regressions (1952-2019)

	Monthly			Quarterly			Annual		
	$k = 30$	$k = 60$	$k = 90$	$k = 30$	$k = 60$	$k = 90$	$k = 10$	$k = 20$	$k = 30$
b/m	2.166	0.382	0.446	0.319	0.095	0.147	0.340	0.026	1.131
d/e	0.017	0.479	0.148	0.398	1.815	0.006	0.001	0.088	0.286
dfy	0.649	0.549	0.021	0.075	0.315	0.640	0.033	1.419	0.755
d/p	1.333	0.086	0.012	0.000	0.070	0.115	0.049	0.022	1.736
d/y	1.536	0.112	0.014	0.028	0.052	0.018	0.009	0.139	2.130
e/p	0.509	0.106	0.131	0.132	0.788	0.160	0.060	0.077	1.891
infl	3.623*	6.527**	2.509	1.963	3.615*	0.881	0.368	0.005	0.162
lty	0.874	1.185	0.869	0.642	0.829	0.053	0.981	0.009	0.044
ntis	0.254	0.248	0.073	0.134	2.523	6.685***	0.471	2.691	0.112
tbl	0.664	3.751*	1.628	1.403	4.801**	0.997	1.804	0.232	1.120
tms	0.090	2.821*	0.589	0.748	9.601***	5.262**	1.027	1.332	6.016**

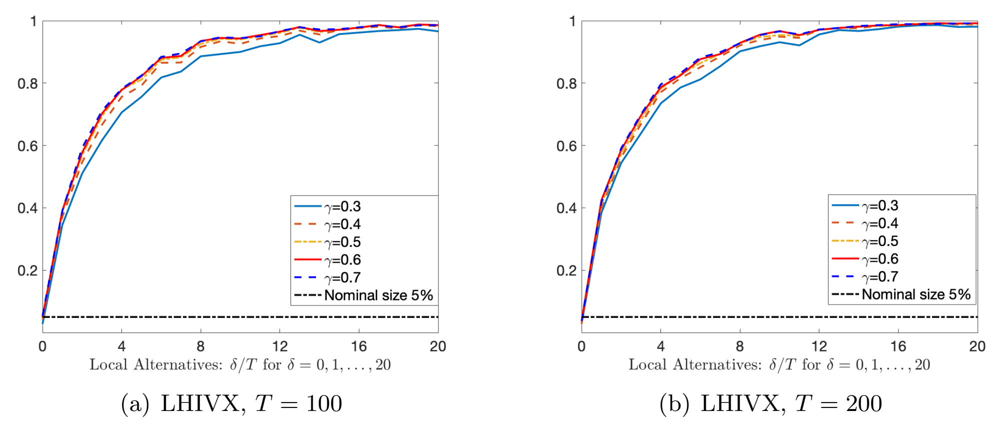
[a] The sample period is January 1952 – December 2019.
[b] *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level, respectively.



(a) IVX with $\delta/T^{\frac{1+\gamma}{2}}$ (b) IVX with δ/T

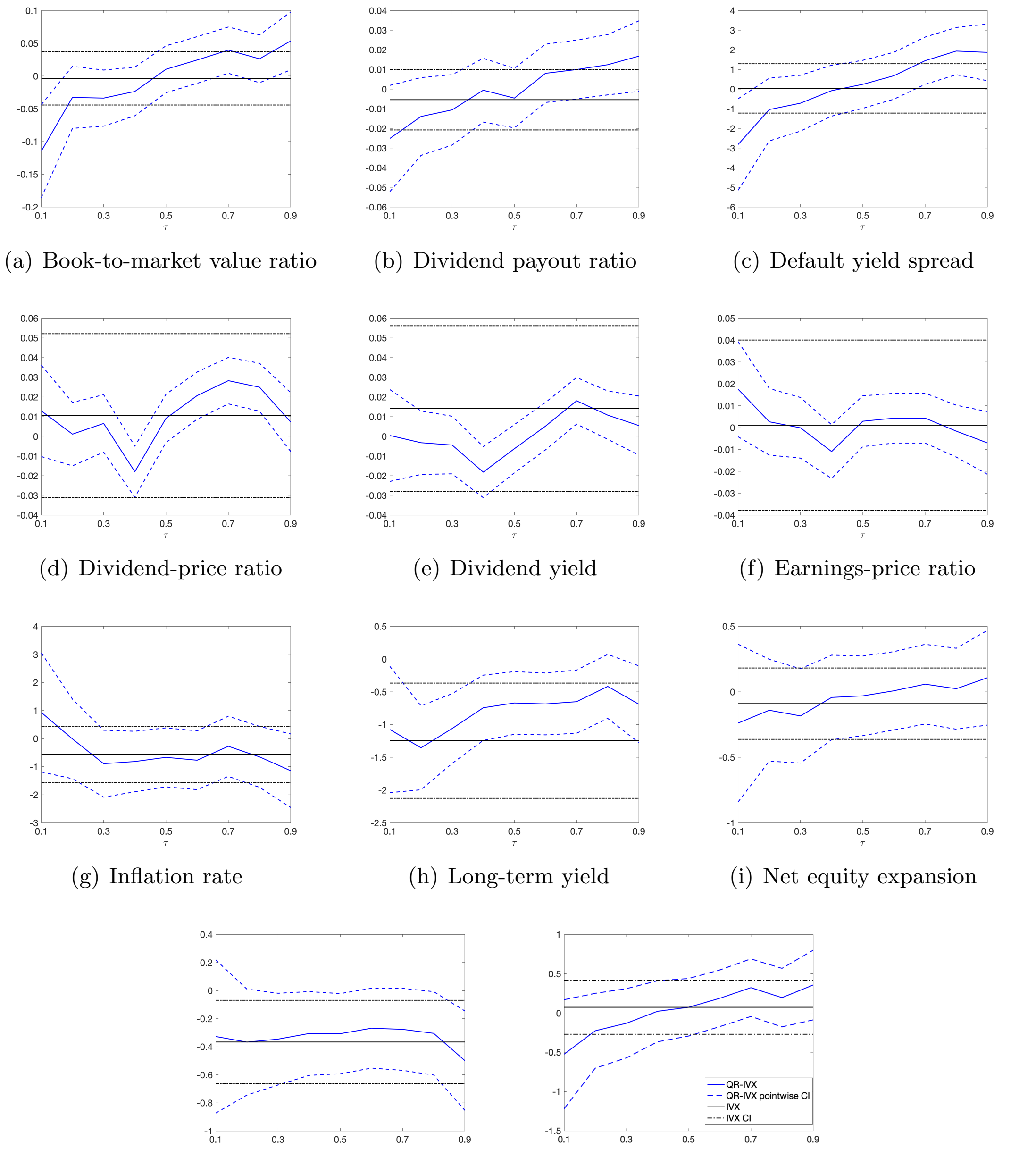
(c) QR-IVX with $\delta/T^{\frac{1+\gamma}{2}}$ (d) QR-IVX with δ/T

Figure 1: Local power functions of the QR-IVX-Wald and IVX-Wald tests with mixed-root regressors ($c_1 = -1, a_1 = 1, c_2 = 0, a_2 = 0, \alpha_2 = 0.75, T = 200$) for local deviations $\delta/T^{\frac{1+\gamma}{2}}$ and δ/T.



(a) LHIVX, $T = 100$ (b) LHIVX, $T = 200$

Figure 2: Power functions for the univariate LHIVX-Wald test with ($c_1 = 1, a_1 = 1, T = 100, 200$).



(a) Book-to-market value ratio (b) Dividend payout ratio (c) Default yield spread

(d) Dividend-price ratio (e) Dividend yield (f) Earnings-price ratio

(g) Inflation rate (h) Long-term yield (i) Net equity expansion

(j) T-bill rate (k) Term spread

Figure 3: IVX and QR-IVX slope estimates and 95% confidence intervals for univariate empirical predictive regressions with monthly data over 1952-2019.

References

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *The Review of Financial Studies*, 20(3):651–707.

Bekaert, G., Harvey, C. R., and Lundblad, C. (2007). Liquidity and expected returns: Lessons from emerging markets. *The Review of Financial Studies*, 20(6):1783–1831.

Bossaerts, P. and Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: what do we learn? *The Review of Financial Studies*, 12(2):405–428.

Cai, Z., Chen, H., and Liao, X. (2020). A new robust inference for predictive quantile regression. *Available at SSRN 3593817*.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.

Cenesizoglu, T. and Timmermann, A. (2008). Is the distribution of stock returns predictable? *Available at SSRN 1107185*.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics*, 33(2):806–839.

Chernozhukov, V., Fernández-Val, I., and Kaji, T. (2016). Extremal quantile regression: An overview. *arXiv preprint arXiv:1612.06850*.

Cogley, T. and Sargent, T. J. (2001). Evolving post-World War II US inflation dynamics. *NBER macroeconomics annual*, 16:331–373.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302.

Demetrescu, M., Georgiev, I., Rodrigues, P. M., and Taylor, A. R. (2020). Testing for episodic predictability in stock returns. *Journal of Econometrics*.

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.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fan, R. and Lee, J. H. (2019). Predictive quantile regressions under persistence and conditional heteroskedasticity. *Journal of Econometrics*, 213(1):261–280.

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.

Gonzalo, J. and Pitarakis, J.-Y. (2012). Regime-specific predictability in predictive regressions. *Journal of Business & Economic Statistics*, 30(2):229–241.

Granger, C. W. and Swanson, N. R. (1997). An introduction to stochastic unit-root processes. *Journal of Econometrics*, 80(1):35–62.

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.*

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2020). Taking stock of long-horizon predictability tests: Are factor returns predictable? *Working paper.*

Kothari, S. P. and Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics*, 44(2):169–203.

Lamont, O. (1998). Earnings and expected returns. *The Journal of Finance*, 53(5):1563–1587.

Lee, J. H. (2016). Predictive quantile regression with persistent covariates: IVX-QR approach. *Journal of Econometrics*, 192(1):105–118.

Lieberman, O. and Phillips, P. C. B. (2017). A multivariate stochastic unit root model with an application to derivative pricing. *Journal of Econometrics*, 196(1):99–110.

Lieberman, O. and Phillips, P. C. B. (2018). IV and GMM inference in endogenous stochastic unit root models. *Econometric Theory*, 34(5):1065–1100.

Lieberman, O. and Phillips, P. C. B. (2020a). Hybrid stochastic local unit roots. *Journal of Econometrics*, 215(1):257–285.

Lieberman, O. and Phillips, P. C. B. (2020b). Understanding temporal aggregation effects on kurtosis in financial indices. *Journal of Econometrics (forthcoming).*

Lin, Y. and Tu, Y. (2020). Robust inference for spurious regressions and cointegrations involving processes moderately deviated from a unit root. *Journal of Econometrics*, 219(1):52–65.

McCabe, B. P. and Smith, R. J. (1998). The power of some tests for difference stationarity under local heteroscedastic integration. *Journal of the American Statistical Association*, 93(442):751–761.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547.

Phillips, P. C. B. (1989). Partially identified econometric models. *Econometric Theory*, pages 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. (2021). Estimation and inference with near unit roots. *Cowles Foundation Discussion Paper No. 2304, Yale University.*

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 Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, pages 971–1001.

Rapach, D. E., Ringgenberg, M. C., and Zhou, G. (2016). Short interest and aggregate stock returns. *Journal of Financial Economics*, 121(1):46–65.

Tao, Y., Phillips, P. C. B., and Yu, J. (2019). Random coefficient continuous systems: Testing for extreme sample path behavior. *Journal of Econometrics*, 209(2):208–237.

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.

Xiao, Z. (2009). Quantile cointegrating regression. *Journal of Econometrics*, 150(2):248–260.

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.

Yoon, G. (2006). A note on some properties of STUR processes. *Oxford Bulletin of Economics and Statistics*, 68(2):253–260.