A Simple Approach for Diagnosing Instabilities in Predictive Regressions*

JEAN-YVES PITARAKIS

Department of Economics
University of Southampton
Southampton, SO17 1BJ. UK
(e-mail: j.pitarakis@soton.ac.uk)

February 1, 2017

Abstract

We introduce a method for detecting the presence of structural breaks in the parameters of predictive regressions linking noisy variables such as stock returns to persistent predictors such as valuation ratios. Our approach relies on the least squares based squared residuals of the predictive regression and is straightforward to implement. The distributions of the two test statistics we introduce are shown to be free of nuisance parameters, valid under dependent errors, already tabulated in the literature and robust to the degree of persistence of the chosen predictor. Our proposed method is subsequently applied to the predictability of US stock returns. .

JEL: C12, C22, C53, C58

Keywords: Predictability of Stock Returns, Structural Breaks, CUSUMSQ, Predictive Regressions.

Word Count: Approximately 8000 words

*I am grateful to two anonymous referees and the editor Anindya Banerjee for detailed comments and valuable suggestions that helped improve this paper considerably.

1 Introduction

Models where quantities such as stock returns are regressed on lagged values of predictors such as valuation ratios, interest rates, investor sentiment or other economic and financial variables have been at the core of a vast body of applied and theoretical research in financial economics. The key goal of such specifications is the detection of stock return predictability with important implications for asset pricing theories and the use of conditional asset pricing models which rely on the existence of such predictors. Inferences in the context of these predictive regressions are complicated due to the joint interaction of the highly persistent nature of the commonly used predictors (e.g. dividend yields, price to earnings ratios) with endogeneity problems arising from the correlation of the innovations of the predictors with the predictive regression errors. This has typically led to nonstandard inferences and a growing literature aiming to develop valid and reliable inferences in such settings (see Valkanov (2003), Lewellen (2004), Campbell and Yogo (2006), Jansson and Moreira (2006), Kasparis, Andreou and Phillips (2015) and more recently Kostakis *et al.* (2015) amongst numerous others).

In parallel to this methodological literature on inferences in predictive regressions it has also been recognised that predictability itself may be a time varying phenomenon and that the impact of predictors such as dividend yields, interest rates and others may be evolving over time. In their comprehensive study on the predictability of the equity premium for instance Welch and Goyal (2008) have documented significant instabilities in predictability as also highlighted in Rapach and Wohar (2006), Timmermann (2008), Lettau and Van Nieuwerburgh (2008) and numerous others. The sensitivity analysis conducted in Kostakis *et al.* (2015) also highlighted significant variations in test conclusions depending on whether one considers pre or post 50s data.

Most existing methods used to assess time variation and breaks in the parameters of regression models are typically designed for purely stationary or purely nonstationary settings and are not necessarily suitable for the specificities of predictive regressions. The Brownian Bridge type of asymptotics of the most commonly used SupWald and related tests of Andrews (1993) would no longer be valid for instance when considering nearly integrated predictors. In Rapach and Wohar (2006) the authors used the standard SupWald based test together with bootstrap approximations to infer predictability on US return data. Even with methods specifically designed to address the econometric difficulties characterising predictive regressions instabilities have been mainly highlighted through ad-hoc sub-period analyses. In Kostakis *et al.* (2015) the authors developed a method for testing predictability designed to be immune to the degree of persistence of the predictors and through an ad-hoc sub-period implementation of their methodology documented significant changes in predictability over particular periods.

The goal of this paper is to propose a formal method for uncovering instability in predictive regressions that is specifically designed to handle the presence of nearly integrated predictors in addition to accommodating conditional heteroskedasticity and possible endogeneity in the form of contemporaneous correlations between the innovations driving the predictors and the errors of the predictive regressions. Our method is simple to implement and relies on a simple construct that uses the cumulated squared residuals of a linear predictive regression. More importantly and unlike most of the literature that models persistence via nearly integrated processes the limiting distributions of our proposed test statistics are free of nuisance parameters, already tabulated and do not depend on the unknown non-centrality parameter driving the degree of persistence of the predictors. We view this robustness of our inferences to the noncentrality parameter tuning the persistence of the predictor as a particularly noteworthy and unique feature. In summary our methods offer a straightforward and easy to implement diagnostic tool for exploring potential instabilities prior to conducting further inferences. When applied to the detection of instabilities in the context of the predictability of aggregate US stock market returns over the 1927-2013 period parameter stability is strongly rejected across all considered predictors.

The plan of the paper is as follows. Section 2 introduces our operating model, assumptions and test statistics and obtains its large sample properties. Section 3 focuses on the finite sample size and power porperties of our tests in addition to establishing their consistency. Section 4 applies our methodology to the predictability of aggregate US returns using the recently extended Goyal and Welch (2013) dataset. Section 5 concludes.

2 Cumulative Squared Residuals Based Tests

Our operating model is given by the following predictive regression

$$y_{t+1} = \alpha + \beta x_t + u_{t+1} \tag{1}$$

with the predictor x_t modelled as the nearly integrated process

$$x_t = \left(1 - \frac{c}{T}\right) x_{t-1} + v_t \tag{2}$$

with $c > 0$ and u_t and v_t denoting stationary disturbances. Our key goal is to develop inferences for detecting the presence of a structural break in the conditional mean parameters of (1) that are immune to the unknown non-centrality parameter c and valid under sufficiently weak assumptions on the random disturbance terms and their interactions. The probabilistic properties of our specification are collected in the following set of assumptions.

Assumptions (i) $v_t = \Psi(L)\epsilon_t$ with $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ such that $\Psi(1) \neq 0$, $\Psi_0 = 1$ and absolutely summable coefficients. (ii) $w_t = (u_t, \epsilon_t)'$ is a martingale difference sequence with respect to the natural filtration $\mathcal{F}_t = \sigma(w_t, w_{t-1}, \ldots)$ such that $E||w_t||^4 < \infty$ and $E[w_t w_t'] = \Sigma_w \equiv \{\{\sigma_u^2, \sigma_{u\epsilon}\}, \{\sigma_{u\epsilon}, \sigma_\epsilon^2\}\} > 0$. (iii) The sequence $\eta_t = u_t^2 - \sigma_u^2$ has autocovariances γ_j^η such that

$\sum_{j=1}^{\infty} |\gamma_j^{\eta}| < \infty$ and satisfies the invariance principle $T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \eta_t \Rightarrow \phi\, W(r)$ with $W(r)$ denoting a standard Brownian Motion and $\phi^2 = \gamma_0^{\eta} + 2\sum_{j=1}^{\infty} \gamma_j^{\eta}$.

Assumptions (i)-(ii) are the norm in the *linear* predictive regression literature. They allow v_t to be a very general short memory linear process. They require w_t to be a martingale difference sequence with finite fourth order moments allowing it to be conditionally heteroskedastic with a possibly time varying conditional variance $E[w_t w_t' | \mathcal{F}_{t-1}]$ (see for instance Campbell and Yogo (2006, pp. 57-58), Kostakis *et al.* (2015) and references therein). The unconditional variance-covariance matrix Σ_w may be non-diagonal allowing the shocks to y_t and x_t to be contemporaneously correlated, a fundamental feature in predictive regressions linking stock returns to valuation ratios such as dividend yields and price-to-earnings ratios. Although it is possible to relax assumption (ii) to also allow u_t to be serially correlated via a strong mixing or β mixing setting the specific predictive regression environment and the fact that the most common choice for y_t involves asset returns makes the m.d.s setting with possible conditional heteroskedasticity a natural choice, unlike for instance in single equation cointegrating relationships.

Assumption (iii) is presented as a high level assumption that takes the form of an invariance principle for variances which is needed for establishing the distributional properties of our test statistics formed using the squared residuals obtained from (1). The absolute summability of the autocovariances of η_t ensures in turn that ϕ^2 the limit of $V[\sum_{t=1}^{T} \eta_t / T]$ exists. Examples of processes and primitive assumptions satisfying the above involve a rich set of conditionally heteroskedastic ARCH/GARCH and related specifications under suitable conditions on their parameterisations (e.g. existence of moments restrictions). For a broad range of primitive settings ensuring that (iii) holds see Giraitis, Kokoszka and Leipus (2000, Theorem 5.1), Giraitis, Kokoszka and Leipus (2001, Example 2.2 and Theorem 2.1), Berkes, Hörmann and Horvàth (2008), Lindner (2009) and references therein.

We next introduce our test statistics for testing the null model in (1) against departures from parameter stability as in (5) below. Our proposed inferences will rely on the fluctuations of the squared residuals and more specifically on functionals of the quantity $C_k = \sum_{t=1}^{k} \hat{u}_t^2 - (k/T)\sum_{t=1}^{T} \hat{u}_t^2$. We consider two such functionals. The first one given by the well known CUSUM of squares statistic formulated as

$$C_{SQ} = \max_{1 \leq k \leq T} \frac{1}{\hat{\phi}} \left| \frac{C_k}{\sqrt{T}} \right| \tag{3}$$

and the second one given by

$$A_{SQ} = \frac{1}{T\hat{\phi}^2} \sum_{k=1}^{T} \frac{C_k^2}{T} \tag{4}$$

where $\hat{\phi}^2$ denotes an estimator of the long run variance ϕ^2.

Since the early work of Brown, Durbin and Evans (1975) the use of CUSUM and CUSUM of squares types of test statistics have had a long history in the changepoint literature and both statistics and their multiple variants have been extensively used in applications in virtually all scientific fields. In Xiao and Phillips (2002) the authors used the CUSUM principle to develop a test for detecting the presence of cointegration within a single equation setting. In Deng and Perron (2008a, 2008b) the authors developed a comprehensive analysis of the power properties of CUSUM and related statistics in the context of detecting parameter shifts in regression models under stationarity. Kasparis (2008) and more recently Berenguer-Rico and Nielsen (2016) introduced a CUSUM based approach for detecting misspecification in nonlinear models with non-stationary regressors. The idea behind a test statistic such as (3) or (4) is that any omitted time variation within the predictive regression will contaminate the standard least squares residuals and their squares and hence should be detectable by analysing how $\hat{u}_t$ and $\hat{u}_t^2$ fluctuate.

Test statistics such as C_{SQ} and A_{SQ} were originally developed as exploratory tools for detecting departures from the null of parameter stability with no particular alternative in mind, nevertheless such tests can be shown to be consistent against a broad range of alternatives including single and multiple changepoint scenarios (see Deng and Perron (2008a, 2008b)). The recent changepoint literature is also rich in examples where CUSUM type statistics have been fine-tuned to capture *specific* departures from the null (see Xu (2013a, 2013b), Kejriwal (2012), Juhl and Xiao (2009)) via the formulation of long run variance estimators that use features from the alternative model (e.g. residuals under the alternative of interest or residulas under a nonparametrically specified alternative). In this paper the particular alternative we will consider when focusing on the power properties of C_{SQ} and A_{SQ} is the single changepoint predictive regression given by

$$y_{t+1} = (\alpha_1 + \beta_1 x_t) I(t \leq k) + (\alpha_2 + \beta_2 x_t) I(t > k) + u_{t+1} \tag{5}$$

with $k = [T\pi]$ referring to the location of the break-point in the slopes, intercepts or both and $\pi \in (0, 1)$ the break fraction.

Letting $\hat{\eta}_t = \hat{u}_t^2 - \hat{\sigma}_u^2$ and $\hat{\gamma}_\ell = \sum_{t=\ell+1}^{T} \hat{\eta}_t \hat{\eta}_{t-\ell}/T$ the long run variance estimator of ϕ^2 used in (3)-(4) is given by

$$\hat{\phi}^2 = \hat{\gamma}_0^\eta + 2 \sum_{\ell=1}^{T-1} k(\ell/M_T) \hat{\gamma}_\ell^\eta \tag{6}$$

with $k(.)$ denoting a kernel function and M_T the bandwidth. Under the maintained null model in (1) and assumptions (i)-(iii) consistency of $\hat{\phi}^2$ for ϕ^2 is ensured provided that $M_T \to \infty$ and $M_T/T \to 0$ together with some mild conditions on the kernel function (see Newey and West (1987), Andrews (1991)). Our formulation of the long run variance also parallels the setting in Deng and Perron (2008a, 2008b). Throughout the remainder of this paper we will follow the recent literature and use the Bartlett kernel $k(x) = (1 - |x|) I(|x| \leq 1)$ when estimating ϕ^2. Bandwidths can be selected to be fixed as in Kostakis *et al.* (2015) (e.g. $M_T = T^{1/3}$) or data dependent along the lines of Andrews (1991) with $M_T = 1.1447((4\hat{\lambda}/(1-\hat{\lambda}^2)^2)\ T)^{1/3}$ and $\hat{\lambda}$ denoting the AR(1) parameter estimate from regressing $\hat{\eta}_t$ on $\hat{\eta}_{t-1}$. If one wishes to restrict the presence of dependence in the $\eta_t's$ by forcing the $u_t's$ to be conditionally homoskedastic then the estimator of the long run variance $\phi^2 = E[\eta_t^2]$ is given by $\hat{\gamma}_0^{\eta} \equiv \hat{\phi}^2_{hom}$.

Proposition 1. Under Assumptions (i)-(iii), model (1)-(2) and as $T \to \infty$ we have

$$C_{SQ} \;\Rightarrow\; \sup_{\pi \in [0,1]} |W^0(\pi)| \tag{7}$$

$$A_{SQ} \;\Rightarrow\; \int_0^1 W^0(\pi)^2 d\pi \tag{8}$$

with $W^0(\pi) = W(\pi) - \pi W(1)$ a standard Brownian Bridge.

A *fundamental* feature of our limiting results in Proposition 1 is that the unknown noncentrality parameter c characterising the degree of persistence of x_t does not enter into their expression in (7)-(8) making the practical implementation of our approach particularly straightforward. Note also that these limiting distributions are unaffected by the presence or absence of endogeneity as captured by a possibly nonzero off-diagonal element of Σ_w and are valid regardless of whether a fixed or data dependent bandwidth has been used in the formulation of $\hat{\phi}^2$.

Another convenient aspect of these distributions is the wide availability of their quantiles for inference purposes. From Billingsley (1986) we also have

$$P\left(\sup_{\pi \in [0,1]} |W^0(\pi)| > u\right) \;=\; 2\sum_{j=1}^{\infty}(-1)^{j+1} e^{-2j^2u^2} \tag{9}$$

which can easily be used to construct suitable p-values. Alternatively, the 1%, 5% and 10% critical values of the distribution are given by 1.628, 1.358 and 1.224 respectively. The random variable in the right hand side of (8) is commonly known as the Cramer Von Mises distribution and is also widely tabulated in the literature. In this instance we have

$$P\left(\int_0^1 W^0(\pi)^2 d\pi > u\right) \;=\; \frac{1}{\pi}\sum_{j=1}^{\infty}(-1)^{j+1}\int_{(2j-1)^2\pi^2}^{4j^2\pi^2}\sqrt{\frac{-\sqrt{y}}{\sin(\sqrt{y})}}\,\frac{e^{-xy}}{y}dy \tag{10}$$

with the 10%, 5%, 2.5% and 1% critical values given by 0.347, 0.461, 0.581 and 0.744.

Before proceeding with the size, consistency and finite sample power properties of our two test statistics it is useful to assess the finite sample adequacy of the above asymptotic quantiles via a simulation exercise. Our chosen DGP (DGP1 thereafter) is given by (1) with u_t set to follow the ARCH(1) process $u_t = \zeta_t \sqrt{h_t}$ with $h_t = \omega + \delta u_{t-1}^2$. We also let the shocks to the predictor variable x_t follow the AR(1) process $v_t = \rho_v v_{t-1} + \epsilon_t$. The shocks $(\zeta_t, \epsilon_t)'$ are modelled as a bivariate normal random variable with covariance $\{\{1, \sigma_{\zeta\epsilon}\}, \{\sigma_{\zeta\epsilon}, 1\}\}$. Table 1 below presents the *5% cutoffs* across three non-centrality parameter configurations $c = 1$, $c = 10$ and $c = 40$ using $\{\alpha, \beta\} = \{0.025, 0.01\}$, $\rho_v = 0$, $\sigma_{\zeta\epsilon} = 0$, $\delta \in \{0.0, 0.2, 0.4\}$, $\omega = 1$ and 5000 replications. Alternative parameterisations imposing $\rho_v = 0.5$, $\sigma_{\zeta\epsilon} \in \{-0.5, -0.9\}$ and alternative magnitudes for $\{\alpha, \beta\}$ led to virtually identical outcomes as expected by the asymptotic distribution theory.

Table 1 about here

Our experiments have been conducted using two alternative estimators of the long run variance $\hat{\phi}^2$ in addition to considering the standard homoskedastic version $\hat{\phi}^2_{hom}$ when operating under *known* $\delta = 0$. Using the Bartlett kernel we considered both a fixed bandwidth given by $M = T^{1/3}$ with the associated long run variance referred to as $\hat{\phi}^2_{hac1}$ and a data dependent bandwidth obtained as in Andrews (1991) and with the associated long run variance denoted $\hat{\phi}^2_{hac2}$.

The first important point to infer from Table 1 is the robustness of the finite sample critical values to various magnitudes of the non-centrality parameter c regardless of the sample size. A second important observation under this null setting is the robustness of critical values to the bandwidth choice regardless of whether the DGP includes ($\delta = 0.2$, $\delta = 0.4$) or excludes ARCH effects ($\delta = 0$). Despite these desirable robustness features we also note that under small to moderate sample sizes the finite sample critical values of C_{SQ} tend to lie mildly to the left of their asymptotic counterparts (e.g. 1.287 versus 1.358 under $T = 200$ and $\delta = 0$) suggesting a potentially undersized test for such sample sizes. The same is not true for the A_{SQ} statistic which is characterised by an excellent match between its asymptotic and finite sample critical values even for sample sizes as small as $T = 100$.

Our next objective is to comprehensively explore the finite sample size properties of our tests followed by a formal evaluation of their consistency and finite sample power.

3 Empirical Size, Test Consistency and Finite Sample Power Properties of C_{SQ} and A_{SQ}

Finite Sample Size Properties

We operate within the same DGP as above (DGP1) and evaluate the sensitivity of empirical sizes to the magnitudes of c, $\sigma_{\zeta,\epsilon}$ and ρ_v with a particular emphasis on the ability of our long run variance estimators $\hat{\phi}_{hac1}$ (fixed bandwidth) and $\hat{\phi}_{hac2}$ (data dependent bandwidth) to annihilate and correct for the presence of ARCH effects. Our experiments are based on a 5% nominal size. To highlight potential distortions induced by the use of $\hat{\phi}^2_{hac1}$ and $\hat{\phi}^2_{hac2}$ we initially conduct our size experiments under conditional homoskedasticity ($\delta = 0$) using $\hat{\phi}^2_{hom}$ with results presented in Table 2 while Tables 3-4 present size estimates obtained using the two alternative long run variance estimators across $c = 1$ and $c = 40$ respectively.

Tables 2-4 about here

Under conditional homoskedasticity (Table 2) we note that the empirical sizes of both C_{SQ} and A_{SQ} remain highly robust to alternative parameterisations involving ρ_v, $\sigma_{\zeta\epsilon}$ and c. With sizes in the vicinity of 4%-4.5% under moderately sized samples C_{SQ} is mildly undersized while A_{SQ} displays empirical sizes very close to the nominal 5%. For both the C_{SQ} and A_{SQ} statistics, a given choice of the long run variance estimator and sample size we can note that the resulting empirical sizes continue to show only mild variations across alternative parameterisations including magnitudes of c, strength of endogeneity and persistence of the error process driving the predictor. Across all scenarios C_{SQ} based inferences are typically undersized for small to moderate sample sizes (e.g. in the region of 4%-4.5% under a 5% nominal size for T=600 but closer to 3% for T=200) while A_{SQ} based size estimates appear to match closely their nominal counterparts even under very small sample sizes.

Next, focusing on the specific role played by the long run variance estimators under ARCH errors ($\delta = 0.2$ and $\delta = 0.4$) in Tables 3-4 we note that both $\hat{\phi}_{hac1}$ and $\hat{\phi}_{hac2}$ are able to correct for dependence and lead to test statistics whose size properties parallel the conditionally homoskedastic case. Test statistics constructed using long run variances with a fixed bandwidth (i.e. $\hat{\phi}_{hac1}$) do occasionally display mildly smaller empirical sizes compared with the use of $\hat{\phi}_{hac2}$ but our overall results suggest no particular advantage of using one over the other in terms of their behaviour under the null hypothesis.

In summary, A_{SQ} based inferences offer good size properties in both small and moderately large sample sizes across all scenarios considered while C_{SQ} based inferences may suffer from mild undersizeness for small to moderately sized samples. This also makes it important to assess the finite sample power of C_{SQ} using the finite sample critical values of Table 1 rather than their asymptotic counterparts.

Test Consistency and Finite Sample Power Properties

We next evaluate the ability of the C_{SQ} and A_{SQ} statistics to detect departures from (1) towards models characterised by the presence of a structural break in their intercept and/or slope parameters as in (5). Power properties are particularly important to explore in this context as tests statistics such as the ones considered here have both their numerators and denominators diverge with the sample size. Within a conditionally homoskedastic environment and the use of $\hat{\phi}^2_{hom}$ it is straightforward to establish that the normalised numerators of both C_{SQ} and A_{SQ} diverge faster than $\hat{\phi}_{hom}$ and $\hat{\phi}^2_{hom}$ respectively, ensuring test consistency. When using long run variance estimators as in (6) however the particular choice of bandwidth may impact the speed of divergence of $\hat{\phi}^2$ possibly leading to inconsistent tests or tests with non-monotonic power - in the case of data dependent bandwidths in particular (see Crainiceanu and Vogelsang (2007), Deng and Perron (2008b), Juhl and Xiao (2009) amongst others).

In the following Proposition we initially characterise the large sample behaviour of C_{SQ} and A_{SQ} under the alternative model of interest with the understanding that the bandwidth of the long run variance estimators is fixed and of the type $M_T \equiv M = T^{\kappa}$ as considered in Kostakis *et al.* (2015), Xiao and Phillips (2002) amongst others. It is also understood that in the alternative specification in (5) we may have both the intercept and slope shift at the same time or each shifting individually with the other remaining constant.

Proposition 2. (a) Under model (5), assumptions (i)-(iii), $E[w_t w_t'|\mathcal{F}_{t-1}] = \Sigma_w$ and $\hat{\phi}^2 = \sum_{t=1}^T \hat{\eta}_t^2/T$ we have $C_{SQ} = O_p(\sqrt{T})$ and $A_{SQ} = O_p(\sqrt{T})$. (b) Under model (5), assumptions (i)-(iii) and $\hat{\phi}^2$ obtained from (6) we have $C_{SQ} = O_p(\sqrt{T/M})$ and $A_{SQ} = O_p(\sqrt{T/M})$.

Proposition 2 establishes the large sample behaviour of our tests under the breakpoint alternative and highlights the important influence the bandwidth parameter plays on their power properties when using HAC type long run variance estimators. Under conditional homoskedasticity (assumed known) and using $\hat{\phi}^2_{hom}$ our tests are consistent because their normalised numerator $C_k/\sqrt{T}$ diverges at a faster speed than $\hat{\phi}_{hom}$. HAC type long run variance estimators on the other hand diverge at a faster rate leading to overall test statistics that diverge at a slower rate than the benchmark scenario of conditional homoskedasticity and use of $\hat{\phi}^2_{hom}$. Under the fixed bandwidth $M = T^{1/3}$ for instance we note that both A_{SQ} and C_{SQ} diverge at a rate $T^{1/3}$ compared to $T^{1/2}$ under conditional homoskedasticity. Similarly with $M = T^{1/2}$ both test statistics diverge as $O_p(T^{1/4})$. Nevertheless tests are consistent provided that $T/M \to \infty$. The above rates of divergence of our test statistics under the alternative also coincide with the behaviour of CUSUM type tests in the context of cointegrated regressions as documented in Xiao and Phillips (2002).

The use of a data dependent bandwidth in the above context is more problematic raising the well known phenomenon of non-monotonic power. Intuitively M_T constructed as in Andrews (1991) grows too fast, potentially making the denominator of the test statistics grow faster than their numerator with the trade-off depending on how far away the alternative model is from the null (i.e. on the size of the parameter shifts). The impact of this issue within the context of a predictive regression is explored and discussed in greater depth below.

Our finite sample power experiments proceed as in Deng and Perron (2008b) with a focus on non-local to zero breaks. We consider breaks that occur early on ($\pi_0 = 0.3$), at the middle ($\pi_0 = 0.5$) and towards the end of the sample ($\pi_0 = 0.7$). We set $\{\alpha_{10}, \beta_{10}\} = \{0.025, 0.01\}$ and vary $\{\alpha_{20}, \beta_{20}\}$ across a broad range of magnitudes away from the null. The nominal size is set to 5% and $T = 200$. For our inferences based on C_{SQ} we use the $T = 200$ based finite sample critical values of Table 1 while A_{SQ} based inferences are conducted using their asymptotic critical values. For a broad range of scenarios and a given intercept/slope jump size we have also repeated our experiments across growing sample sizes with $T \in \{100, 200, 600, 1000\}$. Those results are compiled within a supplementary appendix accompanying the paper and further corroborate our empirical findings below.

Table 5 below considers a conditionally homoskedastic scenario with a shift in both the intercept and slope parameters and evaluates performance under $\sigma_{\zeta\epsilon} = 0$, $\rho_v \in \{0.0, 0.5\}$ and $c \in \{1, 10\}$ for both C_{SQ} and A_{SQ} using $\hat{\phi}^2_{hom}$. Across all experiments power is clearly monotonic and both C_{SQ} and A_{SQ} display qualitatively and quantitatively similar patterns. The two factors significantly affecting power when both the intercept and slope shift are ρ_v and c.

Table 5 about here

For both test statistics power is larger under $\rho_v = 0.5$ compared with the case $\rho_v = 0.0$ (i.e. stronger persistence in the shocks driving the predictor x_t leads to greater power). Similarly, power is also substantially larger under $c = 1$ than $c = 10$ i.e. the near integratedness of the predictor leads to significantly more favourable power outcomes than in a purely stationary context (note that under $c = 10$ and $T = 200$ the AR(1) coefficient driving x_t is equal to 0.950 compared with 0.995 under $c = 1$). This latter point is particularly important in our context as predictive regressions typically involve highly persistent predictors such as dividend yields or other valuation ratios with autocorrelation coefficients in the vicinity of 1. The drop in power for purely stationary as opposed to a near unit root context can be substantial (e.g. as large as 50%) and highlights the usefulness and particular suitability of CUSUM type inferences in our predictive regression context. The location of the true break-point also appears to influence power differently depending on whether $c = 1$ or $c = 10$. In the latter case power drops considerably under $\pi_0 = 0.5$ when the true break point is located in the middle of the sample whereas locating the break in the middle of the sample has a much milder impact on power when $c = 1$. Finally, for large enough shifts in α and β power is close to 1 regardless of the magnitudes of c and ρ_v .

Our supplementary appendix provides further sensitivity analyses surrounding the above experiments. We note for instance the robustness of the above outcomes to alternative magnitudes of $\sigma_{\zeta,\epsilon}$ (e.g. $\sigma_{\zeta,\epsilon} = -0.9$). We have also repeated the same analysis by considering solely intercept shifts. In this latter case the drop in power throughout all parameterisations is considerable. Both C_{SQ} and A_{SQ} are unable to detect intercept shifts unless the latter are very large. Power drops even further under $\pi_0 = 0.5$ when the intercept shift is located in the middle of the sample regardless of whether $c = 1$ or $c = 10$. Nevertheless power does increase monotonically as we move further away from the null setting. Further intuition about the $\pi_0 = 0.5$ case under intercept shifts can be gained from equation (15) in the appendix. Although both C_{SQ} and A_{SQ} diverge, the first and last terms in the right hand side of (15) vanish when $\pi_0 = 0.5$ resulting in a substantially smaller statistic. Unlike the scenario where *both* the intercept and slope parameters shift it is here also interesting to point out that different magnitudes of ρ_v or c do not appear to lead to any distinctive power patterns (e.g. lower power under $\rho_v = 0$) when only the intercept is allowed to shift. This further supports the view that the predictive regression environment is particularly suitable for CUSUM type test statistics when interest lies in inferring the time varying predictability induced by a persistent predictor as opposed to say being driven by intercept shifts.

We next focus on the power properties of C_{SQ} and A_{SQ} when evaluated using $\hat{\phi}_{hac1}$ and $\hat{\phi}_{hac2}$ under conditional heteroskedasticity. Tables 6-7 present results for the case of a joint intercept and slope shift under $\rho_v = 0$ and $\sigma_{\zeta,\epsilon} = -0.9$. Our results clearly illustrate the negative impact the choice of bandwidth may have on inferences that rely on HAC type long run variances. Under a joint intercept and slope shift inferences based on a data dependent bandwidth via the use of $\hat{\phi}_{hac2}$ exhibit non-monotonic power across all parameterisations.

Tables 6-7 about here

This is due to the fact that the misspecified $\hat{\eta}_t = \hat{u}_t^2 - \hat{\sigma}^2$ sequence obtained from (1) and which is contaminated with the omitted parameter shifts leads to a data dependent bandwidth $M_T = 1.1447((4\hat{\lambda}/(1-\hat{\lambda}^2)^2)T)^{1/3}$ with $\hat{\lambda} = \sum \hat{\eta}_t \hat{\eta}_{t-1} / \sum \hat{\eta}_{t-1}^2$ that is too large. It is straightforward to show for instance that under a joint intercept and slope shift and with $c = 1$, $\hat{\lambda} \to 1$ as $T \to \infty$ making M_T very large even under a sample size such as $T = 200$ considered here. For a given T, the larger the intercept *and* slope jump the closer $\hat{\lambda}$ is to 1. Within the context of the simulations in Table 6 under $\delta = 0.4$ for instance the configuration $\{\alpha_{10}, \alpha_{20}, \beta_{10}, \beta_{20}\} = \{0.025, 0.73, 0.01, 0.72\}$ led to an average across replication of 8.50 for M_T sharply increasing to 17.58 under $\{\alpha_{10}, \alpha_{20}, \beta_{10}, \beta_{20}\} = \{0.025, 2.15, 0.01, 2.13\}$.

Inferences based on $\hat{\phi}_{hac1}$ on the other hand lead to significantly more favourable power profiles and do not suffer from the same power non-monotonicity problem. The case $\delta = 0$ does however illustrate a clear drop in power compared to the use of $\hat{\phi}_{hom}$ suitable solely under conditional homoskedasticity.

In supplementary simulations presented in our online appendix we have also reconsidered the scenarios of Tables 6-7 under $\rho_v = 0.5$ across $\sigma_{\zeta,\epsilon} = 0$ and $\sigma_{\zeta,\epsilon} = -0.9$ and additional DGPs with intercept shifts only. Our findings continue to highlight the detrimental impact that large magnitudes of ρ_v have on power while the magnitude of $\sigma_{\zeta,\epsilon}$ continues to have little influence. Interestingly the power non-monotonicity problem characterising the use of $\hat{\phi}_{hac2}$ does not arise when the alternative model has an intercept shift only. In such instances inferences based on either $\hat{\phi}_{hac1}$ or $\hat{\phi}_{hac2}$ lead to quantitatively similar but poor empirical power magnitudes especially under $\pi_0 = 0.5$.

4 Time Varying Return Predictability

We apply our methodology to the predictability of US equity returns with valuation ratios as recently explored in Kostakis *et al.* (2015) where the authors developed a novel methodology designed to test the presence of predictability via a Wald type test of the hypothesis $\beta = 0$ in (1). Their key contribution was to propose an IV based Wald statistic whose limiting distribution remains unaffected by the noncentrality parameter c driving the degree of persistence of the predictor. Using monthly data spanning the period 1927-2011 the authors documented a statistically and economically significant predictability of excess returns using the dividend yield, earnings-price, dividend-price and book-to-market value ratios. At the same time and through a sub-period analysis using the same methodology the authors highlighted the sensitivity of their results to particular time periods and more specifically showed that virtually all of the conventionally used predictors lose their predictive ability in a post 50s sample. The robustness of our C_{SQ} and A_{SQ} statistics to the magnitude of the noncentrality parameter c makes them particularly suitable for diagnosing predictive instability in a simple way.

The source of our data is an updated version of the monthly dataset used in Welch and Goyal (2008) (see Goyal and Welch (2013)) as also considered in Kostakis *et al.* (2015) and covers the period 1927:1-2013:12. US market returns are proxied by the CRSP value-weighted returns in excess of the 1-month T-bill rate. The predictors we consider are the dividend yield (DY) expressed as the natural log of dividends over lagged prices, the earnings price ratio (EP) expressed as the natural log of earnings over prices, the dividend price ratio (DP) expressed as the natural log of dividends over prices, the dividend payout ratio (DPO) expressed as the natural log of dividends over earnings and finally the book-to-market ratio (BM) expressed as the natural log of book value over market value. For each of the above predictors we have estimated a simple linear predictive regression as in (1) and calculated the magnitude of C_{SQ} and A_{SQ} as expressed in (3) and (4) across alternative choices of long run variance estimators including a homoskedasticity based variance. Results are presented in Table 8.

Table 8 about here

Our empirical results highlight a strong presence of instability in all five of the predictive regressions when considering the full sample period of 1927-2013. This outcome is robust to the use of alternative variance estimators and holds across both C_{SQ} and A_{SQ}. It is also clear however that this instability vanishes as we exclude pre-mid 40s data or beyond. Looking at the 1940-2013 outcomes for instance we note that across all five predictors the A_{SQ} based inferences are unable to reject the null of parameter stability at any reasonable significance level using either $\hat{\phi}_{hom}$, $\hat{\phi}_{hac1}$ or $\hat{\phi}_{hac2}$. The same conclusions continue to hold as the sample is moved further ahead and using $\hat{\phi}_{hac1}$ based inferences in particular. Interestingly a standard least squares based estimator of the location of the breakpoint also led to $\hat{k} = 1940:05$ using the full sample. Our results are in line with the sensitivity analysis conducted in Kostakis *et al.* (2015) and highlight the usefulness of our procedures for uncovering instability in predictive regressions. They are trivial to implement, they rely on existing tabulated distributions and are robust to the nearly integrated nature of predictors. Furthermore, they have non trivial power provided that the predictors are sufficiently persistent.

5 Conclusions

We have introduced a method for uncovering time variation in the parameters of a predictive regression that relies on two functionals of their squared least squares residuals. Besides the simplicity of their implementation an important feature of our two test statistics is the convenience of their limiting distributions that do not depend on the unknown noncentrality parameter used to parameterise the persistent nature of predictors driving the predictive regression and their readily available quantiles.

Numerous extensions to this research are currently under investigation. A particularly interesting avenue is the generalisation of our specification in (1) to a setting that includes multiple predictors with possibly different degrees of persistence. A second extension involves modelling the coexistence of instabilities in both the conditional mean and error variances along the lines studied in the earlier work of Pitarakis (2004).

TABLES

TABLE 1

5% Finite sample critical values of C_{SQ} and A_{SQ} under conditional homoskedasticity ($\delta = 0$) and conditional heteroskedasticity ($\delta \in \{0.2, 0.4\}$)

	$\delta = 0$			$\delta = 0.2$		$\delta = 0.4$	
	$\hat{\phi}_{hom}$	$\hat{\phi}_{hac1}$	$\hat{\phi}_{hac2}$	$\hat{\phi}_{hac1}$	$\hat{\phi}_{hac2}$	$\hat{\phi}_{hac1}$	$\hat{\phi}_{hac2}$
C_{SQ}				$c = 1$			
T=100	1.274	1.262	1.280	1.273	1.309	1.252	1.281
T=200	1.287	1.274	1.286	1.279	1.322	1.279	1.300
T=600	1.334	1.327	1.332	1.309	1.340	1.300	1.304
T=1000	1.340	1.328	1.342	1.350	1.369	1.309	1.307
A_{SQ}							
T=100	0.468	0.456	0.467	0.477	0.501	0.472	0.491
T=200	0.448	0.450	0.453	0.450	0.474	0.484	0.486
T=600	0.456	0.454	0.459	0.462	0.483	0.482	0.486
T=1000	0.463	0.464	0.459	0.468	0.487	0.481	0.482
C_{SQ}				$c = 10$			
T=100	1.264	1.254	1.265	1.249	1.288	1.245	1.273
T=200	1.306	1.280	1.298	1.288	1.315	1.260	1.277
T=600	1.328	1.326	1.332	1.334	1.355	1.310	1.317
T=1000	1.336	1.320	1.335	1.313	1.331	1.322	1.322
A_{SQ}							
T=100	0.460	0.451	0.461	0.463	0.493	0.453	0.467
T=200	0.447	0.449	0.451	0.467	0.485	0.461	0.472
T=600	0.458	0.460	0.462	0.478	0.493	0.474	0.478
T=1000	0.471	0.461	0.470	0.458	0.470	0.480	0.481
C_{SQ}				$c = 40$			
T=100	1.249	1.238	1.253	1.241	1.281	1.249	1.273
T=200	1.297	1.282	1.292	1.293	1.318	1.269	1.278
T=600	1.318	1.303	1.324	1.318	1.352	1.310	1.316
T=1000	1.337	1.330	1.334	1.335	1.353	1.294	1.296
A_{SQ}							
T=100	0.457	0.447	0.457	0.441	0.467	0.462	0.479
T=200	0.463	0.451	0.464	0.477	0.491	0.475	0.479
T=600	0.450	0.448	0.445	0.470	0.486	0.478	0.485
T=1000	0.453	0.452	0.454	0.470	0.477	0.462	0.463

TABLE 2

Empirical size under conditional homoskedasticity (5% Nominal)

C_{SQ}	$\rho_v = 0$			$\rho_v = 0.5$		
$\sigma_{\zeta\epsilon} = 0.0$	$c = 1$	$c = 10$	$c = 40$	$c = 1$	$c = 10$	$c = 40$
T=100	2.5%	2.6%	3.0%	2.5%	2.8%	3.2%
T=200	3.4%	3.6%	3.4%	3.3%	3.3%	3.5%
T=600	3.6%	4.1%	4.4%	4.1%	4.1%	4.1%
T=1000	4.2%	4.0%	4.4%	4.1%	4.2%	4.0%
$\sigma_{\zeta\epsilon} = -0.5$						
T=100	2.9%	2.8%	3.0%	2.8%	2.9%	2.7%
T=200	3.3%	3.4%	3.1%	3.4%	3.4%	3.4%
T=600	3.9%	4.3%	4.5%	4.0%	4.4%	4.3%
T=1000	4.2%	4.6%	4.4%	4.5%	4.8%	4.5%
$\sigma_{\zeta\epsilon} = -0.9$						
T=100	2.6%	2.7%	2.7%	2.7%	2.7%	2.7%
T=200	3.5%	3.4%	3.5%	3.3%	3.6%	2.9%
T=600	4.4%	3.8%	3.9%	4.1%	4.2%	4.0%
T=1000	4.1%	4.4%	4.2%	4.4%	4.3%	4.4%

A_{SQ}	$\rho_v = 0$			$\rho_v = 0.5$		
$\sigma_{\zeta\epsilon} = 0.0$	$c = 1$	$c = 10$	$c = 40$	$c = 1$	$c = 10$	$c = 40$
T=100	4.5%	4.5%	4.7%	4.1%	5.0%	4.9%
T=200	4.7%	4.9%	4.9%	4.7%	4.8%	4.9%
T=600	4.6%	5.0%	5.0%	5.0%	4.7%	5.3%
T=1000	4.8%	4.6%	5.2%	4.7%	4.8%	5.0%
$\sigma_{\zeta\epsilon} = -0.5$						
T=100	4.7%	4.6%	5.0%	4.3%	4.9%	4.5%
T=200	4.4%	4.8%	4.6%	4.6%	4.8%	4.7%
T=600	4.5%	5.0%	5.1%	4.9%	5.2%	4.9%
T=1000	4.9%	5.3%	5.1%	5.2%	5.2%	5.1%
$\sigma_{\zeta\epsilon} = -0.9$						
T=100	4.6%	4.3%	4.5%	4.5%	4.2%	4.6%
T=200	4.8%	4.6%	4.7%	4.7%	5.1%	4.3%
T=600	5.3%	4.7%	4.9%	5.0%	5.1%	5.0%
T=1000	5.0%	4.9%	5.1%	4.7%	5.3%	4.9%

TABLE 3

Empirical size under conditional heteroskedasticity (5% Nominal) and $c = 1$

$c = 1$	$\rho_v = 0.0$						$\rho_v = 0.5$					
	$\hat{\phi}_{hac1}$			$\hat{\phi}_{hac2}$			$\hat{\phi}_{hac1}$			$\hat{\phi}_{hac2}$		
C_{SQ}	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$
$\sigma_{\zeta\epsilon} = 0.0$												
$T = 100$	1.97%	2.06%	1.77%	2.52%	3.15%	2.58%	1.87%	2.01%	1.99%	2.41%	2.92%	2.65%
$T = 200$	2.86%	2.66%	2.82%	3.23%	3.57%	3.27%	2.88%	3.06%	2.70%	3.15%	3.89%	3.16%
$T = 600$	3.35%	3.78%	3.71%	3.58%	4.40%	4.00%	3.96%	3.96%	3.59%	4.19%	4.66%	3.83%
$T = 1000$	3.80%	4.37%	4.29%	4.27%	4.85%	4.29%	3.94%	4.31%	4.09%	4.03%	4.89%	4.16%
$\sigma_{\zeta\epsilon} = -0.5$												
$T = 100$	2.09%	1.96%	2.12%	2.76%	2.91%	2.88%	1.89%	2.10%	1.87%	2.71%	3.07%	2.69%
$T = 200$	2.96%	2.83%	3.01%	3.18%	3.79%	3.41%	2.61%	3.07%	2.57%	3.34%	3.75%	2.89%
$T = 600$	3.48%	3.97%	3.36%	3.85%	4.52%	3.67%	3.73%	3.70%	3.41%	3.88%	4.38%	3.66%
$T = 1000$	4.03%	4.44%	3.78%	4.19%	5.07%	3.81%	4.25%	4.16%	4.11%	4.44%	4.66%	4.07%
$\sigma_{\zeta\epsilon} = -0.9$												
$T = 100$	1.98%	2.13%	2.11%	2.48%	3.12%	2.93%	1.89%	1.92%	1.84%	2.40%	2.82%	2.65%
$T = 200$	2.91%	2.97%	2.62%	3.22%	3.82%	2.90%	2.79%	2.98%	2.96%	3.16%	3.77%	3.30%
$T = 600$	4.48%	4.04%	3.52%	4.51%	4.88%	3.84%	3.71%	3.98%	3.74%	4.13%	4.74%	3.90%
$T = 1000$	3.96%	3.97%	3.84%	4.20%	4.60%	3.87%	4.12%	4.03%	4.20%	4.26%	4.71%	4.11%
A_{SQ}	$\delta = 0.0$	$\delta = 0.2$	$\delta = 03$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$
$\sigma_{\zeta\epsilon} = 0.0$												
$T = 100$	4.42%	4.73%	4.93%	4.55%	5.68%	5.45%	4.09%	4.71%	5.11%	4.20%	5.51%	5.52%
$T = 200$	4.60%	5.08%	5.74%	4.78%	5.77%	5.92%	4.70%	5.22%	5.54%	4.77%	5.97%	5.67%
$T = 600$	4.63%	5.31%	5.54%	4.60%	5.71%	5.71%	4.72%	5.13%	5.32%	4.91%	5.73%	5.39%
$T = 1000$	4.77%	5.37%	5.63%	4.82%	5.78%	5.59%	4.71%	5.47%	5.65%	4.61%	5.85%	5.56%
$\sigma_{\zeta\epsilon} = -0.5$												
$T = 100$	4.42%	4.56%	5.47%	4.63%	5.50%	5.95%	4.04%	4.83%	5.26%	4.26%	5.57%	5.62%
$T = 200$	4.30%	4.84%	5.56%	4.44%	5.75%	5.85%	4.34%	5.15%	5.12%	4.75%	5.81%	5.32%
$T = 600$	4.55%	5.46%	5.16%	4.52%	6.02%	5.26%	4.77%	5.24%	5.13%	4.91%	5.80%	5.23%
$T = 1000$	4.80%	5.47%	5.67%	4.83%	5.81%	5.65%	5.12%	5.47%	5.73%	5.21%	5.92%	5.63%
$\sigma_{\zeta\epsilon} = -0.9$												
$T = 100$	4.18%	4.91%	5.27%	4.57%	5.58%	5.65%	4.34%	4.33%	4.99%	4.53%	5.24%	5.61%
$T = 200$	4.57%	5.21%	5.12%	4.80%	5.81%	5.33%	4.40%	4.90%	5.91%	4.66%	5.67%	6.26%
$T = 600$	5.27%	5.39%	5.35%	5.31%	6.05%	5.44%	4.80%	5.42%	5.77%	4.88%	5.88%	5.89%
$T = 1000$	4.76%	5.05%	5.30%	4.89%	5.43%	5.30%	4.66%	5.11%	5.64%	4.78%	5.46%	5.56%

TABLE 4

Empirical size under conditional heteroskedasticity (5% Nominal) and

$c = 40$

$c = 40$	$\rho_v = 0.0$						$\rho_v = 0.5$					
	$\hat{\phi}_{hac1}$			$\hat{\phi}_{hac2}$			$\hat{\phi}_{hac1}$			$\hat{\phi}_{hac2}$		
C_{SQ}	$\delta = 0.0$	$\delta = 0.2$	$\delta = 03$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$
$\sigma_{\zeta\epsilon} = 0.0$												
$T = 100$	2.00%	2.30%	1.79%	2.75%	3.25%	2.54%	2.37%	2.09%	1.84%	3.09%	3.07%	2.66%
$T = 200$	3.03%	2.98%	2.85%	3.32%	3.77%	3.41%	2.84%	3.08%	2.75%	3.35%	4.07%	3.19%
$T = 600$	4.07%	3.83%	3.86%	4.33%	4.48%	4.00%	3.77%	3.74%	3.89%	4.14%	4.52%	4.18%
$T = 1000$	4.22%	4.25%	3.99%	4.44%	4.88%	4.07%	3.99%	4.08%	4.15%	3.99%	4.58%	4.17%
$\sigma_{\zeta\epsilon} = -0.5$												
$T = 100$	2.47%	2.17%	1.88%	2.93%	3.19%	2.61%	2.02%	2.38%	2.11%	2.59%	3.55%	2.85%
$T = 200$	2.67%	3.11%	2.86%	3.13%	4.09%	3.34%	2.98%	2.83%	2.81%	3.36%	3.68%	3.27%
$T = 600$	4.16%	3.99%	3.76%	4.43%	4.72%	3.91%	3.66%	4.15%	3.53%	4.29%	4.97%	3.74%
$T = 1000$	4.14%	4.35%	4.08%	4.42%	4.83%	4.13%	4.26%	4.05%	3.67%	4.55%	4.65%	3.74%
$\sigma_{\zeta\epsilon} = -0.9$												
$T = 100$	2.05%	2.16%	2.06%	2.42%	3.34%	2.77%	1.95%	2.13%	1.82%	2.51%	3.11%	2.58%
$T = 200$	2.79%	3.29%	2.92%	3.31%	4.04%	3.40%	2.63%	2.98%	2.66%	2.87%	4.10%	2.89%
$T = 600$	3.85%	3.69%	3.31%	3.87%	4.33%	3.49%	3.80%	3.92%	3.62%	4.11%	4.67%	3.91%
$T = 1000$	3.97%	4.08%	3.83%	4.31%	4.70%	3.85%	3.95%	4.15%	4.03%	4.35%	4.71%	4.13%
A_{SQ}	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$
$\sigma_{\zeta\epsilon} = 0.0$												
$T = 100$	4.28%	4.68%	5.10%	4.77%	5.65%	5.73%	4.25%	4.64%	4.94%	4.76%	5.48%	5.67%
$T = 200$	4.73%	5.37%	5.63%	4.88%	6.14%	5.87%	4.53%	5.26%	5.48%	4.82%	5.97%	5.87%
$T = 600$	4.73%	5.12%	5.80%	4.85%	5.67%	5.94%	5.08%	5.06%	5.82%	5.27%	5.70%	5.94%
$T = 1000$	5.08%	5.37%	5.77%	5.21%	5.79%	5.70%	5.02%	5.29%	5.90%	4.96%	5.60%	5.88%
$\sigma_{\zeta\epsilon} = -0.5$												
$T = 100$	4.70%	4.77%	4.72%	4.96%	5.44%	5.41%	4.36%	5.39%	5.36%	4.46%	6.18%	5.97%
$T = 200$	4.35%	5.03%	5.31%	4.41%	5.75%	5.53%	4.72%	5.29%	5.61%	4.77%	6.09%	5.87%
$T = 600$	5.12%	4.99%	6.03%	5.16%	5.55%	6.13%	4.73%	5.20%	5.64%	4.87%	5.73%	5.75%
$T = 1000$	5.04%	5.29%	5.71%	5.16%	5.63%	5.73%	4.89%	5.16%	5.38%	5.09%	5.49%	5.41%
$\sigma_{\zeta\epsilon} = -0.9$												
$T = 100$	4.27%	4.88%	5.06%	4.53%	5.89%	5.74%	4.18%	4.54%	5.02%	4.48%	5.23%	5.70%
$T = 200$	4.42%	5.44%	5.89%	4.58%	6.11%	6.07%	4.18%	5.28%	5.31%	4.38%	6.04%	5.72%
$T = 600$	4.72%	5.09%	5.59%	4.87%	5.65%	5.66%	4.90%	4.97%	5.18%	5.02%	5.53%	5.30%
$T = 1000$	4.96%	5.23%	5.46%	5.09%	5.63%	5.42%	4.74%	5.46%	5.80%	4.93%	5.86%	5.76%

TABLE 5

Empirical power under conditional homoskedasticity - Intercept and Slope

Shift: $\{\alpha_{10}, \beta_{10}\} = \{0.025, 0.01\}$, $\sigma_{\zeta\epsilon} = 0.0$

α_{20}	0.17	0.31	0.45	0.59	0.73	0.87	1.01	1.16	1.30	1.44	1.58	1.72	1.86	2.00	2.15
β_{20}	0.15	0.29	0.43	0.58	0.72	0.86	1.00	1.14	1.28	1.42	1.57	1.71	1.85	1.99	2.13
C_{SQ}							$c = 1$ $(\rho_v = 0)$								
$\pi_0 = 0.3$	13.8	45.9	73.2	88.0	94.7	97.6	98.6	99.0	99.4	99.5	99.5	99.6	99.6	99.7	99.8
$\pi_0 = 0.5$	9.3	36.7	62.8	78.0	86.0	91.1	93.6	95.5	96.5	97.0	97.4	97.8	98.1	98.1	98.2
$\pi_0 = 0.7$	20.0	51.9	73.9	85.5	91.2	94.3	96.0	96.8	97.6	98.1	98.2	98.3	98.4	98.8	98.8
A_{SQ}															
$\pi_0 = 0.3$	13.2	44.5	71.2	86.9	93.8	96.8	98.2	98.7	99.1	99.2	99.2	99.4	99.4	99.6	99.6
$\pi_0 = 0.5$	8.0	32.8	59.1	73.7	82.3	87.4	90.9	93.1	94.2	95.3	95.6	95.9	96.6	96.5	96.6
$\pi_0 = 0.7$	19.3	50.2	71.2	82.9	89.1	91.9	94.0	94.9	95.7	96.5	96.9	97.0	97.1	97.4	97.6
C_{SQ}							$c = 1$ $(\rho_v = 0.5)$								
$\pi_0 = 0.3$	43.7	84.8	96.5	98.9	99.6	99.7	99.7	99.9	99.8	99.8	99.9	99.9	99.9	99.8	99.9
$\pi_0 = 0.5$	35.6	77.3	90.8	95.6	97.3	98.1	98.6	99.0	99.1	99.1	99.4	99.3	99.5	99.6	99.5
$\pi_0 = 0.7$	51.3	83.9	94.0	96.9	98.1	98.5	98.9	99.1	99.3	99.3	99.5	99.4	99.5	99.4	99.5
A_{SQ}															
$\pi_0 = 0.3$	42.0	83.6	95.8	98.4	99.4	99.6	99.5	99.7	99.7	99.7	99.8	99.8	99.8	99.7	99.8
$\pi_0 = 0.5$	31.4	72.8	87.7	93.3	95.8	96.8	97.6	97.9	98.0	98.2	98.4	98.5	98.6	98.8	98.6
$\pi_0 = 0.7$	49.1	81.1	91.8	95.1	96.6	97.2	97.8	98.1	98.2	98.4	98.7	98.5	98.6	98.4	98.6
C_{SQ}							$c = 10$ $(\rho_v = 0)$								
$\pi_0 = 0.3$	5.5	13.2	35.5	62.3	81.3	90.5	94.9	96.7	97.6	98.2	98.5	98.7	98.9	99.1	99.0
$\pi_0 = 0.5$	4.7	6.6	15.1	29.0	45.0	57.6	67.1	73.9	78.3	81.8	84.7	86.4	88.1	89.4	90.3
$\pi_0 = 0.7$	5.6	13.8	35.7	60.6	77.5	88.1	92.1	94.4	95.7	96.5	96.9	97.4	97.6	98.3	98.3
A_{SQ}															
$\pi_0 = 0.3$	5.3	13.0	33.9	59.3	78.4	88.3	93.2	95.3	96.5	97.2	97.5	97.9	98.3	98.3	98.5
$\pi_0 = 0.5$	4.7	6.0	12.5	24.2	38.6	50.1	58.9	66.6	70.8	74.8	77.8	79.7	81.7	82.6	84.3
$\pi_0 = 0.7$	5.7	13.8	34.5	58.0	75.2	86.0	90.0	92.5	94.0	95.2	95.6	96.1	96.3	97.2	97.0
C_{SQ}							$c = 10$ $(\rho_v = 0.5)$								
$\pi_0 = 0.3$	10.4	52.8	85.3	94.8	97.5	98.3	98.9	99.3	99.3	99.4	99.3	99.4	99.5	99.5	99.5
$\pi_0 = 0.5$	5.7	25.8	55.0	73.5	82.8	87.3	91.0	93.0	93.8	94.7	95.3	95.3	95.9	95.7	95.9
$\pi_0 = 0.7$	10.7	52.1	82.8	92.8	95.9	97.3	97.9	98.3	98.4	98.8	98.6	99.0	99.0	98.9	99.0
A_{SQ}															
$\pi_0 = 0.3$	10.1	50.1	82.7	93.0	96.2	97.5	98.2	98.6	98.7	98.9	98.8	98.9	99.0	99.0	99.0
$\pi_0 = 0.5$	5.3	21.6	47.8	65.4	75.3	80.7	84.9	87.2	89.1	90.2	91.2	91.1	92.1	91.5	92.3
$\pi_0 = 0.7$	10.8	49.5	79.8	90.6	94.0	95.8	96.5	97.2	97.4	97.7	97.5	97.9	98.1	98.0	98.0

TABLE 6

Empirical power of C_{SQ} with long run variance estimators - Intercept and Slope shift: $\{\alpha_{10}, \beta_{10}\} = \{0.025, 0.01\}$, $\sigma_{\zeta\epsilon} = -0.9$, $\rho_v = 0.0$, $c = 1$

α_{20}	0.17	0.31	0.45	0.59	0.73	0.87	1.01	1.16	1.30	1.44	1.58	1.72	1.86	2.00	2.15
β_{20}	0.15	0.29	0.43	0.58	0.72	0.86	1.00	1.14	1.28	1.42	1.57	1.71	1.85	1.99	2.13
$\pi_0 = 0.3$								$\delta = 0$							
$\hat{\phi}_{hom}$	13.9	42.0	66.7	82.7	91.8	95.5	97.4	98.7	99.2	99.2	99.6	99.5	99.5	99.6	99.6
$\hat{\phi}_{hac1}$	9.4	27.4	45.1	59.7	69.5	76.0	80.8	83.9	86.5	87.7	88.1	88.2	89.4	90.0	89.8
$\hat{\phi}_{hac2}$	12.1	31.5	42.8	47.3	45.5	41.7	38.6	34.4	31.8	29.7	26.6	25.4	24.7	23.4	22.4
								$\delta = 0.2$							
$\hat{\phi}_{hac1}$	8.2	22.1	37.3	51.9	63.2	70.0	76.5	80.7	83.5	84.9	86.9	87.7	88.7	89.0	88.8
$\hat{\phi}_{hac2}$	10.3	24.5	35.1	41.7	42.1	39.9	38.5	36.6	32.3	31.2	28.6	26.8	26.0	23.3	22.7
								$\delta = 0.4$							
$\hat{\phi}_{hac1}$	6.4	15.8	29.1	41.8	52.7	62.5	68.5	73.8	78.0	81.7	83.9	84.8	86.0	87.4	88.0
$\hat{\phi}_{hac2}$	7.3	16.9	27.0	33.3	36.6	38.7	36.8	36.0	35.0	32.9	31.2	29.8	27.9	26.6	26.0
$\pi_0 = 0.5$								$\delta = 0$							
$\hat{\phi}_{hom}$	10.0	37.2	61.2	75.9	86.2	91.0	93.7	95.7	96.3	97.0	97.3	97.7	98.1	98.3	98.6
$\hat{\phi}_{hac1}$	5.0	11.2	19.3	26.0	30.5	33.3	36.4	37.4	38.4	39.2	40.3	41.1	40.8	41.6	41.6
$\hat{\phi}_{hac2}$	7.6	17.7	22.3	21.6	18.9	15.4	13.8	12.0	10.7	9.9	9.3	8.1	7.5	7.8	7.2
								$\delta = 0.2$							
$\hat{\phi}_{hac1}$	4.2	9.2	16.0	22.5	28.0	31.4	34.0	35.4	36.5	38.3	40.0	40.1	39.5	40.3	41.0
$\hat{\phi}_{hac2}$	6.4	13.8	18.6	19.2	18.8	16.6	14.9	13.3	11.7	10.1	10.0	9.5	8.8	7.9	7.9
								$\delta = 0.4$							
$\hat{\phi}_{hac1}$	4.6	7.2	13.0	18.0	23.8	27.7	30.5	33.5	35.1	36.7	37.7	38.5	39.6	39.0	40.3
$\hat{\phi}_{hac2}$	5.7	9.9	14.5	16.7	17.3	16.4	15.4	14.1	13.3	12.0	11.0	10.3	9.6	8.8	8.5
$\pi_0 = 0.7$								$\delta = 0$							
$\hat{\phi}_{hom}$	25.1	59.7	80.4	90.1	94.3	96.7	97.6	98.1	98.6	98.8	98.9	99.1	98.9	99.1	99.2
$\hat{\phi}_{hac1}$	21.2	49.3	65.6	73.7	77.4	79.0	80.3	80.4	80.3	80.5	80.7	80.3	80.4	81.0	80.5
$\hat{\phi}_{hac2}$	23.8	53.6	66.9	69.1	65.0	59.8	53.9	48.2	43.0	39.2	34.5	32.6	30.6	28.4	26.6
								$\delta = 0.2$							
$\hat{\phi}_{hac1}$	15.8	41.3	58.7	68.2	74.2	76.9	78.4	79.4	80.2	80.2	80.6	80.5	81.1	80.6	81.0
$\hat{\phi}_{hac2}$	18.3	45.3	59.9	65.8	64.9	61.9	56.5	51.8	46.7	42.5	39.1	36.0	33.4	32.0	29.0
								$\delta = 0.4$							
$\hat{\phi}_{hac1}$	11.6	31.4	47.7	59.5	67.7	72.1	75.6	77.3	78.3	79.3	79.5	79.6	80.4	80.2	81.4
$\hat{\phi}_{hac2}$	12.9	33.9	49.4	57.9	61.3	60.6	59.1	54.6	51.1	47.7	44.2	41.1	38.0	35.6	33.2

TABLE 7

Empirical power of A_{SQ} with long run variance estimators - Intercept and Slope shift: $\{\alpha_{10}, \beta_{10}\} = \{0.025, 0.01\}$, $\sigma_{\zeta\epsilon} = -0.9$, $\rho_v = 0.0$, $c = 1$

α_{20}	0.17	0.31	0.45	0.59	0.73	0.87	1.01	1.16	1.30	1.44	1.58	1.72	1.86	2.00	2.15
β_{20}	0.15	0.29	0.43	0.58	0.72	0.86	1.00	1.14	1.28	1.42	1.57	1.71	1.85	1.99	2.13
$\pi_0 = 0.3$								$\delta = 0$							
$\hat{\phi}_{hom}$	13.1	40.4	64.9	81.4	90.8	94.8	96.9	98.3	98.7	98.9	99.3	99.2	99.2	99.4	99.3
$\hat{\phi}_{hac1}$	9.6	27.5	45.7	61.3	71.3	77.7	81.9	84.9	87.2	88.2	88.4	88.9	89.9	90.2	90.1
$\hat{\phi}_{hac2}$	11.6	30.5	43.6	50.4	51.3	49.8	47.5	45.0	43.1	41.5	38.8	37.2	37.4	35.7	35.3
								$\delta = 0.2$							
$\hat{\phi}_{hac1}$	8.4	22.1	37.8	52.9	64.1	71.3	77.7	82.2	84.0	85.6	87.3	88.5	88.9	89.3	89.3
$\hat{\phi}_{hac2}$	9.9	23.9	36.1	43.8	46.6	46.5	47.1	45.6	42.6	42.3	39.8	39.5	37.6	36.0	35.2
								$\delta = 0.4$							
$\hat{\phi}_{hac1}$	7.5	16.2	29.3	42.6	53.8	63.2	69.1	75.0	79.6	82.5	84.6	85.6	86.4	87.9	88.3
$\hat{\phi}_{hac2}$	8.0	17.2	27.7	35.6	40.2	43.7	43.5	44.6	44.4	43.0	42.3	41.4	39.1	38.2	37.9
$\pi_0 = 0.5$								$\delta = 0$							
$\hat{\phi}_{hom}$	7.7	31.8	55.3	71.1	82.1	87.7	90.7	93.4	94.1	94.8	95.3	96.3	96.6	96.9	97.0
$\hat{\phi}_{hac1}$	3.9	9.3	17.3	24.4	29.8	33.1	35.7	37.0	38.3	39.0	40.0	41.4	41.2	41.9	42.0
$\hat{\phi}_{hac2}$	5.4	14.0	18.9	19.1	18.1	15.7	14.6	13.7	12.5	12.2	12.2	11.4	11.0	11.0	10.4
								$\delta = 0.2$							
$\hat{\phi}_{hac1}$	4.2	7.6	14.2	20.9	26.1	30.3	33.7	35.1	37.3	38.1	39.9	40.2	40.2	40.7	41.4
$\hat{\phi}_{hac2}$	5.7	10.6	15.5	17.1	16.9	15.8	14.9	14.3	13.6	12.5	13.0	12.5	11.4	11.0	11.0
								$\delta = 0.4$							
$\hat{\phi}_{hac1}$	5.8	6.6	11.6	16.7	22.1	27.0	29.4	33.2	34.9	36.8	37.3	39.0	40.0	39.5	41.3
$\hat{\phi}_{hac2}$	6.4	8.5	13.0	14.6	15.6	15.5	15.1	14.9	13.8	13.4	12.3	12.5	11.7	11.6	11.6
$\pi_0 = 0.7$								$\delta = 0$							
$\hat{\phi}_{hom}$	23.8	57.6	77.7	87.8	92.2	94.7	95.8	96.4	97.1	97.5	97.7	98.0	97.8	98.0	98.2
$\hat{\phi}_{hac1}$	20.8	48.0	64.2	72.1	76.2	77.6	78.9	79.2	79.3	79.6	79.6	79.4	79.2	79.7	79.4
$\hat{\phi}_{hac2}$	23.0	51.7	65.4	68.8	67.5	64.2	61.2	57.5	53.5	51.5	48.8	46.5	45.1	42.9	41.5
								$\delta = 0.2$							
$\hat{\phi}_{hac1}$	16.3	40.7	58.0	66.6	73.0	76.2	77.1	78.2	79.1	79.2	79.5	79.5	80.2	79.6	79.9
$\hat{\phi}_{hac2}$	18.4	43.4	58.6	64.8	65.9	65.3	61.6	60.0	56.7	53.5	52.0	49.2	48.1	46.3	44.6
								$\delta = 0.4$							
$\hat{\phi}_{hac1}$	13.0	32.2	47.9	59.2	66.8	71.2	74.8	76.1	77.0	78.4	78.7	78.6	79.2	79.1	80.1
$\hat{\phi}_{hac2}$	14.0	34.1	48.9	57.8	61.8	62.6	62.7	60.7	58.7	56.6	54.7	52.1	50.4	49.4	46.9

TABLE 8

Return Predictability

	C_{SQ}			A_{SQ}		
	$\hat{\phi}_{hom}$	$\hat{\phi}_{hac1}$	$\hat{\phi}_{hac2}$	$\hat{\phi}_{hom}$	$\hat{\phi}_{hac1}$	$\hat{\phi}_{hac2}$
	1927-2013			1927-2013		
DY	3.538***	1.994***	2.192***	3.807***	1.210***	1.461***
EP	3.458***	1.976***	2.154***	3.636***	1.187***	1.411***
DP	3.521***	1.988***	2.179***	3.780***	1.205***	1.447***
DPO	3.480***	1.982***	2.165***	3.726***	1.209***	1.442***
BM	3.538***	1.997***	2.202***	3.878***	1.205***	1.465***
	1940-2013			1940-2013		
DY	1.069	0.869	0.940	0.201	0.133	0.155
EP	1.069	0.862	0.936	0.200	0.130	0.153
DP	1.067	0.866	0.937	0.201	0.132	0.155
DPO	1.005	1.816	0.879	0.175	0.115	0.133
BM	1.066	0.861	0.932	0.198	0.129	0.151
	1950-2013			1950-2013		
DY	1.528**	1.136	1.244*	0.515**	0.285	0.341
EP	1.503**	1.112	1.223	0.494**	0.270	0.327
DP	1.521**	1.129	1.234*	0.512**	0.282	0.337
DPO	1.431**	1.072	1.171	0.441 *	0.248	0.295
BM	1.479**	1.098	1.206	0.477**	0.263	0.317
	1960-2013			1960-2013		
DY	1.037	0.786	0.854	0.187	0.107	0.127
EP	1.032	0.780	0.851	0.187	0.107	0.128
DP	1.034	0.782	0.850	0.187	0.107	0.126
DPO	1.011	0.771	0.836	0.177	0.103	0.121
BM	1.027	0.778	0.849	0.187	0.107	0.128

APPENDIX

In what follows we make extensive use of existing results on the large sample properties of sample moments of highly persistent processes as in (2) without explicitly appealing to first principles. From Phillips (1987) for instance $\sum_{t=1}^{[T\pi]} x_{t-1}^2/T^2 \Rightarrow \int_0^\pi J_c^2(r)dr$, $\sum_{t=1}^{[T\pi]} x_{t-1}/T\sqrt{T} \Rightarrow \int_0^\pi J_c(r)dr$ (see also Sandberg (2009)) and $\sum_{t=1}^{[T\pi]} x_{t-1}u_t/T = O_p(1)$. Here $J_c(r)$ is a diffusion process such that $dJ_c(r) = cJ_c(r)dr + dB_v(r)$ with $B_v(r)$ denoting the Brownian Motion associated with v_t. More generally, using the continuous mapping theorem and our assumptions on the finiteness of moments in (i)-(iii) we also have $\sum_{t=1}^{[T\pi]} x_{t-1}^m/T^{1+\frac{m}{2}} = O_p(1)$, $\sum_{t=1}^{[T\pi]} x_{t-1}^m u_t/T^{(1+m)/2} = O_p(1)$ for $m = 1, 2, 3, 4$, $\sum_{t=1}^{[T\pi]} x_{t-1}^2 u_t^2/T^2 = O_p(1)$ and $\sum_{t=1}^{[T\pi]} x_{t-1}u_t^3/T^{3/2} = O_p(1)$ also leading to $T(\hat{\beta} - \beta) = O_p(1)$ and $\sqrt{T}(\hat{\alpha} - \alpha) = O_p(1)$ under the null model in (1) and $\hat{\alpha}$ and $\hat{\beta}$ denoting the least squares estimators of α and β (see Valkanov (2003) for explicit expressions for these limiting distributions). The above large sample properties also directly imply that under model (1)-(2) we have $\sum_{t=1}^T \hat{u}_t^4/T = \sum_{t=1}^T u_t^4 + o_p(1)$ and $\sum_{t=1}^T \hat{u}_t^2/T = \sum_{t=1}^T u_t^2 + o_p(1)$ ensuring that $\hat{\phi}_{hom}^2 \xrightarrow{p} \phi^2 \equiv [\eta_t^2]$. Provided that the bandwidth M_T is such that $M_T \to \infty$ and $M_T/T \to 0$ it follows from Andrews (1991) and the above that we also have $\hat{\phi}_{hac1}^2 \xrightarrow{p} \phi^2$ and $\hat{\phi}_{hac2}^2 \xrightarrow{p} \phi^2$.

Proof of Proposition 1. Letting $\tilde{x}_t = (1, x_t)'$ and $\theta = (\alpha, \beta)'$ we can write $\sum_{t=1}^k \hat{u}_t^2 = \sum_{t=1}^k u_t^2 + (\hat{\theta} - \theta)' \sum_{t=1}^k \tilde{x}_{t-1}\tilde{x}_{t-1}'(\hat{\theta} - \theta) - 2\sum_{t=1}^k u_t \tilde{x}_{t-1}'(\hat{\theta} - \theta)$. It is also convenient to introduce the normalising matrix $D_T = diag(\sqrt{T}, T)$ so that

$$\frac{\sum_{t=1}^k \hat{u}_t^2}{\sqrt{T}} = \frac{\sum_{t=1}^k u_t^2}{\sqrt{T}} + \frac{1}{\sqrt{T}}[D_T(\hat{\theta} - \theta)]' \left(D_T^{-1} \sum_{t=1}^k \tilde{x}_{t-1}\tilde{x}_{t-1}' D_T^{-1} \right) D_T(\hat{\theta} - \theta) - 2\frac{1}{\sqrt{T}} \sum_{t=1}^k u_t \tilde{x}_{t-1}' D_T^{-1}(D_T(\hat{\theta} - \theta)) \quad (11)$$

also noting that

$$\frac{\sum_{t=1}^k \hat{u}_t^2}{\sqrt{T}} - \frac{k}{T}\frac{\sum_{t=1}^T \hat{u}_t^2}{\sqrt{T}} = \left(\frac{\sum_{t=1}^k u_t^2}{\sqrt{T}} - \frac{k}{T}\frac{\sum_{t=1}^T u_t^2}{\sqrt{T}} \right) + \left(\frac{\sum_{t=1}^k \hat{u}_t^2}{\sqrt{T}} - \frac{\sum_{t=1}^T u_t^2}{\sqrt{T}} \right) - \frac{k}{T}\left(\frac{\sum_{t=1}^T \hat{u}_t^2}{\sqrt{T}} - \frac{\sum_{t=1}^T u_t^2}{\sqrt{T}} \right) \quad (12)$$

Given our chosen process in (2) and the results stated above it is clear that $D_T(\hat{\theta} - \theta) = O_p(1)$ leading to $\sum_{t=1}^T \hat{u}_t^2/\sqrt{T} = \sum_{t=1}^T u_t^2/\sqrt{T} + o_p(1)$. Next, the boundedness of $\max_k D_T^{-1} \sum_{t=1}^k \tilde{x}_{t-1}\tilde{x}_{t-1}' D_T^{-1}$ together with $D_T(\hat{\theta} - \theta) = O_p(1)$ also ensures that $T^{-1/2}(D_T(\hat{\theta} - \theta))'(\max_k D_T^{-1} \sum \tilde{x}_{t-1}\tilde{x}_{t-1}' D_T^{-1})(D_T(\hat{\theta} - \theta)) \xrightarrow{p} 0$. Finally combining with $T^{-1/2} \max_k |\sum_{t=1}^k u_t x_{t-1}'(\hat{\theta} - \theta)D_T| \xrightarrow{p} 0$ we have $\max_k |\sum_{t=1}^k \hat{u}_t^2/\sqrt{T} - \sum_{t=1}^k u_t^2/\sqrt{T}| \xrightarrow{p} 0$ leading to

$$\max_k \left| \frac{\sum_{t=1}^k \hat{u}_t^2}{\sqrt{T}} - \frac{k}{T}\frac{\sum_{t=1}^T \hat{u}_t^2}{\sqrt{T}} \right| = \left| \frac{\sum_{t=1}^k u_t^2}{\sqrt{T}} - \frac{k}{T}\frac{\sum_{t=1}^T u_t^2}{\sqrt{T}} \right| + o_p(1) \quad (13)$$

Assumptions (iii), the continuous mapping theorem combined with the consistency of the long run variance estimators leads to the desired result in (7) and similarly for (8). □

Proof of Proposition 2. We initially treat scenario (i) under conditional homoskedasticity allowing both the intercept and slope parameters to shift jointly. The alternative model is given by (5) and we let $k_0 = [T\pi_0]$ denote the true break point location and π_0 the associated break fraction. Defining $\mu_0 \equiv (\alpha_{20} - \alpha_{10})$ and $\lambda_0 \equiv (\beta_{20} - \beta_{10})$ it is also convenient to rewrite the specification in (5) as $y_t = \alpha_{10} + \mu_0 I(t > k_0) + \beta_{10}x_{t-1} + \lambda_0 x_{t-1}I(t > k_0) + u_t$. Letting $C_{k0} = \sum_{t=1}^{k0} \hat{u}_t^2 - (k_0/T)\sum_{t=1}^T \hat{u}_t^2$ and using $\hat{u}_t = \alpha_{10} + \beta_{10}x_{t-1} + \mu_0 I(t > k_0) + \lambda_0 x_{t-1}I(t > k_0) - \hat{\alpha} - \hat{\beta}x_{t-1}$ lengthy but standard algebra combined with successive uses of the continuous mapping theorem leads to $|C_{k0}/T^2| \Rightarrow$

$|Z_{1\infty}(J_c, \pi_0, \lambda_0)|$ where

$$\begin{aligned} Z_{1\infty}(J_c, \pi_0, \lambda_0) &= \lambda_0^2 \left(Q_{1\infty}^2 (1-\pi_0) \int_0^{\pi_0} J_c(r)^2 - (1-Q_{1\infty})^2 \pi_0 \int_{\pi_0}^1 J_c(r)^2 \right) + \\ & 2\lambda_0^2 \left(Q_{1\infty}(1-\pi_0) \int_0^{\pi_0} J_c(r) - (1-Q_{1\infty})\pi_0 \int_{\pi_0}^1 J_c(r) \right) \times \\ & \left(\int_{\pi_0}^1 J_c(r) - Q_{1\infty} \int_0^1 J_c(r) \right) \end{aligned} \quad (14)$$

with $Q_{1\infty} \equiv [\int_{\pi_0}^1 J_c(r)^2 - \int_0^1 J_c(r) \int_{\pi_0}^1 J_c(r)] / \int_0^1 J_c^*(r)^2$ and $J_c^*(r) = J_c(r) - \int_0^1 J_c(r)$ establishing that $|C_{k0}/\sqrt{T}| = O_p(T\sqrt{T})$. Proceeding similarly for $\hat{\phi}_{hom}^2 = \sum_{t=1}^T \hat{\eta}_t^2/T$ we also obtain $\hat{\phi}_{hom}^2 = O_p(T^2)$ so that $|C_{k0}/\sqrt{T}|/\hat{\phi}_{hom} = O_p(\sqrt{T})$. As $\max_k |C_k/\sqrt{T}|/\hat{\phi}_{hom} \geq |C_{k0}/\sqrt{T}|/\hat{\phi}_{hom}$ the required result follows. The case of an intercept shift only follows identical lines. In this instance we have $|C_{k0}/T| \Rightarrow |Z_{2\infty}(J_c, \pi_0, \mu_0)|$ with

$$\begin{aligned} Z_{2\infty}(J_c, \pi_0, \mu_0) &= \mu_0^2 \pi_0 (1-\pi_0)(1-2\pi_0) + \mu_0^2 Q_{2\infty}^2 \left(\int_0^{\pi_0} J_c^*(r)^2 - \pi_0 \int_0^1 J_c^*(r)^2 \right) \\ &+ 2\mu_0^2 (1-2\pi_0) Q_{2\infty} \int_0^{\pi_0} J_c^*(r) \end{aligned} \quad (15)$$

$Q_{2\infty} \equiv \int_{\pi_0}^1 J_c^* / \int_0^1 J_c^*(r)^2$ and $\hat{\phi}_{hom}^2 = O_p(1)$ so that $|C_{k0}/\sqrt{T}|/\hat{\phi}_{hom} = O_p(\sqrt{T})$ as stated. (ii) Under conditional heteroskedasticity we continue to have $|C_{k0}/\sqrt{T}| = O_p(T\sqrt{T})$ when both the intercept and slope parameters shift and $|C_{k0}/\sqrt{T}| = O_p(\sqrt{T})$ when only the intercept is allowed to shift. Operating with a fixed (non data dependent) bandwidth such that $M \to \infty$ the result follows by noting that $1 + 2\sum_{\ell=1}^M (1 - \ell/(M+1)) = M$ and $\sum \hat{\eta}_t^2/T^2 \approx \sum \hat{\eta}_t \hat{\eta}_{t-\ell}/T^2$ leading to $\hat{\phi}_{hac1}^2 = O_p(T^2 M)$. The intercept only case follows from $\hat{\phi}_{hac1}^2/M = O_p(1)$. $\square$

REFERENCES

Andrews, D. W. K. (1991). 'Heteroskedasticity and Autocorrelation consistent covariance matrix estimation', *Econometrica*, Vol. 59, pp. 817-858.

Andrews, D. W. K. (1993). 'Tests for Parameter Instability and Structural Change with Unknown Changepoint', *Econometrica*, Vol. 61, pp. 821-856.

Berenguer-Rico, V., and Nielsen, B. (2015). 'Cumulated sum of squares statistics for non-linear and non-stationary regressions', Unpublished Manuscript, Oxford University.

Berkes, I., Hörmann, S., Horváth, L. (2008). 'The functional central limit theorem for a family of GARCH models', *Statistics and Probability Letters*, Vol. 78, pp. 2725-2730.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, 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.

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

Crainiceanu, C. M. and Vogelsang, T. (2007). 'Nonmonotonic power for tests of a mean shift in a time series', *Journal of Statistical Computation and Simulation*, Vol. 77, pp. 457-476.

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.

Giraitis L., Kokoszka P. and Leipus R. (2000). 'Stationary ARCH models: dependence structure and Central Limit Theorem', *Econometric Theory*, Vol. 16, pp. 3-22.

Giraitis, L., Kokoszka, P. and Leipus, R. (2001). ' Testing for Long Memory in the Presence of a General Trend', *Journal of Applied Probability*, Vol. 38, pp. 1033-1054.

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

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

Juhl, T. and Xiao, Z. (2009). 'Tests for changing mean with monotonic power', *Journal of Econometrics*, Vol. 148, pp. 14-24.

Kasparis, I. (2008). ‘Detection of functional form misspecication in cointegrating relations’, *Econometric Theory*, Vol. 24, pp. 1373-1403.

Kasparis, I., Andreou, E. and Phillips, P.C.B. (2015). ‘Nonparametric Predictive Regressions’, *Journal of Econometrics*, Vol. 185, pp. 468494.

Kejriwal, M.(2009). ‘Tests for a mean shift with good size and monotonic power’, *Economics Letters*, Vol. 102, pp. 78-82.

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

Lettau, M., and Van Nieuwerburgh, S. (2008). ‘Reconciling the return predictability evidence’, *Review of Financial Studies,* Vol. 21, pp. 1607-1652.

Lewellen, J. (2004). ‘Predicting returns with financial rations’, *Journal of Financial Economics*, Vol. 74, pp. 209-235.

Lindner, A.M. (2009). ‘Stationarity, Mixing, Distributional Properties and Moments of GARCH(p,q) Processes’, in T.G. Anderson et al., *Handbook of Financial Time Series*, Vol. 43, Springer-Verlag Berlin Heidelberg.

Newey, W. K. and West, K. D. (1987). ‘A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix’, *Econometrica*, Vol. 55, pp. 703-708.

Phillips, P. C. B. (1987). ‘Time Series Regression with a Unit Root’, *Econometrica*, Vol. 55, pp. 227-301.

Pitarakis, J., (2004). ‘Least Squares Estimation and Tests of Breaks in Mean and Variance under Misspecification’, *Econometrics Journal,* Vol. 7, pp. 32-54.

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. and M. E. Wohar (2006). ‘Structural Breaks and Predictive Regression Models of Aggregate US Stock Returns’, *Journal of Financial Econometrics,* Vol. 4, pp. 238-274.

Sandberg, R. (2009). ‘Convergence to Stochastic Power Integrals For Dependent Heterogeneous Processes’, *Econometric Theory,* Vol. 25, pp. 739-747.

Timmermann, A. (2008). ‘Elusive Return Predictability’, *International Journal of Forecasting*, Vol. 24, pp. 1-18.

Valkanov, R. (2003). ‘Long-horizon regressions: theoretical results and applications’, *Journal of Financial Economics,* Vol. 68, pp. 201-232.

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.

Xiao, Z., and Phillips, P.C.B. (2002). 'A CUSUM test for cointegration using regression residuals', *Journal of Econometrics*, Vol. 108, pp. 4361.

Xu, K. (2013a). 'Powerful tests for structural changes in volatility', *Journal of Econometrics*, Vol. 173, pp. 126-142.

Xu, K. (2013b). 'Power monotonicity in detecting volatility levels change', *Economics Letters*, Vol. 121, pp. 64-69.