ELSEVIER

Journal of Econometrics 109 (2002) 365–387

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Unit root tests with a break in innovation variance

Tae-Hwan Kim, Stephen Leybourne, Paul Newbold *

School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK

Abstract

It is shown that an abrupt change in the innovation variance of an integrated process can generate spurious rejections of the unit root null hypothesis in routine applications of Dickey–Fuller tests. We develop and investigate modified test statistics, based on unit root tests of Perron for a time series with a changing level, or changing intercept and slope, which are applicable when there is a change in innovation variance of an unknown magnitude at an unknown location.

JEL classification: C12; C15

Keywords: Dickey–Fuller tests; Integrated processes; Perron tests; Structural break

1. Introduction

A few authors have analysed the possibility of breaks in the variance of a time series. For example, Wichern et al. (1976) considered maximum likelihood estimation of an unknown break point in the variance of a first order autoregression, while Hsu (1977) proposed tests for the existence of a break, at an unknown point in time, in the variance of a sequence of independent normal random variables. Inclán (1993) used Bayesian methods to detect multiple breaks in variance in a time series. However, relatively little attention has been paid to the possibility of a break in the innovation variance of an integrated process, and to the impact of such a break on testing the null hypothesis of a unit autoregressive root. An exception is Hamori

* Corresponding author. Tel.: +0115-951-5151; fax: +0115-951-4159.

E-mail addresses: tae-hwan.kim@nottingham.ac.uk (T.-H. Kim), steve.leybourne@nottingham.ac.uk (S. Leybourne), paul.newbold@nottingham.ac.uk (P. Newbold).

PII: S0304-4076(02)00118-5

and Tokihisa (1997). These authors considered Dickey–Fuller tests based only on the regression with no constant and trend, concentrating on the case of an *increase* in innovation variance, reporting a moderate tendency to spuriously reject the unit root hypothesis.

The no constant, no trend model is of very limited practical value, as it implies that, under the alternative hypothesis of trend stationarity, the generating process is known to have mean zero. Unfortunately, the results reported by Hamori and Tokihisa for the simple model turn out to be unreliable predictors, both qualitatively and quantitatively, of what is found when either a constant or a linear trend is incorporated into the Dickey–Fuller regression. In Section 2 of the paper, we analyse the former case in detail and note simulation evidence of very similar conclusions for the latter. In short, we find quite severe spurious rejections of the unit root null hypothesis when there is a relatively early *decrease* in the innovation variance. [1]

Having demonstrated the phenomenon of spurious rejections by Dickey–Fuller tests in the presence of an innovation variance shift, a result which complements the analysis of a trend shift in Leybourne et al. (1998), the remainder of the paper is devoted to the development of modified tests that allow for a possible change in innovation variance at an unknown point in time. Section 3 of the paper discusses the estimation of the break point when a break occurs and also considers the behaviour of break point estimators when there is no break, while Section 4 develops a modified Perron-style unit root test and derives the asymptotic null distribution of this test statistic in the presence of a change in variance. We assess the finite sample size and power of the new test through simulation experiments. Section 5 discusses the extension to the case where a linear trend is incorporated in the model.

2. Spurious rejections in Dickey–Fuller tests

Consider a DGP given by

$$y_t = \mu + z_t, \quad t = 1, \ldots, T, \tag{1}$$

$$z_t = \rho z_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta z_{t-j} + \varepsilon_t, \quad \varepsilon_t = \sigma_t \eta_t, \tag{2}$$

[1] In a footnote, Hamori and Tokihisa appear to suggest, on theoretical grounds, that for their simple model a decrease in variance will lead to *under-rejection* of the null hypothesis. In fact, in simulations, not reported here, of that model based on series of 100 observations, we were unable to confirm that prediction, finding instead a modest tendency to over-reject, particularly for relatively early breaks. Again, however, the phenomenon is far less severe for the simple model than occurs for the more widely used models discussed in Section 2.

where lagged changes are incorporated in (2) to account for serial correlation, all roots of $1 - \sum_{j=1}^{p-1} \phi_j x^j = 0$ have modulus greater than unity, σ_t^2 is defined by

$$\sigma_t^2 = \sigma_1^2 1[t \leqslant \tau^* T] + \sigma_2^2 1[t > \tau^* T] \tag{3}$$

and η_t is assumed to satisfy the following assumption:

Assumption 1. η_t is a martingale difference sequence and satisfies $\mathrm{E}(\eta_t^2|\eta_{t-1},\ldots) = 1$ and $\mathrm{E}(|\eta_t|^{4+\gamma}|\eta_{t-1},\ldots) = \kappa < \infty$ *for some* $\gamma > 0$.

Thus there is a break in the variance of the innovation process ε_t at time $\tau^* T$, the variance changing from σ_1^2 to σ_2^2. The t-ratio variant of the Dickey–Fuller test is based on a fitted autoregression. The asymptotic distribution of this statistic under the null $\rho = 1$ will be derived for illustration for the case where the Dickey–Fuller regression includes a constant but no trend and $p = 1$. That limiting distribution, which involves both the break fraction τ^* and the ratio of the innovation standard deviations σ_2/σ_1, leads to a prediction that the test will spuriously reject the unit root null hypothesis when there is an abrupt *decrease* in innovation variance, most seriously so when the break is relatively early.

2.1. Asymptotic null distribution of the Dickey–Fuller statistic

The t-ratio variant of the Dickey–Fuller statistic, denoted t_{O}, in the case $p = 1$ for testing $\rho = 1$, with a constant term included, is based on estimating the OLS regression model

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t. \tag{4}$$

The following theorem gives the asymptotic null distribution of t_{O} under the DGP (1)–(3).

Theorem 1. *Under the DGP* (1)–(3) *with* $\rho = 1$, $p = 1$, *and Assumption* 1,

$$t_{\mathrm{O}} \Rightarrow \frac{1}{\{\tau^* + (1-\tau^*)\delta^2\}^{1/2}} \frac{a(\delta,\tau^*)}{b(\delta,\tau^*)^{1/2}}, \tag{5}$$

where $\delta = \sigma_2/\sigma_1$,

$$a(\delta,\tau^*) = \frac{1}{2}\{W(1)^2 - 1\} - (\delta^2 - 1)\frac{1}{2}\{W(1)^2 - W(\tau^*)^2 - (1-\tau^*)\}$$
$$- \delta(\delta - 1)W(\tau^*)\{W(1) - W(\tau^*)\}$$

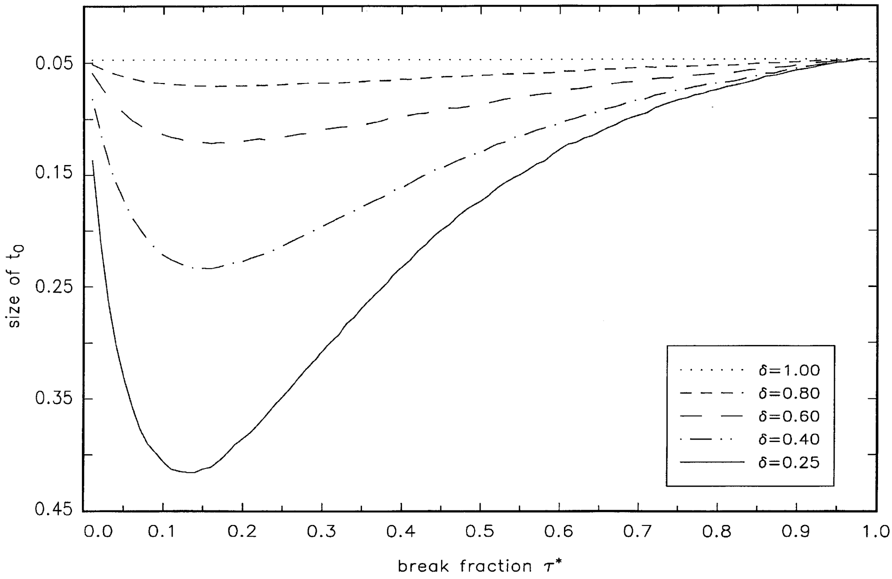

Fig. 1. Asymptotic size of nom. 0.05 level t_0-test with decrease in variance.

$$-\left[\int_0^1 W(r)\,\mathrm{d}r+(\delta-1)\left\{\int_{\tau^*}^1 W(r)\,\mathrm{d}r-(1-\tau^*)W(\tau^*)\right\}\right]$$

$$\times[W(1)+(\delta-1)\{W(1)-W(\tau^*)\}],$$

$$b(\delta,\tau^*)=\int_0^1 W(r)^2\,\mathrm{d}r+(\delta^2-1)\int_{\tau^*}^1 W(r)^2\,\mathrm{d}r$$

$$-2\delta(\delta-1)W(\tau^*)\int_{\tau^*}^1 W(r)\,\mathrm{d}r+(\delta-1)^2(1-\tau^*)W(\tau^*)^2$$

$$-\left[\int_0^1 W(r)\,\mathrm{d}r+(\delta-1)\left\{\int_{\tau^*}^1 W(r)\,\mathrm{d}r-(1-\tau^*)W(\tau^*)\right\}\right]^2.$$

Here $W(r)$ is a standard Brownian motion process.

2.2. Simulation evidence

We simulated the limiting functionals given in Theorem 1, using 40,000 replications and series of 5,000 Gaussian white-noise innovations, and thus computed the asymptotic size of nominal 0.05-level Dickey–Fuller tests based on regression (4). Results for a full range of break fractions τ^*, and for various values of the standard deviations ratio $\delta \leqslant 1$ are graphed in Fig. 1. These imply potentially serious spurious rejections when there is a relatively early decrease in innovation variance. [2] However, results not reported here for the case $\delta > 1$ indicated no substantial deviation of asymptotic sizes from the nominal 0.05-level.

3. Break estimation

In view of the spurious rejection problem associated with the Dickey–Fuller test uncovered in the previous section, we now seek an alternative test that is valid in the presence of a break in innovation variance. We do not assume that the location of any break is known, but consider in this section the problem of estimating that location, proposing two possible estimators. While the problem of testing for a break in mean at an unknown point has received considerable attention in the literature, rather less attention has been paid to the problem of estimating the unknown break point in mean: procedures that have been proposed and analysed include the MLE method (Picard, 1985; Bhattacharya, 1987; Fu and Curnow, 1990), the LS method (Bai, 1993; Bai and Perron, 1998), the LAD method (Bai, 1995) and the QMLE method (Bai et al., 1998; Nunes et al., 1995). We adopt the QMLE method and the LS method to estimate the unknown break point in innovation variance.

3.1. QMLE method

We begin by treating ε_t as if it is normally distributed with zero mean and variance σ_t^2 in (2), (3). Then, the quasi-log likelihood is a function of τ, σ_1^2 and σ_2^2, but we can concentrate this function by writing it as a function of τ alone. It can be shown that the negative of the concentrated quasi-log likelihood function is proportional to

$$Q_T(\tau) = \tau \ln \sigma_1(\tau)^2 + (1-\tau) \ln \sigma_2(\tau)^2,$$

where

$$\sigma_1(\tau)^2 = (\tau T)^{-1} \sum_{1}^{\tau T} \varepsilon_t^2, \qquad \sigma_2(\tau)^2 = \{(1-\tau)T\}^{-1} \sum_{\tau T+1}^{T} \varepsilon_t^2. \tag{6}$$

This objective function can be calculated only when we know ε_t. In reality, it is not known and we modify the concentrated quasi-log likelihood function by replacing ε_t

[2] To verify the applicability of these findings in moderate-sized samples, we simulated series of $T = 100$ observations from the DGP (1)–(3), with $\rho = 1$, $p = 1$ and η_t generated as standard normal, using 40,000 replications. Empirical sizes of nominal 0.05-level tests were virtually indistinguishable from those of Fig. 1. As a check on the generality of this phenomenon, this experiment was repeated with a linear trend term included in regression (4), and the usual Dickey–Fuller with trend critical values. The results for rejection rates of nominal 0.05-level tests were both qualitatively and quantitatively similar to those for the constant only case.

with the residual e_t from the following OLS regression:

$$y_t = \hat{\alpha} + \hat{\rho} y_{t-1} + \sum_{j=1}^{p-1} \hat{\phi}_j \Delta y_{t-j} + e_t. \tag{7}$$

Let $0 < \tau_1 < \tau_2 < 1$. Then the break estimator is defined by the solution to the following minimisation problem:

$$\hat{\tau} = \arg\min_{\tau \in [\tau_1, \tau_2]} \hat{Q}_T(\tau),$$

$$\hat{Q}_T(\tau) = \tau \ln \hat{\sigma}_1(\tau)^2 + (1-\tau) \ln \hat{\sigma}_2(\tau)^2,$$

where

$$\hat{\sigma}_1(\tau)^2 = (\tau T)^{-1} \sum_{1}^{\tau T} e_t^2, \qquad \hat{\sigma}_2(\tau)^2 = \{(1-\tau)T\}^{-1} \sum_{\tau T+1}^{T} e_t^2. \tag{8}$$

Once we obtain the break estimator $\hat{\tau}$ in this manner, the respective variance estimators are defined by

$$\hat{\sigma}_1(\hat{\tau})^2 = (\hat{\tau} T)^{-1} \sum_{1}^{\hat{\tau} T} e_t^2, \qquad \hat{\sigma}_2(\hat{\tau})^2 = \{(1-\hat{\tau})T\}^{-1} \sum_{\hat{\tau} T+1}^{T} e_t^2. \tag{9}$$

The following theorem, proved in Appendix A, demonstrates consistency of these estimators.

Theorem 2. *Suppose that* $\tau^* \in (\tau_1, \tau_2)$. *Under the DGP* (1)–(3) *with* $\rho = 1$ *and Assumption* 1, *we have* (i) $\hat{\tau} - \tau^* = o_p(1)$, (ii) $\hat{\sigma}_1(\hat{\tau})^2 - \sigma_1^2 = o_p(1)$, *and* (iii) $\hat{\sigma}_2(\hat{\tau})^2 - \sigma_2^2 = o_p(1)$, *where* $\hat{\tau}$ *is the QMLE estimator.*

3.2. LS method

As in the QMLE method, we first consider the estimation of a break fraction τ^* assuming that we observe ε_t, and then we relax this assumption later. The basic idea is to transform the structural break in the variance of ε_t into a structural break in the mean of ε_t^2. Define $\xi_t = \varepsilon_t^2 - \mathrm{E}(\varepsilon_t^2)$. Then, we have

$$\varepsilon_t^2 = \sigma_t^2 + \xi_t,$$

where σ_t^2 is given in (3). Hence, we have a structural break in the mean of ε_t^2 at time $\tau^* T$. In this case, using the LS method suggested by Bai (1993) and Nunes et al. (1995), the LS break estimator $\hat{\tau}$ is defined as:

$$\hat{\tau} = \arg\min_{\tau \in [\tau_1, \tau_2]} S_T(\tau),$$

$$S_T(\tau) = \sum_{1}^{\tau T} \{\varepsilon_t^2 - \sigma_1(\tau)^2\}^2 + \sum_{\tau T+1}^{T} \{\varepsilon_t^2 - \sigma_2(\tau)^2\}^2,$$

where $\sigma_1(\tau)^2$ and $\sigma_2(\tau)^2$ are given in (6). It can be shown that the above LS minimization is equivalent to the following maximization problem:

$$\hat{\tau} = \arg\min_{\tau\in[\tau_1,\tau_2]} V_T(\tau)^2,$$

$$V_T(\tau)^2 = \{\tau(1-\tau)\}\{\sigma_2(\tau)^2 - \sigma_1(\tau)^2\}^2.$$

In the practically important case where ε_t is not observable, we use the residuals e_t from regression (7). Then, the break estimator is defined by

$$\hat{\tau} = \arg\min_{\tau\in[\tau_1,\tau_2]} \hat{V}_T(\tau)^2,$$

$$\hat{V}_T(\tau)^2 = \{\tau(1-\tau)\}\{\hat{\sigma}_2(\tau)^2 - \hat{\sigma}_1(\tau)^2\}^2,$$

where $\hat{\sigma}_1(\tau)^2$ and $\hat{\sigma}_2(\tau)^2$ are given in (8). The variance estimators are again calculated using the formulae in (9). The following theorem, proved in Appendix A, demonstrates consistency of these estimators.

Theorem 3. *Suppose that* $\tau^* \in (\tau_1, \tau_2)$. *Under the DGP* (1)–(3) *with* $\rho = 1$ *and Assumption* 1, *we have* (i) $\hat{\tau} - \tau^* = o_p(1)$, (ii) $\hat{\sigma}_1(\hat{\tau})^2 - \sigma_1^2 = o_p(1)$, *and* (iii) $\hat{\sigma}_2(\hat{\tau})^2 - \sigma_2^2 = o_p(1)$, *where* $\hat{\tau}$ *is the LS estimator.*

In this LS approach, we can also show the limiting behaviour of the break estimator $\hat{\tau}$ when there is no break in the data generating process in (3); that is $\tau^* \in \{0,1\}$. This issue is addressed in the following theorem.

Theorem 4. *Suppose that* $\tau^* \in \{0,1\}$. *Under the DGP* (1)–(3) *with* $\rho = 1$ *and Assumption* 1, *and if the process* $\{\eta_t^2 - 1\}$ *obeys the functional CLT*

(i) *If* $0 < \tau_1 < \tau_2 < 1$, *then*

$$\hat{\tau} \Rightarrow \arg\min_{\tau\in[\tau_1,\tau_2]} \frac{\sigma^4\gamma^2\{B(\tau) - \tau B(1)\}^2}{\tau(1-\tau)},$$

where $\gamma^2 = var(\eta_t^2)$ *and* $B(\tau)$ *is a standard Brownian motion defined as the limit of* $\gamma^{-1}T^{-1/2}\sum_1^{\tau T}(\eta_t^2 - 1)$.

(ii) *If* $\tau_1 = 0$ *and* $\tau_2 = 1$, *then* $\hat{\tau} \xrightarrow{p} \{0,1\}$.

It can further be shown, following part (ii) of Theorem 4, that in the no break case convergence to either endpoint is equally likely—that is the two probabilities are each 0.5. (That this is the case follows from noting that here the reversed series follows the same process as the series in forward time)

Theorems 2–4 establish consistency under the unit root null $\rho = 1$. Straightforward modifications to their proofs establish that our estimators are also consistent under the stationary alternative $|\rho| < 1$. In results not reported here we confirmed the conclusions of these theorems by simulating series of 1,000 observations with Gaussian innovations, noting further that the conclusion of Theorem 4 for the no break case appears also to apply to the QMLE estimator. The importance of these findings in the no break case is that a search for a break should be conducted over the entire range $0 \leqslant \tau \leqslant 1$. This is particularly so, since our unit root test, developed in the following section, has critical values that shift towards zero as $\hat{\tau}$ moves further from 0.5. In consequence, any trimming of the contemplated break interval is likely to lead to some loss in power of the test.

4. Tests based on GLS

We now turn to a modification of the Dickey–Fuller test, in circumstances where a break in variance, possibly of an uncertain extent and at an unknown location, has occurred. Perhaps a natural approach is through generalised least squares. However, we begin by showing that, even if the break date and two innovation variances are *known*, GLS generates a test statistic whose limiting null distribution still depends on the two innovation variances. We next show how the procedure can be amended to avoid this problem using the results of Perron (1989, 1990). Finally, a feasible procedure, employing the estimators of Section 3 is proposed and assessed.

4.1. Standard GLS

In this subsection we demonstrate that a direct application of the standard GLS does not deliver a desirable solution even when we know all the nuisance parameters τ^*, σ_1 and σ_2. This is because the asymptotic distribution of the t-statistic from the standard GLS regression still depends on the nuisance parameters. It is sufficient for this demonstration to consider just the case of no lagged changes, taking $p=1$ in (2). The GLS-transformed representation of (4) is seen to be

$$\tilde{y}_t(\tau^*) = \alpha d_t(\tau^*) + \rho \dot{y}_{t-1}(\tau^*) + \eta_t, \tag{10}$$

where

$$\tilde{y}_t(\tau^*) = \sigma_1^{-1} y_t 1[t \leqslant \tau^* T] + \sigma_2^{-1} y_t 1[t > \tau^* T],$$

$$d_t(\tau^*) = \sigma_1^{-1} 1[t \leqslant \tau^* T] + \sigma_2^{-1} 1[t > \tau^* T],$$

$$\dot{y}_{t-1}(\tau^*) = \sigma_1^{-1} y_{t-1} 1[t \leqslant \tau^* T] + \sigma_2^{-1} y_{t-1} 1[t > \tau^* T].$$

The GLS analogue of the Dickey–Fuller test t_{O} is the t-ratio for testing $\rho = 1$ when (10) is estimated by OLS. We denote this statistic as t_{G}. The following lemma shows that the asymptotic null distribution of t_{G} depends on τ^*, σ_1 and σ_2.

Lemma 1. *Under the DGP* (1)–(3) *with* $\rho = 1,\ p = 1$, *and Assumption* 1,

$$t_{\mathrm{G}} \Rightarrow \frac{c}{d^{1/2}},$$

where

$$c = (1/2)\{W(1)^2 - 1\} - \kappa W(\tau^*)\{W(1) - W(\tau^*)\}$$
$$- \left\{\sigma_1^{-1}\int_0^{\tau^*} W(r)\,\mathrm{d}r + \sigma_2^{-1}\int_{\tau^*}^1 W(r)\,\mathrm{d}r - \sigma_2^{-1}(1-\tau^*)\kappa W(\tau^*)\right\}$$
$$\times[\sigma_1^{-1}W(\tau^*) + \sigma_2^{-1}\{W(1) - W(\tau^*)\}],$$

$$d = \int_0^1 W(r)^2\,\mathrm{d}r - 2\kappa W(\tau^*)\int_\tau^1 W(r)\,\mathrm{d}r + \kappa^2(1-\tau^*)W(\tau^*)^2$$
$$- \left\{\sigma_1^{-1}\int_0^{\tau^*} W(r)\,\mathrm{d}r + \sigma_2^{-1}\int_{\tau^*}^1 W(r)\,\mathrm{d}r - \sigma_2^{-1}(1-\tau^*)\kappa W(\tau^*)\right\}^2$$

and

$$\kappa = \delta^{-1}(\delta - 1), \qquad \delta = \sigma_2/\sigma_1.$$

The proof of this lemma is quite straightforward once we express $\tilde{y}_t(\tau^*)$ in the form

$$\tilde{y}_t(\tau^*) = (\mu + z_0)d_t + w_t - \kappa w_{\tau^* T}1[t > \tau^* T], \tag{11}$$

where $w_t = \sum_{i=1}^t \eta_i$. Furthermore, we may rewrite (11) as

$$\tilde{y}_t(\tau^*) = \tilde{y}_{t-1}(\tau^*) + c_1 1[t = \tau^* T + 1] + \eta_t,$$
$$c_1 = -\kappa\{(\mu + z_0)\sigma_1^{-1} + w_{\tau^* T}\} \tag{12}$$

which allows us to see heuristically why the asymptotic distribution of t_{G} will depend on κ. This is caused by the presence at time $\tau^* T + 1$ of the *stochastic* quantity $-\kappa w_{\tau^* T}$ of c_1. This term is $O_p(T^{1/2})$ and its effect does not vanish asymptotically. Hence, a straightforward GLS treatment of the usual Dickey–Fuller regression is inadequate, as different critical values would be required for each κ. It is, therefore, necessary to modify the GLS regression to allow for a change in level. We do this in the next subsection.[3]

[3] We note that the same conclusion applies also in the case of a Dickey–Fuller regression with no intercept or with intercept and trend, where the limiting distribution of the test statistic still depends on the magnitude of any change in innovation variance.

4.2. Feasible modified GLS

In view of (12) we are led to consider a modified GLS-transformed regression. We deal here with the more general generating process (1)–(3) for any finite positive integer p. Consider the behaviour of $\tilde{y}_t(\tau^*)$, defined in the previous subsection. Clearly we have, after adding $\sum_{j=1}^{p-1} \phi_j \Delta y_{t-j}$ to the right-hand side of (4),

$$\tilde{y}_t(\tau^*) = \begin{cases} \alpha\sigma_1^{-1} + \rho\tilde{y}_{t-1}(\tau^*) + \sum_{j=1}^{p-1} \phi_j \Delta\tilde{y}_{t-j}(\tau^*) + \eta_t, \ t \leqslant \tau^* T, \\ \alpha\sigma_2^{-1} + \rho\tilde{y}_{t-1}(\tau^*) + \sum_{j=1}^{p-1} \phi_j \Delta\tilde{y}_{t-j}(\tau^*) + \eta_t, \ t \geqslant \tau^* T + p + 1. \end{cases} \tag{13}$$

The two regimes in (13) are identical except that their intercepts differ. However, in the intermediate segment, $t = \tau^* T + 1, \ldots, \tau^* T + p$, the quantity $\tilde{y}_t(\tau^*) - \rho\tilde{y}_{t-1}(\tau^*) - \sum_{j=1}^{p-1} \phi_j \Delta\tilde{y}_{t-j}(\tau^*) - \eta_t$ is stochastic, taking a different value at each time period, and so belongs to neither of the regimes in (13). This conclusion can be summarised by introducing a one-time dummy variable and its lags to cover the intermediate period. We can then write, incorporating also a dummy for the change in intercept,

$$\tilde{y}_t(\tau^*) = \alpha_0 + \alpha_1 d_{1t}(\tau^*) + \alpha_2 d_{2t}(\tau^*) + \sum_{j=1}^{p-1} \theta_j d_{2,t-j}(\tau^*) + \rho\tilde{y}_{t-1}(\tau^*) + \sum_{j=1}^{p-1} \phi_j \Delta\tilde{y}_{t-j}(\tau^*) + \eta_t, \tag{14}$$

where

$$d_{1t}(\tau^*) = 1[t > \tau^* T], \qquad d_{2t}(\tau^*) = 1[t = \tau^* T + 1].$$

This specification is precisely that of Perron (1990) in the case of a change in level of the innovational outlier type, except that we require also lags of the one time dummy when $p > 1$. This latter factor does not alter the limiting null distribution of unit root test statistics as it implies in effect simply the removal of a further $(p-1)$ central data points from the Perron specification. Note that the need for one-time dummies here is similar in spirit to the same need in the additive outlier model considered by Perron and Vogelsang (1992).

As a practical matter, the true break and the standard deviations σ_1 and σ_2 will be unknown. However, these can be consistently estimated through the procedures of the previous section, so that tests are based on the fitted regression

$$\tilde{y}_t(\hat{\tau}) = \hat{\alpha}_0 + \hat{\alpha}_1 d_{1t}(\hat{\tau}) + \hat{\alpha}_2 d_{2t}(\hat{\tau}) + \sum_{j=1}^{p-1} \hat{\theta}_j d_{2,t-j}(\hat{\tau})$$

$$+ \hat{\rho}\tilde{y}_{t-1}(\hat{\tau}) + \sum_{j=1}^{p-1} \hat{\phi}_j \Delta\tilde{y}_{t-j}(\hat{\tau}) + \hat{\eta}_t, \tag{15}$$

where

$$\tilde{y}_t(\hat{\tau}) = \hat{\sigma}_1^{-1} y_t 1[t \leqslant \hat{\tau}T] + \hat{\sigma}_2^{-1} y_t 1[t > \hat{\tau}T].$$

Use of consistent estimators in place of unknown parameters will not affect the limiting null distribution of unit root test statistics, which remain as given by Perron (1990). For this feasible procedure, let t_F denote the t-ratio associated with the test of the null hypothesis $\rho = 1$ when (15) is estimated by OLS. The following theorem gives the limiting null distribution of this statistic.

Theorem 5. *Under the DGP* (1)–(3) *with* $\rho = 1$ *and Assumption* 1, *given a break fraction* $\tau^* \in (0,1)$,

$$t_F \Rightarrow \frac{f(\tau^*)}{g(\tau^*)^{1/2}}$$

where

$$f(\tau^*) = (1/2)\{W(1)^2 - 1\} - \tau^{*-1} W(\tau^*) \int_0^{\tau^*} W(r)\,\mathrm{d}r$$

$$- (1-\tau^*)^{-1}\{W(1) - W(\tau^*)\} \int_{\tau^*}^1 W(r)\,\mathrm{d}r,$$

$$g(\tau^*) = \int_0^1 W(r)^2\,\mathrm{d}r - \tau^{*-1}\left\{\int_0^{\tau^*} W(r)\,\mathrm{d}r\right\}^2 - (1-\tau^*)^{-1}\left\{\int_{\tau^*}^1 W(r)\,\mathrm{d}r\right\}^2.$$

The limit distribution is that given in Eq. (8) of Perron (1990) and depends only on τ^*, and not the innovation variances (σ_1^2, σ_2^2). Critical values of this limit distribution are provided in Perron (1990), Table 4. One can thus base the test on Perron's critical values, which depend to some extent on τ^*, but not on σ_2/σ_1. Appropriate critical values for our test are found by reference to the consistent estimator $\hat{\tau}$ of τ^*. When no break occurs, as we saw in the previous section $\hat{\tau}$ tends in probability to $\{0,1\}$, so that our test statistic tends to the same limiting distribution as the Dickey–Fuller statistic.

4.3. Monte Carlo simulation

In this section we consider the finite sample size and power properties of the test t_F. That is, we assume that τ^*, σ_1 and σ_2 are unknown and need to be estimated.

Table 1
Size of t_F at nominal 0.05-level critical value, $p = 1$

δ	1.00	0.80	0.60	0.40	0.25
(a) $T = 100$					
τ^*					
0.20	0.050	0.061	0.068	0.063	0.063
0.40	—	0.054	0.052	0.055	0.054
0.60	—	0.047	0.047	0.055	0.056
0.80	—	0.045	0.043	0.050	0.053
(b) $T = 200$					
τ^*					
0.20	0.048	0.060	0.061	0.056	0.056
0.40	—	0.052	0.048	0.050	0.052
0.60	—	0.048	0.049	0.055	0.055
0.80	—	0.046	0.049	0.051	0.053

Data were generated from the DGP (1)–(3) with standard normal η_t, for values of δ chosen to match those used in Fig. 1. In estimating τ^*, we report results only for the QMLE estimator of Section 3. Results based on the LS estimator were very similar, though here nominal and empirical sizes of unit root tests differed a little more. We searched the entire range $0 \leqslant \tau \leqslant 1$ of possible break fractions, which produces tests that are a little more powerful than if this interval is truncated. Finite sample critical values of the test statistic for $\tau^* = 0.1(0.1)0.9$ are given in Table 4 of Perron (1990), while corresponding values of the Dickey–Fuller statistic are appropriate for $\tau^* = \{0, 1\}$. For any estimate $\hat{\tau}$, we then interpolated between these critical values.

We first considered the simplest case, where lagged changes are not required in regression (15); that is, $p=1$ is appropriately chosen. To verify that our statistic has the correct size in particular cases, series of 100 and 200 observations were generated from random walks with various break fractions τ^* and ratios of standard deviations δ. Here and in subsequent simulations, results are based on 5,000 replications. Empirical sizes of our test statistic are given in Table 1 for nominal 0.05-level tests. Empirical and nominal sizes are satisfactorily close. Table 2 reports powers of the same test when data are generated by a first order autoregression with $\rho=0.8, 0.9$ for $T=100$ and $\rho=0.9, 0.95$ for $T = 200$. As one would expect, power increases both with increasing sample size and increasing distance from 1 of the autoregressive parameter. The test power depends quite heavily on δ, and to a lesser extent on τ^*. The former follows from the fact that, the larger is any break, the more precisely will the break fraction be estimated, the latter may in part reflect the apparent relationship between test sizes and τ^* seen in Table 1.

Although it can be seen from Table 1 that the statistic t_F has the correct size when there is no break in innovation variance ($\delta = 1.00$), there is necessarily some loss in

Table 2
Power of t_F at nominal 0.05-level critical value, $p = 1$

	δ	1.00	0.80	0.60	0.40	0.25
(a) $T = 100$						
τ^*						
0.20	$\rho = 0.90$	0.256	0.267	0.279	0.340	0.465
	$\rho = 0.80$	0.772	0.767	0.770	0.814	0.870
0.40	$\rho = 0.90$	—	0.243	0.249	0.280	0.387
	$\rho = 0.80$	—	0.742	0.710	0.752	0.826
0.60	$\rho = 0.90$	—	0.244	0.224	0.258	0.350
	$\rho = 0.80$	—	0.746	0.710	0.749	0.813
0.80	$\rho = 0.90$	—	0.253	0.242	0.264	0.308
	$\rho = 0.80$	—	0.760	0.752	0.784	0.834
(b) $T = 200$						
τ^*						
0.20	$\rho = 0.95$	0.261	0.269	0.285	0.343	0.472
	$\rho = 0.90$	0.770	0.763	0.768	0.816	0.871
0.40	$\rho = 0.95$	—	0.237	0.246	0.288	0.404
	$\rho = 0.90$	—	0.719	0.710	0.763	0.831
0.60	$\rho = 0.95$	—	0.236	0.239	0.274	0.369
	$\rho = 0.90$	—	0.718	0.705	0.754	0.826
0.80	$\rho = 0.95$	—	0.254	0.263	0.280	0.322
	$\rho = 0.90$	—	0.749	0.748	0.796	0.844

power compared with the usual Dickey–Fuller test, which would be appropriate in this case. For example, we found for the latter with $T=100$ and $\rho=0.8$ power to be 0.862. This is of course to be expected as t_F is not constructed with optimality in this situation foremost in mind, but from consideration of size robustness under an unknown break in innovation variance. Its gain over the Dickey–Fuller test in size reliability under such a break far outweighs the modest sacrifice in power when there is no break—that is, it seems well worth accepting slightly diminished power for the avoidance of spurious rejections.

Tables 3 and 4 report on the case where a lagged difference is included in the fitted model, so that $p=2$ in (15). For Table 3, the generating model was the ARIMA$(1,1,0)$ process

$$\Delta y_t = 0.7\Delta y_{t-1} + \varepsilon_t$$

with break specifications exactly as in the previous two tables. Again our test appears to exhibit good size reliability. The generating process for Table 4 is the same first order autoregression as for Table 2. Unsurprisingly, we note a small drop in power when the augmented variant of the test is employed.

Table 3
Size of t_F at nominal 0.05-level critical value, $p=2$

δ	1.00	0.80	0.60	0.40	0.25
(a) $T=100$					
τ^*					
0.20	0.048	0.063	0.068	0.062	0.058
0.40	—	0.056	0.061	0.063	0.060
0.60	—	0.049	0.053	0.058	0.061
0.80	—	0.046	0.048	0.048	0.049
(b) $T=200$					
τ^*					
0.20	0.051	0.060	0.058	0.058	0.056
0.40	—	0.052	0.051	0.054	0.051
0.60	—	0.045	0.050	0.056	0.055
0.80	—	0.046	0.051	0.054	0.054

Table 4
Power of t_F at nominal 0.05-level critical value, $p=2$

	δ	1.00	0.80	0.60	0.40	0.25
(a) $T=100$						
τ^*						
0.20	$\rho=0.90$	0.229	0.241	0.255	0.300	0.428
	$\rho=0.80$	0.653	0.636	0.646	0.701	0.781
0.40	$\rho=0.90$	—	0.224	0.225	0.249	0.338
	$\rho=0.80$	—	0.617	0.602	0.640	0.721
0.60	$\rho=0.90$	—	0.223	0.203	0.227	0.290
	$\rho=0.80$	—	0.614	0.593	0.631	0.702
0.80	$\rho=0.90$	—	0.218	0.207	0.226	0.265
	$\rho=0.80$	—	0.635	0.617	0.647	0.707
(b) $T=200$						
τ^*						
0.20	$\rho=0.95$	0.240	0.246	0.267	0.315	0.423
	$\rho=0.90$	0.703	0.700	0.708	0.757	0.824
0.40	$\rho=0.95$	—	0.223	0.220	0.262	0.377
	$\rho=0.90$	—	0.638	0.624	0.683	0.774
0.60	$\rho=0.95$	—	0.213	0.212	0.245	0.328
	$\rho=0.90$	—	0.644	0.637	0.678	0.771
0.80	$\rho=0.95$	—	0.225	0.228	0.257	0.303
	$\rho=0.90$	—	0.676	0.689	0.724	0.786

5. The linear trend case

As it stands, the test t_F is invariant to μ for $\rho \leqslant 1$. However, in practice, we often require invariance to a linear trend. We therefore consider a second test statistic, which we denote t_{tF}. In the contemplated data-generating process, (1) is replaced by

$$y_t = \mu + \beta t + z_t, \quad t = 1, \ldots, T,$$

while retaining (2) and (3). In the case where $\rho = 1$ and $p = 1$, we may show that

$$\tilde{y}_t(\tau^*) = c_2 + \tilde{y}_{t-1}(\tau^*) + c_3 1[t = \tau^* T + 1] + c_4 1[t > \tau^* T] + \eta_t,$$

$$c_2 = \beta\sigma_1^{-1},$$

$$c_3 = -\kappa\{w_{\tau^* T} + \sigma_1^{-1}(\mu + z_0 + \beta\tau^* T)\},$$

$$c_4 = -\beta\sigma_1^{-1}\kappa$$

which demonstrates the need to allow now for a change in drift in addition to a change in level.

Thus, following through the same argument as in Section 4.2, the two right-hand expressions in (13) need to be augmented to allow for a change in slope as well as intercept. Thus, following the addition to the right-hand side of (4) of $\lambda t + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j}$, the terms $\lambda\sigma_i^{-1} t$ (i=1, 2) are added to the respective right-hand side expressions in (13). The argument for the inclusion of a one-time dummy and its lags remains unchanged, so that in place of (15) the fitted test regression is

$$\tilde{y}_t(\hat{\tau}) = \hat{\alpha}_0 + \hat{\alpha}_1 d_{1t}(\hat{\tau}) + \hat{\alpha}_2 d_{2t}(\hat{\tau}) + \hat{\alpha}_3 d_{3t}(\hat{\tau}) + \hat{\alpha}_4 t + \sum_{j=1}^{p-1} \hat{\theta}_j d_{2,t-j}(\hat{\tau})$$

$$+ \hat{\rho}\tilde{y}_{t-1}(\hat{\tau}) + \sum_{j=1}^{p-1} \hat{\phi}_j \Delta\tilde{y}_{t-j}(\hat{\tau}) + \hat{\eta}_t$$

where

$$d_{3t}(\tau^*) = (t - \tau^* T)1[t > \tau^* T].$$

Here, the unknown break fraction τ^* can be consistently estimated by the two procedures discussed in Section 3, with the obvious modification that a term $\hat{\lambda} t$ is added to the right-hand side of (7). The statistic t_{tF} is then the usual t-ratio for testing $\rho = 1$. Apart from the lagged one time dummies, whose necessary inclusion again does not affect the limiting null distribution, the set-up here is precisely that of the "Model (C)" unit root test of Perron (1989). The limiting null distribution is given in Theorem 2 of that paper, and critical values of that distribution are given there in Table VI.B.

In Tables 5 and 6, we investigate finite sample size and power of this test for samples of 100 and 200 observations. Our approach is precisely as in Tables 1 and 2 of Section 4.3, the data-generating processes being exactly as there. We simulated finite sample critical values of Perron statistics for $\tau^* = 0.1(0.1)0.9$, rather than using the asymptotic values, and again used the Dickey–Fuller (now with trend) critical values for $\tau^* = \{0, 1\}$. Table 5 suggests that our test has quite good size properties, while comparison of Table 6 with Table 2 indicates, as is generally found with Dickey–Fuller-type tests, a substantial fall in power when a linear trend term is included.

Table 5
Size of t_{tF} at nominal 0.05-level critical value, $p = 1$

δ	1.00	0.80	0.60	0.40	0.25
(a) $T = 100$					
τ^*					
0.20	0.044	0.044	0.050	0.051	0.055
0.40	—	0.041	0.041	0.047	0.050
0.60	—	0.040	0.043	0.047	0.051
0.80	—	0.038	0.037	0.040	0.041
(b) $T = 200$					
τ^*					
0.20	0.050	0.056	0.050	0.051	0.053
0.40	—	0.048	0.048	0.050	0.051
0.60	—	0.046	0.045	0.049	0.050
0.80	—	0.044	0.046	0.048	0.049

Table 6
Power of t_{tF} at nominal 0.05-level critical value, $p = 1$

	δ	1.00	0.80	0.60	0.40	0.25
(a) $T = 100$						
τ^*						
0.20	$\rho = 0.90$	0.141	0.145	0.136	0.137	0.183
	$\rho = 0.80$	0.505	0.494	0.490	0.525	0.560
0.40	$\rho = 0.90$	—	0.130	0.119	0.123	0.140
	$\rho = 0.80$	—	0.461	0.404	0.428	0.503
0.60	$\rho = 0.90$	—	0.126	0.111	0.115	0.127
	$\rho = 0.80$	—	0.466	0.400	0.412	0.460
0.80	$\rho = 0.90$	—	0.123	0.116	0.126	0.137
	$\rho = 0.80$	—	0.470	0.442	0.464	0.478
(b) $T = 200$						
τ^*						
0.20	$\rho = 0.95$	0.156	0.143	0.134	0.146	0.172
	$\rho = 0.90$	0.483	0.482	0.458	0.481	0.558
0.40	$\rho = 0.95$	—	0.130	0.111	0.127	0.142
	$\rho = 0.90$	—	0.419	0.368	0.400	0.475
0.60	$\rho = 0.95$	—	0.126	0.107	0.118	0.120
	$\rho = 0.90$	—	0.408	0.366	0.385	0.442
0.80	$\rho = 0.95$	—	0.135	0.123	0.134	0.138
	$\rho = 0.90$	—	0.452	0.448	0.463	0.477

6. Conclusion

We have seen that, in the presence of a decrease in innovation variance, Dickey–Fuller tests for unit roots can lead to serious spurious rejections of the null hypothesis. We propose instead tests based on the prior estimation of the break point, if such a point exist, and of the variances of the innovations of the two parts of the series. These estimates are then employed in modified variants of the tests of Perron (1989, 1990) for unit roots in the presence of changes in mean or trend. Simulation evidence indicates that our tests have good size and power properties in conventionally employed sample sizes.

Acknowledgements

We are very grateful for the comments of an associate editor and two referees on an earlier version of this paper. These led to substantial improvements.

Appendix A. Proofs of Theorems

Proof of Theorem 1. First, note that y_t can be written in the form

$$y_t = \sigma_1 w_t 1[t \leqslant \tau^* T] + (\sigma_2 w_t - \lambda w_{\tau^* T}) 1[t > \tau^* T],$$

where $\lambda = \sigma_2 - \sigma_1$ and $w_t = \sum_{i=1}^{t} \eta_i$. Also, t_{O} can be written as

$$t_{\mathrm{O}} = \frac{T^{-1} \sum_{t=2}^{T} (y_{t-1} - \bar{y})\varepsilon_t}{\hat{\sigma}\{T^{-2} \sum_{t=2}^{T} (y_{t-1} - \bar{y})^2\}^{1/2}}.$$

Dealing with the numerator term first, we have

$$T^{-1} \sum_{2}^{T} (y_{t-1} - \bar{y})\varepsilon_t = T^{-1} \sum_{2}^{T} y_{t-1}\varepsilon_t - T^{-1/2}\bar{y}T^{-1/2} \sum_{2}^{T} \varepsilon_t$$

and using the above representation for y_t, we find that

$$T^{-1} \sum_{2}^{T} y_{t-1}\varepsilon_t = \sigma_1^2 T^{-1} \sum_{2}^{T} w_{t-1}\eta_t + (\sigma_2^2 - \sigma_1^2) T^{-1} \sum_{\tau^* T+2}^{T} w_{t-1}\eta_t$$

$$- \sigma_2 \lambda T^{-1/2} w_{\tau^* T} T^{-1/2} (w_T - w_{\tau^* T+1}) + \mathrm{o_p}(1)$$

where

$$T^{-1} \sum_{2}^{T} w_{t-1}\eta_t \Rightarrow \frac{1}{2}\{W(1)^2 - 1\},$$

$$T^{-1} \sum_{\tau^* T+2}^{T} w_{t-1}\eta_t \Rightarrow \frac{1}{2}\{W(1)^2 - W(\tau^*)^2 - (1 - \tau^*)\},$$

$$T^{-1/2} w_{\tau^* T} T^{-1/2} (w_T - w_{\tau^* T+1}) \Rightarrow W(\tau^*)\{W(1) - W(\tau^*)\}.$$

Next,

$$T^{-1/2}\bar{y} = \sigma_1 T^{-3/2}\sum_1^T w_t + \lambda T^{-3/2}\sum_{\tau^* T+1}^T w_t - \lambda(1-\tau^*)T^{-1/2}w_{\tau^* T},$$

where

$$T^{-3/2}\sum_1^T w_t \Rightarrow \int_0^1 W(r)\,\mathrm{d}r, \quad T^{-3/2}\sum_{\tau^* T+1}^T w_t \Rightarrow \int_{\tau^*}^1 W(r)\,\mathrm{d}r,$$

$$T^{-1/2}w_{\tau^* T} \Rightarrow W(\tau^*)$$

and

$$T^{-1/2}\sum_2^T \varepsilon_t = \sigma_1 T^{-1/2}w_T + \lambda T^{-1/2}(w_T - w_{\tau^* T}),$$

where

$$T^{-1/2}w_T \Rightarrow W(1), \qquad T^{-1/2}(w_T - w_{\tau^* T}) \Rightarrow W(1) - W(\tau^*).$$

Gathering together these results then shows that

$$T^{-1}\sum_2^T (y_{t-1} - \bar{y})\varepsilon_t \Rightarrow \sigma_1^2 a(\delta, \tau^*). \tag{A.1}$$

Using a similar argument, we have

$$T^{-2}\sum_2^T (y_{t-1} - \bar{y})^2 \Rightarrow \sigma_1^2 b(\delta, \tau^*). \tag{A.2}$$

Finally, it is straightforward to show that

$$\begin{aligned}\hat{\sigma}^2 &= T^{-1}\sum_{t=1}^T \varepsilon_t^2 + \mathrm{o_p}(1)\\ &= \sigma_1^2 T^{-1}\sum_1^T \eta_t^2 + (\sigma_2^2 - \sigma_1^2)(1-\tau^*)\{(1-\tau^*)T\}^{-1}\sum_{\tau^* T+1}^T \eta_t^2\\ &\Rightarrow \tau^*\sigma_1^2 + (1-\tau^*)\sigma_2^2. \end{aligned} \tag{A.3}$$

Combining (A.1)–(A.3) gives the result in (5). □

Proof of Theorem 2. First, we derive the probability limit of $\hat{\sigma}_1(\tau)^2$. We define $X_t = [1, y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1}]'$, $\hat{\pi}=[\hat{\alpha}, \hat{\rho}, \hat{\phi}_1, \ldots, \hat{\phi}_{p-1}]'$ and $\pi=[0, 1, \phi_1, \ldots, \phi_{p-1}]'$. Then,

$\hat{\sigma}_1(\tau)^2$ can be expressed as

$$\hat{\sigma}_1(\tau)^2 = (\tau T)^{-1}\sum_{t=1}^{\tau T}\varepsilon_t^2 - 2(\tau T)^{-1}\{D_T(\hat{\pi}-\pi)\}' D_T^{-1}\sum_{t=1}^{\tau T} X_t\varepsilon_t$$

$$+ (\tau T)^{-1}\{D_T(\hat{\pi}-\pi)\}' D_T^{-1}\sum_{t=1}^{\tau T} X_t X_t' D_T^{-1}\{D_T(\hat{\pi}-\pi)\}$$

$$= (\tau T)^{-1}\sum_{t=1}^{\tau T}\varepsilon_t^2 + \mathrm{o_p}(1),$$

where $D_T = \mathrm{diag}(T^{1/2}, T, T^{1/2}, \ldots, T^{1/2})$. The last equality is due to the facts that $D_T^{-1}\sum_{t=1}^{\tau T} X_t\varepsilon_t = \mathrm{O_p}(1), D_T^{-1}\sum_{t=1}^{\tau T} X_t X_t' D_T^{-1} = \mathrm{O_p}(1)$ and $D_T(\hat{\pi}-\pi) = \mathrm{O_p}(1)$. We consider two cases: when $\tau \leqslant \tau^*$, we have

$$(\tau T)^{-1}\sum_{t=1}^{\tau T}\varepsilon_t^2 \Rightarrow \sigma_1^2$$

and for the other case $\tau > \tau^*$, the limit is

$$(\tau T)^{-1}\sum_{t=1}^{\tau T}\varepsilon_t^2 \Rightarrow \frac{\tau^*}{\tau}\,\sigma_1^2 + \frac{\tau-\tau^*}{\tau}\,\sigma_2^2.$$

Therefore, we have

$$\hat{\sigma}_1(\tau)^2 \Rightarrow \sigma_1^2 1[\tau \leqslant \tau^*] + \left(\frac{\tau^*}{\tau}\sigma_1^2 + \frac{\tau-\tau^*}{\tau}\sigma_2^2\right) 1[\tau > \tau^*].$$

The same kind of argument can be used to show that

$$\hat{\sigma}_2(\tau)^2 \Rightarrow \left(\frac{\tau^*-\tau}{1-\tau}\,\sigma_1^2 + \frac{1-\tau^*}{1-\tau}\sigma_2^2\right) 1[\tau \leqslant \tau^*] + \sigma_2^2 1[\tau > \tau^*].$$

Given the functional form of $\hat{Q}_T(\tau) = \tau\ln\hat{\sigma}_1(\tau)^2 + (1-\tau)\ln\hat{\sigma}_2(\tau)^2$, we have the following uniform convergence result:

$$\hat{Q}_T(\tau) \Rightarrow Q(\tau),$$

where the limit function $Q(\tau)$ is a non-random function given by

$$Q(\tau) = \left[\tau\ln(\sigma_1^2) + (1-\tau)\ln\left\{\frac{\tau^*-\tau}{1-\tau}\,\sigma_1^2 + \frac{1-\tau^*}{1-\tau}\,\sigma_2^2\right\}\right] 1[\tau \leqslant \tau^*]$$

$$+ \left[\tau\ln\left\{\frac{\tau^*}{\tau}\,\sigma_1^2 + \frac{\tau-\tau^*}{\tau}\,\sigma_2^2\right\} + (1-\tau)\ln(\sigma_2^2)\right] 1[\tau > \tau^*].$$

In order to use a similar argument to that of Theorem 3.4 in White (1996), we only need to show that the non-random limit function $Q(\tau)$ is minimised at the true break point τ^*. We consider the case: $\tau \leqslant \tau^*$ and we want to show that for any $\tau \in (0,\tau^*)$, $Q(\tau) - Q(\tau^*) \geqslant 0$. Using some basic algebra, one can show that $Q(\tau) - Q(\tau^*) \geqslant 0$ if and only if

$$\omega + (1-\omega)\left(\frac{\sigma_2^2}{\sigma_1^2}\right) \geqslant 1^{\omega}\left(\frac{\sigma_2^2}{\sigma_1^2}\right)^{(1-\omega)},$$

where $\omega = (\tau^* - \tau)/(1-\tau)$. The latter is the inequality between the weighted arithmetic mean and the weighted geometric mean of 1 and σ_2^2/σ_1^2 and the equality holds only when $1 = \sigma_2^2/\sigma_1^2$. The same argument can be used for the other case: $\tau > \tau^*$. Hence, as long as there is a break in innovation variance, the strict inequality holds and we have the required result:

$$\hat{\tau} - \tau^* = \mathrm{o_p}(1). \tag{A.4}$$

In order to prove the other two results, we note that

$$\hat{\sigma}_1(\tau^*)^2 \Rightarrow \sigma_1^2, \qquad \hat{\sigma}_2(\tau^*)^2 \Rightarrow \sigma_2^2.$$

This uniform convergence together with the result in (A.4) implies that $\hat{\sigma}_1(\hat{\tau})^2$ and $\hat{\sigma}_2(\hat{\tau})^2$ converge to σ_1^2 and σ_2^2 in probability. □

Proof of Theorem 3. Given the results from Theorem 2 on the limits of $\hat{\sigma}_1(\tau)^2$ and $\hat{\sigma}_2(\tau)^2$, it is now easy to show that the limit of the objective function $\hat{V}_T(\tau)^2$ is given by

$$\hat{V}_T(\tau)^2 \Rightarrow V(\tau)^2,$$

where

$$V(\tau)^2 = (\sigma_2^2 - \sigma_1^2)\frac{\tau}{1-\tau}(1-\tau^*)1[\tau \leqslant \tau^*] + (\sigma_2^2 - \sigma_1^2)\frac{1-\tau}{\tau}\tau^{*2}1[\tau > \tau^*].$$

Using the fact that $(\tau/(1-\tau))((1-\tau)/\tau)$ is a monotone increasing (decreasing) function, it is easy to show that the non-random limit function $V(\tau)$ is maximised at the true break point τ^*. Hence, we have the required result: $\hat{\tau} - \tau^* = \mathrm{o_p}(1)$ which also implies that $\hat{\sigma}_1(\hat{\tau})^2$ and $\hat{\sigma}_2(\hat{\tau})^2$ converge to σ_1^2 and σ_2^2 in probability. □

Proof of Theorem 4. First, we assume that $0 < \tau_1 < \tau_2 < 1$. Note that it can be shown that

$$T^{1/2}\hat{V}_T(\tau) = \frac{1}{\{\tau(1-\tau)\}^{1/2}}\left\{\tau\frac{1}{T^{1/2}}\sum_{t=1}^{T}(e_t^2 - \sigma^2) - \frac{1}{T^{1/2}}\sum_{t=1}^{\tau T}(e_t^2 - \sigma^2)\right\},$$

where $\sigma^2 = \sigma_1^2 = \sigma_2^2$ is the common innovation variance. We consider the partial sum:

$$\frac{1}{T^{1/2}}\sum_{t=1}^{\tau T}(e_t^2-\sigma^2) = \frac{1}{T^{1/2}}\sum_{t=1}^{\tau T}(\varepsilon_t^2-\sigma^2) - 2\,\frac{1}{T^{1/2}}\{D_T(\hat{\pi}-\pi)\}' D_T^{-1}\sum_{t=1}^{\tau T} X_t\varepsilon_t$$

$$+\frac{1}{T^{1/2}}\{D_T(\hat{\pi}-\pi)\}' D_T^{-1}\sum_{t=1}^{\tau T} X_t X_t' D_T^{-1}\{D_T(\hat{\pi}-\pi)\}$$

$$= \sigma^2\,\frac{1}{T^{1/2}}\sum_{t=1}^{\tau T}(\eta_t^2-1) + \mathrm{o_p}(1) \Rightarrow \sigma^2\gamma B(\tau).$$

Applying the above result, we obtain

$$T^{1/2}\hat{V}_T(\tau) \Rightarrow \frac{\sigma^2\gamma\{\tau B(1)-B(\tau)\}}{\{\tau(1-\tau)\}^{1/2}}.$$

Since

$$\hat{\tau} = \underset{\tau\in[\tau_1,\tau_2]}{\arg\max}\,\{T^{1/2}\hat{V}_T(\tau)\}^2$$

we have by the continuous mapping theorem

$$\hat{\tau} \Rightarrow \underset{\tau\in[\tau_1,\tau_2]}{\arg\max}\,\frac{\sigma^4\gamma^2\{B(\tau)-\tau B(1)\}^2}{\tau(1-\tau)}$$

which completes the proof of the first part of the theorem. □

The second part of the theorem follows from Corollary 1 in Andrews (1993). Using the first part of the theorem, together with the law of the iterated logarithm for Brownian motion it can be shown that $\max_{\tau\in[0,1]} T\hat{V}_T(\tau)^2 \Rightarrow \infty$. Since $\max_{\tau\in[\tau_1,\tau_2]} T\hat{V}_T(\tau)^2 = \mathrm{O_p}(1)$ for $0<\tau_1<\tau_2<1$, these results together imply that $\hat{\tau}\xrightarrow{\mathrm{p}}\{0,1\}$.

Proof of Theorem 5. Suppose we know all the nuisance parameters τ^*, σ_1^2 and σ_2^2. In this case, the specification in (14) can be directly estimated. Let $t_{F\{\tau^*,\sigma_1,\sigma_2\}}$ denote the t-ratio for testing $\rho=1$ when (14) is estimated by OLS. Then, by a straightforward application of the arguments in Perron (1990), we have

$$t_{F\{\tau^*,\sigma_1,\sigma_2\}} \Rightarrow \frac{f(\tau^*)}{g(\tau^*)^{1/2}}.$$

This is because the specification is the same as given in Perron (1990), except that there are additional lagged one-time dummy variables whose role is essentially to remove $(p-1)$ central observations. Next, we assume that τ^* is known, but σ_1^2 and σ_2^2 are unknown. In this case, we run the following regression:

$$\breve{y}_t(\tau^*) = \hat{\alpha}_0 + \hat{\alpha}_1 d_{1t}(\tau^*) + \hat{\alpha}_2 d_{2t}(\tau^*) + \sum_{j=1}^{p-1} \hat{\theta}_j d_{2,t-j}(\tau^*)$$

$$+ \hat{\rho}\breve{y}_{t-1}(\tau^*) + \sum_{j=1}^{p-1} \hat{\phi}_j \Delta \breve{y}_{t-j}(\tau^*) + \hat{\eta}_t, \tag{A.5}$$

where

$$\breve{y}_t(\tau^*) = \hat{\sigma}_1(\tau^*)^{-1} y_t 1[t \leqslant \tau^* T] + \hat{\sigma}_2(\tau^*)^{-1} y_t 1[t > \tau^* T]$$

and $\hat{\sigma}_1(\tau^*)^2$ and $\hat{\sigma}_2(\tau^*)^2$ are defined as in (9) with τ^* in place of $\hat{\tau}$. Letting $t_{F\{\tau^*,\hat{\sigma}_1(\tau^*),\hat{\sigma}_2(\tau^*)\}}$ denote the t-ratio for testing $\rho = 1$ when (A.5) is estimated by OLS, we have the following result

$$t_{F\{\tau^*,\hat{\sigma}_1(\tau^*),\hat{\sigma}_2(\tau^*)\}} \Rightarrow \frac{f(\tau^*)}{g(\tau^*)^{1/2}}. \tag{A.6}$$

All the terms in $t_{F\{\tau^*,\hat{\sigma}_1(\tau^*),\hat{\sigma}_2(\tau^*)\}}$ are exactly the same as in $t_{F\{\tau^*,\sigma_1,\sigma_2\}}$ except that they are multiplied by $\sigma_1^2\hat{\sigma}_1(\tau^*)^{-2}$ or $\sigma_2^2\hat{\sigma}_2(\tau^*)^{-2}$ depending on which subsample they belong to. Therefore, result (A.6) easily follows from the fact that $\sigma_1^2\hat{\sigma}_1(\tau^*)^{-2} \Rightarrow 1$, $\sigma_2^2\hat{\sigma}_2(\tau^*)^{-2} \Rightarrow 1$. Then, the claim in Theorem 5, where $t_F = t_{F\{\hat{\tau},\hat{\sigma}_1(\hat{\tau}),\hat{\sigma}_2(\hat{\tau})\}}$, follows from the uniform convergence result in (A.6) and the consistency result of $\hat{\tau}$ in Section 3. □

References

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Bai, J., 1993. Least squares estimation of a shift in linear processes. Journal of Time Series Analysis 15, 453–472.

Bai, J., 1995. Least absolute deviation estimation of a shift. Econometric Theory 11, 403–436.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–79.

Bai, J., Lumsdaine, R., Stock, J., 1998. Testing for and dating common breaks in multivariate time series. Review of Economic Studies 65, 395–432.

Bhattacharya, P.K., 1987. Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case. Journal of Multivariate Analysis 23, 183–208.

Fu, Y., Curnow, R.N., 1990. Maximum likelihood estimation of multiple change points. Biometrika 77, 563–573.

Hamori, S., Tokihisa, A., 1997. Testing for a unit root in the presence of a variance shift. Economics Letters 57, 245–253.

Hsu, S., 1977. Tests for variance shift at an unknown time point. Applied Statistics 26, 279–284.

Inclán, C., 1993. Detection of multiple changes of variance using posterior odds. Journal of Business and Economic Statistics 11, 289–300.

Leybourne, S.J., Mills, T.C., Newbold, P., 1998. Spurious rejections by Dickey–Fuller tests in the presence of a break under the null. Journal of Econometrics 87, 191–203.

Nunes, LC., Kuan, C.M., Newbold, P., 1995. Spurious break. Econometric Theory 11, 736–749.

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

Perron, P., 1990. Testing for a unit root in a time series with a changing mean. Journal of Business and Economic Statistics 8, 153–162.

Perron, P., Vogelsang, T.J., 1992. Testing for a unit root in a time series with a changing mean: corrections and extensions. Journal of Business and Economic Statistics 10, 467–470.

Picard, D., 1985. Testing and estimating change-points in time series. Advances in Applied Probability 176, 841–867.

White, H., 1996. Estimation, Inference and Specification Analysis. Cambridge University Press, Cambridge.

Wichern, D.W., Miller, R., Hsu, D., 1976. Changes of variance in first-order autoregressive time series models-with an application. Applied Statistics 25, 248–256.