Least Squares Estimation and Tests of Breaks in Mean and Variance under Misspecification

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

October 2003

Abstract

In this paper we investigate the consequences of misspecification on the large sample properties of change-point estimators and the validity of tests of the null hypothesis of linearity versus the alternative of a structural break. Specifically this paper concentrates on the interaction of structural breaks in the mean and variance of a time series when either of the two is omitted from the estimation and inference procedures. Our analysis considers the case of a break in mean under omitted regime dependent heteroscedasticity and that of a break in variance under an omitted mean shift. The large and finite sample properties of the resulting least squares based estimators are investigated and the impact of the two types of misspecification on inferences about the presence or absence of a structural break subsequently analyzed.

Keywords: Structural Breaks, Misspecification, Variance shifts, Bootstrapping.

JEL: C22.

[1]Address for Correspondence: Jean-Yves Pitarakis, Department of Economics, University of Southampton, Highfield, Southampton, SO17 1BJ, United-Kingdom. Tel:+44-23-80592631.

1 Introduction

The recent time series literature provides considerable evidence suggesting that many economic time series are characterized by single or multiple shifts in the parameters of their conditional mean (see Stock and Watson (1996, 1999), Bai (1997), Altissimo and Corradi (1999), Hansen (2001) among numerous others). This upsurge in the econometrics of structural change followed the developments in the theory of testing for breaks of unknown location developed in Andrews (1993) and Andrews and Ploberger (1994) (see also Csörgo and Horvath (1997) and references therein for earlier developments in the statistics literature).

Despite the growing applied and theoretical interest in structural change issues it is also the case that most work in this area has operated under the assumption that structural instability may only occur in the mean of a series, restricting higher order moments such as the variance to remain constant. Formal theoretical research that concentrated on the potential presence of breaks in variance includes among others the extension of cusum type tests to detecting shifts in the variance of stationary independent series proposed in Inclan and Tiao (1994), the detection and estimation of multiple breaks in the variance of independent gaussian series investigated in Chen and Gupta (1997) and more recently the work of Wang and Zivot (2000) who proposed a Bayesian approach for the joint treatment of structural instability in both the mean and variance. From an applied perspective the existence of breaks in variance has also attracted considerable interest following the work of McConnell and Perez-Quiros (2000) who documented the existence of a break in US output volatility occurring in the early mid 80s. Building on this line of research, Van Dijk and Sensier (2001) also explored the existence of a break in the volatility of a large database of US macroeconomic series and found that the vast majority of the real series were also characterized by a variance shift that occurred during the early mid 80s (see also Stock and Watson (2002, 2003a, 2003b)).

Despite strong empirical evidence about the presence of breaks in both the mean and variance of economic time series our reading of the literature suggests that those two aspects have been mainly investigated in isolation. Tests for a structural break in the regression parameters for instance are typically conducted under the assumption of a constant error variance while tests of a break in variance (usually implemented as tests of a shift in the mean of the squared residuals sequence obtained from a model fitted via least squares) ignore the possibility of a break in the mean. The same also holds for procedures used to estimate the timing of a shift in either the mean or variance (see Hansen (2001) for an overview of the economics and econometrics of structural change estimation and detection techniques). Note that throughout this paper references to a break in either the mean or variance of a series are also understood to refer to changes in the regression parameters characterizing the conditional mean and error variance respectively.

A natural question that arises therefore is the extent to which the omission of the presence of a break occurring in either the mean or variance of a series distorts inferences about structural break type nonlinearities based on the conventional distributional theory developed in Andrews (1993). Similarly it is also of interest to know how the large sample properties of the corresponding least squares based change-point location estimators are affected when the estimation procedure fails to take into account the possibility that both the mean and variance may shift at the same or different time periods. In both McConnell and Perez-Quiros (2000) and Van Dijk and Sensier (2001) for instance the authors recognized the possibility that inferences about the presence of a break in variance may be contaminated by an omitted break in the mean of the series.

The plan of the paper is as follows. Section 2 focuses on the estimation of the location of a break in mean when the analysis fails to take the presence of regime dependent heteroscedasticity into account and that of the estimation of the location of a break in variance when a potential break in mean is ignored. Section 3 concentrates on the distortions induced by the omission of either a break in mean or variance when testing the null hypothesis of mean or variance linearity against the alternative of a structural break. Section 4 concludes and all proofs are relegated to the appendix.

2 Least Squares Estimation of the Change-Point in Mean or Variance under Misspecification

We consider the following data generating process

$$y_t = x_t'\beta_t + \sigma_t\epsilon_t \quad t = 1, \ldots, T \tag{1}$$

where y_t is the dependent variable, x_t is a $K \times 1$ vector of regressors and β_t the corresponding $K \times 1$ time varying coefficient vector written as $\beta_t = \beta_1 I(t \leq k_1^0) + \beta_2 I(t > k_1^0)$ with $I(.)$ denoting the indicator function and k_1^0 the timing of the break. Writing

$$\sigma_t = \sigma_1 I(t \leq k_2^0) + \sigma_2 I(t > k_2^0) \tag{2}$$

with $\sigma_1 > 0$ and $\sigma_2 > 0$, the disturbance term $\sigma_t \epsilon_t$ is also characterized by a break occuring in its variance at some time k_2^0. The model in (1) is thus characterized by a break in its conditional mean parameters and error variance occurring at times k_1^0 and k_2^0 respectively. For simplicity in what follows we will refer to the break in the $\beta's$ and $\sigma's$ as the break in mean and variance respectively. Throughout this paper we operate under the following set of assumptions

Assumptions

(i) $\epsilon_t|\Omega_t \sim i.i.d.(0,1)$ *with* $E(\epsilon_t)^4 < \infty$,

(ii) $k_1^0 = [T\pi_1^0]$ *and* $k_2^0 = [T\pi_2^0]$ *with* $\pi_i^0 \in \Pi_i$ *and* $\Pi_i = [\underline{\pi}_i, \bar{\pi}_i] \subset (0,1)$ *for i=1,2,*

(iii) $\sup_{\pi \in [0,1]} \left| \frac{\sum_{t=1}^{[T\pi]} x_t \epsilon_t}{T} \right| = o_p(1)$,

(iv) $\sup_{\pi \in [0,1]} \left| \frac{\sum_{t=1}^{[T\pi]} x_t x_t'}{T} - Q(\pi) \right| = o_p(1)$ *with* $Q(\pi)$ *denoting a positive definite matrix, absolutely continuous and monotonically increasing function of* π,

where Ω_t denotes the information set at time t and [.] is the greatest integer function. Assumptions (i) and (ii) are standard in the change-point literature (see Bai (1997), Chong (2001)). Assumption (ii) for instance requires the true change-point not to be located in the extreme top or bottom of the sample. This ensures that there are enough observations at both ends of the sample so as to ensure the identifiability of regime specific parameters (see Andrews (1993)). In applied work it is common practice for instance to set the lower bound $\underline{\pi}_1$ to 5% or 10% and use $\bar{\pi}_1 = 1 - \underline{\pi}_1$. Assumptions (iii) and (iv) are high-level and correspond to unifom law of large number type requirements. They allow for a wide range of specifications for the conditional mean equation, including stationary autoregressive processes while trending and integrated regressors are ruled out.

At this stage it is also important to emphasise the limiting behaviour of the partial sum of the second moments of the regressors as represented by the matrix functional $Q(\pi)$ in (iv). Within the specification given by (1)-(2) and unless the regressors are assumed to be exogenous or the size of the shift in the $\beta's$ and $\sigma's$ is assumed to converge to zero with T, $Q(\pi)$ will not be a linear functional of π. The matrix functional $Q(\pi)$ will typically be expressed as a nonlinear function of π with kinks occurring at $\pi = \pi_1^0$ and $\pi = \pi_2^0$. Under $\pi_1^0 = \pi_2^0 \equiv \pi^0$ for instance we will let

$$Q(\pi) \quad = \quad \pi Q_1 I(\pi \leq \pi^0) + [\pi^0 Q_1 + (\pi - \pi^0) Q_2] I(\pi > \pi^0) \tag{3}$$

with $Q_1 = E(x_t x_t')$ for $t \leq k^0$, $Q_2 = E(x_t x_t')$ for $t > k^0$. Under $x_t = y_{t-1}$ for instance we will have

$Q_1 = \sigma_1^2/(1-\beta_1^2)$ and $Q_2 = \sigma_2^2/(1-\beta_2^2)$. This also highlights the fact that restricting $Q(\pi)$ to be linear in π as say when $Q(\pi) = \pi Q$ for some $Q \succ 0 \; \forall t$ would lead to a more restrictive framework for the possible choice of the regressors in (1). Our formulation of $Q(\pi)$ also extends naturally to the case where the breaks in mean and variance do not occur at the same time. Under $k_1^0 < k_2^0$ for instance and letting $Q_1 = E(x_t x_t')$ for $t \leq k_1^0$, $Q_2 = E(x_t' x_t)$ for $k_1^0 < t \leq k_2^0$ and $Q_3 = E(x_t x_t')$ for $t > k_2^0$ we write

$$\begin{aligned} Q(\pi) &= \pi Q_1 I(\pi \leq \pi_1^0) + [\pi_1^0 Q_1 + (\pi - \pi_1^0) Q_2] I(\pi_1 < \pi \leq \pi_2^0) \\ &+ [\pi_1^0 Q_1 + (\pi_2^0 - \pi_1^0) Q_2 + (\pi - \pi_2^0) Q_3] I(\pi > \pi_2^0). \end{aligned} \tag{4}$$

Throughout this paper our analyis will explicitly distinguish between the two scenarios described in (3) and (4) together with the case where $Q(\pi) = \pi Q$. The formulation in (4) also extends naturally to the case where $k_2^0 < k_1^0$.

2.1 Estimation of the Change-Point in the Mean under an Omitted Variance Shift

Our initial objective is to evaluate the properties of $\hat{k}_1$ an estimator of the location of the change-point k_1^0 in the slope parameters and intercept when the regime dependent heterogeneous structure of the error process is ignored in the fitted specification. The estimator of the change-point is obtained as the minimizer of the (misspecified) concentrated sum of squared errors function

$$S_{1T}(k) = \sum_{t=1}^{k} (y_t - x_t' \hat{\beta}_1(k))^2 + \sum_{t=k+1}^{T} (y_t - x_t' \hat{\beta}_2(k))^2 \tag{5}$$

where $\hat{\beta}_1(k)$ and $\hat{\beta}_2(k)$ denote the least squares estimators of the slope parameters within each regime for given k. Alternatively and for greater technical convenience we can also reformulate $\hat{k}_1$ as $\hat{k}_1 = \arg\max_k G_{1T}(k)$ with $G_{1T}(k) = S_T - S_{1T}(k)$ and $S_T = \sum_{t=1}^{T} (y_t - x_t' \hat{\beta})^2$ denoting the full sample sum of squared errors. In Bai and Perron (1998) the authors established the weak consistency of $\hat{\pi}_1 = \hat{k}_1/T$ under a set of assumptions which allow for the error variances to also shift across regimes. Although not explicitly stated the weak consistency result $\hat{\pi}_1^0 \xrightarrow{p} \pi_1^0$ continues to hold if the break in the conditional mean parameters and that in the variance do not occur at the same time. Despite this desirable limiting property of $\hat{\pi}_1$, it is difficult to analytically quantify the loss of efficiency that arises from the omission of the break in variance since the limiting distribution of change-point estimators is not easily tractable under fixed shift magnitudes and is not invariant to numerous model specific features such as the distribution of the $\epsilon_t' s$ (see Hinkley (1970)). Instead the common approach in the literature has been to proceed within a small shift asymptotic framework, allowing the magnitude of the jump to converge to zero at a prespecified rate.

Being aware of the finite sample properties of $\hat{\pi}_1$ is important not only because of the direct economic implications that the accurate dating of a structural break in the mean may entail but also for the subsequent analysis such as the search for further breaks in the variance. In this latter case for instance obtaining residuals that are not contaminated by an omitted break in the conditional mean is crucial for properly locating the potential presence of a break in volatility (see Stock and Watson (2002)).

Here we initially aim to highlight the important distortions that characterise $\hat{\pi}_1$ even in moderately large sample sizes, when the data generating process is as in (1)-(2). Specifically we consider an autoregressive process of order one given by $y_t = (\beta_0^{(1)} + \beta_1^{(1)} y_{t-1}) I(t \leq k_1^0) + (\beta_0^{(2)} + \beta_1^{(2)} y_{t-1}) I(t > k_1^0) + \sigma_t \epsilon_t$ and where $\sigma_t = \sigma_1 I(t \leq k_2^0) + \sigma_2 I(t > k_2^0)$. In McConnell and Perez-Quiros (2000) for instance the authors have advocated such a specification for the growth rate in US GDP. Our experiments are conducted across $N = 5000$ replications using $T \in \{250, 500, 1000\}$ and $(\sigma_1, \sigma_2) \in \{(1, 1), (1, 2), (2, 1)\}$ thus covering both increases and decreases in the error variances together with the constant variance benchmark case. Regarding the positioning of the structural breaks we consider two scenarios. One that imposes the break date to be the same, setting $\pi_1^0 = \pi_2^0 \in \{0.25, 0.50, 0.75\}$ and one that allows the breaks to occur at different periods with $(\pi_1^0, \pi_2^0) \in \{(0.25, 0.75), (0.75, 0.25)\}$. Throughout all our experiments the random error term ϵ_t is taken to be a standard normally distributed random variable. For the choice of the autoregressive parameters we set $\beta_0^{(1)} = 1$, $\beta_0^{(2)} = 2$, $\beta_1^{(1)} = 0.4$ and $\beta_1^{(2)} = 0.1$. The latter correspond to an unconditional mean equal to 1.67 for the first regime and 2.22 for the second regime.

Results are presented in Table 1 which displays the empirical means and standard deviations of the change-point estimator $\hat{\pi}_1$. For the correctly specified fitted model corresponding to the DGPs with $\sigma_1 = \sigma_2 = 1$ we can observe that the empirical biases of $\hat{\pi}_1$ are small, virtually vanishing for $T \geq 500$. The corresponding empirical standard deviations also decline with T reaching a common magnitude of approximately 0.024 across all true locations of the break dates. The picture is different however when the true models are characterized by a break in variance and the latter is ignored during the estimation stage. Despite the clear progression of the estimator $\hat{\pi}_1$ towards its true value π_1^0 as T grows we can observe important finite sample biases even for sample sizes as large as T=500. The direction of the biases depends jointly on whether the underlying error variance increased or decreased from one regime to the other and on the location of the true break point. Under the common break date scenarios, the changepoint estimator displays a tendency to overestimate the location of its true counterpart when $\sigma_1 < \sigma_2$ and to underestimate it when $\sigma_1 > \sigma_2$. The most pronounced negative bias occurs when the true break point occurs towards the end of the sample (i.e. under $\pi_1^0 = 0.75$) while the greatest positive bias occurs when the true break point locates at the beginning of the sample (i.e. $\pi_1^0 = 0.25$). In this latter case and under T=250 the bias might translate into mislocating the timing of the structural break by more than a decade when the data are quarterly for instance. We also note that the magnitudes of the biases are affected by whether the breaks in mean and variance occur at the same time compared with scenarios where one precedes or succedes the other.

Table 1 about here

Regarding the behaviour of the empirical standard deviations of $\hat{\pi}_1$ we can also note that for both small and large sample sizes they are substantially higher than their counterparts under $\sigma_1 = \sigma_2$. These differences in the magnitudes of the standard deviations persist as T is allowed to grow. Regardless of the true locations of the break points they are always substantially higher than their i.i.d errors based counterparts even under T=1000. This also reflects the fact that the limiting distribution of $\hat{\pi}_1$ is different from that obtained under the i.i.d errors case. Overall, omitting the presence of a break in the error variance induces an important increase in the variability of $\hat{\pi}_1$. The magnitude of the effect appears to depend jointly on the location of the true break points and on whether the error variance increased or decreased following the occurence of the break.

2.2 Estimator of the Change-Point in the Variance under an Omitted Mean-Shift

We next focus on the estimation of the change-point in the error variance when the potential presence of a break in mean is ignored. A common approach for estimating k_2^0 involves treating the break in variance problem as a break in the mean of a relevant squared sequence (see Csörgo and Horváth (1997), Hansen (2001)). This is justified by the fact that under a correctly specified conditional mean equation the ensuing residuals are consistent for the true errors. It is also important to stress that our key interest here is the estimation of the location of the change-point rather than that of the various parameters characterizing each regime.

When ignoring the presence of the shift in the slopes the squared residuals are given by $z_t = (y_t - x_t'\hat{\beta})^2$. The estimator of the change-point in the variance is then defined as

$$\hat{k}_2 = \arg\max_k G_{2T}(k) \tag{6}$$

with

$$G_{2T}(k) \quad = \quad \frac{k(T-k)}{T}(\bar{z}_2 - \bar{z}_1)^2 \tag{7}$$

and where $\bar{z}_1 = \sum_{t=1}^{k} z_t/k$ and $\bar{z}_2 = \sum_{t=k+1}^{T} z_t/(T-k)$. Note that the objective function in (7) is equivalent to one that would arise if we were to estimate the location of a *mean shift* in the scalar sequence z_t. Under the present scenario where z_t absorbs the omitted shift in the $\beta's$ that occurs at time k_1^0 the limiting behaviour of $\hat{\pi}_2$ will crucially depend on whether or not the breaks in mean and variance occurred at the sime time (i.e. whether $k_1^0 = k_2^0$ or $k_1^0 \neq k_2^0$) and on the location of k_1^0. Under $k_1^0 \neq k_2^0$ for instance we can intuitively expect that the z_t sequence will be characterised by two structual breaks in its mean locating at k_1^0 and k_2^0 respectively. Thus by proceeding as in (6)-(7) to obtain the timing of the break in z_t our analysis will be omitting the presence of a second underlying break. It is then unclear whether the resulting break date estimator will be consistent for its true counterpart or that that has been ignored.

2.2.1 Common Break Dates

We initially concentrate on the case where the breaks in mean and variance occurred at the same time, setting $k_1^0 = k_2^0 \equiv k^0$ in model (1) and first focus on the limiting behaviour of a normalised version of the objective function in (7). This preliminary result is summarised in the following lemma.

Lemma 1 *Under assumptions (i)-(iv) with $Q(\pi)$ as in (3) we have*

$$\sup_{\pi \in [0,1]} \left| \frac{G_{2T}([T\pi])}{T} - G_{2\infty}(\pi) \right| \quad \xrightarrow{p} \quad 0$$

as $T \to \infty$, where $G_{2\infty}(\pi)$ is a nonstochastic continuous function of π given by

$$\begin{aligned} G_{2\infty}(\pi) \quad &= \quad \left[\frac{\pi(1-\pi^0)^2}{1-\pi} + \left(\frac{(\pi^0)^2(1-\pi)^2 - \pi^2(1-\pi^0)^2}{\pi(1-\pi)} \right) I(\pi > \pi^0) \right] \\ &\times \quad \left[\Delta + (\pi^0)^2 \lambda' Q_1 M^{-1} Q_2 M^{-1} Q_1 \lambda - (1-\pi^0)^2 \lambda' Q_2 M^{-1} Q_1 M^{-1} Q_2 \lambda \right]^2 \end{aligned} \tag{8}$$

with $\Delta = (\sigma_2^2 - \sigma_1^2)$, $\lambda = (\beta_2 - \beta_1)$, and $M = [\pi^0 Q_1 + (1-\pi^0) Q_2] \equiv Q(1)$.

The above lemma summarises the uniform probability limit of the objective function in (7) which will play a key role in the evaluation of the asymptotic properties of $\hat{\pi}_2$. From the expression in (8) we also note that under a framework where $Q(\pi) = \pi Q$ in assumption (iv), $G_{2\infty}(\pi)$ is given by

$$\begin{aligned} G_{2\infty}(\pi) \quad &= \quad \left[\frac{\pi(1-\pi^0)^2}{1-\pi} + \left(\frac{(\pi^0)^2(1-\pi)^2 - \pi^2(1-\pi^0)^2}{\pi(1-\pi)} \right) I(\pi > \pi^0) \right] \\ &\times \quad \left[\Delta - (1-2\pi^0) \lambda' Q \lambda \right]^2 . \end{aligned} \tag{9}$$

The next proposition presents the limiting properties of $\hat{\pi}_2 = \hat{k}/T$ when estimated as in (6)-(7).

Proposition 2 *Under assumptions (i)-(iv) and letting* $k_1^0 = k_2^0 \equiv k^0$ *in model (1) we have* $\hat{\pi}_2 \xrightarrow{p} \pi^0$ *as* $T \rightarrow \infty$.

According to the above proposition omitting the presence of a break in the $\beta' s$ does not affect the consistency of the change-point estimator in the variance when the latter is estimated using the squared residuals of the fitted mispecified model and provided that the omitted break in mean occurs at the same time as that characterizing the error variance.

To further explore the properties of the estimator described in proposition 2 we reconsider the same Monte-Carlo exercise as in Table 1 but here we wrongly fit a linear AR(1) model to the data and consider the sampling properties of the resulting change-point estimator obtained from the (misspecified) squared residuals. Results are presented in Table 2.

Table 2 about here

For small sample sizes we can note that the biases in the estimates of $\hat{\pi}_2$ are positive under $\sigma_1 < \sigma_2$ and negative when $\sigma_1 > \sigma_2$. We also note that in absolute terms finite sample biases are smaller when the omitted break in mean occurs in the middle of the sample with $\pi^0 = 0.5$. This latter point also holds true for the corresponding empirical standard deviations of $\hat{\pi}_2$ suggesting that the location of the break in variance is more accurately estimated when $\pi^0 = 0.5$. Overall however we observe a clear decline in the empirical biases of $\hat{\pi}_2$ as the sample size increases. Under all scenarios for π^0 for instance the biases are virtually zero under $T = 1000$. Similarly all empirical standard deviations are seen to have stabilised around a common value of approximately 0.012 across all parameter configurations. Comparing the empirical standard deviations presented in Table 1 and Table 2 it is also interesting to note that the same break location is estimated much more accurately when estimating it as a break in variance (omitting the break in mean) rather than as a break in mean (omitting the break in variance).

2.2.2 Distinct Break Dates

Before proceeding with the limiting behaviour of $\hat{\pi}_2$ under the case where the break in mean and the break in variance do not occur at the same time (i.e. $k_1^0 \neq k_2^0$) we initially concentrate on the corresponding limiting behaviour of the normalised version of the objective function $G_{2T}(k)$ in (7), summarised in the following lemma

Lemma 2 *Under assumption (i)-(iv) with* $k_1^0 < k_2^0$, $Q(\pi)$ *as in (4) we have*

$$\sup_{\pi\in[0,1]} \left| \frac{G_{2T}([T\pi])}{T} - G_{2\infty}(\pi) \right| \overset{p}{\rightarrow} 0$$

as $T \rightarrow \infty$, *where* $G_{2\infty}(\pi)$ *is a nonstochastic continuous function of* π *given by*

$$\begin{aligned} G_{2\infty}(\pi) &= \frac{\pi}{1-\pi}\left[(1-\pi_2^0)\Delta + (\pi_1^0)^2\lambda' C_1 \lambda - (1-\pi_1^0)\lambda' C_2 \lambda\right]^2 \;\; for \; \pi \leq \pi_1^0 \\ &= \pi(1-\pi)\left[\frac{1-\pi_2^0}{1-\pi}\Delta + \frac{(\pi_1^0)^2}{1-\pi}\lambda' C_1\lambda - \frac{\pi_1^0}{\pi}\lambda' C_2 \lambda - \frac{(\pi_1^0)^2(\pi-\pi_1^0)}{\pi(1-\pi)}\lambda' C_3 \lambda\right]^2 \;\; for \; \pi_1^0 < \pi < \pi_2^0 \\ &= \frac{1-\pi}{\pi}\left[\pi_2^0\Delta - (\pi_1^0)^2\lambda' C_1\lambda - (\pi_1^0)\lambda' C_2\lambda + (\pi_1^0)^2\lambda' C_4 \lambda\right]^2 \;\; for \; \pi > \pi_2^0, \end{aligned} \tag{10}$$

where $C_1 = Q_1 M^{-1} H M^{-1} Q_1$, $C_2 = H M^{-1} Q_1 M^{-1} H$, $C_3 = Q_1 M^{-1} Q_2 M^{-1} Q_1$, $C_4 = Q_1 M^{-1} Q_3 M^{-1} Q_1$ *with* $M = \pi_1^0 Q_1 + (\pi_2^0 - \pi_1^0) Q_2 + (1-\pi_2^0) Q_3 \equiv Q(1)$ $H = [(\pi_2^0 - \pi_1^0) Q_2 + (1-\pi_2^0) Q_3]$.

At this stage it is also useful to specialise $G_{2\infty}(\pi)$ in (10) to the case where $Q(\pi) = \pi Q$. Under this scenario we have $C_1 = (1-\pi_1^0)Q$, $C_2 = (1-\pi_1^0)^2 Q$ and $C_3 = C_4 = Q$ leading to

$$\begin{aligned} G_{2\infty}(\pi) &= \frac{\pi}{1-\pi}\left[(1-\pi_2^0)\Delta - (1-\pi_1^0)(1-2\pi_1^0)\lambda' Q \lambda\right]^2 \;\; for \; \pi \leq \pi_1^0 \\ &= \pi(1-\pi)\left[\frac{1-\pi_2^0}{1-\pi}\Delta - \frac{\pi_1^0}{\pi}(1-2\pi_1^0)\lambda' Q\lambda\right]^2 \;\; for \; \pi_1^0 < \pi < \pi_2^0 \\ &= \frac{1-\pi}{\pi}\left[\pi_2^0\Delta - \pi_1^0(1-2\pi_1^0)\lambda' Q\lambda\right]^2 \;\; for \; \pi \geq \pi_2^0. \end{aligned} \tag{11}$$

The following proposition presents the limiting properties of $\hat{\pi}_2$ corresponding to the case where $Q(\pi) = \pi Q$.

Proposition 3 *Under assumptions (i)-(iv) with* $Q(\pi) = \pi Q$, $\pi_1^0 < \pi_2^0$ *and as* $T \rightarrow \infty$ *we have (a)* $\hat{\pi}_2 \overset{p}{\rightarrow} \pi_2^0$ *if* $\pi_1^0 = \frac{1}{2}$, *(b)* $\hat{\pi}_2 \overset{p}{\rightarrow} \pi_2^0$ *if* $\pi_1^0 \neq \frac{1}{2}$ *and* $(\Delta/\lambda' Q\lambda)^2 > (1-2\pi_1^0)^2 \pi_1^0(1-\pi_1^0)/\pi_2^0(1-\pi_2^0)$, *(c)* $\hat{\pi}_2 \overset{p}{\rightarrow} \pi_1^0$ *if* $\pi_1^0 \neq \frac{1}{2}$ *and* $(\Delta/\lambda' Q\lambda)^2 < (1-2\pi_1^0)^2 \pi_1^0(1-\pi_1^0)/\pi_2^0(1-\pi_2^0)$.

According to the above proposition omitting the presence of a break in the $\beta' s$ may adversely affect the limiting properties of the resulting change-point in the variance $\hat{\pi}_2$ in the sense that the latter may no longer be consistent for its true counterpart π_2^0 but converge to the location of the omitted break in the $\beta' s$ instead. This will not happen however if the omitted break in the regression parameters occurs in the middle of the sample with $\pi_1^0 = \frac{1}{2}$.

To shed further light on the specific condition on the parameters required under scenarios (b) and (c) of proposition 3 we can focus on a simple mean-variance shift model, setting $x_t = 1$ in (1) and for which $\lambda' Q\lambda \equiv (\beta_2 - \beta_1)^2$. From proposition 3 in order for $\hat{\pi}_2$ to be consistent for π_1^0 instead of π_2^0 the parameter configuration must be such that $\Delta^2 < (\beta_2 - \beta_1)^4 (1-2\pi_1^0)^2 \pi_1^0 (1-\pi_1^0)/\pi_2^0(1-\pi_2^0)$. It is then straightforward to note that this outcome will only occur if the size of the jump in the means is substantially higher than that characterising the variances. This can be illustrated via a small Monte-Carlo experiment in which we chose to set $(\sigma_1, \sigma_2) = (1, 1.5)$ and $(\beta_1, \beta_2) = (1, 3)$ with $\pi_1^0 = 0.25$ and $\pi_2^0 = 0.35$. Using samples of size $T = 1000$ across $N = 5000$ replications led to an empirical mean of 0.24 for $\hat{\pi}_2$, clearly illustrating the fact that omitting the presence of a break in mean may lead us to wrongly view it as a break in variance.

We next extend our results to a more general framework allowing $Q(\pi)$ to be as in (4). Equivalently we are now interested in the limiting behaviour of $\hat{\pi}_2$ when $G_{2\infty}(\pi)$ is given by (10) rather than (11). This scenario is relevant for instance if the DGP contains lagged dependent variables as in an autoregressive specification. Using (10) we initially introduce a condition on the parameters ensuring that $G_{2\infty}(\pi_2^0) > G_{2\infty}(\pi_1^0)$

$$\left[\Delta - (\pi_1^0)^2 \lambda'(C_3 - C_4)\lambda\right]^2 \;>\; \frac{\pi_1^0(1-\pi_1^0)}{\pi_2^0(1-\pi_2^0)} \left[\lambda' C_2 \lambda - (\pi_1^0)^2 \lambda' C_3 \lambda\right]^2 \tag{12}$$

and the following proposition summarises the limiting behaviour of $\hat{\pi}_2$ under this setting.

Proposition 4 *Under assumptions (i)-(iv) with $\pi_1^0 < \pi_2^0$ and as $T \to \infty$ we have $\hat{\pi}_2 \xrightarrow{p} \pi_2^0$ if $G_{2\infty}(\pi_2^0) > G_{2\infty}(\pi_1^0)$ and $\hat{\pi}_2 \xrightarrow{p} \pi_1^0$ if $G_{2\infty}(\pi_1^0) > G_{2\infty}(\pi_2^0)$.*

At this stage it is also interesting to relate the above results to the analysis in Bai (1997) where the author investigated the limiting properties of a single change-point location estimator obtained as the minimiser of $S_{1T}(k)$ with $x_t = 1$ when the underlying series is characterised by more than one break. An important result established in that paper is that the change-point location estimator obtained from a model that ignored the presence of multiple breaks will remain consistent for one of the true break points. Since the omission of a break in mean (assumed to occur at a different time than the break in variance) will translate into a break in the squared residuals sequence z_t our results in Propositions 3 and 4 can also be interpreted along the same lines as the analysis of Bai (1997). The condition presented in (12) is then equivalent to requiring that one of the two breakpoints dominates in the sample in the sense of contributing the most to the maximisation of the objective function. Although the results presented in Lemma 2 and the consistency properties that followed in propositions 3 and 4 have been established for the case where the break in mean is assumed to occur prior to that characterising the error variances it is a simple algebraic exercise to reformulate the same results for the case where $\pi_1^0 > \pi_2^0$ via an appropriate reparameterization of (4) and details are therefore omitted.

To further illustrate the empirical properties of $\hat{\pi}_2$ when the break in mean and variance occur at different time periods we conducted a simulation experiment using a DGP similar to that considered in Table 2 but in which we allowed π_1^0 and π_2^0 to be such that $(\pi_1^0, \pi_2^0) = \{(0.25, 0.75), (0.75, 0.25)\}$. Within the first configuration the break in mean precedes that in the variance while in the second configuration the break variance occurs before that in the mean. Results for this experiment are presented in Table 3. Note that the chosen parameter configuration is such that in the context of proposition 3 we have $G_{2\infty}(\pi_2^0) > G_{2\infty}(\pi_1^0)$.

Table 3 about here

Across both panels of Table 3 we can clearly observe that despite the omitted break in mean that occured either before or after the break in variance the estimator of the break fraction $\hat{\pi}_2$ converges to its true counterpart as T is allowed to increase.

3 Testing for a Structural Break under Misspecification

Our previous analysis has dealt with the estimation of the timing of a structural break in either the slope parameters or error variance within a misspecified framework that ignored the possibility that both may shift at some unknown time periods denoted k_1^0 and k_2^0 respectively. We found that omitting the presence of a break in either the conditional mean parameters or error variance may have important adverse effects on the limiting and finite sample properties of the estimated break locations. At the same time we also highlighted the occurence of a wide range of scenarios for which the omission of either of the breaks does not affect the consistency of estimator obtained from a misspecified model. Our next objective is to evaluate the influence of the same type of misspecification when testing the null hypotheses given by $H_0 : \beta_1 = \beta_2$ and $H_0 : \sigma_1^2 = \sigma_2^2$ respectively.

3.1 Testing for a Structural Break in Regression Slopes

We are initially interested in the consequences of wrongly imposing the homogeneity of error variance (i.e. $\sigma_1 = \sigma_2 = \sigma$ say) when testing the null hypothesis given by $H_0 : \beta_1 = \beta_2$ in (1). The commonly used test statistics for conducting such inferences are the supremum, average and exponential functionals of the Wald, LM or LR statistics proposed in Andrews (1993) and Andrews and Ploberger (1994). The limiting distributions of these test statistics are nonstandard and are given by corresponding functionals of normalized squared brownian bridge processes (see also Hansen (1997) for an overview of their limiting properties and practical implementation using simulation based approximations to p-values).

Within our present context we can express the LM test statistic as $LM_T(\pi_1) = G_{1T}([T\pi_1])/\hat{\sigma}_u^2$ with $\hat{\sigma}_u^2 = \sum_{t=1}^T \hat{u}_t^2/T$ and $\hat{u}_t = y_t - x_t'\hat{\beta}$. Note that here we are operating under the null hypothesis of no change in regression parameters, setting $\beta_1 = \beta_2 = \beta \ \forall t$. As shown in Andrews (1993) and Andrews and Ploberger (1994) under a wide range of regularity conditions and assuming that $\sigma_1 = \sigma_2$ the asymptotic null distribution of the supremum version of $LM_T(\pi)$ is given by

$$\sup_{\pi \in \Pi_1} LM_T(\pi) \quad \stackrel{d}{\to} \quad \sup_{\pi \in \Pi_1} \frac{[W(\pi) - \pi W(1)]'[W(\pi) - \pi W(1)]}{\pi(1-\pi)} \tag{13}$$

where $W(.)$ denotes a K dimensional standard Brownian Motion.

If we proceed as above when evaluating the test statistic but ignore the fact that $\sigma_1 \neq \sigma_2$, it is straightforward to establish that although the right hand side of (13) remains stochastically bounded it is no longer given by the normalized quadratic form of a Brownian Bridge process. For the clarity of the exposition this latter point can be illustrated using a simple variance shift framework with a constant term as the sole regressor, setting $x_t = 1$ in (1). The key point affecting the distributional results is the fact that under the behaviour of u_t in (1) with $x_t = 1$, $\sum_{t=1}^{[T\pi]} u_t^2/T$ no longer converges to a functional that is linear in π. More specifically the limit functional is given by $\pi\sigma_1^2 + (\sigma_2^2 - \sigma_1^2)(\pi - \pi_2^0)I(\pi > \pi_2^0)$ with a kink at $\pi = \pi_2^0$. Under this scenario, the limiting behaviour of $LM_T(\pi)$ is given by

$$\begin{aligned} LM_T(\pi) \quad &\stackrel{d}{\to} \quad \frac{1}{\pi(1-\pi)\zeta}\left[\sigma_1(W(\pi) - \pi W(1)) - (\sigma_2 - \sigma_1)\pi(W(1) - W(\pi_2^0))\right]^2 I(\pi \leq \pi_2^0) \\ &+ \quad \frac{1}{\pi(1-\pi)\zeta}\left[\sigma_2(W(\pi) - \pi W(1)) - (\sigma_2 - \sigma_1)(1-\pi)W(\pi_2^0)\right]^2 I(\pi > \pi_2^0) \end{aligned} \tag{14}$$

with $\zeta = \sigma_1^2\pi_2^0 + \sigma_2^2(1-\pi_2^0)$. We can note from (14) that the correct limiting distribution of the supremum of the LM statistic (or any other functional such as the average or exponential) will now depend on the location of the ignored break fraction π_2^0 that characterizes the variance of the error process together with the magnitudes of σ_1 and σ_2. More importantly this highlights the fact that basing inferences on the asymptotic p-values tabulated in Hansen (1997) will result in misleading conclusions. To further explore this latter point we conducted a simulation experiment designed to evaluate the magnitudes and direction of the distortions that may arise when facing this misspecification scenario.

Table 4 presents the empirical sizes of the SupLM based tests when the fitted model ignores the presence of regime dependent heteroscedasticity. Specifically we used the iid errors based critical values while the process that generated the data although satisfying the null given by $H_0 : \beta_1 = \beta_2$ is characterized by an error process of the form $u_t = \sigma_t \epsilon_t$ with σ_t as in (2). For comparison purposes the first column of Table 4 also presents the corresponding figures under a DGP with iid errors, imposing $\sigma_1 = \sigma_2 = 1$. All experiments have been conducted across $N = 5000$ replications.

Table 4 about here

Focusing first on the case where the tests are evaluated within a correctly specified framework we can note that the test is slightly undersized in finite samples but clearly tends towards the correct nominal sizes of 2.5% and 5% as T is allowed to grow (see Diebold and Chen (1997) for a comprehensive Monte-Carlo evaluation of the finite sample size properties of change-point tests). Note that since our tests are conducted using p-values obtained via simulations rather than exact critical values we may not expect an exact match between nominal and empirical sizes even under very large sample sizes.

The remaining figures presented in Table 4 clearly highlight the important distortions that may arise in practice when the regime dependent nature of the error variance is ignored. Under all scenarios the empirically obtained sizes are greater than their nominal counterparts clearly suggesting that the direction of the distortions is towards a spurious detection of a break-point in the regression parameters. Equivalently, failing to take the shift in error variances into account translates into frequent spurious detections of a break in β. The magnitudes of the size distortions depend *jointly* on the location of the break in variance captured by π_2^0 and on whether the variance increased or decreased following the occurence of the break. When the omitted break in variance locates towards the bottom of the sample (e.g. when $\pi_2^0 = 0.75$) the greatest size distortions occur under $\sigma_1 < \sigma_2$ and vice-versa when the omitted break occurs at the top of the sample (e.g. when $\pi_2^0 = 0.25$). Unanimously however we can note that the direction of the distortions is towards spurious rejections of the null hypothesis of no structural break in the regression parameters. The key practical implication of the above findings is that one should interpret inferences about the presence of a structural break in mean very cautiously since the latter can be seriously contaminated by an underlying shift in variance and the finding of a break in mean may in fact be due to an underlying break in the variance of the process under study. To our knowledge most recent applied work in this area ignored this potential source of misspecification.

It is a notoriously difficult problem to design good test procedures about the equality of regression slopes while not necessarily maintaining the equality of variances assumption. One possible amendment to the test procedure evaluated above is to proceed with a traditional GLS type transformation of the original model. Suppose for instance that σ_1, σ_2 and k_2^0 are all known parameters. Since the model under the null is here given by $y_t = x_t'\beta + \sigma_t\epsilon_t$ we can define $\tilde{y}_t = y_t/\sigma_t$ and $\tilde{x}_t = x_t/\sigma_t$ and rewrite the model as $\tilde{y}_t = \tilde{x}_t'\beta + \epsilon_t$. Although the new error process is now i.i.d. an important feature still present in this GLS-transformed model is the nonstationarity of the regressor matrix $\tilde{x}_t$ which violates the regularity conditions ensuring a limiting behaviour as in the right hand side of (13). Further insight into this latter point can be obtained by evaluating the limiting behaviour of the partial sums of the sample moment matrix. Using straightforward algebra and setting $x_t = 1$ for notational simplicity we have

$$\frac{1}{T}\sum_{t=1}^{[T\pi]} \tilde{x}_t \tilde{x}_t' \quad \Rightarrow \quad \frac{\pi}{\sigma_1^2} - \left[\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right](\pi - \pi_2^0) I(\pi > \pi_2^0). \tag{15}$$

Note that when $\sigma_1^2 = \sigma_2^2$, (15) is linear in π whereas for $\sigma_1^2 \neq \sigma_2^2$ it is characterised by a kink occurring at $\pi = \pi_2^0$. The above arguments illustrate that the null limiting distribution of the LM statistic for testing $H_0 : \beta_1 = \beta_2$ in the GLS transformed model will be affected by the structural break in $\tilde{x}_t$.

Here we initially explore an alternative approach for testing the null $H_0 : \beta_1 = \beta_2$ by proceeding via a fixed regressor bootstrap (see Hansen (2000)) implemented on the GLS transformed model. We let $\hat{\sigma}_t = \hat{\sigma}_1 I(t \leq \hat{k}_2) + \hat{\sigma}_2 I(t > \hat{k}_2)$ where $\hat{\sigma}_1^2 = \sum_{t=1}^{\hat{k}_2}(y_t - x_t'\hat{\beta})^2/\hat{k}_2$, $\hat{\sigma}_2^2 = \sum_{t=\hat{k}_2+1}^{T}(y_t - x_t'\hat{\beta})^2/(T-\hat{k}_2)$ and $\hat{k}_2$ is obtained as in (6)-(7). The fixed regressor bootstrap has been proposed in Hansen (2000) for testing the null hypothesis of no structural change in regression slopes when the regressors may themselves be characterised by a break and is therefore appropriate for our specific framework. To evaluate its properties in the context of the above GLS transformed specification we implement the procedure on the same model as that considered in Table 4. Specifically the DGP is given by a linear AR(1) process with an error variance characterised by a break occurring at time k_2^0. Results are displayed in Table 5 in which we present the empirical size of the bootstrap based test when implemented on both the GLS transformed and untransformed models. Here our simulation based results have been obtained using 1000 bootstrap draws across $N = 5000$ Monte-Carlo replications. For comparison purposes the empirical sizes based on Andrews's limiting distribution (using Hansen's (1997) asymptotic p-value approximations) are also presented.

Table 5 about here

Focusing first on the homoscedastic version of the bootstrap within the GLS transformed model we can note substantial improvements to the size properties of the test with the bootstrap based empirical sizes remaining very close to their nominal counterparts of 2.5% and 5% for both large and moderately small sample sizes. Comparing the figures of Table 5 with Table 4 under $\pi_2^0 = 0.75$ and

$(\sigma_1, \sigma_2) = (1, 2)$ for instance we can note that a strongly inflated empirical size of 19.24% for T=250 has an homoscedastic bootstrap based counterpart of 5.60%. Similar improvements can also be noted throughout all other parameter configurations. It is also interesting to note that although theoretically inappropriate inferences based on the standard limiting distributions using Hansen's (1997) p-values also lead to notable improvements to the size properties of the test when implemented on the GLS transformed model. In the first column of Table 5 we also present empirical sizes corresponding to a scenario where the GLS transformation is spurious in the sense that the underlying DGP is not characterised by any jump in variances.

From the figures presented in Table 5, it is also important to note the inappropriateness of the bootstrap when applied to the untransformed model, although its heteroscedastic counterpart (designed for models with conditional but not regime dependent heteroscedasticity) appears to improve upon the raw asymptotic p-value based inferences or the homoscedastic counterpart of the bootstrap. We conjecture that the ability of the heteroscedastic bootstrap to provide such an improvement when implemented on the untransformed model may be due to the fact that an omitted break in unconditional variance may be mistaken for an ARCH type error (see Hendry (1995, pp. 574-576)).

Besides the above bootstrap based approach for assessing the presence of a break in regression slopes a more practical and less computationally intensive strategy involves the use of robust (heteroskedasticity consistent) versions of the LM statistic implemented on the untransformed model. This is for instance the approach adopted in Stock and Watson (2002) when exploring breaks in the volatility of macroeconomic time series. It is well known however that the use of heteroskedasticity consistent covariance matrix estimators when constructing the test statistics may lead to substantial finite sample distortions in practice (see for instance MacKinnon and White (1985), Cribari-Neto and Zarkos (2001) among others).

The use of robust test statistics that allows us to ignore the potential presence of a jump in the error variance lead to tests with severe size distortions in sample sizes most commonly encountered in applied research. More specifically the empirical sizes characterising the tests appear to be substantially lower than their nominal counterparts, typically less than half the nominal size. Under T=250 for instance and regardless of the magnitude of π_2^0 the empirical size corresponding to a nominal size of 2.5% was approximately 1%. Comparing the use of robust test statistics with the bootstrap based approach it is also clear that the latter performs substantially better.

3.2 Testing for a Structural Break in Error Variances

Here we consider the null hypothesis of no structural break in the error variance given by $H_0 : \sigma_1^2 = \sigma_2^2$ in (1)-(2). Under a correctly specified conditional mean equation the test can be viewed as a simple mean-shift test implemented on the squared residual sequence. Letting $y_t = x_t'\beta + u_t$ with $u_t = \sigma\epsilon_t$ denote the true model under the null hypothesis and defining the corresponding squared residuals sequence as $z_t = (y_t - x_t'\hat{\beta})^2$ the LM statistic can here be written as

$$LM_T(k) = \frac{T}{k(T-k)}\frac{1}{\hat{\sigma}_z^2}\left[\sum_{t=1}^{k} z_t - \frac{k}{T}\sum_{t=1}^{T} z_t\right]^2 \tag{16}$$

with $\hat{\sigma}_z^2 = \sum_{t=1}^{T}(z_t - \bar{z})^2/T$. Using the invariance principle for variances proposed in Phillips and Solo (1992) and noting that under our operating assumptions and the law of large numbers we have $\hat{\sigma}_z^2 \xrightarrow{p} (\kappa - 1)\sigma^4$ with $\kappa \equiv E(\epsilon_t^4)$ it follows that under the null hypothesis the limiting distribution of the test statistic in (16) is given by

$$LM_T(\pi) \xrightarrow{d} \frac{[W(\pi) - \pi W(1)]^2}{\pi(1-\pi)}. \tag{17}$$

Thus inferences based on the supremum, average or exponential versions of $LM_T(\pi)$ over $\pi \in \Pi_2$ can be conducted in a manner identical to the traditional change-point tests using the p-value approximations obtained in Hansen (1997).

Next, suppose that the above procedure ignores the fact that a structural break occurred in the slope parameters. In other words the model under the null hypothesis continues to have a constant error variance given by $u_t = \sigma\epsilon_t \; \forall t$ but $\beta_1 \neq \beta_2$ in (1) and the test ignores this latter feature. Since $\hat{u}_t^2$ absorbs the ignored break in the $\beta's$ we intuitively expect increasingly spurious rejections of the null of variance homogeneity as $T \to \infty$. As shown below however this result is not general and will depend on the magnitude of π_1^0, the true location of the omitted break in the regression slopes together with the nature of the regressors. The following proposition initially summarises the limiting behaviour of a normalised version of $LM_T(\pi)$.

Proposition 5 *Under assumptions (i)-(iv) and model (1)-(2) with $\sigma_1 = \sigma_2 = \sigma$ but $\beta_1 \neq \beta_2$ we have*

$$\begin{aligned} \frac{LM_T(\pi)}{T} &\xrightarrow{p} \frac{1}{\pi(1-\pi)\tau}\left[(1-\pi_1^0)^2\lambda' H_1\lambda - (\pi_1^0)^2\lambda' H_2\lambda\right]^2 \\ &\times \left[\pi(1-\pi_1^0) - (\pi - \pi_1^0)I(\pi > \pi_1^0)\right]^2 \end{aligned} \tag{18}$$

uniformly over $\pi \in \Pi_2$ as $T \to \infty$ and where $H_1 = Q_2Q(1)^{-1}Q_1Q(1)^{-1}Q_2$, $H_2 = Q_1Q(1)^{-1}Q_2Q(1)^{-1}Q_1$ with $Q(1) = [\pi_1^0 Q_1 + (1-\pi_1^0)Q_2]$ and $\tau = plim\ \hat{\sigma}_z^2$.

Note that the right hand side of (18) has been obtained under the assumption that $Q(\pi)$ is as in (3) and a direct implication of this result is that when testing the null hypothesis of no structural break in variance under an omitted break in mean we will have $LM_T(\pi) = O_p(T)$, implying as expected that the size of the test for the null of homogeneity of variances will tend towards one with T leading to systematic rejections of the true null hypothesis, except for some special cases. Equivalently, failing to take the break in mean into account will translate into a spurious break in the error variance, regardless of whether the supremum, average or exponential versions of the test statistic are used.

The expression presented in (18) also indicates that the location of the omitted break in mean as given by the magnitude of π_1^0 will play a key role in determining the severity of the size distortions. For some specific values of π_1^0 it is in fact possible that the empirical implementation of the test may show virtually no distortions at all. To illustrate this latter point we initially focus on the special case that arises under $Q(\pi) = \pi Q$, implying that $H_1 = H_2 = Q$ in (18). This formulation would be valid for instance if the regressors x_t are exogenous. The corresponding limiting behaviour of $LM_T(\pi)/T$ is now given by

$$\frac{LM_T(\pi)}{T} \stackrel{p}{\to} \frac{(\lambda' Q \lambda)^2}{\pi(1-\pi)\tau}(1-2\pi_1^0)^2 \left[\pi(1-\pi_1^0) - (\pi - \pi_1^0) I(\pi > \pi_1^0)\right]^2. \tag{19}$$

From the expression in (19) it is then clear that under this scenario if the omitted break in mean locates in the middle of the sample with $\pi_1^0 = \frac{1}{2}$ then $LM_T(\pi)/T$ will converge to zero in probability and the resulting finite sample size distortions will be much less pronounced. This will not happen however if $Q(\pi)$ is expressed as in (3) since for the right hand side of (19) to equal zero under this setting we would require $H_1 = H_2$ or equivalently $Q_1 = Q_2$ which is ruled out by assumption. If $\pi_1^0 \neq \frac{1}{2}$ however, it is possible that a particular parameter configuration of the DGP may be such that $(1-\pi_1^0)^2 H_1 = (\pi_1^0)^2 H_2$ thus making the right hand side of (19) to equal zero. When the DGP is given by an AR(1) process with $y_t = \beta_1 y_{t-1} I(t \leq k_1^0) + \beta_2 y_{t-1} I(t > k_1^0) + \epsilon_t$ for instance then choosing β_1, β_2 and π_1^0 so that $(1-\pi_1^0)^2(1-\beta_1^2) = (\pi_1^0)^2(1-\beta_2^2)$ would lead to $LM_T(\pi)/T \stackrel{p}{\to} 0$.

To further illustrate and quantify the properties described above we conducted a set of simulation experiments that focused on the empirical size of the test of the null of variance homogeneity when the fitted models ignore the presence of a break in the regression slopes. The null model is now given by (1)-(2) with $\sigma_1 = \sigma_2$ and although $\beta_1 \neq \beta_2$ in the underlying true model the squared residuals are obtained imposing the homogeneity of the $\beta's$. The DGP is here given by an AR(1) with a constant error variance but a break in both its constant and slope parameters. Results are presented in Table

6 which displays the empirical sizes of the test of no structural break in variance under 2.5% and 5% nominal sizes. As expected the figures confirm the fact that the empirical size of the test increases with T leading to spurious rejections of the true null hypothesis.

Table 6 about here

Under $\pi_1^0 = 0.75$ for instance and a nominal size of 5% the corresponding empirical size is given by 11.08% under T=500 and 16.56% under T=1000. Confirming our above discussion we also observe a notably different size behaviour when the omitted break in the $\beta's$ occurs in the middle of the sample with $\pi_1^0 = 0.50$. Under this scenario the empirical sizes remain very close to their nominal counterpart.

On possible strategy for avoiding important size distortions when testing for the possible presence of a break in variance is to control for the break in mean, regardless of whether the latter is present or not. Specifically, using $z_t = (y_t - x_t'\hat{\beta}_1(\hat{k}_1)I(t \leq \hat{k}_1) - x_t'\hat{\beta}_2(\hat{k}_1)I(t > \hat{k}_1))^2$ in (18) it is straightforward to note that the resulting limiting distribution of the LM statistic will be as in (19), thus validating the use of the asymptotic p-values tabulated in Hansen (1997). To further explore the properties of this approach, Table 7 presents corresponding empirical size estimates when the latter are obtained while controlling for the break in mean. The improvements relative to Table 6 are clearly substantial.

Table 7 about here

Under $T = 500$ and $\pi_1^0 = 0.75$ for instance and without controlling for the break in mean the standard implementation of the SupLM test led to an empirical size of 11.08% for a nominal counterpart of 5%. Once we controlled for the break in mean the corresponding empirical size was reduced to 5.02%, virtually equal to its nominal counterpart. The first column of Table 7 also demonstrates that the empirical sizes that arise when the controlled shift in mean is spurious (i.e. when the underlying DGP has no break in mean) remain close to their nominal counterparts.

4 Conclusions

In this paper we formally evaluated the consequences of omitting the presence of a structural break in either the conditional mean parameters or error variance of a series on the resulting change-point estimators and the size properties of parameter constancy tests. We first derived the limiting properties of the least squares based estimators of the location of a break in the mean or variance within models that ignored the possibility that both breaks may occur at the same or different times. Subsequently we analysed the consequences of this type of misspecification on the size properties of parameter constancy tests.

Table 1: Estimator of the Break in Mean Under an Omitted Break in Variance

DGP: $y_t = (1 + 0.4y_{t-1})I(t \leq [T\pi_1^0]) + (2 + 0.1y_{t-1})I(t > [T\pi_1^0]) + \sigma_t\epsilon_t$

$$\sigma_t = \sigma_1 I(t \leq [T\pi_2^0]) + \sigma_2 I(t > [T\pi_2^0])$$

	$\hat{\pi}_1$			STDEV		
	$\pi_1^0 = 0.25, \pi_2^0 = 0.25$					
(σ_1, σ_2)	T=250	T=500	T=1000	T=250	T=500	T=1000
(1,1)	0.268	0.249	0.246	0.132	0.068	0.024
(1,2)	0.434	0.341	0.278	0.226	0.170	0.082
(2,1)	0.213	0.222	0.232	0.099	0.065	0.035
	$\pi_1^0 = 0.50, \pi_2^0 = 0.50$					
(1,1)	0.484	0.491	0.495	0.108	0.051	0.024
(1,2)	0.582	0.541	0.513	0.143	0.096	0.046
(2,1)	0.382	0.436	0.473	0.138	0.101	0.056
	$\pi_1^0 = 0.75, \pi_2^0 = 0.75$					
(1,1)	0.698	0.733	0.744	0.145	0.066	0.025
(1,2)	0.762	0.765	0.760	0.108	0.060	0.033
(2,1)	0.506	0.608	0.695	0.229	0.195	0.121
	$\pi_1^0 = 0.25, \pi_2^0 = 0.75$					
(1,1)	0.268	0.249	0.246	0.132	0.068	0.024
(1,2)	0.449	0.324	0.255	0.284	0.204	0.073
(2,1)	0.272	0.257	0.248	0.135	0.093	0.050
	$\pi_1^0 = 0.75, \pi_2^0 = 0.25$					
(1,1)	0.698	0.733	0.744	0.145	0.066	0.025
(1,2)	0.698	0.723	0.741	0.141	0.095	0.050
(2,1)	0.479	0.624	0.727	0.288	0.233	0.101

Table 2: Estimator of the Break in Variance Under an Omitted Break in Mean (Common Break Dates)

DGP: $y_t = (1 + 0.4y_{t-1})I(t \leq [T\pi^0]) + (2 + 0.1y_{t-1})I(t > [T\pi^0]) + \sigma_t \epsilon_t$

$\sigma_t = \sigma_1 I(t \leq [T\pi^0]) + \sigma_2 I(t > [T\pi^0])$

	$\hat{\pi}_2$			STDEV		
	$\pi^0 = 0.25$					
(σ_1, σ_2)	T=250	T=500	T=1000	T=250	T=500	T=1000
(1,2)	0.300	0.268	0.257	0.114	0.049	0.015
(2,1)	0.220	0.235	0.242	0.036	0.022	0.012
	$\pi^0 = 0.50$					
(1,2)	0.530	0.513	0.505	0.062	0.031	0.011
(2,1)	0.462	0.483	0.492	0.062	0.031	0.012
	$\pi^0 = 0.75$					
(1,2)	0.773	0.762	0.756	0.036	0.023	0.011
(2,1)	0.697	0.730	0.742	0.105	0.041	0.015

Table 3: Estimator of the Break in Variance Under an Omitted Break in Mean (Distinct Break Dates)

DGP: $y_t = (1 + 0.4y_{t-1})I(t \leq [T\pi_1^0]) + (2 + 0.1y_{t-1})I(t > [T\pi_1^0]) + \sigma_t\epsilon_t$

$\sigma_t = \sigma_1 I(t \leq [T\pi_2^0]) + \sigma_2 I(t > [T\pi_2^0])$

	$\hat{\pi}_2$			STDEV		
	$\pi_1^0 = 0.25, \pi_2^0 = 0.75$					
(σ_1, σ_2)	T=250	T=500	T=1000	T=250	T=500	T=1000
(1,2)	0.774	0.762	0.756	0.037	0.024	0.012
(2,1)	0.696	0.762	0.756	0.105	0.039	0.013
	$\pi_1^0 = 0.75, \pi_2^0 = 0.25$					
(1,2)	0.304	0.269	0.257	0.120	0.049	0.015
(2,1)	0.219	0.234	0.242	0.036	0.022	0.012

Table 4: Empirical Size (Null of No Structural Break in Mean)

DGP: No shift in Mean (omitted variance shift)

$$y_t = 1 + 0.4y_{t-1} + u_t$$

$$u_t = \sigma_1\epsilon_t I(t \leq [T\pi_2^0]) + \sigma_2\epsilon_t I(t > [T\pi_2^0])$$

	Correct Fit		$\pi_2^0 = 0.25$		$\pi_2^0 = 0.50$		$\pi_2^0 = 0.75$	
Nominal	2.5%	5%	2.5%	5%	2.5%	5%	2.5%	5%
	$\sigma_1 = 1, \sigma_2 = 1$		$\sigma_1 = 1, \sigma_2 = 2$		$\sigma_1 = 1, \sigma_2 = 2$		$\sigma_1 = 1, \sigma_2 = 2$	
T=250	1.58%	3.52%	2.78%	5.28%	6.74%	10.86%	13.56%	19.24%
T=500	1.88%	4.36%	3.40%	6.00%	7.24%	12.00%	14.04%	19.76%
T=1000	2.36%	4.66%	3.60%	6.60%	8.06%	12.70%	15.26%	21.54%
	$\sigma_1 = 1, \sigma_2 = 1$		$\sigma_1 = 2, \sigma_2 = 1$		$\sigma_1 = 2, \sigma_2 = 1$		$\sigma_1 = 2, \sigma_2 = 1$	
T=250	1.58%	3.52%	13.12%	18.98%	6.06%	10.10%	2.88%	5.00%
T=500	1.88%	4.36%	14.34%	21.08%	6.44%	11.06%	3.24%	6.36%
T=1000	2.36%	4.66%	15.12%	21.20%	7.86%	12.10%	3.79%	6.70%

Table 5: Empirical Size (Null of No Structural Break in Mean)

DGP: No shift in Mean (omitted variance shift)

$$y_t = 1 + 0.4 y_{t-1} + u_t$$

$$u_t = \sigma_1 \epsilon_t I(t \leq [T\pi_2^0]) + \sigma_2 \epsilon_t I(t > [T\pi_2^0])$$

$$\sigma_1 = 1, \ \sigma_2 = 2$$

GLS Transformed Model								
	$\sigma_1 = \sigma_2 = 1$		$\pi_2^0 = 0.25$		$\pi_2^0 = 0.50$		$\pi_2^0 = 0.75$	
T=250	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%
Andrews	2.00%	4.40%	1.70%	3.70%	2.30%	4.70%	3.00%	5.60%
Homboot	2.30%	4.50%	2.00%	4.20%	2.30%	4.70%	3.10%	5.70%
Hetboot	2.90%	5.70%	2.60%	5.20%	2.90%	5.50%	3.10%	6.10%
T=500	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%
Andrews	2.10%	4.70%	1.80%	3.90%	1.90%	4.60%	2.80%	5.80%
Homboot	2.41%	5.10%	2.30%	4.40%	2.10%	4.70%	2.60%	5.50%
Hetboot	2.60%	5.80%	2.70%	4.80%	2.20%	5.30%	3.00%	6.00%
T=1000	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%
Andrews	2.40%	4.80%	2.20%	4.90%	2.70%	5.40%	3.30%	5.90%
Homboot	2.30%	4.70%	2.30%	5.40%	2.80%	5.20%	3.20%	5.30%
Hetboot	2.80%	5.20%	2.30%	5.70%	3.20%	5.70%	3.20%	5.60%
Untransformed Model								
T=250	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%
Andrews	2.20%	4.40%	3.40%	6.40%	7.90%	12.50%	15.50%	21.30%
Robust	1.00%	2.70%	1.06%	2.68%	0.94%	2.58%	0.82%	2.32%
Homboot	2.30%	4.80%	3.50%	6.50%	8.30%	12.90%	16.10%	22.40%
Hetboot	2.70%	5.40%	3.40%	6.50%	4.70%	8.10%	5.70%	9.40%
T=500	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%
Andrews	2.30%	5.00%	3.60%	6.90%	7.50%	10.90%	11.20%	16.70%
Robust	1.68%	3.60%	1.54%	3.90%	1.54%	3.38%	1.30%	3.06%
Homboot	2.50%	5.30%	3.30%	6.50%	7.10%	10.60%	11.60%	16.80%
Hetboot	2.80%	5.80%	3.60%	5.80%	3.60%	5.60%	3.80%	6.70%
T=1000	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%
Andrews	2.50%	4.90%	4.10%	7.20%	6.90%	11.00%	12.90%	18.90%
Robust	1.98%	4.16%	2.02%	4.62%	1.92%	4.18%	1.84%	3.34%
Homboot	2.60%	5.20%	4.20%	6.00%	6.90%	11.13%	14.50%	19.50%
Hetboot	2.70%	5.50%	3.30%	5.60%	3.00%	5.90%	3.20%	5.90%

Table 6: Empirical Size (Null of No Structural Break in Variance)

DGP: No shift in Variance (omitted shift in mean)

$$y_t = (1 + 0.4\ y_{t-1})I(t \leq k_1^0) + (2 + 0.1\ y_{t-1})I(t > k_1^0) + \epsilon_t$$

	$\pi_1^0 = 0.25$		$\pi_1^0 = 0.50$		$\pi_1^0 = 0.75$		No Break in Mean	
Nominal	2.5%	5%	2.5%	5%	2.5%	5%	2.5%	5%
T=250	3.78%	6.10%	2.32%	4.54%	4.58%	7.90%		
T=500	5.50%	9.06%	2.54%	5.16%	6.48%	11.08%		
T=1000	7.94%	12.52%	3.24%	5.76%	11.08%	16.56%		
T=5000	35.90%	46.86%	5.78%	10.12%	52.24%	62.16%		

Table 7: Empirical Size (Null of No Structural Break in Variance)

DGP: No shift in Variance (omitted shift in mean)

$$y_t = 1I(t \leq k_1^0) + 2I(t > k_1^0) + \epsilon_t$$

	$\pi_1^0 = 0.25$		$\pi_1^0 = 0.50$		$\pi_1^0 = 0.75$		No Break in Mean	
Nominal	2.5%	5%	2.5%	5%	2.5%	5%	2.5%	5%
T=250	26.96%	36.44%	1.50%	2.98%	27.28%	36.36%	2.62%	4.68%
T=500	57.54%	66.80%	1.36%	3.50%	57.06%	67.28%	2.80%	4.90%
T=1000	90.64%	94.26%	1.80%	3.62%	90.72%	94.36%	2.76%	5.38%
T=5000	100.00%	100.00%	1.86%	3.82%	100.00%	100.00%	2.60%	4.96%

Table 8: Empirical Size (Null of No Structural Break in Variance)

DGP: No shift in Variance (controlled shift in mean)

$$y_t \quad = \quad (1 + 0.4\ y_{t-1})I(t \leq k_1^0) + (2 + 0.1\ y_{t-1})I(t > k_1^0) + \epsilon_t$$

	Correct Fit		$\pi_1^0 = 0.25$		$\pi_1^0 = 0.50$		$\pi_1^0 = 0.75$	
Nominal	2.5%	5%	2.5%	5%	2.5%	5%	2.5%	5%
T=250	2.46%	4.32%	2.62%	4.68%	2.92%	4.98%	2.70%	4.50%
T=500	2.64%	4.56%	2.80%	4.72%	3.14%	4.98%	2.84%	5.02%
T=1000	2.99%	5.14%	3.14%	5.46%	3.10%	5.72%	2.92%	5.40%

APPENDIX

PROOF OF LEMMA 1 Here we are operating under the case where the break in mean and variance occur at the same time k^0. We initially obtain the limiting behaviour of the normalised objective function given by $G_{2T}([T\pi])/T$ for $k \leq k^0$. We have $\bar{z}_1 = \frac{1}{k}\sum_{t=1}^{k}(\sigma_1\epsilon_t - x_t(\hat{\beta}-\beta_1))^2$ and $\bar{z}_2 = \frac{1}{T-k}[\sum_{t=k+1}^{k^0}(\sigma_1\epsilon_t - x_t(\hat{\beta}-\beta_1))^2 + \sum_{t=k^0+1}^{T}(\sigma_2\epsilon_t - x_t(\hat{\beta}-\beta_2))^2]$ which we can rewrite as

$$\bar{z}_1 \quad = \quad \sigma_1^2\frac{\sum_{t=1}^{k}\epsilon_t^2}{k} + (\hat{\beta}-\beta_1)\frac{\sum_{t=1}^{k}x_t x_t'}{k}(\hat{\beta}-\beta_1) - 2\sigma_1(\hat{\beta}-\beta_1)\frac{\sum_{t=1}^{k}x_t\epsilon_t}{k} \tag{A.1}$$

and

$$\begin{aligned}\bar{z}_2 \quad &= \quad \sigma_1^2\frac{\sum_{t=k+1}^{k^0}\epsilon_t^2}{T-k} + \sigma_2^2\frac{\sum_{t=k^0+1}^{T}\epsilon_t^2}{T-k} + (\hat{\beta}-\beta_1)\frac{\sum_{t=k+1}^{k^0}x_t x_t'}{T-k}(\hat{\beta}-\beta_1) \\ &+ \quad (\hat{\beta}-\beta_2)\frac{\sum_{t=k^0+1}^{T}x_t x_t'}{T-k}(\hat{\beta}-\beta_2) - 2\sigma_1(\hat{\beta}-\beta_1)\frac{\sum_{t=k+1}^{k^0}x_t\epsilon_t}{T-k} - 2\sigma_2(\hat{\beta}-\beta_2)\frac{\sum_{t=k^0+1}^{T}x_t\epsilon_t}{T-k} \end{aligned} \tag{A.2}$$

with $\hat{\beta}$ given by

$$\hat{\beta} \quad = \quad \left(\frac{\sum_{t=1}^{T}x_t x_t'}{T}\right)^{-1}\left(\frac{\sum_{t=1}^{k^0}x_t x_t'}{T}\beta_1 + \frac{\sum_{t=k^0+1}^{T}x_t x_t'}{T}\beta_2 + \sigma_1\frac{\sum_{t=1}^{k^0}\epsilon_t}{T} + \sigma_2\frac{\sum_{t=k^0+1}^{T}\epsilon_t}{T}\right). \tag{A.3}$$

Using assumptions (iii)-(iv) on (A.3) together with standard algebra gives $\hat{\beta} \stackrel{p}{\rightarrow} Q(1)^{-1}[Q(\pi^0)\beta_1 + (Q(1)-Q(\pi^0))\beta_2]$ from which we can also obtain $\hat{\beta}-\beta_1 \stackrel{p}{\rightarrow} Q(1)^{-1}(Q(1)-Q(\pi^0))\lambda$ and $\hat{\beta}-\beta_2 \stackrel{p}{\rightarrow} -Q(1)^{-1}Q(\pi^0)\lambda$. Using the expressions of $\bar{z}_1$ and $\bar{z}_2$ in (A.1)-(A.2) together with assumptions $(iii)-(iv)$ gives

$$\begin{aligned}\bar{z}_2 - \bar{z}_1 \quad &\stackrel{p}{\rightarrow} \quad \left(\frac{1-\pi^0}{1-\pi}\right)\Delta + \lambda'(Q(1)-Q(\pi^0))Q(1)^{-1}\frac{(\pi Q(\pi^0)-Q(\pi))}{\pi(1-\pi)}Q(1)^{-1}(Q(1)-Q(\pi^0))\lambda \\ &+ \quad \lambda Q(\pi^0)Q(1)^{-1}\frac{(Q(1)-Q(\pi^0))}{1-\pi}Q(1)^{-1}Q(\pi^0)\lambda \end{aligned} \tag{A.4}$$

where $\Delta = (\sigma_2^2 - \sigma_1^2)$ and the convergence is uniform over $\pi \leq \pi^0$.

Noting that we are operating under the assumption where $Q(\pi)$ is as in (3) with $Q(1) = \pi^0 Q_1 + (1-\pi^0)Q_2$, (A.4) can be rewritten as

$$\bar{z}_2 - \bar{z}_1 \quad \stackrel{p}{\rightarrow} \quad \frac{1-\pi^0}{1-\pi}\left[\Delta - (1-\pi^0)^2\lambda' Q_2 M^{-1}Q_1 M^{-1}Q_2\lambda + (\pi^0)^2\lambda' Q_1 M^{-1}Q_2 M^{-1}Q_1\lambda\right] \tag{A.5}$$

from which it follows that $\sup_{\pi\leq\pi^0}|\frac{G_{2T}([T\pi])}{T} - G_{2\infty}(\pi\leq\pi^0)| \stackrel{p}{\rightarrow} 0$ with $G_{2\infty}(\pi\leq\pi^0)$ given by

$$G_{2\infty}(\pi\leq\pi^0) \quad = \quad \frac{\pi(1-\pi^0)^2}{1-\pi}\left[\Delta - (1-\pi^0)^2\lambda' Q_2 M^{-1}Q_1 M^{-1}Q_2\lambda + (\pi^0)^2\lambda' Q_1 M^{-1}Q_2 M^{-1}Q_1\lambda\right]^2. \tag{A.6}$$

Proceeding similarly for the case $k \geq k^0$ we obtain $\sup_{\pi\geq\pi^0}|\frac{G_{2T}([T\pi])}{T} - G_{2\infty}(\pi\geq\pi^0)| \stackrel{p}{\rightarrow} 0$ with $G_{2\infty}(\pi\geq\pi^0)$ given by

$$G_{2\infty}(\pi\geq\pi^0) \quad = \quad \frac{(\pi^0)^2(1-\pi)}{\pi}\left[\Delta - (1-\pi^0)^2\lambda' Q_2 M^{-1}Q_1 M^{-1}Q_2\lambda + (\pi^0)^2\lambda' Q_1 M^{-1}Q_2 M^{-1}Q_1\lambda\right]^2. \tag{A.7}$$

Writing $G_{2\infty}(\pi) = G_{2\infty}(\pi\leq\pi^0)I(\pi\leq\pi^0) + G_{2\infty}(\pi\geq\pi^0)I(\pi>\pi^0)$ and using (A.6) with (A.7) leads to the desired result in (10).

PROOF OF PROPOSITION 2 We initially show that $G_{2\infty}(\pi)$ in (7) is uniquely maximised at $\pi = \pi^0$. Letting $C = [\Delta + (\pi^0)^2\lambda' Q_1 M^{-1}Q_2 M^{-1}Q_1\lambda - (1-\pi^0)^2\lambda' Q_2 M^{-1}Q_1 M^{-1}Q_2\lambda]^2$ we have that $dG_{2\infty}(\pi<\pi^0)/d\pi = [(1-\pi^0)/(1-\pi)]^2 C^2 > 0$ together with $dG_{2\infty}(\pi<\pi^0)/d\pi|_{\pi=\pi^0} = C^2 > 0$. Thus $G_{2\infty}(\pi)$ is strictly increasing on $[\underline{\pi}, \pi^0]$ and uniquely maximised at $\pi = \pi^0$. Similarly $dG_{2\infty}(\pi)/d\pi = -(\pi^0/\pi)^2 C^2 < 0$ and $dG_{2\infty}(\pi)/d\pi|_{\pi=\pi^0} = -C^2$ implying that $G_{2\infty}(\pi)$ is strictly decreasing on $[\pi^0, \bar{\pi}]$ and uniquely maximised at $\pi = \pi^0$. Thus uniformly over $[\underline{\pi}, \bar{\pi}]$ the normalised objective function in (7) converges uniformly to a continuous function that is uniquely maximised at $\pi = \pi^0$. Since $\hat{\pi}_2$ maximises $G_{2T}([T\pi])$ it follows from Theorem 2.1 in Newey and McFadden (1994) that $\hat{\pi}_2 \xrightarrow{p} \pi^0$.

PROOF OF PROPOSITION 3 We initially evaluate the extrema of $G_{2\infty}(\pi)$ given in (11). Under $\pi_1^0 < \pi_2^0$ and for $\pi_1^0 = \frac{1}{2}$ we have $G_{2\infty}(\pi < \pi_1^0) = \pi(1-\pi_2^0)^2\Delta^2/(1-\pi)$, $G_{2\infty}(\pi = \pi_1^0) = (1-\pi_2^0)^2\Delta^2$, $G_{2\infty}(\pi_1 < \pi < \pi_2^0) = \pi(1-\pi_2^0)^2\Delta^2/(1-\pi)$, $G_{2\infty}(\pi = \pi_2^0) = \pi_2^0(1-\pi_2^0)\Delta^2$ and $G_{2\infty}(\pi > \pi_2^0) = (1-\pi)(\pi_2^0)^2\Delta^2/\pi$. Next, we note $G_{2\infty}(\pi = \pi_1^0) - G_{2\infty}(\pi < \pi_1^0) = (1-2\pi)(1-\pi_2^0)^2\Delta^2/(1-\pi) > 0$ since $\pi < \frac{1}{2}$ and $G_{2\infty}(\pi = \pi_2^0) - G_{2\infty}(\pi > \pi_2^0) = \pi_2^0(\pi - \pi_2^0)\Delta^2/\pi > 0$ since $\pi > \pi_2^0$. We also note that $G_{2\infty}(\pi = \pi_2^0) - G_{2\infty}(\pi = \pi_1^0) = (1-\pi_2^0)(2\pi_2^0 - 1)\Delta^2 > 0$ since $\pi_2^0 > \frac{1}{2}$. Finally $G_{2\infty}(\pi = \pi_2^0) - G_{2\infty}(\pi_1^0 < \pi < \pi_2^0) = (\pi_2^0 - \pi)(1-\pi_2^0)/(1-\pi) > 0$ from which we can conclude that $G_{2\infty}(\pi)$ is uniquely maximised at $\pi = \pi_2^0$. Since $\hat{\pi}_2$ maximises $G_{2\infty}([T\pi])/T$ and given that this objective function converges uniformly in probability to a nonstochastic continuous function of π that is uniquely maximised at $\pi = \pi_2^0$ the result in (a) follows from Theorem 2.1 in Newey and McFadden (1994). For $\pi_1^0 \neq \frac{1}{2}$ from (11) it is also straightforward to observe that $G_{2\infty}(\pi = \pi_1^0) - G_{2\infty}(\pi < \pi_1^0) > 0$ and $G_{2\infty}(\pi = \pi_2^0) - G_{2\infty}(\pi > \pi_2^0) > 0$. Furthermore, under this scenario we also have

$$G_{2\infty}(\pi = \pi_1^0) - G_{2\infty}(\pi = \pi_2^0) = (\pi_2^0 - \pi_1^0)\left[\frac{\pi_1^0(1-2\pi_1^0)^2(\lambda'Q\lambda)^2}{\pi_2^0} - \frac{1-\pi_2^0}{1-\pi_1^0}\Delta^2\right]^2 \tag{A.8}$$

thus implying that $G_{2\infty}(\pi = \pi_2^0) > G_{2\infty}(\pi = \pi_1^0)$ iff

$$\left(\frac{\Delta}{\lambda'Q\lambda}\right)^2 > \frac{\pi_1^0(1-\pi_1^0)(1-2\pi_1^0)^2}{\pi_2^0(1-\pi_2^0)}. \tag{A.9}$$

Noting that

$$G_{2\infty}(\pi = \pi_2^0) - G_{2\infty}(\pi_1^0 < \pi < \pi_2^0) = (\pi_2^0 - \pi)\left[\Delta^2\frac{1-\pi_2^0}{1-\pi} - (\lambda'Q\lambda)^2\frac{(\pi_1^0)^2(1-2\pi_1^0)^2}{\pi\pi_2^0}\right] \tag{A.10}$$

the expression in (A.10) is strictly positive iff

$$\left(\frac{\Delta}{\lambda'Q\lambda}\right)^2 > \frac{(1-2\pi_1^0)^2(\pi_1^0)^2}{\pi\pi_2^0}. \tag{A.11}$$

but since (A.11) is automatically satisfied under the requirement in (A.9) it follows that $G_{2\infty}(\pi)$ is uniquely maximised at $\pi = \pi_2^0$ and the result in (b) is established. If the condition in (A.9) is reversed so that $G_{2\infty}(\pi = \pi_1^0) > G_{2\infty}(\pi = \pi_2^0)$ it also follows that $G_{2\infty}(\pi = \pi_1^0) - G_{2\infty}(\pi_1 < \pi < \pi_2^0) > 0$ thus implying that $G_{2\infty}(\pi)$ is uniquely maximised at $\pi = \pi_1^0$, leading to the result in (c).

PROOF OF PROPOSITION 4 We initially evaluate the extrema of $G_{2\infty}(\pi)$ as given by (10). Letting $\rho_1 = \sigma_1^2 + \lambda' C_2\lambda$, $\rho_2 = \sigma_1^2 + (\pi_1^0)^2\lambda' C_3\lambda$ and $\rho_3 = \sigma_2^2 + (\pi_1^0)^2\lambda' C_4\lambda$ we can reformulate (10) as $G_{2\infty}(\pi \leq \pi_1^0) = [(1-\pi_2^0)(\rho_3 - \rho_2) + (1-\pi_1^0)(\rho_2 - \rho_1)]^2\pi/(1-\pi)$, $G_{2\infty}(\pi_1^0 < \pi \leq \pi_2^0)\pi(1-\pi)[\frac{1-\pi_2^0}{1-\pi}(\rho_3 - \rho_2) + \frac{\pi_1^0}{\pi}(\rho_2 - \rho_1)]^2$ and $G_{2\infty}(\pi \geq \pi_2^0) = [\pi_2^0(\rho_3 - \rho_2)\pi_2^0 + \pi_1^0(\rho_2 - \rho_1)]^2(1-\pi)/\pi$. Next, noting that $dG_{2\infty}(\pi \leq \pi_1^0)/d\pi > 0$ and $dG_{2\infty}(\pi \geq \pi_2^0)/d\pi < 0$ we have that $G_{2\infty}(\pi)$ is increasing over $[\underline{\pi}, \pi_1^0]$ and decreasing over $[\pi_2^0, \bar{\pi}]$ thus $G_{2\infty}(\pi_1^0) > G_{2\infty}(\pi < \pi_1^0)$ and $G_{2\infty}(\pi_2^0) > G_{2\infty}(\pi > \pi_2^0)$ implying that the maximum of $G_{2\infty}(\pi)$ cannot occur on $[\underline{\pi}, \pi_1^0)$ or $(\pi_2^0, \bar{\pi}]$. Next, we have

$$G_{2\infty}(\pi_2^0) - G_{2\infty}(\pi_1^0 < \pi < \pi_2^0) = \frac{(1-\pi_2^0)(\pi_2^0 - \pi)}{1-\pi}(\rho_3 - \rho_2)^2 - \frac{(\pi_1^0)^2(\pi_2^0 - \pi)}{\pi\pi_2^0}(\rho_2 - \rho_1)^2 \tag{A.12}$$

and

$$G_{2\infty}(\pi_1^0) - G_{2\infty}(\pi_1^0 < \pi < \pi_2^0) = \frac{\pi_1^0(\pi - \pi_1^0)}{\pi}(\rho_2 - \rho_1)^2 - \frac{(1-\pi_2^0)^2(\pi - \pi_1^0)}{(1-\pi)(1-\pi_1^0)}(\rho_3 - \rho_2)^2 \tag{A.13}$$

We initially assume that $G_{2\infty}(\pi_2^0) > G_{2\infty}(\pi_1^0)$. Within our new notations this requirement can be reformulated as

$$G_{2\infty}(\pi_2^0) - G_{2\infty}(\pi_1^0) \quad = \quad \frac{1-\pi_2^0}{1-\pi_1^0}(\rho_3-\rho_2)^2 - \frac{\pi_1^0}{\pi_2^0}(\rho_2-\rho_1)^2 > 0. \tag{A.14}$$

Given that (A.14) holds it is then straightforward to note that (A.12) is strictly positive since $\pi > \pi_1^0$, thus implying that $G_{2\infty}(\pi)/T$ is uniquely maximised at π_2^0. Since $\hat{\pi}_2$ maximises $G_{2T}([T\pi])$ it follows from Theorem 2.1 in Newey and McFadden (1994) that $\hat{\pi}_2 \stackrel{p}{\to} \pi_2^0$ as required. Proceeding similarly for the case where (A.14) is negative so that $G_{2\infty}(\pi_1^0) > G_{2\infty}(\pi_2^0)$ we again have that (A.13) is strictly positive, implying that under this scenario $G_{2\infty}(\pi)$ is uniquely maximised at π_1^0. Since $\hat{\pi}_2$ maximises $G_{2T}([T\pi])/T$ it again follows from Theorem 2.1 in Newey and McFadden (1994) that $\hat{\pi}_2 \stackrel{p}{\to} \pi_1^0$ as required.

PROOF OF PROPOSITION 5 We first consider the case $k \leq k_1^0$ under which we have $\sum_{t=1}^k z_t/T = \sum_{t=1}^k (y_t - x_t'\hat{\beta})^2/T = \sum_{t=1}^k (\sigma\epsilon_t - x_t'(\hat{\beta}-\beta_1))^2$ from which we can also write

$$\frac{\sum_{t=1}^k z_t}{T} \quad = \quad \sigma^2\frac{\sum_{t=1}^k \epsilon_t^2}{T} + (\hat{\beta}-\beta_1)'\frac{\sum_{t=1}^k x_t x_t'}{T}(\hat{\beta}-\beta_1) + o_p(1). \tag{A.15}$$

Letting $\lambda = (\beta_2 - \beta_1)$ and using assumptions (iii)-(iv) we have

$$\hat{\beta} - \beta_1 \quad \stackrel{p}{\to} \quad Q(1)^{-1}(Q(1) - Q(\pi_1^0))\lambda. \tag{A.16}$$

and since we are operating under the assumption that $\pi \leq \pi_1^0$ here $Q(\pi)$ is given by $Q(\pi) = \pi Q_1$ and (A.16) can be rewritten as $\hat{\beta}_1 - \beta_1 \stackrel{p}{\to} (1-\pi_1^0)Q(1)^{-1}Q_2$ with $Q(1) = \pi_1^0 Q_1 + (1-\pi_1^0)Q_2$. Upon rearranging and using standard algebra we thus have

$$\frac{\sum_{t=1}^k z_t}{T} \quad \stackrel{p}{\to} \quad \sigma^2\pi + \pi(1-\pi_1^0)^2\lambda' Q_2 Q(1)^{-1}Q_1 Q(1)^{-1}Q_2\lambda \tag{A.17}$$

where the convergence is uniform over $\pi \leq \pi_1^0$. Proceeding similarly for the limiting behaviour of $\sum_{t=1}^k z_t/T$ under the case $k \geq k_1^0$ we have $\sum_{t=1}^k z_t = \sum_{t=1}^{k_1^0} z_t + \sum_{t=k_1^0+1}^k z_t$ leading to

$$\begin{aligned}\frac{\sum_{t=1}^k z_t}{T} \quad &= \quad \sigma^2\frac{\sum_{t=1}^k \epsilon_t^2}{T} + (\hat{\beta}-\beta_1)'\frac{\sum_{t=1}^{k_1^0} x_t x_t'}{T}(\hat{\beta}-\beta_1) \\ &+ \quad (\hat{\beta}-\beta_2)'\frac{\sum_{t=k_1^0+1}^k x_t x_t'}{T}(\hat{\beta}-\beta_2) + o_p(1)\end{aligned} \tag{A.18}$$

which follows from assumption (iii). Under this scenario (A.16) continues to hold and we also have $\hat{\beta}-\beta_2 \stackrel{p}{\to} -\pi_1^0 Q(1)^{-1}Q_1\lambda$ leading to

$$\begin{aligned}\frac{\sum_{t=1}^k z_t}{T} \quad &\stackrel{p}{\to} \quad \sigma^2\pi + (1-\pi_1^0)^2\pi_1^0\lambda' Q_2 Q(1)^{-1}Q_1 Q(1)^{-1}Q_2\lambda \\ &+ \quad (\pi_1^0)^2(\pi-\pi_1^0)\lambda' Q_1 Q(1)^{-1}Q_2 Q(1)^{-1}Q_1\lambda\end{aligned} \tag{A.19}$$

with the convergence being uniform over $\pi \geq \pi_1^0$. We next turn to the limiting behaviour of $\sum_{t=1}^T z_t/T$ which we can write as

$$\begin{aligned}\frac{\sum_{t=1}^T z_t}{T} \quad &= \quad \sigma^2\frac{\sum_{t=1}^T \epsilon_t^2}{T} + (\hat{\beta}-\beta_1)'\frac{\sum_{t=1}^{k_1^0} x_t x_t'}{T}(\hat{\beta}-\beta_1) \\ &+ \quad (\hat{\beta}-\beta_2)'\frac{\sum_{t=k_1^0+1}^T x_t x_t'}{T}(\hat{\beta}-\beta_2) + o_p(1)\end{aligned} \tag{A.20}$$

from which we obtain

$$\begin{aligned}\frac{\sum_{t=1}^T z_t}{T} \quad &\stackrel{p}{\to} \quad \sigma^2 + \pi_1^0(1-\pi_1^0)^2\lambda' Q_2 Q(1)^{-1}Q_1 Q(1)^{-1}Q_2\lambda + \\ &+ \quad (\pi_1^0)^2(1-\pi_1^0)\lambda' Q_1 Q(1)^{-1}Q_2 Q(1)^{-1}Q_1\lambda.\end{aligned} \tag{A.21}$$

Using (A.17), (A.19) and (A.21) in (16) and upon rearranging leads to the desired result in (18).

REFERENCES

ALTISSIMO and V. CORRADI (2000) "Strong rules for detecting the presence of multiple breaks in time series," Discussion Paper No. 00/11, University of Exeter.

ANDREWS, D. W. K. (1993) "Tests for parameter instability and structural change with unknown change point," Econometrica, 61, 821-856.

ANDREWS, D.W.K. and W. PLOBERGER (1994) "Optimal tests when a nuisance parameter is present only under the alternative," Econometrica 62, 1383-1414.

BAI, J. (1997) "Estimation of a change-point in multiple regression models," Review of Economics and Statistics, 79, 674-679.

BAI, J. (1997) "Estimating multiple breaks one at a time," Econometric Theory, 13, 315-352.

BAI, J. and P. PERRON (1998) "Estimating and Testing Linear Models with Multiple Structural Changes," Econometrica, 66, pp. 47-78.

CHEN, J. and A. K. GUPTA (1997) "Testing and Locating Variance Changepoints With Application to Stock Prices," Journal of the American Statistical Association, 92, pp. 739-747.

CHONG, T. T. (2001) "Structural Change in AR(1) Models," Econometric Theory, 17, pp. 87-155.

CSÖRGO, M. and L. HORVÀTH (1997) *Limit Theorems in Change-Point Analysis,* John Wiley and Sons: New-York.

DIEBOLD, F. X. and C. CHEN (1996) "Testing Structural Stability with Endogenous Breakpoint - A size comparison of analytic and bootstrap procedures," Journal of Econometrics, 70, 221-241.

HANSEN, B. E. (1997) "Approximate asymptotic p-values for structural change tests," Journal of Business and Economic Statistics, 15, 60-67.

HANSEN, B. E. (2000) "Testing for structural change in conditional models," Journal of Econometrics, 97, 93-115.

HANSEN, B. E. (2001) "The New Econometrics of Structural Change," Journal of Economic Perspectives, 15, 117-128.

HENDRY, D. F. (1995) Dynamic Econometrics, Oxford University Press, Oxford.

HINKLEY, D. V. (1970) "Inference about the Change Point in a Sequence of Random Variables," Biometrika, 57, pp. 1-17.

INCLAN, C. (1993) "Detection of Multiple Changes of Variance Using Posterior Odds," Journal of Business and Economic Statistics, Vol. 11, pp. 289-300.

INCLAN, C. and G. C. TIAO (1994) "Use of cumulative sums of squares for retrospective detection of changes of variances," Journal of the American Statistical Association, Vol. 89, pp. 913-923.

McCONNELL, M. M. and G. PEREZ-QUIROS (2000) "Output Fluctuations in the United-States: What has changed since the early 1980s?" American Economic Review, 90, 1464-1476.

NEWEY, W. K. and D. L. McFADDEN (1994) "Large Sample Estimation and Hypothesis Testing," in R. F. Engle and D. L. McFadden, eds., Handbook of Econometrics, Vol. IV (North-Holland), 2111-2245.

STOCK, J. H. and M. W. WATSON (1996) "Evidence on structural instability in macroeconomic time series relations," Journal of Business and Economic Statistics, 14, 11-30.

STOCK, J. H. and M. W. WATSON (1999) "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in R. F. Engle and H. White, eds., Cointegration, Causality and Forecasting: A Festschrift in Honour of C. W. J. Granger, pp. 1-44, Oxford University Press, Oxford.

STOCK, J.H. and M. W. WATSON (2002) "Has the Business Cycle Changed and Why?" NBER Macroeconomics Annual, 2002, Cambridge: MIT Press.

STOCK, J.H. and M. W. WATSON (2003a) "Has the Business Cycle Changed and Why? Evidence and Explanations" Harvard University, Working Paper.

STOCK, J.H. and M. W. WATSON (2003b) "Understanding changes in international business cycle dynamics," NBER Working Paper.

VANDIJK, D. and M. SENSIER (2001) "Short Term Volatility versus Long-Term Growth: Evidence in US Macroeconomic Time Series," Unpublished Manuscript.

WANG and E. ZIVOT (2001) "A Bayesian Time Series Models of Multiple Structural Changes in Level, Trend and Variance," Journal of Business and Economic Statistics, Forthcoming.