Lanne, Markku; Lütkepohl, Helmut; Saikkonen, Pentti

Working Paper

Unit root tests in the presence of innovational outliers

SFB 373 Discussion Paper, No. 2001,82

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Lanne, Markku; Lütkepohl, Helmut; Saikkonen, Pentti (2001) : Unit root tests in the presence of innovational outliers, SFB 373 Discussion Paper, No. 2001,82, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10050593

This Version is available at:
https://hdl.handle.net/10419/62701

Unit Root Tests in the Presence of Innovational Outliers*

by

Markku Lanne
University of Helsinki

Helmut Lütkepohl
Humboldt University Berlin

and
Pentti Saikkonen
University of Helsinki

Address for correspondence: Helmut Lütkepohl, Institut für Statistik und Ökonometrie, Humboldt-Universität, Spandauer Str. 1, D-10178 Berlin, Germany

Abstract

Unit root tests are considered for time series with innovational outliers. The function representing the outliers can have a very general nonlinear form and additional deterministic mean and trend terms are allowed for. Prior to the tests the deterministic parts and other nuisance parameters of the data generation process are estimated in a first step. Then the series are adjusted for these terms and unit root tests of the Dickey-Fuller type are applied to the adjusted series. The properties of previously suggested tests of this sort are analyzed and modifications are proposed which take into account estimation errors in the nuisance parameters. An important result is that estimation under the null hypothesis is preferable to estimation under local alternatives. This contrasts with results obtained by other authors for time series without outliers. A comparison with additive outlier models is also performed.

JEL classification: C22, C12

Keywords: Univariate time series, unit root, structural shift, autoregression

*We are grateful to Ralf Brüggemann for helping with the computations and to Christian Müller for comments. Moreover, we thank the Deutsche Forschungsgemeinschaft, SFB 373, the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213) and the Yrjö Jahnsson Foundation for financial support. The third author also acknowledges financial support by the Alexander von Humboldt Foundation under a Humboldt research award. Part of this research was done while the first and third authors were visiting the Humboldt University in Berlin.

1 Introduction

Interventions that occur regularly in economic systems often cause outliers and structural shifts in the observed time series. Therefore, modeling time series with such features has become an issue of importance in recent years. In particular, testing for unit roots in the presence of outliers and structural shifts has attracted considerable attention. Examples of articles dealing with these issues are Perron (1989, 1990), Perron & Vogelsang (1992), Banerjee, Lumsdaine & Stock (1992), Zivot & Andrews (1992), Amsler & Lee (1995), Leybourne, Newbold & Vougas (1998), Montañés & Reyes (1998), Vogelsang & Perron (1998), Saikkonen & Lütkepohl (2001) and Lütkepohl, Müller & Saikkonen (2001) (henceforth LMS). In some of the literature the time where the intervention or shift occurs is assumed to be known and in other articles it is assumed unknown. In this study we assume that the intervention point is known. Such an assumption is often reasonable in practice. For example, on January 1, 1999, a common currency was introduced in a number of European countries or the German unification is known to have occurred in 1990. These events have had an impact on some economic time series which may be useful to take into account in unit root testing.

We will follow LMS and model the intervention as an innovational outlier which is represented by a very general nonlinear deterministic shift function. These authors propose tests for unit roots based on the idea that the deterministic part is estimated in a first step and is subtracted from the series. In the estimation procedure quasi-differenced variables are used so that estimation is done under local alternatives as in Elliot, Rothenberg & Stock (1996). Standard unit root tests are then applied to the adjusted series.* The purpose of this study is to propose modifications of these tests which are expected to work well in small sample situations and we will perform Monte Carlo comparisons of the properties of the tests. We will also compare our tests to those proposed by Lanne, Lütkepohl & Saikkonen (2001) (henceforth LLS) who consider an additive intervention model. The results lead to useful recommendations for applied work.

The structure of this study is as follows. The basic model from LMS is presented in Sec. 2 together with the assumptions needed for asymptotic derivations. Estimation of the nuisance parameters within these models is discussed in Sec. 3 and a range of unit root tests is presented in Sec. 4 including the asymptotic distributions of the test statistics. In Sec. 5

*For a brief exposition of such tests see also Hansen (1993).

we present some local power simulations. A small sample comparison of the tests based on a Monte Carlo experiment is reported in Sec. 6 and conclusions are summarized in Sec. 7. Some proofs are provided in the Appendix.

In the following, L and Δ are the lag and differencing operators, respectively, so that, for a time series variable y_t, $Ly_t = y_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$. Convergence in probability and in distribution are denoted by $\xrightarrow{p}$ and $\xrightarrow{d}$, respectively. Independently, identically distributed will be abbreviated as $iid(\cdot,\cdot)$, where the first and second moments are indicated in parentheses. The normal distribution is signified by $N(\cdot,\cdot)$. Furthermore, $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ and $o_p(\cdot)$ are the usual symbols for the order of convergence and convergence in probability, respectively, of a sequence. The symbol $\lambda_{min}(A)$ denotes the minimal eigenvalue of a matrix A and $\|\cdot\|$ denotes the Euclidean norm. The abbreviations sup and inf are used as usual for supremum and infimum, respectively. The n-dimensional Euclidean space is signified as $\mathbf{R}^n$. DGP abbreviates data generation process, DF is short for Dickey-Fuller and OLS and GLS are used for ordinary least squares and generalized least squares, respectively. Moreover, AR abbreviates autoregressive or autoregressive process.

2 The Models

We consider the model

$$b(L)y_t = \mu_0 + \mu_1 t + f_t(\theta)'\gamma + v_t, \qquad t = 1, 2, \ldots, \tag{2.1a}$$

where the error term v_t is assumed to be an AR process of order 1,

$$v_t = \rho v_{t-1} + \varepsilon_t, \tag{2.1b}$$

starting with $v_0 = 0$, for convenience. Here $\varepsilon_t \sim iid(0, \sigma^2)$ and $-1 < \rho \leq 1$ with $\rho = 1$ implying a unit root in y_t. The operator $b(L) = 1 - b_1 L - \cdots - b_p L^p$ is a polynomial in L with roots bounded away from the unit circle. More precisely, for some $\epsilon > 0$, $b(L) \neq 0$ for $|L| \leq 1 + \epsilon$. For simplicity, we assume that a suitable number of presample values of the observed series y_t is available. The parameters μ_0, μ_1 and γ in these models are supposed to be unrestricted.

The term $f_t(\theta)'\gamma$ describes the form of the innovational outlier. In its general form it consists of the unknown $(k \times 1)$ parameter vector γ and the $(k \times 1)$ vector of deterministic

sequences f_t depending on the $(m \times 1)$ parameter vector θ. A simple version of this term is obtained by defining

$$f_t(\theta) = d_{1t} = \begin{cases} 0, & t < T_1 \\ 1, & t \geq T_1 \end{cases}$$

so that the function f_t is a simple shift dummy variable and the corresponding γ is a scalar. In that case the term $f_t(\theta)'\gamma$ generates a level shift of the form $\gamma b(L)^{-1} d_{1t}$. Thus, depending on the AR operator, even if the innovational outlier term consists of a simple shift dummy, it can generate a smooth shift to a new level of the series. This behavior is the typical characteristic of innovational outlier models as opposed to additive outlier models which often include abrupt shifts. In our framework the outlier term can be much more general than the simple shift dummy. It is assumed to be generated by a function $f_t(\theta)$ where the functional form is known and the parameter vector θ may be unknown. Knowledge of the functional form of $f_t(\theta)$ implies in particular that the shift date T_1 where the innovational outlier occurs is known. Conditions required for the parameters θ and the sequence $f_t(\theta)$ are collected in the following set of assumptions which are taken from Saikkonen & Lütkepohl (2001), LMS and LLS.

Assumption 1

(a) The parameter space of θ, denoted by Θ, is a compact subset of $\mathbf{R}^m$.

(b) For each $t = 1, 2, \ldots$, the term $f_t(\theta)$ is a continuously differentiable function in an open set containing the parameter space Θ and, denoting by $F_t(\theta)$ the vector of all partial derivatives of $f_t(\theta)$,

$$\sup_T \sum_{t=1}^{T} \sup_{\theta \in \Theta} \|\Delta f_t(\theta)\| < \infty \quad \text{and} \quad \sup_T \sum_{t=1}^{T} \sup_{\theta \in \Theta} \|\Delta F_t(\theta)\| < \infty$$

where $f_0(\theta) = 0$ and $F_0(\theta) = 0$.

(c) $f_1(\theta) = \cdots = f_{p+1}(\theta) = 0$ for all $\theta \in \Theta$. Moreover, defining $G_t(\theta) = [f_t(\theta)' : F_t(\theta)']'$ for $t = 1, 2, \ldots$, there exists a real number $\epsilon > 0$ and an integer T_* such that, for all $T \geq T_*$,

$$\inf_{\theta \in \Theta} \lambda_{min} \left\{ \sum_{t=2}^{T} \Delta G_t(\theta) \Delta G_t(\theta)' \right\} \geq \epsilon.$$

□

As mentioned earlier, these conditions are taken from earlier literature where they are also discussed and explained in some detail. Whereas a compact parameter space Θ and the continuity requirement in Assumption 1(b) are standard assumptions in nonlinear estimation and testing problems, the summability conditions in Assumption 1(b) for the function $f_t(\theta)$ and its partial derivatives $F_t(\theta)$ are not restrictive in the present context because they hold in the applications we have in mind, if the parameter space Θ is defined in a suitable way. The conditions in Assumption 1(b) and (c) are formulated for differences of the sequences $f_t(\theta)$, $g_t(\theta)$ and the partial derivatives because our aim is to study unit root tests. Hence, estimation of the parameters μ_i $(i = 0, 1)$, θ and γ is considered under the null hypothesis that the error process contains a unit root. Efficient estimation then requires that the variables are differenced.

Assumption 1(c) guarantees that our estimators of the deterministic part of the model are well-defined. Notice, however, that consistent estimation of θ and γ is not possible because, by Assumption 1(b), the variation of (the differenced) regressors does not increase as $T \to \infty$. In LLS it is also emphasized that our assumptions imply that, for each value of θ, the sequence $g_t(\theta)$ defines a slowly evolving trend in the terminology of Elliott, Rothenberg & Stock (1996, Condition B). Our conditions are stronger than those of Elliott et al., however. No attempt has been made here to weaken Assumption 1 because it is convenient for our purposes and applies to the models of interest in the following. For more discussion of Assumption 1 see in particular LLS.

We compare unit root tests within the model (2.1). More precisely, we consider tests of the pair of hypotheses

$$H_0 : \rho = 1 \qquad \text{vs.} \qquad H_1 : |\rho| < 1.$$

The idea is to estimate the nuisance parameters first and then remove the deterministic part and the serial dependence induced by the AR operator $b(L)$. Then a test is performed on the adjusted series. In the next section we therefore discuss estimation of the nuisance parameters.

3 Estimation of Nuisance Parameters

Suppose that the process v_t is near integrated with

$$\rho = \rho_T = 1 + \frac{c}{T}, \tag{3.1}$$

where $c \leq 0$ is a fixed real number. The estimation procedure proposed by LMS uses an empirical counterpart of the parameter ρ replacing c by a chosen value $\bar{c}$ and pretending that $\bar{c} = c$ although we do not assume that this presumption is actually true. The idea is to apply a GLS procedure by first transforming the variables in (2.1a) by the filter $1 - \bar{\rho}_T L$, where $\bar{\rho}_T = 1 + \frac{\bar{c}}{T}$ and then applying OLS to the transformed model. In other words, we use quasi-differencing for our variables. The choice of $\bar{c}$ will be discussed later.

Using ρ_T instead of ρ, the generating process of v_t can be written as

$$v_t = \rho_T v_{t-1} + \varepsilon_t, \qquad t = 1, 2, \ldots \tag{3.2}$$

Employing matrix notation and defining

$$Y = [y_1 : (y_2 - \bar{\rho}_T y_1) : \cdots : (y_T - \bar{\rho}_T y_{T-1})]',$$

$$Z_1 = \begin{bmatrix} 1 & 1 - \bar{\rho}_T & \cdots & 1 - \bar{\rho}_T \\ 1 & (2 - \bar{\rho}_T) & \cdots & (T - \bar{\rho}_T(T-1)) \end{bmatrix}',$$

$$Z_2(\theta) = [f_1(\theta) : f_2(\theta) - \bar{\rho}_T f_1(\theta) : \cdots : f_T(\theta) - \bar{\rho}_T f_{T-1}(\theta)]',$$

$Z(\theta) = [Z_1 : Z_2(\theta)]$ and $W(\theta) = [Z(\theta) : V]$, where V is the $(T \times p)$ matrix containing lagged values of y_t transformed in the same way as the other variables, the quasi-differenced form of $(2.1a)$ can be written as

$$Y = W(\theta)\beta + \mathcal{E}. \tag{3.3}$$

Here $\beta = [\mu_0 : \mu_1 : \gamma' : b']'$ and $\mathcal{E} = [e_1 : \cdots : e_T]'$ is an error term such that $e_t = v_t - \bar{\rho}_T v_{t-1} = \varepsilon_t + T^{-1}(c - \bar{c})v_{t-1}$. We shall consider a nonlinear OLS estimation of (3.3) by proceeding in the same way as in the case $c = 0$, that is, $e_t = \varepsilon_t$ or under the null hypothesis. Our estimators are thus obtained by minimizing the sum of squares function

$$S_T(\theta, \beta) = (Y - W(\theta)\beta)'(Y - W(\theta)\beta). \tag{3.4}$$

The estimator of β can be written as

$$\tilde{\beta} = (W(\tilde{\theta})'W(\tilde{\theta}))^{-1}W(\tilde{\theta})'Y, \tag{3.5}$$

where θ is the value of θ which minimizes (3.4) jointly with β.

One may also wish to consider a modification of the above approach to avoid potential adverse finite sample effects from treating the first observation differently than the other observations in (3.3). Thus, we delete the first element in (3.3) and define

$$Y^* = [(y_2 - \bar{\rho}_T y_1) : \cdots : (y_T - \bar{\rho}_T y_{T-1})]',$$

$$Z_1^* = \begin{bmatrix} 1 & \cdots & 1 \\ (2 - \bar{\rho}_T) & \cdots & (T - \bar{\rho}_T(T-1)) \end{bmatrix}'$$

and

$$Z_2^*(\theta) = [(f_2(\theta) - \bar{\rho}_T f_1(\theta)) : \cdots : (f_T(\theta) - \bar{\rho}_T f_{T-1}(\theta))]'.$$

Moreover, the $((T-1) \times p)$ matrix V^* and the $((T-1) \times 1)$ vector $\mathcal{E}^*$ are defined by deleting the first row and first component from V and $\mathcal{E}$, respectively. Instead of (3.3) we now consider

$$Y^* = W^*(\theta)\beta^* + \mathcal{E}^*, \tag{3.6}$$

where $W^*(\theta) = [Z^*(\theta) : V^*]$ with $Z^*(\theta) = [Z_1^* : Z_2^*(\theta)]$ and $\beta^* = [\mu_0^* : \mu_1 : \gamma' : b']'$ with $\mu_0^* = \mu_0(1 - \bar{\rho}_T)$. For simplicity the notation ignores the dependence of the quantities on the chosen value of $\bar{c}$ and on the sample size. In this approach we do not try to estimate the parameter μ_0. This means that we cannot obtain an empirical counterpart of the process v_t but only of $v_t + \mu_0$. This feature will be taken into account in constructing unit root tests in the next section.

We estimate the parameters β^* and θ in (3.6) by minimizing the obvious analog of the sum of squares function in (3.4). If $\bar{c} = 0$ (or $\bar{\rho}_T = 1$) the two columns of the matrix Z_1^* are identical so that the regression model (3.6) is not of full column rank. Then we shall delete the first column of Z_1^* and accordingly delete μ_0^* from the parameter vector β^*. Since the treatment of this special case is fairly obvious it will not be discussed here in more detail. We may also impose the restriction $\mu_1 = 0$ if a linear trend term is not needed.

4 The Tests

Once the nuisance parameters in (2.1a) have been estimated, the residual series $\tilde{v}_t = \tilde{b}(L)y_t - \tilde{\mu}_0 - \tilde{\mu}_1 t - f_t(\tilde{\theta})'\tilde{\gamma}$ may be used to obtain unit root tests. There are several possible choices.

LMS suggest using DF t-tests like, for instance, Elliott et al. (1996). In the following we shall also consider these tests.

Consider the auxiliary regression model

$$\tilde{v}_t = \rho \tilde{v}_{t-1} + e_t^*, \qquad t = 2, \ldots, T. \tag{4.1}$$

If $\tilde{v}_t$ is replaced by v_t, the error term in (4.1) becomes ε_t so that we can use OLS to obtain a test statistic. LMS consider the usual t-statistic for testing $\rho = 1$ in (4.1). In the following this statistic will be denoted by $\mathbf{t}_{LMS}$. Note that LMS use the model (4.1) for $t = 1, \ldots, T$ with $\tilde{v}_0 = 0$.

Notice, however, that the error term in the auxiliary regression model (4.1) also contains estimation errors caused by replacing the nuisance parameters $b = [b_1 : \cdots : b_p]'$, μ_0, μ_1, θ and γ by their OLS estimators. As far as the finite sample properties of the above test and particularly the performance of the asymptotic size approximation are concerned it may therefore be worthwhile to try to allow for this feature. To investigate this possibility, suppose the null hypothesis holds and note that, by straightforward calculation, one can readily see that

$$e_t^* = \varepsilon_t - [\tilde{b}(L) - b(L)]\Delta y_t - (\tilde{\mu}_1 - \mu_1) - [\Delta f_t(\tilde{\theta})'\tilde{\gamma} - \Delta f_t(\theta)'\gamma], \quad t = 2, \ldots, T. \tag{4.2}$$

For simplicity, consider first the special case where the function $f_t(\theta)$ is defined by the step dummy d_{1t} so that it is independent of the parameter θ. As is clear from equation (4.2), the estimation errors caused by using estimators of nuisance parameters can then be allowed for by augmenting the auxiliary regression model (4.1) by the impulse dummy Δd_{1t}, the lagged differences $\Delta y_{t-1}, \ldots, \Delta y_{t-p}$, and an intercept term. After this modification the test statistic can be defined on the basis of the OLS estimator of ρ in the same way as before. The inclusion of an impulse dummy in (4.1) will not change the limiting distribution of the resulting unit root test but the inclusion of an intercept term does. We shall consider both modifications. It should be noted, however, that since the mean value of the lagged differences $\Delta y_{t-1}, \ldots, \Delta y_{t-p}$ is generally nonzero the inclusion of these variables as additional regressors in (4.1) will change the limiting distribution of the resulting unit root test. The reason is that these lagged differences are not asymptotically orthogonal to the variable $\tilde{v}_{t-1}$. It turns out, however, that this feature can be allowed for by using the mean-adjusted variables $\Delta y_{t-j} - \tilde{\mu}_*$ $(j = 1, \ldots, p)$, where $\tilde{\mu}_* = \tilde{\mu}_1/\tilde{b}(1)$.

If the function $f_t(\theta)$ depends on the parameter vector θ the treatment of the fourth term on the right hand side of (4.2) becomes slightly more complicated than in the foregoing special case. We shall then assume that the function $f_t(\theta)$ is continuously differentiable in an open set containing the parameter space Θ and use the Taylor series approximation $\Delta f_t(\tilde{\theta}) - \Delta f_t(\theta) \approx \Delta(\partial f_t(\tilde{\theta})/\partial\theta')(\tilde{\theta}-\theta)$. Thus, instead of (4.1) we shall consider the auxiliary regression model

$$\tilde{v}_t = \rho\tilde{v}_{t-1} + \Delta f_t(\tilde{\theta})'\pi_1 + \Delta F_t(\tilde{\theta})'\pi_2 + \tilde{q}_t'\pi_3 + e_t^{\dagger}, \quad t = 2, \ldots, T, \tag{4.3}$$

where $F_t(\tilde{\theta})$ is a $(mk \times 1)$ vector containing the partial derivatives in the matrix $\partial f_t(\tilde{\theta})/\partial\theta$ and $\tilde{q}_t' = [\Delta y_{t-1} - \tilde{\mu}_* : \cdots : \Delta y_{t-p} - \tilde{\mu}_*]$. Let $\mathbf{t}_{adj}$ be the usual t-statistic for the null hypothesis $\rho = 1$ based on the OLS estimator of ρ in (4.3).

Including an intercept term in the auxiliary regression gives

$$\tilde{v}_t = \nu + \rho\tilde{v}_{t-1} + \Delta f_t(\tilde{\theta})'\pi_1 + \Delta F_t(\tilde{\theta})'\pi_2 + \tilde{q}_t'\pi_3 + e_t^{\dagger}, \quad t = 2, \ldots, T, \tag{4.4}$$

and the relevant t-statistic will be denoted by $\mathbf{t}_{int}$.

Denoting the OLS estimators of b, μ_1, γ and θ based on the model (3.6) by $\tilde{b}^*$, $\tilde{\mu}_1^*$, $\tilde{\gamma}^*$ and $\tilde{\theta}^*$, respectively, we can form the series $\tilde{v}_t^* = \tilde{b}^*(L)y_t - \tilde{\mu}_1^* t - f_t(\tilde{\theta}^*)'\tilde{\gamma}$ $(t = 2, \ldots, T)$. Its theoretical counterpart is $v_t^* = v_t + \mu_0$ for which we have $v_t^* = \nu + \rho v_{t-1}^* + \varepsilon_t$ where $\nu = (1-\rho)\mu_0$. Thus, yet another unit root test may be based on the auxiliary regression model

$$\tilde{v}_t^* = \nu + \rho\tilde{v}_{t-1}^* + e_t^{**}, \quad t = 2, \ldots, T. \tag{4.5}$$

Our test statistic, denoted by $\mathbf{t}_{LMS}^*$, is the t-statistic for the null hypothesis $\rho = 1$ in (4.5) based on OLS estimation.

It is also possible to include terms to take care of estimation errors and base the unit root test on an auxiliary regression similar to (4.4),

$$\tilde{v}_t^* = \nu + \rho\tilde{v}_{t-1}^* + \Delta f_t(\tilde{\theta}^*)'\pi_1 + \Delta F_t(\tilde{\theta}^*)'\pi_2 + \tilde{q}_t^{*\prime}\pi_3 + e_t^{*\dagger}, \quad t = 2, \ldots, T, \tag{4.6}$$

The resulting unit root test statistic will be denoted by $\mathbf{t}_{int}^*$.

Moreover, if we have the a priori restriction $\mu_1 = 0$ the estimation procedures in Section 3 and the definitions of $\tilde{v}_t$ and $\tilde{v}_t^*$ are adjusted accordingly. Since in this case the limiting distributions of the corresponding unit root tests change, we augment the test statistics with

a superscript 0 to distinguish them from the statistics which allow for a linear time trend. In other words, the test statistics are denoted by $\mathbf{t}^0_{LMS}$, $\mathbf{t}^0_{adj}$, $\mathbf{t}^0_{int}$, $\mathbf{t}^{*0}_{LMS}$ and $\mathbf{t}^{*0}_{int}$, respectively. The limiting null distributions of all the test statistics are given in the following theorem which partly summarizes known results. The remaining unknown parts are proven in the Appendix.

Theorem

Suppose that Assumption 1 holds and that the matrices $Z(\theta)$ and $Z^*(\theta)$ are of full column rank for all $T \geq k+1$ and all $\theta \in \Theta$. Then,

$$\mathbf{t}^0_{LMS}, \mathbf{t}^0_{adj} \xrightarrow{d} \left(\int_0^1 B_c(s)^2 ds\right)^{-1/2} \int_0^1 B_c(s) dB_c(s), \tag{4.7}$$

where $B_c(s) = \int_0^s \exp\{c(s-u)\} dB_0(u)$ with $B_0(u)$ a standard Brownian motion,

$$\mathbf{t}^0_{int}, \mathbf{t}^{*0}_{LMS}, \mathbf{t}^{*0}_{int} \xrightarrow{d} \left(\int_0^1 \bar{B}_c(s)^2 ds\right)^{-1/2} \int_0^1 \bar{B}_c(s) dB_c(s), \tag{4.8}$$

where $\bar{B}_c(s)$ is the mean-adjusted version of $B_c(s)$,

$$\mathbf{t}_{LMS}, \mathbf{t}_{adj} \xrightarrow{d} \left(\int_0^1 G_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 G_c(s;\bar{c}) dG_c(s;\bar{c}), \tag{4.9}$$

where $G_c(s;\bar{c}) = B_c(s) - sK_c(\bar{c})$ with

$$K_c(\bar{c}) = h(\bar{c})^{-1} \int_0^1 (1-\bar{c}s) dB_0(s) + h(\bar{c})^{-1}(c-\bar{c}) \int_0^1 (1-\bar{c}s) B_c(s) ds$$

and $h(\bar{c}) = 1 - \bar{c} + \bar{c}^2/3$. Here the stochastic integral is a short-hand notation for $\int_0^1 G_c(s;\bar{c}) dB_c(s) - K_c(\bar{c}) \int_0^1 G_c(s;\bar{c}) ds$. Moreover,

$$\mathbf{t}_{int} \xrightarrow{d} \left(\int_0^1 \bar{G}_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 \bar{G}_c(s;\bar{c}) dG_c(s;\bar{c}), \tag{4.10}$$

where $\bar{G}_c(s;\bar{c})$ is a mean-adjusted version of $G_c(s;\bar{c})$. Furthermore,

$$\mathbf{t}^*_{LMS}, \mathbf{t}^*_{int} \xrightarrow{d} \left(\int_0^1 \bar{G}^*_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 \bar{G}^*_c(s;\bar{c}) dG^*_c(s;\bar{c}) \tag{4.11}$$

where $G^*_c(s;\bar{c}) = B_c(s) - sK^*_c(\bar{c})$, with $K^*_c(0) = B_c(1)$ and, for $\bar{c} < 0$,

$$K^*_c(\bar{c}) = \frac{12}{\bar{c}} \int_0^1 \left(\frac{1}{2} - s\right) dB_0(s) + \frac{12(c-\bar{c})}{\bar{c}} \int_0^1 \left(\frac{1}{2} - s\right) B_c(s) ds,$$

$\bar{G}^*_s(s;\bar{c})$ is a mean-adjusted version of $G^*_s(s;\bar{c})$ and the stochastic integral is a short-hand notation for $\int_0^1 \bar{G}^*_c(s;\bar{c}) dB_c(s) - K^*_c(\bar{c}) \int_0^1 \bar{G}^*_c(s;\bar{c}) ds$. □

The conditions for the ranks of the matrices $Z(\theta)$ and $Z^*(\theta)$ are included in the theorem because they are plausible. It can be shown using Assumption 1 that they hold for T large enough. The asymptotic distributions in (4.7) - (4.10) are the same as those in Theorem 1 of LLS for corresponding tests derived in an additive outlier framework. Thus, some critical values are available in Table 2 of LLS. For $c = 0$, the null distributions in (4.7) and (4.8) are conventional Dickey-Fuller (DF) distributions for unit root tests in models without deterministic terms and with intercept, respectively. The distribution of $\mathbf{t}_{LMS}$ in (4.9) is, of course, the one obtained by LMS.

The limiting null distribution of the test statistic $\mathbf{t}_{int}$ is again obtained by setting $c = 0$. It is free of unknown nuisance parameters but depends on the quantity $\bar{c}$. It differs from that of $\mathbf{t}_{LMS}$ and $\mathbf{t}_{adj}$ in that $G_c(s;\bar{c})$ is replaced by a mean-adjusted version. This difference is due to the intercept term included in the auxiliary regression model (4.4).

Obviously, the asymptotic distribution of the test statistics $\mathbf{t}^*_{LMS}$ and $\mathbf{t}^*_{int}$ also differs from the other ones. Instead of $G_c(s;\bar{c})$ in (4.10) we have $G^*_c(s;\bar{c})$ in (4.11). The difference between these two quantities is due to the different limiting distributions of the estimators $\tilde{\mu}_1$ and $\tilde{\mu}^*_1$. This difference results from a different treatment of the intercept term in the regression models (3.3) and (3.6) and, in the special case $\bar{c} = 0$, this difference vanishes.

Simulations shall be performed to study the power properties of our test statistics. Even without such simulations it is clear, however, that in terms of asymptotic local power the test statistics in (4.10) and (4.11) are inferior to those in (4.9) because they are not asymptotically equivalent to $\mathbf{t}_{LMS}$ and the asymptotic local power of $\mathbf{t}_{LMS}$ is indistinguishable from optimal (see Elliott et al. (1996)). This result is based on a specific initial value assumption, however, which may be unrealistic in some cases (see Elliott et al. (1996, pp. 819-820)). Therefore the performance of the $\mathbf{t}_{int}$, $\mathbf{t}^*_{LMS}$ and $\mathbf{t}^*_{int}$ tests may be preferable in some finite sample situations. We will report local power results and small sample comparisons for these tests in the following section.

Alternative approaches such as point optimal tests are possible in the present context. These tests would be based on the statistics $\hat{\sigma}^2(1)$ and $\hat{\sigma}^2(\bar{\rho}_T)$ defined by replacing $\tilde{\rho}$ in the variance estimator by unity and $\bar{\rho}_T$, respectively. According to the simulation results of Elliott et al. (1996) the overall properties of their DF t-statistic appeared somewhat better than those of the point optimal tests. Therefore we use the DF test versions. It may

also be worth noting that seasonal dummies may be included without affecting the limiting distributions of our tests.

5 Local Power Simulations

All the tests considered in the previous section are summarized in Table 1 for the case where no a priori restriction is available for μ_1. In order to investigate the local power of the tests we have generated time series

$$v_t = \rho_T v_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, T, \quad v_0 = 0, \quad \rho_T = 1 + c/T, \quad \varepsilon_t \sim iid\, N(0,1). \tag{5.1}$$

We assume $y_t = v_t$ so that $p = 0$ and there is no additional dynamics. Moreover, there is no deterministic part and we can use the generated series to investigate the tests with and without the restriction $\mu_1 = 0$. For this purpose we use again $\bar{\rho}_T = 1 + \bar{c}/T$ and consider the following $\tilde{v}_t$ series:

- $\tilde{v}_t^{(0)} = y_t - \tilde{\mu}_0$ $(t = 1, \ldots, T)$, where $\tilde{\mu}_0$ is obtained from a regression $(1 - \bar{\rho}_T L)y_t = \mu_0 z_{0t} + error_t$ $(t = 1, \ldots, T)$ with $z_{01} = 1$ and $z_{0t} = 1 - \bar{\rho}_T$ for $t = 2, \ldots, T$.
- $\tilde{v}_t^{(1)} = v_t - \tilde{\mu}_0 - \tilde{\mu}_1 t$ $(t = 1, \ldots, T)$, where $\tilde{\mu}_0$ and $\tilde{\mu}_1$ are obtained from a regression $(1 - \bar{\rho}_T L)y_t = \mu_0 z_{0t} + \mu_1(t - \bar{\rho}_T(t-1)) + error_t$ $(t = 1, \ldots, T)$.

Moreover, the $\tilde{v}_t^*$ series are obtained as:

- $\tilde{v}_t^{*(0)} = y_t$ $(t = 1, \ldots, T)$.
- $\tilde{v}_t^{*(1)} = y_t - \tilde{\mu}_1 t$ $(t = 1, \ldots, T)$, where $\tilde{\mu}_1$ is obtained from a regression $(1 - \bar{\rho}_T L)y_t = \nu_0 + \mu_1(t - \bar{\rho}_T(t-1)) + error_t$ $(t = 2, \ldots, T)$.

The series $\tilde{v}_t^{(i)}$ $(i = 0, 1)$ are used to compute t-statistics for the null hypothesis $\rho = 1$ based on the regression model (4.1). The series $\tilde{v}_t^{(1)}$ is also used to compute the t-statistic for $\rho = 1$ in $\tilde{v}_t^{(1)} = \nu + \rho \tilde{v}_{t-1}^{(1)} + u_t^*$ and the series $\tilde{v}_t^{*(i)}$ $(i = 0, 1)$ are used for the same purpose in conjunction with model (4.5). For large sample size T and $c = 0$ (i.e., $\rho_T = 1$) we get realizations of the null distributions corresponding to (4.7) - (4.11) in this way.

Notice that the asymptotic distributions in (4.7) and (4.8) do not depend on $\bar{c}$. Simulating (4.7) via $\tilde{v}_t^{(0)}$ there is a potential small sample dependence on $\bar{c}$, however. Therefore we use

Table 1. Summary of Tests

Test statistic	Underlying auxiliary regression	Asymptotic distribution
$\mathbf{t}_{LMS}$	$\tilde{v}_t = \rho\tilde{v}_{t-1} + e_t^*$	$\left(\int_0^1 G_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 G_c(s;\bar{c}) dG_c(s;\bar{c})$
$\mathbf{t}_{adj}$	$\tilde{v}_t = \rho\tilde{v}_{t-1} + \Delta f_t(\tilde{\theta})'\pi_1 + \Delta F_t(\tilde{\theta})'\pi_2 + \tilde{q}_t'\pi_3 + e_t^\dagger$	$\left(\int_0^1 G_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 G_c(s;\bar{c}) dG_c(s;\bar{c})$
$\mathbf{t}_{int}$	$\tilde{v}_t = \nu + \rho\tilde{v}_{t-1} + \Delta f_t(\tilde{\theta})'\pi_1 + \Delta F_t(\tilde{\theta})'\pi_2 + \tilde{q}_t'\pi_3 + e_t^\dagger$	$\left(\int_0^1 \bar{G}_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 \bar{G}_c(s;\bar{c}) dG_c(s;\bar{c})$
$\mathbf{t}_{LMS}^*$	$\tilde{v}_t^* = \nu + \rho\tilde{v}_{t-1}^* + e_t^{**}$	$\left(\int_0^1 \bar{G}_c^*(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 \bar{G}_c^*(s;\bar{c}) dG_c^*(s;\bar{c})$
$\mathbf{t}_{int}^*$	$\tilde{v}_t^* = \nu + \rho\tilde{v}_{t-1}^* + \Delta f_t(\tilde{\theta}^*)'\pi_1 + \Delta F_t(\tilde{\theta}^*)'\pi_2 + \tilde{q}_t^{*\prime}\pi_3 + e_t^{*\dagger}$	$\left(\int_0^1 \bar{G}_c^*(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 \bar{G}_c^*(s;\bar{c}) dG_c^*(s;\bar{c})$

$\bar{c} = -7$ as recommended by Elliott et al. (1996) in local power comparisons. In contrast, simulating (4.8) via $\tilde{v}_t^{*(0)} = y_t$, the latter distribution is, of course, invariant to $\bar{c}$ so that $\bar{c} = 0$ may be used without loss of generality. For the asymptotic distribution in (4.9), the optimal value of $\bar{c}$ found by Elliott et al. (1996) is $\bar{c} = -13.5$ and, hence, we also use that value later.

It remains to consider the role of $\bar{c}$ for the local power of the tests with asymptotic distributions (4.10) and (4.11). To determine a suitable $\bar{c}$ for these distributions we have generated critical values for a 5% significance level based on 10 000 drawings with sample size $T = 500$ using $c = 0$ and then we have simulated the local power curves in Figures 1 and 2. In Figure 1 it is seen that the local power associated with the distribution in (4.10) is almost invariant to the value of $\bar{c}$. Hence, $\bar{c} = 0$ may just as well be used. This feature was also reported by LLS. In other words, the deterministic terms may be estimated under the null rather than local alternatives in order to get optimal local power for $\mathbf{t}_{int}$. The same also holds for $\mathbf{t}_{LMS}^*$ and $\mathbf{t}_{int}^*$ as is seen in Figure 2. In that figure it is also obvious that for the latter statistics the value of $\bar{c}$ matters. However, optimal local power is achieved for $\bar{c} = 0$, at least for $c = 0, -3, \ldots, -30$. Thus, critical values for all relevant cases are available in Table 2 of LLS.

We will now consider the local power properties resulting from the five distributions in the Theorem with $\bar{c} = -7$ for (4.7), $\bar{c} = -13.5$ for (4.9) and $\bar{c} = 0$ for the remaining distributions. One would expect the tests based on the $\tilde{v}_t^{(i)}$ series to have better power than those based on the $\tilde{v}_t^{*(i)}$ because the former use an extra observation in estimating the nuisance parameters. Of course, tests based on the assumption $\mu_1 = 0$ are expected to have more power than the

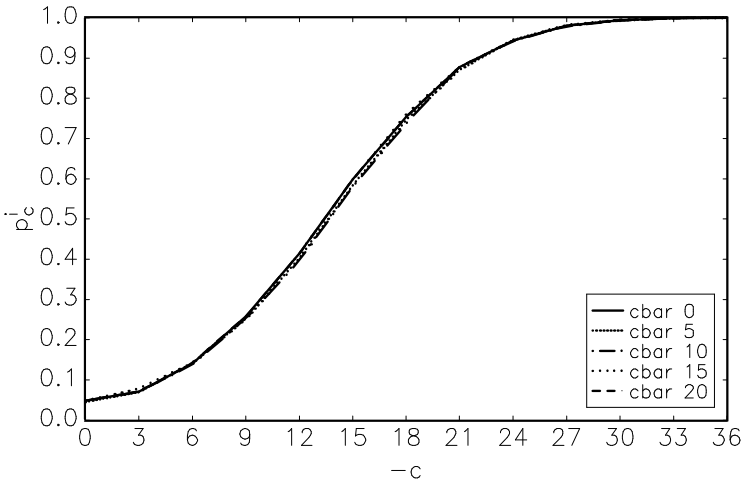

Figure 1. Local power associated with (4.10) ($\mathbf{t}_{int}$) ($-\bar{c} = 0, 5, 10, 15, 20$).

corresponding tests which do not use that a priori restriction. To explore these issues we have generated v_t as in (5.1) with different values of c and sample size $T = 500$. Comparing the resulting test values to the 5% critical values in Table 2 of LLS gives the empirical local power of the tests. The corresponding local power curves are plotted in Figure 3. They are again based on 10 000 replications of the simulation experiment.

The results in the figure are as expected. The tests which use the restriction $\mu_1 = 0$ are relatively more powerful than the corresponding ones which do not take the restriction into account. Moreover, tests which include an intercept term in the auxiliary regression (based on $\tilde{v}_t^{*(i)}$) tend to be less powerful than the corresponding tests which include the first observation in estimating the nuisance parameters (based on the $\tilde{v}_t^{(i)}$). Except for $\mathbf{t}_{LMS}^0$ and $\mathbf{t}_{adj}^0$, the differences in local power are in fact not very substantial. In other words, if a linear trend term cannot be excluded a priori, the price in terms of local power for dropping the first observation is not very high. On the other hand, substantial gains in local power are possible if $\mu_1 = 0$ can be assumed. In this case using (3.6) instead of (3.3) in estimating the nuisance parameters has a quite high price. Generally tests which include an intercept term in the underlying regression model have reduced local power. Of course, local power is a concept based on asymptotic considerations. In small samples the situation may be quite different, in particular, if the assumptions underlying the tests are not fully satisfied for a time series of interest. Therefore we will explore the small sample properties of the different

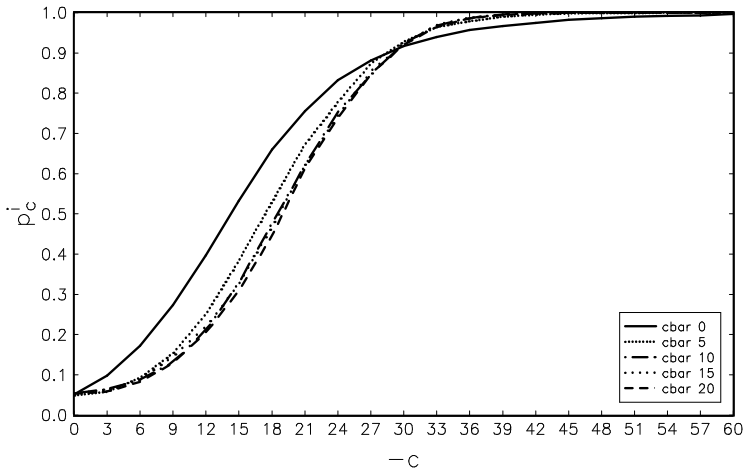

Figure 2. Local power associated with (4.11) $(\mathbf{t}^*_{LMS}, \mathbf{t}^*_{int})$ $(-\bar{c} = 0, 5, 10, 15, 20)$.

variants of the tests in the next section.

6 Small Sample Comparison

We have performed some simulations to investigate the properties of the tests in small samples based on the following two processes:

$$(1 - b_1 L)y_t = d_{1t} + v_t, \qquad v_t = \rho v_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{6.1}$$

and

$$y_t = d_{1t} + x_t, \qquad (1 - b_1 L)(1 - \rho L)x_t = \varepsilon_t, \qquad t = 1, \ldots, T, \tag{6.2}$$

with $\varepsilon_t \sim iid\, N(0, 1)$, $\rho = 1, 0.9, 0.8$, $T = 100, 200$. In some of the simulations we also generated 100 presample values which were discarded except that presample values were used in the estimations. Furthermore, we use $T_1/T = 0.5$, that is, the break point is in the middle of the sample. Preliminary simulations indicated that the location of the break point is not critical for the results as long as it is not very close to the beginning or the end of the sample. Therefore placing it in the middle does not imply a loss of generality for the situations we have in mind. The first process (6.1) is a special case of the model (2.1) and, for $b_1 \neq 0$, it generates a smooth shift in the deterministic term. For this process our $\mathbf{t}$ tests are appropriate. In contrast, the DGP (6.2) represents an additive outlier model

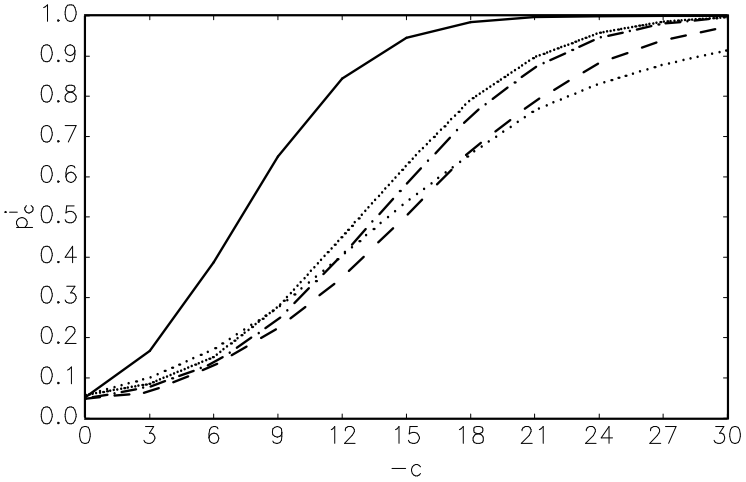

Figure 3. Local power of tests ($T = 500$).
(— (4.7) ($\mathbf{t}^0_{LMS}$, $\mathbf{t}^0_{adj}$), $\cdot - \cdot$ (4.8) ($\mathbf{t}^0_{int}$, $\mathbf{t}^{*0}_{LMS}$, $\mathbf{t}^{*0}_{int}$),
$\cdots\cdots$ (4.9) ($\mathbf{t}_{LMS}$, $\mathbf{t}_{adj}$), $- - -$ (4.10) ($\mathbf{t}_{int}$),
$\cdots$ (4.11) ($\mathbf{t}^*_{LMS}$, $\mathbf{t}^*_{adj}$))

with an abrupt shift at time T_1 so that in general the model underlying the $\mathbf{t}$ tests can only approximate the DGP (6.2). For this process, applying our tests may give some indication of the flexibility of the framework and of the consequences of violating the assumptions underlying the derivations of our tests.

For comparison purposes we also include the tests of LLS (denoted by τ) in our study. These tests are based on processes with potentially abrupt level shift as in (6.2). Specifically the tests used in the following are τ^0_{adj}, τ^{0+}_{adj}, τ_{int} and τ^+_{int} which were found to have the best small sample properties in LLS. These tests correspond to $\mathbf{t}^0_{adj}$ and $\mathbf{t}_{int}$, respectively. In particular, they have the same asymptotic properties. The difference between the τ_{int} and τ^+_{int} statistics is that the short-term dynamics are dealt with in a slightly different way and a similar comment applies for τ^0_{adj} and τ^{0+}_{adj} (see LLS for details). The test τ^+_{int} was shown to outperform other unit root tests for time series with level shifts such as those proposed by Perron (1989, 1990) in terms of size precision and power (see Lanne & Lütkepohl (2001)). Including τ^+_{int} in the present study we can thereby investigate the performance of our new tests more generally.

To capture the smooth transition from one regime to another as in DGP (6.1) the τ tests

may be combined with a smooth shift function. For both types of tests we use the shift functions given in Table 3 for both processes. The last two shift functions allow for smooth deterministic shifts. All three shift functions can be shown to satisfy Assumption 1. For some of the tests the derivatives of the shift functions are needed. They are also given in Table 3. Since $f_t^{(1)}$ does not depend on θ the derivative $F_t^{(1)}$ is zero. Hence, no extra terms $\Delta F_t^{(1)}(\theta)$ appear in those auxiliary regressions which generally contain this term. In the simulations we use a range of $0 < \theta < 2$ for $f_t^{(2)}(\theta)$ and $0 < \theta < 0.8$ for $f_t^{(3)}(\theta)$ in estimating the parameters of the deterministic term. Although there is no linear trend term in the DGPs we allow for such a term in computing some of the test statistics.

Table 3. Shift Functions and Their Derivatives

Shift function $f_t(\theta)$	Derivatives $F_t(\theta)$
$f_t^{(1)}(\theta) = d_{1t} = \begin{cases} 0, & t < T_1 \\ 1, & t \geq T_1 \end{cases}$	$F_t^{(1)}(\theta) = 0$
$f_t^{(2)}(\theta) = \begin{cases} 0, & t < T_1 \\ 1 - \exp\{-\theta(t - T_1)\}, & t \geq T_1 \end{cases}$	$F_t^{(2)}(\theta) = \begin{cases} 0, & t < T_1 \\ (t - T_1)\exp\{-\theta(t - T_1)\}, & t \geq T_1 \end{cases}$
$f_t^{(3)}(\theta) = \begin{bmatrix} \frac{d_{1,t}}{1-\theta L} \\ \frac{d_{1,t-1}}{1-\theta L} \end{bmatrix}$	$F_t^{(3)}(\theta) = \begin{bmatrix} \frac{d_{1,t-1}}{(1-\theta L)^2} \\ \frac{d_{1,t-2}}{(1-\theta L)^2} \end{bmatrix}$

As in LLS we found that nonzero $\bar{c}$ values lead to gross size distortions and in particular to substantial overrejection in some situations. Therefore, nonzero $\bar{c}$ cannot be recommended for applied work. Hence, in the following discussion we exclusively focus on results for $\bar{c} = 0$, that is, estimation of the nuisance parameters is done under the null hypothesis.

Relative rejection frequencies from 1000 replications of the experiment are given in Tables 4 - 7. Results are given for selected tests only. We will first comment on Tables 4 and 5 where the initial values used in the simulations are randomized by simulating 100 presample values as described previously. In the tables we only show the results for those tests which performed overall best in terms of small sample power within their respective groups, the

Table 4. Relative Rejection Frequencies of Tests, $T = 100$, $T_1 = 50$, $\bar{c} = 0$, Nominal Significance Level 5%, Random Initial Values

Shift function	Test	DGP (6.1), $b_1 = 0.5$			DGP (6.1), $b_1 = 0.8$			DGP (6.2), $b_1 = 0.5$			DGP (6.2), $b_1 = 0.8$		
		$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8
$f_t^{(1)}$	$\mathbf{t}_{int}^0$	0.060	0.289	0.577	0.065	0.227	0.345	0.054	0.292	0.561	0.067	0.234	0.354
	τ_{adj}^0	0.040	0.285	0.527	0.020	0.140	0.275	0.039	0.291	0.535	0.016	0.156	0.315
	τ_{adj}^{+0}	0.061	0.343	0.575	0.053	0.287	0.382	0.063	0.353	0.590	0.050	0.292	0.436
	$\mathbf{t}_{int}$	0.079	0.216	0.468	0.079	0.161	0.269	0.081	0.217	0.455	0.077	0.159	0.268
	τ_{int}	0.022	0.091	0.305	0.001	0.004	0.029	0.020	0.090	0.302	0.000	0.006	0.034
	τ_{int}^+	0.075	0.216	0.499	0.064	0.149	0.262	0.080	0.233	0.526	0.065	0.167	0.286
$f_t^{(2)}$	$\mathbf{t}_{int}^0$	0.095	0.306	0.535	0.088	0.227	0.366	0.100	0.286	0.547	0.099	0.238	0.361
	τ_{adj}^0	0.064	0.259	0.486	0.046	0.158	0.281	0.063	0.247	0.486	0.036	0.145	0.263
	τ_{adj}^{+0}	0.072	0.266	0.491	0.049	0.176	0.299	0.069	0.253	0.496	0.042	0.166	0.281
	$\mathbf{t}_{int}$	0.134	0.290	0.505	0.134	0.244	0.355	0.135	0.288	0.509	0.141	0.237	0.330
	τ_{int}	0.051	0.157	0.358	0.022	0.040	0.085	0.055	0.142	0.348	0.018	0.034	0.059
	τ_{int}^+	0.059	0.160	0.371	0.026	0.051	0.116	0.059	0.150	0.362	0.031	0.048	0.080
$f_t^{(3)}$	$\mathbf{t}_{int}^0$	0.108	0.268	0.445	0.101	0.217	0.304	0.110	0.259	0.435	0.105	0.208	0.293
	τ_{adj}^0	0.060	0.268	0.426	0.082	0.217	0.293	0.064	0.266	0.417	0.079	0.223	0.302
	τ_{adj}^{+0}	0.056	0.252	0.418	0.036	0.140	0.243	0.059	0.249	0.404	0.037	0.144	0.249
	$\mathbf{t}_{int}$	0.135	0.279	0.474	0.129	0.229	0.322	0.134	0.278	0.468	0.140	0.213	0.314
	τ_{int}	0.062	0.146	0.325	0.072	0.091	0.134	0.060	0.141	0.322	0.074	0.086	0.133
	τ_{int}^+	0.052	0.129	0.317	0.014	0.029	0.068	0.048	0.120	0.314	0.016	0.028	0.064

groups being $\mathbf{t}^0$ tests (i.e., tests without linear trend term) and $\mathbf{t}$ tests (i.e., tests with linear trend). We are only presenting the best tests in the tables to avoid covering up the most important findings by the large volume of results for all the tests and simulation designs. It may be worth noting, however, that some of the other tests were nearly as good as the tests shown in the tables whereas some other tests performed very poorly indeed. Thus, some of the other tests are not very useful for applied work whereas some other ones are almost as good as those presented in the tables.

In the following, we consider only $\mathbf{t}_{int}^0$, τ_{adj}^0, τ_{adj}^{+0}, $\mathbf{t}_{int}$, τ_{int} and τ_{int}^+. From LLS we know that in the group of tests which exclude the deterministic trend term, τ_{adj}^0 and τ_{adj}^{+0} were generally best in terms of power, each having advantages in some situations. Among the $\mathbf{t}^0$ tests, $\mathbf{t}_{int}^0$ was overall clearly best with highest power most of the time and close to the maximum in the other situations. Note also that its empirical size tends to exceed the nominal 5% slightly but not very much. More precisely, in most cases with $\rho = 1$ its relative rejection frequency is less than 10% and in no case it is more than 13%.

In the group of τ tests which allow for a linear trend term, τ_{int} and τ_{int}^+ dominated the

Table 5. Relative Rejection Frequencies of Tests, $T = 200$, $T_1 = 100$, $\bar{c} = 0$, Nominal Significance Level 5%, Random Initial Values

Shift function	Test	DGP (6.1), $b_1 = 0.5$			DGP (6.1), $b_1 = 0.8$			DGP (6.2), $b_1 = 0.5$			DGP (6.2), $b_1 = 0.8$		
		$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8
$f_t^{(1)}$	$\mathbf{t}_{int}^0$	0.053	0.703	0.964	0.058	0.519	0.781	0.055	0.698	0.966	0.054	0.495	0.799
	τ_{adj}^0	0.033	0.633	0.850	0.017	0.472	0.650	0.035	0.653	0.867	0.017	0.494	0.726
	τ_{adj}^{+0}	0.044	0.678	0.869	0.040	0.556	0.706	0.050	0.693	0.878	0.041	0.591	0.774
	$\mathbf{t}_{int}$	0.061	0.568	0.913	0.062	0.394	0.687	0.059	0.556	0.910	0.058	0.387	0.675
	τ_{int}	0.024	0.466	0.899	0.007	0.125	0.476	0.028	0.469	0.907	0.009	0.130	0.513
	τ_{int}^+	0.053	0.613	0.933	0.061	0.415	0.732	0.050	0.617	0.943	0.064	0.434	0.753
$f_t^{(2)}$	$\mathbf{t}_{int}^0$	0.073	0.631	0.907	0.067	0.476	0.708	0.071	0.629	0.904	0.068	0.474	0.724
	τ_{adj}^0	0.053	0.608	0.810	0.028	0.464	0.656	0.051	0.610	0.819	0.028	0.472	0.684
	τ_{adj}^{+0}	0.053	0.614	0.812	0.029	0.474	0.664	0.051	0.614	0.824	0.033	0.482	0.691
	$\mathbf{t}_{int}$	0.074	0.559	0.867	0.077	0.393	0.658	0.071	0.558	0.877	0.075	0.398	0.682
	τ_{int}	0.047	0.491	0.896	0.011	0.161	0.511	0.046	0.489	0.844	0.011	0.142	0.509
	τ_{int}^+	0.043	0.497	0.889	0.020	0.195	0.534	0.044	0.492	0.891	0.020	0.174	0.531
$f_t^{(3)}$	$\mathbf{t}_{int}^0$	0.065	0.536	0.802	0.067	0.389	0.596	0.068	0.534	0.809	0.074	0.385	0.625
	τ_{adj}^0	0.045	0.553	0.742	0.050	0.492	0.637	0.049	0.571	0.747	0.051	0.507	0.656
	τ_{adj}^{+0}	0.043	0.542	0.737	0.026	0.447	0.619	0.046	0.563	0.741	0.028	0.463	0.635
	$\mathbf{t}_{int}$	0.079	0.552	0.861	0.082	0.399	0.645	0.075	0.545	0.862	0.075	0.389	0.634
	τ_{int}	0.042	0.488	0.861	0.075	0.237	0.529	0.051	0.484	0.854	0.075	0.241	0.512
	τ_{int}^+	0.034	0.473	0.861	0.013	0.170	0.497	0.036	0.479	0.856	0.012	0.167	0.486

other tests in the study of LLS. Whereas τ_{int} turned out to be preferable in conjunction with shift function $f_t^{(3)}$, τ_{int}^+ clearly dominated for $f_t^{(1)}$. Both tests performed poorly for $b_1 = 0.8$ and $T = 100$. Finally, $\mathbf{t}_{int}$ is overall the best $\mathbf{t}$ test allowing for a trend. Its power is usually very close to that of $\mathbf{t}_{int}^*$, though. In fact, the two tests often produce identical rejection frequencies. Therefore, we present results for just one of them. For $T = 100$, both $\mathbf{t}_{int}$ and $\mathbf{t}_{int}^*$ reject a bit too often if they are used in conjunction with $f_t^{(2)}$ and $f_t^{(3)}$. This may not be too surprising given that using these shift functions for the presently considered DGP means that we are fitting a misspecified model. The tests are doing quite well if the correct shift function $f_t^{(1)}$ is used.

The following further conclusions emerge from Tables 4 and 5. In line with the local power results, excluding a linear trend term from the models when such a restriction is correct results in substantially better power. Although there are power differences between the tests which allow for a linear trend, there is no clear winner. In other words, each of the tests is advantageous in some situations. On the one hand, $\mathbf{t}_{int}$ has often more power than τ_{int} and τ_{int}^+ and, on the other hand, $\mathbf{t}_{int}$ tends to reject a bit too often. The same is true for

tests excluding a linear trend term.

It is also apparent that it is not essential to use a test designed for a particular model when that model is in fact the true DGP. In other words, the performance of the tests is similar for the alternative DGPs (6.1) and (6.2). This may not be very surprising given that the two models are in fact quite close in many respects. The results in Tables 4 and 5 show that the performance of the tests depends more strongly on the shift functions than on the type of DGP. Furthermore, changing b_1 from 0.5 to 0.8 has a substantial effect. It implies a sizable decline in power in most cases. Again, this behaviour of the tests may not be too surprising because for b_1 close to 1 the processes have two roots close to unity and therefore are difficult to distinguish from unit root processes. Finally, the performance of all the tests improves markedly if T is increased from 100 to 200.

It is noteworthy that the tests based on (3.6) in estimating the nuisance parameters ($\mathbf{t}^{*0}_{LMS}$, $\mathbf{t}^{*0}_{int}$, $\mathbf{t}^{*}_{LMS}$, $\mathbf{t}^{*}_{int}$), were not as powerful as those shown in Tables 4 and 5. This result is in line with the local power results. To explore the impact of the initial values we have controlled them in some of our simulations. In Table 6, results for $T = 100$ and zero initial values are provided. Clearly, the power of the tests tends to be larger than in the corresponding entries in Table 4, especially for those tests which do not allow for a linear trend. Thus, using zero initial values helps to improve power in samples of size $T = 100$ although they have no impact asymptotically.

Since $T = 100$ is obviously too small to ensure the validity of asymptotic properties, it is also not surprising that the power tends to be smaller if unusually large initial values are considered. In Table 7 we show results for the situation where the initial values are all set to 5. For $\rho = 0.9$ and 0.8, the standard deviations of the y_t generated by (6.2) range from about 3 to almost 10, depending on b_1. Hence, initial values of 5 may be regarded as moderate or large compared to the randomly chosen values in Tables 4 and 5. Using identical values for y_{-1} and y_0 may be reasonable given the large correlation in the y_t. In Table 7 the power of the tests tends to be lower than in the corresponding Table 4. In some situations the power decline is particularly strong for tests that do not include an intercept term in the test regression ($\tau^{0}_{adj}, \tau^{+0}_{adj}$). A similar problem was also observed for some of the other tests based on regressions without an intercept and for which results are not shown in the tables. It may also be worth noting that the relative performance of the tests changed if zero initial

Table 6. Relative Rejection Frequencies of Tests, $T = 100$, $T_1 = 50$, $\bar{c} = 0$, Nominal Significance Level 5%, Zero Initial Values

Shift function	Test	DGP (6.1), $b_1 = 0.5$			DGP (6.1), $b_1 = 0.8$			DGP (6.2), $b_1 = 0.5$			DGP (6.2), $b_1 = 0.8$		
		$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8
$f_t^{(1)}$	$\mathbf{t}_{int}^0$	0.050	0.271	0.594	0.053	0.208	0.329	0.047	0.257	0.596	0.046	0.197	0.327
	τ_{adj}^0	0.044	0.446	0.730	0.017	0.210	0.404	0.049	0.512	0.845	0.015	0.259	0.584
	τ_{adj}^{+0}	0.069	0.560	0.785	0.067	0.427	0.580	0.067	0.617	0.889	0.066	0.487	0.751
	$\mathbf{t}_{int}$	0.067	0.244	0.561	0.067	0.170	0.319	0.067	0.245	0.583	0.073	0.171	0.310
	τ_{int}	0.019	0.091	0.317	0.002	0.004	0.032	0.021	0.094	0.325	0.001	0.005	0.043
	τ_{int}^+	0.070	0.229	0.559	0.070	0.190	0.317	0.072	0.238	0.571	0.072	0.181	0.321
$f_t^{(2)}$	$\mathbf{t}_{int}^0$	0.088	0.296	0.560	0.095	0.203	0.344	0.083	0.288	0.570	0.085	0.219	0.336
	τ_{adj}^0	0.057	0.414	0.668	0.032	0.246	0.435	0.055	0.459	0.782	0.024	0.248	0.500
	τ_{adj}^{+0}	0.064	0.427	0.681	0.034	0.271	0.451	0.059	0.476	0.786	0.024	0.269	0.517
	$\mathbf{t}_{int}$	0.129	0.316	0.562	0.126	0.251	0.370	0.136	0.340	0.588	0.130	0.260	0.379
	τ_{int}	0.046	0.142	0.349	0.011	0.037	0.088	0.041	0.150	0.366	0.009	0.033	0.068
	τ_{int}^+	0.053	0.155	0.373	0.021	0.045	0.115	0.052	0.157	0.391	0.017	0.038	0.083
$f_t^{(3)}$	$\mathbf{t}_{int}^0$	0.106	0.230	0.409	0.115	0.170	0.257	0.112	0.213	0.400	0.120	0.174	0.250
	τ_{adj}^0	0.061	0.402	0.559	0.055	0.312	0.463	0.058	0.420	0.630	0.049	0.319	0.496
	τ_{adj}^{+0}	0.054	0.375	0.550	0.027	0.237	0.401	0.052	0.394	0.624	0.023	0.245	0.457
	$\mathbf{t}_{int}$	0.128	0.272	0.475	0.136	0.224	0.318	0.126	0.276	0.467	0.132	0.211	0.308
	τ_{int}	0.054	0.153	0.360	0.051	0.094	0.135	0.054	0.157	0.379	0.046	0.076	0.133
	τ_{int}^+	0.043	0.132	0.348	0.009	0.029	0.078	0.043	0.140	0.365	0.009	0.026	0.072

Table 7. Relative Rejection Frequencies of Tests, $T = 100$, $T_1 = 50$, $\bar{c} = 0$, Nominal Significance Level 5%, Initial Values 5

Shift function	Test	DGP (6.1), $b_1 = 0.5$			DGP (6.1), $b_1 = 0.8$			DGP (6.2), $b_1 = 0.5$			DGP (6.2), $b_1 = 0.8$		
		$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8
$f_t^{(1)}$	$\mathbf{t}_{int}^0$	0.067	0.329	0.659	0.062	0.243	0.410	0.064	0.261	0.593	0.066	0.183	0.311
	τ_{adj}^0	0.042	0.044	0.053	0.016	0.166	0.151	0.049	0.126	0.104	0.015	0.180	0.234
	τ_{adj}^{+0}	0.064	0.066	0.070	0.046	0.349	0.272	0.067	0.168	0.147	0.066	0.372	0.405
	$\mathbf{t}_{int}$	0.076	0.166	0.311	0.080	0.133	0.177	0.070	0.208	0.421	0.077	0.169	0.263
	τ_{int}	0.021	0.060	0.155	0.001	0.002	0.017	0.021	0.084	0.265	0.001	0.004	0.034
	τ_{int}^+	0.068	0.189	0.292	0.062	0.222	0.298	0.072	0.214	0.429	0.072	0.179	0.293
$f_t^{(2)}$	$\mathbf{t}_{int}^0$	0.095	0.355	0.665	0.104	0.270	0.413	0.094	0.292	0.580	0.095	0.204	0.330
	τ_{adj}^0	0.062	0.041	0.059	0.037	0.164	0.105	0.055	0.162	0.153	0.024	0.190	0.260
	τ_{adj}^{+0}	0.069	0.045	0.063	0.055	0.194	0.120	0.059	0.165	0.168	0.024	0.203	0.270
	$\mathbf{t}_{int}$	0.116	0.190	0.272	0.146	0.181	0.197	0.122	0.243	0.373	0.133	0.222	0.300
	τ_{int}	0.047	0.096	0.166	0.011	0.032	0.060	0.041	0.117	0.231	0.009	0.032	0.066
	τ_{int}^+	0.051	0.116	0.160	0.025	0.106	0.140	0.052	0.126	0.226	0.017	0.042	0.082
$f_t^{(3)}$	$\mathbf{t}_{int}^0$	0.111	0.333	0.585	0.125	0.231	0.362	0.112	0.268	0.506	0.120	0.185	0.288
	τ_{adj}^0	0.072	0.053	0.050	0.185	0.279	0.137	0.058	0.182	0.170	0.048	0.256	0.296
	τ_{adj}^{+0}	0.056	0.047	0.045	0.027	0.142	0.087	0.052	0.161	0.157	0.023	0.189	0.245
	$\mathbf{t}_{int}$	0.117	0.253	0.391	0.139	0.199	0.260	0.126	0.268	0.435	0.132	0.216	0.298
	τ_{int}	0.057	0.144	0.216	0.173	0.283	0.264	0.053	0.137	0.262	0.046	0.081	0.121
	τ_{int}^+	0.045	0.111	0.199	0.012	0.055	0.121	0.043	0.124	0.248	0.009	0.031	0.071

values instead of random initial values are used. In that case some tests without intercept term in the test regression did a little better than in the nonzero initial value case. Thus, in particular if unusual initial values are suspected, using one of the tests with intercept term in the test regression is advisable. Alternatively one may remove the first values of a time series under consideration if they appear to be unusual.

The results in Tables 4, 6 and 7 also show that the tests are generally not very reliable if time series with $T = 100$ observations are under consideration. Moreover, the performance of the tests tends to be inferior if one of the misspecified and more complicated shift functions $f_t^{(2)}$ or $f_t^{(3)}$ is used.

7 Conclusions

Standard unit root tests are known to have reduced power if they are applied to time series with outliers and structural shifts. Therefore we have considered unit root tests that explicitly allow for innovational outliers of a very general, possibly nonlinear form at a known point in time. We have argued that knowing the timing of the shift is quite common in practice whereas the precise form of the shift is usually unknown. Therefore, allowing for general and flexible shift functions is important. In this study we have focussed on models where the shift is regarded as part of the deterministic component of the DGP. Building on proposals by Saikkonen & Lütkepohl (2001), LMS and LLS it is suggested to estimate the deterministic part in a first step by a GLS procedure which may proceed under local alternatives or under the unit root null hypothesis. The original series is adjusted in a second step by subtracting the estimated deterministic part and accounting for the stationary dynamics. Then DF type tests are applied to the adjusted series. A number of modifications of previously proposed tests of this sort are considered. In particular, tests are proposed that take into account estimation errors in the nuisance parameters. Local power and small sample properties of the tests are obtained.

The following general results emerge from our study. Some of the suggested modifications work clearly better in small samples than the original tests proposed by LMS in that they have superior size and power properties. Although local power gains are possible for some of the tests if the nuisance parameters are estimated under local alternatives rather than

under the null hypothesis, substantial size distortions may result in small samples in the former case. Therefore we recommend estimating the nuisance parameters under the null hypothesis.

A comparison with similar tests based on additive outlier models with a possibly abrupt shift shows that neither of the test versions dominates in all situations. Therefore it may be reasonable to use both of them in practice although the tests proposed in this study have some disadvantages in terms of size in small samples.

Initial values are found to have an impact on the small sample power of the tests. It turns out that including an intercept term in the test regression is important to guard against undesirable effects of large initial values. In practice, it may be worth discarding unusual values at the very beginning of a time series under consideration to avoid a loss in power due to untypical initial values.

If a deterministic linear time trend can be excluded on a priori grounds, it is recommended to perform tests in models without a linear trend term because excluding it may result in sizable power gains. Finally, using test versions with the best power properties is of particular importance in the present context because in some situations the tests do not perform very well for samples of size as large as $T = 100$.

Although we have focussed on a single shift in a time series, the tests can in principle be extended to allow for more than one shift. Of course, the small sample behaviour may be different in this case and needs to be explored in the future if applied researchers wish to use the tests in this more general context. In future research it may also be of interest to consider the situation where the timing of the shift is unknown and has to be determined from the data. We leave these issues for future investigations.

Appendix. Proof of Theorem

We focus on the case where μ_1 is not zero a priori. The result for $\mathbf{t}_{LMS}$ in (4.9) can be obtained from LMS so that we consider $\mathbf{t}_{adj}$.

We shall first study the appropriately standardized moment matrix in the OLS estimation of the parameters in (4.3). By Lemma A.1 of LMS, we have $\tilde{\mu}_* = \mu_* + O_p(T^{-1/2})$, where $\mu_* = \mu_1/b(1)$. This result in conjunction with (12.A.10) and arguments similar to those in

(12.A.14) of LMS, can be used to show that

$$T^{-1}\sum_{t=1}^{T}\tilde{q}_t\tilde{q}_t' \xrightarrow{p} \sigma^2\Sigma(b),$$

where $\Sigma(b) = \sigma^{-2}\mathrm{Cov}[b(L)^{-1}(\varepsilon_1,\ldots,\varepsilon_p)]$. The same arguments and the representation given for $\tilde{v}_t$ in the proof of Theorem 1 of LMS yield

$$T^{-3/2}\sum_{t=1}^{T}\tilde{v}_{t-1}\tilde{q}_t' = o_p(1).$$

Finally, when the assumptions made for $f_t(\theta)$ and $F_t(\theta)$ are also used, we get

$$T^{-1}\sum_{t=1}^{T}\tilde{v}_{t-1}\left[\Delta f_t(\tilde{\theta})' : \Delta F_t(\tilde{\theta})'\right] = o_p(1)$$

and

$$T^{-1/2}\sum_{t=1}^{T}\tilde{q}_t\left[\Delta f_t(\tilde{\theta})' : \Delta F_t(\tilde{\theta})'\right] = o_p(1).$$

Thus, the appropriately standardized moment matrix between the three regressors $\tilde{v}_{t-1}$, $\left[\Delta f_t(\tilde{\theta})' : \Delta F_t(\tilde{\theta})'\right]'$ and $\tilde{q}_t$ is asymptotically block diagonal. It is also asymptotically positive definite, as can be seen by using the assumptions and arguments similar to those in the proof of Lemma A.1 of LMS.

We shall next consider the error term $e_t^\dagger$ in (4.3) and show how it is related to the error term e_t^* in (4.1). First, recall that $\rho_T(L) = 1 - \rho_T L$ and observe that, for $t \geq 2$,

$$\begin{aligned} e_t^* &= \varepsilon_t + \left(\tilde{b}(L) - b(L)\right)\left(\Delta y_t - \tfrac{c}{T}y_{t-1}\right) + \tfrac{c}{T}\left(\tilde{\mu}_0 - \mu_0\right) \\ &\quad - \left(\tilde{\mu}_1 - \mu_1\right)\left(1 - \tfrac{c(t-1)}{T}\right) - \left(\rho_T(L)f_t(\tilde{\theta})'\tilde{\gamma} - \rho_T(L)f_t(\theta)'\gamma\right). \end{aligned} \tag{A.1}$$

Here we have used the definitions of v_t and $\tilde{v}_t$ and the identities $\rho_T(L)y_t = \Delta y_t - \frac{c}{T}y_{t-1}$ and $\rho_T(L)t = 1 - \frac{c(t-1)}{T}$. Identifying the parameter vector π_3 in (4.3) with $-(\tilde{b} - b)$ shows that the inclusion of the regressors $\Delta y_{t-j} - \tilde{\mu}_*$ $(j = 1,\ldots,p)$ in (4.1) changes the second term on the r.h.s. of $(A.1)$ to $(\tilde{b}(L) - b(L))(\tilde{\mu}_* - \frac{c}{T}y_{t-1})$. It is also easy to see that, as far as the limiting distribution of the test statistics $\mathbf{t}_{LMS}$ and $\mathbf{t}_{adj}$ is concerned, the contribution of the third and fifth terms on the r.h.s. of $(A.1)$ is negligible. In the same way one can conclude that, from our point of view, terms which are added to the error term $e_t^\dagger$ by including the regressors $\Delta f_t(\tilde{\theta})$ and $\Delta F_t(\tilde{\theta})$ are asymptotically negligible. Thus, we can conclude that for our purposes the error term $e_t^\dagger$ can be treated by using the approximation

$$e_t^\dagger \approx \varepsilon_t - \left(\tilde{b}(L) - b(L)\right)\left(\tilde{\mu}_* - \frac{c}{T}y_{t-1}\right) - \left(\tilde{\mu}_1 - \mu_1\right)\left(1 - \frac{c(t-1)}{T}\right). \tag{A.2}$$

Using Lemma A.1 and equation (12.A.10) of LMS in conjunction with the representation given for $\tilde{v}_t$ in the proof of Theorem 1 of the same paper it can be shown that $e_t^\dagger$ can be further approximated by replacing $\tilde{\mu}_*$ on the r.h.s. of $(A.2)$ first by μ_* and then by Δy_t. Since the third and fifth terms on the r.h.s. of $(A.1)$ can be ignored we have thus demonstrated that the above approximation of $e_t^\dagger$ becomes $e_t^\dagger \approx e_t^*$ and, since the appropriately standardized moment matrix in the OLS estimation of (4.3) is asymptotically block diagonal, it follows that the OLS estimators obtained for the parameter ρ from (4.1) and (4.3) are asymptotically equivalent. Since it is straightforward to show that the same is true for the related error variance estimators the limiting distribution of $\mathbf{t}_{adj}$ follows.

Using the definition of the test statistic $\mathbf{t}_{int}$ and the above arguments it is straightforward to prove (4.10). In order to prove (4.11) we first need the following result.

Lemma A.1.

Suppose that the assumptions of the Theorem hold. Then,

$$\tilde{b}^* \xrightarrow{p} b, \tag{A.3}$$

$$\tilde{\theta}^* = \theta + O_p(1), \tag{A.4}$$

$$\tilde{\gamma}^* = \gamma + O_p(1) \tag{A.5}$$

and

$$T^{1/2}\left(\tilde{\mu}_1^* - \tilde{b}^*(1)\mu_1/b(1)\right) \xrightarrow{d} \sigma K_c^*(\bar{c}), \tag{A.6}$$

where

$$K_c^*(\bar{c}) = \frac{12}{\bar{c}}\int_0^1 \left(\frac{1}{2} - s\right) dB_0(s) + \frac{12(c-\bar{c})}{\bar{c}}\int_0^1 \left(\frac{1}{2} - s\right) B_c(s)ds$$

for $\bar{c} < 0$ and $K_c^*(0) = B_c(1)$.

Proof: For $\bar{c} = 0$ the result is fairly obvious because the considered regression model differs from that used in Lemma A.1 of LMS only in that the first observation is omitted.

Now suppose that $\bar{c} < 0$. Arguments entirely similar to those used in the proof of Lemma A.1 of LMS show that the estimation of the parameters γ and θ is asymptotically orthogonal to the estimation of the other parameters in the regression model (3.6). Given this, $(A.4)$ and $(A.5)$ can be obtained in the same way as in the aforementioned previous case, whereas

the asymptotic properties of the estimators $\tilde{b}^*$ and $\tilde{\mu}_1^*$ can be obtained as if the matrix $Z_2^*(\theta)$ were omitted from $Z^*(\theta)$ and hence from $W^*(\theta)$. Since the first column of the matrix Z_1^* is constant, this implies that we only need to repeat the derivations given in the proof of Lemma A.1 of LMS for the counterparts of the estimators $\tilde{b}^*$ and $\tilde{\mu}_1^*$ by using mean-adjusted versions of the relevant variables. This argument gives $(A.3)$ and $(A.6)$ because the omission of the first observation has no effect on these (asymptotic) results.

Using equation (12.A.11) of LMS it is straightforward to check that the mean-adjusted versions of the second sample moments in (12.A.6), (12.A.14) and (12.A.15) of LMS converge to the same limits as before except that the constant $h(\bar{c}) = 1 - \bar{c} + \bar{c}^2/3$ has to be replaced by $\bar{c}^2/12$, the limit of the mean-adjusted version of the l.h.s. of (12.A.6) of LMS. Thus, it follows that instead of (12.A.19) of LMS we can now write

$$\begin{bmatrix} T^{1/2}(\tilde{b}^* - b) \\ T^{1/2}(\tilde{\mu}_1^* - \mu_1) \end{bmatrix} = \bar{R}_{11} \begin{bmatrix} T^{-1/2}\bar{V}'\mathcal{E} \\ T^{-1/2}\bar{Z}_{12}'\mathcal{E} \end{bmatrix} + o_p(1), \qquad (A.7)$$

where $\bar{Z}_{12}$ is the mean-adjusted version of the second column of the matrix Z_1 with a typical component $\bar{c}(\frac{1}{2} - \frac{t}{T}) + \frac{\bar{c}}{2T}$ and

$$\bar{R}_{11} = \begin{bmatrix} \sigma^2\Sigma(b) + \mu_*^2\frac{\bar{c}^2}{12}\mathbf{1}_p\mathbf{1}_p' & \mu_*\frac{\bar{c}^2}{12}\mathbf{1}_p \\ \mu_*\frac{\bar{c}^2}{12}\mathbf{1}_p' & \frac{\bar{c}^2}{12} \end{bmatrix}.$$

Here $\mathbf{1}_p = [1 : \cdots : 1]'$ is $(p \times 1)$ and the other notation is as before. Analogously to the treatment of equation (12.A.19) of LMS we can now premultiply equation $(A.7)$ by $[\mu_*\mathbf{1}_p' : 1]$ to obtain

$$T^{1/2}\left(\tilde{\mu}_1^* - \tilde{b}^*(1)\mu_1/b(1)\right) = \frac{12}{\bar{c}^2}T^{-1/2}\bar{Z}_{12}'\mathcal{E} + o_p(1). \qquad (A.8)$$

Now recall that a typical component of the vector $\bar{Z}_{12}$ is $\bar{c}(\frac{1}{2} - \frac{t}{T}) + \frac{\bar{c}}{2T}$. Using this fact, the definition of the vector $\mathcal{E}$ and well-known limit theorems for stationary and nearly integrated processes, it is straightforward to derive $(A.6)$ from $(A.8)$. As for $(A.3)$, the proof can be readily obtained by using $(A.7)$ and equation (12.A.11) of LMS which defines the elements of the matrix V. This completes the proof of Lemma A.1. □

Once the result of Lemma A.1 is available, the limiting distribution of the test statistic $\mathbf{t}_{LMS}^*$ can be obtained by following the arguments in the proof of Theorem 1 of LMS. Specifically, a representation similar to that obtained for $\tilde{v}_t$ in LMS can now be used for $\tilde{v}_t^*$

to get

$$T^{-1/2}\tilde{v}^*_{[Ts]} = T^{-1/2}v^*_{[Ts]} - T^{1/2}\left(\tilde{\mu}^*_1 - \tilde{b}(1)\mu_1/b(1)\right)\frac{[Ts]}{T} + o_p(1) \xrightarrow{d} \sigma G^*_c(s;\bar{c}).$$

Given this result, it is straightforward to proceed in the same way as in the proofs of Theorem 1 in Saikkonen & Lütkepohl (2001) and Theorem 1 of LMS to show that the limiting distribution of the test statistic $\mathbf{t}^*_{LMS}$ is a "mean-adjusted version" of that of $\mathbf{t}_{LMS}$. In other words, $\mathbf{t}^*_{LMS}$ has the limiting distribution given in (4.11). It follows from the fact that the regressors $\Delta f_t(\tilde{\theta}^*)$, $\Delta F_t(\tilde{\theta}^*)$ and $\tilde{q}^*_t$ are asymptotically orthogonal to $\tilde{v}^*_{t-1}$ and the constant term that the test statistic $\mathbf{t}^*_{int}$ has the same limiting distribution. These facts can be seen by using the results in the appendix of LMS in the same way as in the case of the test statistic $\mathbf{t}_{adj}$ and in the proof of Lemma A.1.

Finally, the arguments used for the test statistic $\mathbf{t}^*_{int}$ above also apply to the test statistic $\mathbf{t}_{int}$. However, since the test statistic $\mathbf{t}_{int}$ is based on the series $\tilde{v}_t$ instead of $\tilde{v}^*_t$, the function $\bar{G}_c(s;\bar{c})$ will appear in place of the function $\bar{G}^*_c(s;\bar{c})$.

References

Amsler, C. & J. Lee (1995), An LM test for a unit root in the presence of a structural change, *Econometric Theory*, 11, 359 - 368.

Banerjee, A., R.L. Lumsdaine & J.H. Stock (1992), Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and international evidence, *Journal of Business & Economic Statistics*, 10, 271 - 287.

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

Fuller, W.A. (1976), *Introduction to Statistical Time Series*, New York: Wiley.

Hansen, G. (1993), *Quantitative Wirtschaftsforschung*, München: Vahlen.

Lanne, M. & H. Lütkepohl (2001), Unit root tests for time series with level shifts: A comparison of different proposals, SFB Discussion Paper No. 5, Humboldt University Berlin.

Lanne, M., H. Lütkepohl & P. Saikkonen (2001), Comparison of unit root tests for time series with level shifts, *Journal of Time Series Analysis*, forthcoming.

Leybourne, S., P. Newbold & D. Vougas (1998), Unit roots and smooth transitions, *Journal of Time Series Analysis*, 19, 83 - 97.

Lütkepohl, H., C. Müller & P. Saikkonen (2001), Unit root tests for time series with a structural break when the break point is known, in C. Hsiao, K. Morimune & J. Powell (eds.), *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, Cambridge: Cambridge University Press, pp. 327 - 348.

Montañés, A. & M. Reyes (1998), Effect of a shift in the trend function on Dickey-Fuller unit root tests, *Econometric Theory*, 14, 355 - 363.

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 & Economic Statistics*, 8, 153 - 162.

Perron, P. & T.J. Vogelsang (1992), Nonstationarity and level shifts with an application to purchasing power parity, *Journal of Business & Economic Statistics*, 10, 301 - 320.

Saikkonen, P. & H. Lütkepohl (2001), Testing for unit roots in time series with level shifts, *Allgemeines Statistisches Archiv*, 85, 1 - 25.

Vogelsang, T.J. & P. Perron (1998), Additional tests for a unit root allowing for a break in the trend function at an unknown time, *International Economic Review*, 39, 1073 - 1100.

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