Uncovering regimes in out of sample forecast errors from predictive regressions*

Anibal Emiliano Da Silva Neto
University of Southampton
Department of Economics
aeds1n15@soton.ac.uk

Jesús Gonzalo
Universidad Carlos III de Madrid
Department of Economics
jgonzalo@uc3m.es

Jean-Yves Pitarakis
University of Southampton
Department of Economics
j.pitarakis@soton.ac.uk

Abstract

We introduce a set of test statistics for assessing the presence of regimes in out of sample forecast errors produced by recursively estimated linear predictive regressions that can accommodate multiple highly persistent predictors. Our tests statistics are designed to be robust to the chosen starting window size and are shown to be both consistent and locally powerful. Their limiting null distributions are also free of nuisance parameters and hence robust to the degree of persistence of the predictors. Our methods are subsequently applied to the predictability of the value premium whose dynamics are shown to be characterised by state dependence.

JEL: C12, C22, C53, C58.

Keywords: Predictive Regressions, Predictability, Out of Sample Forecast Errors, CUSUM, Thresholds.

*We wish to thank the Editor and two anonymous referees for valuable comments and suggestions. We also thank seminar participants at the Zaragoza Time Series Workshop 2018, Bilgi University Workshop on Advances in Econometric Methods 2018, Southampton Econometrics Workshop 2018, University of Cologne, University of Essex and University of Nottingham; Pitarakis thanks the British Academy for financial support through grant SRG170220.

1 Introduction

A vast body of recent empirical research documented the presence of state dependence in the forecast errors produced by models used to generate forecasts of a broad range of economic and financial variables such as stock and bond returns, commodity returns, rates of inflation, currency returns among many others. State dependence in this context takes the form of forecast errors having different quality characteristics such as lower variances in periods of economic recessions versus expansions. In Golez and Koudijs (2017) for instance the authors considered century long stock market data and documented the considerable strengthening of the in-sample and out of sample predictive power of dividend yields for stock returns during recessions. Chauvet and Potter (2013) remarked that predictability of output growth is much harder during recessions while Gargano et al. (2017) established that commodity returns are predictable using macroeconomic information but solely during recessions.

This state dependence in the behaviour of forecast errors has been typically documented through a descriptive comparison of prediction errors (e.g. lower MSEs during recessions than expansions) or the use of recession dummies within the underlying forecasting models. Numerous papers concerned with the predictability of the equity premium with valuation ratios documented important differences in out of sample goodness of fit metrics across NBER business cycle dates (see Li and Tsiakas (2016), Rapach, Strauss and Zhou (2010) amongst others).

The main goal of this paper is to introduce formal diagnostic tools for explicitly testing for the presence of broadly defined regimes in the out-of-sample prediction errors generated from predictive regression models. We are interested in both the levels of forecast errors and their squares as considering the two series can convey useful information on both misspecification issues and regime specificity in MSEs. Rather than thinking of regimes as matching business cycle dates we take a broader view of the notion of state dependence and associate regimes with observed proxies of the state of the economy exceeding or falling below particular levels. Our proposed methods require solely the computation of recursive least squares residuals which are then used within a CUSUM type construct and are therefore very easily implementable. Our operating framework is also flexible enough to accommodate predictive regressions with multiple highly persistent predictors of possibly different persistence strengths. Suppose for instance that one wishes to evaluate the predictability of the equity premium with the commonly used Goyal and Welch predictors (Welch and Goyal (2008), Goyal and Welch (2014)). These include quantities such as dividend yields, price-to-earnings ratios, interest rates all known to be highly persistent variables with potentially different degrees of persistence and typically modelled as nearly integrated processes with a nuisance parameter that parameterises persistence strength. How does one go about formally testing whether forecasts generated from such models lead to forecast errors that behave differently across the business cycle?

The issue is of great practical importance as the presence of regime specificity in prediction errors would call for a reassessment of the models used to generate forecasts and in particular motivate a switch to nonlinear specifications that are explicitly able to capture episodic predictability as for instance in Gonzalo and Pitarakis (2012, 2017) where the authors considered the inclusion of threshold effects within predictive regressions driven by a single highly persistent predictor. Such piecewise linear structures are particularly convenient as they allow the forecaster to control the particular indicator used for proxying economic times or more generally sentiment. As such they are not necessarily restricted to a rigid regime structure dictated by formal externally provided business cycle dates. We view the testing procedures introduced in this paper as useful practical diagnostic tools that can be used to motivate the explicit inclusion of regime dependence within the predictive model itself. Although post-diagnostic re-evaluation issues are beyond the scope of this paper such specifications have been shown to lead to considerable gains in prediction accuracy as demonstrated in an in-sample and single predictor based equity premium forecasting context in Gonzalo and Pitarakis (2012, 2017). More recent research by Farmer et al. (2018) also explores the idea of *pockets of predictability* by proposing novel non-parametric methods to model and detect such behaviour and offering interesting insights and formal evidence on its causes in the context of stock returns. This notion of episodic predictability has also been recently revisited via new techniques in Demetrescu et al. (2020).

In the context of predictive regressions an important issue that has attracted considerable attention in the literature is the sensitivity of asymptotic distributions to DGP parameterisations and to the non-centrality parameter used to model high persistence in particular, rendering the conduct of inferences difficult (see for instance Campbell and Yogo (2006), Jansson and Moreira (2006), Breitung and Demetrescu (2015), Kostakis et al. (2015), Phillips (2015), Pitarakis (2017), Georgiev et al. (2018)). Although the same problem also arises in our present setting our proposed methods are able to accommodate unknown persistence in addition to being able to handle the presence of multiple predictors without the need to appeal to instrumental variable, bootstrapping or Bonferroni type methods.

The structure of the paper is as follows. Section 2 introduces our main operating model together with the proposed test statistics. Section 3 develops the asymptotic theory of our tests together with their consistency and local power properties. Section 4 investigates their finite sample size and power properties. Section 5 applies our methodology to the predictability of the value premium and Section 6 concludes. All proofs are relegated to the appendix.

2 The Forecasting Model and Test Statistics

Our baseline specification is given by the following linear multiple predictive regression

$$y_{t+1} \quad = \quad \beta_0 + \boldsymbol{x}_t' \boldsymbol{\beta}_1 + u_{t+1} \tag{1}$$

where $\boldsymbol{x}_t$ is a p-vector of highly persistent predictors parameterised as

$$\boldsymbol{x}_t \quad = \quad \left(I_p - \frac{\boldsymbol{C}}{T}\right) \boldsymbol{x}_{t-1} + \boldsymbol{v}_t \tag{2}$$

with $\boldsymbol{C} = diag(c_1, \dots, c_p)$, $c_i > 0$ for $i = 1, \dots, p$ and u_t and $\boldsymbol{v}_t$ denoting stationary disturbances. For subsequent notational purposes it is also convenient to reformulate (1) as $y_{t+1} = \boldsymbol{w}_t' \boldsymbol{\beta} + u_{t+1}$ with $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')$, $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{1p})'$ and $\boldsymbol{w}_t = (1, \boldsymbol{x}_t)'$. In order to use (1) for out of sample forecast evaluation purposes we focus on a recursive least squares based approach whereby the model is re-estimated within an expanding window. More specifically, letting $\hat{\boldsymbol{\beta}}_t = (\sum_{s=1}^t \boldsymbol{w}_{s-1} \boldsymbol{w}_{s-1}')^{-1} (\sum_{s=1}^t \boldsymbol{w}_{s-1} y_s)$ denote the least squares estimator of $\boldsymbol{\beta}$ obtained using data up to time period t the one-step ahead forecast of y made at time t is obtained as $\hat{y}_{t+1|t} = \boldsymbol{w}_t' \hat{\boldsymbol{\beta}}_t$, leading to the forecast error sequence

$$e_{t+1|t} \quad = \quad y_{t+1} - \boldsymbol{w}_t' \hat{\boldsymbol{\beta}}_t, \quad t = k, \dots, T-1. \tag{3}$$

As it stands the above approach for generating predictions assumes an initially available training sample of say k observations used to initiate the recursions so that predictions can then be generated over the remaining $T - k$ periods by re-estimating the model with an additional observation in each step. Given a choice of k, say k_0, recursive forecasts are obtained for $t = k_0, k_0 + 1, \dots, T - 1$. Throughout this paper the initial estimation sample is viewed as a fraction $\pi \in (0, 1)$ of the full sample by setting $k = [T\pi]$, the largest integer smaller than $T\pi$, so that the sequence of out of sample forecast errors $\{e_{t+1|t}\}_{t=k}^{T-1}$ is understood to be of length $T - k$.

Given our operating model in (1) our main objective is to develop a simple approach for assessing the presence of economically meaningful regimes in the forecast errors in (3) and their squares. Throughout this paper we take a broad view of the notion of state dependence, not necessarily equating it with precise business cycle phases. More specifically we will be interested in assessing the behaviour of the $e_{t+1|t}$'s and $e_{t+1|t}^2$'s across different regimes driven by an observable threshold variable lying above or below an unknown cut-off. The choice of the specific threshold variable is naturally dictated by the application of interest. Commonly used options for capturing business cycle movements and the state of the economy include the growth rate in industrial production or GDP, diffusion indices combining multiple macroeconomic indicators, sentiment and confidence indicators etc.

Depending on the application and question of interest one may focus solely on assessing the presence of regimes in $e_{t+1|t}$ if interest is about uncovering potential misspecification in the conditional mean of the forecasting model or alternatively the focus could be on $e_{t+1|t}^2$ if one simply wishes to uncover potential state dependence in out of sample MSEs with both scenarios being viewed as departures of interest from (1). As an omitted regime in conditional means may contaminate forecast error variances it will also be important to assess the information provided by both tests jointly, an issue we explore comprehensively below when we evaluate the large sample properties of our tests under departures that account for regime changes in the conditional means and error variances.

Given the potential sensitivity of the accuracy of forecasts to the choice of the length k of the initial sample used for initiating the recursive forecasts, in what follows we will be interested in assessing the presence of regimes in the $e_{t+1|t}$'s and $e_{t+1|t}^2$'s under both a fixed/given $k = k_0$ scenario commonly used in practice but also a more general setting whereby $k = [T\pi]$ is allowed to vary over an interval $[\pi_a, \pi_b] \subset (0,1)$. The motivation of this latter framework is to render inferences robust to data mining along the lines of Rossi and Inoue (2012). The practical relevance of this issue has been recently highlighted in the context of the predictability of the equity premium in Kolev and Karapandza (2017) where the authors demonstrate that for a given set of predictors alternative data splits may lead to contradictory outcomes about return predictability.

Our proposed inferences about the presence of regimes within $e_{t+1|t}$ and $e^2_{t+1|t}$ will rely on suitably normalised versions of functionals of the following two quantities

$$C_{1T}(k,\gamma) = \sum_{t=k}^{T-1} (e_{t+1|t} - \bar{e}_{T-k}) \mathbb{I}(q_t \leq \gamma) \tag{4}$$

$$C_{2T}(k,\gamma) = \sum_{t=k}^{T-1} (e^2_{t+1|t} - \bar{\tau}^2_{T-k}) \mathbb{I}(q_t \leq \gamma) \tag{5}$$

where q_t denotes the threshold variable, $\bar{e}_{T-k} = \sum_{t=k}^{T-1} e_{t+1|t}/(T-k)$ and $\bar{\tau}^2_{T-k} = \sum_{t=k}^{T-1}(e_{t+1|t} - \bar{e}_{T-k})^2/(T-k)$. Note that the quantity $\bar{e}_{T-k}$ is maintained in (4) as the $e_{t+1|t}$'s should not be confused with full sample residuals which would have an exact zero mean. Throughout this paper we also write

$$q_t = \mu_q + u_{qt} \tag{6}$$

and this threshold variable is understood to be stationary with marginal distribution function $F(.)$ so that when necessary and convenient we make use of the property $\mathbb{I}(q_t \leq \gamma) \equiv \mathbb{I}(F(q_t) \leq \lambda)$ with $F(q_t) \sim U[0,1]$ and refer to the threshold parameter as γ or $\lambda \equiv F(\gamma)$ interchangeably.

Note that (4) and (5) are indexed by both the unknown threshold parameter γ as well as $k = [T\pi]$ which captures the location of the initial sample size used to initiate the recursive forecasts. As highlighted in Rossi and Inoue (2012) forecast accuracy can vary greatly across alternative choices of π. Our formulations in (4)-(5) allow us to construct test statistics that take this dependence on π into account and hence lead to inferences that are less prone to data mining. Nevertheless in what follows we will consider both scenarios (i.e. π fixed and given, say $\pi = \pi_0$ and $\pi \in [\pi_a, \pi_b] \subset (0,1)$).

We consider two alternative functionals of the $C_{iT}(\pi, \lambda)$'s across the two scenarios on π as formulated in the following test statistics. For the scenario where k is taken as given, say $k = k_0 = [T\pi_0]$ we define

$$Sup_{iT} \equiv \sup_{\lambda \in \Lambda} \left| \frac{C_{iT}(\pi_0, \lambda)}{\sqrt{T}\hat{\phi}_i} \right| \quad i = 1,2 \tag{7}$$

$$Ave_{iT} \equiv \underset{\lambda \in \Lambda}{ave} \left(\frac{C^2_{iT}(\pi_0, \lambda)}{T\hat{\phi}^2_i} \right) \quad i = 1,2 \tag{8}$$

where the indexing $i = 1,2$ distinguishes between the statistic implemented on the level of the forecast errors and their squares respectively and with $\hat{\phi}^2_1 = \sum_{t=k}^{T-1}(e_{t+1|t} - \bar{e}_{T-k})^2/T$ and $\hat{\phi}^2_2 = \sum_{t=k}^{T-1}(e^2_{t+1|t} - \bar{\tau}^2_{T-k})^2/T$. Note that for notational simplicity we have suppressed the dependence of Sup_{iT} and Ave_{iT} on λ.

In order to robustify our inferences to the specific choice of k_0 we also consider a framework where k is allowed to take a broad range of values (e.g. $\pi \in \Pi = [0.25, 0.75]$) by proceeding à la Rossi and Inoue (2012). For this purpose we introduce the following alternative test statistic formulations indexed by both π and λ

$$SupSup_{iT} \equiv \sup_{\pi \in \Pi} \sup_{\lambda \in \Lambda} \left| \frac{C_{iT}(\pi, \lambda)}{\sqrt{T}\hat{\phi}_i} \right| \quad i = 1, 2 \tag{9}$$

$$AveAve_{iT} \equiv \underset{\pi \in \Pi}{ave}\ \underset{\lambda \in \Lambda}{ave} \left(\frac{C_{iT}^2(\pi, \lambda)}{T\hat{\phi}_i^2} \right) \quad i = 1, 2. \tag{10}$$

The above statistics bear strong resemblance with traditional CUSUM (under $i = 1$) and CUSUMSQ (under $i = 2$) formulations commonly used in the changepoint literature and developed in the early work of Page (1954), Brown et al. (1975) amongst others. Instead of cumulating the quantity of interest up to a potential changepoint we here focus on its random sum as dictated by the magnitude of q_t. Although such test statistics have often been viewed as exploratory tools for assessing parameter stability in regression models and were developed with no particular alternative in mind we have here adapted them to our specific context of threshold effects in forecast errors and their variances and therefore expect them to display good power properties against such scenarios. More specifically, the model against which we will be interested in confronting the out of sample forecast errors estimated from (1) is given by

$$y_{t+1} = \boldsymbol{\beta}'\boldsymbol{w}_t + \delta_0 \mathbb{I}(q_t > \gamma_0) + \boldsymbol{\delta}_1'\boldsymbol{x}_t \mathbb{I}(q_t > \gamma_0) + u_{t+1} \tag{11}$$

where δ_0 and $\boldsymbol{\delta}_1$ capture the presence of threshold effects associated with the intercept and slope parameters respectively. We initially concentrate on departures from (1) towards instabilities induced solely by conditional mean parameters as in (11) but subsequently also explore scenarios where the variance of the $u_t's$ is itself characterised by threshold effects triggered by q_t. This variance based state-dependence may occur in isolation or in conjunction with state-dependence in the slope parameters.

3 Limiting Distributions and Asymptotic Power Properties

Limiting Distributions

Our main objective here is to obtain the limiting distributions of the test statistics based on $C_{1T}(\pi, \lambda)$ and $C_{2T}(\pi, \lambda)$ when the underlying model is given by the linear predictive regression in (1). As our inferences involve both the level and squares of forecast errors we formulate our operating assumptions accordingly as the latter scenario requires further restrictions on the dynamics of the u_t^2 sequence. Assumption **A1** below outlines the probabilistic environment under which we establish the large sample properties of (7)-(10) for $i = 1$.

Assumption A1. *(i)* $\boldsymbol{v}_t = \boldsymbol{\Psi}(L)\boldsymbol{\epsilon}_{vt}$ *with* $\boldsymbol{\Psi}(L) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j L^j$ *such that* $\sum_{j=0}^{\infty} \boldsymbol{\Psi}_j$ *has full rank,* $\boldsymbol{\Psi}_0 = I_p$ *and* $\sum_{j=0}^{\infty} ||\boldsymbol{\Psi}_j|| < \infty$*. (ii)* $\boldsymbol{\eta}_{1t} = (u_t, \boldsymbol{\epsilon}_{vt})'$ *is a martingale difference sequence with respect to the filtration*

$\mathcal{F}_{1,t} = \sigma(\boldsymbol{\eta}_{1s}, u_{qs}|s \leq t)$ satisfying $E[\boldsymbol{\eta}_{1t}\boldsymbol{\eta}'_{1t}|\mathcal{F}_{1,t-1}] = \boldsymbol{\Sigma}_{\eta_1} > 0$ and $E||\boldsymbol{\eta}_{11}||^4 < \infty$. (iii) The probability density function $f_q(.)$ of the threshold variable $q_t = \mu_q + u_{qt}$ is bounded away from zero and ∞ over each bounded set. (iv) The zero mean sequence $\{u_{qt}\}$ is strictly stationary, ergodic, strong mixing with mixing numbers α_m such that $\sum_{m=1}^{\infty} \alpha^{\frac{1}{2}-\frac{1}{r}} < \infty$ for some $r > 2$.

Assumption **A1** mimics closely the environment considered in Gonzalo and Pitarakis (2012, 2017) and excluding the probabilistic properties of q_t has been the operating standard in the linear predictive regression literature. Both $\boldsymbol{v}_t$ and q_t are allowed to display a rich dependence structure while u_t is restricted to be a conditionally homoskedastic martingale difference sequence. The threshold variable is assumed to be stationary throughout and its innovations possibly contemporaneously correlated with the shocks driving the predictive regression in (1). The strong mixing assumption on q_t is in line with Caner and Hansen (2001) and allows the sequence to follow very general stationary ARMA type specifications. As our focus is on using q_t to capture business cycle movements stationarity of q_t is naturally intuitive in the context of this paper. Finally, it is also important to point out that the covariance between the $u_t's$ and the shocks $\boldsymbol{\epsilon}_{vt}$ associated with the predictors, say $\boldsymbol{\Sigma}_{\eta_1} = \{\{\sigma_u^2, \boldsymbol{\sigma}'_{u\boldsymbol{\epsilon}_v}\}, \{\boldsymbol{\sigma}_{u\boldsymbol{\epsilon}_v}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_v\boldsymbol{\epsilon}_v}\}\}$, can be non-diagonal allowing them to be correlated as it is commonly observed in applications involving returns and dividend yields for instance.

An implication of the above assumptions is that a functional central limit theorem (FCLT) holds for $\boldsymbol{z}_t = (u_t, u_t I(q_{t-1} \leq \lambda), \boldsymbol{v}_t)'$ which we write as

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \boldsymbol{z}_t \quad \Rightarrow \quad (B_1(r), B_1(r,\lambda), \boldsymbol{B}_v(r))' \equiv BM(\Omega) \tag{12}$$

where $\Omega = \sum_{=-\infty}^{\infty} E[\boldsymbol{z}_0\boldsymbol{z}'_k] > 0$. Here $B_1(r,\lambda)$ is a two-parameter Brownian Motion i.e. a zero mean Gaussian process with covariance kernel $\sigma_u^2(r_1 \wedge r_2)(\lambda_1 \wedge \lambda_2)$ as introduced in a related context in Caner and Hansen (2001). Our assumptions under **A1**(ii) also imply a particular structure for Ω without forcing it to be diagonal as both serial correlation and heteroskedasticity are ruled out from the dynamics of the $u_t's$. More specifically we can formulate Ω as

$$\Omega = \begin{bmatrix} \sigma_u^2 & \lambda\sigma_u^2 & \boldsymbol{\sigma}'_{u\boldsymbol{\epsilon}_v}\Psi(1) \\ \lambda\sigma_u^2 & \lambda\sigma_u^2 & \lambda\boldsymbol{\sigma}'_{u\boldsymbol{\epsilon}_v}\Psi(1) \\ \boldsymbol{\sigma}_{u\boldsymbol{\epsilon}_v}\Psi(1) & \lambda\boldsymbol{\sigma}_{u\boldsymbol{\epsilon}_v}\Psi(1) & \Psi(1)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_v\boldsymbol{\epsilon}_v}\Psi(1)' \end{bmatrix} \tag{13}$$

and for later use we also write $B_1(r,\lambda) = \phi_1 W_1(r,\lambda)$ with $\phi_1^2 \equiv \sigma_u^2 \equiv E[u_t^2]$ and $W_1(r,\lambda)$ a two parameter standard Brownian Motion.

Assumption **A2** below introduces the conditions under which we obtain the limiting distributions of (7)-(10) for $i = 2$ based on the squared forecast errors. In order to handle the asymptotics associated with the use of squared CUSUMs as in (5) we supplement Assumption **A1** with additional restrictions involving the dynamics of the u_t^2 sequence and its interactions with the remaining random disturbances in the system.

Assumption A2. *Assumption* **A1** *continues to hold with* $\boldsymbol{\eta}_{1t}$ *replaced with* $\boldsymbol{\eta}_{2t} = (u_t, u_t^2 - \sigma_u^2, \boldsymbol{\epsilon}_{vt})'$.

We note that Assumption **A2** essentially imposes further conditional and unconditional moment restrictions on the u_t sequence. This is needed to ensure that an FCLT type result holds for the marked empirical process $Y_T(r,\lambda) = T^{-1/2}\sum_{t=1}^{[Tr]}(u_t^2 - \sigma_u^2)I(q_t \leq \lambda)$. Assumption **A2** ensures for instance that $E[(u_t^2 - \sigma_u^2)I(q_t \leq \lambda)] = \lambda\phi_2^2$ for $\phi_2^2 = E[(u_t^2 - \sigma_u^2)]^2$ while the finiteness of the fourth moments of $(u_t^2 - \sigma^2)$ is needed to establish the stochastic equicontinuity of $Y_T(r,u)$ (see Caner and Hansen (2001)). These assumptions ensure for instance that

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Tr]}(u_t^2 - \sigma_u^2)\mathbb{I}(q_{t-1} \leq \lambda) \quad \Rightarrow \quad B_2(r,\lambda) \equiv \phi_2 W_2(r,\lambda) \tag{14}$$

where $W_2(r,\lambda)$ is a two parameter standard Brownian Motion.

We initially focus on the large sample behaviour of (7)-(8) for $\pi = \pi_0$, taking the starting point of the expanding window used to initiate forecasts as given. Their limiting behaviour is summarised in Proposition 1 below.

Proposition 1. *Under assumptions* **A1** *for* $i = 1$ *and under assumptions* **A2** *for* $i = 2$ *we have as* $T \to \infty$

$$Sup_{iT} \quad \Rightarrow \quad \sup_{\lambda \in \Lambda}|W_i^0(\lambda)| \quad i = 1,2 \tag{15}$$

$$Ave_{iT} \quad \Rightarrow \quad \int_0^1 W_i^0(\lambda)^2 d\lambda \quad i = 1,2 \tag{16}$$

with $W_i^0(\lambda)$ *denoting a standard Brownian Bridge process.*

It is here interesting to note that given our operating model in (1)-(2) the above limiting distributions are free of any nuisance parameters, including the magnitudes of the underlying non-centrality parameters appearing in **C**. This is quite a unique outcome in the time series literature involving nearly integrated processes where the handling of the unknown $c_i's$ has become an important topic in itself. Also noteworthy is the fact that the limiting distributions are the same regardless of whether we implement our tests on the $e_{t+1|t}$'s or $e_{t+1|t}^2$'s.

Another important and convenient feature of (15)-(16) is that their tabulations are readily available in the literature, including the possibility of obtaining exact p-values. These distributions are well defined for $\lambda \in [0,1]$ (see Billingsley (1986)). In the case of (15) the 10%, 5% and 1% cutoffs are given by 1.224, 1.358 and 1.628. For the distribution in (16) the corresponding cutoffs are 0.347, 0.461 and 0.744. It is also worth pointing out that restricting mildly the $[0,1]$ intervals by taking the supremum in (15) over a subset such as $[0.1, 0.9]$ or $[0.2, 0.8]$ leads to almost identical critical values (e.g. the 1.224 cut-off decreases to 1.222 under $\lambda \in [0.2, 0.8]$ while under $[0.1, 0.9]$ it remains unchanged at the chosen precision level).

Next, we focus on the case where the practitioner does not wish to take a stance on where to start the build-up of the recursive forecast errors. The parameter π is now allowed to be such that $\pi \in \Pi$ so that the test statistics are evaluated for each possible magnitude of π (or k), say for instance starting from the 25% of the sample up to 75% of the sample. The test statistics are now given by (9)-(10) and their limiting distribution is summarised in Proposition 2 below.

Proposition 2. *Under Assumptions* **A1** *for* $i = 1$ *and under Assumptions* **A2** *for* $i = 2$ *we have as* $T \to \infty$

$$SupSup_{iT} \Rightarrow \sup_{\pi \in \Pi} \sup_{\lambda \in [0,1]} \left| \frac{W_i(1-\pi,\lambda) - \lambda W_i(1-\pi,1)}{\sqrt{1-\pi}} \right| \quad i = 1,2 \tag{17}$$

$$AveAve_{iT} \Rightarrow \frac{1}{\pi_b - \pi_a} \int_{\Pi} \int_0^1 \left(\frac{W_i(1-\pi,\lambda) - \lambda W_i(1-\pi,1)}{\sqrt{1-\pi}} \right)^2 d\pi d\lambda \quad i = 1,2 \tag{18}$$

with $W_i(1-\pi,\lambda)$ *denoting a two-parameter standard Brownian Motion.*

As in Proposition 1 and for notational simplicity we have here suppressed the dependence of $SupSup_{iT}$ and $AveAve_{iT}$ on λ and π. We continue to note that the limiting distributions in (17)-(18) are free of nuisance parameters. Both distributions are functionals of a process of the type $K(\zeta,\lambda) = W(\zeta,\lambda) - \lambda W(\zeta,1)$ commonly known as a Kiefer-Müller process which is a zero mean Gaussian processes with covariance kernel $Cov[K(\zeta_1,\lambda_1), K(\zeta_2,\lambda_2)] = (\zeta_1 \wedge \zeta_2)(\lambda_1 \wedge \lambda_2 - \lambda_1\lambda_2)$.

Although free of nuisance parameters it is clear that critical points of these distributions will depend on the interval $\Pi = [\pi_a, \pi_b]$. Tabulations of (17)-(18) are to our knowledge not available in the literature. To approximate these distributions we have generated a sequence of i.i.d random variables of length $T = 500$ that are uniformly distributed on $[0,1]$. $K(1-\pi,\lambda)$ is then approximated by $\sum_{t=[T\pi]}^{T}(\mathbb{I}(\mathcal{U}_t \leq \lambda) - \lambda)/\sqrt{T}$. Equivalently, $K(1-\pi,\lambda)/\sqrt{1-\pi}$ is approximated via $\sum_{t=[T\pi]}^{T}(\mathbb{I}(\mathcal{U}_t \leq \lambda) - \lambda)/\sqrt{T - [T\pi]}$. The supremum and average of the approximating process is obtained by maximising over π and λ with the process being repeated across $N = 10000$ replications. Key quantiles of (17)-(18) are presented in Table 1 across a selection of Π intervals.

Consistency and Local Power Properties

Having established the asymptotic distributions of our test statistics when the underlying model is given by (1) we next assess their ability to detect fixed as well as local departures from (1) driven by the presence of threshold effects as in (11). In the case of local departures $\boldsymbol{\delta}$ is understood to be parameterised as $\boldsymbol{\delta}_T = (\delta_{0T}, \boldsymbol{\delta}_{1T}')'$ with $\boldsymbol{\delta}_T \to 0$ at suitable rates as $T \to \infty$.

We initially focus on the consistency properties of our four test statistics establishing their large sample behaviour under a fixed magnitude of $\boldsymbol{\delta}$. The main result is summarised in Proposition 3 below.

Proposition 3. *Under (11) with* $\boldsymbol{\delta}_T = \boldsymbol{\delta} \neq 0$ *and the same assumption structure as in Propositions 1-2* $\{Sup_{iT}, Ave_{iT}, SupSup_{iT}, AveAve_{iT}\}$ *diverge to infinity for* $i = 1$ *and provided that* $\lambda_0 \neq 0.5$ *for* $i = 2$.

Proposition 3 establishes the consistency of all four test statistics, noting also that their stated divergence occurs regardless of whether solely the intercept δ_0, the slopes $\boldsymbol{\delta}_1$ or both are allowed to shift. There is one instance however under which our squared error based statistics (i.e. for $i = 2$) will display little power. This happens when the true threshold parameter is such that $P(q_t \leq \gamma_0) = P(q_t > \gamma_0)$, equivalently when $\lambda_0 = 0.5$. A similar phenomenon has also been documented in the context of the changepoint literature and inferences based on CUSUMSQ type statistics (see for instance Theorem 1 in Deng and Perron (2008)). It is also important to point out that the power loss that occurs under $\lambda_0 = 0.5$ when dealing with test statistics involving squared forecast errors is not unique to our CUSUMSQ type formulations as the same phenomenon will also occur if one were to use a Wald type statistic instead (e.g. a Wald statistic for detecting shifts in the mean of the $e^2_{t+1|t}$ sequence).

We next focus on the local power properties of our test statistics. For the simplicity of the exposition and with no loss of generality we concentrate our attention on the quantities $|C_{iT}(\pi_0, \lambda)|/\sqrt{T}\hat{\phi}_i$ for $i = 1, 2$ in (7) as the supremum and average functionals of the resulting limits as well as those associated with $C_{iT}(\pi_0, \lambda)^2/T\hat{\phi}_i^2$ would follow straightforwardly.

Proposition 4. *(i) Suppose model (11) holds with $\boldsymbol{\delta}_T = (\delta_0/\sqrt{T}, \mathbf{0}')$ (intercept shifts only). Under assumption **A1**, $\lim_{|\delta_0|\to\infty} \lim_{T\to\infty} |C_{1T}(\pi_0; \lambda, \lambda_0)/\sqrt{T}\hat{\phi}_1| = \infty$ in probability. (ii) Suppose model (11) holds with $\boldsymbol{\delta}_T = (0, \boldsymbol{\delta}_1/T')$ (slope shifts only). Under assumption **A1**, $\lim_{||\boldsymbol{\delta}_1||\to\infty} \lim_{T\to\infty} |C_{1T}(\pi_0; \lambda, \lambda_0)/\sqrt{T}\hat{\phi}_1| = \infty$ in probability.*

Proposition 5. *(i) Suppose model (11) holds with $\boldsymbol{\delta}_T = (\delta_0/T^{1/4}, \mathbf{0}')$ (intercept shifts only). Under assumption **A2**, $\lim_{|\delta_0|\to\infty} \lim_{T\to\infty} |C_{2T}(\pi_0; \lambda, \lambda_0)/\sqrt{T}\hat{\phi}_2| = \infty$ in probability. (ii) Suppose model (11) holds with $\boldsymbol{\delta}_T = (0, \boldsymbol{\delta}_1'/T^{3/4})$ (slope shifts only). Under assumption **A2**, $\lim_{||\boldsymbol{\delta}_1||\to\infty} \lim_{T\to\infty} |C_{2T}(\pi_0; \lambda, \lambda_0)/\sqrt{T}\hat{\phi}_2| = \infty$ in probability. (iii) Suppose model (11) holds with $\boldsymbol{\delta}_T = (\delta_0/\sqrt{T}, \mathbf{0}')$ and/or $\boldsymbol{\delta}_T = (0, \boldsymbol{\delta}_1'/T)$ then $C_{2T}(\pi_0; \lambda, \lambda_0)/\sqrt{T}\hat{\phi}_2 \Rightarrow W_2^0(\lambda)$ as $T \to \infty$.*

Proposition 4 focused on the local power properties of our test statistics associated with the levels of the forecast errors while Proposition 5 focused on the corresponding outcomes for the test statistics based on squared forecast errors instead. We note that C_{1T} based inferences are able to detect local departures of the type $(\delta_0/\sqrt{T}, \boldsymbol{\delta}_1'/T)$ while C_{2T} based statistics require the mildly slower vanishing rates of $(\delta_0/T^{1/4}, \boldsymbol{\delta}_1'/T^{3/4})$ in order to be able to detect them as otherwise their power will equal their size as implied by Proposition 5(iii).

Overall our proposed tests statistics based on either C_{1T} or C_{2T} are expected to display good power properties provided that the presence of regimes comes from the slope parameters while C_{2T} based inferences are unlikely to pickup regimes if the latter are solely originating from small shifts in intercepts. This distinct behaviour of C_{1T} and C_{2T} will also be particularly important when it comes to detecting departures from (1) that also include the potential presence of threshold effects in the variances of the u_t's. Before proceeding with such generalisations it is instructive to also present the explicit form of the limiting random variables associated with the large sample behaviour of $C_{iT}(\pi_0; \lambda, \lambda_0)/\sqrt{T}\hat{\phi}_i$ in Propositions 4 and 5 as this will help to explicitly highligh the key parameters affecting the power properties of our tests.

Focusing first on the C_{1T} based inferences as in Proposition 4 we have

$$\left|\frac{C_{1T}(\pi_0;\lambda,\lambda_0)}{\sqrt{T}\hat{\phi}_1}\right| \quad\Rightarrow\quad \left|W_1^0(\lambda)-\frac{\sqrt{1-\pi_0}(\lambda\wedge\lambda_0-\lambda\lambda^0)\delta_0}{\sigma_u}\right| \ \text{ for } \boldsymbol{\delta}_T=(\delta_0/\sqrt{T},\mathbf{0}')', \tag{19}$$

$$\left|\frac{C_{1T}(\pi_0;\lambda,\lambda_0)}{\sqrt{T}\hat{\phi}_1}\right| \quad\Rightarrow\quad \left|W_1^0(\lambda)-\frac{(\lambda\wedge\lambda_0-\lambda\lambda^0)}{\sigma_u\sqrt{1-\pi_0}}\ \boldsymbol{\delta}_1{}'\int_{\pi_0}^1\boldsymbol{J}_c\right| \ \text{ for } \boldsymbol{\delta}_T=(0,\boldsymbol{\delta}_1'/T)' \tag{20}$$

as $T\to\infty$. As stated in Proposition 4 we note that both quantities in the right hand side of (19)-(20) diverge with δ_0 and $\boldsymbol{\delta}_1$ respectively. Similarly under the setting of Proposition 5 for C_{2T} we have

$$\left|\frac{C_{2T}(\pi_0;\lambda,\lambda_0)}{\sqrt{T}\hat{\phi}_1}\right| \quad\Rightarrow\quad \left|W_2^0(\lambda)-\frac{\sqrt{1-\pi_0}(1-2\lambda_0)(\lambda\wedge\lambda_0-\lambda\lambda^0)\delta_0^2}{\phi_2}\right| \ \text{ for } \boldsymbol{\delta}_T=(\delta_0/T^{1/4},\mathbf{0}')', \tag{21}$$

$$\left|\frac{C_{2T}(\pi_0;\lambda,\lambda_0)}{\sqrt{T}\hat{\phi}_1}\right| \quad\Rightarrow\quad \left|W_2^0(\lambda)-\frac{(\lambda\wedge\lambda_0-\lambda\lambda^0)(1-2\lambda_0)}{\phi_2\sqrt{1-\pi_0}}\ \boldsymbol{\delta}_1{}'\int_{\pi_0}^1\boldsymbol{J}_c\boldsymbol{J}_c'\ \boldsymbol{\delta}_1\right| \ \text{ for } \boldsymbol{\delta}_T=(0,\boldsymbol{\delta}_1'/T^{3/4})' \tag{22}$$

as $T\to\infty$.

The results in (19)-(20) indicate that the power of C_{1T} based inferences are affected by $(\pi_0,\lambda_0,\sigma_u,\delta_0)$ under intercept shifts only scenarios and by $(\pi_0,\lambda_0,\sigma_u,\delta_0,c_1,\dots,c_p)$ under slope shifts. For both scenarios power is expected to improve for lower magnitudes of σ_u (holding all else equal), larger magnitudes of δ_0 and/or $\boldsymbol{\delta}_1$ and a smaller fraction of the sample size used to initiate the recursive forecasts. Power is of course also expected to be more favourable under slope shifts than it is under intercept shifts alone as the tests can detect departures from (1) that are *more local* (e.g. $\delta_0/\sqrt{T}$ vs $\boldsymbol{\delta}_1/T$ for C_{1T} based inferences).

The results in (21)-(22) indicate that inferences based on C_{2T} will be affected by $(\pi_0,\lambda_0,\sigma_u,E[u_t^4],\delta_0)$ and $(\pi_0,\lambda_0,\sigma_u,E[u_t^4],\boldsymbol{\delta}_1,c_1,\dots,c_p)$ so that all other things being equal a high kurtosis in the $u_t's$ is expected to also deteriorate power via $\phi_2^2=E[u_t^2-\sigma_u^2]^2$. It is also clear that C_{2T} based tests will have no meaningful power when $\lambda_0=0.5$ due to the appearance of the $(1-2\lambda_0)$ factor on both numerators.

Another important point that follows from the above outcomes is related to the role played by the non-centrality parameter c on the power properties of the tests. From (19) and (21) we note that the degree of persistence of the predictors is not expected to have any influence on the detection ability of our tests when the models under consideration have intercept shifts only. In contrast, persistence is expected to play an important role when it comes to detecting slope shifts as can be seen from the distinct appearance of the $\boldsymbol{J}_c$ process in (20) versus (22).

It is here also interesting to point out that although C_{1T} based inferences are able to detect less distant slope departures than C_{2T} based inferences (i.e. $\boldsymbol{\delta}_1/T$ versus $\boldsymbol{\delta}_1/T^{3/4}$) it is not clear whether this will necessarily translate into better finite sample power properties for the associated statistics (provided that $\lambda_0\neq 0.5$). Indeed, comparing (20) with (22) and focusing on a single predictor scenario we note the presence of the term $\delta_1\int J_c$ in (20) versus $\delta_1^2\int J_c^2$ in (22). Depending on the magnitude of δ_1 this distinction introduces a degree of ambiguity on whether C_{1T} or C_{2T} based statistics should lead to more favourable power outcomes.

The Case of Variance Shifts

An issue we hinted at earlier and which we may be confronted with in applied work is the possibility that the shocks driving the predictive regression in (1) may themselves be subject to regime specific volatility. An important issue that arises in this context is the behaviour of our C_{1T} and C_{2T} based statistics under such a scenario and whether our current toolkit can help detect such occurrences and possibly be used to disentangle between regimes induced by the conditional mean and regimes induced by threshold effects in variances. To explore these issues we consider a setting where u_{t+1} in (1) is now formulated as

$$u_{t+1} \quad = \quad \sigma_1 \epsilon_{t+1} + (\sigma_2 - \sigma_1)\epsilon_{t+1} \mathbb{I}(q_t > \gamma_{v0}) \tag{23}$$

where the threshold parameter γ_{v0} may or may not equal the threshold parameter γ_0 associated with (potential) changes in the intercept/slope parameters of the forecasting model in (11). The ϵ_{t+1}'s are here understood to satisfy assumptions **A1** and/or **A2** with ϵ_t replacing u_t in $\boldsymbol{\eta}_{1t}$ and $\boldsymbol{\eta}_{2t}$ respectively as summarised under Assumptions **B1** and **B2** below.

Assumption B1. Assumption **A1** holds with $\boldsymbol{\eta}_{1t} = (\epsilon_t, \boldsymbol{\epsilon}_{vt})'$ and where $E[\epsilon_t^2 | \mathcal{F}_{1,t-1}] = 1$ in $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}$.

Assumption B2. Assumption **A2** holds with $\boldsymbol{\eta}_{2t} = (\epsilon_t, \epsilon_t^2 - 1, \boldsymbol{\epsilon}_{vt})'$ and where $E[\epsilon_t^2 | \mathcal{F}_{2,t-1}] = 1$, $E[(\epsilon_t^2 - 1)^2 | \mathcal{F}_{2,t-1}] = \kappa_4$, $E[\epsilon_t(\epsilon_t^2 - 1) | \mathcal{F}_{2,t-1}] = \kappa_3$ in $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_2}$.

The formulation in (23) imposes a piecewise linear structure on the variances of the $u_t's$ and has been a commonly used parameterisation considered in the structural break literature where q_t is typically given by a deterministic time trend. Despite its simplicity the piecewise linear structure of u_t is able to capture very rich dynamics including serial correlation in the u_t^2. It is indeed straightforward to observe from (23) that although u_{t+1} continues to be a martingale difference sequence, unlike the ϵ_{t+1}'s, it no longer satisfies assumptions **A1**(ii) or **A2** as it is no longer conditionally homoskedastic. Using (23) we note for instance that $E[u_{t+1}^2 | \mathcal{F}_{i,t}] = \sigma_1^2 + (\sigma_2^2 - \sigma_1^2) I(q_t > \gamma_{vo})$. As a consequence the null limiting distributions of our test statistics as stated in Propositions 1 and 2 will no longer hold.

Before proceeding further it is useful to recall that the main objective behind our proposed tests is to detect the presence of state dependence in forecast errors. This state dependence may originate from omitted regimes in conditional mean parameters as in (11), from variances as in (23) or from both. Although our main concern is for our inferences based on C_{1T} and C_{2T} to be able to detect such departures from (1) regardless of where they originate, it may also be possible to disentangle between the two through the joint use of both statistics. We may for instance want our C_{1T} based inferences not to be sensitive to the presence of variance shifts so that a non-rejection based on C_{1T} followed by a rejection based on C_{2T} may be used to argue that state dependence originates in the variances.

Borrowing from the changepoint literature where it is customary to explore test properties within small shift asymptotic frameworks we adapt the same idea to (23), parameterising it as

$$u_{t+1} = \sigma_1 \epsilon_{t+1} + \frac{(\sigma_2 - \sigma_1)}{T^\theta} \epsilon_{t+1} \mathbb{I}(q_t > \gamma_{v0}) \tag{24}$$

for some $0 < \theta \leq 1/2$. We can interpret (24) as a *limited* heterogeneity framework similar to that considered in a different context in Steland (2020). The formulation in (24) is essentially saying that the difference in variances across the two regimes gets small as the sample size increases. Differently put we operate under an asymptotic approximation that is valid for small magnitudes of the spread $(\sigma_2 - \sigma_1)$ with the parameter θ controlling the rate at which this spread decreases to zero, i.e., how small we are forcing $(\sigma_2 - \sigma_1)$ to be. Smaller values of θ are thus less restrictive.

We next focus on the limiting behaviour of our test statistics based on C_{1T} and C_{2T} under Assumptions **B1** and **B2** respectively. Model (1) is understood to hold with u_{t+1} as in (24) so that the forecasting regression is characterised by threshold effects in the variance of its innovations while having a correctly specified conditional mean.

Proposition 6. *(i) Suppose model (1) holds with u_{t+1} as in (24) and $\theta > 0$. As $T \to \infty$ we have $|C_{1T}(\pi_0; \lambda, \lambda_{v0})/\sqrt{T}\hat{\phi}_1| \Rightarrow |W_1^0(\lambda)|$. (ii) Suppose model (1) holds with u_{t+1} as in (24) and $0 < \theta < 1/2$, then $|C_{2T}(\pi_0; \lambda, \lambda_{v0})/\sqrt{T}\hat{\phi}_2|$ diverges to infinity in probability as $T \to \infty$. (iii) Suppose model (1) holds with u_{t+1} as in (24) and $\theta = 1/2$, then $\lim_{|\sigma_2 - \sigma_1| \to \infty} \lim_{T \to \infty} |C_{2T}(\pi_0; \lambda, \lambda_{v0})/\sqrt{T}\hat{\phi}_2| = \infty$.*

The results in Proposition 6 are particularly important as they imply that inferences based on C_{1T} will remain as in Proposition 1, unaffected by the presence of regime specific heteroskedasticity. On the other hand under the same setting, C_{2T} based inferences should be able to detect departures from (1) as in (24) due to our results in Proposition 6(ii)-(iii). This suggests for instance that a scenario whereby C_{1T} based test statistics fail to reject the null while C_{2T} based inferences lead to rejections would indicate that state dependence in forecast errors originates solely in the variances. Our Monte-Carlo based results do indeed demonstrate that even under large spreads in variances (i.e. large magnitudes of $|\sigma_2 - \sigma_1|$), C_{1T} based inferences remain very close to their nominal size as expected from Proposition 6(i).

If C_{1T} based inferences result in rejections due to shifts in slopes then from our result in Proposition 5(ii) it is also almost certainly the case that C_{2T} based inferences should also follow suit. A rejection by both statistics would then indicate that state dependence in forecast errors originates from shifts in the slopes of (1) and potentially also shifts in variances, the latter being picked up by C_{2T}. Finally if C_{1T} based inferences lead to rejections while C_{2T} based inferences do not then based on our results in Propositions 5(iii) such a scenario is likely to indicate that state dependence originates solely from intercept shifts.

4 Experimental Evidence

Our initial objective is to document the finite sample adequacy of the distributions presented in (15)-(18) when the DGP is given by (1)-(2). We subsequently assess the ability of our test statistics to reject the null of a linear predictive regression when the true specification is given by the threshold model in (11), illustrating our key outcomes presented in Propositions 4 and 5. In analogy to the structure of Section 3 above these size and power properties of our tests are in a first instance evaluated in the context of conditionally homoskedastic $u_t's$ satisfying Assumptions **A1** and **A2**. We subsequently concentrate on the case of variance shifts as in (24) and illustrate our key results in Proposition 6 within a DGP given by (1)-(2) but with the u_t's characterised by threshold effects in their variances. In this latter context we are particularly interested in documenting the ability of our test statistics to disentangle between threshold effects in the slopes and threshold effects in variances.

DGPs

The main DGP for our size experiments is given by (1)-(2) with $x_{it} = (1-c_i/T)x_{it-1}+v_{it}$, $v_{it} = 0.5v_{it-1}+\epsilon_{ivt}$, $\epsilon_{ivt} \sim (0, \sigma^2_{\epsilon_{iv}} = 1)$ for $i = 1, 2, 3$ so as to accommodate up to three predictors. This allows us to explore the robustness of our limiting distributions in Propositions 1-2 to alternative magnitudes of the $c_i's$ and the influence of the latter on test powers. Throughout both our size and power experiments the threshold variable as described in (6) is taken to follow the centered AR(1) process $u_{qt} = 0.5u_{qt-1} + \epsilon_{qt}$, $\epsilon_{qt} \sim (0, \sigma^2_{\epsilon_q} = 1)$. The random disturbance driving (1) is such that $u_t \sim (0, \sigma^2_u = 1)$.

The covariance matrix of the $(p + 2)$ dimensional vector $(u_t, \epsilon_{1vt}, \epsilon_{2vt}, \ldots, \epsilon_{pvt}, \epsilon_{qt})$ collecting all random disturbances driving the system allows for non-zero contemporaneous correlations between these disturbances. In line with the empirical literature on the predictability of returns with valuation ratios the covariances between the $u_t's$ in (1) and the $\epsilon_{ivt}'s$ are chosen to ensure a strong negative correlation between the shocks to the $y's$ and the shocks to the predictors. For the single predictor case ($p = 1$) we set $\sigma_{u,\epsilon_{1v}} = -0.7$, $\sigma_{u,\epsilon_q} = 0.4$ and $\sigma_{\epsilon_{1v},\epsilon_q} = 0.2$. For $p = 2$ we set $\sigma_{u,\epsilon_{1v}} = -0.7$, $\sigma_{u,\epsilon_{2v}} = -0.5$, $\sigma_{u,\epsilon_q} = 0.3$, $\sigma_{\epsilon_{1v},\epsilon_{2v}} = 0.3$, $\sigma_{\epsilon_{2v},\epsilon_q} = 0.2$ and $\sigma_{\epsilon_{1v},\epsilon_q} = 0.2$. Finally for $p = 3$ we use the same covariances as under $p = 2$ together with $\sigma_{u,\epsilon_{3v}} = -0.2$, $\sigma_{\epsilon_{1v},\epsilon_{3v}} = 0.3$, $\sigma_{\epsilon_{2v},\epsilon_{3v}} = 0.3$, $\sigma_{\epsilon_{2v},\epsilon_q} = 0.3$ and $\sigma_{\epsilon_{3v},\epsilon_q} = 0.4$. Throughout this paper all simulations are conducted using normally distributed random variables across $N = 5000$ replications.

Our power experiments are conducted with a DGP as in (11), distinguishing between intercept shifts only scenarios ($\boldsymbol{\delta}_1 = 0$) and slope shifts only scenarios ($\delta_0 = 0$) for $\gamma_0 = \{-0.5, 0\}$. Here we note that the case $\gamma_0 = 0$ corresponds to a regime structure with a 50/50 split between the q_t observations that lie above and below γ_0 so that $\lambda_0 = 0.5$. Such a parameterisation is important for illustrating the power failures affecting inferences based on $e^2_{t+1|t}$'s when $\lambda_0 = 0.5$ as theoretically established in (21)-(22).

Size experiments

We initially focus on the Sup_{iT} and Ave_{iT} statistics in (7)-(8) which assume a given/known starting point for the start of the recursions to generate the out of sample forecast errors (say $k_0 = [T\pi_0]$ with $\pi_0 = 0.25$). For each generated $(y_t, \boldsymbol{x}_t', q_t)$, the sequence of forecast errors and their squares is first obtained as in (3). These sequences are subsequently used to construct the test statistics in (7)-(8). Results are presented in Tables 2-3 which compare empirical sizes with the chosen nominal size of 5%. Table 2 presents results for predictive regressions with $p = 1, 2, 3$ predictors when the latter are constrained to have the same degree of persistence. Table 3 provides additional outcomes when each predictor has a different non-centrality parameter.

Empirical sizes are seen to match their nominal counterparts closely across all sample sizes except perhaps for a very mild undersizeness under $T = 400$ characterising the Sup_{2T} based tests that rely on the squared forecast errors. It is also generally the case that the Sup_{iT} based tests display mildly lower empirical sizes compared to the Ave_{iT} based statistics which marginally overshoot the nominal size of 5%. Overall our proposed tests can be seen to be accurately sized across a rich set of alternative scenarios and sample sizes. As expected from our asymptotics it is also particularly important to note the robustness of outcomes to alternative choices of the non-centrality parameters and the number of predictors included in the predictive regressions.

We next consider the case of the $SupSup_{iT}$ and $AveAve_{iT}$ statistics in (9)-(10) which are designed to robustify inferences to the chosen starting period of the recursions. Our experiments use the range $\Pi = [0.50, 0.75]$ for scanning across recursion starting points. The corresponding 5% quantile cut-offs of the two test statistics are given by 1.573 and 0.415 respectively (see Table 1) and results on specifications with $p = 1, 2, 3$ predictors are presented in Table 4. We continue to note good to excellent matches of nominal sizes across all parameterisations. Empirical sizes associated with the $SupSup_{iT}$ statistics typically lie mildly below those associated with the $AveAve_{iT}$ statistics as documented previously. Also noteworthy is the robustness of finite sample sizes to the number of predictors and their associated non-centrality parameters as it was the case for the Sup_{iT}/Ave_{iT} statistics above.

Power experiments

We next explore the power properties of our test statistics under intercept only and slope shift only scenarios. Our empirical results in Tables 5-8 are designed to illustrate both the correct decision frequencies associated with each statistic as T grows but also the evolution of power for fixed T as the parameterisations of the DGP move further away from linearity.

Tables 5-6 focus on the power of the Sup_{iT} and Ave_{iT} statistics across intercept shift only and slope shift only scenarios. Focusing first on the case of intercept shifts (Table 5) we note empirical correct decision frequencies at or very near 100% for samples of size $T = 400$ or above when inferences are based on the Sup_{1T}/Ave_{1T} statistics under moderately large departures such as $\delta_0 \geq 0.5$. Even under $\delta_0 = 0.25$ we note empirical powers above 80% for $T = 1000$. These outcomes are in line with our results in Proposition 4 where we documented the ability of C_{1T} based statistics to detect local intercept departures of the type $\delta_0/\sqrt{T}$. Table 5 also highlights the robustness of power outcomes to the magnitude of non-centrality parameters as expected from (19) and (21) where c is seen not to play any role under intercept shift only scenarios.

The empirical power outcomes associated with intercept shifts only change significantly when inferences are based on the squared forecast errors via the Sup_{2T}/Ave_{2T} statistics. These squared forecast error based statistics are unable to detect intercept shifts unless the latter are very large and $\lambda_0 \neq 0.5$. This is in total agreement with our result in Proposition 5(i) from which we note that to be detectable by Sup_{2T}/Ave_{2T}, local departures should be of the form $\delta_0/T^{1/4}$ as opposed to $\delta_0/\sqrt{T}$ in the case of the Sup_{1T}/Ave_{1T} statistics.

Table 6 repeats the same exercise for slope shifts only. The first two vertical panels distinguish between p=1 and p=2 predictor scenarios while the third panel reconsiders the case of p=2 predictors but with only one of the two associated slopes shifting. As in the intercept shift only case, Sup_{1T}/Ave_{1T} based tests display correct decision frequencies either close to 100% or quickly converging to 100% as δ_1 increases. Power is also seen to improve with the degree of persistence of predictors with the most favourable outcomes occurring under $c = 1$. This is in total agreement with our result in (20) which documents the explicit role played by the non-centrality parameter c on the local power properties of the tests.

In this context of slope shifts and unlike the earlier intercept shift only scenarios, the squared forecast errors based statistics Sup_{2T}/Ave_{2T} also display good power properties provided that $\lambda_0 \neq 0.5$. They can in fact also be seen to dominate Sup_{1T}/Ave_{1T} based inferences in many instances (under $c = 20$ and $c = 40$ in particular) despite the fact that the latter are able to detect departures of the type $\boldsymbol{\delta}_1/T$ versus $\boldsymbol{\delta}_1/T^{3/4}$ for the former. Our results in (20) and (22) and the terms involving $\int J_c$ versus $\int J_c^2$ in particular provide the intuition for these empirical occurrences. It is also interesting to observe that the power properties of the Sup_{2T}/Ave_{2T} based tests appear to deteriorate much less compared with Sup_{1T}/Ave_{1T} when the predictors move closer to the stationarity region (e.g. under $c = 20$ and $c = 40$).

Tables (7)-(8) extend the above analysis to our class of $SupSup_{iT}/AveAve_{iT}$ based tests and broadly corroborate our findings based on the Sup/Ave statistics. The use of $SupSup/AveAve$ based tests does not appear to lead to any meaningful finite sample power losses relative to the Sup/Ave statistics which were based on a fixed recursion starting point.

Variance shifts

The last set of experiments focuses on the behaviour of our C_{1T} and C_{2T} based statistics under variance shifts and aim to illustrate our key results in Proposition 6. Specifically, we expect that provided the shifts in variance are not too large, C_{1T} based statistics will maintain their correct size. At the same time we expect C_{2T} based tests to display good power properties for either small or large variance shifts. We view these features as particularly useful as they may resolve the issue of identifying whether state dependence originates from conditional mean parameters or variances (e.g. if C_{1T} based inferences fail to reject while C_{2T} based inferences lead to rejections).

Results are presented in Table 9. As expected from Proposition 6(ii) we note that the C_{2T} based statistics display excellent power to detect variance shifts even under moderately sized samples while C_{1T} based inferences tend to remain close to nominal size even for moderately large shifts, as expected from Proposition 6(i). Thus although rejections by C_{2T} may be caused by either variance shifts and/or slope parameter shifts, a non-rejection by C_{1T} followed by a rejection by C_{2T} would indicate state dependence induced solely by variance shifts.

5 Application: Regime Specificity in the Value Premium

In this section we use our methods to assess the potential presence of state dependence in the forecasts of the value premium, a series which has been the subject of considerable research in the asset pricing literature and which refers to the commonly observed out-performance of cheap/value stocks (e.g. stocks having high Book-to-Market value (BM), high Earnings-to-Price ratios) relative to more expensive growth stocks (e.g. low Book-to-Market stocks). Although there is broad consensus on its existence across most developed and emerging markets the debate on whether this premium reflects compensation for risk or a behavioural anomaly is still very much on-going (see Lakonishok et al. (1994) , Fama and French (2012) and references therein). A related agenda that parallels our own objectives here has also been concerned with establishing whether such risk premiums are time varying and their relationship with the business cycle (see Petkova and Zhang (2005), Long et al. (2008), Gülen et al. (2011)).

Although there is no unique way of defining the value premium a standard approach in the literature has involved forming decile or quintile portfolios on firm specific characteristics such as Book-to-Market Value, Dividend Yields, Ernings-to-Price ratios and evaluating the return spreads between upper and lower deciles in each time period and for a broad cross-section. This is also the approach we adopt here using Kenneth French's publicly available data library. More specifically we focus on monthly US data covering the period 1930-2017 and construct our value premium series as the difference between the top and bottom decile portfolios formed on Book-to-Market (D10BM-D1BM). In addition to this decile based series we also consider the well-known value premium factor known as HML (High minus Low) and which is one of the three factors used in Fama and French's three factor model (see Fama and French (2012)) to capture the sensitivity and exposure of a given portfolios or individual stock to value. Our choice of predictors is based on a selection of variables that have been commonly considered in the equity premium forecasting literature and given by the well known Goyal and Welch data set (see Goyal and Welch (2014)). We restrict our analysis to a selection of these variables which have been commonly found to play a significant role in the predictability of the equity premium. These are the log dividend yield (LDY), the treasury bill rate (TBL), the term spread (TMS), the long term government bond returns (LTR) and the default yield spread (DEF). Finally our choice of threshold variable (and business cycle proxy) q_t is given by the monthly growth rate in the US index of industrial production (IPGR). A standard ADF test applied to this series led to a strong rejection of the unit root hypothesis, supporting our setting that operates under a stationary threshold variable. As our test statistics are able to accommodate predictive regressions with both single and multiple predictors our empirical analysis considers the above variables both individually and in a selection of groups. All our inferences that are based on a given π_0 use $\pi_0 = 0.25$ throughout while our $SupSup_{iT}$ and $AveAve_{iT}$ based outcomes use the range $\Pi = [0.25, 0.75]$.

Results are presented in Table 10. Focusing first on the single predictor based specifications we note that our empirical results provide a consistent picture for the two series both across the selection of predictors and the two types of test statistics being considered. The big picture that comes across is the obvious absence of any threshold effects induced by our business cycle proxy on the level of the out of sample forecast errors suggesting no evidence of state dependence in the conditional mean specifications. This finding is supported by both our Sup_{iT} and Ave_{iT} based statistics across all predictors as well as our $SupSup_{iT}$ and $AveAve_{iT}$ based inferences. Indeed we can note that none of the inferences based on Sup_{1T}/Ave_{1T} and $SupSup_{1T}/AveAve_{1T}$ lead to any significant rejections at or above a 5% nominal level.

Perhaps more interestingly our inferences strongly and convincingly point to the presence of state dependence in the volatility of the value premium instead which in turn leads to distinct MSEs in good versus bad times, a phenomenon also pointed out in Gülen et al. (2011). In the context of the HML sequence we note that inferences based on Sup_{2T}/Ave_{2T} and $SupSup_{2T}/AveAve_{2T}$ fully coincide across all predictors, leading to strong rejections of linearity. For the D10BM-D1BM sequence it continues to be the case that both Ave_{2T} and $AveAve_{2T}$ lead to strong rejections with consistent outcomes throughout. The stronger power properties of $AveAve_{iT}$ based statistics relative to their $SupSup_{iT}$ versions may also explain the inability of $SupSup_{2T}$ to point to threshold effects for this series when $AveAve_{2T}$ based inferences point to strong rejections.

More generally our results highlight the need for inferences about predictability to account for the possibility of regime specific heteroskedasticity. Although predictability is often assessed via the statistical significance of estimated slope coefficients and the use of standard heteroskedasticity consistent standard errors the latter are not necessarily able to correct for such effects as documented for instance in Pitarakis (2002).

6 Conclusions and Discussion

The goal of this paper was to propose a set of diagnostic tools for assessing the presence of regimes in the out of sample forecast errors generated by commonly used predictive regressions. An important motivation for this research was driven by the observation in numerous empirical studies that predictability may often be a cyclical phenomenon kicking in during particular periods while vanishing in other times. The tests developed in this paper were therefore designed to provide a simple toolkit for formally assessing the occurrence of such phenomena.

Unique features of our methods include their ability to accommodate multiple predictors with varying persistence strengths and their ability to robustify inferences to the size of the starting sample used to initiate the recursive forecasts. Equally importantly, the practical implementation of these methods is particularly straightforward as inferences rely on nuisance parameter free distributions whose critical values are available. This, despite the fact that predictors are parameterised with unknown persistence parameters in the underlying DGP. Finally, another particularly noteworthy feature of our tests is their ability, under certain circumstances, to identify whether state-dependence originates in the mean or variance related parameters of the process.

The recent literature on predictive regressions has devoted considerable attention to the issue of time variation in the conditional mean parameters of such specifications. These developments have typically differed in the type of parameterisations used to capture parameter instability with an important emphasis placed on traditional structural breaks and more recently on more general forms of time variation via random coefficient models specified in a way that accommodates slowly evolving changes as for instance in Georgiev et al. (2018) or smoothly varying coefficients as in Farmer et al. (2018). A common objective behind the bulk of this research agenda has been the development of in-sample tests of linearity (i.e. parameter stability) against such time-varying alternatives with a particular effort devoted to robustifying inferences to unknown forms of conditional and/or unconditional heteroskedasticity via bootstrap based approaches. In contrast to this rich agenda our own interest has instead been on using out of sample forecast errors *and* their squares to uncover the presence of regimes in their respective dynamics. This has also allowed us to explicitly address the important issue of distinguishing between time variation driven by conditional mean parameters and/or error variances which is of interest in its own right. Another distinctive feature of our environment stems from our use of threshold effects to capture parameter variation. This freedom to place a cause to what drives regimes via a user-selected threshold variable enhances the scope of our proposed methods and depending on its dynamics can accommodate both noisy and smoother parameter variations.

TABLES

TABLE 1

Quantiles of the Asymptotic Distributions of SupSup and AveAve statistics

	10%	5%	2.5%	1%
		$\pi \in [0.25, 0.75]$		
SupSup	1.504	1.632	1.736	1.874
AveAve	0.317	0.412	0.518	0.638
		$\pi \in [0.50, 0.75]$		
SupSup	1.446	1.573	1.684	1.816
AveAve	0.325	0.415	0.517	0.658
		$\pi \in [0.50, 0.90]$		
SupSup	1.550	1.670	1.785	1.916
AveAve	0.315	0.392	0.484	0.603

TABLE 2

Empirical sizes of Sup_i and Ave_i statistics under single and multiple predictors (5% Nominal)

	p=1				p=2				p=3			
	Sup_1	Sup_2	Ave_1	Ave_2	Sup_1	Sup_2	Ave_1	Ave_2	Sup_1	Sup_2	Ave_1	Ave_2
						c=1						
T=400	4.68	4.22	6.26	5.06	4.96	3.88	6.62	5.38	5.78	3.48	7.28	5.92
T=600	4.42	4.18	5.78	5.20	5.16	4.86	5.78	6.16	5.68	4.20	6.98	5.46
T=1000	5.22	4.30	6.08	5.32	4.92	4.22	5.58	4.88	4.96	4.26	6.22	5.52
						c=20						
T=400	4.18	4.06	5.24	5.60	4.90	4.18	6.42	5.62	5.26	3.54	6.80	5.40
T=600	4.86	3.92	5.64	5.06	4.90	4.26	5.62	5.60	4.60	3.88	6.04	5.08
T=1000	4.24	4.28	5.52	5.62	4.42	4.52	5.34	5.68	5.10	3.80	6.24	4.70
						c=40						
T=400	4.36	3.66	5.74	4.84	4.46	3.90	6.00	5.62	4.74	3.92	6.10	5.26
T=600	4.18	3.96	5.16	5.26	4.58	4.30	6.08	5.22	4.80	4.16	6.10	5.34
T=1000	4.42	4.18	5.38	5.20	4.56	4.46	5.76	5.54	4.52	4.06	5.44	4.92

TABLE 3

Empirical sizes of Sup_i and Ave_i with heterogeneous predictors (5% Nominal)

	p=2				p=3			
	Sup_1	Sup_2	Ave_1	Ave_2	Sup_1	Sup_2	Ave_1	Ave_2
	$(c_1, c_2) = (1, 20)$				$(c_1, c_2, c_3) = (1, 1, 20)$			
T=400	5.20	3.62	6.74	5.76	4.72	3.94	6.68	5.62
T=600	4.86	4.14	6.22	5.30	5.44	3.78	6.44	4.82
T=1000	4.94	3.88	5.84	5.08	4.88	4.24	5.94	5.24
	$(c_1, c_2) = (1, 40)$				$(c_1, c_2, c_3) = (1, 20, 40)$			
T=400	4.68	3.96	5.92	5.32	5.08	4.34	6.70	5.80
T=600	4.34	4.16	5.84	5.50	4.88	4.52	5.72	5.70
T=1000	4.56	4.14	5.52	4.74	4.76	4.38	5.68	5.50
	$(c_1, c_2) = (20, 40)$				$(c_1, c_2, c_3) = (20, 20, 40)$			
T=400	4.44	3.84	5.86	5.00	4.94	3.90	6.68	5.20
T=600	5.06	4.18	5.92	5.82	4.74	4.22	6.26	5.74
T=1000	4.82	4.14	5.92	5.08	4.76	4.76	6.08	5.32

TABLE 4

Empirical sizes of SupSup_i and AveAve_i under $\Pi = [0.50, 0.75]$ (5% nominal)

	p=1				p=2				p=3			
	SupSup_1	SupSup_2	AveAve_1	AveAve_2	SupSup_1	SupSup_2	AveAve_1	AveAve_2	SupSup_1	SupSup_2	AveAve_1	AveAve_2
						c=1						
T=400	3.86	3.50	5.56	6.30	4.74	3.86	6.32	6.70	4.12	3.80	6.02	6.82
T=600	4.52	4.22	5.94	6.24	4.94	4.44	5.80	5.86	4.76	4.48	6.02	6.70
T=1000	5.22	4.44	6.42	6.00	4.94	4.76	6.04	6.40	5.04	4.08	5.90	5.66
						c=20						
T=400	4.26	3.68	6.68	6.30	3.62	3.62	6.02	5.98	4.30	3.36	6.04	6.30
T=600	4.64	3.54	5.72	5.46	5.46	3.60	6.66	5.74	5.08	4.02	6.42	5.74
T=1000	4.96	4.20	5.80	5.46	4.98	4.48	5.60	5.76	4.98	4.58	5.92	6.30
						c=40						
T=400	4.16	4.02	5.44	6.82	4.18	3.50	5.78	5.62	4.54	3.14	6.22	5.64
T=600	4.94	3.88	6.20	5.82	4.38	3.60	5.94	6.14	4.64	3.62	6.04	5.78
T=1000	4.74	4.62	6.18	6.08	5.12	4.78	5.84	6.26	5.24	4.36	6.38	5.70

TABLE 5

Empirical power of Sup_i and Ave_i under intercept shifts of magnitude δ_0

	$\gamma_0 = 0$						$\gamma_0 = -0.5$					
δ_0	0.25	0.50	0.75	1.00	1.25	1.50	0.25	0.50	0.75	1.00	1.25	1.50
c=1						T=400						
Sup_1	0.37	0.95	1.00	1.00	1.00	1.00	0.38	0.86	1.00	1.00	1.00	1.00
Sup_2	0.04	0.05	0.06	0.09	0.13	0.18	0.06	0.04	0.08	0.34	0.40	0.60
Ave_1	0.40	0.95	1.00	1.00	1.00	1.00	0.40	0.90	1.00	1.00	1.00	1.00
Ave_2	0.06	0.06	0.09	0.12	0.16	0.22	0.00	0.06	0.14	0.40	0.46	0.64
						T=600						
Sup_1	0.58	1.00	1.00	1.00	1.00	1.00	0.46	1.00	1.00	1.00	1.00	1.00
Sup_2	0.04	0.05	0.07	0.10	0.14	0.18	0.02	0.06	0.18	0.48	0.54	0.64
Ave_1	0.60	1.00	1.00	1.00	1.00	1.00	0.52	1.00	1.00	1.00	1.00	1.00
Ave_2	0.06	0.06	0.09	0.12	0.16	0.21	0.04	0.06	0.20	0.48	0.62	0.74
						T=1000						
Sup_1	0.83	1.00	1.00	1.00	1.00	1.00	0.68	1.00	1.00	1.00	1.00	1.00
Sup_2	0.04	0.05	0.07	0.10	0.15	0.19	0.04	0.02	0.30	0.58	0.80	0.96
Ave_1	0.84	1.00	1.00	1.00	1.00	1.00	0.68	1.00	1.00	1.00	1.00	1.00
Ave_2	0.06	0.06	0.08	0.12	0.17	0.22	0.00	0.10	0.32	0.60	0.74	0.96
c=20						T=400						
Sup_1	0.38	0.95	1.00	1.00	1.00	1.00	0.20	0.82	1.00	1.00	1.00	1.00
Sup_2	0.04	0.04	0.06	0.08	0.10	0.14	0.02	0.02	0.16	0.36	0.50	0.70
Ave_1	0.41	0.95	1.00	1.00	1.00	1.00	0.24	0.82	1.00	1.00	1.00	1.00
Ave_2	0.06	0.06	0.08	0.10	0.12	0.17	0.00	0.06	0.20	0.34	0.52	0.78
						T=600						
Sup_1	0.58	1.00	1.00	1.00	1.00	1.00	0.46	1.00	1.00	1.00	1.00	1.00
Sup_2	0.04	0.04	0.06	0.07	0.10	0.15	0.02	0.12	0.14	0.24	0.64	0.76
Ave_1	0.61	0.99	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00
Ave_2	0.06	0.06	0.07	0.09	0.13	0.17	0.02	0.10	0.16	0.28	0.66	0.80
						T=1000						
Sup_1	0.84	1.00	1.00	1.00	1.00	1.00	0.84	1.00	1.00	1.00	1.00	1.00
Sup_2	0.04	0.04	0.06	0.08	0.11	0.16	0.02	0.04	0.28	0.62	0.78	0.90
Ave_1	0.85	1.00	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.00	1.00	1.00
Ave_2	0.05	0.06	0.08	0.10	0.13	0.17	0.02	0.06	0.34	0.70	0.76	0.94
c=40						T=400						
Sup_1	0.39	0.95	1.00	1.00	1.00	1.00	0.32	0.78	1.00	1.00	1.00	1.00
Sup_2	0.04	0.04	0.05	0.07	0.10	0.14	0.06	0.06	0.12	0.28	0.44	0.64
Ave_1	0.42	0.95	1.00	1.00	1.00	1.00	0.36	0.82	1.00	1.00	1.00	1.00
Ave_2	0.05	0.07	0.07	0.09	0.12	0.16	0.12	0.06	0.18	0.30	0.48	0.70
						T=600						
Sup_1	0.57	1.00	1.00	1.00	1.00	1.00	0.40	0.98	1.00	1.00	1.00	1.00
Sup_2	0.04	0.05	0.05	0.07	0.11	0.14	0.00	0.10	0.14	0.40	0.66	0.84
Ave_1	0.60	1.00	1.00	1.00	1.00	1.00	0.40	0.98	1.00	1.00	1.00	1.00
Ave_2	0.06	0.06	0.07	0.09	0.13	0.17	0.04	0.12	0.14	0.44	0.70	0.90
						T=1000						
Sup_1	0.84	1.00	1.00	1.00	1.00	1.00	0.70	1.00	1.00	1.00	1.00	1.00
Sup_2	0.04	0.04	0.06	0.08	0.11	0.16	0.04	0.06	0.22	0.52	0.72	0.96
Ave_1	0.84	1.00	1.00	1.00	1.00	1.00	0.72	1.00	1.00	1.00	1.00	1.00
Ave_2	0.05	0.05	0.08	0.10	0.13	0.18	0.06	0.08	0.20	0.58	0.82	0.96

TABLE 6

Empirical power of Sup_i and Ave_i under slope shifts of magnitude δ_1

	$\gamma_0=0$			$\gamma_0=-0.5$			$\gamma_0=0$			$\gamma_0=-0.5$			$\gamma_0=0$			$\gamma_0=-0.5$		
	p=1						p=2						p=2 (one slope shift)					
δ_1	0.50	0.75	1.00	0.50	0.75	1.00	0.50	0.75	1.00	0.50	0.75	1.00	0.50	0.75	1.00	0.50	0.75	1.00
c=1	T=400																	
Sup_1	0.93	0.93	0.94	0.92	0.93	0.94	0.93	0.93	0.94	0.93	0.94	0.94	0.92	0.94	0.94	0.91	0.94	0.93
Sup_2	0.44	0.54	0.56	0.96	0.99	0.99	0.51	0.55	0.56	0.98	0.99	0.99	0.44	0.51	0.56	0.95	0.98	0.99
Ave_1	0.93	0.94	0.94	0.92	0.93	0.94	0.93	0.94	0.94	0.93	0.94	0.94	0.93	0.95	0.94	0.92	0.94	0.93
Ave_2	0.46	0.56	0.58	0.97	0.99	0.99	0.53	0.57	0.58	0.98	0.99	0.99	0.47	0.54	0.58	0.96	0.99	0.99
	T=600																	
Sup_1	0.94	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.94	0.95	0.95
Sup_2	0.49	0.56	0.59	0.99	1.00	1.00	0.55	0.57	0.59	1.00	1.00	1.00	0.48	0.56	0.58	0.99	1.00	1.00
Ave_1	0.94	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.96
Ave_2	0.51	0.57	0.60	0.99	1.00	1.00	0.57	0.59	0.61	1.00	1.00	1.00	0.50	0.58	0.60	0.99	1.00	1.00
	T=1000																	
Sup_1	0.96	0.96	0.97	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.97	0.96	0.96	0.96
Sup_2	0.54	0.58	0.59	1.00	1.00	1.00	0.57	0.58	0.60	1.00	1.00	1.00	0.54	0.57	0.59	1.00	1.00	1.00
Ave_1	0.96	0.96	0.97	0.95	0.97	0.96	0.96	0.96	0.96	0.96	0.96	0.97	0.96	0.97	0.97	0.96	0.96	0.96
Ave_2	0.55	0.59	0.60	1.00	1.00	1.00	0.58	0.59	0.61	1.00	1.00	1.00	0.55	0.58	0.60	1.00	1.00	1.00
c=20	T=400																	
Sup_1	0.57	0.68	0.73	0.56	0.66	0.70	0.67	0.72	0.74	0.64	0.71	0.72	0.56	0.69	0.72	0.54	0.66	0.70
Sup_2	0.13	0.32	0.41	0.59	0.91	0.95	0.26	0.41	0.46	0.86	0.95	0.96	0.14	0.32	0.41	0.59	0.91	0.95
Ave_1	0.59	0.70	0.74	0.58	0.68	0.71	0.68	0.74	0.75	0.65	0.72	0.73	0.58	0.70	0.74	0.56	0.68	0.71
Ave_2	0.16	0.35	0.44	0.63	0.92	0.95	0.30	0.44	0.48	0.87	0.95	0.96	0.17	0.34	0.43	0.62	0.92	0.95
	T=600																	
Sup_1	0.68	0.76	0.78	0.67	0.75	0.76	0.75	0.78	0.79	0.73	0.76	0.77	0.69	0.76	0.78	0.66	0.74	0.77
Sup_2	0.20	0.40	0.46	0.86	0.98	0.99	0.34	0.47	0.50	0.96	0.99	0.99	0.18	0.40	0.44	0.85	0.98	0.99
Ave_1	0.69	0.77	0.79	0.68	0.76	0.76	0.76	0.79	0.80	0.74	0.76	0.77	0.69	0.77	0.79	0.67	0.75	0.78
Ave_2	0.22	0.43	0.48	0.87	0.98	0.99	0.36	0.49	0.52	0.97	0.99	0.99	0.21	0.41	0.46	0.86	0.98	0.99
	T=1000																	
Sup_1	0.78	0.84	0.84	0.76	0.81	0.83	0.81	0.84	0.84	0.80	0.82	0.82	0.79	0.83	0.83	0.77	0.81	0.83
Sup_2	0.29	0.45	0.50	0.99	1.00	1.00	0.42	0.51	0.52	1.00	1.00	1.00	0.28	0.45	0.48	0.99	1.00	1.00
Ave_1	0.78	0.84	0.84	0.77	0.82	0.84	0.82	0.84	0.85	0.81	0.82	0.83	0.79	0.83	0.84	0.78	0.82	0.84
Ave_2	0.31	0.47	0.51	0.99	1.00	1.00	0.43	0.52	0.53	1.00	1.00	1.00	0.30	0.46	0.50	0.99	1.00	1.00
c=40	T=400																	
Sup_1	0.35	0.52	0.59	0.32	0.52	0.57	0.47	0.61	0.62	0.46	0.60	0.62	0.35	0.51	0.58	0.31	0.51	0.56
Sup_2	0.07	0.23	0.34	0.30	0.82	0.92	0.16	0.35	0.44	0.68	0.92	0.94	0.07	0.21	0.34	0.28	0.80	0.92
Ave_1	0.37	0.53	0.61	0.35	0.53	0.58	0.49	0.62	0.64	0.48	0.61	0.62	0.36	0.53	0.60	0.33	0.52	0.58
Ave_2	0.09	0.27	0.37	0.35	0.83	0.93	0.19	0.38	0.46	0.72	0.93	0.95	0.09	0.24	0.37	0.33	0.82	0.93
	T=600																	
Sup_1	0.49	0.64	0.67	0.47	0.62	0.66	0.60	0.68	0.71	0.58	0.67	0.70	0.50	0.63	0.67	0.46	0.62	0.66
Sup_2	0.12	0.31	0.40	0.63	0.95	0.98	0.24	0.42	0.47	0.91	0.98	0.99	0.12	0.30	0.40	0.60	0.95	0.98
Ave_1	0.52	0.65	0.68	0.48	0.63	0.66	0.61	0.69	0.73	0.59	0.68	0.70	0.51	0.64	0.68	0.48	0.63	0.66
Ave_2	0.14	0.34	0.43	0.66	0.96	0.98	0.26	0.44	0.48	0.92	0.98	0.99	0.14	0.33	0.43	0.64	0.96	0.98
	T=1000																	
Sup_1	0.65	0.74	0.77	0.63	0.73	0.74	0.72	0.76	0.78	0.71	0.75	0.76	0.65	0.74	0.75	0.63	0.72	0.75
Sup_2	0.19	0.38	0.46	0.94	0.99	1.00	0.34	0.47	0.50	0.99	1.00	1.00	0.18	0.40	0.44	0.94	1.00	1.00
Ave_1	0.66	0.75	0.77	0.63	0.73	0.75	0.73	0.77	0.78	0.71	0.76	0.76	0.66	0.75	0.76	0.64	0.73	0.76
Ave_2	0.21	0.40	0.48	0.95	1.00	1.00	0.37	0.49	0.50	0.99	1.00	1.00	0.21	0.42	0.46	0.95	1.00	1.00

TABLE 7

Empirical power of SupSup_i and AveAve_i under intercept shifts of magnitude δ_0 and $\Pi = [0.50, 0.75]$

	$\gamma_0 = 0$						$\gamma_0 = -0.5$					
δ_0	0.25	0.5	0.75	1	1.25	1.5	0.25	0.5	0.75	1	1.25	1.5
c=1						T=400						
SupSup_1	0.22	0.77	1.00	1.00	1.00	1.00	0.17	0.67	0.96	1.00	1.00	1.00
SupSup_2	0.04	0.05	0.06	0.08	0.11	0.16	0.04	0.06	0.10	0.20	0.32	0.47
AveAve_1	0.27	0.79	0.99	1.00	1.00	1.00	0.23	0.68	0.95	1.00	1.00	1.00
AveAve_2	0.06	0.07	0.10	0.11	0.14	0.19	0.07	0.09	0.15	0.26	0.38	0.52
						T=600						
SupSup_1	0.37	0.94	1.00	1.00	1.00	1.00	0.27	0.88	1.00	1.00	1.00	1.00
SupSup_2	0.04	0.06	0.07	0.09	0.11	0.16	0.05	0.07	0.13	0.25	0.43	0.65
AveAve_1	0.40	0.94	1.00	1.00	1.00	1.00	0.31	0.87	1.00	1.00	1.00	1.00
AveAve_2	0.06	0.08	0.09	0.12	0.14	0.18	0.07	0.10	0.17	0.31	0.48	0.68
						T=1000						
SupSup_1	0.61	1.00	1.00	1.00	1.00	1.00	0.50	0.99	1.00	1.00	1.00	1.00
SupSup_2	0.04	0.06	0.07	0.09	0.13	0.17	0.05	0.08	0.18	0.40	0.64	0.83
AveAve_1	0.61	1.00	1.00	1.00	1.00	1.00	0.52	0.99	1.00	1.00	1.00	1.00
AveAve_2	0.06	0.08	0.09	0.11	0.15	0.18	0.07	0.10	0.22	0.42	0.66	0.83

TABLE 8

Empirical power of SupSup_i and AveAve_i under slope shifts of magnitude δ_1 and $\Pi = [0.50, 0.75]$

	$\gamma_0 = 0$				$\gamma_0 = -0.5$				$\gamma_0 = 0$				$\gamma_0 = -0.5$			
	p=1								p=2							
δ_1	0.50	0.75	1.00	1.25	0.50	0.75	1.00	1.25	0.50	0.75	1.00	1.25	0.50	0.75	1.00	1.25
c=1									T=400							
SupSup_1	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	1.00	0.98	0.99	0.99	0.99
SupSup_2	0.48	0.62	0.63	0.69	0.92	0.98	0.98	0.99	0.60	0.69	0.70	0.72	0.96	0.98	0.99	0.99
AveAve_1	0.97	0.98	0.98	0.98	0.96	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.97	0.98	0.98	0.98
AveAve_2	0.46	0.57	0.59	0.64	0.92	0.97	0.98	0.99	0.56	0.63	0.64	0.64	0.95	0.98	0.99	0.99
									T=600							
SupSup_1	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
SupSup_2	0.52	0.65	0.67	0.70	0.97	1.00	1.00	1.00	0.65	0.72	0.73	0.76	0.99	1.00	1.00	1.00
AveAve_1	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99
AveAve_2	0.47	0.59	0.61	0.63	0.97	0.99	1.00	1.00	0.60	0.64	0.67	0.69	0.99	0.99	1.00	1.00
									T=1000							
SupSup_1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SupSup_2	0.62	0.70	0.72	0.72	1.00	1.00	1.00	1.00	0.72	0.75	0.77	0.77	1.00	1.00	1.00	1.00
AveAve_1	0.99	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00
AveAve_2	0.54	0.60	0.62	0.64	1.00	1.00	1.00	1.00	0.63	0.66	0.67	0.68	1.00	1.00	1.00	1.00

TABLE 9

Empirical size (Sup_1, Ave_1) and power (Sup_2, Ave_2) under variance shifts

	$\gamma_{0v} = 0$					$\gamma_{0v} = -0.5$					$\gamma_{0v} = 0.5$				
$\sigma_2 - \sigma_1$	0.00	1.00	2.00	3.00	4.00	0.00	1.00	2.00	3.00	4.00	0.00	1.00	2.00	3.00	4.00
c=1								T=400							
Sup_1	0.04	0.07	0.08	0.09	0.09	0.04	0.06	0.06	0.06	0.07	0.05	0.05	0.06	0.06	0.06
Sup_2	0.04	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00
Ave_1	0.06	0.06	0.06	0.06	0.07	0.05	0.05	0.05	0.05	0.04	0.06	0.04	0.04	0.04	0.04
Ave_2	0.05	1.00	1.00	1.00	1.00	0.06	1.00	1.00	1.00	1.00	0.06	1.00	1.00	1.00	1.00
								T=600							
Sup_1	0.04	0.07	0.09	0.09	0.10	0.04	0.05	0.06	0.06	0.07	0.05	0.06	0.07	0.06	0.07
Sup_2	0.04	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00
Ave_1	0.05	0.06	0.07	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.06	0.04	0.05	0.04	0.05
Ave_2	0.06	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00
								T=1000							
Sup_1	0.05	0.07	0.09	0.09	0.10	0.05	0.05	0.06	0.07	0.07	0.05	0.05	0.06	0.07	0.06
Sup_2	0.04	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00
Ave_1	0.06	0.06	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.06	0.04	0.04	0.04	0.04
Ave_2	0.05	1.00	1.00	1.00	1.00	0.06	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00
c=20								T=400							
Sup_1	0.04	0.07	0.09	0.09	0.10	0.04	0.06	0.06	0.05	0.06	0.05	0.05	0.06	0.06	0.06
Sup_2	0.04	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00
Ave_1	0.06	0.05	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.04	0.06	0.04	0.04	0.04	0.04
Ave_2	0.05	1.00	1.00	1.00	1.00	0.06	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00
								T=600							
Sup_1	0.05	0.07	0.09	0.09	0.10	0.05	0.05	0.06	0.07	0.06	0.05	0.05	0.06	0.06	0.06
Sup_2	0.04	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00
Ave_1	0.06	0.06	0.07	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.05	0.04	0.04	0.04	0.04
Ave_2	0.05	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00	0.06	1.00	1.00	1.00	1.00
								T=1000							
Sup_1	0.04	0.07	0.08	0.09	0.10	0.05	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.06
Sup_2	0.05	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00
Ave_1	0.05	0.06	0.06	0.05	0.06	0.06	0.04	0.04	0.03	0.04	0.05	0.04	0.04	0.04	0.03
Ave_2	0.06	1.00	1.00	1.00	1.00	0.06	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00
c=40								T=400							
Sup_1	0.05	0.06	0.08	0.09	0.09	0.05	0.05	0.06	0.07	0.06	0.04	0.05	0.06	0.06	0.07
Sup_2	0.04	1.00	1.00	1.00	1.00	0.03	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00
Ave_1	0.06	0.05	0.06	0.06	0.06	0.06	0.05	0.04	0.05	0.04	0.05	0.04	0.04	0.04	0.04
Ave_2	0.05	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00
								T=600							
Sup_1	0.05	0.07	0.08	0.09	0.10	0.04	0.06	0.06	0.07	0.07	0.05	0.06	0.06	0.06	0.06
Sup_2	0.04	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00
Ave_1	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.03	0.05	0.04	0.06	0.05	0.04	0.04	0.04
Ave_2	0.05	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00
								T=1000							
Sup_1	0.04	0.07	0.10	0.09	0.10	0.05	0.06	0.06	0.07	0.07	0.05	0.05	0.06	0.06	0.07
Sup_2	0.04	1.00	1.00	1.00	1.00	0.04	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00
Ave_1	0.05	0.05	0.06	0.05	0.06	0.05	0.05	0.04	0.04	0.04	0.06	0.04	0.04	0.03	0.04
Ave_2	0.05	1.00	1.00	1.00	1.00	0.06	1.00	1.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00

TABLE 10

State Dependence in Value Premium Forecasts (***, ** and * indicate significance at 2.5%, 5% and 10% levels respectively)

	Value Premium (D10BM-D1BM)				Value Premium (HML)			
	Sup_1	Sup_2	Ave_1	Ave_2	Sup_1	Sup_2	Ave_1	Ave_2
LDY	0.693	1.287*	0.104	0.594***	0.930	1.713***	0.268	0.773***
TBL	0.721	1.274*	0.119	0.556**	0.893	1.714***	0.278	0.746***
TMS	0.661	1.311*	0.071	0.606***	0.824	1.678***	0.194	0.763***
DEF	0.879	1.407**	0.225	0.723***	1.078	1.734***	0.452*	0.838***
LTR	0.717	1.256*	0.153	0.557**	1.091	1.694***	0.387*	0.787***
INFL	0.712	1.253*	0.117	0.545**	0.979	1.655***	0.307	0.713***
LDY+TBL	0.805	1.211	0.130	0.516***	0.960	1.663***	0.296	0.692***
LDY+TMS	0.823	1.215	0.117	0.492**	0.981	1.620***	0.291	0.663***
LDY+INFL	0.686	1.291*	0.098	0.598***	0.916	1.713***	0.259	0.772***
LDY+TBL+LTR	0.799	1.244*	0.139	0.549**	1.025	1.712***	0.335	0.756***
	$SupSup_1$	$SupSup_2$	$AveAve_1$	$AveAve_2$	$SupSup_1$	$SupSup_2$	$AveAve_1$	$AveAve_2$
LDY	1.106	1.392	0.100	0.594***	0.953	1.764***	0.076	0.627***
TBL	1.118	1.388	0.102	0.583***	0.931	1.774***	0.079	0.610***
TMS	1.179	1.417	0.122	0.622***	0.841	1.743***	0.052	0.643***
DEF	0.986	1.455	0.095	0.694***	1.097	1.795***	0.139	0.697***
LTR	1.065	1.370	0.090	0.583***	1.104	1.793***	0.116	0.659***
INFL	1.090	1.397	0.095	0.589***	1.001	1.749***	0.086	0.613***
LDY+TBL	1.117	1.392	0.108	0.582***	0.989	1.742***	0.084	0.605***
LDY+TMS	1.162	1.411	0.121	0.600***	0.986	1.725**	0.076	0.614***
LDY+INFL	1.108	1.392	0.100	0.596***	0.940	1.764***	0.074	0.623***
LDY+TBL+LTR	1.095	1.377	0.099	0.592***	1.050	1.789***	0.102	0.646***

PROOFS

In what follows we use $\mathbb{I}(F(q_t) \leq F(\gamma)) \equiv \mathbb{I}(\tilde{q}_t \leq \lambda)$ and let $\mathbf{w}_t(\lambda) = \mathbf{w}_t \mathbb{I}(\tilde{q}_t \leq \lambda)$ denote the $(p+1)$ vector of regressors whose elements interact on an element by element basis with the indicator vector $\mathbb{I}(\tilde{q}_t \leq \lambda)$.

PROOF OF PROPOSITION 1. Both our Sup_{1T} and Ave_{1T} statistics rely on the quantity

$$\begin{aligned}
\frac{C_{1T}(\pi_0, \lambda)}{\sqrt{T}} &= \frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} (e_{t+1|t} - \bar{e}_{T-[T\pi_0]}) \mathbb{I}(\tilde{q}_t \leq \lambda) \\
&= \frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} e_{t+1|t} \mathbb{I}(\tilde{q}_t \leq \lambda) - \left(\frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} e_{t+1|t} \right) \frac{1}{T - [T\pi_0]} \sum_{t=[T\pi_0]}^{T-1} \mathbb{I}(\tilde{q}_t \leq \lambda). \quad (25)
\end{aligned}$$

where $e_{t+1|t} = u_{t+1} - \mathbf{w}_t'(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})$. For the first component of (25) we have

$$\frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} e_{t+1|t} \mathbb{I}(\tilde{q}_t \leq \lambda) = \frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} u_{t+1} \mathbb{I}(\tilde{q}_t \leq \lambda) - \frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} \mathbf{w}_t(\lambda)'(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}). \quad (26)$$

Under our assumptions **A1** it is a standard result that $\boldsymbol{x}_{[Tr]}/\sqrt{T} \Rightarrow \boldsymbol{J}_c(r)$ with $\boldsymbol{J}_c(r)$ denoting a p-dimensional Ornstein-Uhlenbeck process (see Phillips (1987)). Letting $\bar{\boldsymbol{J}}_c = (1 \ \boldsymbol{J}_c)'$ and introducing the (p+1) dimensional normalisation matrix $\boldsymbol{D}_T = diag(\sqrt{T}, T, \ldots, T)$ standard FCLT based arguments combined with the CMT lead to

$$\boldsymbol{D}_T(\hat{\boldsymbol{\beta}}_{[Tr]} - \boldsymbol{\beta}) \Rightarrow \left(\int_0^r \bar{\boldsymbol{J}}_c(s) \bar{\boldsymbol{J}}_c(s)' ds \right)^{-1} \int_0^r \bar{\boldsymbol{J}}_c(s) dB_u \equiv \boldsymbol{Q}_\infty(r; c) \quad r \in [\pi_0, 1]. \quad (27)$$

From Lemma 1(f) in Gonzalo and Pitarakis (2012) it also follows that

$$\frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} \boldsymbol{w}_t(\lambda)' \boldsymbol{D}_T^{-1} \boldsymbol{D}_T (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}) = \lambda \frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} \boldsymbol{w}_t' \boldsymbol{D}_T^{-1} \boldsymbol{D}_T (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}) + o_p(1) \quad (28)$$

where

$$\frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} \boldsymbol{D}_T^{-1} \boldsymbol{w}_t \Rightarrow \int_{\pi_0}^1 \bar{\boldsymbol{J}}_c(r) dr \quad (29)$$

so that

$$\frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} \boldsymbol{w}_t(\lambda)' \boldsymbol{D}_T^{-1} \boldsymbol{D}_T (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}) \Rightarrow \lambda \int_{\pi_0}^1 \bar{\boldsymbol{J}}_c(r)' \boldsymbol{Q}_\infty(r; c) dr. \quad (30)$$

For the first term in the right hand side of (26), it follows from Theorem 1 in Caner and Hansen (2001) that

$$\frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} u_{t+1} \mathbb{I}(\tilde{q}_t \leq \lambda) \Rightarrow B_1(1 - \pi_0, \lambda) \equiv \sigma_u W_1(1 - \pi_0, \lambda) \quad (31)$$

which for π_0 given is also $\sigma_u \sqrt{1 - \pi_0} W_1(\lambda)$ and therefore by the CMT

$$\frac{1}{\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} e_{t+1|t} \mathbb{I}(\tilde{q}_t \leq \lambda) \Rightarrow \sqrt{1 - \pi_0} \sigma_u W_1(\lambda) - \lambda \int_{\pi_0}^1 \bar{\boldsymbol{J}}_c' \boldsymbol{Q}_\infty(r; c) dr. \quad (32)$$

For the second term in (25), we have

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t} = \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} u_{t+1} - \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} \boldsymbol{w}_t'(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}), \quad (33)$$

and using similar arguments to above, we obtain

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t} \Rightarrow \sigma_e\sqrt{1-\pi_0}W_1(1) - \int_{\pi_0}^{1} \bar{\boldsymbol{J}}_c(r)'\boldsymbol{Q}_\infty(r;c)dr. \quad (34)$$

By Lemma 1(a) in Gonzalo and Pitarakis (2012), we have

$$\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1} \mathbb{I}(\tilde{q}_t \leq \lambda) \xrightarrow{p} \lambda. \quad (35)$$

Therefore, by the CMT, combining (32),(34) and (35) gives

$$\frac{C_{1T}(\pi_0,\lambda)}{\sqrt{T}} \Rightarrow \sigma_u\sqrt{1-\pi_0}[W_1(\lambda) - \lambda W_1(1)].$$

Next, for the denominator of the test statistic, $\hat{\phi}_1$, we have

$$\begin{aligned}\hat{\phi}_1^2 &= \frac{1}{T}\sum_{T=[T\pi_0]}^{T-1}\left(e_{t+1|t} - \bar{e}_{T-[T\pi_0]}\right)^2 \\ &= \frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t}^2 - \frac{T-[T\pi_0]}{T}\left(\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t}\right)^2. \quad (36)\end{aligned}$$

Recalling that $e_{t+1|t} = u_{t+1} - \boldsymbol{w}_t'(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})$ it follows from our assumptions **A1** that

$$\frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t}^2 = \frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} u_{t+1}^2 + O_p\left(\frac{1}{T}\right) \quad (37)$$

and

$$\frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t} = \frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} u_{t+1} + O_p\left(\frac{1}{\sqrt{T}}\right) \quad (38)$$

so that

$$\hat{\phi}_1^2 \xrightarrow{p} (1-\pi_0)\sigma_u^2 \quad (39)$$

leading to

$$\frac{C_{1T}(\pi_0,\lambda)}{\hat{\phi}_1\sqrt{T}} \Rightarrow W_1(\lambda) - \lambda W_1(1) \equiv W_1^0(\lambda).$$

Finally, since *sup*, *ave* and $|.|$ are continuous transformations, the results in (15)-(16) for $i=1$ follow by successively applying the CMT.

We next treat the case of the Sup_{2T} and Ave_{2T} statistics that are based on the squared forecast errors. We write

$$\begin{aligned}\frac{C_{2T}(\pi_0,\lambda)}{\sqrt{T}} &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}\left(e^2_{t+1|t}-\bar{\tau}^2_{T-[T\pi_0]}\right)\mathbb{I}(\tilde{q}_t\leq\lambda)\\ &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}e^2_{t+1|t}\mathbb{I}(\tilde{q}_t\leq\lambda)-\left(\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(e_{t+1|t}-\bar{e}_{T-[T\pi_0]})^2\right)\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1}\mathbb{I}(\tilde{q}_t\leq\lambda)\end{aligned} \quad (40)$$

For the first term in (40), we have

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}e^2_{t+1|t}\mathbb{I}(\tilde{q}_t\leq\lambda) &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}u^2_{t+1}\mathbb{I}(\tilde{q}_t\leq\lambda)+\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta})'\boldsymbol{w}_t(\lambda)\boldsymbol{w}_t(\lambda)'(\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta})\\ &- 2\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta})'\boldsymbol{w}_t(\lambda)u_{t+1}.\end{aligned} \quad (41)$$

From (27)-(28) and the CMT it is straightforward to observe that the second term in the right hand side of (41) is $O_p(1/\sqrt{T})$ and thus vanishes asymptotically. Similarly, Theorem 2 in Caner and Hansen (2001) together with the CMT ensure that

$$\sum_{t=[T\pi_0]}^{T-1}(\boldsymbol{D}_T(\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}))'\boldsymbol{D}_T^{-1}\boldsymbol{w}_t(\lambda)u_{t+1} \;\Rightarrow\; \int_{\pi_0}^{1}\boldsymbol{Q}_\infty(r;c)\bar{\boldsymbol{J}}_c(r)dB_u(r,\lambda) \quad (42)$$

so that the third term in the right hand size of (41) is also $O_p(1/\sqrt{T})$ and vanishes asymptotically. Hence, we can write (41) as

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}e^2_{t+1|t}\mathbb{I}(\tilde{q}_t\leq\lambda)=\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}u^2_{t+1}\mathbb{I}(\tilde{q}_t\leq\lambda)+o_p(1) \quad (43)$$

and therefore

$$\frac{C_{2T}(\pi_0,\lambda)}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(u^2_{t+1}-\sigma^2_u)-\lambda\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T}(u^2_{t+1}-\sigma^2_u)+o_p(1) \quad (44)$$

so that from our assumptions **A2** we have

$$\frac{C_{2T}(\pi_0,\lambda)}{\sqrt{T}} \;\Rightarrow\; \phi_2(W_2(1-\pi_0,\lambda)-\lambda W_2(1-\pi_0,1)). \quad (45)$$

For the denominator, $\hat{\phi}_2$, we have

$$\begin{aligned}\hat{\phi}^2_2 &= \frac{1}{T}\sum_{t=[T\pi_0]}^{T-1}\left(e^2_{t+1|t}-\bar{\tau}^2_{T-[T\pi]}\right)^2\\ &= \frac{1}{T}\sum_{t=[T\pi_0]}^{T-1}\left(e^2_{t+1|t}-\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1}e^2_{t+1|t}\right)^2\\ &= \frac{T-[T\pi_0]}{T}\left(\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1}e^4_{t+1|t}\right)-\frac{T-[T\pi_0]}{T}\left(\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1}e^2_{t+1|t}\right)^2.\end{aligned} \quad (46)$$

Proceeding as above and invoking suitable Laws of Large Numbers, we have that

$$\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t}^4 \to_p \mathbb{E}(u_{t+1}^4) \tag{47}$$

and

$$\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t}^2 \to_p \mathbb{E}(u_{t+1}^2). \tag{48}$$

As $T-[T\pi_0]/T \to 1-\pi_0$ as $T\to\infty$ it follows that $\hat{\phi}_2^2 \xrightarrow{p} (1-\pi_0)[E[u_{t+1}^4]-[E[u_{t+1}^2]^2] \equiv (1-\pi_0)\phi_2^2$. Combining with (45) it follows that $C_{2T}(\pi_0,\lambda)/\hat{\phi}_2\sqrt{T} \Rightarrow W_2(\lambda)-\lambda W_2(1) \equiv W_2^0(\lambda)$. Finally, since *sup*, *ave* and $|.|$ are continuous transformations, the results in (15)-(16) for $i=2$ follow by successively applying the CMT.

PROOF OF PROPOSITION 2: The results in (17)-(18) are obtained using identical arguments as in the proof of Proposition 1, replacing π_0 with π so that the test statistics are viewed as functionals of both $\pi\in\Pi$ and $\lambda\in\Lambda$. The CMT then ensures that (17)-(18) hold.

PROOF OF PROPOSITION 3: We provide a proof for the $C_{1T}(\pi_0,\lambda)$ based statistics as the remaining cases follow identical lines. The DGP is given by (11) while the fitted model is (1). As $Sup_{1T} \geq |C_{1T}(\pi_0,\lambda_0)/\sqrt{T}\hat{\phi}_i|$ we only need to establish that $|C_{1T}(\pi_0,\lambda_0)/\sqrt{T}\hat{\phi}_i|$ diverges to infinity as $T\to\infty$. We intially treat the case of intercept shifts only with the DGP given by $y_{t+1} = \boldsymbol{\beta}'\boldsymbol{w}_t + \delta_0 I(q_t>\gamma_0)+u_{t+1}$. Under this setting it is straightforward to establish that $[(\hat{\beta}_{0,[Tr]}-\beta_0), \sqrt{T}(\hat{\boldsymbol{\beta}}_{1,[Tr]}-\boldsymbol{\beta}_1)] \Rightarrow [\delta_0(1-\lambda_0) \;\; \mathbf{0}]$ which we use within $e_{t+1|t} = u_{t+1}+\delta_0\mathbb{I}(\tilde{q}_t>\lambda_0)-(\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta})'\boldsymbol{w}_t$ and $e_{t+1|t}\mathbb{I}(\tilde{q}_t\leq\lambda_0) = u_{t+1}\mathbb{I}(\tilde{q}_t\leq\lambda_0) - (\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta})'\boldsymbol{w}_t(\lambda_0)$. From standard FCLT and CMT arguments together with $\sum_{t=[T\pi_0]+1}(\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta})'\boldsymbol{w}_t/T \Rightarrow 0$ we have $\sum_{t=[T\pi_0]+1} e_{t+1|t}/T \Rightarrow \delta_0(1-\pi_0)(1-\lambda_0)$ and

$$\frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t}\mathbb{I}(\tilde{q}_t\leq\lambda_0) \quad\Rightarrow\quad 0 \tag{49}$$

so that

$$\frac{1}{\sqrt{T}}\left|\frac{C_{1T}(\pi_0,\lambda_0)}{\sqrt{T}}\right| \quad\Rightarrow\quad (1-\pi_0)\lambda_0(1-\lambda_0)\delta_0. \tag{50}$$

The large sample behaviour of the normalising variance $\hat{\phi}_1^2$ also follows using similar steps with

$$\hat{\phi}_1^2 \quad\Rightarrow\quad (1-\pi_0)[\sigma_u^2+\delta_0^2\lambda_0(1-\lambda_0)]. \tag{51}$$

Combining (50) and (51) establishes the divergence to infinity of $C_{1T}(\pi_0,\lambda_0)/\hat{\phi}_1\sqrt{T}$ at rate $\sqrt{T}$.

For the slope shift only scenario the DGP is given by $y_{t+1} = \boldsymbol{\beta}'\boldsymbol{w}_t + \boldsymbol{\delta}_1'\boldsymbol{x}_t I(q_t>\gamma_0)+u_{t+1}$ and the OLS estimators obtained from (1) now satisfy $[(\hat{\beta}_{0,[Tr]}-\beta_0)/\sqrt{T}, (\hat{\boldsymbol{\beta}}_{1,[Tr]}-\boldsymbol{\beta}_1)] \Rightarrow [0 \;\; \boldsymbol{\delta}_1(1-\lambda_0)]$. This in turn allows us to establish that

$$\frac{1}{T\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1|t}\mathbb{I}(\tilde{q}_t\leq\lambda_0) \quad\Rightarrow\quad 0 \tag{52}$$

and

$$\frac{1}{T\sqrt{T}} \sum_{t=[T\pi_0]}^{T-1} e_{t+1|t} \;\;\Rightarrow\;\; -\lambda_0(1-\lambda_0)\boldsymbol{\delta}_1' \int_{\pi_0}^{1} \boldsymbol{J}_c(r)dr \tag{53}$$

so that

$$\frac{1}{T}\left|\frac{C_{1T}(\pi_0,\lambda_0)}{\sqrt{T}}\right| \;\;\Rightarrow\;\; \lambda_0(1-\lambda_0)\left|\boldsymbol{\delta}_1' \int_{\pi_0}^{1} \boldsymbol{J}_c(r)dr\right|. \tag{54}$$

The large sample behaviour of the normalising variance $\hat{\phi}_1^2$ now also follows using similar steps. Here we have

$$\frac{1}{T}\hat{\phi}_1^2 \;\;\Rightarrow\;\; \lambda_0(1-\lambda_0)\boldsymbol{\delta}_1' \int_{\pi_0}^{1} \boldsymbol{J}_c \boldsymbol{J}_c' \boldsymbol{\delta}_1 \tag{55}$$

so that combining (54)-(55) leads again to the result that $|C_{1T}(\pi_0,\lambda_0)/\hat{\phi}_1\sqrt{T}|$ diverges to infinity at rate $\sqrt{T}$.

PROOF OF PROPOSITION 4: (i) Under the local intercept shift only scenario the DGP is given by $y_{t+1} = \boldsymbol{\beta}'\boldsymbol{w}_t + (\delta_0/\sqrt{T})\mathbb{I}(\tilde{q}_t > \lambda_0) + u_{t+1}$ so that $e_{t+1|t} = u_{t+1} + \delta_0\mathbb{I}(\tilde{q}_t > \lambda_0) - (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})'\boldsymbol{w}_t$ and $e_{t+1|t}\mathbb{I}(\tilde{q}_t \leq \lambda) = u_{t+1}\mathbb{I}(\tilde{q}_t \leq \lambda) + \delta_0\mathbb{I}(\tilde{q}_t > \lambda_0)\mathbb{I}(\tilde{q}_t \leq \lambda) - (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})'\boldsymbol{w}_t(\lambda)$. Unless otherwise indicated in what follows all summations range from $t = [T\pi_0]$ to $T-1$. Using standard algebra and (16) we have

$$\begin{aligned} \frac{C_{1T}(\pi_0;\lambda,\lambda_0)}{\sqrt{T}} &= \left(\frac{\sum u_{t+1}\mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}} - \lambda\frac{\sum u_{t+1}}{\sqrt{T}}\right) - \delta_0(1-\pi_0)(\lambda \wedge \lambda_0 - \lambda\lambda_0) \\ &- \frac{1}{\sqrt{T}}\sum(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})'(\boldsymbol{w}_t(\lambda) - \lambda\boldsymbol{w}_t) + o_p(1) \end{aligned} \tag{56}$$

where me made use of $\sum \mathbb{I}(\tilde{q}_t \leq \lambda)\mathbb{I}(\tilde{q}_t > \lambda_0)/T \xrightarrow{p} (1-\pi_0)(\lambda - \lambda \wedge \lambda_0)$. From Lemma A1 in Gonzalo and Pitarakis (2012) the last term in the right hand side of (56) vanishes asymptotically while the first term converges weakly to $\sigma_u\sqrt{1-\pi_0}\; W_1^0(\lambda)$ from our result in Proposition 1. As under this local framework we also have $\hat{\phi}_1^2 \xrightarrow{p} (1-\pi)\sigma_u^2$ the result in Proposition 4(i) and the corresponding formulation in (19) follow directly. (ii) Under the local slope shift only scenario the DGP is given by $y_{t+1} = \boldsymbol{\beta}'\boldsymbol{w}_t + \boldsymbol{\delta}_1'\boldsymbol{x}_t\mathbb{I}(\tilde{q}_t > \lambda_0)/T + u_{t+1}$, leading to $e_{t+1|t} = u_{t+1} + \boldsymbol{\delta}_1'\boldsymbol{x}_t\mathbb{I}(\tilde{q}_t > \lambda_0)/T - (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})'\boldsymbol{w}_t$ and $e_{t+1|t}\mathbb{I}(\tilde{q}_t \leq \lambda) = u_{t+1}\mathbb{I}(\tilde{q}_t \leq \lambda) + \boldsymbol{\delta}_1'\boldsymbol{x}_t\mathbb{I}(\tilde{q}_t > \lambda_0)\mathbb{I}(\tilde{q}_t \leq \lambda)/T - (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})'\boldsymbol{w}_t(\lambda)$. We now have

$$\begin{aligned} \frac{C_{1T}(\pi_0;\lambda,\lambda_0)}{\sqrt{T}} &= \left(\frac{\sum u_{t+1}\mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}} - \lambda\frac{\sum u_{t+1}}{\sqrt{T}}\right) - (\lambda \wedge \lambda_0 - \lambda\lambda_0)\;\boldsymbol{\delta}_1'\frac{\sum \boldsymbol{x}_t}{T\sqrt{T}} \\ &- \frac{1}{\sqrt{T}}\sum(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})'(\boldsymbol{w}_t(\lambda) - \lambda\boldsymbol{w}_t) + o_p(1) \end{aligned} \tag{57}$$

with the last term in (57) vanishing asymptotically as in (56). It now follows that

$$\frac{C_{1T}(\pi_0;\lambda,\lambda_0)}{\sqrt{T}} \;\;\Rightarrow\;\; \sigma_u\sqrt{1-\pi_0}W_1^0(\lambda) - (\lambda \wedge \lambda_0 - \lambda\lambda_0)\boldsymbol{\delta}_1' \int_{\pi_0}^{1} \boldsymbol{J}_c \tag{58}$$

which when normalised with $\hat{\phi}_1$ leads to the desired result in Proposition 4(ii) and the associated formulation in (20).

PROOF OF PROPOSITION 5: The line of proof for establishing (i) and (ii) and the associated expressions in (21)-(21) parallels closely the arguments in the proof of Proposition 4. In what follows we concentrate on the slope shift scenario in (ii) as the case of an intercept shift follows from identical arguments. Under the local slope shift only scenario the DGP is given by $y_{t+1} = \boldsymbol{\beta}'\boldsymbol{w}_t + \boldsymbol{\delta}_1'\boldsymbol{x}_t\mathbb{I}(\tilde{q}_t > \lambda_0)/T^{3/4} + u_{t+1}$ leading to $e_{t+1|t} = u_{t+1} + \boldsymbol{\delta}_1'\boldsymbol{x}_t\mathbb{I}(\tilde{q}_t > \lambda_0)/T^{3/4} - (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})'\boldsymbol{w}_t$ and $e_{t+1|t}\mathbb{I}(\tilde{q}_t \leq \lambda) = u_{t+1}\mathbb{I}(\tilde{q}_t \leq \lambda) + \boldsymbol{\delta}_1'\boldsymbol{x}_t\mathbb{I}(\tilde{q}_t \leq \lambda)/T^{3/4} - (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})'\boldsymbol{w}_t(\lambda)$. Proceeding as in the proof of Proposition 1, appealing to standard FCLT and CMT arguments we have $T^{-1/4}\boldsymbol{D}_T(\hat{\boldsymbol{\beta}}_{[Tr]} - \boldsymbol{\beta}) \Rightarrow [0 \ \ (1-\lambda_0)\boldsymbol{\delta}_1]'$, leading to

$$\begin{aligned}\frac{\sum e_{t+1|t}^2}{\sqrt{T}} &= \frac{\sum u_{t+1}^2}{\sqrt{T}} + \boldsymbol{\delta}_1'\frac{\sum \boldsymbol{x}_t\boldsymbol{x}_t'\mathbb{I}(\tilde{q}_t > \lambda_0)}{T^2}\boldsymbol{\delta}_1 \\ &+ \sum(\frac{\boldsymbol{D}_T}{T^{1/4}}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}))'(\boldsymbol{D}_T^{-1}\ \boldsymbol{w}_t\boldsymbol{w}_t'\ \boldsymbol{D}_T^{-1})(\frac{\boldsymbol{D}_T}{T^{1/4}}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})) \\ &- \frac{2}{T}\sum(\frac{\boldsymbol{D}_T}{T^{1/4}}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}))'\boldsymbol{D}_T^{-1}\boldsymbol{w}_t\boldsymbol{x}_t'\ \mathbb{I}(\tilde{q}_t > \lambda_0)\boldsymbol{\delta}_1 + o_p(1), \end{aligned} \tag{59}$$

and

$$\begin{aligned}\frac{\sum e_{t+1|t}^2\mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}} &= \frac{\sum u_{t+1}^2\mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}} + \boldsymbol{\delta}_1'\frac{\sum \boldsymbol{x}_t\boldsymbol{x}_t'\mathbb{I}(\tilde{q}_t > \lambda_0)\mathbb{I}(\tilde{q}_t \leq \lambda)}{T^2}\boldsymbol{\delta}_1 \\ &+ \sum(\frac{\boldsymbol{D}_T}{T^{1/4}}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}))'(\boldsymbol{D}_T^{-1}\ \boldsymbol{w}_t(\lambda)\boldsymbol{w}_t(\lambda)'\ \boldsymbol{D}_T^{-1})(\frac{\boldsymbol{D}_T}{T^{1/4}}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})) \\ &- \frac{2}{T}\sum(\frac{\boldsymbol{D}_T}{T^{1/4}}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}))'\boldsymbol{D}_T^{-1}\boldsymbol{w}_t\boldsymbol{x}_t'\ \mathbb{I}(\tilde{q}_t > \lambda_0)\mathbb{I}(\tilde{q}_t \leq \lambda)\boldsymbol{\delta}_1 + o_p(1). \end{aligned} \tag{60}$$

Proceeding as in the proof of Proposition 1, appealing to CMT arguments we now have

$$\frac{\sum(e_{t+1|t}^2 - \sigma_u^2)}{\sqrt{T}} \Rightarrow \phi_2 W_2(1-\pi_0) + \lambda_0(1-\lambda_0)\boldsymbol{\delta}_1'\int_{\pi_0}^1 \boldsymbol{J}_c\boldsymbol{J}_c'\boldsymbol{\delta}_1 \tag{61}$$

and

$$\begin{aligned}\frac{\sum(e_{t+1|t}^2 - \sigma_u^2)\mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}} &\Rightarrow \phi_2 W_2(1-\pi_0, \lambda) - (\lambda - \lambda \wedge \lambda_0)(1-2\lambda_0)\boldsymbol{\delta}_1'\int_{\pi_0}^1 \boldsymbol{J}_c\boldsymbol{J}_c'\boldsymbol{\delta}_1 \\ &+ \lambda(1-\lambda_0)^2\boldsymbol{\delta}_1'\int_{\pi_0}^1 \boldsymbol{J}_c\boldsymbol{J}_c'\boldsymbol{\delta}_1. \end{aligned} \tag{62}$$

Recalling that

$$\frac{C_{2T}(\pi_0; \lambda, \lambda_0)}{\sqrt{T}} = \frac{\sum(e_{t+1|t}^2 - \sigma_u^2)\mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}} - \lambda\frac{\sum(e_{t+1|t}^2 - \sigma_u^2)}{\sqrt{T}} + o_p(1) \tag{63}$$

and since within our local slope shift setting we continue to have $\hat{\phi}_2^2 \xrightarrow{p} (1-\pi_0)\phi_2^2$, the result in (22) follows directly using (61)-(62) within (63) and rearranging. As a consequence, our statement in Proposition 5(ii) also follows.

PROOF OF PROPOSITION 6: (i) Here we have $e_{t+1|t} = u_{t+1} - (\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})'\boldsymbol{w}_t$ with u_{t+1} as in (24) so that

$$\begin{aligned}
\frac{C_{1T}(\pi_0; \lambda, \lambda_{0v})}{\sqrt{T}} &= \frac{\sum u_{t+1}\mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}} - \lambda \frac{\sum u_{t+1}}{\sqrt{T}} + o_p(1) \\
&= \sigma_1 \left(\frac{\sum \epsilon_{t+1}\mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}} - \lambda \frac{\sum \epsilon_{t+1}}{\sqrt{T}} \right) \\
&+ \frac{\sigma_2 - \sigma_1}{T^{\theta}} \left(\frac{\sum \epsilon_{t+1}(\mathbb{I}(\tilde{q}_t \leq \lambda) - \mathbb{I}(\tilde{q}_t \leq \lambda \wedge \lambda_{vo}))}{\sqrt{T}} - \lambda \frac{\sum \epsilon_{t+1}\mathbb{I}(\tilde{q}_t > \lambda_{v0})}{\sqrt{T}} \right) \\
&\Rightarrow \sigma_1 \sqrt{1 - \pi_0}(W_1(\lambda) - \lambda W_1(1)) \qquad (64)
\end{aligned}$$

which is an immediate consequence of assumption **B1**. Combining (64) with $\hat{\phi}_1^2 \xrightarrow{p} (1 - \pi_0)\sigma_1^2$ the result in Proposition 6(i) follows. (ii) With u_{t+1} specified as in (24) it continues to be the case (as in Proposition 1) that

$$\frac{C_{2T}(\pi_0; \lambda_{v0}, \lambda)}{\sqrt{T}} = \frac{\sum u_{t+1}^2 \mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}} - \left(\frac{\sum \mathbb{I}(\tilde{q}_t \leq \lambda)}{T - [T\pi_0]} \right) \frac{\sum u_{t+1}^2}{\sqrt{T}} + o_p(1) \qquad (65)$$

and upon rearranging using (24) we have

$$\begin{aligned}
\left| \frac{C_{2T}(\pi_0; \lambda_{v0}, \lambda)}{\sqrt{T}\hat{\phi}_2} \right| &= \Bigg| \sigma_1^2 \left(\frac{\sum(\epsilon_{t+1}^2 - 1)\mathbb{I}(\tilde{q}_t \leq \lambda)}{\sqrt{T}\hat{\phi}_2} - \lambda \frac{\sum(\epsilon_{t+1}^2 - 1)}{\sqrt{T}\hat{\phi}_2} \right) \\
&+ \frac{(\sigma_2 - \sigma_1)^2}{\hat{\phi}_2 T^{2\theta - \frac{1}{2}}} \left(\frac{\sum \epsilon_{t+1}^2 \mathbb{I}(\tilde{q}_t \leq \lambda)\mathbb{I}(\tilde{q}_t > \lambda_{v0})}{T} - \lambda \frac{\sum \epsilon_{t+1}^2 \mathbb{I}(\tilde{q}_t > \lambda_{v0})}{T} \right) \\
&+ \frac{2\sigma_1(\sigma_2 - \sigma_1)}{\hat{\phi}_2 T^{\theta - \frac{1}{2}}} \left(\frac{\sum \epsilon_{t+1}^2 \mathbb{I}(\tilde{q}_t \leq \lambda)\mathbb{I}(\tilde{q}_t > \lambda_{v0})}{T} - \lambda \frac{\sum \epsilon_{t+1}^2 \mathbb{I}(\tilde{q}_t > \lambda_{v0})}{T} \right) \Bigg| + o_p(1). \qquad (66)
\end{aligned}$$

Using Assumption **B2** and standard algebra we also have $\hat{\phi}_2^2 \xrightarrow{p} (1 - \pi_0)\sigma_1^4 E[\epsilon_{t+1}^2 - 1]^2$ so that the first term in the right hand side converges to $W_2^0(\lambda)$ while the remaining two terms trivially determine our results in Proposition 6(ii) and (iii).

REFERENCES

Billingsley, P. (1986). *Probability and Measure.* Wiley, 3^{rd} edition, New-York.

Breitung, J. and Demetrescu, M. (2015). 'Instrumental Variable and Variable Addition Based Inference in Predictive Regressions', *Journal of Econometrics*, 187, pp. 358-375.

Brown, R.L., Durbin, J., and Evans, J., (1975). 'Techniques for testing the constancy of regression relationship over time', *Journal of Royal Statistical Society Series (B)*, Vol. 37, pp. 149-163.

Campbell, J. Y. and Yogo, M. (2006). 'Efficient tests of stock return predictability', *Journal of Financial Economics,* Vol. 81, pp. 27-60.

Caner, M. and Hansen, B. E. 2001. 'Threshold Autoregression With a Unit Root,'. Econometrica', Vol. 69, pp 1555 - 1596.

Chauvet, M. and Potter, S. (2013). 'Forecasting Output', Handbook of Economic Forecasting vol 2, edited by Allan Timmermann and Graham Elliot, 2013, pp. 1-56.

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

Deng, A. and P. Perron (2008). 'A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change', *Journal of Econometrics*, Vol. 142, pp. 212-240.

Fama, E. F. and French, K. R. (2012). 'Size, value, and momentum in international stock returns', *Journal of Financial Economics*, Vol. 105, 457-472

Farmer, L., Schmidt, S., and Timmermann, A. (2018). 'Pockets of Predictability,' Unpublished Manuscript, SSRN No. 3152386.

Gargano, A., Pettenuzzo, D., and Timmermann, A. (2107). 'Bond Return Predictability: Economic Value and Links to the Macroeconomy', *Management Science,* forthcoming.

Georgiev, I., Harvey, D. I., Leybourne, S. J., and Taylor, R. A. M. (2018). 'Testing for parameter instability in predictive regression models', *Journal of Econometrics,* Vol. 204, 101-118.

Golez, B. and Koudijs, P. (2018). 'Four centuries of return predictability', *Journal of Financial Economics*, Vol. 127, pp. 248-263.

Gonzalo, J. and Pitarakis, J. (2012). 'Regime specific predictability in predictive regressions', *Journal of Business and Economic Statistics*, Vol. 30, 229-241.

Gonzalo, J. and Pitarakis, J. (2017). 'Inferring the predictability induced by a persistent regressor in a predictive threshold model,' *Journal of Business and Economic Statistics*, Vol. 35, 202-217.

Goyal, A. and Welch, I. (2014). 'Update to *A Comprehensive Look at the Empirical Performance of Equity Premium Prediction*', Unpublished Manuscript.

Gülen, H., Xing, Y., and Zhang, L. (2011). 'Value versus growth: Time-varying expected stock returns,' Financial Management, 2011, 381-407.

Jansson, M. and Moreira, M. J. (2006). 'Optimal Inference in Regression Models with Nearly Integrated Regressors', *Econometrica*, Vol. 74, pp. 681-714.

Kolev, G. I. and Karapandza, R., (2017). 'Out-of-sample equity premium predictability and sample-split invariance,' *Journal of Banking and Finance*, 34, 188-201.

Kostakis, A., Magdalinos, A., and Stamatogiannis, M. (2015). 'Robust Econometric Inference for Stock Return Predictability', *Review of Financial Studies,* Vol. 28, pp. 1506-1553.

Lakonishok, J., Shleifer, A., and Vishny, R.W., (1994). 'Contrarian investment, extrapolation, and risk', *Journal of Finance* 49, 1541 - 1578.

Li, J. and I. Tsiakas (2017). "Equity Premium Prediction: The Role of Economic and Statistical Constraints", Journal of Financial Markets 36, 56-75.

Long, C., Petkova, R., and Zhang, L. (2008) 'The expected value premium', *Journal of Financial Economics*, Vol. 87, pp. 269-280.

Page, E. S., (1954). 'Continuous inspection schemes,' *Biometrika,* Vol. 41, 100-114.

Petkova, R. and Zhang, L. (2005) 'Is value riskier than growth', *Journal of Financial Economics*, 78, pp. 187-202.

Phillips, P. C. B. (2015). 'Pitfalls and possibilities in predictive regression,' Cowles Foundation Discussion Paper No. 2003.

Pitarakis, J., (2002). 'Least Squares Estimation and Tests of Breaks in Mean and Variance under Misspecification', *Econometrics Journal,* Vol. 7, pp. 32-54.

Pitarakis, J., (2017). 'A Simple Approach for Diagnosing Instabilities in Predictive Regressions', *Oxford Bulletin of Economics and Statistics,* Vol. 79, 851-874.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). " Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy', *Review of Financial Studies*, Vol. 23, pp 821 - 862.

Rossi, B. and Inoue, A. (2012). 'Out-of-Sample Forecast Tests Robust to the Choice of Window Size', *Journal of Business and Economic Statistics*, 30, 432-453.

Steland, A. (2020). 'Testing and Estimating Change-Points in the Covariance Matrix of a High-Dimensional

Time Series', *Journal of Multivariate Analysis*, In Press.

Welch, I. and Goyal, A. (2008). 'A Comprehensive Look at the Empirical Performance of Equity Premium Prediction', *Review of Financial Studies*, Vol. 21, pp. 1455-1508.