Studies in Nonlinear Dynamics & Econometrics

Volume 11, *Issue* 3 2007 *Article* 2

Detecting Multiple Changes in Persistence

Stephen Leybourne* Tae-Hwan Kim†

A.M. Robert Taylor‡

*University of Nottingham, UK, steve.leybourne@nottingham.ac.uk

†Yonsei University, tae-hwan.kim@yonsei.ac.kr

‡University of Nottingham, UK, Robert.Taylor@nottingham.ac.uk

Detecting Multiple Changes in Persistence*

Stephen Leybourne, Tae-Hwan Kim, and A.M. Robert Taylor

Abstract

This paper considers the problem of testing for and dating changes (at unknown points) in the order of integration of a time series between different trend-stationary and difference-stationary regimes. While existing procedures in the literature are designed for processes displaying only a single such change in persistence, our proposed methodology is also valid in the presence of multiple changes in persistence. Our procedure is based on sequences of doubly-recursive implementations of the regression-based unit root statistic of Elliott et al. (1996). The asymptotic validity of our procedure is demonstrated analytically. We use Monte Carlo methods to simulate both finite sample and asymptotic critical values for our proposed testing procedure and to simulate the finite sample behaviour of our procedure against a variety of single and multiple persistence change series. The procedure is shown to work well in practice. The impact of deterministic level and trend breaks on our procedure is also discussed. An empirical application of the procedure to interest rate data is considered.

---

*We are grateful to two anonymous referees for their helpful comments on an earlier version of this paper. Address for Correspondence: Robert Taylor, School of Economics, The Sir Clive Granger Building, University of Nottingham, Nottingham NG7 2RD, U.K. Email: Robert.Taylor@nottingham.ac.uk

1 Introduction

The conventional assumption of a constant order of integration for a time series is contentious and there is a growing body of evidence to suggest that certain economic and financial time-series display changes in persistence, varying between difference-stationarity, $I(1)$, and trend-stationarity, $I(0)$, regimes[1]; see, for example, the literature reviews in Kim (2000) and Leybourne *et al.* (2003) and the citations therein. Notably, Leybourne *et al.* (2003), Busetti and Taylor (2004), Taylor (2005) and Harvey *et al.* (2006) all find significant evidence of a change in persistence from $I(1)$ to $I(0)$ behaviour in the U.S. inflation rate occurring in the early 1980s, in line with changes in the Fed's monetary policy under the Volcker-Greenspan era; see, for example, Clarida *et al.* (2000). Recently, Sollis (2006) reports significant evidence of a change from $I(0)$ to $I(1)$ behaviour in the S&P Composite dividend-price ratio with the breakpoint located in the mid-1950s. As demonstrated in Craine (1993) (see also Campbell and Shiller, 1987) a unit root in this series violates the restriction of no rational stock market bubble. Consequently, Sollis' results suggest the presence a rational bubble from the S&P Composite data from the mid-1950s onwards.

The usual augmented Dickey-Fuller [ADF] unit root test statistic will not diverge to minus infinity with the sample size when applied to persistence change series, since the $I(1)$ part of the series will dominate asymptotically; see, *inter alia*, Chong (2001) and Taylor (2005). In this sense, the ADF test therefore does not provide a consistent testing procedure for discriminating between fixed $I(1)$ processes and persistence change series. Similarly, as shown in Busetti and Taylor (2004), the Kwiatkowski *et al.* (1992) [KPSS] test statistic will diverge to plus infinity, again because of the $I(1)$ part of the series, and so is, in parallel with the ADF test, unable to discern between fixed $I(1)$ processes and persistence change series.

In a seminal paper Banerjee *et al.* (1992) [BLS] propose regression-based tests of the constant $I(1)$ null against the alternative of a single change in persistence, either from $I(0)$ to $I(1)$ or *vice versa*. Their tests are formed from the minimum of a sequence of ADF statistics computed by recursive least squares across changing sub-samples of the data and they tabulate critical values, obtained under the unit root null hypothesis. Recently Leybourne *et al.* (2003) [LKSN] have extended the work of BLS to allow for ADF tests based on local GLS de-trending and have also provided methods of consistently estimating the break fraction between the $I(0)$ and $I(1)$ regimes. Residual-based tests of the constant $I(0)$ null against a change in persistence have been proposed in, *inter alia*, Kim (2000), Kim *et al.* (2002), Busetti and Taylor (2004) and Harvey *et al.* (2006), with consistent break fraction estimators provided in Kim *et al.* (2002) and Busetti and Taylor (2004).

[1] Other authors have found evidence supportive of breaks between separate $I(0)$ regimes; see, for example, Stock and Watson (1996) and Bai and Perron (2003). However, since techniques for identifying such behaviour are already well-established in the literature (see, *inter alia*, Bai and Perron, 1998, 2003, and Bai, 2000) and because our focus in this paper is on processes which display changes in persistence between $I(1)$ and $I(0)$ regimes, we shall not consider such processes further.

The foregoing persistence change tests share the property that they are designed to test for a single change in persistence. Bai (2000,p.304) argues that ' ... multiple changes may be a more accurate characterization of economic time series' than single change-point models because '... a myriad of political and economic factors may alter the data generating process ... ', while Pesaran *et al.* (2006,p.1) argue that 'Structural changes ... appear to affect models for the evolution in key economic and financial times series such as output growth, inflation, exchange rates, interest rates and stock returns. This could reflect legislative, institutional or technological changes, shifts in economic policy ... ' suggesting that the assumption of a single change is likely to be overly restrictive in practice. Indeed, for the U.S. Treasury bill rate series which they analyse, Pesaran *et al.* (2006) delimit seven separate autoregressive regimes, most of which can be identified with well-documented changes in the Fed's monetary policy, the first and third of which appear to be characterised by $I(1)$ behaviour. As a second example, although Kim (2000) finds evidence of a change in persistence from $I(0)$ to $I(1)$ behaviour in the U.S. Government budget deficit he subsequently argues that 'It would be more reasonable to think that a change from stationarity to nonstationarity ... is a temporary one because corrective measures would be instituted at some point in time', *op. cit.*,p.110, implying a multiple persistence change process. Moreover, although the results in Sollis (2006) using tests for a single change in persistence are suggestive of a rational bubble from the S&P Composite data starting from the mid-1950s onwards, the contention of, for example, Shiller (2000) and Campbell and Shiller (2001) that this series is mean-reverting suggests that this bubble would not be sustainable in the long run, implying a subsequent reversion to $I(0)$ behaviour. It is also worth noting that the smooth transition autoregressive and self-exciting autoregressive models with a central $I(1)$ regime, used to model real interest rates and real exchange rates by Kapetanios and Shin (2003), and Kapetanios *et al.* (2003), both imply multiple changes in persistence between $I(0)$ and $I(1)$ behaviour.

In general, the tests for a single change in persistence will not be consistent against processes which display multiple changes in persistence. Where multiple changes in persistence occur these procedures also cannot be used in general to consistently partition the data into its separate $I(0)$ and $I(1)$ regimes. Moreover, even for a single change model they require different tests and breakpoint estimators depending on the direction of change which the data display; that is, whether the data display a change from $I(0)$ to $I(1)$, or *vice versa*. In the light of these important drawbacks with existing approaches, in this paper we develop regression-based tests based on doubly-recursive sequences of ADF-type unit root statistics and associated breakpoint estimators. Our proposed procedure can accommodate processes which display multiple changes in persistence and are valid regardless of the direction of the change(s).

The plan of the paper is as follows. In Section 2 we introduce our null and alternative hypotheses using a simple time-varying [TV] $AR(1)$ model. We start with the simplest model to facilitate a clear and intuitive exposition of our results, which is extended subsequently to a more general model. This model allows for multiple changes in persistence and generalises the models used by the foregoing authors for a single change in persistence. In Section 3 we introduce and motivate our testing and estimation procedure for multiple changes in persistence. In Section 4 we derive representations for the limiting null distribution of our proposed test statistics and provide both finite sample and asymptotic critical values from these distributions. In Section 5 we analyse the properties of our proposed testing strategy under multiple changes in persistence. Here we show that our tests are consistent and that our procedure can be used to consistently partition the data into its separate $I(0)$ and $I(1)$ regimes. We extend our results to include additional serial correlation in the model in Section 6. Section 7 contains some Monte Carlo experiments to investigate the finite sample properties of our proposed test. In section 8 we consider the impact on our procedure of level and trend breaks in the deterministic kernel d_t occurring at known or unknown points in the sample. Section 9 provides an empirical application, using long-run interest rate data for several countries. We find that persistence change patterns differ quite markedly across countries, which has important implications for any attempt to establish empirical support for the uncovered interest rate parity hypothesis using such data. Section 10 concludes. A mathematical appendix contains the proofs of our theoretical results.

2 The Multiple Persistence Change Model

Consider the following TV $AR(1)$ data generation process (DGP) for the time series y_t, observed for $t = 1, ..., T$:

$$y_t = d_t + u_t \tag{2.1}$$
$$u_t = \rho_i u_{t-1} + \varepsilon_t. \tag{2.2}$$

In (2.1), the deterministic kernel, $d_t = z_t'\beta$, is assumed to be either a constant, $z_t = 1$ and $\beta = \beta_0$, or a constant plus linear time trend, $z_t = (1, t)'$ and $\beta = [\beta_0, \beta_1]'$. This is subsequently weakened in section 8 to allow for breaks in level and trend in d_t. In (2.2), the stochastic component u_t is taken for the present to be a TV $AR(1)$ process. This is done in order to simplify our exposition but will be subsequently weakened to allow for TV $AR(p)$ behaviour in Section 6. Precise details on the TV autoregressive parameter, ρ_i, in (2.2) will be given below, while following BLS (Assumption A, p.273) we make the following assumption regarding the innovation process ε_t in (2.2):

Assumption 1: *The error term ε_t is a martingale difference sequence and satisfies* $E(\varepsilon_t^2|\, \varepsilon_{t-1}, ...) = \sigma^2$, $\quad E(|\varepsilon_t|^i|\varepsilon_{t-1}, ...) = \kappa_i$ $(i = 3, 4)$, *and* $\sup_t E(|\varepsilon_t|^{4+\gamma}|\varepsilon_{t-1}, ...) = \kappa < \infty$ *for some* $\gamma > 0$.

Without loss of generality, we assume that the initial value $u_0 = 0$.

The null hypothesis we look to test in this paper is that y_t is $I(1)$ throughout the sample period; that is, $\rho_i = 1$ for all time points in (2.2). We denote this as H_0 in what follows. The alternative hypothesis, H_1, is that y_t is subject to one or more regime shifts between $I(0)$ and $I(1)$ behaviour. Specifically, our alternative DGP is one where the TV autoregressive coefficient, ρ_i, is subject to m change points, where $m \geq 1$ is not assumed known but is assumed not to depend on T, and hence admits $m+1$ regimes, with change-point fractions given by $\tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{m-1} < \tau_m < \tau_{m+1}$, where $\tau_0 = 0$ and $\tau_{m+1} = 1$. Precisely, we specify (2.2) under H_1 as follows. For the i^{th} regime in $(\tau_{i-1}, \tau_i]$,

$$\left.\begin{array}{c} u_t = u_{\lfloor \tau_{i-1}T \rfloor} + h_t, \\ h_t = \rho_i h_{t-1} + \varepsilon_t, \\ h_{\lfloor \tau_{i-1}T \rfloor} = 0 \end{array}\right\} \qquad t = \lfloor \tau_{i-1}T \rfloor + 1, \lfloor \tau_{i-1}T \rfloor + 2, ..., \lfloor \tau_i T \rfloor, \tag{2.3}$$

where $\lfloor \tau T \rfloor$ denotes the integer part of τT: where there is no confusion, we will use τT in place of $\lfloor \tau T \rfloor$ in what follows. Notice that (2.3) is designed to avoid the problem of spurious sharp jumps to zero in u_t at the break points between $I(1)$ and $I(0)$ regimes, noted on p.278 of BLS who use a similar device. In other words, it ensures a 'joining up' of the consecutive $I(1)$ and $I(0)$ regimes in the series. In the context of (2.3), an $I(0)$ regime is one associated with a $|\rho_i| < 1$, while an $I(1)$ regime has $\rho_i = 1$, and the changes between $I(0)$ and $I(1)$ regimes satisfy $\rho_i = 1$ if $|\rho_{i-1}| < 1$ and $|\rho_i| < 1$ if $\rho_{i-1} = 1$. Finally, observe that the single change models considered by, *inter alia*, Kim (2000), BLS and LKSN have $m = 1$ with either $|\rho_1| < 1$ and $\rho_2 = 1$ (a change from $I(0)$ to $I(1)$) or $\rho_1 = 1$ and $|\rho_2| < 1$ (a change from $I(0)$ to $I(1)$), with change-point fraction τ_1.

3 The Test Procedure

Our aim in this paper is both to test for the presence of multiple regime shifts and to consistently estimate the associated change-point fractions. In other words, our goal is to partition y_t, $t = 1, 2, ..., T$, into its separate $I(0)$ and $I(1)$ regimes. As we will subsequently demonstrate, a test statistic appropriate for this purpose is based on a doubly-recursive application of a unit root statistic. Here, we employ the local GLS de-trended ADF unit root testing methodology of Elliott *et al.* (1996), used for detecting a single change in persistence by LKSN.

Let $\tau \in (\lambda, 1]$, for a given λ in $(0, 1)$, and denote the local GLS de-trended ADF unit root statistic that uses the sample observations between λT and τT as $DF_G(\lambda, \tau)$. This statistic is therefore the standard ADF $t-$statistic associated with $\hat{\rho}$ in the fitted OLS regression

$$\Delta y_t^d = \hat{\rho} y_{t-1}^d + \hat{\varepsilon}_t, \qquad t = \lambda T, \lambda T + 1, ..., \tau T \tag{3.1}$$

where $y_t^d \equiv y_t - z_t'\hat{\beta}$, with $\hat{\beta}$ the OLS estimate of β obtained from regressing

$$y_{\lambda,\tau} \equiv [y_{\lambda T}, y_{\lambda T+1} - \bar{\alpha} y_{\lambda T}, ..., y_{\tau T} - \bar{\alpha} y_{\tau T - 1}]' \tag{3.2}$$

on

$$Z_{\lambda,\tau} \equiv [z_{\lambda T}, z_{\lambda T+1} - \bar{\alpha} z_{\lambda T}, ..., z_{\tau T} - \bar{\alpha} z_{\tau T - 1}]' \tag{3.3}$$

with $\bar{\alpha} = 1 + \bar{c}/T$, for some $\bar{c} < 0$; cf. Elliott *et al.* (1996).

Our testing procedure will be based on the statistic

$$M(\lambda) \equiv \inf_{\tau \in (\lambda,1]} DF_G(\lambda, \tau). \tag{3.4}$$

In order to motivate (3.4) as the basis of our test statistic, observe that the testing procedure proposed in LKSN uses (3.4) with $\lambda = \lambda_0 \in [T^{-1}, 2T^{-1})$, so that the first observation used in (3.1) is y_1.[2] This test is designed to detect a DGP with a single change in persistence from $I(0)$ to $I(1)$ behaviour at time $T\tau_1$. The logic behind using $M(\lambda_0)$ in this situation is that with $\tau < \tau_1$, and as $\tau \to \tau_1$, then $DF_G(\lambda_0, \tau)$ becomes increasingly negative (diverging to minus infinity) since an increasing proportion of $I(0)$ observations are added to the regression. However, once $\tau > \tau_1$, a proportion of $I(1)$ observations enter the statistic and prevent it from diverging (i.e. it becomes bounded in probability). This implies that, asymptotically,

$$M(\lambda_0) \equiv \inf_{\tau \in (\lambda_0,1]} DF_G(\lambda_0, \tau) \tag{3.5}$$

will be equal to $DF_G(\lambda_0, \tau_1)$, and hence, $\hat{\tau}_s \equiv \arg\inf_{\tau \in (\lambda_0,1]} DF_G(\lambda_0, \tau) \xrightarrow{p} \tau_1$, so that the change fraction τ_1 is estimated consistently; see LKSN for detailed proofs.

Now suppose that the direction of change is reversed so that a (single) change in persistence from $I(1)$ to $I(0)$ occurs at time $T\tau_1$. The $M(\lambda_0)$ test is now invalid because the initial $I(1)$ regime clearly renders $M(\lambda_0)$ non-divergent (bounded in probability) across all τ. This, therefore, also renders $\hat{\tau}_s$ inconsistent for τ_1. Consequently, LKSN propose a test based on the statistic

$$M^*(\lambda_0) \equiv \inf_{\tau \in (\lambda_0,1]} DF_G^*(\lambda_0, \tau) \tag{3.6}$$

where $DF_G^*(\lambda_0, \tau)$ is the analogue of the $DF_G(\lambda_0, \tau)$ statistic when (3.1) is run using the time-reversed data, $\tilde{y}_t = y_{T-t+1}$, $t = 1, ..., T$. As shown in LKSN, $M^*(\lambda_0)$ diverges and the breakpoint estimator $\tilde{\tau}_s \equiv \arg\inf_{\tau \in (\lambda_0,1]} DF_G^*(\lambda_0, \tau)$ is consistent for τ_1. Where the direction of change is unknown, LKSN suggest basing inference on the statistic $\bar{M} \equiv \min\{M(\lambda_0),\ M^*(\lambda_0)\}$, which they demonstrate to be consistent against either direction of change.

As the previous example shows, important problems exist with the approach of BLS and LKSN even where only a single change in persistence occurs. Now, however, suppose that (2.3) specifies that y_t is $I(1)$ changing to $I(0)$ and back to $I(1)$ at times $T\tau_1$ and $T\tau_2$, respectively; that is, a process which displays two changes in persistence with two $I(1)$ regimes surrounding a single $I(0)$ regime. The initial $I(1)$ regime again renders $M(\lambda_0)$ non-divergent across all τ, but now the terminal $I(1)$ regime also renders

[2] BLS use the same regression as LKSN but applied to the OLS de-trended data, $\tilde{y}_t^d \equiv y_t - z_t'\tilde{\beta}$, where $\tilde{\beta}$ is the OLS estimate of β obtained from regressing y_t on z_t, $t = 1, ..., T$. LKSN show that, as with a comparison of the full-sample ADF tests based on OLS and local GLS de-trending, the tests based on local GLS de-trending are more powerful than those based on OLS de-trending.

$M^*(\lambda_0)$ non-divergent across all τ. Consequently, $\hat{\tau}_s$ and $\tilde{\tau}_s$ will also now be inconsistent for τ_1 and τ_2, respectively. The procedure used in LKSN will not therefore be effective in this situation. However, if we consider instead (3.4), which is indexed by both λ and τ, it follows that $DF_G(\lambda, \tau)$ will be divergent for $\tau_1 \leq \lambda < \tau_2$ and $\tau < \tau_2$, but non-divergent otherwise, since it will contain a proportion of $I(1)$ observations. Moreover, this statistic will become increasingly negative (divergent to minus infinity) as the proportion of $I(0)$ observations grows; that is, as $\lambda \to \tau_1$ and $\tau \to \tau_2$. This implies that the statistic which minimises the doubly-recursive sequence $\{DF_G(\lambda, \tau),\ \lambda \in (0, 1),\ \tau \in (\lambda, 1]\}$ over λ and τ, *viz*,

$$\begin{aligned} M &\equiv \inf_{\lambda \in (0,1)} M(\lambda) \\ &\equiv \inf_{\lambda \in (0,1)} \inf_{\tau \in (\lambda,1]} DF_G(\lambda, \tau) \end{aligned} \tag{3.7}$$

will, asymptotically, be equal to $DF_G(\tau_1, \tau_2)$ and, as a consequence, the estimators $(\hat{\lambda}, \hat{\tau}) \equiv \arg\inf_{\lambda \in (0,1)} \inf_{\tau \in (\lambda,1]} DF_G(\lambda, \tau)$ will be consistent for the change fractions τ_1 and τ_2, respectively; that is, the start and end points of the $I(0)$ regime.

Notice that the testing and change-point estimation procedure based on the doubly-recursive statistic M of (3.7) will also be valid in the single break case even where the direction of change is not known. That is, M will diverge and the estimators $\hat{\lambda}$ and $\hat{\tau}$ will converge in probability to 0 and τ_1 respectively in the case of a change from $I(0)$ to $I(1)$ at time $\tau_1 T$, and to τ_1 and 1 respectively in the case of a change from $I(1)$ to $I(0)$ at time $\tau_1 T$.

The multiple change example given above is perhaps the simplest possible to consider because there is only one $I(0)$ regime for our procedure to delineate. In the more general case where there are multiple changes between $I(0)$ and $I(1)$ regimes (that is, where $m \geq 3$), it is a fairly natural extension of the foregoing logic to anticipate that, asymptotically, M of (3.7) will be equal to $DF_G(\tau_{n-1}, \tau_n)$, and, hence, $(\hat{\lambda}, \hat{\tau}) \equiv \arg\inf_{\lambda \in (0,1)} \inf_{\tau \in (\lambda,1]} DF_G(\lambda, \tau) \xrightarrow{p} (\tau_{n-1}, \tau_n)$. Here τ_{n-1} and τ_n are the start and end points of the 'most prominent' $I(0)$ regime, which is an increasing function of the durations $\tau_i - \tau_{i-1}$ and a decreasing function of the corresponding dominant roots ρ_i (< 1), in each case taken over all the $I(0)$ regimes. Precise details and a proof of this result are given in Section 5.

After having delineated the most prominent $I(0)$ regime, using the procedure above, the presence of any further $I(0)$ regimes can be detected sequentially by analysing the appropriate subintervals of the data. Since the initial application of M detects the most prominent $I(0)$ regime, it must be the case that, asymptotically at least, the interval $[\hat{\lambda}, \hat{\tau}]$ contains no further regime changes, since it cannot include any periods of $I(1)$ behaviour. Hence we need only look for additional regime changes in the two subintervals $[0, \hat{\lambda}]$ and $[\hat{\tau}, 1]$. The same procedure, based on M of (3.7), applied to each of these two subintervals will then find the most prominent (in exactly the same sense as above) $I(0)$ (if any) regime remaining in each. Clearly, then, each of the two subintervals can be further partitioned and examined for additional $I(0)$ regimes. In this way, asymptotically, all $I(0)$ regimes together with their start and end points can be identified. Since these partitions are non-overlapping, the period between the end point of one $I(0)$ regime and the start point of the next $I(0)$ regime must represent an $I(1)$ regime.

4 The Null Distribution of the Statistic

The following theorem establishes the limiting distribution of the test statistic M, via that of $DF_G(\lambda,\tau)$, under the null hypothesis that y_t is $I(1)$ throughout. Here, "$\Rightarrow$" is used to denote weak convergence. The result is presented for the linear trend case, $z_t = (1,\ t)'$. When there is a constant but no time trend in the DGP (2.1) and y_t^d is calculated assuming a constant only, i.e. $z_t = 1$, the limiting null distributions of $DF_G(\lambda,\tau)$ and M obtain as special cases of the results given in Theorem 1.

Theorem 1 *Let y_t be generated according to (2.1)-(2.2) and Assumption 1. Then, under the constant $I(1)$ null hypothesis, H_0, and for $z_t = (1,\ t)'$,*

$$
\begin{aligned}
DF_G(\lambda,\tau) &\Rightarrow \frac{\frac{1}{2}\{V^2_{\tau,\lambda,\tau} - V^2_{\lambda,\lambda,\tau} - (\tau-\lambda)\} + H_{\lambda,\tau}}{(\int_\lambda^\tau V^2_{s,\lambda,\tau} ds)^{1/2}} \\
&\equiv L(\lambda,\tau), \\
M &\Rightarrow \inf_{\lambda\in(0,1)} \inf_{\tau\in(\lambda,1]} L(\lambda,\tau),
\end{aligned}
$$

where, with $W(s)$ a standard Brownian Motion process,

$$
\begin{aligned}
V_{s,\lambda,\tau} &= W(s) - B_{\lambda,\tau} - sB^*_{\lambda,\tau}, \\
B_{\lambda,\tau} &= c_1^{-1}\{W(\lambda)c_3 + \lambda^2 L_1 - \lambda L_2 - \lambda W(\lambda)c_2\}, \\
B^*_{\lambda,\tau} &= c_1^{-1}\left\{L_2 + \bar{c}^2\lambda(\tau-\lambda)W(\lambda) - \lambda L_1 - W(\lambda)c_2\right\}, \\
c_1 &= c_3 + \lambda\bar{c}(2-\bar{c}\tau)(\tau-\lambda), \\
c_2 &= -\bar{c}(\tau-\lambda) + \frac{1}{2}\bar{c}^2(\tau^2-\lambda^2), \\
c_3 &= \tau - \lambda - \bar{c}(\tau^2-\lambda^2) + \frac{1}{3}\bar{c}^2(\tau^3-\lambda^3), \\
L_1 &= -\bar{c}\left\{W(\tau) - W(\lambda)\right\} + \bar{c}^2\int_\lambda^\tau W(s)ds, \\
L_2 &= W(\tau) - W(\lambda) - \bar{c}\left\{\tau W(\tau) - \lambda W(\lambda)\right\} + \bar{c}^2\int_\lambda^\tau sW(s)ds, \\
H_{\lambda,\tau} &= B_{\lambda,\tau}\left\{(\tau-\lambda)B^*_{\lambda,\tau} - W(\tau) + W(\lambda) + V_{\tau,\lambda,\tau} - V_{\lambda,\lambda,\tau}\right\}.
\end{aligned}
$$

Selected finite sample and asymptotic critical values for the test which rejects H_0 for large negative values of the statistic M, obtained by numerical simulation methods, are provided in Section 7.1 below.

5 The Behaviour of the Statistic under H_1

In this section we examine how our proposed statistic M and the associated breakpoint estimators $\hat{\lambda}$ and $\hat{\tau}$ of Section 3 behave under the alternative hypothesis H_1 of at least one regime shift between $I(0)$ and $I(1)$ behaviour.

First let I_0 be the set of i, $2 \leq i \leq m$ such that in (2.3) the regime changes from $I(1)$ to $I(0)$ at τ_{i-1} and back to $I(1)$ at τ_i. Then, u_t is $I(0)$ with parameter $|\rho_i| < 1$ between τ_{i-1} and τ_i. We may then state the following theorem which details the large sample behaviour of the statistic M when applied to the $I(0)$ intervals within the sample.

Theorem 2 *Let y_t be generated according to (2.1)-(2.3) and Assumption 1. Then, for any $i \in I_0$,*

$$T^{-1/2} \inf_{\lambda \in [\tau_{i-1}, \tau_i)} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) \Rightarrow -(\tau_i - \tau_{i-1})^{1/2} \sqrt{\frac{1-\rho_i}{1+\rho_i}} \equiv M_i^*.$$

Remark 1: Notice that M_i^* is strictly negative, with the infimum achieved at $\lambda = \tau_{i-1}$, $\tau = \tau_i$.

Remark 2: Theorem 2 tells us that, for any $i \in I_0$, asymptotically, $\inf_{\lambda \in [\tau_{i-1}, \tau_i)} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) = DF_G(\tau_{i-1}, \tau_i)$. That is, the most negative of the sequence of statistics $\{M(\lambda),\ \lambda \in [\tau_{i-1}, \tau_i)\}$ is $DF_G(\tau_{i-1}, \tau_i)$ which uses all of the data in the $I(0)$ interval.

Now let I_1 be the set of i^*, $2 \leq i^* \leq m$, such that the regime changes from $I(0)$ to $I(1)$ at τ_{i^*-1} and back to $I(0)$ at τ_{i^*}. We may then state the following theorem which details the large sample behaviour of the statistic M when applied to the $I(1)$ intervals within the sample.

Theorem 3 *Let y_t be generated according to (2.1)-(2.3) and Assumption 1. Then, for any $i^* \in I_1$,*

$$T^{-1/2} \inf_{\lambda \in [\tau_{i^*-1}, \tau_{i^*})} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) \Rightarrow 0.$$

With only a trivial modification to the proofs, the results of Theorems 2 and 3 can also be shown to hold when the first regime is $I(0)$ ($i = 1$) or $I(1)$ ($i^* = 1$). Consequently, combining these results, and using the continuous mapping theorem [CMT], we obtain the following corollary when we consider the entire interval for $\lambda \in (0, 1)$.

Corollary 1 *Let y_t be generated according to (2.1)-(2.3) and Assumption 1. Then* *under* H_1,

$$\begin{aligned} T^{-1/2}M &\equiv T^{-1/2} \inf_{\lambda\in(0,1)} \inf_{\tau\in(\lambda,1]} DF_G(\lambda,\tau) \\ &\Rightarrow \inf_{i\in I_0} \left\{ M_i^* = -(\tau_i - \tau_{i-1})^{1/2} \sqrt{\frac{1-\rho_i}{1+\rho_i}} \right\} \\ &= -(\tau_n - \tau_{n-1})^{1/2} \sqrt{\frac{1-\rho_n}{1+\rho_n}} \equiv M_n^*. \end{aligned} \tag{5.1}$$

where n *is the value of* $i \in I_0$ *such that* M_i^* *attains its minimum value at* $i = n$.

Remark 3: This result implies, asymptotically, that $M = DF_G(\tau_{n-1}, \tau_n)$. This infimum is clearly an increasing function of the duration of the $I(0)$ interval, $\tau_i - \tau_{i-1}$, and a decreasing function of the associated parameter ρ_i ($|\rho_i| < 1$). If all of the $I(0)$ intervals are of equal duration, then the infimum will be associated with the interval for which the dominant autoregressive root, ρ_i, is at a minimum. Moreover, where the ρ_i ($|\rho_i| < 1$) take the same value, the infimum will be associated with the $I(0)$ interval for which $\tau_i - \tau_{i-1}$, the duration of the regime, is maximised. We refer to $i = n$ as the *most prominent* $I(0)$ regime that occurs in the data.

Remark 4: The result in Corollary 1 also applies when the data follow a constant $I(0)$ process; that is, $\rho_i = \rho$ for all time points with $|\rho| < 1$ in (2.2). Here $\rho_n = \rho$, $\tau_n = 1$ and $\tau_{n-1} = 0$ in (5.1). In cases where the process is constant $I(0)$, our procedure will therefore (asymptotically) correctly identify the most prominent $I(0)$ regime as being the entire available sample.

Given the result in Corollary 1, the following lemma defines consistent estimators, obtained directly from the statistic M, for the start and end points of the most prominent $I(0)$ regime.

Lemma 1 *Under the conditions of Corollary 1,*

$$\begin{aligned} (\hat{\lambda}, \hat{\tau}) &\equiv \arg \inf_{\lambda\in(0,1)} \inf_{\tau\in(\lambda,1]} DF_G(\lambda,\tau) \\ &\Rightarrow (\tau_{n-1}, \tau_n). \end{aligned}$$

Remark 5: The above procedure therefore allows us to identify the start and finish points of the 'most prominent' $I(0)$ regime that occurs in the data (this may be the full sample as in Remark 4). As noted at the end of Section 3, additional $I(0)$ regimes can be identified by applying the M statistic to subintervals of the data. That is, conditional on M rejecting the null hypothesis, we would, as appropriate, re-apply M to each subinterval $[0, \hat{\lambda}]$ and $[\hat{\tau}, 1]$, and so on. This allows us to identify all the $I(0)$ regimes asymptotically given the consistency results of $(\hat{\lambda}, \hat{\tau})$ demonstrated in

Lemma 1. Moreover, if the subintervals $[0, \hat{\lambda}]$ and $[\hat{\tau}, 1]$ contain no further $I(0)$ regimes, asymptotically the same critical values apply to each subinterval as for the series as whole. This is a direct consequence of the fact that $\hat{\lambda}$ and $\hat{\tau}$ are consistent estimators that allow the $I(0)$ regimes to be isolated, and effectively removed, meaning that subsequent tests are eventually applied to regimes of purely $I(1)$ data.

6 Extension to Higher-Order Time-Varying Serial Correlation

Suppose that, instead of (2.1) and (2.2), y_t is generated by the following TV $AR(p)$ process:

$$\begin{aligned} y_t &= d_t + u_t, \\ u_t &= \sum_{j=1}^{p} \phi_{i,j} u_{t-j} + \varepsilon_t \end{aligned} \tag{6.1}$$

The same conditions as in Assumption 1 are imposed on the error term ε_t. Observe that (6.1) can be equivalently written as

$$u_t = \rho_i u_{t-1} + \sum_{j=1}^{p-1} \varsigma_{ij} \Delta u_{t-j} + \varepsilon_t \tag{6.2}$$

where $\rho_i \equiv \sum_{j=1}^{p} \phi_{ij}$ and $\varsigma_{is} \equiv -(\phi_{i,s+1} + \phi_{i,s+2} + \cdots + \phi_{i,p})$, $i = 1, ..., m+1$, $s = 1, ..., p-1$. Notice that ρ_i is the dominant autoregressive root in the ith regime. Without loss of generality we set $u_0 = u_{-1} = \cdots = u_{1-p} = 0$. We also impose the standard (stability) assumption,

Assumption 2: *All of the roots of the* $(m+1)$ *equations* $1 - \varsigma_{i1} z - \varsigma_{i2} z^2 - \cdots - \varsigma_{i,p-1} z^{p-1} = 0$, $i = 1, ..., m+1$, *lie outside the unit circle.*

Remark 6: Notice that although both the dominant autoregressive root, ρ_i, and the lag coefficients, $\varsigma_{i,j}$, are permitted to differ across the $(m+1)$ separate regimes, for simplicity we have set a common lag length, p, across regimes. In principle the lag length could also be regime-dependent, taking the value p_i, say, in the ith regime, $i = 1, ..., m+1$. This situation can, however, be accommodated within the above framework simply by defining $p \equiv \max_{i=1,...,m+1} p_i$, and setting $\varsigma_{i,j} = 0$, $j = p_i, ..., p-1$, in (6.2) for $i = 1, ..., m+1$.

The null hypothesis H_1 is, as before, that y_t is $I(1)$ throughout the sample period; that is, $\rho_i = 1$ for the entire sample period. The alternative hypothesis, H_1, is again that y_t is subject to one or more regime shifts between $I(0)$ to $I(1)$, with the formulation given in (2.3) appropriately generalised so that the alternative DGP for the i^{th} regime in (τ_{i-1}, τ_i) is given by:

$$\left.\begin{array}{c} u_t = u_{\tau_{i-1}T} + h_t, \\ h_t = \rho_i h_{t-1} + \sum_{j=1}^{p-1} \varsigma_{ij} \Delta h_{t-j} + \varepsilon_t, \\ h_{\tau_{i-1}T} = \cdots = h_{\tau_{i-1}T-p+1} = 0 \end{array}\right\} t = \tau_{i-1}T+1, \tau_{i-1}T+2, ..., \tau_i T, \ i = 1, ..., m+1. \tag{6.3}$$

Our ADF statistic, $DF_G(\lambda, \tau)$, is now obtained from the appropriately augmented regression equation

$$\Delta y_t^d = \hat{\rho} y_{t-1}^d + \sum_{j=1}^{p-1} \hat{\varsigma}_j \Delta y_{t-j} + \hat{\varepsilon}_t, \qquad t = \lambda T, \lambda T+1, ..., \tau T \tag{6.4}$$

where y_t^d is as defined before. The ADF statistic, $DF_G(\lambda, \tau)$, is then fed into (3.7) to obtain the M statistic. The test statistic M, thus obtained, has the same asymptotic null distribution as given in Theorem 1; that is, $M \Rightarrow \inf_{\lambda \in (0,1)} \inf_{\tau \in (\lambda, 1]} L(\lambda, \tau)$. The proof is straightforward, but lengthy and follows along the lines of the proof of the standard full-sample ADF unit root statistic.

We now consider the behaviour of the M statistic under the alternative hypothesis, H_1. Lemma 2 generalises the results in Theorems 2 and 3 to allow for the higher-order serial correlation introduced in (6.3). In the lemma, $\gamma_{i,j}$ and $\gamma_{i,j}^{\Delta}$ are used to denote the j^{th} autocovariances of u_t and Δu_t for the ith stationary regime, $i \in I_0$, respectively, while $e_1' \equiv (1, 0, ..., 0)_{1 \times p}$.

Lemma 2 *Suppose that y_t is generated according to (2.1) and (6.3) and Assumptions 1 and 2. Then, for any $i \in I_0$,*

$$T^{-1/2} \inf_{\lambda \in [\tau_{i-1}, \tau_i)} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) \Rightarrow -(\tau_i - \tau_{i-1})^{1/2} \Theta_i,$$

where

$$\begin{aligned} \Theta_i &= -\frac{e_1' A_i^{-1} B_i}{\sigma \{e_1' A_i^{-1} e_1\}^{1/2}} > 0, \\ A_i &= \begin{bmatrix} \gamma_{i,0} & \gamma_{i,0} - \gamma_{i,1} & \cdots & \gamma_{i,p-2} - \gamma_{i,p-1} \\ \gamma_{i,0} - \gamma_{i,1} & \gamma_{i,0}^{\Delta} & \cdots & \gamma_{i,p-2}^{\Delta} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{i,p-2} - \gamma_{i,p-1} & \gamma_{i,p-2}^{\Delta} & \cdots & \gamma_{i,0}^{\Delta} \end{bmatrix}, \\ B_i &= \begin{bmatrix} \gamma_{i,1} - \gamma_{i,0} \\ \gamma_{i,1}^{\Delta} \\ \cdots \\ \gamma_{i,p-1}^{\Delta} \end{bmatrix}. \end{aligned}$$

Moreover, for any $i^* \in I_1$,

$$T^{-1/2} \inf_{\lambda \in [\tau_{i^*-1}, \tau_{i^*})} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) \Rightarrow 0.$$

As in Corollary 1, when we consider the entire unit interval for $\lambda \in (0,1)$, we have

$$\begin{aligned} T^{-1/2}M &\equiv T^{-1/2} \inf_{\lambda \in (0,1)} \inf_{\tau \in (\lambda,1]} DF_G(\lambda,\tau) \\ &\Rightarrow \inf_{i \in I_0} \left\{ -(\tau_i - \tau_{i-1})^{1/2} \left(-\frac{e_1' A_i^{-1} B_i}{\sigma \{e_1' A_i^{-1} e_1\}^{1/2}} \right) \right\} \end{aligned}$$

The autocovariances $\gamma_{i,j}$ and $\gamma_{i,j}^{\Delta}$ can be expressed solely in terms of $\phi_{i,j}$, the coefficients of the stationary $AR(p)$ process of the i^{th} $I(0)$ regime. Using Maple for symbolic calculation, one obtains that, for any finite $p > 0$,

$$-\frac{e_1' A_i^{-1} B_i}{\sigma \{e_1' A_i^{-1} e_1\}^{1/2}} = \sqrt{\frac{1-\rho_i}{\sum_{j=0}^{p^*} (p-2j)(\phi_{i,p-j} - \phi_{i,j})}} \equiv c_i$$

where $p^* = p/2$ if p is even and $(p-1)/2$ if p is odd and $\phi_{i,0} = -1$. Consequently,

$$\begin{aligned} T^{-1/2}M &\Rightarrow \inf_{i \in I_0} \left\{ -(\tau_i - \tau_{i-1})^{1/2} c_i \right\} \\ &= -(\tau_n - \tau_{n-1})^{1/2} c_n. \end{aligned} \tag{6.5}$$

As before, n is the value of $i \in I_0$ such that $-(\tau_i - \tau_{i-1})^{1/2} c_i$ attains its minimum value at $i = n$. On comparing the general result in (6.5) with the result in Corollary 1, it is easily seen that the same observations that were made in Remark 3 for the simple case still hold true; that is, the limit is an increasing function of the duration of the $I(0)$ interval, $\tau_i - \tau_{i-1}$, and a decreasing function of associated dominant autoregressive root, ρ_i ($|\rho_i| < 1$).

7 Numerical Results

7.1 Critical Values

Table 1 reports finite sample and asymptotic null lower tail critical values for M. These were obtained by Monte Carlo simulation using pseudo-data generated according to the random walk DGP:

$$y_t = y_{t-1} + \varepsilon_t, \quad t = 1, ..., T,$$

with $y_0 = 0$ and $\varepsilon_t \sim NIID(0,1)$. Finite sample critical values are reported for various values of $T \leq 400$ while the row labelled '∞' reports asymptotic critical values for the statistics, obtained by direct simulation of the appropriate limiting functionals given in Theorem 1, using discrete approximations for $T = 1000$. The simulations were performed using $20,000$ Monte Carlo replications and the RNDN function of Gauss 3.1. Results are reported for both the constant only case ($z_t = 1$) and the constant and trend case ($z_t = [1,\ t]'$). We use the value $\bar{c} = -10$ in both situations and throughout the remainder of this paper.[3] Since, when calculating $DF_G(\lambda, \tau)$ we require $\tau > \lambda$, here and in all subsequent applications of the tests used in this paper we set $\tau \geq \lambda + 0.2$.

Table 1: Null Critical Values for M.

	$z_t = 1$			$z_t = [1, t]'$		
T	10%	5%	1%	10%	5%	1%
20	-4.736	-5.369	-7.530	-5.559	-6.149	-7.458
30	-4.391	-5.024	-6.618	-5.459	-6.036	-7.400
40	-4.082	-4.515	-5.662	-5.176	-5.697	-7.024
50	-3.954	-4.351	-5.292	-4.970	-5.450	-6.541
60	-3.883	-4.240	-5.133	-4.846	-5.265	-6.219
70	-3.803	-4.143	-4.974	-4.781	-5.157	-6.069
80	-3.744	-4.083	-4.811	-4.712	-5.078	-5.904
90	-3.735	-4.088	-4.781	-4.671	-5.019	-5.829
100	-3.718	-4.049	-4.734	-4.649	-4.968	-5.654
120	-3.699	-4.026	-4.669	-4.581	-4.904	-5.612
140	-3.669	-3.992	-4.614	-4.553	-4.857	-5.517
160	-3.668	-3.963	-4.558	-4.527	-4.802	-5.403
180	-3.657	-3.973	-4.512	-4.494	-4.773	-5.375
200	-3.662	-3.964	-4.536	-4.480	-4.751	-5.323
250	-3.645	-3.917	-4.513	-4.422	-4.685	-5.238
300	-3.646	-3.926	-4.466	-4.405	-4.670	-5.169
350	-3.642	-3.925	-4.463	-4.400	-4.667	-5.159
400	-3.627	-3.900	-4.438	-4.385	-4.633	-5.099
∞	-3.616	-3.885	-4.421	-4.367	-4.627	-5.078

7.2 Finite Sample Size Properties

In this subsection we use Monte Carlo simulation methods to investigate the behaviour of the multiple persistence change test M of Section 3 when applied to a constant $I(1)$ process driven by serially dependent shocks. Specifically, we follow BLS, p.278 and consider the constant parameter $ARIMA(1, 1, 0)$ DGP:

$$\begin{aligned} y_t &= y_{t-1} + u_t, \quad t = 1, ..., T && (7.1) \\ u_t &= \phi u_{t-1} + \varepsilon_t \quad t = -100, ..., T, && (7.2) \end{aligned}$$

with $y_0 = 0$ and $\varepsilon_t \sim NIID(0, 1)$. As in BLS, we vary the autoregressive design parameter among $\phi \in \{0.4, 0.6\}$. Notice that H_0 holds in all cases.

For the remainder of this section, results are reported for the constant only case, $z_t = 1$.[4] Notice that, due to invariance we have set $d_t = 0$ (cf (2.1)) with no loss of generality. Following standard empirical practice, we have employed data-based lag selection methods to fit the lag truncation order in (6.4). The sequential approach of Ng and Perron (1995) was used with a maximal lag order of four and a significance

[3] The finite sample size and power properties of the tests did not appear to be particularly sensitive to the choice of $\bar{c}$; cf Elliott *et al.* (1996).

[4] Results for $z_t = (1, t)'$ are qualitatively similar and available on request.

level of 10% on the sequential t-tests on the highest lag. This procedure was applied to each sub-sample regression computed.[5]

Panel A of Table 2 reports empirical rejection frequencies, for 200 and 400, of the M test, together with the $M(\lambda_0)$, $M^*(\lambda_0)$ and $\bar{M}$ tests of LKSN (see Section 3), and the full-sample ADF-type test of Elliott *et al.* (1996), $DF_G(0,1)$, when applied to data generated by (7.1)-(7.2). All tests reported in Table 2 were run at the nominal (lower tail) 5% significance level, with critical values taken from Table 1 for the M test, and generated by Monte Carlo simulation for the $M(\lambda_0)$, $M^*(\lambda_0)$, $\bar{M}$ and $DF_G(0,1)$ tests using the same methodology as outlined in Section 7.1. All reported size and power simulation experiments were based on 10,000 replications.

From the results in Panel A of Table 2, it is seen that some size distortion is seen in all of the statistics with the effects slightly larger for the M test than for the $M(\lambda_0)$, $M^*(\lambda_0)$ and $\bar{M}$ tests which in turn display higher empirical size than $DF_G(0,1)$. In all cases the size distortions diminish as T is increased, as would be expected, but do not vary greatly with ϕ. It is well-known that data-dependent lag selection methods are not neutral in finite samples, leading to a degree of over-sizing in the resulting unit root tests; see Taylor (2000). That the M test is most affected is no surprise, given that this is based on a double recursion of sub-sample unit root statistics each using the same data-dependent lag selection procedure.

7.3 Finite Sample Properties Under H_1

In this section we report the empirical rejection frequencies of the tests and breakpoint estimators of Section 3 when the data are generated according to a variety of persistence change DGPs. Specifically, we consider three separate models each of which is a special case of (2.1)-(2.3):

1. **Model I** is a single change-point model; that is, (2.3) with $m = 1$ and the single breakpoint $0 < \tau_1 < 1$. In this case we consider the following values of the first and second sub-sample autoregressive and breakpoint parameters: $\rho_1 \in \{1.0, 0.8\}$, $\rho_2 \in \{1.0, 0.8\}$, with $\rho_1 \neq \rho_2$ and $\tau_1 \in \{0.3, 0.5\}$, respectively.

2. **Model II** is a double change-point model; that is (2.3) with $m = 2$, breakpoints $\tau \equiv (\tau_1, \tau_2)$ such that $0 < \tau_1 < \tau_2 < 1$, and autoregressive regime parameters $\rho \equiv (\rho_1, \rho_2, \rho_3)$. We report results for the following cases: $\tau \in \{(0.25, 0.75),\ (0.35, 0.75)\}$ with $\rho \in (1.0, 0.8, 1.0); \tau = \{(0.25, 0.60),\ (0.35, 0.50)\}$ with $\rho \in (0.8, 1.0, 0.8)$.

3. **Model III** is a quadruple change-point model; that is, (2.3) with $m = 4$, breakpoints $\tau \equiv (\tau_1, ..., \tau_4)$, such that $0 < \tau_1 < \tau_2 < \cdots < \tau_4 < 1$, and autoregressive regime parameters $\rho \equiv (\rho_1, ..., \rho_5)$. We report results for $\tau \in \{(0.1, 0.3, 0.5, 0.9),$ $(0.1, 0.3, 0.4, 0.9)\}$, and $\rho \in \{(1.0, 0.8, 1.0, 0.3, 1.0),\ (1.0, 0.6,\ 1.0, 0.3, 1.0)\}$.

[5] BLS use a fixed lag length of four.

Table 2. Size and power of nominal 5%-level tests against a change in persistence. $z_t = 1$.

Panel A. Size for $ARIMA(1,1,0)$; DGP (7.1)-(7.2)

	$T=200$					$T=400$				
ϕ	M	$M(\lambda_0)$	$M^*(\lambda_0)$	$\bar{M}$	$DF_G(0,1)$	M	$M(\lambda_0)$	$M^*(\lambda_0)$	$\bar{M}$	$DF_G(0,1)$
0.4	0.10	0.08	0.08	0.09	0.06	0.08	0.06	0.06	0.07	0.06
0.6	0.10	0.09	0.08	0.09	0.06	0.08	0.06	0.06	0.07	0.06

Panel B. Power against Model I

			$T=200$							$T=400$						
τ_1	ρ_1	ρ_2	M	M_1	M_2	$M(\lambda_0)$	$M^*(\lambda_0)$	$\bar{M}$	$DF_G(0,1)$	M	M_1	M_2	$M(\lambda_0)$	$M^*(\lambda_0)$	$\bar{M}$	$DF_G(0,1)$
0.30	1.0	0.8	0.82	0.05	0.04	0.16	0.90	0.82	0.36	1.00	0.05	0.04	0.19	0.98	0.97	0.37
0.50	1.0	0.8	0.52	0.04	0.05	0.06	0.82	0.65	0.21	0.99	0.04	0.04	0.08	0.96	0.93	0.23
0.30	0.8	1.0	0.16	0.06	0.03	0.57	0.06	0.41	0.16	0.69	0.05	0.02	0.98	0.06	0.93	0.17
0.50	0.8	1.0	0.53	0.05	0.02	0.89	0.08	0.67	0.32	0.99	0.05	0.02	1.00	0.08	1.00	0.35

Panel C. Power against Model II

					$T=200$							$T=400$						
τ_1	τ_2	ρ_1	ρ_2	ρ_3	M	M_1	M_2	$M(\lambda_0)$	$M^*(\lambda_0)$	$\bar{M}$	$DF_G(0,1)$	M	M_1	M_2	$M(\lambda_0)$	$M^*(\lambda_0)$	$\bar{M}$	$DF_G(0,1)$
0.25	0.75	1.0	0.8	1.0	0.61	0.05	0.05	0.19	0.19	0.24	0.21	0.99	0.05	0.05	0.23	0.23	0.32	0.23
0.35	0.75	1.0	0.8	1.0	0.46	0.04	0.05	0.12	0.17	0.17	0.16	0.96	0.04	0.05	0.15	0.20	0.24	0.17
0.25	0.60	0.8	1.0	0.8	0.96	0.11	0.05	0.49	0.99	0.97	0.36	1.00	0.47	0.03	0.93	1.00	1.00	0.37
0.35	0.50	0.8	1.0	0.8	0.99	0.18	0.04	0.72	1.00	0.99	0.58	1.00	0.75	0.03	0.99	1.00	1.00	0.60

Panel D. Power against Model III

		$T=200$							$T=400$						
τ_3	ρ_2	M	M_1	M_2	$M(\lambda_0)$	$M^*(\lambda_0)$	$\bar{M}$	$DF_G(0,1)$	M	M_1	M_2	$M(\lambda_0)$	$M^*(\lambda_0)$	$\bar{M}$	$DF_G(0,1)$
0.50	0.8	1.00	0.15	0.07	0.29	0.58	0.60	0.33	1.00	0.44	0.05	0.33	0.52	0.60	0.34
0.50	0.6	1.00	0.45	0.07	0.40	0.58	0.66	0.35	1.00	0.96	0.05	0.39	0.52	0.64	0.35
0.40	0.8	1.00	0.17	0.07	0.38	0.60	0.67	0.44	1.00	0.44	0.04	0.39	0.54	0.66	0.43
0.40	0.6	1.00	0.41	0.07	0.47	0.60	0.71	0.46	1.00	0.95	0.04	0.45	0.54	0.69	0.45

Note: For all of the reported results pertaining to Model III, $\tau_1 = 0.1$, $\tau_2 = 0.3$, $\tau_4 = 0.9$, $\rho_1 = \rho_3 = \rho_5 = 1$ and $\rho_4 = 0.3$.

Table 3. Monte Carlo mean and standard deviation of changepoint estimators. Models I, II and III. $z_t = 1$. $T = 200$.

											Model I									
τ_1		ρ_1	ρ_2		$\hat{\lambda}$	$se(\hat{\lambda})$	$\hat{\tau}$	$se(\hat{\tau})$	$\hat{\lambda}_1$	$se(\hat{\lambda}_1)$	$\hat{\tau}_1$	$se(\hat{\tau}_1)$	$\hat{\lambda}_2$	$se(\hat{\lambda}_2)$	$\hat{\tau}_2$	$se(\hat{\tau}_2)$	$\hat{\tau}_s$	$se(\hat{\tau}_s)$	$\tilde{\tau}_s$	$se(\tilde{\tau}_s)$
0.30		1.0	0.8		0.27	0.14	0.94	0.11	0.09	0.08	0.19	0.13	0.90	0.11	0.97	0.05	0.56	0.25	0.31	0.14
0.50		1.0	0.8		0.44	0.15	0.93	0.12	0.15	0.12	0.30	0.15	0.91	0.09	0.97	0.04	0.46	0.24	0.46	0.14
0.30		0.8	1.0		0.09	0.21	0.41	0.23	0.04	0.05	0.13	0.13	0.68	0.19	0.85	0.15	0.37	0.14	0.58	0.21
0.50		0.8	1.0		0.07	0.13	0.55	0.16	0.04	0.04	0.10	0.11	0.70	0.15	0.86	0.12	0.53	0.13	0.54	0.24
											Model II									
τ_1	τ_2	ρ_1	ρ_2	ρ_3	$\hat{\lambda}$	$se(\hat{\lambda})$	$\hat{\tau}$	$se(\hat{\tau})$	$\hat{\lambda}_1$	$se(\hat{\lambda}_1)$	$\hat{\tau}_1$	$se(\hat{\tau}_1)$	$\hat{\lambda}_2$	$se(\hat{\lambda}_2)$	$\hat{\tau}_2$	$se(\hat{\tau}_2)$	$\hat{\tau}_s$	$se(\hat{\tau}_s)$	$\tilde{\tau}_s$	$se(\tilde{\tau}_s)$
0.25	0.75	1.0	0.8	1.0	0.23	0.12	0.77	0.13	0.08	0.07	0.16	0.11	0.84	0.11	0.93	0.08	0.61	0.23	0.41	0.24
0.35	0.75	1.0	0.8	1.0	0.30	0.13	0.77	0.13	0.10	0.09	0.21	0.12	0.84	0.11	0.92	0.08	0.53	0.25	0.44	0.23
0.25	0.60	0.8	1.0	0.8	0.54	0.14	0.96	0.11	0.09	0.12	0.29	0.13	0.93	0.09	0.98	0.03	0.37	0.18	0.57	0.09
0.35	0.50	0.8	1.0	0.8	0.44	0.15	0.97	0.08	0.08	0.09	0.32	0.12	0.93	0.09	0.98	0.03	0.49	0.21	0.47	0.11
											Model III									
τ_3		ρ_2			$\hat{\lambda}$	$se(\hat{\lambda})$	$\hat{\tau}$	$se(\hat{\tau})$	$\hat{\lambda}_1$	$se(\hat{\lambda}_1)$	$\hat{\tau}_1$	$se(\hat{\tau}_1)$	$\hat{\lambda}_2$	$se(\hat{\lambda}_2)$	$\hat{\tau}_2$	$se(\hat{\tau}_2)$	$\hat{\tau}_s$	$se(\hat{\tau}_s)$	$\tilde{\tau}_s$	$se(\tilde{\tau}_s)$
0.50		0.8			0.48	0.06	0.91	0.04	0.11	0.09	0.33	0.10	0.94	0.04	0.98	0.03	0.47	0.24	0.50	0.14
0.50		0.6			0.48	0.08	0.91	0.05	0.09	0.06	0.31	0.08	0.94	0.05	0.98	0.03	0.45	0.23	0.50	0.14
0.40		0.8			0.38	0.06	0.91	0.04	0.10	0.08	0.28	0.09	0.94	0.04	0.98	0.03	0.55	0.25	0.43	0.15
0.40		0.6			0.37	0.08	0.91	0.04	0.09	0.06	0.28	0.08	0.94	0.04	0.98	0.03	0.53	0.25	0.42	0.16

Note: For all of the reported results pertaining to Model III, $\tau_1 = 0.1$, $\tau_2 = 0.3$, $\tau_4 = 0.9$, $\rho_1 = \rho_3 = \rho_5 = 1$ and $\rho_4 = 0.3$.

Table 4. Monte Carlo mean and standard deviation of changepoint estimators. Models I, II and III. $z_t = 1$. $T = 400$.

												Model I										
τ_1		ρ_1	ρ_2		$\hat{\lambda}$	$se(\hat{\lambda})$	$\hat{\tau}$	$se(\hat{\tau})$	$\hat{\lambda}_1$	$se(\hat{\lambda}_1)$	$\hat{\tau}_1$	$se(\hat{\tau}_1)$	$\hat{\lambda}_2$	$se(\hat{\lambda}_2)$	$\hat{\tau}_2$	$se(\hat{\tau}_2)$	$\hat{\tau}_s$	$se(\hat{\tau}_s)$	$\tilde{\tau}_s$	$se(\tilde{\tau}_s)$		
0.30		1.0	0.8		0.28	0.08	0.98	0.05	0.09	0.07	0.18	0.09	0.96	0.06	0.99	0.03	0.59	0.26	0.28	0.08		
0.50		1.0	0.8		0.48	0.09	0.97	0.05	0.15	0.12	0.31	0.13	0.95	0.06	0.99	0.03	0.49	0.25	0.46	0.09		
0.30		0.8	1.0		0.04	0.09	0.33	0.13	0.02	0.02	0.04	0.06	0.58	0.17	0.79	0.16	0.34	0.09	0.55	0.22		
0.50		0.8	1.0		0.03	0.05	0.52	0.09	0.02	0.02	0.04	0.05	0.68	0.13	0.84	0.12	0.53	0.09	0.52	0.25		
												Model II										
τ_1	τ_2	ρ_1	ρ_2	ρ_3	$\hat{\lambda}$	$se(\hat{\lambda})$	$\hat{\tau}$	$se(\hat{\tau})$	$\hat{\lambda}_1$	$se(\hat{\lambda}_1)$	$\hat{\tau}_1$	$se(\hat{\tau}_1)$	$\hat{\lambda}_2$	$se(\hat{\lambda}_2)$	$\hat{\tau}_2$	$se(\hat{\tau}_2)$	$\hat{\tau}_s$	$se(\hat{\tau}_s)$	$\tilde{\tau}_s$	$se(\tilde{\tau}_s)$		
0.25	0.75	1.0	0.7	1.0	0.24	0.07	0.76	0.08	0.07	0.06	0.15	0.08	0.85	0.08	0.93	0.06	0.65	0.22	0.37	0.23		
0.35	0.75	1.0	0.7	1.0	0.34	0.08	0.76	0.08	0.10	0.08	0.21	0.10	0.85	0.08	0.93	0.07	0.55	0.26	0.39	0.22		
0.25	0.60	0.8	1.0	0.8	0.58	0.06	0.99	0.03	0.04	0.08	0.26	0.10	0.97	0.04	0.99	0.01	0.32	0.14	0.58	0.06		
0.35	0.50	0.8	1.0	0.8	0.48	0.10	0.99	0.03	0.04	0.06	0.35	0.09	0.97	0.04	0.99	0.01	0.46	0.18	0.48	0.07		
												Model III										
τ_3		ρ_2			$\hat{\lambda}$	$se(\hat{\lambda})$	$\hat{\tau}$	$se(\hat{\tau})$	$\hat{\lambda}_1$	$se(\hat{\lambda}_1)$	$\hat{\tau}_1$	$se(\hat{\tau}_1)$	$\hat{\lambda}_2$	$se(\hat{\lambda}_2)$	$\hat{\tau}_2$	$se(\hat{\tau}_2)$	$\hat{\tau}_s$	$se(\hat{\tau}_s)$	$\tilde{\tau}_s$	$se(\tilde{\tau}_s)$		
0.50		0.8			0.49	0.02	0.91	0.02	0.09	0.06	0.32	0.07	0.94	0.03	0.97	0.02	0.50	0.24	0.43	0.13		
0.50		0.6			0.49	0.03	0.91	0.02	0.09	0.04	0.31	0.05	0.94	0.03	0.97	0.02	0.48	0.23	0.43	0.14		
0.40		0.8			0.39	0.03	0.91	0.02	0.09	0.06	0.30	0.06	0.94	0.03	0.97	0.02	0.56	0.25	0.35	0.12		
0.40		0.6			0.39	0.04	0.91	0.02	0.09	0.04	0.30	0.05	0.94	0.03	0.97	0.02	0.55	0.25	0.35	0.12		

Note: For all of the reported results pertaining to Model III, $\tau_1 = 0.1$, $\tau_2 = 0.3$, $\tau_4 = 0.9$, $\rho_1 = \rho_3 = \rho_5 = 1$ and $\rho_4 = 0.3$.

Under Model I the process displays a single change in persistence from $I(1)$ to $I(0)$ ($I(0)$ to $I(1)$) at time $\tau_1 T$ for $\rho_1 = 1$ ($\rho_2 = 1$). Under Model II we allow for either an $I(1) - I(0) - I(1)$ or an $I(0) - I(1) - I(0)$ persistence change profile. In the latter case our choice of the design parameters ρ and τ imply that the final $I(0)$ regime is, in the terminology of Section 5, the most prominent. Under Model III the process displays an $I(1) - I(0) - I(1) - I(0) - I(1)$ persistence change profile, with the second of the $I(0)$ regions always the most prominent. Recall that the regression-based tests for a single change in persistence and change-point estimators of LKSN are consistent only against Model I and the $I(0) - I(1) - I(0)$ version of Model II, in both cases conditional upon the direction of change being known. Under neither Model I, nor Model II or Model III is the full-sample ADF-type test of Elliott *et al.* (1996), $DF_G(0, 1)$, consistent.

The tests were calculated in the same way as outlined in Section 6.2, except that no lagged dependent variables were included in the test regressions.[6] The data were generated with $d_t = 0$ (again due to invariance), $\varepsilon_t \sim NIID(0, 1)$ and $u_0 = 0$. Panels B, C and D of Table 2 report results from Models I, II and III, respectively. Tables 3 and 4 report the empirical mean and standard deviation of the estimators for $T = 200$ and $T = 400$, respectively.

The procedure based on the M test is reported as follows. In Panels B, C and D of Table 2, the column headed M reports the rejection frequency for the full-sample M test, as in Panel A. The subsequent columns, labelled M_1 and M_2, denote the rejection frequencies of the second stage M tests, applied to the first and second sub-samples of the data after an initial $I(0)$ region has been delineated. As in Remark 5, we only apply the M_1 and M_2 tests in cases where a significant outcome of the full-sample M statistic occurs. In Tables 3 and 4, $\hat{\lambda}$ and $\hat{\tau}$ denote the estimated start and end points of the most prominent $I(0)$ regime corresponding to the full-sample M test, while $\hat{\lambda}_1$, $\hat{\tau}_1$, and $\hat{\lambda}_2$, $\hat{\tau}_2$ are the corresponding change-point estimators based on the M_1 and M_2 tests, respectively. Also reported are results for the change-point estimators $\hat{\tau}_s$ and $\tilde{\tau}_s$ of LKSN.

Consider first the results for Model I, the single breakpoint model. Here we would expect the tests and estimators of LKSN to dominate, they being specifically designed for the single change case. Broadly speaking this is the case, although when comparing the power properties of the tests it is important to remember that of the tests proposed in LKSN only $\bar{M}$ does not require the user to know the direction of change (i.e. whether there is an $I(0)$-$I(1)$ shift or vice versa). Consequently, the most meaningful power comparison is between M and $\bar{M}$ and we see here that the M test is quite competitive, except for cases where the process displays an early $I(0)$-$I(1)$ shift. For an $I(0)$-$I(1)$ ($I(1)$-$I(0)$) shift, the sub-sample M_2 (M_1) tests are both approximately on the nominal level, as is predicted by the large sample theory. The change-point estimators $\hat{\lambda}$ and $\hat{\tau}$ from the M procedure are also competitive, especially when one bears in mind that, unlike LKSN's estimators $\hat{\tau}_s$ and $\tilde{\tau}_s$, they do not require knowledge of the direction of change. Indeed, the results in Tables 3-4 clearly demonstrate the inconsistency of

[6]This was done because computing time for these experiments with lag selection was found to be prohibitive.

using $\hat{\tau}_s$ ($\tilde{\tau}_s$) when the direction of change is $I(1)$-$I(0)$ ($I(0)$-$I(1)$).

Consider now the results for Model II, the double change model. The results in the first two rows of Panel C are cases where a single $I(0)$ regime is interior to the sample, and here the inconsistency of the tests and estimators of LKSN is quite clear. In contrast, our new M procedure works very well. The M test applied to the full sample displays good power increasing rapidly with the sample size, while the second stage M_1 and M_2 tests are both approximately on the nominal level, again in accordance with the large sample theory. The estimators of the start and end of the interior $I(0)$ regime, $\hat{\lambda}$ and $\hat{\tau}$, respectively, obtained from the M procedure behave well with standard errors decreasing as the sample size increases. Turning to the final two rows of Panel C, both cases where the process displays an $I(0)-I(1)-I(0)$ regime profile, we see that the M and $\bar{M}$ tests both reject the constant $I(1)$ null virtually all of the time. Although $\tilde{\tau}_s$ does a reasonable job in picking up the start point of the second $I(0)$ regime, $\hat{\tau}_s$ does a very poor job picking up the end point of the first $I(0)$ regime (note that both estimators are theoretically consistent). Interestingly, the properties of $\tilde{\tau}_s$ and $\hat{\tau}_s$ in these two cases are not so very different from those for Model II with an $I(0)-I(1)$ shift. Moreover, in the case where $(\tau_1, \tau_2) = (0.35, 0.50)$, so that the $I(1)$ region is small, $\hat{\tau}_s$ is, worryingly, centred around roughly the same fraction as $\tilde{\tau}_s$. In practice then it seems that where multiple changes in persistence occur the LKSN approach may not be especially useful, even in cases where it is theoretically consistent. In contrast the $\hat{\lambda}$ and $\hat{\tau}$ estimators from the M procedure do an excellent job in delineating the 'most prominent' (second) $I(0)$ regime. The second stage M_2 test is then approximately on nominal size (as expected given that, asymptotically, it will be run on purely $I(1)$ data), while the M_1 test shows power increasing with both T and τ_1 to reject the (incorrect) null hypothesis of constant $I(1)$ behaviour in the first sub-sample; notice that for $T = 200$ the initial $I(0)$ region contains only 50 data points for $\tau_1 = 0.25$. The corresponding estimators $\hat{\lambda}_1$ and $\hat{\tau}_1$ also perform well especially for the larger sample size considered.

We now turn to the results for the quadruple change-point Model III. The 'most prominent' $I(0)$ region (that lying between either $0.4T$ or $0.5T$ and $0.9T$) is easily picked up by the full-sample M test and its start and end points are very well estimated by $\hat{\lambda}$ and $\hat{\tau}$, respectively, even for $T = 200$. The interval from $0.9T$ to T is a purely $I(1)$ region and the M_2 test is approximately on the nominal level in all cases, as expected. The M_1 test does a reasonable job in picking up the second interior $I(0)$ region between $0.1T$ and $0.3T$ (especially when one notes that for $T = 200$ this region contains only 40 observations), with power increasing with T and as ρ_2 decreases, other things equal. The estimators of the start and end points of the second interior $I(0)$ regime, $\hat{\lambda}_1$ and $\hat{\tau}_1$, respectively, perform excellently even for $T = 200$. In contrast the inconsistency of the tests and estimators of LKSN is clearly seen.

Finally, the inconsistency of the full sample ADF-type test of Elliott *et al.* (1996), $DF_G(0,1)$ is also clearly demonstrated on comparing the results for $T = 200$ and $T = 400$ for each of Models I,I and III.

8 Structural Breaks in Level and Trend

Thus far we have assumed that either $d_t = \beta_0$ (a constant) or $d_t = \beta_0 + \beta_1 t$ (a linear trend) in (2.1). Our procedure can be generalised to the case where z_t is a fixed sequence satisfying the weak regularity conditions of, for example Phillips and Xiao (1999), simply by using the appropriate local GLS de-trended residuals from the regression of $y_{\lambda,\tau}$ on $Z_{\lambda,\tau}$. Two leading examples (Perron, 1989) are given by: (i) the broken level case, $d_t = \beta_0 + \beta_0^* h_t(\lambda_0)$, and (ii) the broken trend case, $d_t = \beta_0 + \beta_1 t + \beta_0^* h_t(\lambda_0) + \beta_1^*(t - \lfloor \lambda_0 T \rfloor) h_t(\lambda_0)$, where in each case the indicator variable $h_t(\lambda_0) = 1(t > \lfloor T\lambda_0 \rfloor)$ and $\lambda_0 \in (0, 1)$ is a known deterministic breakpoint. Multiple breaks in level and trend are also permitted under the conditions of Phillips and Xiao (1999). In the broken level and broken trend examples given above, $Z_{\lambda,\tau}$ of (3.3) would therefore need to be constructed from $z_t = (1, t, h_t(\lambda_0))'$ and $z_t = (1, t, h_t(\lambda_0), (t - \lfloor T\lambda_0 \rfloor) h_t(\lambda_0))'$, respectively.[7] As noted in Perron and Rodriguez (2003), the broken level case above is in fact an example of a slowly evolving deterministic component satisfying Condition B of Elliott *et al.* (1996, p.816) and, hence, the large sample properties, and hence asymptotic critical values, of the M statistic formed in this way will be identical to those given previously for the constant case, $z_t = 1$. This is also true for the case of multiple breaks in level, but would not be true for the broken trend example, however, and here new critical values would be required.

Of considerably greater practical interest is the case where the timing of deterministic breaks in level or trend are unknown, and here the analysis is more complicated. The conventional unit root literature has tackled this problem in two ways, either by minimising the sequence of unit root statistics over all possible breakpoints, as in Zivot and Andrews (1992), or by estimating the breakpoint and using that estimate as if it were the true value, as in Perron (1997); see also Perron and Rodríguez (2003). In principle the same approaches could be applied to the testing problem considered in this paper, such that for every sub-sample regression used in constructing M these procedures are applied. Although feasible, such procedures would clearly be highly computationally demanding and an analysis of the properties of such methods is beyond the scope of the present paper.

In practice it is probably not unreasonable to assume that structural breaks in the deterministic kernel d_t occur at the same point(s) in the sample as changes in persistence. Interestingly, graphical evidence presented in Kurozumi (2005,p.184) suggests that data generated according to a persistence change process with a fixed level/trend coefficient across the sample (his model, Kurozumi, 2005, Equation (1), is essentially the same as that given in (2.1)-(2.2) in section 2) tends to display a simultaneous shift in level/trend at the persistence change date. As Kurozumi (2005) argues “As we can see from the figures, variance of the process changes before/after the break point. In addition, the figures appear to show a structural break in constant and/or linear trend. These two phenomena sometimes appear in macroeconomic time series and hence, the

[7]When constructing these sub-sample residuals, any indicator variables should, of course, be omitted from z_t if $h_t(\lambda_0)$ assumes a fixed value throughout a given sub-sample.

model (1) may be seen as an alternative to the usual trend-break model." Even so, it is still of interest to investigate the behaviour of our procedure when there are genuine breaks in level/trend.

Regardless of where they occur in the sample, level change(s) satisfy Condition B of Elliott *et al.* (1996) and, hence, have no impact on the large sample properties of our procedure which are therefore as given previously (see sections 4,5 and 6) under both H_0 and H_1. As for trend breaks, owing to the inconsistency of unit root tests in the presence of neglected trend breaks (see, for example, Perron, 1989) it should be clear that the large sample results pertaining to H_1, given in section 5 and in Lemma 2, remain valid in the presence of breaks in trend, provided changes in persistence occur at the same point(s) in the sample as the trend break(s). This because the sub-sample local GLS de-trended ADF statistics used in constructing the M statistic only assume that the deterministic component, d_t, has constant parameters within the specific sample period they are estimated over. Indeed, we might expect our breakpoint estimators to be more accurate in finite samples because of the impact of the trend breaks on the underlying ADF statistics. Under H_0 (where the process is constant $I(1)$) the limiting distribution of M will be non-pivotal, as are the limiting distributions of the standard unit root tests of Zivot and Andrews (1992), Perron (1997) and Perron and Rodríguez (2003). Perron and Rodríguez (2003) show that the standard tests tend to be undersized in such cases and we might therefore expect the same in our tests.

In order to explore the finite sample impact of the level and trend breaks on our proposed procedure, we now report results from a small Monte Carlo experiment. Data were generated according to the DGP

$$y_t = d_t + u_t \tag{8.1}$$
$$u_t = \rho_t u_{t-1} + \varepsilon_t \sim NIID(0,1), \quad u_0 = 0, \tag{8.2}$$

with $\rho_t = 1$, $t = 1, ..., \lfloor T/2 \rfloor$, and $\rho_t = \rho$, $t = \lfloor T/2 \rfloor + 1, ..., T$. We report results for $\rho \in \{1, 0.7\}$ allowing for a constant $I(1)$ process (H_0) and a process which displays a shift from $I(1)$ to $I(0)$ behaviour (H_1) half way through the sample. We allow for a simultaneous change in the level or trend of the process via $d_t = \beta_0^* h_t(0.5) + \beta_1^*(t - \lfloor T/2 \rfloor) h_t(0.5)$. Results are reported for $\beta_0^* = 5$, $\beta_1^* = 0$ and $\beta_0^* = 0$, $\beta_1^* = 0.5$, a break in level of five standard deviations and a break in trend of half a standard deviation, respectively, both of which constitute sizeable breaks in these parameters; cf. Perron (1989,pp.1369-1370). Qualitatively similar conclusions were drawn for other values of these parameters and for other break locations. Results are reported in Tables 5 and 6, where the nomenclature used has the same meaning as in Tables 2-4, for $z_t = 1$ and $z_t = (1, t)'$ for the break in level case, and for $z_t = (1, t)'$ for the break in trend case.

As predicted by the large sample theory discussed above, a change in level has rather little impact on either the size or power of our proposed tests, either for $z_t = 1$ or $z_t = (1, t)'$. The M tests are very marginally under-sized for $T = 200$ with an associated small loss of power relative to the constant level case. The size properties of the second-stage M_1 and M_2 tests appear unaffected by whether there is a level shift or not. However, the breakpoint estimator $\hat{\lambda}$, which estimates the start of the $I(0)$

regime, is markedly more accurate under a simultaneous break in level and change in persistence *vis-à-vis* a change in persistence only, as expected. The accuracy of the endpoint estimator, $\hat{\tau}$, does not appear to vary according to whether there is a simultaneous level change or not. Turning to the results for the broken trend case we again see marginal under-sizing in the M test with power comparable to that seen in the broken level case. Again the size properties of the second-stage M_1 and M_2 tests appear largely unaffected by the trend break. As with the broken level case, the accuracy of the breakpoint estimator $\hat{\lambda}$ is improved under a simultaneous break in trend, with the standard errors further decreased relative to the broken level case. Again $\hat{\tau}$ appears unaffected relative to the constant trend case.

Table 5. Size and power of nominal 5%-level tests against a change in persistence under breaks in level/trend.

Panel A. Size: DGP (8.1)-(8.2), $\rho = 1$									
				$T = 200$			$T = 400$		
z_t	ρ	β_0^*	β_1^*	M			M		
1	1	0	0	0.05			0.05		
		5	0	0.04			0.05		
$(1,\, t)'$	1	0	0	0.05			0.05		
		5	0	0.04			0.05		
		0	0.5	0.04			0.04		
Panel B. Power against DGP (8.1)-(8.2), $\rho = 0.7$									
				$T = 200$			$T = 400$		
z_t	ρ	β_0^*	β_1^*	M	M_1	M_2	M	M_1	M_2
1	0.7	0	0	0.88	0.04	0.04	1.00	0.04	0.03
		5	0	0.82	0.04	0.04	1.00	0.05	0.03
$(1,\, t)'$	0.7	0	0	0.58	0.05	0.05	1.00	0.04	0.04
		5	0	0.51	0.05	0.06	1.00	0.04	0.04
		0	0.5	0.52	0.04	0.06	1.00	0.05	0.05

Table 6. Monte Carlo mean and standard deviation of change-point estimators. DGP (8.1)-(8.2), $\rho = 0.7$.

				$T = 200$				$T = 400$			
z_t	ρ	β_0^*	β_1^*	$\hat{\lambda}$	$se(\hat{\lambda})$	$\hat{\tau}$	$se(\hat{\tau})$	$\hat{\lambda}$	$se(\hat{\lambda})$	$\hat{\tau}$	$se(\hat{\tau})$
1	0.7	0	0	0.458	0.013	0.961	0.006	0.480	0.003	0.987	0.001
		5	0	0.518	0.011	0.962	0.006	0.514	0.002	0.987	0.001
$(1,\, t)'$	0.7	0	0	0.439	0.020	0.937	0.012	0.474	0.005	0.983	0.001
		5	0	0.496	0.020	0.937	0.013	0.503	0.004	0.983	0.001
		0	0.5	0.507	0.009	0.937	0.013	0.505	0.001	0.983	0.001

9 Empirical Application

As an empirical illustration of our test procedure, we consider the problem of detecting $I(0)$ and $I(1)$ regimes in long-run interest rate data. Specifically, we examine the log yields on 10 year Government bonds for the United Kingdom, United States, Canada and Australia. The data, taken from the OECD/MEI database, are monthly and cover the period 1978:1-2001:12 (288 observations), with the exception of the Canadian data, which covers the period 1982:6-2001:12 (235 observations).

Since a visual inspection of the data generally suggests a general long-term reduction in interest rates that is not inconsistent with a downward linear trend, we permitted this possibility by employing the M test fitting a constant and trend, $z_t = (1, t)'$. Data-based lag selection was carried out using the Ng and Perron (1995) procedure outlined in Section 7.2. The results are given in Table 7, where a bold entry signifies a rejection at the 10% significance level or below.

Table 7: Application to Interest Rate Data.

	period (obs)	M	$I(0)$ regime start (obs)	$I(0)$ regime end (obs)
U.K.	1978:01-2001:12 (1-288)	**-4.845**	1982:09 (57)	1988:10 (130)
U.S.A.	1978:01-2001:12 (1-288)	-3.780	-	-
Canada	1982:06-2001:12 (1-235)	**-4.613**	1994:07 (146)	1998:12 (199)
	1982:06-1994:06 (1-145)	**-5.397**	1990:04 (95)	1992:09 (124)
Australia	1978:01-2001:12 (1-288)	**-4.489**	1982:12 (60)	1987:11 (119)
	1978:01-1981:11 (1-59)	**-6.673**	1978:12 (12)	1980:03 (27)

The initial application of the M test uncovers interior $I(0)$ regimes for each country except the United States; the United States data appearing $I(1)$ throughout. For the United Kingdom and Australia, the dates of these $I(0)$ regimes are fairly similar, lasting throughout most of the 1980s. For Canada, the $I(0)$ regime lasts from the mid to late 1990s. No further $I(0)$ regimes were uncovered for the United Kingdom using sub-sample M tests. For Canada, a second $I(0)$ regime is found, interior to the first sub-sample, this time in the early 1990s. Similarly, for Australia, a second $I(0)$ regime is found, interior to the first sub-sample, in the late 1970s. No other sub-samples yielded significant test values. Thus, United Kingdom interest rates appear to have switched regimes in the order $I(1) - I(0) - I(1)$, while both Canadian and Australian rates have followed an $I(1) - I(0) - I(1) - I(0) - I(1)$ pattern of regimes.[8] When the tests $M(\lambda_0)$, $M^*(\lambda_0)$ and $\bar{M}$ are applied to this data, no rejections are found.[9] This however, is not surprising as when $I(0)$ regimes have been uncovered they all appear to occur in the interior of the relevant sample, and, as we have demonstrated, these latter tests are not capable of consistently detecting this sort of behaviour. If nothing else, these results serve to clearly highlight the differing patterns of persistence of long-run interest series across major economies over a common period of time. This, in turn, has important implications for empirical tests of the uncovered interest rate parity (UIP)

[8]The number of lagged difference terms identified for the DF_G statistic in each regime is as follows. UK: 1,1,3; US: 1; Canada: 0,3,0,4,0; Australia: 3,4,4,1,0.

[9]The numerical results are available on request.

hypothesis. The UIP hypothesis requires that the difference between long-run interest rates of two countries should be $I(0)$. However, this obviously cannot be the case if two countries long-run interest rate series experience differing patterns of persistence, in terms of the numbers, timings and durations of the $I(0)$ and $I(1)$ regimes. Such data is simply incapable of lending any empirical support to the UIP hypothesis. We conjecture that this is a firm possibility in explaining why so many empirical studies have failed to yield any support for the UIP hypothesis.

10 Conclusions

In this paper we have extended the literature on testing for and dating persistence change to allow for processes which display multiple changes in persistence. Our proposed procedure is based on doubly-recursive implementations of the well-known ADF-type unit root test of Elliott *et al.* (1996). A first pass of the procedure across the complete sample allows one to identify the 'most prominent' $I(0)$ region in the data. Subsequent applications of the procedure can then be used to detect any further $I(0)$ regimes in the sample. Our proposed procedure can therefore consistently partition the sample data into its separate $I(0)$ and $I(1)$ regimes. The limiting null distribution of our proposed tests were derived and their consistency properties, together with those of the associated change-point estimators, against processes which display persistence change were established. Using numerical methods we have provided both finite sample and asymptotic critical values for our proposed tests and investigated the finite sample behaviour of our proposed methodology against a variety of both single and multiple persistence change DGPs. The impact on our proposed procedure of level and/or trend breaks occurring at known or unknown points in the sample was also explored. Finally, an empirical application of our proposed procedure to interest data uncovered evidence of differing patterns of $I(0)$ and $I(1)$ regime switching in the majority of the series analysed.

A Appendix

PROOF OF THEOREM 1. Since the test statistic is invariant to β_0 and β_1 we assume, without loss of generality, that $\beta_0 = \beta_1 = 0$ in what follows. It can be shown, using the Functional Central Limit Theorem [FCLT] and the CMT, that under

Assumption 1

$$\begin{aligned} T^{-1/2}\sum_{t=\lambda T}^{\tau T}\Delta y_t &\Rightarrow \sigma\{W(\tau)-W(\lambda)\}, \qquad T^{-3/2}\sum_{t=\lambda T}^{\tau T} y_t \Rightarrow \sigma\int_\lambda^\tau W(s)ds, \\ T^{-3/2}\sum_{t=\lambda T}^{\tau T} t\Delta y_t &\Rightarrow \sigma\{\tau W(\tau)-\lambda W(\lambda)\}-\sigma\int_\lambda^\tau W(s)ds, \quad T^{-5/2}\sum_{t=\lambda T}^{\tau T} t y_t \\ &\Rightarrow \sigma\int_\lambda^\tau sW(s)ds. \end{aligned}$$

Using these results, we obtain that $T^{-1/2}\hat{\beta}_0 \Rightarrow \sigma B_{\lambda,\tau}$ and $T^{1/2}\hat{\beta}_1(\tau) \Rightarrow \sigma B^*_{\lambda,\tau}$ where $B_{\lambda,\tau}$ and $B^*_{\lambda,\tau}$ are as defined in the theorem.

Consider the statistic $DF_G(\lambda,\tau)$ from (3.1). This can be written as

$$DF_G(\lambda,\tau) = \hat{\sigma}_T^{-1}A_{1,T}(\lambda,\tau)^{-1/2}A_{2,T}(\lambda,\tau) \tag{A.1}$$

where

$$\begin{aligned} A_{1,T}(\lambda,\tau) &= T^{-2}\sum_{t=\lambda T}^{\tau T}(y_{t-1}^d)^2, \\ A_{2,T}(\lambda,\tau) &= T^{-1}\sum_{t=\lambda T}^{\tau T} y_{t-1}^d\Delta y_t^d, \\ \hat{\sigma}_T^2 &= \{(\tau-\lambda)T\}^{-1}\sum_{t=\lambda T}^{\tau T}\hat{\varepsilon}_t^2. \end{aligned} \tag{A.2}$$

Using the fact that $T^{-1/2}y_{sT}^d \Rightarrow \sigma V_{s,\lambda,\tau}$ where $V_{s,\lambda,\tau} = W(s) - B_{\lambda,\tau} - sB^*_{\lambda,\tau}$, it is straightforward to demonstrate that

$$\begin{aligned} A_{1,T}(\lambda,\tau) &\Rightarrow \sigma^2\int_\lambda^\tau V_{s,\lambda,\tau}^2 ds \\ A_{2,T}(\lambda,\tau) &\Rightarrow \sigma^2\frac{1}{2}\left\{V_{\tau,\lambda,\tau}^2 - V_{\lambda,\lambda,\tau}^2 - (\tau-\lambda)\right\} + H_{\lambda,\tau} \end{aligned}$$

with $H_{\lambda,\tau}$ as defined in the theorem. These results, together with the fact that $\hat{\sigma}_T^2 \Rightarrow \sigma^2$, yield the result that $DF_G(\lambda,\tau) \Rightarrow L(\lambda,\tau)$. Consequently, via an application of the CMT we obtain that

$$\inf_{\lambda\in(0,1)}\inf_{\tau\in(\lambda,1]} DF_G(\lambda,\tau) \Rightarrow \inf_{\lambda\in(0,1)}\inf_{\tau\in(\lambda,1]} L(\lambda,\tau)$$

as required.

PROOF OF THEOREM 2. The remaining proofs are presented for $z_t = 1$. The corresponding proofs for $z_t = (1,t)'$ are very similar but lengthier and yield identical results and, hence, are omitted. Since the statistic is invariant to β_0 we assume, without loss of generality, that $\beta_0 = 0$. Under Assumption 1, it can be shown that the OLS estimator $\hat{\beta}_0$ from the regression of $y_{\lambda,\tau}$ on $Z_{\lambda,\tau}$ has the following property; for a value $\lambda \in [\tau_{i-1}, \tau_i)$,

$$\hat{\beta}_0 = y_{\tau_{i-1}T} + h_{\lambda T} 1[\lambda \in (\tau_{i-1}, \tau_i)] + O_p(T^{-1/2}). \tag{A.3}$$

For some $\lambda \in [\tau_{i-1}, \tau_i)$, consider regression (3.1) which uses observations between λT and τT. First, we examine the case where $\lambda < \tau \leq \tau_i$. Here $DF_G(\lambda, \tau)$, scaled by $T^{-1/2}$, can be written as

$$T^{-1/2} DF_G(\lambda, \tau) = (\tau - \lambda)^{1/2} \hat{\sigma}_T^{-1} A_{1,T}^*(\lambda, \tau)^{-1/2} A_{2,T}^*(\lambda, \tau)$$

where

$$\begin{aligned} A_{1,T}^*(\lambda, \tau) &= \{(\tau - \lambda)T\}^{-1} \sum_{t=\lambda T}^{\tau T} (y_{t-1}^d)^2, \\ A_{2,T}^*(\lambda, \tau) &= \{(\tau - \lambda)T\}^{-1} \sum_{t=\lambda T}^{\tau T} y_{t-1}^d \Delta y_t^d \end{aligned}$$

with $\hat{\sigma}_T^2$ as defined in (A.2). The following results follow straightforwardly using (A.3):

$$\begin{aligned} A_{1,T}^*(\lambda, \tau) &\Rightarrow \gamma_{i,0} + u_{i,\infty}^2 1[\lambda \in (\tau_{i-1}, \tau_i)] \\ &\equiv a_{i,1}(\lambda), \\ A_{2,T}^*(\lambda, \tau) &\Rightarrow -(\gamma_{i,0} - \gamma_{i,1}) \\ &\equiv a_{i,2}, \\ \hat{\sigma}_T^2 &\Rightarrow -2a_{i,2} - a_{i,2}^2 a_{i,1}(\lambda)^{-1}, \end{aligned}$$

where $\gamma_{i,j} = \sigma^2 \rho_i^j / (1 - \rho_i^2)$, and $u_{i,\infty}$ is a random drawing from the $I(0)$ regime between τ_{i-1} and τ_i. Combining these results, and simplifying, we obtain that

$$T^{-1/2} DF_G(\lambda, \tau) \quad \Rightarrow \quad -(\tau - \lambda)^{1/2} H_i(\lambda),$$

where

$$H_i(\lambda) = (1 + 2u_{i,\infty}^2 1[\lambda \in (\tau_{i-1}, \tau_i)])^{-1/2} \sqrt{\frac{1 - \rho_i}{1 + \rho_i}}.$$

Since $H_i(\lambda) > 0$, it follows immediately that the limit is negative.

Next, consider the case where $\lambda < \tau_i < \tau$. From (A.1) we have that $DF_G(\lambda, \tau) = \hat{\sigma}_T^{-1} A_{1,T}(\lambda, \tau)^{-1/2} A_{2,T}(\lambda, \tau)$. Now since

$$A_{1,T}(\lambda, \tau) \equiv T^{-2} \sum_{t=\lambda T}^{\tau_i T} (y_{t-1}^d)^2 + T^{-2} \sum_{t=\tau_i T+1}^{\tau T} (y_{t-1}^d)^2, \tag{A.4}$$

it is seen that the second term in the right member of (A.4) involves a non-zero proportion of $I(1)$ variables. Consequently, $A_{1,T}(\lambda, \tau)$ is of $O_p(1)$. That $A_{2,T}(\lambda, \tau)$ and $\hat{\sigma}_T^2$ are also of $O_p(1)$ follows similarly. Hence, $DF_G(\lambda, \tau) = O_p(1)$, and so $T^{-1/2} DF_G(\lambda, \tau) \Rightarrow 0$.

Combining the results for both $\lambda < \tau \leq \tau_i$ and $\lambda < \tau_i < \tau$, we have that

$$T^{-1/2} DF_G(\lambda, \tau) \Rightarrow -(\tau - \lambda)^{1/2} H_i(\lambda) 1[\tau \in (\lambda, \tau_i]]. \tag{A.5}$$

Using an application of the CMT, we therefore obtain that

$$T^{-1/2} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) \Rightarrow \inf_{\tau \in (\lambda, \tau_i]} - \left\{ (\tau - \lambda)^{1/2} H_i(\lambda) \right\}.$$

The function $H_i(\lambda)$ does not depend on τ and so $-(\tau - \lambda)^{1/2} H_i(\lambda)$ is monotonically decreasing in τ over the interval $(\lambda, \tau_i]$. Hence, the infimum over τ is achieved at the upper change point τ_i. Thus, we obtain that

$$T^{-1/2} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) \Rightarrow -(\tau_i - \lambda)^{1/2} H_i(\lambda).$$

Finally, since $\lambda \in [\tau_{i-1}, \tau_i)$, we note that $-(\tau_i - \lambda)^{1/2} H_i(\lambda)$ takes its infimum value λ at the lower change point τ_{i-1}. That is,

$$\begin{aligned} T^{-1/2} \inf_{\lambda \in [\tau_{i-1}, \tau_i)} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) &\Rightarrow -(\tau_i - \tau_{i-1})^{1/2} H_i(\tau_{i-1}) \\ &= -(\tau_i - \tau_{i-1})^{1/2} (1 - \rho_i)^{1/2} (1 + \rho_i)^{-1/2}. \end{aligned}$$

PROOF OF THEOREM 3. Using (A.1) and (A.2) again, we see that for any $\lambda \in [\tau_{i^*-1}, \tau_{i^*})$, $A_{1,T}(\lambda, \tau)$ involves a non-zero proportion of $I(1)$ variables and, hence, is of $O_p(1)$ under Assumption 1. That $A_{2,T}(\lambda, \tau)$ and $\hat{\sigma}_T^2$ are also of $O_p(1)$ follows similarly. Hence, $DF_G(\lambda, \tau) = O_p(1)$ and it follows that $\inf_{\lambda \in [\tau_{i^*-1}, \tau_{i^*})} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) = O_p(1)$.

PROOF OF LEMMA 1. Let $\gamma = (\lambda, \tau)'$, $\hat{\gamma} = (\hat{\lambda}, \hat{\tau})'$ and $\gamma^* = (\tau_{n-1}, \tau_n)'$. Using the same logic as was used in deriving the result in (A.5), it can be shown that

$$Q_T(\gamma) \Rightarrow Q(\gamma) \tag{A.6}$$

where $Q_T(\gamma) = T^{-1/2} DF_G(\lambda, \tau)$ and $Q(\gamma) = -(\tau - \lambda)^{1/2} H_i(\lambda) 1[\lambda \in [\tau_{i-1}, \tau_i)] 1[\tau \in (\lambda, \tau_i)] 1[i \in I_0]$. Defining $\Gamma = \{\gamma \in (0,1)^2 : \lambda < \tau\}$, we note that $\hat{\gamma} = \arg\inf_{\gamma \in \Gamma} Q_T(\gamma)$ and $\gamma^* = \arg\inf_{\gamma \in \Gamma} Q(\gamma)$ by the definition of $\gamma^* = (\tau_{n-1}, \tau_n)'$ in Corollary 1. Now since the objective function $Q_T(\gamma)$ approaches the limit function $Q(\gamma)$, the solution $\hat{\gamma}$ from $Q_T(\gamma)$ should also approach the solution from $Q(\gamma)$. We formalise this idea below.

Define N to be an open ball in Γ centred at γ^*; that is, for some value of a, $N = \{(\lambda, \tau) : (\lambda - \tau_{n-1})^2 + (\tau - \tau_n)^2 < a\} \subset \Gamma$, and let N^c be the complement of N in Γ. Then, it can be shown that $\inf_{\gamma \in \bar{N}} Q(\gamma) = Q(\gamma_a)$ where $\gamma_a \in \{(\lambda, \tau) : (\lambda - \tau_{n-1})^2 + (\tau - \tau_n)^2 = a\}$.

Let $\xi = Q(\gamma_a) - Q(\gamma^*)$ and E_T be the event "$|Q_T(\gamma) - Q(\gamma)| < \frac{\xi}{2}$ for all γ." With this definition of ξ, it is entirely obvious that for any γ satisfying $Q(\gamma) - Q(\gamma^*) < \xi$, we have $\gamma \in N$. Below, we show that $\hat{\gamma} \in N$ with probability 1. Since N can be made arbitrarily small and centred at γ^*, it implies that $\hat{\gamma} \xrightarrow{p} \gamma^*$.

It can be shown that the event E_T implies

$$Q_T(\gamma^*) < Q(\gamma^*) + \frac{\xi}{2}, \tag{A.7}$$

and

$$Q(\hat{\gamma}) - \frac{\xi}{2} < Q_T(\hat{\gamma}). \tag{A.8}$$

Now since $Q_T(\hat{\gamma}) \leq Q_T(\gamma^*)$, the inequality in (A.8) implies that

$$Q(\hat{\gamma}) - \frac{\xi}{2} < Q_T(\gamma^*). \tag{A.9}$$

Combining (A.7) and (A.9), we have that $Q(\hat{\gamma}) - Q(\gamma^*) < \xi$ which in turn implies that $\hat{\gamma} \in N$. Therefore, $\Pr(E_T) \leq \Pr(\hat{\gamma} \in N)$. Since $\Pr(E_T) \to 1$ by the uniform convergence in (A.6), we have $\Pr(\hat{\gamma} \in N) \to 1$, which completes the proof.

PROOF OF LEMMA 2. The proof is very similar to that of Theorem 2, so we only provide a brief sketch of the proof. Define $X_t = [y^d_{t-1}, \Delta y^d_{t-1}, ..., \Delta y^d_{t-p+1}]'$, $\hat{\delta} = [\hat{\rho}, \hat{\varsigma}_1, ..., \hat{\varsigma}_{p-1}]'$ and $\delta = [\rho_i - 1, \varsigma_{i,1}, ..., \varsigma_{i,p-1}]'$.

First, consider the case where $\tau \leq \tau_i$. Observe first that the ADF-type t-statistic from (6.4), scaled by $T^{-1/2}$, can be written as

$$\begin{aligned} & T^{-1/2} DF_G(\lambda, \tau) \\ = \; & (\tau - \lambda)^{1/2} \frac{e_1' \left\{ ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} X_t X_t'^{-1} \right\}^{-1} ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} X_t \Delta y_t^d}{\hat{\sigma}_T \left[e_1' \left\{ ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} X_t X_t' \right\}^{-1} e_1 \right]^{1/2}} \end{aligned}$$

where $\hat{\sigma}^2_T$ is the usual OLS variance estimator from the regression residuals. It is easily demonstrated that

$$\begin{aligned} ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} X_t X_t'^{-1} \; & \Rightarrow \; A_i(\lambda), \\ ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} X_t \Delta y_t^d \; & \Rightarrow \; B_i, \\ \hat{\sigma}^2_T \; & \Rightarrow \; -2e_1' B_i - B_i' A_i(\lambda)^{-1} B_i \equiv \sigma^2(\lambda), \end{aligned}$$

where B_i is defined in the theorem and $A_i(\lambda)$ coincides with A_i (defined in the theorem), excepting its $(1,1)$ element which is given by $\gamma_{i,0} + u^2_{i,\infty} 1[\lambda \in (\tau_{i-1}, \tau_i)]$. These results imply that $T^{-1/2} DF_G(\lambda, \tau) \Rightarrow -(\tau - \lambda)^{1/2} \Theta_i(\lambda)$ where $\Theta_i(\lambda) = -\frac{e_1' A_i(\lambda)^{-1} B_i}{\sigma(\lambda) \left[e_1' A_i(\lambda)^{-1} e_1 \right]^{1/2}}$.

Consider now the case where $\tau > \tau_i$. Here it is straightforward to show that $DF_G(\lambda, \tau) = O_p(1)$ because a non-zero portion of $I(1)$ observations are included in $DF_G(\lambda, \tau)$. Therefore, it follows that $T^{-1/2} DF_G(\lambda, \tau) \Rightarrow 0$.

Combining the foregoing results, we have that

$$T^{-1/2} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) \Rightarrow \inf_{\tau \in (\lambda, \tau_i]} \left\{ -(\tau - \lambda)^{1/2} \Theta_i(\lambda) \right\}.$$

The function $\Theta_i(\lambda)$ does not depend on τ, moreover it is positive the limit. To show the latter result, notice first that $T^{1/2}(\hat{\delta} - \delta) = O_p(1)$, and, hence,

$$T^{-1/2} DF_G(\lambda, \tau) = (\rho_i - 1) \hat{\sigma}_T^{-1} \left[e_1' \left\{ ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} X_t X_t' \right\}^{-1} e_1 \right]^{-1/2} + o_p(1).$$

Now since $\rho_i - 1 < 0$ and the terms $\hat{\sigma}_T$ and $e_1' \left\{ ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} X_t X_t' \right\}^{-1} e_1$ are both positive and bounded in probability, the limit $-(\tau - \lambda)^{1/2} \Theta_i(\lambda)$ is therefore negative which, in turn, implies that $\Theta_i(\lambda)$ is positive. Consequently, $-(\tau - \lambda)^{1/2} \Theta_i(\lambda)$ is monotonically decreasing in τ over the interval $(\lambda, \tau_i]$ and, hence,

$$T^{-1/2} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) \Rightarrow -(\tau_i - \lambda)^{1/2} \Theta_i(\lambda).$$

Now since $\lambda \in [\tau_{i-1}, \tau_i)$, we observe that $-(\tau_i - \lambda)^{1/2} \Theta_i(\lambda)$ takes its infimum value λ at the lower change point, τ_{i-1}. Therefore,

$$T^{-1/2} \inf_{\lambda \in [\tau_{i-1}, \tau_i)} \inf_{\tau \in (\lambda, 1]} DF_G(\lambda, \tau) \Rightarrow -(\tau_i - \tau_{i-1})^{1/2} \Theta_i(\tau_{i-1}).$$

Introducing the notation $\Theta_i = \Theta_i(\tau_{i-1})$ and $A_i = A_i(\tau_{i-1})$, the proof of the lemma is completed provided we can show that $\sigma^2(\tau_{i-1}) = \sigma^2$. To demonstrate this result, notice from (A.3) that for $\lambda = \tau_{i-1}$ we have that $\hat{\beta}_0 = y_{\tau_{i-1}T} + O_p(T^{-1/2})$, which implies that $y_t^d = h_t + O_p(T^{-1/2})$. We therefore have that at the infimum points $\lambda = \tau_{i-1}$, $\tau = \tau_i$,

$$\begin{aligned}
\hat{\sigma}_T^2 &= ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} \hat{\varepsilon}_t^2 \\
&= ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} (\Delta y_t^d - X_t' \hat{\delta})^2 \\
&= ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} (\Delta h_t - X_t' \hat{\delta})^2 + o_p(1) \\
&= ((\tau - \lambda)T)^{-1} \sum_{t=\lambda T}^{\tau T} \varepsilon_t^2 + o_p(1) \Rightarrow \sigma^2
\end{aligned}$$

where the last line is due to the fact that $\Delta h_t - X_t' \hat{\delta} = \varepsilon_t + O_p(T^{-1/2})$.

References

Bai, J. (2000). Vector autoregressive models with structural change in regression coefficients and in variance-covariance matrices. *Annals of Economics and Finance* 1, 303-339.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47-78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18, 1-22.

Banerjee, A., Lumsdaine, R. and Stock, J. (1992). Recursive and sequential tests of the unit root and trend break hypotheses: theory and international evidence. *Journal of Business and Economics Statistics* 10, 271-288.

Busetti, F. and Taylor, A.M.R. (2004). Tests of stationarity against a change in persistence. *Journal of Econometrics* 123, 33-66.

Campbell, Y.J. and Shiller, R.J. (1987). Cointegration and tests of present value models. *Journal of Political Economy* 95, 1062-88.

Campbell, Y.J. and Shiller, R.J. (2001). Valuation ratios and the long-run stock market outlook: an update. *National Bureau of Economic Research, Working Paper Series, #8221.*

Chong, T. T-L. (2001). Structural changes in AR(1) models. *Econometric Theory* 17, 87-155.

Craine, R. (1993). Rational bubbles: a test. *Journal of Economic Dynamics and Control* 17, 829-45.

Clarida, R.H., Gali, J. and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *Quarterly Journal of Economics* 115, 147-80.

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64, 813-836.

Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2006). Modified tests for a change in persistence. *Journal of Econometrics* 134, 441-469.

Kapetanios, G. and Shin, Y. (2003). Unit Root Tests in Three Regime SETAR Models. Available from `http://lubswww.leeds.ac.uk/MKB/ys100t/UP.htm`

Kapetanios, G., Shin, Y. and Snell, A. (2003). Testing for a Unit Root in the Nonlinear STAR Framework. *Journal of Econometrics* 112, 359-379.

Kim, J.Y. (2000). Detection of change in persistence of a linear times series. *Journal of Econometrics* 95, 97-116.

Kim, J., Belaire Franch, J. and Badillo Amador, R. (2002). Corrigendum to "Detection of change in persistence of a linear time series." *Journal of Econometrics* 109, 389-392.

Kurozumi, E. (2005). Detection of structural change in the long-run persistence in a univariate time series. *Oxford Bulletin of Economics and Statistics* 67, 181-206.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159-178.

Leybourne, S., Kim, T., Smith, V. and Newbold, P. (2003). Tests for a change in persistence against the null of difference-stationarity. *Econometrics Journal* 6, 291-311.

Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268-281.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361-1401.

Perron, P. (1997). Further evidence of breaking trend functions in macroeconomic variables. *Journal of Econometrics* 80, 355-385.

Perron, P. and Rodríguez, G. (2003). GLS detrending, efficient unit root tests and structural change. *Journal of Econometrics* 115, 1-27.

Pesaran, M.H., Pettenuzzo, D. and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, forthcoming.

Phillips, P.C.B. and Xiao, Z. (1998). A primer on unit root testing. *Journal of Economic Surveys* 12, 423-470.

Shiller, R.J. (2000). *Irrational Exuberance.* Princeton University Press, Princeton, New Jersey.

Sollis, R. (2006). Testing for bubbles: an application of tests for a change in persistence. *Applied Financial Economics* 16, 491-498.

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

Taylor, A.M.R. (2000). The finite sample effects of deterministic variables on conventional methods of lag selection in unit root tests. *Oxford Bulletin of Economics and Statistics* 62, 293-304.

Taylor, A.M.R. (2005). Fluctuation tests for a change in persistence. *Oxford Bulletin of Economics and Statistics* 67, 207-230.

Zivot, E. and Andrews, D.W.K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics* 10, 251-270.