

# Department of Economics

## Tests for Deterministic Parametric Structural Change in Regression Models

George Kapetanios

Working Paper No. 539

May 2005

ISSN 1473-0278



Queen Mary  
University of London

# Tests for Deterministic Parametric Structural Change in Regression Models

George Kapetanios\*  
Queen Mary, University of London

April 8, 2005

## Abstract

The problem of structural change justifiably attracts considerable attention in econometrics. A number of different paradigms have been adopted ranging from structural breaks which are sudden and rare to time-varying coefficient models which exhibit structural change more frequently and continuously. This paper is concerned with parametric econometric models whose coefficients change deterministically and smoothly over time. In particular we provide and discuss tests for the null hypothesis of no structural change versus the alternative hypothesis of smooth deterministic structural change. We provide asymptotic tests for this null hypothesis. However, the finite sample performance of these tests is not good as they overreject significantly. To address this problem we propose and justify bootstrap based tests. These tests perform well in an extensive Monte Carlo study.

*Keywords: Structural Change, Non-Stationarity, Deterministic Time-Variation*  
*JEL code: C10, C14*

---

\*Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS. email: G.Kapetanios@qmul.ac.uk

# 1 Introduction

The investigation of structural change in econometric models has been assuming increasing importance in the literature over the past couple of decades. This focus is not surprising. Assuming wrongly that the structure of a model remains fixed over time, has very significant and adverse implications. The first obvious implication is inconsistency of the parameter estimates. A distinct, yet related, implication is the fact that structural change chance is likely to be responsible for most major forecast failures of time series models.

As a result a huge literature on modelling structural change has emerged. Most of the work assumes that structural changes in parametric models occur rarely and are abrupt. Many tests for the presence of structural change of that form exist in the literature starting with the pathbreaking work of Chow (1960) who assumed knowledge of the point in time at which the change occurred. Other tests that relax this assumption have been developed by Brown, Durbin, and Evans (1974), Ploberger and Kramer (1992) and many others. In this context it is worth noting that little is being said about the cause of structural breaks in either statistical or economic terms. Recent work by Kapetanios and Tzavalis (2004) provides a possible avenue for modelling structural breaks and, thus, addresses partially this issue.

Another more recent strand of the literature takes a different approach. In this approach the coefficients of parametric models are assumed to evolve over time. To achieve this the parameters are assumed to be stochastic processes leading to stochastic time-varying coefficient (STVC) models. Such models bear resemblance to simple nonlinear econometric models such as bilinear models (see Tong (1990)). STVC models have been used recently in applied macroeconomic work by, e.g., Cogley and Sargent (2002), to model the evolution of macroeconomic variables such as US inflation in the post WWII era. In this case coefficients have been assumed to evolve as random walks over time.

An important question arising out of the use of such models goes to the heart of what structural change is. A relatively uncontroversial definition would be a change in the unconditional moments of the process under investigation. If one were to adopt this definition, use of STVC models may be problematic. To see this we note that, as mentioned above, these models can be viewed as nonlinear time series models. But processes following nonlinear models of that form can be strictly stationary under certain assumptions (see, e.g.,

Pourahmadi (1988) and Liu and Brockwell (1988)). Alternatively, such processes may have asymptotically infinite variance like random walk processes. Neither of these alternatives is what is needed in the modelling of structural change. For example, a process such as US inflation whose variance has been falling in the last decade cannot be theoretically consistent with a model whose coefficients follow random walk processes.

These problems may be addressed by assuming that coefficients change but in a smooth deterministic way. Such modelling attempts have a long pedigree in statistics starting with the work of Priestley (1965). This paper suggested that processes may have time-varying spectral densities which change slowly over time. The context of this work is nonparametric. This work has more recently been followed up by Dahlhaus (1996) and others who refer to such processes as semi-stationary processes. A parametric alternative to this approach has been pursued by Robinson (1989) for linear regression models and Robinson (1991) for nonlinear parametric models. Recently, Orbe, Ferreira, and Rodriguez-Poo (2005) extended these results to include time-varying seasonal effects. We will refer to such parametric models as deterministic time-varying coefficient (DTVC) models. A disadvantage of such an approach is that the coefficient change cannot be modelled or, for that matter, forecast. Both of these are theoretically possible with STVC. However, an important assumption underlying DTVC is that coefficients change slowly. As a result forecasting may be carried out by assuming that the coefficients remain at their end-of-observed-sample value.

This paper will focus on testing for stationarity against the alternative of structural change of the DTVC type. This problem is clearly of interest both for modelling and forecasting purposes. A number of asymptotic results available in the literature on the properties of the estimates of time-varying coefficients can be used to provide such tests. We discuss these in detail. We propose two types of tests. One looks at the difference between the estimate of a coefficient under the null and under the alternative hypothesis to construct a test. This has a number of similarities with work in the nonparametric literature such as Ait-Sahalia, Bickel, and Stocker (2001).

We also take another approach. Under the null hypothesis the derivative of the coefficient with respect to time is zero. This hypothesis can be tested. We discuss this test. From an asymptotic point of view this last test appears less useful. It is easy to show that the latter test will be  $Th^3$  consistent whereas the former is  $Th$  consistent where  $T$  is the number of observations and  $h$  is a window length which tends asymptotically to zero. However,

asymptotics are often a poor guide to small sample performance especially in nonparametric analysis. Our extensive Monte Carlo study will show that to be the case here. In particular we will see that the tests overreject significantly. As a result we suggest use of the bootstrap. Further we see that in small samples the derivative-based test is more powerful than the test based on the coefficient difference under the two hypotheses.

The paper is structured as follows: Section 2 discusses the theoretical framework. Section 3 proposes the new tests and discusses their asymptotic properties. Section 4 proposes bootstrap versions of the asymptotic tests. Section 5 presents an extensive Monte Carlo study. Finally, Section 6 concludes.

## 2 Preliminaries

Let the model of interest be given by

$$y_t = \beta(t)'x_t + u_t \quad (1)$$

where  $y_t$  and  $x_t$  are the scalar dependent and  $k$ -dimensional explanatory variables respectively.

The following assumptions provide information on the detailed specification of the above model:

**Assumption 1**  $\beta(t) = \beta_{t/T}$  where each element of  $\beta_\tau$ ,  $\beta_{i,\tau}$ ,  $i = 1, \dots, k$ ,  $\tau \in (0, 1)$ , is continuous and twice differentiable on  $(0, 1)$ .

**Assumption 2**  $x_t$  is an  $\alpha$ -mixing sequence with size  $-4/3$  and finite 8-th moments.  $E(x_{is}x_{jt}) = m_{ij,s,t} = m_{ij}(s/T, t/T) + O(T^{-1})$  where  $m_{ij}(\cdot, \cdot)$  is a twice differentiable function of both its arguments.

**Assumption 3**  $u_t$  is a stationary martingale difference sequence with finite 4-th moments which is independent of  $x_t$  at all leads and lags.

**Assumption 4** The function  $K(\cdot)$  is a second order kernel with compact support  $[-1, 1]$  and absolutely integrable Fourier transform.

Assumption 1 is a crucial assumption. It specifies that  $\beta(t)$  is a smooth deterministic function of time. It is interesting to note that it depends not only on the point in time  $t$  but also on the sample size  $T$ . This is necessary since in order to estimate consistently a

particular parameter one needs the sample size that relates to that parameter to tend to infinity. This is achieved in this context by allowing an increasing number of neighboring observations to be informative about  $\beta$  at time  $t$ . In other words we have to assume that as the sample size grows the function  $\beta_\tau$  stretches to cover the whole period of the sample. This setup has precedents in the statistical literature. For example, the concept of slowly varying processes of Priestley (1965) forms an early instance of similar ideas. Assumptions 2 and 3 are standard mixing and moment conditions for the explanatory variables and the error term. It is important to note that  $x_t$  is allowed to be nonstationary. Note further that the martingale difference assumption for the error term is not crucial and is adopted for simplicity. General forms of stationary weak dependence for the error term can be accommodated with minimal changes in the analysis.

Finally, assumption 4 relates to the kernel function that will be used for estimation.

Following Robinson (1989) and Orbe, Ferreira, and Rodriguez-Poo (2005), we propose the following estimator for  $\beta_\tau$ .

$$\hat{\beta}_\tau = \left( \sum_{t=1}^T K_{t,\tau} x_t x_t' \right)^{-1} \left( \sum_{t=1}^T K_{t,\tau} x_t y_t \right) \quad (2)$$

where  $K_{t,\tau} = (Th)^{-1} K((\tau - t)/Th)$ . This estimator bears close resemblance to the standard OLS estimator and it is easy to see that it is the closed form solution of the following optimisation

$$\min_{\beta} \sum_{t=1}^T K_{t,\tau} (y_t - \beta(t)' x_t)^2$$

### 3 Asymptotic Tests

We wish to test the hypothesis that  $\beta_t$  does not change over time.

$$H_0 : \beta_\tau = \beta, \forall \tau \quad (3)$$

against the alternative hypothesis that  $\beta_\tau$  is non-constant and satisfies assumption 1. We start our analysis by looking at pointwise tests, i.e. tests of the hypothesis

$$H_{0\tau} : \beta_\tau = \beta \quad (4)$$

for a fixed  $\tau$ . Let us denote the estimate of  $\beta$  under the null as  $\tilde{\beta}$ . Depending on the assumptions made about  $u_t$ , standard methods can be used to estimate  $\beta$  under the null. For example, in the case where the disturbances are spherical and uncorrelated from  $x_t$  OLS

is an optimal estimator. We construct two such tests. The first looks at the difference between the estimates of  $\beta_t$  under the two hypotheses. The test statistic takes the form

$$T^{\tau,l} = (\hat{\beta}_\tau - \tilde{\beta}_\tau)' \hat{V}(\hat{\beta}_\tau - \tilde{\beta}_\tau)^{-1} (\hat{\beta}_\tau - \tilde{\beta}_\tau) \quad (5)$$

The second test statistic uses the fact that under the null hypothesis

$$\nabla_\tau \beta_\tau \equiv \frac{\partial \beta_\tau}{\partial \tau} = 0, \forall \tau \quad (6)$$

This test statistic takes the form

$$T^{\tau,d} = \nabla_\tau \hat{\beta}'_\tau \hat{V}(\nabla_\tau \hat{\beta}_\tau)^{-1} \nabla_\tau \hat{\beta}_\tau \quad (7)$$

Details on the asymptotic variances  $V(\hat{\beta}_\tau - \tilde{\beta}_\tau)$ ,  $\hat{V}(\nabla_\tau \hat{\beta}_\tau)$  and their estimators are given in Theorems 1 and 2. Before introducing tests for  $H_0$  we examine the asymptotic properties of these tests.

**Theorem 1** *Under assumptions 1-4 and  $h = o(T^{-1/5})$*

$$T^{\tau,l} \xrightarrow{d} \chi_k^2 \quad (8)$$

*Under the alternative hypothesis the test is consistent.*

**Theorem 2** *Under assumptions 1-4,  $h = o(T^{-1/5})$  and  $h = O(T^{-1/v})$ , for some  $v < 7$*

$$T^{\tau,d} \xrightarrow{d} \chi_k^2 \quad (9)$$

*Under the alternative hypothesis the test is consistent.*

These tests can be used to test  $H_{0\tau}$  but not  $H_0$ . For that we need to jointly consider many points in the interval  $(0, 1)$  where  $\tau$  is defined. To conduct this test we need to use summary statistics for a set of pointwise test statistics. The problem has parallels with the problem of testing when a nuisance parameter is unidentified under the null hypothesis. This problem arises in many areas in econometrics such as linearity testing, tests for structural breaks and others (for more details see Davies (1977) and Andrews and Ploberger (1994)). Let the set of points for which test statistics are available be denoted by  $\mathcal{T}_m = \{\tau_1, \tau_2, \dots, \tau_m\}$ , where  $\tau_1 < \tau_2 < \dots < \tau_m$ . Three summary statistics are usually considered. These are given by

$$T_{AVE}^i = \frac{1}{m} \sum_{j=1}^m T^{\tau_j,i}, \quad i = l, d \quad (10)$$

$$T_{SUP}^i = \sup_j T^{\tau_j,i}, \quad i = l, d \quad (11)$$

$$T_{EXP}^i = \frac{1}{m} \sum_{j=1}^m e^{\frac{T^{\tau_j,i}}{2}}, \quad i = l, d \quad (12)$$

We make the following assumption about  $\mathcal{T}_m$ .

**Assumption 5** (i)  $m \rightarrow \infty$  as  $T \rightarrow \infty$ , (ii)  $mh \rightarrow 0$  as  $T \rightarrow \infty$ , (iii)  $\mathcal{T}_m$  is such that for all  $j \geq 2$ ,  $\tau_j - \tau_{j-1} = o(1)$  and  $\tau_j - \tau_{j-1} = O\left(\frac{1}{m}\right)$ .

We have the following theorem on the asymptotic distribution of the above tests.

**Theorem 3** Under  $H_0$ , assumptions 1-5,  $h = o(T^{-1/5})$  and  $h = O(T^{-1/v})$ , for some  $v < 7$ , we have that  $(T_{SUP}^i - a_m)/b_m$  has a cumulative density function given by  $e^{-e^{-x}}$  where choices for  $a_m$  and  $b_m$  are given in the appendix. Further,

$$\left(\frac{\sqrt{m}(T_{AVE}^i - k)}{2k}\right)^2 \xrightarrow{d} \chi_1^2 \quad (13)$$

and

$$\left(\frac{\sqrt{m}(T_{EXP}^i - \mu^{\chi_k})}{V^{\chi_k}}\right)^2 \xrightarrow{d} \chi_1^2 \quad (14)$$

where  $\mu^{\chi_k} = E(e^{z_{\chi_k}^2})$  and  $V^{\chi_k} = E(e^{2z_{\chi_k}^2})$  and  $z_{\chi_k}^2$  denotes a  $\chi_k^2$  random variable. Under the alternative hypothesis that  $\beta_\tau$  is non-constant and satisfies assumption 1, all tests are consistent.

**Remark 1** We note a few facts about the distribution with cdf  $e^{-e^{-x}}$  which is usually referred to as the extreme value distribution. Its probability density function is given by  $e^{-x-e^{-x}}$ . Its cumulants are given by  $\kappa_r = (-1)^r \psi^{(r)}(1)$ , where  $\psi^{(r)}$  is the  $r$ -th polygamma function, i.e.  $\frac{d^r \ln \Gamma(x)}{dx^r}$ . So,  $E(X) = 1 + \gamma$  where  $\gamma$  is Euler's constant ( $\gamma \simeq 0.57722$ ) and  $Var(X) = 1/6\pi^2$ .

## 4 Bootstrap Tests

Given the slow rate of convergence related to nonparametric asymptotics, it is not surprising to note that asymptotic results may not provide good approximations to small sample behaviour. This is the case for the asymptotic tests we discussed in the previous section. More details on the poor performance of these asymptotic tests will be provided in the Monte Carlo study of Section 5. Note that the bad performance of tests based on nonparametric asymptotic results is documented in the literature. In particular, Fan (1995, 1998) show that asymptotic tests have rejection probabilities that deviate significantly from the nominal significance level.

A possible solution for this is the bootstrap. The bootstrap distribution of a statistic can be defined as the exact finite sample distribution function evaluated at an estimate of the unknown parameters. As discussed by Singh (1981), Hall (1986), Hall (1992) and Brown (2000), bootstrapping a studentized statistic that is asymptotically pivotal will provide a



closer approximation to the true distribution than the standard limiting distribution, with coverage differing from the nominal level by only  $O_p(T^{-1})$  instead of  $O_p(T^{-1/2})$ , for independent observations. Results for dependent observations have been provided by Horowitz and Hall (1996) and Bose (1988) among others. A recent summary of the available asymptotic results for dependent observations may be found in Lahiri (2003).

Hartigan (1986), Hall (1988) and Beran (1988) advocate the use of pivoting as a device to reduce the error in rejection probability. Although much of the asymptotic theory for the bootstrap has been developed for the construction of confidence intervals, the well known duality between hypothesis testing and confidence intervals guarantees that any ranking of bootstrap variants for confidence intervals will hold in the case of hypothesis testing.

The drawback of this method has been noted by a number of authors including Tibshirani (1988) and more recently Horowitz (1995). The principal disadvantage is that studentizing requires an estimate of the standard deviation of the test statistic which in some cases can represent a poor approximation to the true value. Further, a pivoting procedure advocated by Beran (1988) requires the use of an inner bootstrap loop and as such there is an obvious trade-off between reduction in approximation error and the attendant computational burden. In addition, we note that the asymptotic theory is not informative in the absence of pivotalness. Since in most cases statistics are only asymptotically pivotal, then faced with a finite sample there is no theory-based ranking for pivotal versus non-pivotal bootstrap statistics. Given the above, small sample analysis maybe an appropriate guide on whether or not to studentize.

These comments are relevant in our case because the variance estimator for the derivative-based test appears to be badly behaved in small samples. All asymptotic tests are badly behaved as evidenced by the Monte Carlo study but the derivative-based test more so. As a result we, firstly, do not consider an asymptotic version of the derivative test, and secondly choose not to studentise this test unlike the test based on coefficient differences.

Below we give the bootstrap algorithm we suggest.

**Algorithm 1**    1. Estimate  $\hat{\beta}_\tau$  using (2) for all points in  $\mathcal{T}_m$ . Estimate  $\tilde{\beta}$  using OLS if appropriate. Obtain OLS residuals. Denote the set of OLS residuals by  $\{\hat{u}_t\}_{t=1}^T$

2. Generate a bootstrap sample for  $u_t$ , denoted  $u_t^*$ , by resampling with replacement from

$\{\hat{u}_t\}_{t=1}^T$  to obtain  $\{u_t^*\}_{t=1}^T$

3. Generate a bootstrap sample for  $y_t$ , denoted  $y_t^*$  by

$$y_t^* = \tilde{\beta}' x_t + u_t^*, \quad t = 1, \dots, T \quad (15)$$

4. Construct bootstrap version of  $T^{\tau,i}$  and  $T_j^i$ ,  $i = l, d$ ,  $j = SUP, AVE, EXP$ , denoted  $T^{\tau,i*}$  and  $T_j^{i*}$ ,  $i = l, d$ ,  $j = SUP, AVE, EXP$ .

5. Repeat steps 2-4,  $B$  times to obtain the empirical distribution of  $T^{\tau,i}$  and  $T_j^i$ ,  $i = l, d$ ,  $j = SUP, AVE, EXP$ .

We can show the following theorem for the bootstrap procedure defined by the above algorithm.

**Theorem 4** Under assumptions 1-4,  $h = o(T^{-1/5})$  and  $h = O(T^{-1/v})$ , for some  $v < 7$ , the bootstrap provides a consistent estimator of the asymptotic distributions of  $T^{\tau,i}$  and  $T_j^i$ ,  $i = l, d$ ,  $j = SUP, AVE, EXP$ .

We denote the bootstrap tests by  $T^{\tau,i*}$  and  $T_j^{i*}$ ,  $i = l, d$ ,  $j = SUP, AVE, EXP$ .

## 5 Monte Carlo Study

### 5.1 Monte Carlo Setup

In this section we present a Monte Carlo study on the small sample properties of the new testing procedures. We consider the following model under the null hypothesis.

$$y_t = x_t + u_t \quad (16)$$

where  $u_t, x_t \sim N(0, 1)$ .  $T = 100, 200, 400$ . For  $\hat{\beta}$  we set  $h = T^{-1/l}$  where  $l = 4, 6, 8$ . The truncated standard normal kernel is used throughout.<sup>1</sup>

Results, for  $T^{\tau,i}$  and  $T_j^i$ ,  $i = l, d$ ,  $j = SUP, AVE, EXP$  are presented in Tables 1 and 2.

For the power properties of the test we consider two different models ( $P1$  and  $P2$ ). The first model is

$$y_t = \beta_t x_t + u_t \quad (17)$$

---

<sup>1</sup>It is obvious (and stated in Assumption 4) that the kernel must have bounded support and in particular that it must be bounded between -1 and 1. For such cases, there exists work in the literature suggesting that asymmetric kernels may be useful (see, e.g., Scaillet (2004)).

where

$$\beta_t = \sin\left(\frac{4\pi}{T}\right) \left(1 - \frac{t}{T}\right) \quad (18)$$

and  $u_t \sim N(0, 1)$ . For the second model the only difference is that  $u_t \sim N(0, \sigma_t^2)$  where

$$\sigma_t^2 = \sin\left(\frac{2\pi}{T}\right) + 2 \quad (19)$$

This model is strictly speaking not covered by the estimation framework of Section 2, since the error term is not stationary. However, recent work by Kapetanios (2005) suggests that the properties of  $\hat{\beta}_t$ , in the case of a stationary error term, extend to this case. For both size and power experiments we set  $\mathcal{T}_m = (0.05, 0.10, \dots, 0.95, 1)$ ,  $\mathcal{T}_m = (0.04, 0.08, \dots, 0.96, 1)$  and  $\mathcal{T}_m = (0.025, 0.05, \dots, 0.975, 1)$  for  $T = 100, 200, 400$  respectively. Results are reported in Tables 3 and 4.

## 5.2 Monte Carlo Results

Preliminary investigation has shown that the behaviour of the derivative-based asymptotic tests is very bad under the null hypothesis. As a result we do not consider these tests in this section. Results for the other tests make interesting reading. In Table 1, we report the rejection probabilities under the null hypothesis for  $T^{\tau, l}$  for  $T = 200$ . Clearly, these tests overreject extensively. As expected the overrejection occurs most intensively at the start and end of the sample, where lack of data cause the tests to be less well behaved. It is unlikely that useful inference can be provided in this context using the asymptotic properties of these tests.

Moving to Table 2 which reports rejection probabilities for  $T_j^l$  and  $T_j^{i*}$ ,  $i = l, d$ , we see that the same holds for the  $T_j^l$  tests. Again, these tests appear of little use to applied researchers faced with samples of sizes usually encountered in, e.g., macroeconomic analysis. On the contrary, the  $T_j^{i*}$ ,  $i = l, d$ , tests are correctly sized for all  $T$  and  $h$  considered. As a result we choose not to consider the asymptotic tests for the power experiments.

Moving on to the power properties we see that the derivative-based tests are more powerful than the tests based on the differences between the coefficients estimated under the null and under the alternative hypothesis, in most cases. The superiority of the derivative-based tests is quite pronounced. For example, in the case of model  $P1$ ,  $T = 100$  and  $l = 4$ , the derivative based supremum test rejects 38% whereas the coefficient-based test reject only

22% of the time. Similar gaps arise in many other cases. This result is in stark contrast with the fact that the asymptotic rates of convergence imply that the derivative-based tests should be less powerful.

We conclude that the substantial advantage enjoyed by the derivative tests is related to the fact that the mean normalisation is different for the derivative and coefficient-based tests. In particular, the derivative-based tests require no normalisation as under the null hypothesis the derivative is known to be equal to zero. For the coefficient-based tests, the normalisation under the null hypothesis is not known and needs to be estimated via an OLS estimation of the null model. This is exacerbated through the bootstrap algorithm which requires a separate estimation of the null model for every bootstrap replication. No such estimation is needed for the derivative tests.

Other interesting patterns emerge from the power results. The supremum-based test seems to perform better than either the average or the exponential test. All tests get more powerful as  $T$  increases. The pattern of behaviour related to the choice of  $h$  is more complex. For the coefficient-based tests the power performance is not monotonic with respect to  $h$ . The tests are least powerful for  $h = T^{1/6}$ . The derivative-based tests are most powerful for  $h = T^{1/4}$  with the power declining for higher  $h$ . Finally, it is worth pointing out that the biggest performance gap between the derivative-based and coefficient tests appears for the supremum tests. The difference exists for the other summary statistics but is less pronounced.

## 6 Conclusion

Structural change is justifiably a major concern in econometric modelling. A number of different paradigms have been adopted ranging from structural breaks which are sudden and rare to time-varying coefficient models which exhibit structural change more frequently and continuously. This paper is concerned with parametric econometric models whose coefficients change deterministically and smoothly over time.

In particular, we provide and discuss tests for the null hypothesis of no structural change versus the alternative hypothesis of smooth deterministic structural change. We provide asymptotic tests for this null hypothesis. However, the finite sample performance of these tests is not good as they overreject significantly. To address this problem we propose and justify bootstrap based tests. These tests perform well in an extensive Monte Carlo study.

# Appendix

We first prove a preliminary lemma.

**Lemma 1** *Under assumptions 1-4 and under the null hypothesis,  $H_0$*

$$\sqrt{T}(\tilde{\beta} - \beta) \xrightarrow{d} N(0, M^{-1}CM^{-1}) \quad (20)$$

where  $M = \lim_{T \rightarrow \infty} \sum_{t=1}^T E(x_t x_t')$  and

$$C = \lim_{T \rightarrow \infty} \left\{ \sum_{t=1}^T E(u_t^2 x_t x_t') + \sum_{t=1}^T \sum_{i=1}^{t-1} E(u_t u_{t-i} x_t x_{t-i}') \right\} \quad (21)$$

**Proof.** This lemma simply proves asymptotic normality for the OLS estimator in the case where  $x_t$  is not stationary. The only ingredients needed for the proof of the lemma is a law of large numbers for  $\frac{1}{T} \sum_{t=1}^T x_t x_t'$  and a central limit theorem for  $\frac{1}{\sqrt{T}} \sum_{t=1}^T x_t u_t$ . Starting with the law of large numbers we note that  $x_t$  is mixing by assumption 2. Then, by Theorem 19.11 of Davidson (1994) it follows that

$$\frac{1}{T} \sum_{t=1}^T x_t x_t' \xrightarrow{p} M \quad (22)$$

Finally, assumptions 2 and 3 are then sufficient for

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T x_t u_t \xrightarrow{d} N(0, C) \quad (23)$$

via theorem 24.6 of Davidson (1994). This completes the proof of the lemma. ■

## Proof of Theorem 1

By Theorem 2 of Orbe, Ferreira, and Rodriguez-Poo (2005) it follows that

$$\sqrt{T}h(\hat{\beta}_\tau - \beta_\tau) \xrightarrow{d} N(0, c_K M_\tau^{-1}) \quad (24)$$

where  $M_\tau = E(x_{T\tau} x_{T\tau}')$  and  $c_K = \int K^2(u) du$ . By Lemma 1 it also follows that, under the null hypothesis

$$\sqrt{T}h(\tilde{\beta}_\tau - \beta_\tau) = o_p(1) \quad (25)$$

Thus

$$\sqrt{T}h(\hat{\beta}_\tau - \tilde{\beta}_\tau) \xrightarrow{d} N(0, c_K M_\tau^{-1}) \quad (26)$$

The null distribution stated in the Theorem then easily follows. Consistency of the test, easily follows by consistency of  $\hat{\beta}_\tau$

## Proof of Theorem 2

For this Theorem we need to derive the asymptotic distribution of  $\nabla_\tau \hat{\beta}_\tau - \nabla_\tau \beta_\tau$ . We have that

$$\hat{\beta}_\tau = \left( \sum_{t=1}^T K_{t,\tau} x_t x_t' \right)^{-1} \left( \sum_{t=1}^T K_{t,\tau} x_t y_t \right) = \hat{V}_{\tau,xx}^{-1} \hat{V}_{\tau,xy} \quad (27)$$

We also note that we can write

$$\beta_\tau = \left( E(x_{[T\tau]} x_{[T\tau]})' \right)^{-1} E(x_{[T\tau]} y_{[T\tau]}) = V_{\tau,xx}^{-1} V_{\tau,xy} \quad (28)$$

It then follows that

$$\nabla_\tau \hat{\beta}_\tau = \hat{V}_{\tau,xx}^{-1} \left( \nabla_\tau \hat{V}_{\tau,xy} - \nabla_\tau \hat{V}_{\tau,xx} \hat{\beta}_\tau \right) \quad (29)$$

and similarly for  $\nabla_\tau \beta_\tau$ . Preempting a rate of convergence equal to  $Th^3$ , we write

$$\sqrt{Th^3} (\nabla_\tau \hat{\beta}_\tau - \nabla_\tau \beta_\tau) \approx \sqrt{Th^3} \left( \hat{V}_{\tau,xx}^{-1} - V_{\tau,xx}^{-1} \right) \left( \nabla_\tau \hat{V}_{\tau,xy} - \nabla_\tau \hat{V}_{\tau,xx} \hat{\beta}_\tau \right) + \quad (30)$$

$$\hat{V}_{\tau,xx}^{-1} \left( \sqrt{Th^3} \left( \nabla_\tau \hat{V}_{\tau,xy} - \nabla_\tau V_{\tau,xy} \right) - \nabla_\tau \hat{V}_{\tau,xx} \sqrt{Th^3} \left( \hat{\beta}_\tau - \beta_\tau \right) \right) \quad (31)$$

where  $A \approx B$  implies that  $A - B = o_p(1)$ . But we have that

$$\sqrt{Th^3} \left( \hat{V}_{\tau,xx}^{-1} - V_{\tau,xx}^{-1} \right) \left( \nabla_\tau \hat{V}_{\tau,xy} - \nabla_\tau \hat{V}_{\tau,xx} \hat{\beta}_\tau \right) = o_p(1) \quad (32)$$

and

$$\hat{V}_{\tau,xx}^{-1} \nabla_\tau \hat{V}_{\tau,xx} \sqrt{Th^3} \left( \hat{\beta}_\tau - \beta_\tau \right) = o_p(1) \quad (33)$$

Hence, we concentrate on  $\sqrt{Th^3} \left( \nabla_\tau \hat{V}_{\tau,xy} - \nabla_\tau V_{\tau,xy} \right)$ . But, under the assumptions of the theorem including those on  $h$ ,

$$\sqrt{Th^3} \left( \nabla_\tau \hat{V}_{\tau,xy} - \nabla_\tau V_{\tau,xy} \right) = \sqrt{Th} \sum_{t=1}^T \nabla_\tau K_{t,\tau} x_t u_t \xrightarrow{d} N \left( 0, \left( \int (\nabla_\tau K(\tau))^2 \right) M_\tau \right) \quad (34)$$

Overall,

$$\sqrt{Th^3} (\nabla_\tau \hat{\beta}_\tau - \nabla_\tau \beta_\tau) \xrightarrow{d} N \left( 0, \left( \int (\nabla_\tau K(\tau))^2 \right) M_\tau^{-1} \right) \quad (35)$$

The null distribution of the theorem follows immediately. Consistency of the test easily follows by consistency of  $\nabla_\tau \hat{\beta}_\tau$ .

### Proof of Theorem 3

The first step of the proof of this theorem relates to proving independence of the test statistics  $T^{\tau_j, i}$ ,  $j = 1, \dots, m$ ,  $i = l, d$ . Under assumption 5 (iii), it is the case for the set  $\mathcal{T}_m$  that for all but a finite number of points in  $\mathcal{T}_m$ ,  $\tau_i - \tau_{i-1} = O(\frac{1}{m})$ . Given assumptions 5 (ii), 2 and 3 it follows that the data used in the construction of  $T^{\tau_j, i}$  are independent of the data used in the construction of  $T^{\tau_{j-1}, i}$  and  $T^{\tau_{j+1}, i}$ . Hence, as  $T \rightarrow \infty$  and  $m \rightarrow \infty$  the summary statistics are constructed using independent  $\chi^2$  random variables. This is sufficient for (13) and (14).

We need to prove the result for  $T_{SUP}^i$ ,  $i = l, d$ . To obtain this we use results from the asymptotic theory of extreme order statistics. Following Arnold, Balakrishnan, and Nagaraja (1992) or Galambos (1978), there exist only three forms for the asymptotic cumulative distribution function of an appropriate normalisation of this statistic, given by  $(T_{SUP}^i - a_N)/b_N$ . It is not always the case that such an asymptotic representation exists. These cumulative distributions are given by

$$G_1(x, \alpha) = \begin{cases} 0 & \text{if } x \leq 0 \\ e^{-x^{-\alpha}} & x > 0; \alpha > 0 \end{cases} \quad (36)$$

$$G_2(x, \alpha) = \begin{cases} e^{-(-x)^\alpha} & x < 0; \alpha > 0 \\ 1 & x \geq 0 \end{cases} \quad (37)$$

and

$$G_3(x) = e^{-e^{-x}} \quad (38)$$

According to Theorem 8.3.2 of Arnold, Balakrishnan, and Nagaraja (1992) the distribution can be of the form  $G_2$  only if  $F_k^{-1}(1)$  is finite where  $F_k(\cdot)$  is the cdf of a  $\chi_k^2$  random variable. Since  $F_k^{-1}(1) = \infty$ ,  $G_2$  is not the form of the asymptotic distribution.

The distribution is  $G_1$  iff the following condition applies

$$\lim_{t \rightarrow \infty} \frac{1 - F_k(tx)}{1 - F_k(t)} = x^{-\alpha} \quad (39)$$

But by applying L'Hopital's rule we can easily see that this limit is infinity for  $x > 1$  using  $F_k(x) = \Gamma_{x/2}(k/2)/\Gamma(k/2)$  and  $f_k(x) = \frac{1}{2^{k/2}\Gamma(k/2)}e^{-x/2}x^{(k/2)-1}$  where  $\Gamma_a(b) \equiv \int_0^a e^{-t}t^{a-1}dt$  is the incomplete Gamma function. Thus, we need to either verify that  $G_3$  is the appropriate distribution or conclude that no such distribution exists.

To check whether  $G_3$  is the appropriate distribution we use the third von Mises condition given in Theorem 8.3.3 of Arnold, Balakrishnan, and Nagaraja (1992). This condition states

that the asymptotic distribution is  $G_3$  iff

$$\lim_{x \rightarrow F^{-1}(1)} \frac{d}{dx} \left\{ \frac{1 - F_k(x)}{f_k(x)} \right\} = 0 \quad (40)$$

where  $f_k(x)$  is the pdf of a  $\chi_k^2$  random variable.

The above condition is equivalent to

$$\lim_{x \rightarrow \infty} \frac{-f_k''(x)(1 - F_k(x)) + f_k'(x)f(x)}{2f_k(x)f_k'(x)} = 1 \quad (41)$$

where  $f_k'(x)$  and  $f_k''(x)$  are the first and second derivatives of  $f(x)$ . Then, it is easy to see that we need to prove

$$\lim_{x \rightarrow \infty} \frac{-f_k''(x)(1 - F_k(x))}{f_k(x)f_k'(x)} = 1 \quad (42)$$

Simple algebra indicates that for the  $\chi_k^2$  pdf

$$\lim_{x \rightarrow \infty} \frac{-f_k''(x)}{f_k'(x)} = 1/2 \quad (43)$$

Further, by a double application of L'Hopital's rule, it follows that

$$\lim_{x \rightarrow \infty} \frac{1 - F_k(x)}{f_k(x)} = 2 \quad (44)$$

proving that the required distribution is indeed  $G_3$ . Then, by part (iii) of theorem 8.3.4 of Arnold, Balakrishnan, and Nagaraja (1992) we have that possible (but not unique) expressions for  $a_N$  and  $b_N$  are given by:

$$a_N = F^{-1}(1 - N^{-1}), \quad b_N = F^{-1}(1 - (Ne)^{-1}) - F^{-1}(1 - N^{-1}) \text{ or } b_N = [Nf(a_N)]^{-1} \quad (45)$$

Finally, example 8.3.4 of Arnold, Balakrishnan, and Nagaraja (1992) implies that  $b_N \sim \log N$  and that  $T_{SUP}^i = O_p(\log N)$ .

Consistency of the tests follows from assumption 5 (iii), continuity of  $\beta_\tau$ , consistency of  $T^{\tau,i}$  and the rate of convergence of  $m$ .



## Proof of Theorem 4

In order to prove the theorem it is sufficient to show that the resampled residuals have finite fourth moments in the bootstrap probability space, i.e.  $E^*(u_t^{*4}) < \infty$ . If that is the case then assumptions 1-4 are satisfied for every bootstrap sample implying via Theorems 1-3 that the bootstrap can be used to estimate the asymptotic distributions of  $T^{\tau,i}$  and  $T_j^i$ ,  $i = l, d$ ,  $j = SUP, AVE, EXP$ .

We wish to show that

$$\frac{1}{T} \sum_{t=1}^T \left( \hat{u}_t - \frac{1}{T} \sum \hat{u}_t \right)^4 = O_p(1) \quad (46)$$

Write

$$\frac{1}{T} \sum_{t=1}^T \left( \hat{u}_t - \frac{1}{T} \sum \hat{u}_t \right)^4 < c(A_T + 2C_T + C_T) \quad (47)$$

where

$$A_T = \frac{1}{T} \sum_{t=1}^T u_t^4 \quad (48)$$

$$B_T = \frac{1}{T} \sum_{t=1}^T (\hat{u}_t - u_t)^4 \quad (49)$$

$$c_T = \left( \frac{1}{T} \sum_{t=1}^T \hat{u}_t \right)^4 \quad (50)$$

We need to show that  $B_T$  is  $o_p(1)$  and  $A_T, C_T$  are  $O_p(1)$ . The result for  $A_T$  follows by assumption 3. For  $B_T$  we have

$$\frac{1}{T} \sum_{t=1}^T (\hat{u}_t - u_t)^4 = \frac{1}{T} \sum_{t=1}^T \left( (\tilde{\beta} - \beta) x_t \right)^4 \quad (51)$$

The result follows by Lemma 1 and a law of large numbers for the fourth moments of  $x_t$  which follows by assumption 3 and Theorem 19.11 of Davidson (1994). Finally, the result for  $C_T$  follows by a combinations of the results for  $A_T$  and  $B_T$ .

## References

- AIT-SAHALIA, Y., P. J. BICKEL, AND T. M. STOCKER (2001): “Goodness-of-fit Tests for Kernel Regressions with an Application to Option Implied Volatilities,” *Journal of Econometrics*, 105, 363–412.
- ANDREWS, D. W. K., AND W. PLOBERGER (1994): “Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative,” *Econometrica*, 62, 1383–1414.
- ARNOLD, B. C., N. BALAKRISHNAN, AND H. N. NAGARAJA (1992): *A First Course in Order Statistics*. Wiley.
- BERAN, R. (1988): “Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements,” *Journal of the American Statistical Association*, 83(403).
- BOSE, A. (1988): “Edgeworth Correction by Bootstrap in Autoregressions,” *Annals of Statistics*, 16, 1709–1722.
- BROWN, R. L., J. DURBIN, AND J. M. EVANS (1974): “Techniques for Testing the Constancy of Regression Relationships over Time,” *Journal of the Royal Statistical Association, Series A*, 138, 149–163.
- BROWN, W. B. (2000): “Simulation Variance Reduction for Bootstrapping,” in *Simulation-Based Inference: Methods and Applications*, ed. by B. Mariano, T. Schuermann, and M. Weeks. CUP, Cambridge.
- CHOW, A. (1960): “Tests of Equality Between Sets of Coefficients in two Linear Regressions,” *Econometrica*, 28, 591–605.
- COGLEY, T., AND T. SARGENT (2002): “Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US,” Mimeo, Arizona State University.
- DAHLHAUS, R. (1996): “Fitting Time Series Model to Nonstationary Processes,” *Annals of Statistics*, 25, 1–37.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press.
- DAVIES, R. B. (1977): “Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative,” *Biometrika*, 64(2), 247–254.
- FAN, Y. (1995): “Bootstrapping a Consistent Nonparametric Goodness-of-Fit Test,” *Econometric Reviews*, 14, 367–382.

- (1998): “Goodness-of-Fit Tests based on Kernel Density Estimators with Fixed Smoothing Parameters,” *Econometric Theory*, 14, 604–621.
- GALAMBOS, J. (1978): *The Asymptotic Theory of Extreme Order Statistics*. Wiley.
- HALL, P. (1986): “On the Number of Bootstrap Simulations Required to Construct a Confidence Interval,” *The Annals of Statistics*, 14(4).
- (1988): “Theoretical Comparison of Bootstrap Confidence Intervals,” *Annals of Statistics*, 16, 927–953.
- (1992): *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.
- HARTIGAN, J. A. (1986): “Comment on the Paper by Efron and Tibshirani,” *Statistical Science*, 1, 75–76.
- HOROWITZ, J., AND P. HALL (1996): “Bootstrap Critical Values for Tests Based on GMM Estimators,” *Econometrica*, 64, 891–916.
- HOROWITZ, J. L. (1995): “Bootstrap Methods in Econometrics: Theory and Numerical Performance,” manuscript, Department of Economics, University of Iowa.
- KAPETANIOS, G. (2005): “Estimating Deterministically Time Varying Variances in Regression Models,” .
- KAPETANIOS, G., AND E. TZAVALIS (2004): “The Impact of Large Structural Shocks on Economic Relationships: Evidence from Oil Price Shocks,” *Queen Mary, University of London Working Paper no. 524*, 20.
- LAHIRI, S. N. (2003): *Resampling Methods for Dependent Data*. Springer Verlag.
- LIU, J., AND P. J. BROCKWELL (1988): “On the General Bilinear Time Series Model,” *Journal of Applied Probability*, 25, 553–564.
- ORBE, S., E. FERREIRA, AND J. RODRIGUEZ-POO (2005): “Nonparametric Estimation of Time Varying Parameters under Shape Restrictions,” *Journal of Econometrics*, 126, 53–77.
- PLOBERGER, W., AND W. KRAMER (1992): “The CUSUM Test with OLS Residuals,” *Econometrica*, 60, 271–285.

- POURAHMADI, M. (1988): “Stationarity of the Solution of  $X_t = A_t X_{t-1} + \epsilon_t$  and Analysis of Non-Gaussian Dependent Random Variables,” *Journal of Time Series Analysis*, 9, 225–239.
- PRIESTLEY, M. (1965): “Evolutionary Spectra and Nonstationary Processes,” *Journal of Royal Statistical Society, Series B*, 27, 204–237.
- ROBINSON, P. M. (1989): “Nonparametric Estimation of Time Varying Parameters,” in *Statistics, Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl. Springer Berlin.
- (1991): “Time Varying Nonlinear Regression,” in *Statistics, Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl. Springer Berlin.
- SCAILLET, O. (2004): “Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels,” *Journal of Nonparametric Statistics*, 16, 217–226.
- SINGH, K. (1981): “On the Asymptotic Accuracy of Efron’s Bootstrap,” *Annals of Statistics*, 9, 1187–1195.
- TIBSHIRANI, R. J. (1988): “Variance Stabilization and the Bootstrap,” *Biometrika*, 75, 433–444.
- TONG, H. (1990): *Nonlinear time series: A dynamical system approach*. Oxford University Press.

Table 1: Rejection Probabilities under the null hypothesis for  $T^{\tau,l}$  tests at  $T = 200$

$\tau$	h=4	h=6	h=8
0.04	0.688	0.610	0.557
0.08	0.686	0.596	0.549
0.12	0.655	0.574	0.533
0.16	0.629	0.548	0.496
0.20	0.605	0.523	0.462
0.24	0.584	0.493	0.425
0.28	0.556	0.460	0.374
0.32	0.516	0.419	0.324
0.36	0.494	0.367	0.259
0.40	0.486	0.316	0.197
0.44	0.472	0.257	0.144
0.48	0.461	0.223	0.114
0.52	0.469	0.230	0.115
0.56	0.493	0.254	0.143
0.60	0.515	0.302	0.187
0.64	0.538	0.345	0.245
0.68	0.564	0.395	0.303
0.72	0.577	0.448	0.355
0.76	0.599	0.486	0.405
0.80	0.614	0.528	0.444
0.84	0.635	0.551	0.474
0.88	0.647	0.573	0.501
0.92	0.662	0.598	0.524
0.96	0.679	0.617	0.556
1.00	0.665	0.617	0.569

Table 2: Rejection Probabilities under the null hypothesis for  $T_j^l$  and  $T_j^{l*}$ ,  $i = l, d$  tests

	SUP			AVE			EXP		
h/T	100	200	400	100	200	400	100	200	400
	$T_j^l$								
4	0.096	0.268	0.431	0.699	0.882	0.968	0.547	0.792	0.933
6	0.052	0.165	0.251	0.564	0.765	0.906	0.376	0.637	0.822
8	0.015	0.111	0.184	0.465	0.725	0.879	0.296	0.536	0.760
	$T_j^{l*}$								
4	0.048	0.063	0.052	0.047	0.058	0.055	0.048	0.063	0.053
6	0.045	0.059	0.048	0.048	0.061	0.045	0.046	0.058	0.048
8	0.039	0.054	0.061	0.037	0.054	0.064	0.036	0.055	0.060
	$T_j^{d*}$								
4	0.039	0.066	0.053	0.044	0.062	0.054	0.044	0.062	0.054
6	0.052	0.055	0.042	0.045	0.063	0.049	0.045	0.063	0.049
8	0.037	0.058	0.057	0.038	0.056	0.061	0.038	0.056	0.061

Table 3: Rejection Probabilities under the Alternative hypothesis ( $P1$ ) for  $T_j^{l*}$  tests

	SUP			AVE			EXP		
h/T	100	200	400	100	200	400	100	200	400
	$T_j^{l*}$								
4	0.221	0.519	0.919	0.194	0.339	0.712	0.218	0.520	0.918
6	0.201	0.344	0.632	0.211	0.365	0.615	0.204	0.345	0.632
8	0.239	0.391	0.638	0.245	0.421	0.692	0.244	0.394	0.645
	$T_j^{d*}$								
4	0.380	0.798	0.994	0.264	0.635	0.988	0.264	0.635	0.988
6	0.278	0.612	0.936	0.204	0.384	0.730	0.204	0.384	0.730
8	0.269	0.554	0.904	0.222	0.392	0.677	0.222	0.392	0.677

Table 4: Rejection Probabilities under the Alternative hypothesis ( $P2$ ) for  $T_j^{l*}$  tests

	SUP			AVE			EXP		
h/T	100	200	400	100	200	400	100	200	400
	$T_j^{l*}$								
4	0.121	0.301	0.655	0.113	0.213	0.394	0.120	0.301	0.652
6	0.129	0.207	0.392	0.138	0.229	0.371	0.132	0.211	0.391
8	0.130	0.233	0.378	0.131	0.252	0.411	0.131	0.237	0.383
	$T_j^{d*}$								
4	0.181	0.467	0.851	0.140	0.361	0.791	0.140	0.361	0.791
6	0.166	0.358	0.681	0.139	0.240	0.449	0.139	0.240	0.449
8	0.147	0.338	0.598	0.129	0.233	0.398	0.129	0.233	0.398

**This working paper has been produced by  
the Department of Economics at  
Queen Mary, University of London**

**Copyright © 2005 George Kapetanios  
All rights reserved**

**Department of Economics  
Queen Mary, University of London  
Mile End Road  
London E1 4NS  
Tel: +44 (0)20 7882 5096  
Fax: +44 (0)20 8983 3580  
Web: [www.econ.qmul.ac.uk/papers/wp.htm](http://www.econ.qmul.ac.uk/papers/wp.htm)**