UNIVERSITY OF Southampton
School of Social Sciences

Economics Department
University of Southampton
Southampton SO17 1BJ, UK

Discussion Papers in
Economics and Econometrics

Uncovering Regimes in Out of Sample Forecast Errors

Anibal Emiliano Da Silva (University of Southampton), Jesus Gonzalo (Universidad Carlos III de Madrid), Jean-Yves Pitarakis (University of Southampton),

No. 1803

ISSN 0966-4246

Uncovering Regimes in Out of Sample Forecast Errors

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

Jesus 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

April 7, 2018

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 multiple predictive regressions. These predictive regressions can accommodate multiple predictors that are highly persistent with potentially different degrees of persistence. Our method is also designed to be robust to the chosen starting window size so as to avert data mining concerns. Our tests are shown to be consistent and to lead to null distributions that are free of nuisance parameters and hence robust to the degree of persistence of the predictors.

JEL: C12, C22, C53, C58.

Keywords: Predictive Regressions, Predictability, Out of Sample Forecasting.

*Address for Correspondence: Jean-Yves Pitarakis, University of Southampton, Department of Economics, Southampton SO17 1BJ, United-Kingdom

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 recession 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, Pettenuzzo and Timmermann (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. 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 used within a cusum type construct and are therefore very easily implementable. Our operating framework is 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. 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, 2016) who 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. 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, 2016).

The structure of the paper is as follows. Section 2 introduces our main operating model together with the proposed test statistics. This is then followed by their large sample properties under both the null of a linear predictive regression and the alternative of a threshold type specification. Section 3 concentrates on the finite sample size and power properties of the tests across a broad range of relevant scenarios. Section 4 concludes. All proofs are relegated to the appendix.

2 Uncovering Regimes in Multiple Predictive Regressions

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

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

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

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

with $\boldsymbol{C} = diag(c_1, \ldots, c_p)$, $c_i > 0$ for $i = 1, \ldots, p$ 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 $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')$, $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1p})'$ 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

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

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 using $t = k_0, k_0 + 1, \ldots, T - 1$. Throughout this paper the initial estimation sample is viewed as a fraction $\pi \in (0, 1)$ of the full sample by setting $k = [T\pi]$ so that the sequence of out of sample forecast errors $\{e_{k+1|k}, e_{k+2|k+1}, \ldots, e_{T-1|T-2}, e_{T|T-1}\}$ is of length $T - k$. In order to simplify notation and as this paper is solely restricted to one-period ahead forecasts, forecast errors as defined in (3) will be denoted e_{t+1} for $t = k, \ldots, T - 1$. Given the potential sensitivity of the accuracy of forecasts to the choice of k (the length of the initial sample used for estimation) in what follows we will be interested in assessing the presence of regimes in the $e_t's$ 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).

Given our operating model in (1) our main concern is to develop a simple approach to assessing the presence of economically meaningful regimes within key characteristics of the forecast errors in (3). 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}'s and e^2_{t+1}'s across different regimes dictated by a 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 (e.g. growth rate in industrial production, diffusion indices, sentiment indicators etc.).

Our proposed inferences will rely on normalised versions of functionals of the following quantities

$$C_{1T}(k, \gamma) = \sum_{t=k}^{T-1}(e_{t+1} - \bar{e}_{T-k})I(q_t \leq \gamma) \tag{4}$$
$$C_{2T}(k, \gamma) = \sum_{t=k}^{T-1}(e^2_{t+1} - \hat{\tau}^2_{T-k})I(q_t \leq \gamma). \tag{5}$$

where q_t denotes the threshold variable, $\bar{e}_{T-k} = \sum_{t=k}^{T-1} e_{t+1}/(T-k)$ and $\hat{\tau}^2_{T-k} = \sum_{t=k}^{T-1}(e_{t+1} - \bar{e}_{T-k})^2/(T-k)$. Note that the quantity $\bar{e}_{T-k}$ is maintained in (4) as the $e'_t s$ 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 distribution function $F(.)$ so that when necessary and convenient we make use of the property $I(q_t \leq \gamma) \equiv I(F(q_t) \leq \lambda)$ 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 λ 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 so that two alternative choices of π may lead to very different forecast accuracy outcomes. Our formulations in (4)-(5) will allow us to construct test statistics that take this dependence on π into account and hence lead to outcomes 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 $C_i(\pi, \lambda)$'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 $i = 1, 2$ distiinguishes between the statistic implemented on the level of the forecast errors and their squares respectively and with $\hat{\phi}_1$ and $\hat{\phi}_2$ denoting the associated variance estimators with $\hat{\phi}^2_1 = \sum_{t=k}^{T-1}(e_{t+1} - \bar{e}_{T-k})^2/T$ and $\hat{\phi}^2_2 = \sum_{t=k}^{T-1}(e^2_{t+1} - \bar{\tau}_{T-k})^2/T$.

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).

More specifically we introduce the following alternative test statistic formulations

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

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

The above statistics bear strong resemblance with traditional cusum and cusumsq statistics commonly used in the changepoint literature. Instead of cumulating the quantity of interest up to a potential changepoint we 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 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 of (1) is given by

$$y_{t+1} = \boldsymbol{w}_t'\boldsymbol{\beta}_1 I(q_t \leq \gamma) + \boldsymbol{w}_t'\boldsymbol{\beta}_2 I(q_t > \gamma) + u_{t+1}. \tag{11}$$

Before proceeding with the asymptotic properties of the above statistics under the null hypothesis of a linear predictive regression we introduce our first set of operating assumptions.

Assumptions A. *(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* $\sum_j \boldsymbol{\Psi}_j$ *has full rank,* $\boldsymbol{\Psi}_0 = I_p$ *and* $\sum_{j=0}^{\infty} ||\boldsymbol{\Psi}_j|| < \infty$. *(ii)* $\boldsymbol{\eta}_t = (u_t, \boldsymbol{\epsilon}_{vt})'$ *is a martingale difference sequence with respect to the filtration* $\mathcal{F}_t^A = \sigma(\boldsymbol{\eta}_s, u_{qs}|s \leq t)$ *so that* $E[\boldsymbol{\eta}_t|\mathcal{F}_{t-1}^A] = 0$ *and* $E[\boldsymbol{\eta}_t\boldsymbol{\eta}_t'|\mathcal{F}_{t-1}^A] = \boldsymbol{\Sigma}_\eta > 0$. *(iii)* $E||\boldsymbol{\eta}_t||^4 < \infty$ *(iv) The probability density function* $f_q(.)$ *of* q_t *is bounded away from zero and* ∞ *over each bounded set. (v) The sequence* $\{u_{qt}\}$ *is strictly stationary, ergodic, strong mixing with mixing numbers* α_m *such that* $\sum_{m=1}^{\infty} \alpha^{\frac{1}{m}-\frac{1}{r}} < \infty$ *for some* $r > 2$.

Assumptions A(i)-(v) above mimic closely the environment considered in Gonzalo and Pitarakis (2012, 2016) and excluding the probabilistic properties of q_t have 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. Note that the covariance between the $u_t's$ and the shocks $\boldsymbol{\epsilon}_{vt}$ associated with the predictors, say $\boldsymbol{\Sigma}_\eta = \{\{\sigma_u^2, \sigma'_{u\epsilon_v}\}, \{\sigma_{u\epsilon_v}, \Sigma_{\epsilon_v\epsilon_v}\}\}$, can be non-diagonal allowing them to be correlated as commonly observed in applications involving returns and dividend yields for instance. Here it is also important to highlight the fact that our assumptions allow the threshold variable q_t and the shocks underlying the predictive regression in (1) to be contemporaneously correlated.

An implication of the above assumptions is that an FCLT holds for $\boldsymbol{z}_t = (u_t, u_t I(q_{t-1} \leq \lambda), \boldsymbol{v}_t)'$ which we write as $T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \boldsymbol{z}_t \Rightarrow (B_1(r), B_1(r,\lambda), \boldsymbol{B}_v(r))' \equiv BM(\Omega)$ with $\Omega = \sum_{=-\infty}^{\infty} E[\boldsymbol{z}_0\boldsymbol{z}_k']$. Here $B_1(r,\lambda)$ is a two-parameter Brownian Motion as introduced in a related context in Caner and Hansen (2001) i.e. a zero mean Gaussian process with covariance kernel $\sigma_u^2(r_1 \wedge r_2)(\lambda_1 \wedge \lambda_2)$. Our assumptions under **A**(i)-(ii) also imply a particular structure for Ω as both serial correlation and heteroskedasticity are ruled out from the dynamics of the $u_t's$. More specifically we can formulate Ω as

$$\Omega = \begin{bmatrix} \sigma_u^2 & \lambda\sigma_u^2 & \sigma'_{uv}\Psi(1) \\ \lambda\sigma_u^2 & \lambda\sigma_u^2 & \lambda\sigma'_{uv}\Psi(1) \\ \sigma_{uv}\Psi(1) & \lambda\sigma_{uv}\Psi(1) & \Psi(1)\Sigma_{e_v}\Psi(1)' \end{bmatrix} \tag{12}$$

and for later use we also write $B_1(r,\lambda) = \phi_1 W(r,\lambda)$ with $\phi_1^2 \equiv \sigma_u^2$ and $W(r,\lambda)$ a two parameter standard Brownian Motion. In order to handle the asymptotics associated with the use of squared CUSUMs as in (5) we also need to supplement Assumptions **A** with additional restrictions involving the dynamics of the u_t^2 sequence.

Assumptions B. *(i) The zero mean sequence $g_t = (u_t^2 - \sigma_u^2)$ satisfies the invariance principle $T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} (u_t^2 - \sigma_u^2) \Rightarrow B_2(r) \equiv BM(\phi_2^2)$ with $\phi_2 = E[g_t^2]$, (ii) $E[u_{t+1}^4 | \mathcal{F}_t^A] = E[u_{t+1}^4]$.*

The requirement under Assumptions **B**(ii) retricts the higher moments of the $u_t's$ in the way they interact with past values of the threshold variable. From Caner and Hansen (2001) combining Assumptions A and B above allows us to operate with a suitable FCLT type result for the marked empirical process $T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} (u_t^2 - \sigma_u^2) I(q_{t-1} \leq \lambda) \Rightarrow B_2(r,\lambda) \equiv \phi_2 W(r,\lambda)$.

We intially focus on the large sample behaviour of (7)-(8) which take the recursion starting point as given at $\pi = \pi_0$. The main result on their limiting behaviour is summarised in Proposition 1 below.

Proposition 1 *Under assumptions **A** for $i = 1$ and under assumptions **A** and **B** for $i = 2$ we have as $T \to \infty$*

$$Sup_{iT} \Rightarrow \sup_{\lambda \in \Lambda} |W^0(\lambda)| \quad i = 1,2 \tag{13}$$

$$Ave_{iT} \Rightarrow \int_{\Lambda} W^0(\lambda)^2 d\lambda \quad i = 1,2 \tag{14}$$

with $W^0(\lambda)$ 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**. Also noteworthy is the fact that the limiting distributions are the same regardless of whether we implement our tests on the $e_t's$ or $e_t^2{'}s$. Another convenient feature of (13)-(14) 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), Hall and Werner (1980)). In the case of (13) the 10%, 5% and 1% cutoffs are given by 1.224, 1.358 and 1.628. For the distribution in (14) 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 (13) 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 **A** for $i = 1$ and under assumptions **A** and **B** for $i = 2$ we have as $T \to \infty$*

$$SupSup_{iT} \Rightarrow \sup_{\pi \in \Pi} \sup_{\lambda \in \Lambda} \left| \frac{W(1-\pi,\lambda) - \lambda W(1-\pi,1)}{\sqrt{1-\pi}} \right| \quad i = 1,2 \tag{15}$$

$$AveAve_{iT} \Rightarrow \frac{1}{\pi_b - \pi_a} \int_{\Pi} \int_{\Lambda} \left(\frac{W(1-\pi,\lambda) - \lambda W(1-\pi,1)}{\sqrt{1-\pi}} \right)^2 d\pi d\lambda. \quad i = 1,2 \tag{16}$$

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

We continute to note that the limiting distributions in (15)-(16) are free of nuisance parameters. Both distributions are functionals of a process $K(\zeta,\lambda) = W(\zeta,\lambda) - \lambda W(\zeta,1)$ commonly known as a Kiefer-Müller processes 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 the distributions will depend on the chosen interval for $\pi\in[\pi_a,\pi_b]$ and to our knowldge tabulations for such processes are not available in the literature. To approximate these distributions we make use of the approximation $K(1-\pi,\lambda)\approx\sum_{t=[T\pi]}^{T}(I(\mathcal{U}_t\leq\lambda)-\lambda)/\sqrt{T}$ with $\mathcal{U}_t$ denoting a uniform on $[0,1]$ i.i.d. sequence. Once properly normalised this quantity allows us to construct quantiles of the right hand side of (15)-(16) by first obtaining supremum and/or averages across $\lambda\in[0,1]$ for *each* $\pi\in\Pi$ and subsequently taking the supremum of the latter profile across $\pi's$. These simulations are conducted using $T=1000$ across $N=10000$ replications and key quantiles are presented in Table 1 across a selection of Π intervals.

Before proceeding with the finite sample properties of our test statistics it is also important to assess their ability to detect threshold effects when the latter are present. We focus our interest on alternatives as in (11) which we reformulate as

$$y_{t+1} = \boldsymbol{w}_t'\boldsymbol{\beta} + \boldsymbol{w}_t'\boldsymbol{\delta}_0 I(q_t>\gamma_0) + u_{t+1}. \tag{17}$$

and are intially interested in establishing the large sample behaviour of our test statistics in (7)-(8) under such a fixed departure from (1). Proposition 3 below summarises the main result which we subsequently follow up with a more through focus on the factors affecting power.

Proposition 3. *Under (17) and the same assumption structure as in Propositions 1-2 we have $\{Sup_{iT},Ave_{iT}\}=O_p(\sqrt{T})$ and $\{SupSup_{iT},AveAve_{iT}\}=O_p(\sqrt{T})$ for $i=1$ and provided that $\lambda_0\neq 0.5$ for $i=2$.*

In order to gain more tangible insights into the power properties of our test statistics we next follow Deng and Perron (2008) who advocated the usefulness and greater informativeness of adopting a non-local approach to power via the use of suitable expansions of test statistics designed to highlight the role of key DGP parameters. These will be equally useful in our present context when it comes to interpreting our simulation based findings. With no loss of generality we specialise our specification in (17) to a single predictor context and for clarity purposes reformulate it as

$$y_{t+1} = (\beta_0+\beta_1)x_t + (\delta_0+\delta_1 x_t)I(q_t>\lambda_0) + u_{t+1}. \tag{18}$$

In what follows we concentrate our attention on the quantity $|C_{iT}(\pi_0,\lambda)|/\sqrt{T}\hat{\phi}_i$ in (7)-(8) as results for its squared version can be inferred trivially. Furthermore as $Sup_{iT}\geq|C_{iT}(\pi_0,\lambda_0)|/\sqrt{T\hat{\phi}_i^2}$ we focus on the behaviour of $C_{iT}(\pi_0,\lambda_0)/\sqrt{T\hat{\phi}_i^2}$. The factors which affect the large sample behaviour of this latter object will be equally relevant for all our test statistics. We also distinguish across three scenarios depending on whether the intercepts only, slopes only or both intercepts and slopes are allowed to shift. Our main results are summarised in the following Proposition.

Proposition 4. *Under (18) and the same assumption structure as in Propositions 1-2 we have (i) under*

$\beta_1 = \delta_1$ *(intercept shifts only)*

$$\frac{1}{\sqrt{T}}\left|\frac{C_{1T}(\pi_0,\lambda_0)}{\sqrt{T}\hat{\phi}_1}\right| \Rightarrow \left|\frac{\delta_0\lambda_0(1-\lambda_0)\sqrt{1-\pi_0}}{\sqrt{\sigma_u^2+\delta_0^2\lambda_0(1-\lambda_0)}}\right| \tag{19}$$

$$\frac{1}{\sqrt{T}}\left|\frac{C_{2T}(\pi_0,\lambda_0)}{\sqrt{T}\hat{\phi}_2}\right| \Rightarrow \left|\frac{\delta_0^2\lambda_0(1-\lambda_0)(1-2\lambda_0)\sqrt{1-\pi_0}}{\sqrt{E[u_t^4]+m_\infty(\delta_0,\lambda_0,E[u_t^3],\sigma_u^2)}}.\right| \tag{20}$$

and under $\beta_0 = \delta_0$ *(slope shifts only) we have*

$$\frac{1}{\sqrt{T}}\left|\frac{C_{1T}(\pi_0,\lambda_0)}{\sqrt{T}\hat{\phi}_1}\right| \Rightarrow \left|\frac{-\delta_1\lambda_0(1-\lambda_0)\int J_c}{\sqrt{\delta_1^2\lambda_0(1-\lambda_0)\int_{\pi_0}^1 J_c^2}}\right| \equiv \frac{\sqrt{\lambda_0(1-\lambda_0)}|\int_{\pi_0}^1 J_c|}{\sqrt{\int_{\pi_0}^1 J_c^2}} \tag{21}$$

$$\frac{1}{\sqrt{T}}\left|\frac{C_{2T}(\pi_0,\lambda_0)}{\sqrt{T}\hat{\phi}_2}\right| \Rightarrow \left|\frac{\lambda_0(1-\lambda_0)(1-2\lambda_0)\int_{\pi_0}^1 J_c^2}{\sqrt{n_\infty(\lambda_0)\int_{\pi_0}^1 J_c^4}}\right| \tag{22}$$

wherethe functionals $m_\infty(.)$ *and* $n_\infty(.)$ *are defined in the appendix*

The above formulations are particularly informative for assessing the power properties of our test statistics and for providing the theoretical background for our simulation based results documented further below. It is also interesting to highlight the similarities between our expressions in (19)-(20) and the outcomes presented in Theorems 1-2 of Deng and Perron (2008, pp. 216-220) in a structural break context.

Focusing first on the scenario where solely the intercepts are allowed to shift, from (19)-(20) we note that as δ_0 increases the expressions are strictly increasing. Furthermore the magnitude of the non-centrality parameter c does not appear to play any role when solely intercepts are characterised by threshold effects. This suggests that in our finite sample power experiments we should not see much variation in power properties across different magnitudes of the $c_i's$. Perhaps more interestingly we note that our tests based on the squared forecast errors will be problematic when the true threshold parameter splits the sample equally (i.e. $\lambda_0 = 0.5$), a phenomenon which is clearly noted from our simulations further below and is also known to occur in the context of CUSUMSQ type statistics for structural breaks when the true break point lies in the middle of the sample. Note that this power problem that occurs at $\lambda_0 = 0.5$ does not occur for our C_{1T} based statistics designed to conduct inferences about the mean of the forecast errors.

When slopes are characterised by threshold effects we note from (21)-(22) that power is now influenced by the noncentrality parameter c through the role played by the stochastic integrals in the J_c process. Perhaps more importantly our results highlight a potentially problematic feature of the tests in this context as the δ_1 parameter does not appear in their formulations. This implies that no matter how far away we are from the null model along the dimension of the δ_1 parameter and for a given sample size T power will be unaffected by the magnitude of δ_1. Consistency of the test is maintained however as $T \to \infty$ and is a consequence of the fact that λ_0 is away from either 0 or 1 when there are threshold effects. Note also the indirect role played by δ_0 as its magnitude will influence the magnitude of λ_0.

It is important to point out that the tests proposed here do not suffer from power non-monotonicity problems as it may often occur for cusum based tests. There will be instances however where power is greatly affected by the location of the true threshold (i.e. the magnitude of λ_0) an issue we investigate in our finite sample based simulations below.

3 Finite Sample Size and Power Properties

We initially document the finite sample adequacy of the distributions presented in (13)-(16) 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 (17). Particular emphasis is placed on the robustness of outcomes to the magnitude of the $c_i's$. Throughout both our size and power experiments we parameterise the threshold variable as the stationary AR(1) process $q_t = 0.5q_{t-1} + u_{qt}$ while the predictors are given by $x_{it} = (1 - c_i/T)x_{it-1} + v_{it}$ with $v_{it} = 0.5v_{it-1} + \epsilon_{ivt}$. The covariance matrix of $(u_t, \epsilon_{1vt}, \epsilon_{2vt}, \ldots, \epsilon_{pvt}, u_{qt})$ allows for non-zero contemporaneous correlations between all disturbances. In line with the empiricial literature on the predictability of returns with valuation ratios the covariances between the $u_t's$ in (1) or (17) 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.

We initially focus on the Sup_i and Ave_i 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 $\pi_0 = 0.25$). The DGP is given by (1)-(2) and 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 forced to have the same degree of persistence. Table 3 provides additional outcomes for a larger number of predictors and scenarios where each predictor may have a different non-centrality parameter.

Empirical sizes are seen to match their nominal counterparts very closely across all sample sizes except perhaps for a mild undersizeness characterising the Sup_2 based tests that rely on the squared forecast errors. As expected from our asymptotics it is equally important to note the robustness of outcomes to alternative choices of the non-centrality parameter and the chosen number of predictors. Under $T = 1000$ the grand average across all size estimates, number of predictors and all statistics was 4.76% for $c = 1$, 4.94% for $c = 20$ and 4.98% for $c = 40$. Table 3 repeats the exercise by allowing the multiple predictors to be characterised by different magnitudes of the $c_i's$. Overall we continue to note a good adequacy of the finite sample sizes to their nominal counterparts across all configurations with the Ave_i statistics occasionally displaying a mild degree of oversizeness while Sup_2 remains mildly undersized for $T = 400$.

We next consider the case of the $SupSup_i$ and $AveAve_i$ statistics in (9)-(10) which are designed to be robust to the chosen starting period of the recursions. Our experiments use the interval $\Pi = [0.50, 0.75]$ for scanning across all recursion starting points. The corresponding 5% quantile cut-offs of the two test statistics are given by 1.596 and 0.439 respectively (see Table 1) and results on specifications with $p = 1, 2$ and $p = 3$ predictors are presented in Table 4. Although the average based tests continue to display excellent finite sample size properties it is important to recognise the weakness of the $SupSup_2$ statistic which tends to display empirical sizes in the vicinity of 3% under $T = 400$ and $T = 600$ and regardless of the number of predictors used. Nevertheless its empirical size does improve as we approach samples of size $T = 1000$. Also noteworthy is the robustness of finite sample size properties to the number of predictors and their near integration parameters, as supported by the underlying asymptotics.

Table 5 aims to explore the power properties of our test statistics under fixed departures from the null of a linear predictive regression. The key point here is to highlight the consistency of the test when the intercept and slope parameters move away from α_{10} and β_{10} across alternative threshold parameter locations...

4 Conclusions

We have proposed a series of test statistics for assessing the presence of regimes within out of sample forecast errors generated from multiple predictive regressions. Convenient features of our proposed methods include their robustness to the degree of persistence of predictors and to the starting points of forecast recursions. A broad range of simulation experiments subsequently established that our test statistics are well behaved in finite samples, matching closely their asymptotic distributions and displaying good power.

TABLES

TABLE 1

Quantiles of the Asymptotic Distributions of $SupSup_i$ and $AveAve_i$ statistics

	10%	5%	2.5%	1%
	$\pi \in [0.25, 0.75]$			
$SupSup$	1.513	1.643	1.753	1.903
$AveAve$	0.326	0.436	0.540	0.670
	$\pi \in [0.25, 0.90]$			
$SupSup$				
$AveAve$				
	$\pi \in [0.50, 0.75]$			
$SupSup$	1.461	1.596	1.707	1.863
$AveAve$	0.332	0.439	0.557	0.709
	$\pi \in [0.50, 0.90]$			
$SupSup$	1.558	1.685	1.795	1.939
$AveAve$	0.315	0.412	0.506	0.656

TABLE 2

Empirical Size (5% Nominal) of Sup_i and Ave_i statistics under Single and Multiple Predictors ($p = 1, 2, 3$)

	$p=1$				$p=2$				$p=3$			
	Sup_1	Sup_2	Ave_1	Ave_2	Sup_1	Sup_2	Ave_1	Ave_2	Sup_1	Sup_2	Ave_1	Ave_2
						$c=1$						
$T=400$	4.33	3.90	5.46	5.34	4.43	3.58	6.23	4.91	5.48	3.77	7.29	5.11
$T=600$	4.29	3.82	5.94	4.95	4.61	4.26	5.81	5.65	5.36	3.92	6.51	5.40
$T=1000$	4.70	3.87	5.44	4.87	4.35	4.11	5.22	4.84	4.80	4.11	5.70	5.08
						$c=20$						
$T=400$	4.17	3.47	5.58	4.89	4.41	3.93	5.89	5.07	5.53	4.12	6.88	5.25
$T=600$	4.78	4.05	5.59	5.19	4.62	3.89	5.66	5.21	4.91	3.95	6.04	4.98
$T=1000$	4.53	4.19	5.41	5.31	4.74	4.39	5.68	5.20	4.35	4.29	5.91	5.29
						$c=40$						
$T=400$	4.40	3.67	5.87	4.89	4.35	3.65	5.75	4.94	4.67	3.70	6.30	4.68
$T=600$	4.13	4.05	5.16	5.26	4.63	3.63	5.73	5.02	5.06	3.87	6.55	5.02
$T=1000$	4.87	4.40	5.71	5.12	4.39	4.39	5.36	5.37	4.96	4.25	5.86	5.13

TABLE 3

Empirical Size (5% Nominal) under $p = 3, 5$ predictors and different near integration parameters

	$p = 3$				$p = 5$			
	Sup_1	Sup_2	Ave_1	Ave_2	Sup_1	Sup_2	Ave_1	Ave_2
	$\{c_1, c_2, c_3\} = \{1, 20, 40\}$				$\{c_1, c_2, c_3, c_4, c_5\} = \{1, 1, 1, 20, 40\}$			
$T = 400$	5.15	3.7	6.73	4.84	4.73	3.68	5.86	5.21
$T = 600$	5.06	4.03	6.29	5.24	4.97	3.84	6.22	5.12
$T = 1000$	4.43	4.16	5.56	5.08	4.65	4.35	5.58	5.37
	$\{c_1, c_2, c_3\} = \{1, 20, 20\}$				$\{c_1, c_2, c_3, c_4, c_5\} = \{1, 20, 20, 20, 40\}$			
$T = 400$	5.07	3.69	6.72	4.91	4.68	3.87	6.00	5.21
$T = 600$	5.15	4.03	6.5	5.27	4.95	3.80	6.12	5.05
$T = 1000$	4.53	4.14	5.71	5.05	4.67	4.54	5.57	5.48

TABLE 4

Empirical Size of $SupSup_i$ and $AveAve_i$ statistics (5% Nominal)

	$p=1$				$p=2$				$p=3$			
	$SupSup_1$	$SupSup_2$	$AveAve_1$	$AveAve_2$	$SupSup_1$	$SupSup_2$	$AveAve_1$	$AveAve_2$	$SupSup_1$	$SupSup_2$	$AveAve_1$	$AveAve_2$
						$c=1$						
$T=400$	3.75	3.00	4.88	4.91	3.82	3.04	5.44	4.92	4.42	3.04	6.02	4.98
$T=600$	3.97	3.50	4.89	5.00	3.67	3.10	4.68	4.72	3.96	3.34	5.24	4.78
$T=1000$	4.33	3.38	5.16	4.85	4.48	3.76	4.54	4.80	3.92	3.92	4.76	5.12
						$c=20$						
$T=400$	3.70	2.98	4.88	4.80	3.92	2.92	5.74	4.82	4.44	4.14	6.04	4.70
$T=600$	3.56	3.36	4.56	4.92	3.44	3.04	4.34	4.54	4.24	3.36	5.10	4.54
$T=1000$	4.42	3.36	5.38	4.58	4.48	3.80	4.70	4.82	3.92	4.16	4.66	5.16
						$c=40$						
$T=400$	3.78	3.04	4.84	4.80	3.96	2.92	6.10	4.78	3.78	2.82	5.44	4.68
$T=600$	4.26	3.68	5.18	5.26	3.80	3.30	4.84	5.20	4.84	3.52	6.28	5.20
$T=1000$	4.26	3.42	4.94	4.94	3.92	3.90	4.92	4.68	4.30	3.70	5.00	4.72

TABLE 5

Finite Sample power $SupSup_i$ and $AveAve_i$ with T=600

α_{20}	0.17	0.31	0.45	0.59	0.73	0.87	1.01	1.16	1.3	1.44	1.58	1.72	1.86	2	2.15
β_{20}	0.15	0.29	0.43	0.58	0.72	0.86	1	1.14	1.28	1.42	1.57	1.71	1.85	1.99	2.13
c=1								$\pi \in [0.25, 0.75]$							
$SupSup_1$															
$\lambda_0 = -0.3$	0.993	0.999	0.999	1	0.999	1	1	1	1	1	1	1	1	1	0.999
$\lambda_0 = -0.2$	0.993	0.999	1	1	1	1	1	1	1	1	1	1	1	1	1
$\lambda_0 = 0$	0.99	0.998	0.999	1	1	1	1	1	0.999	1	1	1	1	1	1
$SupSup_2$															
$\lambda_0 = -0.3$	0.740	0.938	0.98	0.984	0.991	0.993	0.990	0.995	0.993	0.993	0.994	0.994	0.997	0.996	0.995
$\lambda_0 = -0.2$	0.573	0.829	0.902	0.929	0.943	0.949	0.959	0.942	0.952	0.942	0.969	0.962	0.958	0.948	0.964
$\lambda_0 = 0$	0.301	0.546	0.655	0.703	0.732	0.759	0.784	0.779	0.772	0.793	0.782	0.779	0.768	0.771	0.790
$AveAve_1$															
$\lambda_0 = -0.3$	0.983	0.996	0.995	1	0.999	1	1	0.997	0.999	0.998	0.998	1	0.998	1	1
$\lambda_0 = -0.2$	0.983	0.996	0.999	0.997	1	0.999	1	0.999	0.999	0.998	1	0.998	1	1	1
$\lambda_0 = 0$	0.976	0.994	0.995	1	0.996	0.998	1	0.997	0.999	0.997	1	0.998	1	1	1
$AveAve_2$															
$\lambda_0 = -0.3$	0.706	0.907	0.966	0.970	0.98	0.989	0.987	0.989	0.986	0.989	0.988	0.984	0.996	0.992	0.990
$\lambda_0 = -0.2$	0.536	0.785	0.869	0.901	0.915	0.912	0.921	0.91	0.916	0.918	0.932	0.929	0.935	0.921	0.929
$\lambda_0 = 0$	0.248	0.448	0.557	0.574	0.61	0.619	0.639	0.63	0.616	0.654	0.656	0.653	0.629	0.651	0.655

PROOFS

PROOF OF PROPOSITION 1. For notational simplicity we set $\beta_0 = 0$ and treat the case of a single predictor. Both our Sup_{1T} and Ave_{1T} statistics rely on the quantity

$$\begin{aligned}\frac{C_{1T}(\pi_0,\lambda)}{\sqrt{T}} &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(e_{t+1}-\bar{e}_{T-[T\pi_0]})\mathbb{I}(q_t \leq \lambda) \\ &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1}\mathbb{I}(q_t \leq \lambda) - \left(\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1}\right)\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1}\mathbb{I}(q_t \leq \lambda). \qquad (23)\end{aligned}$$

For the first term in (23), we can write

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1}\mathbb{I}(q_t \leq \lambda) &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}\Big(u_{t+1}-(\hat{\beta}_t-\beta)x_t\Big)\mathbb{I}(q_t \leq \lambda) \\ &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} u_{t+1}\mathbb{I}(q_t \leq \lambda) - \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(\hat{\beta}_t-\beta)x_t\mathbb{I}(q_t \leq \lambda). \qquad (24)\end{aligned}$$

Under our assumptions it is a standard result that $x_{[Tr]}/\sqrt{T} \Rightarrow J_c(r)$ (see Phillips (1987)) and standard FCLT based arguments combined with the CMT lead to

$$T(\hat{\beta}_{[Tr]}-\beta) \;\; \Rightarrow \;\; \frac{\int_0^r J_c(s)dB_u(s)}{\int_0^r J_c(s)^2 ds} \equiv Q_\infty(r;c) \quad r \in [\pi_0, 1] \qquad (25)$$

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

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(\hat{\beta}_t-\beta)x_t\mathbb{I}(q_t \leq \lambda) = \lambda\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(\hat{\beta}_t-\beta)x_t + o_p(1)$$

so that

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(\hat{\beta}_t-\beta)x_t\mathbb{I}(q_t \leq \lambda) \Rightarrow \lambda\int_{\pi_0}^1 Q_\infty(r;c)J_c(r)dr.$$

For the firms term in (24), it follows from Theorem 1 in Caner and Hansen (2001) that

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} u_{t+1}\mathbb{I}(q_t \leq \lambda) \;\; \Rightarrow \sigma_u W(1-\pi_0,\lambda)$$

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

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1}\mathbb{I}(q_t \leq \lambda) \Rightarrow \sqrt{1-\pi_0}\sigma_u W(\lambda) - \lambda\int_{\pi_0}^1 Q_\infty(r;c)J_c(r)dr$$

For the second term in (23), we can write

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1} \;\; = \;\; \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} u_{t+1} - \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(\hat{\beta}_t-\beta)x_t,$$

and using similar arguments to above, we obtain

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1} \quad \Rightarrow \quad \sigma_e\sqrt{1-\pi_0}W(1) - \int_{\pi_0}^{1} Q_\infty(r;c)J_c(r)dr$$

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

$$\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1} \mathbb{I}(q_t \leq \lambda) \to_p \lambda.$$

Therefore, by the CMT,

$$\frac{C_{1T}(\pi_0,\lambda)}{\sqrt{T}} \Rightarrow \sigma_e\sqrt{1-\pi_0}[W(\lambda)-\lambda W(1)]$$

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

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

Recalling that $e_{t+1} = u_{t+1} - (\hat{\beta}_t - \beta)x_t$ it follows that

$$\frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} e_{t+1}^2 \quad = \quad \frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} u_{t+1}^2 + O_p(\frac{1}{T}) \tag{26}$$

and

$$\frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} e_{t+1} \quad = \quad \frac{1}{T}\sum_{t=[T\pi_0]}^{T-1} u_{t+1} + O_p(\frac{1}{\sqrt{T}}) \tag{27}$$

so that

$$\hat{\phi}_1^2 \quad \stackrel{p}{\to} \quad (1-\pi_0)\sigma_u^2 \tag{28}$$

leading to

$$\frac{C_{1T}(\pi_0,\lambda)}{\hat{\phi}_1\sqrt{T}} \Rightarrow W(\lambda) - \lambda W(1)$$

Finally, since *sup*, *ave* and $|.|$ are continuous transformations, the results in (13)-(14) 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

$$\begin{aligned} \frac{C_{2T}(\pi_0,\lambda)}{\sqrt{T}} &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}\left(e_{t+1}^2 - \bar{\tau}^2\right)\mathbb{I}(q_t \leq \lambda) \\ &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1}^2\mathbb{I}(q_t \leq \lambda) - \left(\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(e_{t+1}-\bar{e}_{T-[T\pi_0]})^2\right)\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1}\mathbb{I}(q_t \leq \lambda) \end{aligned} \tag{29}$$

For the first term in (29), we can write

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1}^2\mathbb{I}(q_t\le\lambda) &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}\left((\hat{\beta}_t-\beta)^2x_t^2-2(\hat{\beta}_t-\beta)x_tu_{t+1}+u_{t+1}^2\right)\mathbb{I}(q_t\le\lambda)\\ &= \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(\hat{\beta}_t-\beta)^2x_t^2\mathbb{I}(q_t\le\lambda)-\frac{2}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(\hat{\beta}_t-\beta)x_tu_{t+1}\mathbb{I}(q_t\le\lambda)\\ &+ \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}u_{t+1}^2\mathbb{I}(q_t\le\lambda). \end{aligned}\tag{30}$$

Using $T^2(\hat{\beta}_{[Tr]}-\beta)^2\Rightarrow Q_\infty(r;c)^2$ which follows from (20) and the CMT it is straightforward to observe that the first term in the right hand side of (23) is $O_p(1/\sqrt{T})$ and thus vanishes asymptotically. Similarly, making use of Theorem 2 in Caner and Hansen (2001) which together with Kurtz and Protter (1991) ensures that

$$\frac{1}{T}\sum_{t=[T\pi_0]+1}^{T-1} T(\hat{\beta}_t-\beta)x_tu_{t+1}\mathbb{I}(q_t\le\lambda) \Rightarrow \int_{\pi_0}^1 Q_\infty(r;c)J_c(r)dB_u(r,\lambda)\tag{31}$$

so that the second term in the right hand size of (23) is also $O_p(1/\sqrt{T})$ and vanishes asymptotically. Hence, we can write (30) as

$$\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1} e_{t+1}^2\mathbb{I}(q_t\le\lambda)=\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}u_{t+1}^2\mathbb{I}(q_t\le\lambda)+o_p(1)$$

and therefore

$$\frac{C_{2T}(\pi_0,\lambda)}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T-1}(u_{t+1}^2-\sigma_u^2)-\lambda\frac{1}{\sqrt{T}}\sum_{t=[T\pi_0]}^{T}(u_{t+1}^2-\sigma_u^2)+o_p(1)\tag{32}$$

so that from our Assumption B we have

$$\frac{C_{2T}(\pi_0,\lambda)}{\sqrt{T}} \Rightarrow \phi_2(W(1-\pi_0,\lambda)-\lambda W(1-\pi_0,1))\tag{33}$$

For the denominator, $\hat{\phi}_2^2$, we have

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

By invoking a suitable Law of Large Numbers, we have that

$$\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1}\hat{e}_{t+1}^4\to_p\mathbb{E}(e_{t+1}^4)$$

and

$$\frac{1}{T-[T\pi_0]}\sum_{t=[T\pi_0]}^{T-1}\hat{e}_{t+1}^2 \to_p \mathbb{E}(e_{t+1}^2).$$

As $\frac{T-[T\pi_0]}{T} \to 1-\pi_0$ as $T\to\infty$, the CMT implies that

$$\hat{\phi}_2^2 \to_p (1-\pi_0)\mathbb{E}(e_{t+1}^4) - (1-\pi_0)\mathbb{E}(e_{t+1}^2)^2 = (1-\pi_0)\mathbb{E}\left[e_{t+1}^2 - \mathbb{E}(e_{t+1}^2)^2\right] = (1-\pi_0)\phi_2^2.$$

Therefore, by CMT, we have

$$\frac{C_{2T}(\pi_0,\lambda)}{\sqrt{T}\hat{\phi}_2} \Rightarrow \frac{\sqrt{1-\pi_0}\phi_2(W(\lambda)-\lambda W(1))}{\sqrt{1-\pi_0}\phi_2} = W(\lambda)-\lambda W(1)$$

Finally, since *sup*, *ave* and $|.|$ are continuous transformations, the results for $\tilde{C}_{2T}(\pi_0,\lambda)$ and $\tilde{A}_{2T}(\pi_0,\lambda)$ follow by successively applying the CMT.

PROOF OF PROPOSITION 2: The results in (15)-(16) 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 our statements in (15)-(16) hold.

REFERENCES

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

Brown, R.L., Durbin, J., 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.

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.

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

Deng, A. and P. Perron (2008a). '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.

Deng, A. and P. Perron (2008b). 'The Limit Distribution of the CUSUM of squares test under general mixing conditions', *Econometric Theory*, Vol. 24, pp. 809-822.

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

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

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.

Gulen, H., Yuhang, X. and Zhang, L. (2011). 'Value versus growth: Time-varying expected stock returns', *Financial Management*, Vol. 40, pp. 381-407.

Lakonishok, J., Shleifer, A., 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.

Ploberger, W. and W. Krämer, (1990). 'The local power of the CUSUM and CUSUM of squares test', *Econometric Theory,* Vol. 6, pp. 335-347.

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.