Economics Division
University of Southampton
Southampton SO17 1BJ, UK

Discussion Papers in
Economics and Econometrics

Title: A Simple Approach for Diagnosing Instabilities in Predictive Regressions

By : Jean-Yves Pitarakis (University of Southampton),

No. 1519

ISSN 0966-4246

A Simple Approach for Diagnosing Instabilities in Predictive Regressions*

Jean-Yves Pitarakis
University of Southampton
Department of Economics
Southampton SO17 1BJ, U.K
j.pitarakis@soton.ac.uk

January 18, 2015

Abstract

We introduce a method for detecting the presence of time variation and instabilities in the parameters of predictive regressions linking noisy variables such as stock returns to highly persistent predictors such as stock market valuation ratios. Our proposed approach relies on the least squares based squared residuals of the predictive regression and is trivial to implement. More importantly the distribution of our test statistic is shown to be free of nuisance parameters, is already tabulated in the literature and is robust to the degree of persistence of the chosen predictor. Our proposed method is subsequently applied to the predictability of monthly US stock returns with the dividend yield, dividend payout, earnings-price, dividend-price and book-to-market value ratios. Our results strongly support the presence of instabilities over the 1927-2013 period but also clearly point to the disappearance of these after the mid 50s.

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

*Address for correspondence: Jean-Yves Pitarakis, University of Southampton, School of Social Sciences, Department of Economics, SO17 1BJ, United-Kingdom.

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 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) and more recently Kostakis, Magdalinos and Stamatogiannis (2014) 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 a recent paper by Kostakis, Magdalinos and Stamatogiannis (2014) 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 settings and are not necessarily suitable for the specificities of predictive regressions. It is straightforward to show for instance that the Brownian Bridge type asymptotics of the most commonly used SupWald type test of Andrews (1993) would no longer be valid when considering nearly integrated predictors. In Rapach and Wohar (2006) the authors used the standard SupWald based together with bootstrap based 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, Magdalinos and Stamatogiannis (2014) 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.

Our goal in 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 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 distribution of our proposed test statistic is free of nuisance parameters, tabulated and does not depend on the unknown non-centrality parameter driving the degree of persistence of the predictors. Our method offers 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 it very clearly highlights a significant switch in predictability that occurred near the mid 50s/early 60s after which time predictability vanishes.

The plan of the paper is as follows. Section 2 introduces our operating model, assumptions and test statistic and obtains its large sample properties. Section 3 is a simulation study highlighting the finite sample size and power properties of our procedure. Section 4 applies our methodology to the predictability of aggregate US returns using the recently extended Goyal and Welch (2013) dataset also considered in Kostakis, Magdalinos and Stamatogiannis (2013). Section 5 concludes.

2 A Cumulative Squared Residuals Based Test

Throughout this paper our operating model is given by the following predictive regression

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

with the predictor x_t modelled as a nearly integrated process so as to capture the frequently observed high degree of serial correlation of commonly considered predictors

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

with $c < 0$ and u_t and v_t denoting stationary disturbances. The probabilistic properties of our specification are summarised in the following set of assumptions.

ASSUMPTIONS: *(i)* $v_t = \Psi(L)\epsilon_t$ *with* $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ *having* $\Psi(1) \neq 0$, $\Psi_0 = 1$ *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* $E[w_t w_t' | \mathcal{F}_{t-1}] = \Sigma_w \equiv \{(\sigma_u^2, \sigma_{u\epsilon}), (\sigma_{\epsilon u}, \sigma_\epsilon^2)\}$ *and* $\sup_t E[||w_t||^{4+\delta} | \mathcal{F}_{t-1}] < \infty$ *for some* $\delta > 0$*. (iii)* $\eta_t = u_t^2 - \sigma_u^2$ *is such that* $E[\eta_t^2 | \mathcal{F}_{t-1}] < \infty$*.*

The above assumptions are standard within the predictive regression literature (see for instance Jansson and Moreira (2006), Campbell and Yogo (2006), Kostakis, Magdalinos and Stamatogiannis (2013) and others) possibly with the exception of requiring the existence of sufficiently high order moments for the errors driving the predictive regression. The latter are important in our context since our inferences will be relying on the variance of squared residuals themselves. The martingale difference setting for u_t is a natural choice since in most applications the null hypothesis is typically understood to describe an efficient market in the sense of excess returns being a fair game. Letting $\hat{u}_t$ denote the least squares residuals estimated from (1) in what follows $\hat{\sigma}_u^2$ will denote the residual variance. Setting $\hat{\eta}_t = \hat{u}_t^2 - \hat{\sigma}_u^2$ we also let $\hat{\phi}^2$ denote the variance of $\hat{\eta}_t$. Note that under an NID setting for the $u_t's$ we have $\phi^2 = 2\sigma_u^4$. Throughout the paper we will also let k refer to the potential location of a break-point in the parameters driving (1) and set $k = [T\pi]$ with π denoting the break fraction.

Our inferences about the potential presence of time variation in the predictive regression in (1) will rely on the fluctuations of the squared residuals as captured by a CUSUM of squares type quantity which we formulate as

$$C_{SQ} = \max_{1 \leq k \leq T} \frac{1}{\hat{\phi}\sqrt{T}} \left| \sum_{t=1}^{k} \hat{u}_t^2 - \frac{k}{T} \sum_{t=1}^{T} \hat{u}_t^2 \right|. \tag{3}$$

Since the early work of Brown (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 innumerable variants have been extensively used in applications in virtually all scientific fields. In Xiao and Phillips (2002) for instance the authors used the CUSUM principle to develop a test for detecting the presence of cointegration within a single equation setting while generalisations of the properties of CUSUM of squares have been explored in Deng and Perron (2008a, 2008b) and others. The idea behind a test statistic such as (3) is that any time variation within the predictive regression will contaminate the standard least squares residuals and their squares and hence should be detectable by analysing how $\hat{u}$ and $\hat{u}_t^2$ fluctuate. The following proposition introduces our first result.

PROPOSITION 1. Under Assumptions (i)-(ii) and model (1)-(2) we have $C_{SQ} \Rightarrow \sup_{\pi \in [0,1]} |BB(\pi)|$ *with* $BB(\pi)$ *denoting a standard Brownian Bridge.*

An important and unique feature of our limiting result in Proposition 1 stems from the fact that the unknown noncentrality parameter c characterising the degree of persistence of x_t does not enter into its expression, making the practical implementation of our approach particularly straightforward. The limiting distribution of C_{SQ} is given by the supremum of a Brownian Bridge and is well known and extensively tabulated in the literature. We indeed have

$$P\left(\sup_{\pi \in [0,1]} |BB(\pi)| \leq u \right) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2 u^2} \tag{4}$$

which can easily be used to construct suitable p-values (see Billingsley (1968)). Alternatively, the 1%, 5% and 10% critical values of the distribution are given by 1.63, 1.36 and 1.22 respectively.

We next explore the properties of the above test statistic when the model in (1)-(2) is truly characterised by structural break type of instabilities. This is a particularly important issue in the context of a CUSUM of squares type statistic for which there is an extensive literature that discusses its deficiencies in relation to its local power properties. Interestingly and as we demonstrate further below our environment in (1)-(2) that involves a highly persistent regressor leads to fundamentally different and significantly more favourable local power properties for C_{SQ} compared to purely stationary environments considered in the literature.

Operating within a purely stationary setting Ploberger and Kramer (1990) considered the local power properties of C_{SQ} and highlighted its trivial power in the sense of power asymptotically converging to size for local deviations from the null (e.g. $y_t = \beta x_{t-1} + u_t$ with $\beta_t = \beta + \frac{\delta}{\sqrt{T}} I(t > k_0)$). More recently however, Deng and Perron (2008a), in their comprehensive study of the power properties of CUSUM and CUSUM of quares tests argued that the usefulness and validity of such tests should not be judged on the basis of their local power properties which are not indicative or even relevant when dealing with finite samples. They further show that by considering non-local deviations from the null the vast majority of the conclusions pertaining to the power of C_{SQ} are overturned. In what follows we address these issues and highlight the fact that in our specific setting that involves a nearly-integrated predictor the concerns regarding local power need not be regardless of whether one considers local shifts or more realistic fixed departures from the null of linearity.

To illustrate some of the above points we present the outcome of a small Monte-Carlo experiment in which we parameterise the predictive regression (ignoring intercepts for simplicity) $y_t = \beta_t x_{t-1} + u_t$ with $\beta_t = \beta + (\delta/\sqrt{T}) I(t > k_0)$ and contrast the local power properties of C_{SQ} across a purely stationary predictor (say $x_t = 0.2 x_{t-1} + v_t$) and a significantly more persistent one as in (2) with $c = 1$. Results are displayed in Table 1 below where we used $\beta = 2$ and $\delta = 3$.

Table 1. Local Power of C_{SQ} in Stationary and Persistent Environments

	T=100	T=200	T=400	T=800	T=2500
			$\pi_0 = 0.25$		
Stationary x_t	2.30%	3.50%	3.80%	4.80%	4.80%
Persistent x_t	16.20%	30.30%	43.50%	60.30%	80.80%
			$\pi_0 = 0.50$		
Stationary x_t	2.10%	3.10%	3.30%	4.70%	5.00%
Persistent x_t	8.00%	20.60%	37.40%	56.60%	78.70%
			$\pi_0 = 0.75$		
Stationary x_t	2.40%	2.90%	3.30%	5.30%	5.20%
Persistent x_t	27.80%	45.60%	58.60%	71.30%	87.40%

The above figures corroborate some of the findings of the existing literature (e.g. Ploberger and Kramer (1990), Deng and Perron (2008a)) in addition to highlighting the distinct and favourable behaviour of C_{SQ} under a persistent predictor. It is clear that within a stationary setting the CUSUM of squares test has trivial power (note that we used a 5% nominal size when assessing empirical power) while under our nearly integrated setting power increases with the sample size and converges to 1. Within this persistent context another interesting point we can infer from Table 1 is the sensitivity of local power to the location of the true break point. Power appears to be weakest when the true break point is located at the middle of the sample (i.e. $\pi_0 = 0.5$), a phenomenon also highlighted in Deng and Perron (2008a).

Our next goal is to offer a more formal analysis of the properties highlighted above. Letting $\widetilde{x}_t = (1, x_{t-1})'$ and $\theta = (\alpha, \beta)'$ we rewrite (1) as

$$\begin{aligned} y_t &= \theta_t' \widetilde{x}_{t-1} + u_t \\ \theta_t &= \theta + \frac{\delta}{\sqrt{T}} I(t > k_0) \end{aligned} \tag{5}$$

with k_0 denoting the location of the true break point.

If x_t is taken to be a purely stationary predictor it is straightforward to show that under local departures as in (5) we continue to have

$$\frac{\sum_{t=1}^{T} \hat{u}_t^2}{\sqrt{T}} = \frac{\sum_{t=1}^{T} u_t^2}{\sqrt{T}} + o_p(1) \tag{6}$$

and similarly $\sum \hat{u}_t^2/T \xrightarrow{p} \sigma_u^2$ and $\sum \hat{u}_t^4/T \xrightarrow{p} E[u_t^4]$ so that the limiting behaviour of C_{SQ} remains as in Proposition 1. Naturally and as discussed in Deng and Perron (2008a) this is no longer the case when considering the more realistic setting of fixed and sufficiently large departures from the null. More importantly as we consider a nearly integrated predictor as in (2) and regardless of whether we operate locally or not our test statistic is characterised by non-trivial power. Indeed, we now have

$$\frac{\sum_{t=1}^{T} \hat{u}_t^2}{\sqrt{T}} = \frac{\sum_{t=1}^{T} u_t^2}{\sqrt{T}} + O_p(\sqrt{T}) \tag{7}$$

and $\sum \hat{u}_t^2/T \xrightarrow{p} \sigma_u^2 + O_p(1)$ and $\sum \hat{u}_t^4/T = O_p(1)$ so that $C_{SQ} = O_p(\sqrt{T})$. This result is formalised in the following proposition.

PROPOSITION 2. Under Assumptions (i)-(ii) and model (5) we have $T^{-\frac{1}{2}} C_{SQ} = Z_\infty(c, \pi_0, \delta)$ with $Z_\infty(.)$ denoting a stohastically bounded $O_p(1)$ random variable.

The key point of the above Proposition is the divergence of the test statistic under a time varying setting parameterised as in (5) so as to ensure the test has nontrivial power. The particular expression of the stochastically bounded term is not interesting per se and is given by a random variable that depends on the noncentrality parameter c, the location of the true break point/fraction ($\pi_0 = k_0/T$) and of course the magnitude of δ.

3 Finite Sample Performance: Size and Power Properties

We initially illustrate the finite sample size properties of our test statistic whose asymptotic behaviour has been established in Proposition 1. Our DGP is given by (1)-(2) with $v_t = \phi_v v_{t-1} + \epsilon_t$. We let $w_t = (u_t, \epsilon_t)'$ denote a bivariate iid Gaussian random vector with nondiagonal covariance $\Sigma_w = \{(1, \sigma_{u\epsilon}), (\sigma_{u\epsilon}, 1)\}$. We set $(\alpha, \beta) = (0.15, 0.25)$, $\phi_v = 0.5$ and $\sigma_{u\epsilon} = -0.3$. In order to highlight the robustness of our results to the persistence parameter we repeat all our simulations for $c \in \{1, 10, 40\}$ across $T = 200, 600, 1000, 1400, 1800$. Results are presented in Table 2 below where the first four columns display the finite sample and asymptotic critical values of the null distribution of C_{SQ} while the remaining columns display the associated empirical sizes for a selection of three nominal sizes.

Table 2. Critical Values and Empirical Size Properties of C_{SQ}

	10%	5%	2.50%	1%	10%	5%	2.50%
		c=1				c=1	
T=200	1.165	1.294	1.385	1.525	7.7	3.2	1.4
T=600	1.207	1.348	1.466	1.613	9.1	4.6	2.2
T=1000	1.186	1.319	1.440	1.607	8.2	4.2	2.1
T=1400	1.199	1.321	1.454	1.586	8.7	4.2	2.1
T=1800	1.208	1.348	1.469	1.612	9.3	4.8	2.3
T=∞	1.220	1.360	1.480	1.630	10.0	5.0	2.5
		c=10				c=10	
T=200	1.167	1.305	1.384	1.530	7.6	3.1	1.4
T=600	1.206	1.342	1.469	1.613	9.0	4.7	2.2
T=1000	1.187	1.324	1.445	1.605	8.2	4.3	2.1
T=1400	1.197	1.319	1.457	1.586	8.7	4.2	2.1
T=1800	1.210	1.348	1.463	1.618	9.2	4.8	2.3
T=∞	1.220	1.360	1.480	1.630	10.0	5.0	2.5
		c=40				c=40	
T=200	1.164	1.301	1.383	1.530	7.6	3.2	1.4
T=600	1.207	1.347	1.459	1.609	9.1	4.7	2.3
T=1000	1.186	1.332	1.445	1.604	8.2	4.3	2.0
T=1400	1.196	1.319	1.454	1.588	8.5	4.0	2.1
T=1800	1.208	1.345	1.463	1.604	9.3	4.8	2.3
T=∞	1.220	1.360	1.480	1.630	10.0	5.0	2.5

We note that the empirical size estimates remain virtually identical across the different magnitudes of the near persistence parameter c for all sample sizes as expected by our theory. In small samples the test statistic suffers from a mild undersizeness which progressively corrects as we move towards sample sizes such as $T = 600$ and beyond.

For our power experiments we consider $\theta = (0.15, 0.25)'$ with $\delta = (2, 2)'$ and the departures from the null are parameterised as in (5). In order to highlight the influence of the location of the break point on the power properties of our test we consider three alternative scenarios that place k_0 early in the sample, late in the sample as well as its middle (e.g. $\pi_0 = 0.25, 0.50, 0.75$). Results are presented in Table 3 below where we used the critical values of Table 2 to evaluate size corrected empirical powers.

Table 3. Power Properties (5% Nominal Size)

	c=1	c=10	c=40
	$\pi_0 = 0.25$		
T=200	36.4	11.0	5.2
T=600	61.6	24.1	6.4
T=1000	73.5	37.7	8.1
T=1400	80.0	50.2	11.7
T=1800	84.0	54.5	9.9
	$\pi_0 = 0.5$		
T=200	32.2	6.5	5.2
T=600	62.2	11.6	4.3
T=1000	71.0	18.4	5.0
T=1400	81.5	26.7	5.9
T=1800	82.7	27.7	4.9
	$\pi_0 = 0.75$		
T=200	55.1	12.9	5.7
T=600	76.3	28.8	7.4
T=1000	84.9	45.5	8.4
T=1400	89.7	58.8	13.1
T=1800	91.5	64.5	14.4

Our results in Table 3 clearly highlight the sensitivity of power to c and π_0. For large values of c (i.e. for regressors further away from the nearly integrated scenario) we note that the test has no power as discussed above. It is particularly interesting to note how the power properties of C_{SQ} are significantly altered as we consider highly persistent regressors. Under this latter scenario we note that power increases towards one as we increase T. It is also interesting to highlight the role of the location of the true break point on power. We note for instance that power is overall similar when $\pi_0 = 0.25$ or $\pi_0 = 0.50$ but tends to increase significantly when $\pi_0 = 0.75$ i.e. when the break point is located later on in the sample. Overall, under a strongly persistent scenario (e.g. $c = 1$) and sample sizes typically encountered in financial data C_{SQ} appears to offer a useful diagnostic tool.

4 Time Varying Return Predictability

We apply our methodology to the predictability of US equity returns with valuation ratios as recently explored in Kostakis, Magdalinos and Stamatogiannis (2014) 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). The key contribution of KMS 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} statistic to the magnitude of the noncentrality parameter c makes it particularly suitable for diagnosing predictive instability in a simple way. It is also interesting to point out that the mere finding of instability is itself evidence of predictability since it indicates changes in the values of the parameters driving the predictive regression which cannot be zero in both regimes.

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 KMS 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} as expressed in (3). Results are presented in Table 3 below where *** indicates rejection at 1% level, ** at 5% and * at 10%.

Table 3. C_{SQ} Statistic $(1.63(1\%), 1.36(5\%), 1.22(10\%))$

	1927-2013	1950-2013	1955-2013	1960-2013
DY	3.538***	1.528**	1.283*	1.037
EP	3.458***	1.503**	1.271*	1.032
DP	3.521***	1.521**	1.278*	1.034
DPO	3.480***	1.431**	1.243*	1.011
BM	3.583**	1.479**	1.265*	1.027

The above results demonstrate a strong presence of instability in all five of the predictive regressions when considering the full sample period of 1927-2013. It is also clear however that this instability vanishes as we exclude pre-mid 50s data. Looking at the 1955-2013 results for instance we note that across all five predictors the C_{SQ} statistic leads to a borderline rejection at 10% while when considering the 1960-2013 period the computed test statistic is significantly below the 10% cutoff. Our results fully corroborate the sensitivity analysis conducted in KMS and highlight the usefulness of our procedure for uncovering instability in predictive regressions. The method is trivial to implement, it relies on existing tabulated distributions and is robust to the nearly integrated nature of predictors. Furthermore, it has 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 a cumulative sum of squared least squares residuals. Besides its simplicity, another important feature of our test statistic is the convenience of its limiting distribution that does not depend on the noncentrality parameter used to model the persistent predictors and thus making it a useful diagnostic tool when considering the use of predictive regressions. 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 significantly more challenging extension could also involve the coexistence of instabilities in both the conditional mean and error variances along the lines studied in the earlier work of Pitarakis (2004).

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 it is well known that $\sum_{t=1}^{[T\pi]} x_{t-1}^2/T^2 \Rightarrow \int_0^{\pi} J_c^2(r)dr \equiv q(0,\pi)$ (see also Sandberg (2009)) and $\sum_{t=1}^{[T\pi]} x_{t-1}u_t/T = O_p(1)$. More generally, using the continuous mapping theorem and our assumptions on the finiteness of moments in (ii)-(iii) we also have $\sum_{t=1}^{[T\pi]} x_{t-1}^m/T^{1+\frac{m}{2}} = O_p(1)$, $\sum_{t=1}^{[T\pi]} x_{t-1}^m u_t/T^{(1+m)/2} = O_p(1)$, $\sum_{t=1}^{[T\pi]} x_{t-1}^2 u_t^2/T^2 = O_p(1)$ 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 $\sqrt{T}(\hat{\alpha}-\alpha) = O_p(1)$ (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 $\hat{\phi}^2 \xrightarrow{p} \phi^2 \equiv [\eta_t^2]$.

PROOF OF PROPOSITION 1: We have $\sum_{t=1}^k \hat{u}_t^2 = \sum_{t=1}^k u_t^2 + (\hat{\theta}-\theta)' \sum_{t=1}^k x_{t-1}x_{t-1}'(\hat{\theta}-\theta) - 2\sum_{t=1}^k u_t x_{t-1}'(\hat{\theta}-\theta)$. It is also convenient to introduce the normalising matrix $D_T = diag(\sqrt{T}, T)$ so that we can write

$$\begin{aligned}\frac{\sum_{t=1}^k \hat{u}_t^2}{\sqrt{T}} &= \frac{\sum_{t=1}^k u_t^2}{\sqrt{T}} + \frac{1}{\sqrt{T}}[D_T(\hat{\theta}-\theta)]'\left(D_T^{-1}\sum_{t=1}^k x_{t-1}x_{t-1}'D_T^{-1}\right)D_T(\hat{\theta}-\theta) - \\ & 2\frac{1}{\sqrt{T}}\sum_{t=1}^k u_t x_{t-1}' D_T^{-1}(D_T(\hat{\theta}-\theta)) \end{aligned} \tag{8}$$

Given our chosen process in (2) and our 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 $\max_k D_T^{-1}\sum_{t=1}^k x_{t-1}x_{t-1}'D_T^{-1}$ together with $D_T(\hat{\theta}-\theta) = O_p(1)$ also ensures that $T^{-1/2}(D_T(\hat{\theta}-\theta))'(\max_k D_T^{-1}\sum x_{t-1}x_{t-1}'D_T^{-1})(D_T(\hat{\theta}-\theta)) \xrightarrow{p} 0$. 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 $\max_k|\sum_{t=1}^k \hat{u}_t^2/\sqrt{T} - \sum_{t=1}^k u_t^2/\sqrt{T}| \xrightarrow{p} 0$. Next, letting $K_T(\pi) = (1/\phi\sqrt{T})\sum_{t=1}^{[T\pi]}(u_t^2-\sigma_u^2)$, assumptions (i)-(iii) ensure that an invariance principle holds with $K_T(\pi) \Rightarrow B(\pi)$. With $C_{SQ} = (\phi/\hat{\phi})\sup_\pi|K_T(\pi) - \pi K_T(1)| + o_p(1)$ the result follows from the continuous mapping theorem.

PROOF OF PROPOSITION 2: For simplicity we consider a specification with no intercept so that the true DGP is $y_t = \beta x_{t-1} + (\delta/\sqrt{T})x_{t-1}I(t > k_0) + u_t$ while the residuals are obtained from the fitted model $\hat{u}_t = y_t - \hat{\rho}x_{t-1}$ leading to $\hat{u}_t = u_t - (\hat{\rho}-\beta)x_{t-1} + (\delta/\sqrt{T})x_{t-1}I(t > k_0)$. From $\hat{\rho} = \beta + (\delta/\sqrt{T})(\sum_{t=k_0+1}^T x_{t-1}^2/\sum_{t=1}^T x_{t-1}^2) + \sum_{t=1}^T u_t x_{t-1}/\sum_{t=1}^T x_{t-1}^2$ it follows that $\sqrt{T}(\hat{\rho}-\beta) = \delta(q(\pi_0,1)/q(0,1)) + o_p(1)$ where $q(\pi_0,1) \equiv \int_{\pi_0}^1 J_c^2(r)dr$ and $q(0,1) \equiv \int_0^1 J_c^2(r)dr$. Applying suitable normalisations to $\hat{u}_t^2$ leads to

$$\frac{1}{T}\sum_{t=1}^T \hat{u}_t^2 \quad \Rightarrow \quad \sigma_u^2 + \delta^2\left(q(\pi_0,1) - \frac{q(\pi_0,1)^2}{q(0,1)}\right) \tag{9}$$

so that within this local setting we have $\sum_{t=1}^T \hat{u}_t^2/\sqrt{T} = O_p(\sqrt{T})$. Next, we consider the large sample behaviour of $\sum_{t=1}^k \hat{u}_t^2$ for $k \leq k_0$ and $k > k_0$ respectively. We have

$$\frac{\sum_{t=1}^{[T\pi]} \hat{u}_t^2}{T} \quad \Rightarrow \quad \pi\sigma_u^2 + \delta^2\left(\frac{q(\pi_0,1)}{q(0,1)}\right)^2 q(0,\pi) \quad \pi \leq \pi_0 \tag{10}$$

and

$$\frac{\sum_{t=1}^{[T\pi]} \hat{u}_t^2}{T} \quad \Rightarrow \quad \pi\sigma_u^2 + \delta^2\left(\frac{q(\pi_0,1)}{q(0,1)}\right)^2 q(0,\pi_0) + \delta^2 q(\pi_0,1)\left(1 - \frac{q(\pi_0,1)}{q(0,1)}\right)^2 \quad \pi > \pi_0. \tag{11}$$

Combining the above results establishes that $|(\sum_{t=1}^k \hat{u}_t^2/\sqrt{T}) - \frac{k}{T}\sum_{t=1}^T \hat{u}_t^2| = O_p(\sqrt{T})$. Proceeding as in Proposition 1 it is also straightforward to establish that under our local setting $\hat{\phi}^2 = O_p(1)$, leading to the result that $C_{SQ}/\sqrt{T} = O_p(1)$.

REFERENCES

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

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

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

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

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

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*, 74, 681-714.

Kostakis, A., Magdalinos, A., and Stamatogiannis, M. (2014), "Robust Econometric Inference for Stock Return Predictability," *Review of Financial Studies,* forthcoming.

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

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

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

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

Rapach, D. and M. E. Wohar (2006), "Structural Breaks and Predictive Regression Models of Aggregate US Stock Returns," *Journal of Financial Econometrics,* 4, 238-274.

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

Timmermann, A. (2008), "Elusive Return Predictability," *International Journal of Forecasting*, 24, 1-18.

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

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