Comparison of Unit Root Tests for Time Series with Level Shifts

Lanne, Markku and Lütkepohl, Helmut and Saikkonen, Pentti

2002

August 4, 2000

Comparison of Unit Root Tests for Time Series with Level Shifts *

by

Markku Lanne

University of Helsinki

Helmut Lütkepohl

Humboldt Universität zu Berlin

and

Pentti Saikkonen

University of Helsinki

Address for correspondence: Markku Lanne, Department of Economics, PO Box 54, FIN-00014 University of Helsinki, FINLAND

Abstract

Unit root tests are considered for time series which have a level shift at a known point in time. The shift 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 level shifts.

Key words: Unit root, Nonlinear shift, Autoregressive process

*We are grateful to Ralf Brüggemann for helping with the computations and to Christian Müller and an anonymous referee for comments. Moreover, we thank the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213) for financial support. The third author also thanks the Alexander von Humboldt Foundation for financial support under a Humboldt research award.

1 Introduction

Modeling structural shifts in time series has become an issue of central importance due to the massive interventions that occur regularly in economic systems. In this context testing for unit roots in the presence of structural shifts has attracted considerable attention in the recent literature (see, e.g., Perron (1989, 1990), Perron & Vogelsang (1992), Banerjee, Lumsdaine & Stock (1992), Zivot & Andrews (1992), Amsler & Lee (1995), Leybourne, Newbold & Vougas (1998), Montañés & Reyes (1998)). In some of the literature the time where the structural change occurs is assumed to be known and in other articles it is assumed unknown. In this study we assume that the break point is known. In practice, such an assumption is often reasonable because the timing of many interventions is known when the analysis is performed. 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.

We will follow Saikkonen & Lütkepohl (1999) (henceforth S&L) and consider models with very general nonlinear deterministic shift functions. 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. 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. The results lead to useful recommendations for applied work.

The structure of the study is as follows. The general model is presented in Sec. 2 together with the assumptions needed for asymptotic derivations. Estimation of the nuisance parameters 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. Since some of the tests have distributions under the null hypothesis which are not tabulated, simulated critical values are presented in Sec. 5. A small sample comparison of the tests based on a Monte Carlo experiment is reported in Sec. 6 and conclusions are given in Sec. 7. The proof of a theorem is provided in the Appendix.

In the following the lag and differencing operators are denoted by L and Δ, respectively. The symbol $\xrightarrow{d}$ is used to signify convergence in distribution. The minimal eigenvalue of a matrix A is denoted by $\lambda_{min}(A)$ and $\|\cdot\|$ is the Euclidean norm.

2 The Model

We consider the following general model for a time series variable y_t with a possible unit root and a level shift from S&L:

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

where the scalars μ_0 and μ_1, the $(m \times 1)$ vector θ and the $(k \times 1)$ vector γ are unknown parameters and $f_t(\theta)$ is a $(k \times 1)$ vector of deterministic sequences depending on the parameters θ. The functional form of $f_t(\theta)$ is assumed to be known. If the sequence represents a level shift the timing of the shift is also known. For example, $f_t(\theta)$ may be thought of as a shift dummy variable which has the value zero before some given break period T_1 and the value one from then onwards. In that case, the break date T_1 is assumed to be known. Much more general situations are covered by our framework, however. Examples are considered in Sec. 6.

The quantity x_t represents an unobservable stochastic error term which is assumed to have a finite order AR representation,

$$b(L)(1 - \rho L)x_t = \varepsilon_t, \tag{2.1b}$$

where $\varepsilon_t \sim iid(0, \sigma^2)$ and $b(L) = 1 - b_1 L - \cdots - b_p L^p$ is a polynomial in the lag operator with roots bounded away from the unit circle. More precisely, the parameter space is assumed to be such that for some $\epsilon > 0$, $b(L) \neq 0$ for $|L| \leq 1 + \epsilon$. This restriction will not be taken into account in the estimation procedure, however. Obviously, if $\rho = 1$ and, hence, the DGP of x_t has a unit root, then the same is true for y_t. The initial values of x_t $(t = -p, \ldots, 0)$ are assumed to be from some fixed distribution which does not depend on the sample size. A more detailed discussion of the implications of alternative assumptions regarding the initial values may be found in Elliott, Rothenberg & Stock (1996).

The parameters μ_0, μ_1 and γ in our model are supposed to be unrestricted. Conditions required for the parameters θ and the sequence $f_t(\theta)$ are collected in the following set of assumptions which are partly taken from S&L.

Assumption 1

(a) The parameter space of θ, denoted by Θ, is a compact subset of the m-dimensional Euclidean space.

(b) For each $t = 1, 2, \ldots$, the function $f_t(\theta)$ is continuously differentiable 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, some of these conditions are just repeated from S&L. The extensions are mostly conditions for the partial derivatives of $f_t(\theta)$. They are used here to accommodate the modifications of the estimation procedures and unit root tests considered in the following sections. A compact parameter space Θ and the continuity requirement in Assumption 1(b) are standard assumptions in nonlinear estimation and testing problems. Furthermore, the summability conditions in Assumption 1(b) are needed for the function $f_t(\theta)$ and its partial derivatives $F_t(\theta)$. They hold in the applications we have in mind, if the parameter space Θ is defined in a suitable way. Therefore the condition is not critical for our purposes. The conditions in Assumption 1(b) and (c) are formulated for differences of the sequences $f_t(\theta)$ and $G_t(\theta)$ because our aim is to study unit root tests. Hence, estimation of the parameters μ, 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.

To understand Assumption 1(c), assume first that the value of the parameter θ is known and that the parameters μ and γ are estimated by applying ordinary least squares (OLS) to the differenced models. Then these assumptions guarantee linear independence of the regressors when T is large enough. There is of course no need to include the infimum in the condition of Assumption 1(c) if θ is known. It is needed, however, when the value of θ is unknown and has to be estimated. We have to impose an assumption which guarantees that the above mentioned linear independence of regressors holds whatever the value of θ because consistent estimation of θ is not possible. This is the purpose of Assumption 1(c). The condition $f_1(\theta) = \cdots = f_{p+1}(\theta) = 0$ is not restrictive for the situations and functions we have in mind and which are considered later. This condition together with the last condition in Assumption 1(c) implies that

$$\inf_{\theta\in\Theta} \lambda_{min} \left\{ \sum_{t=p+2}^{T} [b(L)\Delta G_t(\theta)][b(L)\Delta G_t(\theta)'] \right\} \geq \epsilon$$

for $T \geq T_*$ which is needed for some of the estimators used in the following to be well-defined.

Consistent estimation of θ and γ is not possible because, by Assumption 1(b), the variation of (the differenced) regressors does not increase as $T \to \infty$. The present formulation of Assumption 1(b) also applies when the sequence $f_t(\theta)$ depends on T which may be convenient occasionally. This feature is not made explicit in stating the assumption because it is not needed in the present application of Assumption 1 although it may sometimes be useful to allow the shift function to depend on T.

In the terminology of Elliott, Rothenberg & Stock (1996, Condition B), our assumptions imply that, for each value of θ, the sequence $f_t(\theta)$ defines a slowly evolving trend, although our conditions are stronger than those of Elliott et al.. 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. More discussion of Assumption 1 is given in S&L.

We compare unit root tests within the model (2.1). More precisely, we consider tests of the pair of hypotheses $H_0 : \rho = 1$ vs. $H_1 : |\rho| < 1$. The idea is to estimate the parameters related to the deterministic part first and then remove the deterministic part and perform a test on the adjusted series. In the next section we therefore discuss estimation of the nuisance parameters.

3 Estimators of Nuisance Parameters

Suppose that the process x_t specified in $(2.1b)$ is near integrated so that

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

where $c \leq 0$ is a fixed real number. The estimation procedure proposed by S&L employs an empirical counterpart of the parameter c. This means that we shall replace c by a chosen value $\bar{c}$ and pretend that $\bar{c} = c$ although we do not assume that this presumption is actually true. The idea is to apply a generalized least squares (GLS) procedure by first transforming the variables in (2.1) by the filter $1 - \bar{\rho}_T L$ where $\bar{\rho}_T = 1 + \frac{\bar{c}}{T}$ and then applying GLS to the transformed model. The choice of $\bar{c}$ will be discussed later.

For convenience we will use matrix notation and define

$$Y = [y_1 : (y_2 - \bar{\rho}_T y_1) : \cdots : (y_T - \bar{\rho}_T y_{T-1})]', \tag{3.2a}$$

$$Z_1 = \begin{bmatrix} 1 & 1 - \bar{\rho}_T & \cdots & 1 - \bar{\rho}_T \\ 1 & (2 - \bar{\rho}_T) & \cdots & (T - \bar{\rho}_T(T-1)) \end{bmatrix}' \tag{3.2b}$$

and

$$Z_2(\theta) = [f_1(\theta) : f_2(\theta) - \bar{\rho}_T f_1(\theta) : \cdots : f_T(\theta) - \bar{\rho}_T f_{T-1}(\theta)]'. \tag{3.2c}$$

Here, for simplicity, the notation ignores the dependence of the quantities on the chosen value $\bar{c}$. Using this notation, the transformed form of (2.1) can be written as

$$Y = Z(\theta)\phi + U, \tag{3.3}$$

where $Z(\theta) = [Z_1 : Z_2(\theta)]$, $\phi = [\mu_0 : \mu_1 : \gamma']'$ and $U = [u_1 : \cdots : u_T]'$ is an error term such that $u_t = x_t - \bar{\rho}_T x_{t-1} = b(L)^{-1}\varepsilon_t + T^{-1}(c - \bar{c})x_{t-1}$. Our GLS estimation is based on the covariance matrix resulting from $b(L)^{-1}\varepsilon_t$, denoted by $\sigma^2\Sigma(b)$, where $b = [b_1 : \cdots : b_p]'$. The GLS estimators are thus obtained by minimizing the generalized sum of squares function

$$Q_T(\phi, \theta, b) = (Y - Z(\theta)\phi)'\Sigma(b)^{-1}(Y - Z(\theta)\phi). \tag{3.4}$$

They are denoted as $\hat{\phi}$, $\hat{\theta}$ and $\hat{b}$. Assumption 1 ensures that these estimators are well-defined for T large enough (see S&L for details).

4 The Tests

Once the nuisance parameters in (2.1) have been estimated one can form the residual series $\hat{x}_t = y_t - \hat{\mu}_0 - \hat{\mu}_1 t - f_t(\hat{\theta})'\hat{\gamma}$ $(t = 1, \ldots, T)$ and use it to obtain unit root tests. S&L propose to consider the auxiliary regression model

$$\hat{x}_t = \rho \hat{x}_{t-1} + u_t^*, \qquad t = 2, \ldots, T. \tag{4.1}$$

In the previous section it was seen that if $\hat{x}_t$ is replaced by x_t, the covariance matrix of the error term in (4.1) is $\sigma^2\Sigma^*(b)$, where $\Sigma^*(b)$ is a $((T-1)\times(T-1))$ analog of the matrix $\Sigma(b)$. Because the parameter vector b is estimated to obtain $\hat{x}_t$ it seems reasonable to use this estimator also here and base a unit root test on (4.1) with ρ estimated by feasible GLS with weight matrix $\Sigma^*(\hat{b})^{-1}$. We denote the usual t-statistic for testing the null hypothesis $\rho = 1$ associated with the feasible GLS estimator of ρ by $\tau_{S\&L}$ because it is the statistic considered by S&L except that these authors use residuals $\hat{x}_t$ for $t = 1, \ldots, T$ in (4.1) with initial value $\hat{x}_0 = 0$.

The error term in the auxiliary regression model (4.1) also contains estimation errors caused by replacing the nuisance parameters μ_0, μ_1, θ and γ by their GLS estimators. Being able to allow for the effect of these estimation errors might improve the finite sample properties of the above test and particularly the performance of the asymptotic size approximation. To investigate this issue, consider the special case where the shift function is a step dummy variable $f_t(\theta) = d_{1t}$ which is zero up to period $T_1 - 1$ and one from period T_1 onwards. Suppose that the null hypothesis holds. Then it is straightforward to check that $u_t^* = \Delta x_t - (\hat{\mu}_1 - \mu_1) - \Delta d_{1t}(\hat{\gamma} - \gamma)$ $(t = 2, \ldots, T)$. Thus, augmenting the auxiliary regression model (4.1) by an intercept term and the impulse dummy Δd_{1t} would result in an error term which, under the null hypothesis, would not depend on the errors caused by estimating the nuisance parameters μ_1 and γ. It is fairly obvious that the inclusion of the impulse dummy Δd_{1t} has no effect on the asymptotic properties of the GLS estimator of the parameter ρ and, consequently, on the limiting distribution of the resulting test. Below we will see that the inclusion of an intercept term results in a different limiting distribution. Therefore, we will consider tests with and without intercept in the following.

If the step dummy d_{1t} is replaced by the general function $f_t(\theta)$ the above modification becomes slightly more complicated. We then have

$$\begin{aligned} u_t^* &= \Delta x_t - (\hat{\mu}_1 - \mu_1) - \Delta f_t(\hat{\theta})'\hat{\gamma} + \Delta f_t(\theta)'\gamma \\ &= \Delta x_t - (\hat{\mu}_1 - \mu_1) - \Delta f_t(\hat{\theta})'(\hat{\gamma} - \gamma) - \left(\Delta f_t(\hat{\theta}) - \Delta f_t(\theta)\right)'\gamma, \qquad t = 2, \ldots, T. \end{aligned} \tag{4.2}$$

In the last expression the third term can be handled in the same way as in the previously considered case of a step dummy but the fourth term requires additional considerations. A fairly obvious approach is to assume that the function $f_t(\theta)$ satisfies Assumption 1(b) and use the Taylor series approximation $\Delta f_t(\hat{\theta}) - \Delta f_t(\theta) \approx \Delta\left(\partial f_t(\hat{\theta})/\partial\theta'\right)(\hat{\theta} - \theta)$. Instead of (4.1) we then consider the auxiliary regression model

$$\hat{x}_t = \rho\hat{x}_{t-1} + \Delta f_t(\hat{\theta})'\pi_1 + \Delta F_t(\hat{\theta})'\pi_2 + u_t^\dagger, \qquad t = 2, \ldots, T, \tag{4.3}$$

where $F_t(\hat{\theta})$ is a $(mk \times 1)$ vector containing the partial derivatives in $\partial f_t(\hat{\theta})/\partial\theta$. Let τ_{adj} be the usual 't-statistic' based on the GLS estimation of the parameters in (4.3) with weight matrix $\Sigma^*(\hat{b})^{-1}$. Here the subscript indicates that the statistic is obtained from the *adj*usted auxiliary regression model.

In these tests we still do not make adjustments for the fact that the b parameters are also estimated. A possible modification that adjusts for the estimation of b may be obtained as follows. Define $w_t = b(L)x_t$ so that $w_t = \rho w_{t-1} + \varepsilon_t$. Thus, if we condition on $y_1, \ldots, y_p$, a version of the test statistic $\tau_{S\&L}$ may be obtained from the auxiliary regression model $\hat{w}_t = \rho\hat{w}_{t-1} + error_t$, $(t = p+1, \ldots, T)$, where $\hat{w}_t = \hat{b}(L)\hat{x}_t$. Now, to obtain a modification which takes into account estimation errors in $\hat{b}$, consider the identity

$$\begin{aligned} \hat{w}_t &= w_t + \hat{b}(L)\hat{x}_t - b(L)x_t \\ &= w_t + \hat{b}(L)(\hat{x}_t - x_t) + (\hat{b}(L) - b(L))\hat{x}_t - (\hat{b}(L) - b(L))(\hat{x}_t - x_t), \quad t = p+1, \ldots, T. \end{aligned}$$

Multiplying both sides of this equation by $\rho(L) = 1 - \rho L$ and observing that $\rho(L)w_t = \varepsilon_t$ yields

$$\hat{w}_t = \rho\hat{w}_{t-1} + \rho(L)\hat{b}(L)(\hat{x}_t - x_t) + \sum_{j=1}^{p}(\hat{b}_j - b_j)\rho(L)\hat{x}_{t-j} + r_t, \qquad t = p+2, \ldots, T,$$

where $r_t = \varepsilon_t - (\hat{b}(L) - b(L))\rho(L)(\hat{x}_t - x_t)$ is an error term. Since we try to improve the size performance of the test statistic $\tau_{S\&L}$ we now assume that the null hypothesis holds and replace $\rho(L)$ on the r.h.s. by Δ. Thus, we consider the auxiliary regression model

$$\hat{w}_t = \rho\hat{w}_{t-1} + \hat{b}(L)(\Delta\hat{x}_t - \Delta x_t) + \sum_{j=1}^{p}(\hat{b}_j - b_j)\Delta\hat{x}_{t-j} + r_t, \qquad t = p+2, \ldots, T.$$

Note that estimation errors in r_t are expected to be smaller than those in the second and third terms on the r.h.s. of this equation because, under H_0, they are affected through the product $(\hat{b}(L) - b(L))(\Delta\hat{x}_t - \Delta x_t)$ only. To be able to use this auxiliary model we still have to deal with the second term on the r.h.s.. This, however, leads to considerations very similar to those in the previous modifications and expanding the difference $\Delta\hat{x}_t - \Delta x_t$ we get the auxiliary model

$$\hat{w}_t = \rho\hat{w}_{t-1} + [\hat{b}(L)\Delta f_t(\hat{\theta})']\pi_1 + [\hat{b}(L)\Delta F_t(\hat{\theta})']\pi_2 + \sum_{j=1}^{p}\alpha_j\Delta\hat{x}_{t-j} + r_t^{\dagger}, \quad t = p+2,\ldots,T. \quad (4.4)$$

The modified test statistic is obtained as the usual t-statistic for the hypothesis $\rho = 1$ based on OLS estimation of this model. It will be denoted by τ_{adj}^{+}.

Because the actual mean of the $\hat{x}_t$ may be nonzero, it may be reasonable to include an intercept term in the previously considered auxiliary regressions. For instance, instead of (4.3) we may consider

$$\hat{x}_t = \nu + \rho\hat{x}_{t-1} + \Delta f_t(\hat{\theta})'\pi_1 + \Delta F_t(\hat{\theta})'\pi_2 + u_t^{+}, \qquad t = 2,\ldots,T. \qquad (4.5)$$

The relevant unit root t-statistic will be denoted by τ_{int}, where the subscript indicates that an *int*ercept is included in the model. Similarly, if an intercept term is added to (4.4), the resulting unit root test statistic will be denoted by τ_{int}^{+}.

Moreover, if we have the a priori restriction $\mu_1 = 0$ the estimation procedure in Section 3 and the definition of $\hat{x}_t$ 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 based on the restriction $\mu_1 = 0$ are denoted as $\tau_{S\&L}^{0}$, τ_{adj}^{0}, τ_{adj}^{+0}, τ_{int}^{0} and τ_{int}^{+0}, respectively. The limiting null distributions of all the test statistics are given in the following theorem which is partly proven in the Appendix and partly reviews results from the related literature.

Theorem 1.

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

$$\tau_{S\&L}^{0}, \tau_{adj}^{0}, \tau_{adj}^{+0} \xrightarrow{d} \left(\int_0^1 B_c(s)^2 ds\right)^{-1/2} \int_0^1 B_c(s)dB_c(s), \qquad (4.6)$$

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

$$\tau_{int}^{0}, \tau_{int}^{+0} \xrightarrow{d} \left(\int_0^1 \bar{B}_c(s)^2 ds\right)^{-1/2} \int_0^1 \bar{B}_c(s)dB_c(s), \qquad (4.7)$$

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

$$\tau_{S\&L}, \tau_{adj}, \tau_{adj}^{+} \xrightarrow{d} \left(\int_0^1 G_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 G_c(s;\bar{c})dG_c(s;\bar{c}), \qquad (4.8)$$

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,

$$\tau_{int}, \tau_{int}^{+} \xrightarrow{d} \left(\int_0^1 \bar{G}_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 \bar{G}_c(s;\bar{c})dG_c(s;\bar{c}), \tag{4.9}$$

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

Notice that for $c = 0$ the null distributions in (4.6) and (4.7) are conventional Dickey-Fuller (DF) distributions for unit root tests in models without deterministic terms and with intercept, respectively. The distribution in (4.8) was given by S&L for the statistic $\tau_{S\&L}$ in the form

$$\frac{1}{2}\left(\int_0^1 G_c(s;\bar{c})^2 ds\right)^{-1/2}(G_c(1;\bar{c})^2 - 1),$$

where

$$G_c(s;\bar{c}) = B_c(s) - s\left(\lambda B_c(1) + 3(1-\lambda)\int_0^1 sB_c(s)ds\right)$$

with $\lambda = (1-\bar{c})/h(\bar{c})$. It can be shown that this limiting distribution is equivalent to the one in (4.8) (see the Appendix). We use the latter version now because it facilitates a comparison with the other limiting distributions given in the theorem.

The limiting null distribution of the test statistics τ_{int} and τ_{int}^{+} are 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 $\tau_{S\&L}$, τ_{adj} and τ_{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 (4.5). In this sense, for example τ_{int} may be called a "mean-adjusted version" of τ_{adj}.

To the best of our knowledge the asymptotic distribution in (4.9) has not been studied previously so that critical values and suggestions for appropriate values of $\bar{c}$ are not available. Thus, simulations are required to make the test statistics τ_{int} and τ_{int}^{+} applicable and to study their power properties. Even without such simulations it is clear, however, that in terms of asymptotic local power the test statistics in (4.9) are inferior to those in (4.8) because they are not asymptotically equivalent to $\tau_{S\&L}$ and the asymptotic local power of $\tau_{S\&L}$ is indistinguishable from optimal. Analogously, $\tau_{S\&L}^{0}$, τ_{adj}^{0} and τ_{adj}^{+0} have local power which is indistinguishable from optimal and, hence, the local power of the corresponding τ_{int}^{0} and τ_{int}^{+0} tests is inferior (see Elliott et al. (1996)). However, since these results are asymptotic and based on assumptions which may be unrealistic in some cases (see Elliott et al. (1996, pp. 819-820)) the performance of the τ_{int} and τ_{int}^{0} tests may be preferable in finite samples. 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. We will provide critical values and small sample comparisons for the tests in the following sections.

5 Critical Values

In order to investigate the null distributions and local power of the test statistics we have generated time series

$$x_t = \rho_T x_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, T, \quad x_0 = 0, \quad \rho_T = 1 + c/T, \quad \varepsilon_t \sim iid\, N(0,1). \tag{5.1}$$

Thus, $p = 0$ so that 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 $\hat{x}_t$ series:

- $\hat{x}_t^{(0)} = x_t - \hat{\mu}_0$ $(t = 1, \dots, T)$, where $\hat{\mu}_0$ is obtained from a regression $(1 - \bar{\rho}_T L)x_t = \mu_0 z_{0t} + error_t$ $(t = 1, \dots, T)$ with

$$z_{0t} = \begin{cases} 1, & t = 1, \\ 1 - \bar{\rho}_T, & t = 2, \dots, T, \end{cases}$$

- $\hat{x}_t^{(1)} = x_t - \hat{\mu}_0 - \hat{\mu}_1 t$ $(t = 1, \dots, T)$, where $\hat{\mu}_0$ and $\hat{\mu}_1$ are obtained from a regression $(1 - \bar{\rho}_T L)x_t = \mu_0 z_{0t} + \mu_1(t - \bar{\rho}_T(t-1)) + error_t$ $(t = 1, \dots, T)$.

The series $\hat{x}_t^{(i)}$ $(i = 0, 1)$ are used to compute t-statistics for the null hypothesis $\rho = 1$ based on the regression model (4.1) and a corresponding version with an intercept term. For large sample size T and $c = 0$ (i.e., $\rho_T = 1$) we get realizations of the null distributions corresponding to (4.6) - (4.9) in this way.

Since we did not know which $\bar{c}$ value results in optimal local power of the tests with asymptotic distribution (4.9) we first investigated that issue. To this end we generated critical values for a 5% significance level based on 10 000 drawings with sample size $T = 500$

using $c = 0$ and then we simulated local power curves. It turned out that the local power associated with the distribution in (4.9) is almost invariant to the value of $\bar{c}$. Hence, $\bar{c} = 0$ may just as well be used. In other words, the deterministic terms may be estimated under the null rather than local alternatives in order to get optimal local power for τ_{int} and τ_{int}^{+}.

Some quantiles obtained from 10 000 drawings for different sample sizes and different values of $\bar{c}$ are given in Table 2. In the second and second last panel of the table quantiles are given for nonzero $\bar{c}$ values. They are seen to vary markedly with the sample size. In fact, they roughly decline in absolute value with growing T. For (4.6) the critical values correspond to the critical values of a DF t-test without any deterministic components in the DGP for large T (see, e.g., Fuller (1976, Table 8.5.2)). For smaller sample sizes, however, they differ substantially from the asymptotic quantiles because in generating these null distributions we use an estimator for μ_0 which is obtained under local alternatives. In this case we used a transformation based on $\bar{\rho}_T = 1 + \bar{c}/T$ with $\bar{c} = -7$ because this value was recommended by Elliott et al. (1996) for processes without deterministic trend component ($\mu_1 = 0$). Elliott et al. show that this choice results in tests with optimal local power properties. Clearly, if the asymptotic critical values (see $T = 1000$ in the table) were used when the actual sample size is $T = 50$, say, the test would reject considerably more often than indicated by the significance level chosen. For example, the critical value for a 5% level test for $T = 1000$ is -1.96 which roughly corresponds to the 10% quantile of the distribution for $T = 50$. Thus, substantial small sample distortions of the sizes of the tests must be expected given that the present results are simulated under ideal conditions which are not likely to be satisfied in practice. Hence, in practice, additional sources for distortions may be present. The critical values for $\bar{c} = 0$ are less sensitive to the sample size which may be useful in applied work. In the third panel of the table, for all sample sizes, the quantiles are seen to be close to the corresponding quantiles of the DF distributions for data generation processes (DGPs) with constant term (see again Table 8.5.2 of Fuller (1976)). Similarly, the simulated quantiles in the fifth panel ((4.8), $\bar{c} = -13.5$) are very close to those in Table I.C of Elliott et al. (1996) for all sample sizes given in that table.

6 Small Sample Comparison

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

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

and

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

with $\varepsilon_t \sim iid\, N(0, 1)$, $\rho = 1, 0.9, 0.8$, $T = 100, 200$. We also generated 100 presample values which were discarded. Furthermore, we use $T_1/T = 0.5$, that is, the break point is half way through 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 process (6.1) is in line with the model (2.1) with an abrupt shift at time T_1 so that our tests are appropriate. Because we are interested in the situation where the shift is of a more general unknown form, we also consider the DGP (6.2) which generates a smooth shift in the deterministic term. It is sometimes referred to as an innovational outlier model in the related literature. For nonzero b_1 it is not nested in our general model (2.1) although it is very similar to (6.1) in many respects. To capture the smooth transition from one regime to another in the DGP (6.2), the tests have to be combined with a smooth shift function. In the simulations we use the shift functions $f_t^{(1)}(\theta) = d_{1t}$,

$$f_t^{(2)}(\theta) = \begin{cases} 0, & t < T_1 \\ 1 - \exp\{-\theta(t - T_1 + 1)\}, & t \geq T_1 \end{cases}$$

and $f_t^{(3)}(\theta) = \left[\frac{d_{1,t}}{1-\theta L}, \frac{d_{1,t-1}}{1-\theta L}\right]'$. The last two shift functions allow for smooth deterministic shifts. In the context of DGP (6.1) they allow us to explore the sensitivity of the tests to unnecessarily flexible shifts. Note, however, that $f_t^{(2)}(\theta)$ is close to a shift dummy if θ is large and $f_t^{(3)}(\theta)$ represents a one time shift if θ is close to zero and the second component of γ is zero. Thus, both functions can in principle approximate the actual shift in (6.1) well. In addition, they may be appropriate for series generated by DGP (6.2) because they can capture the resulting smooth level shift.

All three shift functions can be shown to satisfy Assumption 1. For some of the tests the derivatives of the shift functions are needed. Because $f_t^{(1)}$ does not depend on θ, the derivative $F_t^{(1)}$ is zero. Hence, no extra terms $\Delta F_t^{(1)}(\theta)$ appear in the auxiliary regressions for τ_{adj}, τ_{adj}^+, τ_{int} and τ_{int}^+ if they are used with $f_t^{(1)}$. 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 the test statistics which account for deterministic linear trends.

Relative rejection frequencies from 1000 replications of the experiment are given in Tables 3 and 4. In Table 3, actual sizes based on the DGP (6.1) are given for tests for which estimation of the deterministic part is done under local alternatives ($\bar{c} = -7$ for $\tau_{S\&L}^0$, τ_{adj}^0, τ_{adj}^{+0} and $\bar{c} = -13.5$ for $\tau_{S\&L}$, τ_{adj}, τ_{adj}^+). Thus, in this case the DGP is in line with the original model for which the tests are derived. The nominal significance level is 5% in all cases. Obviously, all tests reject too often in some situations. Note that asymptotic critical values are used so that some overrejection was to be expected on the basis of the discussion related to Table 2. For some cases unexpectedly large rejection frequencies are observed, however. For example, it is seen in Table 3 that τ_{adj} rejects in more than 30% of the cases if the shift function $f_t^{(3)}$ is used in the test. Even if $T = 200$, the empirical size is markedly in access of 10% in this case.* Some tests do reasonably well in specific situations. For example, τ_{adj}^{+0} and τ_{adj}^+ produce rejection frequencies close to 5% when the correct shift function $f_t^{(1)}$ is used and the same is true for most of the tests when $T = 200$. Still, none of the tests performs satisfactorily for all shift functions and designs for $T = 100$. Therefore the overall message from Table 3 is clear: If the shape of the shift is unknown and, hence, a flexible shift function is considered, using nonzero values of $\bar{c}$, that is, estimating under local alternatives, bears the risk of substantially distorted sizes of the tests in samples of size 100. Thus, these tests cannot be recommended with the nonzero $\bar{c}$ values considered here. Consequently, there is no point in exploring their small sample power for these $\bar{c}$ values. Hence, in the following we focus on the tests with $\bar{c} = 0$, that is, estimation of the nuisance parameters is done under the null hypothesis.

Power results are given in Table 4 for selected tests only. The results show that for $\bar{c} = 0$

*The results are not shown to save space. More detailed results may be found in the discussion paper version of this paper which is available on request.

the test sizes are much better in line with the nominal 5% (see $\rho = 1$) at least for those tests presented in the table. In fact, for $\bar{c} = 0$ some tests tend to be conservative in specific situations and in some cases very much so (see, e.g., τ_{int} in combination with $f_t^{(1)}$). Most of the tests which are not shown in the table tend to be generally conservative and therefore do not have much small sample power. In Table 4 we only show the results for the original tests $\tau^0_{S\&L}$ and $\tau_{S\&L}$ and those tests which performed overall best in terms of small sample power within their respective groups, the groups being τ^0 tests (τ tests without linear trend term) and τ 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.

In the following, we consider only $\tau^0_{S\&L}$, τ^0_{adj}, τ^{+0}_{adj}, $\tau_{S\&L}$, τ_{int} and τ^+_{int}. In the group of τ^0 tests which exclude the deterministic trend term, τ^0_{adj} and τ^{+0}_{adj} were generally best in terms of power, each having advantages in some situations. In the group of τ tests which allow for a linear trend term, τ_{int} and τ^+_{int} dominate the other tests. Again there is no clear winner among the two tests. Whereas τ_{int} is preferable in conjunction with shift function $f_t^{(3)}$, τ^+_{int} dominates for $f_t^{(1)}$ and $f_t^{(2)}$. Both tests are clearly superior to $\tau_{S\&L}$.

It is also interesting that the results for the two DGPs are quite similar. This may not be very surprising given that the two models are in some sense quite close. A model of the type (6.1) with a deterministic linear trend and a general shift function $f_t(\theta)$ has the form $y_t = \mu_0 + \mu_1 t + f_t(\theta)'\gamma + x_t$. Multiplying this equation by $1 - b_1 L$ yields

$$(1 - b_1 L)y_t = \nu_0 + \nu_1 t + f_t(\theta)'(1 - b_1)\gamma + \Delta f_t(\theta)' b_1 \gamma + v_t, \qquad t = 2, \dots, T,$$

where ν_0 and ν_1 are functions of μ_0, μ_1 and the coefficient b_1. Moreover, v_t is as in (6.2). This shows that if we condition on y_1 in model (6.1) we obtain a model of the form (6.2) except that the additional regressor $\Delta f_t(\theta)$ is included and nonlinear parameter restrictions are involved. By Assumption 1(b) the variables $\Delta f_t(\theta)$ are "asymptotically negligible," however.

The following further conclusions emerge from Table 4. Excluding a linear trend term from the models when such a restriction is correct results in substantially better power. Furthermore, changing b_1 from 0.5 to 0.8 has a substantial effect. It implies a sizable decline in power in most cases. 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. The results in Table 4 also show that there are cases where the tests are 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 more complicated shift functions $f_t^{(2)}$ or $f_t^{(3)}$ is used. We note, however, that the performance of all the tests improved markedly when T was increased from 100 to 200. Even in that case the modifications overall dominate the original test versions $\tau^0_{S\&L}$ and $\tau_{S\&L}$.

7 Conclusions and Extensions

Standard unit root tests are known to have reduced power if they are applied to time series with structural shifts. Therefore we have considered unit root tests that explicitly allow for a level shift 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 focused on models where the shift is regarded as part of the deterministic component of the DGP. Building on a proposal by S&L, 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. 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. Small sample properties of the tests are obtained by simulation.

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 S&L in that they have superior size and power properties. Substantial size distortions may result in small samples if the nuisance parameters are estimated under local alternatives. Therefore we recommend estimating the nuisance parameters under the null hypothesis.

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 $T = 100$. Therefore we recommend using the modified test versions τ^0_{adj} and τ^{+0}_{adj} if no deterministic linear trend is present because they have overall best size and power properties. As none of these tests dominates the other one in all situations it may be useful to apply both tests jointly and reject the unit root hypothesis if one of the tests rejects the null hypothesis. If a linear trend term is needed, the modified test versions τ_{int} and τ^{+}_{int} are recommended based on analogous arguments.

We have also explored the possibility of using DGPs of the type (6.2) with potentially more short-term dynamics. As mentioned earlier, they account for shifts which are due to innovational outliers. Models of this type are preferred in parts of the related literature. In the context of these models unit root tests similar to those of S&L were in fact considered by Lütkepohl, Müller & Saikkonen (2000). Extensions similar to those of the present study are possible and are presented in the discussion paper version which is available upon request. In that study we have also performed a detailed investigation of other modifications which did not perform as well as the tests presented here. Therefore they were deleted from the present version of the paper.

Although we have focused 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. Moreover, a comparison with other unit root tests which allow for structural shifts may be worthwhile. We leave these issues for future investigations.

Appendix. Proof of Theorem 1

In the proof of Theorem 1 we focus on the limiting distributions of test statistics for models where μ_1 is not known to be zero a priori. The case where the restriction $\mu_1 = 0$ is imposed follows by making straightforward modifications to these proofs. We begin with the result in (4.8).

The limiting distribution of $\tau_{S\&L}$ is derived in S&L. In that article it is given in a slightly different form, however. To see that the present form is equivalent it may be worth noting that $(A.21)$ of S&L may be written alternatively as

$$\begin{aligned} & T^{-1}\hat{X}'_{-1}\Sigma(\hat{b})^{-1}(\hat{X}-\hat{X}_{-1}) \\ & = T^{-1}\textstyle\sum_{t=p}^{T}[\hat{b}(L)\hat{x}_{t-1}][\hat{b}(L)\Delta\hat{x}_t]+o_p(1) \\ & = T^{-1}\textstyle\sum_{t=p}^{T}[b(1)\{x_{t-1}-(\hat{\mu}_1-\mu_1)(t-1)\}][b(L)\Delta x_t-b(1)(\hat{\mu}_1-\mu_1)]+o_p(1) \\ & \xrightarrow{d} \sigma^2\textstyle\int_0^1 G_c(s;\bar{c})dB_c(s)-\sigma^2K_c(\bar{c})\int_0^1 G_c(s;\bar{c})ds, \end{aligned} \qquad (A.1)$$

where the last relation follows from well-known limit theorems by noting that the limiting distribution of $\hat{\mu}$ given in (3.12) of S&L can be written alternatively as $\omega K_c(\bar{c})$, where $\omega=\sigma/b(1)$,

$$K_c(\bar{c})=h(\bar{c})^{-1}\int_0^1(1-\bar{c}s)dB_0(s)+h(\bar{c})^{-1}(c-\bar{c})\int_0^1(1-\bar{c}s)B_c(s)ds \qquad (A.2)$$

and $h(\bar{c})=1-\bar{c}+\bar{c}^2/3$. From the representation in $(A.1)$ the limiting distribution in (4.8) follows as in the proof of the asymptotic distribution of the test statistic in S&L. Thus, to prove (4.8), it remains to show that τ_{adj} and τ^+_{adj} have the same limiting distribution as $\tau_{S\&L}$.

Using

$$T^{-1/2}\hat{x}_{[Ts]}\xrightarrow{d}\omega G_c(s;\bar{c}) \qquad (A.3)$$

(see $(A.18)$ of S&L) and the fact that $f_t(\theta)$ satisfies Assumption 1(b) it can be seen that

$$\left\|T^{-1}\sum_{t=1}^{T}\hat{x}_{t-1}\Delta f_t(\hat{\theta})\right\|\leq T^{-1}\max_{1\leq t\leq T}|\hat{x}_t|\sum_{t=1}^{T}\sup_{\theta\in\Theta}\|\Delta f_t(\theta)\|=O_p\left(T^{-1/2}\right)$$

and that a similar result also holds with $\Delta f_t(\hat{\theta})$ replaced by $\Delta F_t(\hat{\theta})$. Using these facts and arguments similar to those in the proof of Lemma 1 of S&L it can be shown that the appropriately standardized moment matrix in the GLS estimation of (4.3) is asymptotically block diagonal and also positive definite. Since it is further straightforward to show that $\sum_{t=1}^{T}\Delta f_t(\hat{\theta})u_t^{\dagger}=O_p(1)$ and similarly with $\Delta f_t(\hat{\theta})$ replaced by $\Delta F_t(\hat{\theta})$ it follows that the limiting distribution of the GLS estimator of ρ in (4.3) and hence that of its t-ratio is the same as in the case of the auxiliary regression model (4.1). We have thus shown that (4.8) holds for the test statistic τ_{adj}.

As for test statistic τ^+_{adj}, note first that the arguments used for τ_{adj} above and those in the proof of Theorem 1 of S&L show that the appropriately standardized moment matrix in the auxiliary regression model used to obtain the test statistic τ^+_{adj} is asymptotically positive definite and also block diagonal between $\hat{w}_{t-1}$ and the other regressors. Deriving the expression of the error term in this auxiliary regression model it is further straightforward to show that τ_{adj}^{+} has the same limiting distribution as $\tau_{S\&L}$ and τ_{adj}. Thus, (4.8) is proven.

Since the test statistics τ_{int} and τ_{int}^{+} are obtained by augmenting the auxiliary regression models used to obtain test statistics τ_{adj} and τ_{adj}^{+}, respectively, by an intercept term, (4.9) can be proven by extending the arguments used above in a standard manner.

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.

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 (2000), 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, forthcoming.

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 (1999), Testing for unit roots in time series with level shifts, Discussion Paper, SFB 373, Humboldt-Universität zu Berlin.

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.

Table 1. Summary of Tests

Test statistic	Underlying auxiliary regression
	Asymptotic distribution $\left(\int_0^1 G_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 G_c(s;\bar{c}) dG_c(s;\bar{c})$
$\tau_{S\&L}$	$\hat{x}_t = \rho\hat{x}_{t-1} + u_t^*$
τ_{adj}	$\hat{x}_t = \rho\hat{x}_{t-1} + \Delta f_t(\hat{\theta})'\pi_1 + \Delta F_t(\hat{\theta})'\pi_2 + u_t^\dagger$
τ_{adj}^+	$\hat{w}_t = \rho\hat{w}_{t-1} + [\hat{b}(L)\Delta f_t(\hat{\theta})']\pi_1 + [\hat{b}(L)\Delta F_t(\hat{\theta})']\pi_2 + \sum_{j=1}^p \alpha_j \Delta\hat{x}_{t-j} + r_t^\dagger$
	Asymptotic distribution $\left(\int_0^1 \bar{G}_c(s;\bar{c})^2 ds\right)^{-1/2} \int_0^1 \bar{G}_c(s;\bar{c}) dG_c(s;\bar{c})$
τ_{int}	$\hat{x}_t = \nu + \rho\hat{x}_{t-1} + \Delta f_t(\hat{\theta})'\pi_1 + \Delta F_t(\hat{\theta})'\pi_2 + u_t^+$
τ_{int}^+	$\hat{w}_t = \nu + \rho\hat{w}_{t-1} + [\hat{b}(L)\Delta f_t(\hat{\theta})']\pi_1 + [\hat{b}(L)\Delta F_t(\hat{\theta})']\pi_2 + \sum_{j=1}^p \alpha_j \Delta\hat{x}_{t-j} + r_t^+$

Table 2. Simulated Quantiles of Null Distributions of Test Statistics Based on 10 000 Replications

Distribution	T	$\alpha_{0.01}$	$\alpha_{0.025}$	$\alpha_{0.05}$	$\alpha_{0.1}$
(4.6) ($\bar{c} = 0$)	50	-2.65	-2.26	-1.97	-1.63
	100	-2.61	-2.25	-1.96	-1.62
	200	-2.64	-2.26	-1.94	-1.62
	500	-2.60	-2.25	-1.95	-1.62
	1000	-2.55	-2.24	-1.96	-1.61
(4.6) ($\bar{c} = -7$)	50	-2.93	-2.56	-2.28	-1.98
	100	-2.73	-2.41	-2.15	-1.83
	200	-2.68	-2.34	-2.05	-1.73
	500	-2.64	-2.30	-2.00	-1.67
	1000	-2.56	-2.22	-1.96	-1.63
(4.7) ($\bar{c} = 0$)	50	-3.64	-3.28	-2.99	-2.67
	100	-3.58	-3.22	-2.94	-2.62
	200	-3.58	-3.22	-2.93	-2.62
	500	-3.47	-3.17	-2.90	-2.62
	1000	-3.48	-3.15	-2.88	-2.58
(4.8) ($\bar{c} = 0$)	50	-3.34	-2.96	-2.65	-2.37
	100	-3.23	-2.90	-2.61	-2.33
	200	-3.17	-2.91	-2.64	-2.33
	500	-3.22	-2.92	-2.64	-2.35
	1000	-3.18	-2.86	-2.62	-2.33
(4.8) ($\bar{c} = -13.5$)	50	-3.83	-3.48	-3.21	-2.91
	100	-3.62	-3.30	-3.03	-2.74
	200	-3.51	-3.24	-2.96	-2.66
	500	-3.43	-3.09	-2.84	-2.57
	1000	-3.40	-3.11	-2.85	-2.57
(4.9) ($\bar{c} = 0$)	50	-3.81	-3.45	-3.15	-2.86
	100	-3.73	-3.38	-3.11	-2.80
	200	-3.64	-3.32	-3.06	-2.77
	500	-3.62	-3.32	-3.08	-2.79
	1000	-3.55	-3.28	-3.03	-2.76

Table 3. Empirical Sizes of Tests Based on DGP (6.1), $T = 100$, $T_1 = 50$, $\bar{c} = -7/-13.5$, Nominal Significance Level 5%

Shift function	b_1	$\tau^0_{S\&L}$	τ^0_{adj}	τ^{+0}_{adj}	$\tau_{S\&L}$	τ_{adj}	τ^+_{adj}
$f_t^{(1)}$	0.5	0.077	0.076	0.069	0.085	0.087	0.071
	0.8	0.164	0.165	0.064	0.072	0.073	0.063
$f_t^{(2)}$	0.5	0.186	0.223	0.276	0.163	0.252	0.276
	0.8	0.227	0.301	0.405	0.089	0.155	0.197
$f_t^{(3)}$	0.5	0.193	0.269	0.224	0.158	0.360	0.262
	0.8	0.206	0.533	0.227	0.080	0.501	0.160

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

Shift function	Test	DGP (6.1), $b_1 = 0.5$ $\rho = 1$	0.9	0.8	DGP (6.1), $b_1 = 0.8$ $\rho = 1$	0.9	0.8	DGP (6.2), $b_1 = 0.5$ $\rho = 1$	0.9	0.8	DGP (6.2), $b_1 = 0.8$ $\rho = 1$	0.9	0.8
$f_t^{(1)}$	$\tau^0_{S\&L}$	0.039	0.289	0.533	0.016	0.156	0.314	0.039	0.284	0.524	0.020	0.138	0.275
	τ^0_{adj}	0.039	0.291	0.535	0.016	0.156	0.315	0.040	0.285	0.527	0.020	0.140	0.275
	τ^{+0}_{adj}	0.063	0.353	0.590	0.050	0.292	0.436	0.061	0.343	0.575	0.053	0.287	0.382
	$\tau_{S\&L}$	0.010	0.054	0.190	0.000	0.006	0.030	0.009	0.050	0.174	0.000	0.005	0.025
	τ_{int}	0.020	0.090	0.302	0.000	0.006	0.034	0.022	0.091	0.305	0.001	0.004	0.029
	τ^+_{int}	0.080	0.233	0.526	0.065	0.167	0.286	0.075	0.216	0.499	0.064	0.149	0.262
$f_t^{(2)}$	$\tau^0_{S\&L}$	0.043	0.235	0.423	0.023	0.123	0.243	0.041	0.231	0.415	0.021	0.129	0.248
	τ^0_{adj}	0.064	0.270	0.454	0.045	0.155	0.288	0.065	0.257	0.433	0.037	0.141	0.276
	τ^{+0}_{adj}	0.048	0.254	0.445	0.026	0.142	0.272	0.049	0.246	0.426	0.025	0.140	0.271
	$\tau_{S\&L}$	0.014	0.056	0.179	0.000	0.004	0.030	0.010	0.051	0.177	0.000	0.006	0.028
	τ_{int}	0.048	0.146	0.349	0.021	0.033	0.089	0.045	0.131	0.336	0.019	0.029	0.063
	τ^+_{int}	0.052	0.167	0.367	0.029	0.045	0.115	0.053	0.151	0.348	0.030	0.039	0.080
$f_t^{(3)}$	$\tau^0_{S\&L}$	0.047	0.215	0.378	0.020	0.120	0.220	0.044	0.219	0.384	0.022	0.113	0.217
	τ^0_{adj}	0.064	0.266	0.417	0.079	0.223	0.302	0.060	0.268	0.426	0.082	0.217	0.293
	τ^{+0}_{adj}	0.059	0.249	0.404	0.037	0.144	0.249	0.056	0.252	0.418	0.036	0.140	0.243
	$\tau_{S\&L}$	0.011	0.044	0.173	0.000	0.005	0.024	0.014	0.046	0.165	0.000	0.005	0.022
	τ_{int}	0.060	0.141	0.322	0.074	0.086	0.133	0.062	0.146	0.325	0.072	0.091	0.134
	τ^+_{int}	0.048	0.120	0.314	0.016	0.028	0.064	0.052	0.129	0.317	0.014	0.029	0.068