Testing for a Unit Root in a Time Series with a Level Shift at Unknown Time

Pentti Saikkonen
Department of Statistics
University of Helsinki
P.O. Box 54
SF-00014 University of Helsinki
FINLAND

Tel: +358-9-1918867
Fax: +358-9-1918872
Email: saikkone@valt.helsinki.fi

and

Helmut Lütkepohl
Institut für Statistik und Ökonometrie
Wirtschaftswissenschaftliche Fakultät
Humboldt University
Spandauer Str. 1
D-10178 Berlin
GERMANY
Tel: +49-30-2093-5718
Fax: +49-30-2093-5712
Email: luetke@wiwi.hu-berlin.de

Abstract

Unit root tests for time series with level shifts of general form are considered when the timing of the shift is unknown. It is proposed to estimate the nuisance parameters of the data generation process including the shift date in a first step and apply standard unit root tests to the residuals. The estimation of the nuisance parameters is done in such a way that the unit root tests on the residuals have limiting distributions for which critical values are tabulated elsewhere in the literature. Empirical examples are discussed to illustrate the procedure.

JEL classification: C22, C12
Keywords: Univariate time series, unit root, structural shift, autoregression

*We are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213) for financial support. The first author also thanks the Alexander von Humboldt Foundation for financial support under a Humboldt research award. Part of this research was done while he was visiting the Humboldt University in Berlin. Christian Müller and Markku Lanne helped with the computations.

1 Introduction

There has been some debate in the recent literature whether macroeconomic time series can be modeled adequately by a nonstationary process with a unit root or whether they are better thought of as being generated by a trend-stationary process with stationary fluctuations around a broken trend. The issue is important because, in the unit root case, stochastic shocks to the series have permanent effects whereas in the trend-stationary model only changes in the trend function have a permanent effect while stochastic shocks are transitory. Usually tests are carried out in order to choose between a unit root process and a trend-stationary alternative. Given the importance of the issue for assessing the implications of economic activities it is not surprising that a number of articles consider unit root tests in the presence of possible structural breaks.

In this literature broadly two alternative assumptions regarding the possible dates of the structural breaks have been made. In one part of the literature the break date is assumed to be known by the analyst, that is, the break is assumed to be due to some exogenous shock which has occurred at a known date. Examples of articles where this assumption was made are Perron (1989, 1990), Saikkonen & Lütkepohl (1999) (henceforth SL) and Lütkepohl, Müller & Saikkonen (1999) (henceforth LMS). Another part of the literature assumes that the break date is unknown to the investigator and it may be a random event which may be modeled endogenously. In some of this literature the timing of a structural break is regarded as an additional unknown parameter. For example, Evans (1989), Christiano (1992), Perron & Vogelsang (1992), Zivot & Andrews (1992), Banerjee, Lumsdaine & Stock (1992) as well as Leybourne, Newbold & Vougas (1998) consider shifts at an unknown date.

Different estimators for the break date have been proposed for the case when it is unknown. Some authors take into account that the final objective of the analysis is testing for a unit root and therefore focus on the consequences of using an estimated break date in this situation (e.g., Perron & Vogelsang (1992), Zivot & Andrews (1992), Banerjee, Lumsdaine & Stock (1992)). For instance, the former two articles propose to estimate the break date such that the unit root test becomes least favourable to the null hypothesis of a unit root and consider the asymptotic distribution theory of the resulting test statistic. Leybourne, Newbold & Vougas (1998) estimate the deterministic part of the assumed DGP (data generation process) first, including possible structural shifts. Then they apply unit root tests to the residuals. The approach of Leybourne, Newbold & Vougas (1998) results in tests for which new critical values have to be generated for each individual time series.

In the present study we will use an approach similar to that of Amsler & Lee (1995). More precisely, we propose estimating all nuisance parameters of the process in a first step in such a way that the limiting distribution of the subsequent unit root tests do not depend on the estimator of the break date. Our approach differs from that of Amsler & Lee in some important respects, however. Whereas these authors fix the break date the timing of the shift is estimated in our approach. Moreover, Amsler & Lee model the shift by a simple dummy variable, whereas much more general structural shifts are considered in our framework. In fact, the shift function can be a smooth function from one state of the process to another or it can be of some other nonlinear form. It is argued by Leybourne, Newbold & Vougas (1998) that allowing for general shift functions is important because it is not likely that all agents react simultaneously and instantaneously to changes in the environment. Therefore a smooth transition to a new level may often be more realistic than an instantaneous shift. Finally, we consider another estimator of the nuisance parameters than Amsler & Lee.

The structure of the paper is as follows. In the next section two general models for univariate time series with a shift in the mean and a possible unit root are presented. The models are those treated by SL and LMS for the case of a known shift date. Section 3 considers estimation of the nuisance parameters of the DGP and the tests for unit roots are presented in Section 4. Empirical examples are discussed in Section 5 and conclusions are given in Section 6. Proofs are deferred to the appendix.

The following general notation is used. The lag and differencing operators are denoted by L and Δ, respectively, so that for a time series variable y_t, $Ly_t = y_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$. The symbols $\rightarrow_p$ and $\rightarrow_d$ signify convergence in probability and in distribution, respectively. Independently, identically distributed will be abbreviated as $iid(\cdot,\cdot)$, where the first and second moments are indicated in parentheses in the usual way. 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. We use $\lambda_{\min}(A)$ ($\lambda_{\max}(A)$) to denote the minimal (maximal) eigenvalue of a matrix A. Moreover, $\|\cdot\|$ denotes the Euclidean norm. GLS is used to abbreviate generalized least squares and sup and inf are short for supremum and infimum, respectively. The m-dimensional Euclidean space is denoted by $\mathbb{R}^m$.

2 The Model Framework

SL and LMS consider two alternative general models for the DGP of a time series with a possible unit root and a structural shift. The one investigated by SL has the form

$$y_t = \mu t + g_{t\tau}(\theta)'\gamma + x_t, \qquad t = 1, 2, \dots, \tag{2.1}$$

where the scalar μ, the $(m \times 1)$ vector θ and the $(k \times 1)$ vector γ are unknown parameters and $g_{t\tau}(\theta)$ is a $(k \times 1)$ vector of deterministic sequences depending on the parameters θ and on the break point which is denoted by τ, that is, a shift occurs in or just before period τ. The quantity x_t represents an unobservable stochastic error term which is assumed to have a finite order autoregressive (AR) representation of order p,

$$b(L)(1 - \rho L)x_t = \varepsilon_t, \tag{2.2}$$

where $b(L) = 1 - b_1 L - \cdots - b_{p-1}L^{p-1}$ has all its zeros outside the unit circle if $p > 1$, while $-1 < \rho \leq 1$. A unit root is present, of course, if $\rho = 1$. Assumptions for the initial values will be discussed later. The essential requirement is that they must be independent of the sample size T. The error terms ε_t are assumed to be iid$(0, \sigma^2)$ with

$$E|\varepsilon_t|^\alpha < \infty \quad \text{for some} \quad \alpha > 2. \tag{2.3}$$

With respect to the function $g_{t\tau}(\theta)$ it is assumed that the first component is unity so that the first component of γ defines the level parameter of y_t. Specifically we have,

$$g_{t\tau}(\theta) = [1 : f_{t\tau}(\theta)']' \tag{2.4}$$

where $f_{t\tau}(\theta)$ is a $(k-1)$-dimensional deterministic sequence to be described in more detail below.

The model considered by LMS has the form

$$b(L)y_t = \mu t + g_{t\tau}(\theta)'\gamma + v_t, \qquad t = 1, 2, \dots, \tag{2.5}$$

where

$$v_t = \rho v_{t-1} + \varepsilon_t \tag{2.6}$$

is an AR process of order 1 and the other notation is as before. Again, if $\rho = 1$, v_t and, hence, y_t has a unit root.

A leading example of a sequence $f_{t\tau}$ is a shift dummy variable

$$f_{t\tau}(\theta) = d_{t\tau} := \begin{cases} 0, & t < \tau \\ 1, & t \geq \tau \end{cases} . \tag{2.7}$$

In this special case the sequence $f_{t\tau}$ does not depend on any unknown parameters θ. Further examples will be discussed later. In SL and LMS it is assumed that the shift point τ is known a priori. In the following we will give up this assumption and consider the case of an unknown τ. In other words, the break point τ will be regarded as an unknown integer valued parameter.

To gain generality, we allow $f_{t\tau}(\theta)$ to be of a much more general form than $d_{t\tau}$. In fact, we only assume that $f_{t\tau}(\theta)$ satisfies the conditions stated in the following assumption.

Assumption A.

(a) The parameter space of θ, denoted by Θ, is a compact subset of $\mathbb{R}^m$ and N_T, the space of τ, is a subset of $\{2, \ldots, T-1\}$.

(b) For each $t = 1, 2, \ldots$ and each $\tau \in N_T$, $f_{t\tau}(\theta)$ is a continuous function of θ and

$$\sup_T \sup_{\theta \in \Theta, \tau \in N_T} \sum_{t=1}^{T} \|\Delta f_{t\tau}(\theta)\| < \infty$$

where $f_{0\tau}(\theta) = 0$.

(c) There exists a real number $\varepsilon > 0$ and an integer T_* such that, for all $T \geq T_*$,

$$\inf_{\theta \in \Theta, \tau \in N_T} \lambda_{\min} \left\{ \sum_{t=1}^{T} \Delta g_{t\tau}(\theta) \Delta g_{t\tau}(\theta)' \right\} \geq \varepsilon,$$

where $\Delta g_{1\tau}(\theta) = [1 : f_{1\tau}(\theta)']'$. ■

Assumption A is very similar to Assumption 1 of SL and LMS and since most of the discussion given for the latter also applies here, we will focus on some differences in the following. In the same way as previously, the parameter space Θ is assumed to be compact. Instead of assuming that the space of τ is the whole set $\{2, \ldots, T-1\}$, as supposed in the special case above, we use the slightly more general assumption that N_T may be a subset of $\{2,\ldots,T-1\}$. In this way it is possible to take prior information on the date of the possible level shift into account. For instance, it may be known that the level shift has occurred during the second half of the sample period. Fixing the value of τ in Assumption A(b) gives Assumption 1(b) of SL except that now the supremum over θ is in front of the sum. Thus, even if the value of τ is known a priori, Assumption A(b) is weaker than its counterpart in SL. When the value of τ is unknown it is required that the summability of $\|f_{t\tau}(\theta)\|$ also holds uniformly over τ.

For a fixed value of τ, Assumption A(c) is also similar to Assumption 1(c) of SL and when the value of τ is not fixed the previous condition is required to hold uniformly over τ. Thus, one can say that Assumption A is obtained from Assumption 1 of SL by somewhat weakening Assumption 1(b) and requiring that the previous conditions hold uniformly over both θ and τ.

It is easy to see that Assumption A is satisfied for the shift dummy in (2.7). Since there is no parameter θ in this case, part (a) is trivially satisfied here. Moreover, Assumption A(b) holds because

$$\sum_{t=1}^{T}\|\Delta f_{t\tau}(\theta)\| = \sum_{t=1}^{T}|\Delta d_{t\tau}| = 1$$

for all T and $0<\tau<T$. Furthermore, defining $g_{t\tau}(\theta) = [1 : d_{t\tau}]'$, the smallest eigenvalue in question in Assumption A(c) is unity for all θ, τ and $T>2$ if $1<\tau<T$.

It can also be shown that the assumption holds for sequences

$$f_{t\tau}(\theta) = \begin{cases} 0, & t<\tau \\ 1-\exp\{-\theta(t-\tau)\}, & t\geq\tau \end{cases}$$

or

$$f_{t\tau}(\theta) = \begin{cases} 0, & t<\tau \\ \exp\{-\theta(t-\tau)\}, & t\geq\tau \end{cases}$$

where θ is an unknown parameter with $0<\theta<\text{constant}<\infty$. These sequences were also considered by SL. Another example sequence used in that article is

$$f_{t\tau}(\theta) = \left[\frac{d_{t,\tau}}{\varphi(L)} : \cdots : \frac{d_{t-q,\tau}}{\varphi(L)}\right]',$$

where the components of θ are given by the unknown coefficients of $\varphi(L) = 1-\varphi_1 L-\cdots-\varphi_r L^r$, which is a lag polynomial with all its zeros outside the complex unit circle. This shift function is motivated in SL, where it is also argued that it satisfies Assumption 1 of that article. The arguments given there can be adopted to show that Assumption A is satisfied. We do not give the details here but refer the reader to SL. Another possible choice of a shift sequence may be of the form

$$f_{t\tau}(\theta) = \exp\{-\theta(t-\tau)^2\}, \qquad \theta > 0,$$

which allows for a smooth but temporary level shift (cf. Lin & Teräsvirta (1994) where also alternative specifications are discussed). Leybourne, Newbold & Vougas (1998) consider the logistic smooth transition function

$$f_{t\tau}(\theta) = [1 + \exp\{-\theta(t-\tau)\}]^{-1}, \qquad \theta > 0.$$

As mentioned earlier, they argue that smooth transitions to a new level of a series are often more plausible because agents are not likely to react all simultaneously due to market inefficiencies, for example. Hence, it is important to allow for the more general nonlinear shifts in the present context.

In the example section, we consider, for instance, the series of Polish industrial production which is likely to have a downward shift at the time when Poland switched from a socialist economy to a market economy. Although the official transition date is known, it is unlikely that the related adjustment processes started exactly at that time because the economic agents knew the date well in advance. Hence, assuming a known break date may be problematic in this case. Moreover, allowing for a smooth transition may be more reasonable than assuming an abrupt shift. Of course, in many cases it may be problematic to assume a specific form of the shift if the time of the shift is unknown. In that situation one may want to consider some general shift function. Alternatively, a very simple shift in the level as modeled by (2.7) may be analyzed. In any case, it is of interest to treat the general models because our theoretical results hold in the general situation. Even more generality is possible by allowing for more than one level shift. It is not difficult to adjust our assumptions to that case. For instance, if there are two level shifts, the integer valued parameter τ is replaced by the vector $\tau = [\tau_1 : \tau_2]'$ and the permissible values of τ_1 and τ_2 are, for instance, $\{2, \ldots, [T/2]\}$ and $\{[T/2]+1, \ldots, T-1\}$, respectively. To avoid more complicated notations we will not treat this case in detail in the following but focus on the situation where there is just one shift.

In asymptotic considerations it may often be natural to assume that the 'true' value of τ may depend on the sample size because in this way it is, for example, possible to allow for the fact that the shift occurs around the middle or in the last quarter etc. of the sample. In that case one may wish to replace the integer valued parameter τ by $T\tau_*$ with τ_* a real valued parameter taking values in the interval $[0,1]$ or some subset of it. This formulation has been used in some previous studies (e.g., Zivot & Andrews (1992), Banerjee et al. (1992)). For our purposes the above formulation with integer valued shift date parameter τ is more convenient, however, because it has also been used by SL and LMS. It is therefore used in the following. From a practical point of view the differences in the two alternative assumptions are hardly of importance. We will not make the possible dependence of the parameter τ on the sample size explicit in the notation because it has no effect on the derivations.

For completeness we mention that seasonal dummy variables may be included in both models (2.1) and (2.5). Again this merely complicates the notation without affecting the asymptotic analysis in any substantial way. Therefore we do not include seasonal dummies here. In the next section we will consider estimating the nuisance parameters of the general models (2.1) and (2.5). The unit root tests are presented in Sec. 4.

3 Estimation of Nuisance Parameters

As in SL and LMS we will consider estimating the nuisance parameters of the general models discussed in the previous section first. The difference to the aforementioned articles is that we have the additional integer valued parameter τ which describes the timing of the level shift. In some of the related literature (e.g., Zivot & Andrews (1992) and Perron & Vogelsang (1992)) the date of the level shift has been estimated by computing the value of the unit root test statistic for all permissible values of τ. The estimated break date is then chosen to be the date which results in the value of the test statistic which is least favourable for the null hypothesis of a unit root. Some other estimators have also been discussed (e.g., Perron & Vogelsang (1992)). These estimation methods require estimation of the other nuisance parameters, that is, μ, θ, γ and $b = (b_1, \dots, b_{p-1})'$ in the present context, for all values of $\tau \in N_T$. We will also use this general approach in the following. However, our specific approach will result in estimators and test statistics with quite different properties than those of other approaches.

If the value of τ is fixed the GLS estimation methods considered in SL and LMS can be readily modified for the present context. We will begin with model (2.1)/(2.2). As in SL it is assumed that the true error process x_t is near-integrated so that the parameter in (2.2) satisfies

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

where $c \leq 0$ is a fixed real number. The idea is to replace ρ_T by $\bar{\rho}_T = 1 + \frac{\bar{c}}{T}$ with $\bar{c}$ a chosen number and transform (2.1) by the filter $1 - \bar{\rho}_T L$. This yields the model

$$Y = Z_\tau(\theta)\phi + U, \tag{3.2}$$

where $Y = [y_1 : (y_2 - \bar{\rho}_T y_1) : \cdots : (y_T - \bar{\rho}_T y_{T-1})]'$, $\phi = [\mu : \gamma']'$, $Z_\tau(\theta) = [Z_1 : Z_{2\tau}(\theta)]$ with $Z_1 = [1 : (2 - \bar{\rho}_T) : \cdots : (T - \bar{\rho}_T(T-1))]'$ and $Z_{2\tau}(\theta) = [g_{1\tau}(\theta) : (g_{2\tau}(\theta) - \bar{\rho}_T g_{1\tau}(\theta)) : \cdots : (g_{T\tau}(\theta) - \bar{\rho}_T g_{T-1,\tau}(\theta))]'$. Finally, $U = [u_1 : \cdots : u_T]'$ is an error term such that $u_t = x_t - \bar{\rho}_T x_{T-1}$ and, hence,

$$u_t = b(L)^{-1}\varepsilon_t + T^{-1}(c - \bar{c})x_{t-1} \overset{def}{=} u_t^{(0)} + T^{-1}(c - \bar{c})x_{t-1} \tag{3.3}$$

as in SL.

For any given value of τ the parameters θ and ϕ as well as the parameters b in the error covariance matrix of (3.2) can be estimated by minimizing the generalized sum of squares function

$$Q_{T\tau}(\phi, \theta, b) = (Y - Z_\tau(\theta)\phi)'\Sigma(b)^{-1}(Y - Z_\tau(\theta)\phi), \tag{3.4}$$

where $\Sigma(b) = \sigma^{-2}\text{Cov}(U^{(0)})$ with $U^{(0)} = [u_1^{(0)} : \cdots : u_T^{(0)}]'$. Assume that the matrix $Z_\tau(\theta)$ is of full column rank for all $\theta \in \Theta$ and all $\tau \in N_T$. As shown in the appendix, this condition holds by Assumption A(c) for all T large enough. Then, repeating the argument used in Section 3 of SL, it can be shown that a minimizer of $Q_{T\tau}(\phi, \theta, b)$ exists for any given value of τ. Let $\hat{\phi}_\tau$, $\hat{\theta}_\tau$ and $\hat{b}_\tau$ be such that

$$\hat{Q}_{T\tau} \overset{def}{=} Q_{T\tau}(\hat{\phi}_\tau, \hat{\theta}_\tau, \hat{b}_\tau) \leq Q_{T\tau}(\phi, \theta, b) \tag{3.5}$$

for all ϕ, θ and b such that $\theta \in \Theta$ and b satisfies the compactness assumption implied by Assumption 2 of SL. That assumption states that the roots of $b(L)$ are bounded away from the unit circle. Note that in the same way as in (3.8) of SL we have

$$\hat{\phi}_\tau = [Z_\tau(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1}Z_\tau(\hat{\theta}_\tau)]^{-1}Z_\tau(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1}Y. \tag{3.6}$$

The following lemma describes asymptotic properties of the estimators $\hat{\phi}_\tau$, $\hat{\theta}_\tau$ and $\hat{b}_\tau$. Before presenting this lemma we note, however, that in the present context a natural way to estimate the parameter τ is to minimize the function $\hat{Q}_{T\tau}$ defined in (3.5). The estimators $\hat{\phi}_\tau$, $\hat{\theta}_\tau$ and $\hat{b}_\tau$ corresponding to this value of τ would then give the estimators used for ϕ, θ and b, respectively. We will not consider this approach here because our test procedure can be used in the same way with any estimator of τ, as will be seen later.

Now we can state the following lemma, where $\hat{\phi}_\tau = [\hat{\mu}_\tau : \hat{\gamma}_\tau']'$ conformably with the partition of ϕ. In the same way as in SL, the lemma assumes local alternatives defined by (3.1). Its proof and other proofs are given in the appendix.

Lemma 3.1.
Suppose that Assumption A stated in Section 2 holds and assume that, for some $\epsilon > 0$, $b(L) \neq 0$ for $|L| \leq 1 + \epsilon$, that is, the roots of $b(L)$ are bounded away from the unit circle. Moreover, suppose that the matrix $Z_\tau(\theta)$ is of full column rank for all $\theta \in \Theta$, all $\tau \in N_T$ and all $T \geq k + 1$. Then,

$$\sup_{\tau \in N_T} \|\hat{\theta}_\tau - \theta\| = O_p(1), \tag{3.7}$$

$$\sup_{\tau \in N_T} \|\hat{\gamma}_\tau - \gamma\| = o_p(T^\eta), \quad \text{for any } \eta \text{ with } \quad \frac{1}{\alpha} < \eta \leq \frac{1}{2}, \tag{3.8}$$

$$\sup_{\tau \in N_T} \|\hat{b}_\tau - b\| \xrightarrow{p} 0 \tag{3.9}$$

and

$$\sup_{\tau \in N_T} \|T^{1/2}(\hat{\mu}_\tau - \mu) - \hat{U}_T\| \xrightarrow{p} 0, \tag{3.10}$$

where

$$\begin{aligned} \hat{U}_T &= (T^{-1} Z_1' \Sigma(b)^{-1} Z_1)^{-1} T^{-1/2} Z_1' \Sigma(b)^{-1} U \\ &\xrightarrow{d} \omega \left(\lambda B_c(1) + 3(1-\lambda) \int_0^1 s B_c(s) ds \right) \end{aligned} \tag{3.11}$$

with $\omega = \sigma / b(1)$, $\lambda = (1 - \bar{c})/(1 - \bar{c} + \bar{c}^2/3)$ and $B_c(s) = \int_0^s \exp\{c(s - u)\} dB_0(u)$ with $B_0(u)$ a standard Brownian motion. ■

In the same way as in Lemma 1 of SL we have again preferred to assume that the regressor matrix $Z_\tau(\theta)$ is of full column rank although this follows from our previous assumptions for T large enough. Lemma 3.1 shows how the considered estimators behave asymptotically and uniformly in τ. The first result of the lemma is, of course, trivial because the parameter space of θ is assumed to be compact. The second result shows that the maximum distance between the estimator $\hat{\gamma}_\tau$ and the true parameter value diverges in probability. The rate of divergence is related to the existence of moments of the error term ε_t or, equivalently, of the observed process. When high order moments exist, a slower rate of divergence is obtained. Since $\alpha > 2$, the rate of divergence that is always obtainable is $o_p(T^{1/2})$. We have given this rate of divergence as an upper bound in (3.8) because it is needed to prove (3.9) and (3.10). It is also the worst rate of divergence which still suffices for the development of the next section. If the value of τ is assumed known a considerable improvement is obtained in (3.8) because then the right hand side can be replaced by $O_p(1)$ (see Lemma 1 of SL). However, in (3.9) and (3.10) the situation is different and no improvement is obtained even if the value of τ is known. A convenient feature of Lemma 3.1 is that it shows the asymptotic behaviour of the considered estimators in the case where τ is replaced by any estimator. Except for the estimation of γ nothing is asymptotically lost by using an estimator for τ instead of the true parameter value and even in the case of γ the result is not too bad, as mentioned above and will be seen in the next section.

To gain intuition for the above discussion, consider model (2.1) with $f_{t\tau}$ as in (2.7) and suppose that $\bar{c} = 0$. Then the nuisance parameters estimated from the differenced version of (2.1) and the parameters γ_1 and γ_2 are coefficients of impulse dummies. Thus, the estimation of these parameters is clearly asymptotically orthogonal to the estimation of the other parameters and it is also fairly obvious that the situation does not change even if an entirely incorrect value is chosen for τ. This example suggests that an explanation for the nice results of Lemma 3.1 is that the consequences of using any incorrect value of τ are not substantial because under the null hypothesis and local alternatives the parameters τ, θ and γ describe such aspects of the observed process which are only minor. Despite this remark, ignoring these aspects can have serious consequences on unit root testing.

A similar result is obtained by Amsler & Lee (1995). As mentioned in the introduction, their assumptions differ from ours, however. In particular, they use a different assumption regarding the shift point. In their framework the shift occurs at a fixed fraction of the sample, at least asymptotically. Moreover, the shift date has to be chosen in a deterministic, nonrandom way and they do not discuss how that is actually done. In contrast, in our framework the choice of τ may be data dependent and, as mentioned earlier, our shift function can be much more general than the simple shift dummy considered by Amsler & Lee.

Now consider estimating the parameters of the model (2.5). It is easy to see that, upon multiplication by $(1-\bar{\rho}_T L)$, the model can be written in matrix form as

$$Y = W_\tau(\theta)\beta + \mathcal{E}, \tag{3.12}$$

where $\beta = [b' : \phi']'$, $W_\tau(\theta) = [V : Z_\tau(\theta)]$ with V the $(T \times (p-1))$ matrix containing lagged values of the y_t transformed in the same way as the other variables. Furthermore, $\mathcal{E} = [e_1 : \cdots : e_T]'$ is an error term such that $e_t = v_t - \bar{\rho}_T v_{t-1}$. It follows from the definitions that

$$e_t = \varepsilon_t + T^{-1}(c-\bar{c})v_{t-1}. \tag{3.13}$$

In this case the estimators are obtained by minimizing

$$S_{T\tau}(\theta,\beta) = (Y - W_\tau(\theta)\beta)'(Y - W_\tau(\theta)\beta). \tag{3.14}$$

In the same way as above suppose that the matrix $W_\tau(\theta)$ is of full column rank for all values of $\theta \in \Theta$ and all $\tau \in N_T$. It is seen in the appendix that this is the case for all T large enough. Repeating the arguments from LMS we can see that, for any chosen value of τ, a minimizer of $S_{T\tau}(\theta,\beta)$, denoted by $\tilde{\theta}_\tau$ and $\tilde{\beta}'_\tau = [\tilde{b}'_\tau : \tilde{\phi}'_\tau]'$, exists when Assumption A holds. Thus,

$$\tilde{S}_{T\tau} \stackrel{def}{=} S_{T\tau}(\tilde{\theta}_\tau, \tilde{\beta}_\tau) \leq S_{T\tau}(\theta,\beta) \tag{3.15}$$

for all $\theta \in \Theta$ and all β. Note that here β is treated as a free parameter although the true value of b is supposed to define a stable lag polynomial. The estimator $\tilde{\beta}_\tau$ may be written as

$$\tilde{\beta}_\tau = [W_\tau(\tilde{\theta}_\tau)'W_\tau(\tilde{\theta}_\tau)]^{-1}W_\tau(\tilde{\theta}_\tau)'Y. \tag{3.16}$$

The discussion following (3.6) regarding the estimation of τ applies here with obvious modifications. The following lemma gives asymptotic properties of the estimators $\tilde{\theta}_\tau$ and $\tilde{\beta}_\tau$ with $\tilde{\phi}_\tau = [\tilde{\mu}_\tau : \tilde{\gamma}'_\tau]'$ partitioned conformably to ϕ.

Lemma 3.2.
Suppose that Assumption A holds and that the matrix $W_\tau(\theta)$ is of full column rank for all $\theta \in \Theta$, all $\tau \in N_T$ and all $T \geq k+1$. Then, if (2.3) holds with $\alpha > 4$,

$$\sup_{\tau \in N_T} \|\tilde{\theta}_\tau - \theta\| = O_p(1), \tag{3.17}$$

$$\sup_{\tau \in N_T} \|\tilde{\gamma}_\tau - \gamma\| = o_p(T^\eta), \quad \text{for any } \eta \text{ with } \quad \frac{1}{\alpha} < \eta \leq \frac{1}{4}, \tag{3.18}$$

$$\sup_{\tau \in N_T} \|\tilde{b}_\tau - b\| \xrightarrow{p} 0 \tag{3.19}$$

and

$$\sup_{\tau \in N_T} \|T^{1/2}(\tilde{\mu}_\tau - \tilde{b}_\tau(1)\mu/b(1)) - \tilde{U}_T\| \xrightarrow{p} 0, \tag{3.20}$$

where

$$\tilde{U}_T = (T^{-1}Z_1'Z_1)^{-1}T^{-1/2}Z_1'\mathcal{E} \xrightarrow{d} \sigma\left(\lambda B_c(1) + 3(1-\lambda)\int_0^1 sB_c(s)ds\right). \tag{3.21}$$

■

Compared with Lemma 3.1 we now need a stronger moment condition for the error term ε_t. Consequently, the rate of divergence which is always obtainable in (3.18) is $o_p(T^{1/4})$. We have again made this rate of divergence an upper bound because it is needed to prove (3.19) and (3.20). In other respects the discussion given for Lemma 3.1 and Lemma A.1 of LMS applies to the results of Lemma 3.2 with obvious modifications.

We close this section with a remark on the estimation of the parameter τ. An estimator of τ is, of course, needed to make the estimators considered in Lemmas 3.1 and 3.2 feasible. If $\hat{\tau}$ is an estimator of τ, feasible counterparts of $\hat{\phi}_\tau$, $\hat{\theta}_\tau$ and $\hat{b}_\tau$ are defined in an obvious way. They will be denoted by $\hat{\phi}_{\hat{\tau}}$, $\hat{\theta}_{\hat{\tau}}$ and $\hat{b}_{\hat{\tau}}$, respectively. The estimator of τ used to define feasible counterparts of $\tilde{\beta}_\tau = [\tilde{b}_\tau' : \tilde{\phi}_\tau']'$ and $\tilde{\theta}_\tau$ is denoted by $\tilde{\tau}$ so that the resulting estimators are $\tilde{\beta}_{\tilde{\tau}} = [\tilde{b}_{\tilde{\tau}}' : \tilde{\phi}_{\tilde{\tau}}']'$ and $\tilde{\theta}_{\tilde{\tau}}$. It turns out that the asymptotic properties of the unit root tests to be studied in the next section do not depend on the choice of the estimators $\hat{\tau}$ and $\tilde{\tau}$.

4 Testing Procedures

Once the nuisance parameters (including τ) of the models (2.1) and (2.5) have been estimated the residual series $\hat{x}_t = y_t - \hat{\mu}_{\hat{\tau}}t - g_{t\hat{\tau}}(\hat{\theta}_{\hat{\tau}})'\hat{\gamma}_{\hat{\tau}}$ and $\tilde{v}_t = \tilde{b}_{\tilde{\tau}}(L)y_t - \tilde{\mu}_{\tilde{\tau}}t - g_{t\tilde{\tau}}(\tilde{\theta}_{\tilde{\tau}})'\tilde{\gamma}_{\tilde{\tau}}$ may be computed and used to obtain unit root tests. Here $\tilde{b}_{\tilde{\tau}}(L)$ is defined in terms of $\tilde{b}_{\tilde{\tau}}$ in an obvious way. There are several possible unit root tests that can be used. In the following we will only present Dickey-Fuller type tests but note that other tests can be set up in an analogous manner.

First consider model (2.1) and the auxiliary regression model

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

Similar to SL we define $\hat{X} = [\hat{x}_2 : \cdots : \hat{x}_T]'$ and $\hat{X}_{-1} = [\hat{x}_1 : \cdots : \hat{x}_{T-1}]'$ and we introduce the estimators

$$\hat{\rho} = (\hat{X}'_{-1}\Sigma(\hat{b}_{\hat{\tau}})^{-1}\hat{X}_{-1})^{-1}\hat{X}'_{-1}\Sigma(\hat{b}_{\hat{\tau}})^{-1}\hat{X} \tag{4.2}$$

and

$$\hat{\sigma}^2 = (T-1)^{-1}(\hat{X} - \hat{X}_{-1}\hat{\rho})'\Sigma(\hat{b}_{\hat{\tau}})^{-1}(\hat{X} - \hat{X}_{-1}\hat{\rho}). \tag{4.3}$$

For testing the null hypothesis we can now introduce the 't-statistic'

$$\mathcal{T}_1 = (\hat{X}'_{-1}\Sigma(\hat{b}_{\hat{\tau}})^{-1}\hat{X}_{-1})^{1/2}(\hat{\rho} - 1)/\hat{\sigma}. \tag{4.4}$$

The limiting distribution of this test statistic under the local alternatives (3.1) is given in the following theorem.

Theorem 4.1.
Suppose the assumptions of Lemma 3.1 hold. Then,

$$\mathcal{T}_1 \xrightarrow{d} \frac{1}{2}\left(\int_0^1 G_c(s;\bar{c})^2 ds\right)^{-1/2}(G_c(1;\bar{c})^2 - 1),$$

where

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

and λ and $B_c(s)$ are as in Lemma 3.1. □

The limiting distribution in Theorem 4.1 agrees with that obtained in Theorem 1 of SL in the case where the shift date is known a priori. Thus, the discussion of Theorem 1 given in SL also applies here and is not repeated. In particular, critical values for the unit root test are available from Elliott et al. (1996) for $\bar{c} = -13.5$. These authors found that with this choice of $\bar{c}$ the test is nearly optimal for all values of c. Small sample simulations for the case of a known shift date indicate, however, that using nonzero $\bar{c}$ values may result in severe size distortions. Therefore, $\bar{c} = 0$ is the preferred value in the examples considered in Section 5 where also critical values for this case are presented. It is interesting and seems remarkable that the estimation of the integer valued parameter τ has no effect on the asymptotic properties of our test. Hence, Theorem 4.1 also justifies the commonly used approach in which τ is 'estimated' by a visual inspection of the series.

Now consider model (2.5) for which we introduce the auxiliary regression model

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

Analogously to LMS, we define the estimator

$$\tilde{\rho} = \left(\sum_{t=2}^{T} \tilde{v}_{t-1}^2 \right)^{-1} \sum_{t=2}^{T} \tilde{v}_{t-1} \tilde{v}_t, \qquad (4.6)$$

the associated error variance estimator

$$\tilde{\sigma}^2 = (T-1)^{-1} \sum_{t=2}^{T} (\tilde{v}_t - \tilde{\rho} \tilde{v}_{t-1})^2 \qquad (4.7)$$

and the test statistic

$$\mathcal{T}_2 = \left(\sum_{t=2}^{T} \tilde{v}_{t-1}^2 \right)^{1/2} (\tilde{\rho} - 1)/\tilde{\sigma}. \qquad (4.8)$$

For this test statistic we have the following theorem.

Theorem 4.2.
If the assumptions of Lemma 3.2 hold, the limiting distribution of the test statistic $\mathcal{T}_2$ is the same as that of the statistic $\mathcal{T}_1$ inTheorem 4.1. ■

The discussion given for the test statistic $\mathcal{T}_1$ inthe foregoing applies here as well with obvious modifications.

Finally, note that the tests can also be used with the a priori restriction $\mu = 0$. The above tests remain the same except for the limiting distribution which is then the same as in the case without any deterministic terms. Power gains can be considerable compared to tests whose properties depend on deterministic terms as in Elliott et al. (1996). Moreover, seasonal dummies may be included without affecting the limiting distribution of our test statistics.

5 Illustrations

To illustrate how our testing procedures work in practice we use time series which have been considered in previous studies on unit root tests in the presence of structural shifts. Specifically, we use two quarterly and two annual time series. The quarterly series are German Gross National Product (GNP) from 1975(1) to 1996(4) and Polish Industrial Production (IP) from

1982(1) to 1995(4).[1] Both series are in natural logarithms. The German series was analyzed by both SL and LMS whereas the Polish series was considered by LMS only. The annual series are U.S. Employment (1860–1988) and U.S. Industrial Production (1890–1988) from the well-known Nelson & Plosser (1982) data set extended as in Kleibergen & Hoek (1999).[2] They are also in logs. The unit root properties of similar series were analyzed by Perron (1989), Zivot & Andrews (1992) and Amsler & Lee (1995) among others.

All four series are plotted in Figure 1 where they are seen to have quite different characteristics. The German GNP series has a clear shift at the time of the reunification in the third quarter of 1990. Note that the monetary unification took place on July 1, 1990. Therefore the GNP series is defined for West Germany until 1990(2) and for all of Germany afterwards. Hence, the shift is due to a change in the definition of the series. Clearly the timing of the shift is known in this case.

In Poland a market economy was officially introduced at the beginning of 1990. So the adjustment processes started well before that date. In Figure 1 a change in the series is observed in 1989. Due to the necessary adjustments in the economy the shift is not an abrupt one, however. Also it is not clear that it can be uniquely associated with a particular quarter.

Similarly, for the annual U.S. series a shift is seen at the time of the Great Crash in 1929. It has been questioned, however, if such an exogenous dating of the shift is appropriate (see, e.g., Zivot & Andrews (1992)). Thus, we have one series where the date of the shift is clearly known (German GNP) and three series where the shift date is suspected although it is not fully clear due to the adjustment processes which may have been in operation.

For the same reasons the form of the shift is not clear a priori. Therefore, we will apply our tests with different shift functions. More precisely, the shift functions are $f_{t\tau}^{(1)} = d_{t\tau}$,

$$f_{t\tau}^{(2)}(\theta) = \begin{cases} 0, & t < \tau \\ 1 - \exp\{-\theta(t-\tau)\}, & t \geq \tau \end{cases} \quad \text{and} \quad f_{t\tau}^{(3)}(\theta) = \left[\frac{f_{t\tau}^{(1)}}{1-\theta L} : \frac{f_{t-1,\tau}^{(1)}}{1-\theta L} \right]'.$$

[1] The data sources are: GNP — quarterly, seasonally unadjusted data, 1975(1)–1990(2) West Germany, 1990(3)–1996(4) all of Germany, Deutsches Institut für Wirtschaftsforschung, Volkswirtschaftliche Gesamtrechnung.
IP — quarterly, seasonally unadjusted data from Poland 1982(1)–1995(4), International Monetary Fund.

[2] We thank Frank Kleibergen for providing the data.

These shift functions were also used in LMS. The first one is just a shift dummy variable whereas the last two allow for smooth shifts as well, as discussed in Section 2.

The shift date τ will be estimated in different ways. As mentioned previously, one possibility is to choose the shift date by visual inspection of the graph of the series. In that case institutional knowledge may also be incorporated. Clearly, this is the preferred method if the break date is actually known as in the case of the German GNP series. Another possibility is to view τ as a regular nuisance parameter and minimize the relevant objective function with respect to τ in addition to all other nuisance parameters. In the present case, $Q_{T\tau}$ and $S_{T\tau}$ are the objective functions, depending on which model and test is used. Of course, since in the present cases some prior information on the possible ranges of τ is available, it may be useful to restrict the range of permissible τ values in the estimation procedure.

Yet another possibility for estimation the shift date has been considered by Banerjee et al. (1992), Zivot & Andrews (1992) and Perron & Vogelsang (1992). They propose to choose the estimate of τ such that the least favourable result for the unit root null hypothesis is obtained. In other words, in our case we have to choose the estimate which leads to the smallest values of our unit root test statistics $\mathcal{T}_1$ and $\mathcal{T}_2$, depending on which test is employed. Note that this estimator of the shift date is also permitted by our assumptions so that the asymptotic distributions of the test statistics hold for this procedure. According to our assumptions, the estimate of τ has to be fixed before the test is performed. Minimizing the test statistic in order to choose τ may be viewed as a first step and then the actual test is performed using the $\hat{\tau}$ or $\tilde{\tau}$ obtained in this way.

There are also other possibilities for estimating τ that could be considered. We will confine the analysis to the three options "visual inspection," "minimization of the objective function" and "minimization of the unit root test statistic" in the following. In all cases we include a linear trend and for the quarterly series we also include seasonal dummy variables. The value of $\bar{c}$ is fixed at zero. As mentioned in the previous section, this choice was suggested by preliminary simulations. We use the same lag order p that has been used in previous studies. Note that for all series ordinary ADF tests not allowing for a shift do not reject a unit root at a 5% significance level for the series considered here.

[3]The critical values are simulated with a GAUSS programme as follows: series $x_t = x_{t-1} + \varepsilon_t$ $(t = 1, 2, \ldots, 1000)$, $x_0 = 0$, $\varepsilon_t \sim iid\, N(0,1)$ are generated and trend adjusted as $\hat{x}_t = x_t - \hat{\mu}_0 - \hat{\mu} t$, where $\hat{\mu}_0$ and $\hat{\mu}$ are obtained from a regression $\Delta x_t = \mu_0 z_{0t} + \mu + error_t$

Table 1. Unit Root Tests for German log GNP

Estimation method for shift date	$p-1$	Shift function	Test statistic (shift date) $\mathcal{T}_1$ $(\hat{\tau})$	$\mathcal{T}_2$ $(\tilde{\tau})$
visual inspection	4	$f^{(1)}$	-0.14 (1990(3))	-1.01 (1990(3))
		$f^{(2)}$	-0.43 (1990(3))	-1.01 (1990(3))
		$f^{(3)}$	-0.14 (1990(3))	-0.97 (1990(3))
	5	$f^{(1)}$	-0.99 (1990(3))	-1.22 (1990(3))
		$f^{(2)}$	-0.99 (1990(3))	-1.22 (1990(3))
		$f^{(3)}$	-1.04 (1990(3))	-1.08 (1990(3))
minimal objective function $1979(1) \leq \tau \leq 1995(1)$	4	$f^{(1)}$	-0.14 (1990(3))	-1.01 (1990(3))
		$f^{(2)}$	-0.43 (1990(3))	-1.01 (1990(3))
		$f^{(3)}$	-0.45 (1990(2))	-0.97 (1990(3))
	5	$f^{(1)}$	-0.99 (1990(3))	-1.22 (1990(3))
		$f^{(2)}$	-0.99 (1990(3))	-1.22 (1990(3))
		$f^{(3)}$	-1.04 (1990(3))	-1.08 (1990(3))
minimal test statistic $1979(1) \leq \tau \leq 1995(1)$	4	$f^{(1)}$	-1.26 (1994(4))	-1.56 (1994(4))
		$f^{(2)}$	-1.53 (1994(4))	-2.00 (1995(1))
		$f^{(3)}$	-1.57 (1994(1))	-1.98 (1995(1))
	5	$f^{(1)}$	-1.71 (1989(3))	-1.65 (1994(4))
		$f^{(2)}$	-1.91 (1994(4))	-2.01 (1995(1))
		$f^{(3)}$	-1.92 (1994(2))	-2.00 (1995(1))

Critical values: -3.18 (1%), -2.62 (5%), -2.33 (10%)[3]

Assuming a known break date, SL and LMS were unable to reject a unit root in German log GNP. In Table 1 we show the results obtained under our present assumptions. In the first panel, the actual shift date, 1990(3), is assumed a priori. The actual values of the test statistics are different from those given in LMS because these authors use $\bar{c} = -13.5$ whereas we use $\bar{c} = 0$. The general results are the same, however. A unit root cannot be rejected at a significance level of 10% regardless of the lag order, the shift function and the test statistic used. Estimating the shift date by minimizing

with $z_{0t} = 1$ for $t = 1$ and 0 otherwise. The unit root test statistics are then computed from the $\hat{x}_t$ as in (4.4). The simulated critical values are the relevant percentage points based on 10 000 replications.

the objective function with respect to τ in addition to the other nuisance parameters for a period covering the actual shift date in this case gives the same results as before because in almost all cases the estimate of the shift date is identical to the actual shift date. The only exception results for lag order 4 if the shift function $f^{(3)}$ is used in conjunction with test statistic $\mathcal{T}_1$. In that case the shift date is estimated to be the second quarter of 1990. The test decision remains unchanged, however. Quite different estimates of the shift date are obtained via minimization of the test statistics. Now the shift dates range from 1989(3) to 1995(1). Obviously, the estimate of τ is not only sensitive to the shift function and test statistic used, but also to the lag order. Although the test decision is the same as before, that is, a unit root cannot be rejected, the outcome of this procedure sheds doubt on the usefulness of this strategy for estimating the shift date, given that the date is known to be 1990(3) for this series.

For the Polish log IP series LMS find some evidence against a unit root when the break date is assumed to be in 1989(3), that is, half a year before the official starting point of the market economy in Poland. However, LMS use a $\bar{c}$ of -13.5 and prior simulations indicate that for this choice the actual rejection rate of the tests may exceed the nominal significance level considerably if the null hypothesis is correct. In Table 2 the results obtained under our present scenario are given. Using the same estimate for the break date as in LMS, namely 1989(3), now results in insignificant test values even at a 10% level which sheds doubt on the previous results of LMS.

Estimating the shift date by the minimal objective function criterion in most cases results in 1990(1) which is not obvious from Figure 1 but is the official starting point of the market economy in Poland. Again a unit root cannot be rejected even in those cases where another shift date is estimated. In the last panel in Table 2, the minimal test statistic criterion is used for estimating the shift date. The estimates are again more diverse than in the previous panel of Table 2. Clearly, minimizing the objective function is the more robust criterion for estimating τ and, given the outcome of the previous example, it may also be more reliable. For the remaining two examples we therefore focus on the minimization of the objective function in estimating τ and we compare the outcome of the tests to the results from visual inspection of the series.

For the two annual series from the Nelson–Plosser data set Perron (1989) argues that the series exhibit a level shift but not a break in the trend slope. Hence they are in line with our framework. He rejects the unit root hypothesis for both series. Zivot & Andrews (1992) also reject the unit

Table 2. Unit Root Tests for Polish log IP

Estimation method for shift date	$p-1$	Shift function	Test statistic (shift date) $\mathcal{T}_1$ $(\hat{\tau})$	$\mathcal{T}_2$ $(\tilde{\tau})$
visual inspection	2	$f^{(1)}$	-1.33 (1989(3))	-1.26 (1989(3))
		$f^{(2)}$	-1.39 (1989(3))	-1.87 (1989(3))
		$f^{(3)}$	-1.32 (1989(3))	-1.98 (1989(3))
	4	$f^{(1)}$	-1.21 (1989(3))	-1.28 (1989(3))
		$f^{(2)}$	-0.26 (1989(3))	-1.50 (1989(3))
		$f^{(3)}$	0.55 (1989(3))	-1.58 (1989(3))
minimal objective function $1984(1) \le \tau \le 1994(1)$	2	$f^{(1)}$	-1.35 (1990(1))	-1.30 (1990(1))
		$f^{(2)}$	-1.43 (1990(1))	-1.35 (1988(2))
		$f^{(3)}$	-1.37 (1990(1))	-1.36 (1988(2))
	4	$f^{(1)}$	-1.25 (1990(1))	-1.24 (1990(1))
		$f^{(2)}$	-1.34 (1990(1))	-1.29 (1990(1))
		$f^{(3)}$	-1.14 (1989(4))	-1.29 (1990(1))
minimal test statistic $1984(1) \le \tau \le 1994(1)$	2	$f^{(1)}$	-1.35 (1990(1)	-1.30 (1990(1))
		$f^{(2)}$	-1.71 (1988(2)	-1.87 (1989(3))
		$f^{(3)}$	-1.65 (1988(2))	-1.98 (1989(3))
	4	$f^{(1)}$	-1.31 (1991(2))	-1.28 (1989(3))
		$f^{(2)}$	-1.59 (1988(2))	-1.52 (1988(2))
		$f^{(3)}$	-1.47 (1987(1))	-1.62 (1989(2))

Critical values: -3.18 (1%), -2.62 (5%), -2.33 (10%)

root hypothesis for U.S. log IP but they cannot reject a unit root in the Employment series if their finite sample critical values based on Student-t innovations are used. Amsler & Lee (1995) cannot reject a unit root with any of their tests in Employment and find mixed evidence regarding a unit root in the IP series. In our analysis we use the extended series and employ the lag orders given in Table 6 of Zivot & Andrews.

The graphs of the two series in Figure 1 indicate that there was a shift after the Great Crash in 1929. Due to our definition of the shift date we therefore specify $\hat{\tau} = \tilde{\tau} = 1930$ as our visual inspection estimator. Note that our definition of the shift date is slightly different than in some other literature. Due to this difference the shift year 1929 in Perron (1989) and

Table 3. Unit Root Tests for U.S. log Employment

Estimation method for shift date	$p-1$	Shift function	Test statistic (shift date) $\mathcal{T}_1$ $(\hat{\tau})$	$\mathcal{T}_2$ $(\tilde{\tau})$
visual inspection	7	$f^{(1)}$	-2.41 (1930)	-1.61 (1930)
		$f^{(2)}$	-2.43 (1930)	-1.40 (1930)
		$f^{(3)}$	-2.27 (1930)	-1.35 (1930)
minimal objective	7	$f^{(1)}$	-2.37 (1932)	-1.61 (1946)
function		$f^{(2)}$	-2.43 (1930)	-1.40 (1930)
$1908 \leq \tau \leq 1977$		$f^{(3)}$	-2.37 (1931)	-1.35 (1930)

Critical values: -3.18 (1%), -2.62 (5%), -2.33 (10%)

Zivot & Andrews (1992) corresponds to our year 1930. The test results are given in Table 3. Only if $\mathcal{T}_1$ is used in conjunction with $f^{(2)}$, a unit root can be rejected at a 10% significance level. All other tests favor the null hypothesis. The situation is similar if the shift date is estimated by the minimal objective function criterion. Most estimated shift dates are close to 1930, the only exception being if $\mathcal{T}_2$ is used with $f^{(1)}$. For this case, 1946 is obtained as shift date which is not totally unreasonable given the graph in Figure 1. Again the values of the test statistics are very stable and close to the corresponding values in the upper half of Table 3. The only change in the test decision is obtained when $\mathcal{T}_1$ is used with $f^{(3)}$ at a 10% level. Hence, for this series our results are more in line with Zivot & Andrews (1992) in that we find some weak evidence against a unit root in log Employment.

Finally, test results for U.S. log IP are given in Table 4. In this case none of the tests rejects the null hypothesis so that our results contrast with those of Perron (1989) but are in accordance with Zivot & Andrews (1992) and to some extent also with Amsler & Lee (1995). Despite the unanimous test decision, the shift dates obtained with the minimal objective function criterion are now quite different. Using the $\mathcal{T}_1$ setup, 1921 is obtained with all three shift functions whereas the $\mathcal{T}_1$ framework results in shift dates ranging from 1930 to 1932, that is, they are close to the Great Crash.

Overall the examples are meant to illustrate how our tests work. Clearly, it would be useful to know more about the properties of the procedures for estimating the shift dates which are currently in use. We intend to look into that aspect in future research.

Table 4. Unit Root Tests for U.S. log IP

Estimation method for shift date	$p-1$	Shift function	Test statistic (shift date) $\mathcal{T}_1$ ($\hat{\tau}$)	$\mathcal{T}_2$ ($\tilde{\tau}$)
visual inspection	8	$f^{(1)}$	-1.17 (1930)	-2.03 (1930)
		$f^{(2)}$	-1.14 (1930)	-1.78 (1930)
		$f^{(3)}$	-1.05 (1930)	-1.68 (1930)
minimal objective function	8	$f^{(1)}$	-1.52 (1921)	-1.79 (1932)
		$f^{(2)}$	-1.52 (1921)	-1.78 (1930)
$1884 \leq \tau \leq 1978$		$f^{(3)}$	-1.46 (1921)	-1.69 (1931)

Critical values: -3.18 (1%), -2.62 (5%), -2.33 (10%)

6 Conclusions

In this study we have shown that unit root tests can be constructed which work if there is a level shift in a time series of interest. The general approach is to estimate the nuisance parameters in a first step, remove the corresponding parts of the DGP and apply a unit root test of the Dickey–Fuller type to the residuals. It is shown that the asymptotic distributions of the test statistics do not depend on the nuisance parameters. In particular, they do not depend on the shift date. In fact, they do not even depend on the way the shift date is estimated. Therefore, an estimator may be based on a visual inspection of the graph of a series of interest, for example. Perron (1989) was criticized by some authors for assuming an exogenous break date in his unit root tests (see, e.g., Zivot & Andrews (1992)). In our approach it does not matter whether we condition on the shift date or treat it as endogenous. Some empirical examples are discussed to illustrate how the tests work in practice.

In future research it may be of interest to explore the small sample implications of specific estimators for the shift date. In the examples, optimizing the objective function used to estimate the nuisance parameters worked reasonably well. However, more small sample experience is needed to give well-founded advise on which estimator to use. Also the choice of the specific shift function is not a trivial matter. The fact, that our approach accommodates a great variety of very general shift functions leaves the applied researcher with a range of options. In the examples we have used different shift functions. Fortunately, the results pointed at least in the same direction. If there is uncertainty with respect to an adequate shift function it may be reasonable to allow at least for some flexibility in the form of the shift function.

Appendix. Proofs

A.1 Proof of Lemma 3.1

It will be convenient to use the subscript "o" to indicate true parameter values. This means, for instance, that (3.2) is written as $Y = Z_{\tau_o}(\theta_o)\phi_o + U$, where the components of the error term are supposed to satisfy (3.3), as assumed here. Thus, we have the identity

$$Y = Z_\tau(\hat{\theta}_\tau)\phi_o + \hat{\xi}_\tau, \qquad (A.1)$$

where $\hat{\xi}_\tau = U + (Z_{2\tau_o}(\theta_o) - Z_{2\tau}(\hat{\theta}_\tau))\gamma_o$. From this and (3.6) it follows that

$$D_{1T}(\hat{\phi}_\tau - \phi_o) = [D_{1T}^{-1} Z_\tau(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1} Z_\tau(\hat{\theta}_\tau) D_{1T}^{-1}]^{-1} D_{1T}^{-1} Z_\tau(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1}\hat{\xi}_\tau, \qquad (A.2)$$

where $D_{1T} = \mathrm{diag}[T^{1/2} : I_k]$. We shall study the two factors of the product on the r.h.s. and start by showing that the inverse is asymptotically block diagonal. To this end, we first conclude from the definitions that

$$Z_1 = \begin{bmatrix} 1 \\ 1 - \frac{\bar{c}}{T} \\ \vdots \\ 1 - \frac{\bar{c}(T-1)}{T} \end{bmatrix} \quad \text{and} \quad Z_{2\tau}(\theta) = \begin{bmatrix} g_{1\tau}(\theta)' \\ \Delta g_{2\tau}(\theta)' - \frac{\bar{c}}{T} g_{1\tau}(\theta)' \\ \vdots \\ \Delta g_{T\tau}(\theta)' - \frac{\bar{c}}{T} g_{T-1,\tau}(\theta)' \end{bmatrix}.$$

It will sometimes be convenient to denote by Z_{1t} the t-th component of Z_1 and by $Z_{2t\tau}(\theta)'$ the t-th row of $Z_{2\tau}(\theta)$. Since $\|f_{t\tau}(\theta)\| \leq \|\Delta f_{1\tau}\| + \cdots + \|\Delta f_{t\tau}(\theta)\|$ it follows from Assumption A(b) that $\max_{1\leq t\leq T}\|f_{t\tau}(\theta)\|$ can be bounded by a constant independent of θ, τ and T. This boundedness property will be of frequent use. It implies, for instance, that $g_{t\tau}(\theta)$ and $Z_{2t\tau}(\theta)$ are similarly bounded which in conjunction with Assumption A(b) yields

$$T^{-1/2} \sup_{\theta\in\Theta,\tau\in N_T} \|Z_{2\tau}(\theta)' Z_1\| = O(T^{-1/2}). \qquad (A.3)$$

This result will be used to show that

$$T^{-1/2} \sup_{\tau\in N_T} \|Z_{2\tau}(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1} Z_1\| = O_p(T^{-1/2}). \qquad (A.4)$$

To justify this, proceed in the same way as in $(A.9)$ and $(A.14)$ of SL and use the above mentioned boundedness of $Z_{2t\tau}(\theta)$ to conclude that

$$T^{-1/2} Z_{2\tau}(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1} Z_1 = T^{-1/2}\sum_{t=p}^{T}[\hat{b}_\tau(L) Z_{2t\tau}(\hat{\theta}_\tau)][\hat{b}_\tau(L) Z_{1t}] + O_p(T^{-1/2}), \tag{A.5}$$

where the error term is uniform in τ and $\hat{b}_\tau(L) = 1 - \hat{b}_{1\tau}L - \cdots - \hat{b}_{p-1,\tau}L^{p-1}$ is defined in terms of the estimators $\hat{b}_\tau$. Since the roots of $b(L)$ are bounded away from the unit circle by assumption, the estimators $\hat{b}_{j\tau}$, $j = 1, \ldots, p-1$, belong to a bounded set so that $(A.3)$ makes it clear that the first term on the r.h.s. of $(A.5)$ is of order $O_p(T^{-1/2})$ uniformly in τ. Thus we have established $(A.4)$.

It follows from $(A.4)$ that the matrix which is inverted in $(A.2)$ is asymptotically block diagonal. We also need the result that the smallest eigenvalues of the blocks on the diagonal are bounded away from zero uniformly in τ. To this end, note that, analogously to $(A.3)$ of SL, we can now conclude from Assumption A(b) that

$$Z_{2\tau}(\theta)' Z_{2\tau}(\theta) = \sum_{t=1}^{T}\Delta g_{t\tau}(\theta)\Delta g_{t\tau}(\theta)' + O(T^{-1}), \tag{A.6}$$

where the error term is uniform in both θ and τ. Next recall from $(A.5)$ of SL that $\lambda_{max}(\Sigma(b)) \le \bar{K} < \infty$ so that

$$\begin{aligned} &\lambda_{\min}(Z_{2\tau}(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1} Z_{2\tau}(\hat{\theta}_\tau)) \\ \ge\ &\bar{K}^{-1}\lambda_{\min}(Z_{2\tau}(\hat{\theta}_\tau)' Z_{2\tau}(\hat{\theta}_\tau)) \\ \ge\ &\bar{K}^{-1}\inf_{\theta\in\Theta,\tau\in N_T}\lambda_{\min}(\textstyle\sum_{t=1}^{T}\Delta g_{t\tau}(\theta)\Delta g_{t\tau}(\theta)') + o_p(1) \\ \ge\ &\bar{K}^{-1}\varepsilon + o_p(1), \qquad T \ge T_*. \end{aligned} \tag{A.7}$$

Here the second inequality follows from $(A.6)$ and the continuity of eigenvalues and the third one from Assumption A(c). Thus we have shown that the lower right hand block of the matrix that is inverted in $(A.2)$ has its smallest eigenvalue bounded away from zero uniformly in τ. It follows from

$$\begin{aligned} T^{-1} Z_1'\Sigma(\hat{b}_\tau)^{-1} Z_1 &\ge \bar{K}^{-1} T^{-1} Z_1' Z_1 \\ &= \bar{K}^{-1}(1 - \bar{c} + \bar{c}^2/3) + O(T^{-1}) \end{aligned} \tag{A.8}$$

that the same is true for the upper left hand block. Here the equality is obtained from $(A.1)$ of SL. Now, using $(A.4)$, $(A.7)$ and $(A.8)$ in conjunction with Lemma A.2 of Saikkonen & Lütkepohl (1996) one obtains

$$\begin{aligned}&(D_{1T}^{-1}Z_\tau(\hat\theta_\tau)'\Sigma(\hat b_\tau)^{-1}Z_\tau(\hat\theta_\tau)D_{1T}^{-1})^{-1}\\ =\ &\text{diag}[(T^{-1}Z_1'\Sigma(\hat b_\tau)^{-1}Z_1)^{-1}:(Z_{2\tau}(\hat\theta_\tau)'\Sigma(\hat b_\tau)^{-1}Z_{2\tau}(\hat\theta_\tau))^{-1}]+O_p(T^{-1/2})\end{aligned} \tag{A.9}$$

uniformly in τ. Note that, by $(A.7)$ and $(A.8)$ the first term on the r.h.s. of $(A.9)$ is of order $O_p(1)$.

Now consider the latter factor on the r.h.s. of $(A.2)$. We divide our analysis into two parts according to the partition $Z_\tau(\theta)=[Z_1 : Z_{2\tau}(\theta)]$. First note that $(A.4)$ obviously holds even if $Z_{2\tau}(\hat\theta_\tau)$ is replaced by $Z_{2\tau_o}(\theta_o)$. Thus, using the definition of $\hat\xi_\tau$ and $(A.4)$ we find that

$$T^{-1/2}Z_1'\Sigma(\hat b_\tau)^{-1}\hat\xi_\tau = T^{-1/2}Z_1'\Sigma(\hat b_\tau)^{-1}U+O_p(T^{-1/2})=O_p(1) \tag{A.10}$$

uniformly in τ. The latter equality can be justified by using arguments entirely similar to those used in $(A.9)$, $(A.11)$ and $(A.12)$ of SL. It is easy to see that the dependence of the estimator $\hat b_\tau$ on τ has no effect on these arguments. For the second part of our present analysis we note that, uniformly in τ,

$$\begin{aligned}Z_{2\tau}(\hat\theta_\tau)'\Sigma(\hat b_\tau)^{-1}\hat\xi_\tau &= \sum_{t=p}^{T}[\hat b_\tau(L)Z_{2t\tau}(\hat\theta_\tau)][\hat b_\tau(L)\hat\xi_{t\tau}]+O_p(1)\\ &= \sum_{t=p}^{T}[\hat b_\tau(L)Z_{2t\tau}(\hat\theta_\tau)][\hat b_\tau(L)u_t]+O_p(1),\end{aligned} \tag{A.11}$$

where $\hat\xi_{t\tau}$ and u_t are t-th components of the vectors $\hat\xi_\tau$ and U, respectively. These equalities can be justified by using the argument in $(A.9)$ of SL and the fact that

$$\sup_T \sup_{\theta\in\Theta,\tau\in N_T}\sum_{t=1}^{T}\|Z_{2t\tau}(\theta)\|<\infty \tag{A.12}$$

obtained from Assumption A(b) and the definition of $Z_{2t\tau}(\theta)$. To study the first term in the last expression of $(A.11)$, consider for example the quantity

$$\left\|\sum_{t=p}^{T}Z_{2t\tau}(\hat\theta_\tau)u_t\right\| \le \max_{1\le t\le T}|u_t|\sup_T \sup_{\theta\in\Theta,\tau\in N_T}\sum_{t=p}^{T}\|Z_{2t\tau}(\theta)\|. \tag{A.13}$$

The latter factor on the r.h.s. is finite by $(A.12)$, so we need to consider the first one. To this end, recall from (3.3) that

$$u_t = b(L)^{-1}\varepsilon_t + T^{-1}(c-\bar c)x_{t-1} \overset{def}{=} u_t^{(0)} + T^{-1}(c-\bar c)x_{t-1},$$

where $T^{-1}\max_{1\leq t\leq T}|x_{t-1}| = O_p(T^{-1/2})$ (see SL below (3.5)). As for $u_t^{(0)}$, we have $E|u_t^{(0)}|^\alpha < \infty$ by (2.3) so that

$$P\left\{\max_{1\leq t\leq T}|u_t^{(0)}| > T^\eta\varepsilon\right\} \leq \sum_{t=1}^{T} P\{|u_t^{(0)}|^\alpha > T^{\alpha\eta}\varepsilon^\alpha\} \leq \text{const. } \varepsilon^\alpha T^{1-\alpha\eta},$$

where the latter inequality is Markov's. Since the above result holds for every $\epsilon > 0$ we have

$$\max_{1\leq t\leq T}|u_t^{(0)}| = o_p(T^\eta), \quad \eta > 1/\alpha.$$

Thus, we can conclude that

$$\max_{1\leq t\leq T}|u_t| = o_p(T^\eta), \quad \eta > 1/\alpha, \tag{A.14}$$

which, combined with $(A.12)$, shows that the r.h.s. of $(A.13)$ is of order $o_p(T^\eta)$. Since it is clear that the same conclusion obtains even if u_t and $Z_{2t\tau}(\hat{\theta}_\tau)$ in $(A.13)$ are replaced by lagged values it follows that the first term in the last expression of $(A.11)$ is of order $o_p(T^\eta)$ uniformly in τ. Thus, we have shown that

$$Z_{2\tau}(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1}\hat{\xi}_\tau = o_p(T^\eta), \quad \eta > 1/\alpha, \tag{A.15}$$

uniformly in τ. Since $\alpha > 2$ (see (2.3)) we can proceed by assuming that $\eta \leq 1/2$. Hence, using $(A.9)$, $(A.10)$ and $(A.15)$ we find from $(A.2)$ that

$$D_{1T}(\hat{\phi}_\tau - \phi_o) = \begin{bmatrix} T^{1/2}(\hat{\mu}_\tau - \mu_o) \\ \hat{\gamma}_\tau - \gamma_o \end{bmatrix} = \begin{bmatrix} O_p(1) \\ o_p(T^\eta) \end{bmatrix}, \quad 1/\alpha < \eta \leq 1/2, \tag{A.16}$$

uniformly in τ. This proves (3.8) while (3.7) is obvious because the parameter space Θ is compact by assumption.

Next we shall prove (3.9). The argument is similar to that used to prove the consistency of the estimator $\hat{b}$ in the proof of Lemma 1 of SL. Thus, in the same way as in that proof we introduce the notation

$$r_\tau(\theta,\phi) = Z_\tau(\theta)\phi - Z_{\tau_o}(\theta_o)\phi_o = Z_1(\mu - \mu_o) + Z_{2\tau}(\theta)\gamma - Z_{2\tau}(\theta_o)\gamma_o.$$

Since $U = Y - Z_{\tau_o}(\theta_o)\phi_o$, we have $Y - Z_\tau(\theta)\phi = U - r_\tau(\theta,\phi)$ and, in the same way as in the proof of Lemma 1 of SL, we can write

$$\begin{aligned} Q_{T\tau}(\phi,\theta,b) &= U'\Sigma(b)^{-1}U - 2U'\Sigma(b)^{-1}r_\tau(\theta,\phi) + r_\tau(\theta,\phi)'\Sigma(b)^{-1}r_\tau(\theta,\phi) \\ &\overset{def}{=} Q_{1T}(b) + Q_{2T\tau}(\phi,\theta,b) + Q_{3T\tau}(\phi,\theta,b). \end{aligned}$$

We shall show later that

$$T^{-1}Q_{iT\tau}(\hat{\phi}_\tau, \hat{\theta}_\tau, \hat{b}_\tau) = o_p(1), \qquad i = 2, 3, \tag{A.17}$$

uniformly in τ. Assuming this we have,

$$T^{-1}Q_{T\tau}(\phi_o, \theta_o, b_o) \geq T^{-1}Q_{T\tau}(\hat{\phi}_\tau, \hat{\theta}_\tau, \hat{b}_\tau) = T^{-1}Q_{1T}(\hat{b}_\tau) + o_p(1),$$

uniformly in τ. Here the first relation is based on the definition of the estimators $\hat{\phi}_\tau$, $\hat{\theta}_\tau$ and $\hat{b}_\tau$. Since $Q_{1T}(b)$ is the same as its counterpart in SL we have $T^{-1}Q_{1T}(b) \to_p \bar{Q}_1(b)$, where the convergence is uniform in b and the limit is as explained below $(A.15)$ of SL. In particular, $\bar{Q}_1(b)$ is continuous and $\bar{Q}_1(b) \geq \bar{Q}_1(b_o)$ with equality if and only if $b = b_o$. Thus, since $Q_{T\tau}(\phi_o, \theta_o, b_o) = Q_{1T}(b_o)$ we can conclude from the above inequality that

$$\bar{Q}_1(b_o) \geq \bar{Q}_1(\hat{b}_\tau) + o_p(1) \tag{A.18}$$

uniformly in τ. To see that (3.9) follows from this, denote $\hat{b}_\tau = \hat{b}_{T\tau}$ and suppose that (3.9) does not hold. This means that we can find a subsequence $\hat{b}_{T_j\tau}$, say, such that, for some $\geq \varepsilon > 0$ and $\vartheta > 0$,

$$P\left\{\sup_{\tau \in N_{T_j}} \|\hat{b}_{T_j\tau} - b_o\| \geq \varepsilon\right\} \geq \vartheta$$

for all j. This implies that for some $\tau_j \in N_{T_j}$ we have

$$P\{\|\hat{b}_{T_j\tau_j} - b_o\| \geq \varepsilon\} \geq \vartheta$$

for all j. However, since $\bar{Q}_1(b)$ is continuous and uniquely minimized at $b = b_o$ this implies that we can find some $\varepsilon^* > 0$ such that

$$P\left\{\sup_{\tau \in N_{T_j}} \bar{Q}_1(\hat{b}_{T_j\tau}) - \bar{Q}_1(b_o) \geq \varepsilon^*\right\} \geq \vartheta.$$

This is a contradiction to $(A.18)$.

Thus, to complete the proof of (3.9) we have to justify $(A.17)$. The argument is again similar to its counterpart in the proof of Lemma 1 of SL. Recall from $(A.5)$ of SL that $\lambda_{\min}(\Sigma(b)) \geq \underline{K} > 0$ and conclude from this that, uniformly in τ,

$$(\hat{\mu}_\tau - \mu_o)^2 Z_1' \Sigma(\hat{b}_\tau)^{-1} Z_1 \leq \underline{K}^{-1}(\hat{\mu}_\tau - \mu_o)^2 \|Z_1\|^2 = O_p(1),$$

where the equality follows from $(A.16)$ and the latter relation in $(A.8)$. Similarly, we have uniformly in τ,

$$\begin{aligned}
&[Z_{2\tau}(\hat{\theta}_\tau)\hat{\gamma}_\tau - Z_{2\tau_o}(\theta_o)\gamma_o]'\Sigma(\hat{b}_\tau)^{-1}[Z_{2\tau}(\hat{\theta}_\tau)\hat{\gamma}_\tau - Z_{2\tau_o}(\theta_o)\gamma_o] \\
&\leq \underline{K}^{-1}\|Z_{2\tau}(\hat{\theta}_\tau)\hat{\gamma}_\tau - Z_{2\tau_o}(\theta_o)\gamma_o\|^2 \\
&\leq 2\underline{K}^{-1}\|Z_{2\tau}(\hat{\theta}_\tau)\|^2\|\hat{\gamma}_\tau\|^2 + 2\underline{K}^{-1}\|Z_{2\tau_o}(\theta_o)\|^2\|\gamma_o\|^2 \\
&= o_p(T),
\end{aligned}$$

where the equality is justified by $(A.12)$ and $(A.16)$. To see that $(A.17)$ holds for $i = 3$, insert the latter expression of $r_\tau(\theta, \phi)$ in the definition of $Q_{3T\tau}(\phi, \theta, b)$ and use the above results in conjunction with the Cauchy-Schwarz inequality. That $(A.17)$ holds for $i = 2$ can be deduced from this, the previously mentioned fact that $T^{-1}Q_{1\tau}(b)$ converges in probability and uniformly in b, and the Cauchy-Schwarz inequality. Thus, we have established $(A.17)$ and thereby completed the proof of (3.9).

To complete the proof of the lemma, we still have to establish (3.10) and (3.11). The arguments used to obtain $(A.16)$ readily show that (3.10) holds if b in the definition of $\hat{U}_T$ is replaced by the estimator $\hat{b}_\tau$. Thus, we have to show that the error of replacing $\hat{b}_\tau$ by b $(= b_o)$ is of order $o_p(1)$ uniformly in τ. However, since we have proved (3.9) this can be done by using the argument described in the proof of Lemma 1 of SL [see the end of the proof where (3.12) is established]. This proves (3.10) while (3.11) is obtained from the proof of (3.12) in SL. This completes the proof.

A.2 Proof of Lemma 3.2

First note that (see (3.12))

$$Y = W_\tau(\tilde{\theta}_\tau)\beta_o + \tilde{\xi}_\tau \qquad (A.19)$$

where $\tilde{\xi}_\tau = \mathcal{E} + (Z_{2\tau_o}(\theta_o) - Z_{2\tau}(\tilde{\theta}_\tau))\gamma_o$. Thus, using (3.16) we can write

$$D_T(\tilde{\beta}_\tau - \beta_o) = [D_T^{-1}W_\tau(\tilde{\theta}_\tau)'W_\tau(\tilde{\theta}_\tau)D_T^{-1}]^{-1}D_T^{-1}W_\tau(\tilde{\theta}_\tau)'\tilde{\xi}_\tau, \qquad (A.20)$$

where $D_T = \mathrm{diag}[T^{1/2}I_p : I_k]$. Denote $V_1 = [V : Z_1]$ so that $W_\tau(\theta) = [V_1 : Z_{2\tau}(\theta)]$. We shall demonstrate that

$$\begin{aligned}
&(D_T^{-1}W_\tau(\tilde{\theta}_\tau)'W_\tau(\tilde{\theta}_\tau)D_T^{-1})^{-1} \\
&= \mathrm{diag}[(T^{-1}V_1'V_1)^{-1} : (Z_{2\tau}(\tilde{\theta}_\tau)'Z_{2\tau}(\tilde{\theta}_\tau))^{-1}] + o_p(T^{\eta-1/2})
\end{aligned} \qquad (A.21)$$

uniformly in τ. Note that here as well as below $\eta \leq 1/4$ is assumed. The discussion given in the proof of Lemma A.1 of LMS implies that the first inverse on the r.h.s. of $(A.21)$ is of order $O_p(1)$ while $(A.6)$ and Assumption A(c) imply that the same is true for the second inverse uniformly in τ. Thus, from Lemma A.2 of Saikkonen & Lütkepohl (1996) it follows that it suffices to establish $(A.21)$ without taking inverses. Since $(A.3)$ also holds in the present context we only need to show that

$$T^{-1/2} \sup_{\theta\in\Theta,\tau\in N_T} \|Z_{2\tau}(\theta)'V\| = o_p(T^{\eta-1/2}). \qquad (A.22)$$

To justify this, notice that, in the same way as in the proof of Lemma A.1 of LMS, a typical column of the matrix $T^{-1/2}Z_{2\tau}(\theta)'V$ is

$$\begin{aligned} T^{-1/2}\textstyle\sum_{t=1}^{T} Z_{2t\tau}(\theta)(y_{t-i} - \tilde{\rho}_T y_{t-i-1}) &= T^{-1/2}\textstyle\sum_{t=1}^{T} Z_{2t\tau}(\theta)u_{t-i} + O_p(T^{-1/2}) \\ &= o_p(T^{\eta-1/2}) \qquad (1 \leq i \leq p-1) \end{aligned}$$

uniformly in both θ and τ. To prove these equalities, first observe that a representation given in $(A.11)$ of LMS for $y_t - \rho_T y_{t-i-1}$ also applies here if we only replace $g_t(\theta)$ by $g_{t\tau}(\theta)$. The first one of the above equalities is obtained by using this fact in conjunction with $(A.3)$ and $(A.12)$ while the second one follows by combining $(A.12)$, $(A.14)$ and an obvious modification of $(A.13)$. Thus, we have established $(A.22)$ and thereby $(A.21)$ as well.

Next consider the latter factor on the r.h.s. of $(A.20)$. Using the definition of $\tilde{\xi}_\tau$ together with $(A.3)$ and $(A.22)$ yields

$$T^{-1/2}Z_1'\tilde{\xi}_\tau = T^{-1/2}Z_1'\mathcal{E} + O_p(T^{-1/2})$$

and

$$T^{-1/2}V'\tilde{\xi}_\tau = T^{-1/2}V'\mathcal{E} + o_p(T^{\eta-1/2})$$

uniformly in τ. Since $V_1 = [V : Z_1]$ these results give

$$T^{-1/2}V_1'\tilde{\xi}_\tau = T^{-1/2}V_1'\mathcal{E} + o_p(T^{\eta-1/2}) \qquad (A.23)$$

uniformly in τ. Let e_t denote the t-th component of $\mathcal{E}$ and recall from (3.5) of LMS that $e_t = \varepsilon_t + T^{-1}(c - \bar{c})v_{t-1}$, where $T^{-1}\max_{1\leq t\leq T}|v_t| = O_p(T^{-1/2})$. Similarly to $(A.14)$ we therefore have $\max_{1\leq t\leq T}|e_t| = o_p(T^{\eta})$. From this fact, the definition of $\tilde{\xi}_\tau$ and $(A.12)$ one thus obtains

$$Z_{2\tau}(\tilde{\theta}_\tau)'\tilde{\xi}_\tau = Z_{2\tau}(\tilde{\theta}_\tau)'\mathcal{E} + O_p(1) = o_p(T^{\eta}) \qquad (A.24)$$

uniformly in τ. From the proof of Lemma A.1 of LMS we find that $T^{-1/2}V_1'\mathcal{E} = O_p(1)$. Using this fact, $(A.23)$ and $(A.24)$ it follows that

$$D_T W_\tau(\tilde{\theta}_\tau)'\tilde{\xi}_\tau = \begin{bmatrix} T^{-1/2}V_1'\mathcal{E} \\ o_p(T^\eta) \end{bmatrix} + o_p(T^{\eta-1/2}) \qquad (A.25)$$

uniformly in τ. Now, from $(A.20)$, $(A.21)$ and $(A.25)$ we can conclude that

$$D_T(\tilde{\beta}_\tau - \beta_o) = \begin{bmatrix} (T^{-1}V_1'V_1)^{-1}T^{-1/2}V_1'\mathcal{E} \\ o_p(T^\eta) \end{bmatrix} + o_p(T^{2\eta-1/2})$$

uniformly in τ. When $\eta \leq 1/4$ this implies

$$\begin{bmatrix} T^{1/2}(\tilde{b}_\tau - b_o) \\ T^{1/2}(\tilde{\mu}_\tau - \mu_o) \end{bmatrix} = (T^{-1}V_1'V_1)^{-1}T^{-1/2}V_1'\mathcal{E} + o_p(1) \qquad (A.26)$$

uniformly in τ. Since here $T^{-1}V_1'V_1 = R_{11} + o_p(1)$ with R_{11} as in $(A.13)$ of LMS we can proceed in the same way as after $(A.19)$ of LMS and conclude that (3.19) - (3.21) hold. As to the proof of (3.17) and (3.18), the former is again trivial because the parameter space Θ is compact by assumption while the latter is an immediate consequence of $(A.20)$, $(A.21)$ and $(A.25)$. This completes the proof of Lemma 3.2.

A.3 Proof of Theorem 4.1

The proof follows with similar arguments used in the proof of Theorem 1 of SL. First note that (cf. $(A.17)$ of SL)

$$\hat{x}_t = x_t - (\hat{\mu}_{\hat{\tau}} - \mu_o)t - g_{t\hat{\tau}}(\hat{\theta}_{\hat{\tau}})'\hat{\gamma}_{\hat{\tau}} + g_{t\tau_o}(\theta_o)'\gamma_o. \qquad (A.27)$$

Recall that $\max_{1\leq t\leq T}\|g_{t\tau}(\theta)\|$ is bounded uniformly in θ, τ and T. From this fact and (3.8) of Lemma 3.1 it follows that

$$\max_{1\leq t\leq T}\|g_{t\hat{\tau}}(\hat{\theta}_{\hat{\tau}})'\hat{\gamma}_{\hat{\tau}}\| \leq \max_{1\leq t\leq T}\|g_{t\hat{\tau}}(\hat{\theta}_{\hat{\tau}})\|(\|\hat{\gamma}_{\hat{\tau}} - \gamma_o\| + \|\gamma_o\|) = o_p(T^{1/2}). \quad (A.28)$$

Thus, we can conclude from $(A.27)$ that

$$\begin{aligned} T^{-1/2}\hat{x}_{[Ts]} &= T^{-1/2}x_{[Ts]} - T^{1/2}(\hat{\mu}_{\hat{\tau}} - \mu_o)\frac{[Ts]}{T} + o_p(1) \\ &= T^{-1/2}x_{[Ts]} - \hat{U}_T\frac{[Ts]}{T} + o_p(1) \\ &\xrightarrow{d} \omega G_c(s;\bar{c}). \end{aligned} \qquad (A.29)$$

Here the latter equality is based on (3.10) and the weak convergence is obtained from (3.11) and the argument used to obtain $(A.18)$ of SL.

Next note that $\Delta x_t = T^{-1}cx_{t-1} + b(L)^{-1}\varepsilon_t$ by (3.2) of SL. Using this, $(A.28)$, Lemma 3.1, and Assumption A(b) it is not difficult to conclude from $(A.27)$ that

$$\begin{aligned} T^{-1}\textstyle\sum_{t=p}^{T}\Delta\hat{x}_{t-i}\Delta\hat{x}_{t-j} &= T^{-1}\textstyle\sum_{t=p}^{T}\Delta x_{t-i}\Delta x_{t-j} + o_p(1) \\ &= \textstyle\sum_{t=p}^{T}\hat{u}_{t-i}^{(0)}\hat{u}_{t-j}^{(0)} + o_p(1) \qquad (i,j = 0,...,p-1) \end{aligned} \tag{$A.30$}$$

where again $u_t^{(0)} = b(L)^{-1}\varepsilon_t$. Thus, $(A.29)$, $(A.30)$ and the consistency of the estimator $\hat{b}_{\hat{\tau}}$ (see (3.9)) imply that we can repeat the argument used to obtain $(A.20)$ and $(A.21)$ of SL. Hence, we have

$$T^{-2}\hat{X}_{-1}'\Sigma(\hat{b}_{\hat{\tau}})^{-1}\hat{X}_{-1} \xrightarrow{d} \sigma^2\int_0^1 G_c(s;\bar{c})^2 ds \tag{$A.31$}$$

and

$$T^{-1}\hat{X}_{-1}'\Sigma(\hat{b}_{\hat{\tau}})^{-1}(\hat{X}-\hat{X}_{-1}) \xrightarrow{d} \frac{1}{2}\sigma^2 G_c(s;\bar{c})^2 - \frac{1}{2}\sigma^2. \tag{$A.32$}$$

These results imply $\hat{\rho} = 1 + O_p(T^{-1})$ and further $\hat{\sigma}^2 = \sigma^2 + o_p(1)$ so that the stated result follows in the same way as in SL.

A.4 Proof of Theorem 4.2

Using the representation $b(L) = b(1) + b_*(L)\Delta$ we can show in the same way as at the beginning of the proof of Theorem 1 of LMS that

$$\begin{aligned} \tilde{v}_t = {} & v_t - (\tilde{\mu}_{\tilde{\tau}} - \tilde{b}_{\tilde{\tau}}(1)b_o(1)^{-1}\mu_o)t + \tilde{b}_*(1)b_o(1)^{-1}\mu_o \\ & + \tilde{b}_{\tilde{\tau}}(L)k_t + (\tilde{b}_\tau(1) - b_o(1))x_t + (\tilde{b}_{*\tilde{\tau}}(1) - b_{*o}(L))\Delta x_t - g_{t\tilde{\tau}}(\tilde{\theta}_{\tilde{\tau}})'\tilde{\gamma}_{\tilde{\tau}}, \end{aligned}$$

where k_t is similar to $(A.10)$ of LMS. Thus, this equality, Lemma 3.2, an analog of $(A.28)$ and arguments similar to those in the proofs of Theorem 4.1 and Theorem 1 of LMS yield

$$\begin{aligned} T^{-1/2}\tilde{v}_{[Ts]} &= T^{-1/2}v_{[Ts]} - T^{1/2}(\tilde{\mu}_{\tilde{\tau}} - \tilde{b}_{\tilde{\tau}}(1)b_o(1)^{-1}\mu_o)\frac{[Ts]}{T} + o_p(1) \\ &\xrightarrow{d} \sigma G_c(s;\bar{c}). \end{aligned}$$

Hence, proceeding in the same way as in the proof of Theorem 4.1 we can complete the proof. Details are omitted.

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.

Christiano, L.J. (1992), Searching for a break in GNP, *Journal of Business & Economic Statistics*, 10, 237–250.

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

Evans, G.W. (1989), Output and unemployment dynamics in the United States: 1950 - 1985, *Journal of Applied Econometrics*, 4, 213–237.

Granger, C.W.J. & T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships*, New York: Oxford University Press.

Hwang, J. & P. Schmidt (1996), Alternative methods of detrending and the power of unit root tests, *Journal of Econometrics*, 71, 227–248.

Kleibergen, F. & H. Hoek (1999), Bayesian analysis of ARMA models using noninformative priors, Discussion Paper, Econometric Institute and Tinbergen Institute, Erasmus University Rotterdam.

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

Lin, C.-F.J. & T. Teräsvirta (1994), Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics*, 62, 211–228.

Lütkepohl, H., C. Müller & P. Saikkonen (1999), Unit root tests for time series with a structural break when the break point is known, Discussion Paper No. 33, SFB 373, Humboldt-Universität zu Berlin.

Nelson, C.R. & C.I. Plosser (1982), Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics*, 10 139–162.

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.

Phillips, P.C.B. (1987), Time series regression with a unit root, *Econometrica*, 55, 277–301.

Saikkonen, P. & H. Lütkepohl (1996), Infinite-order cointegrated vector autoregressive processes: Estimation and inference, *Econometric Theory*, 12, 814–844.

Saikkonen, P. & H. Lütkepohl (1999), Testing for unit roots in time series with level shifts, Discussion Paper No. 27, 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.

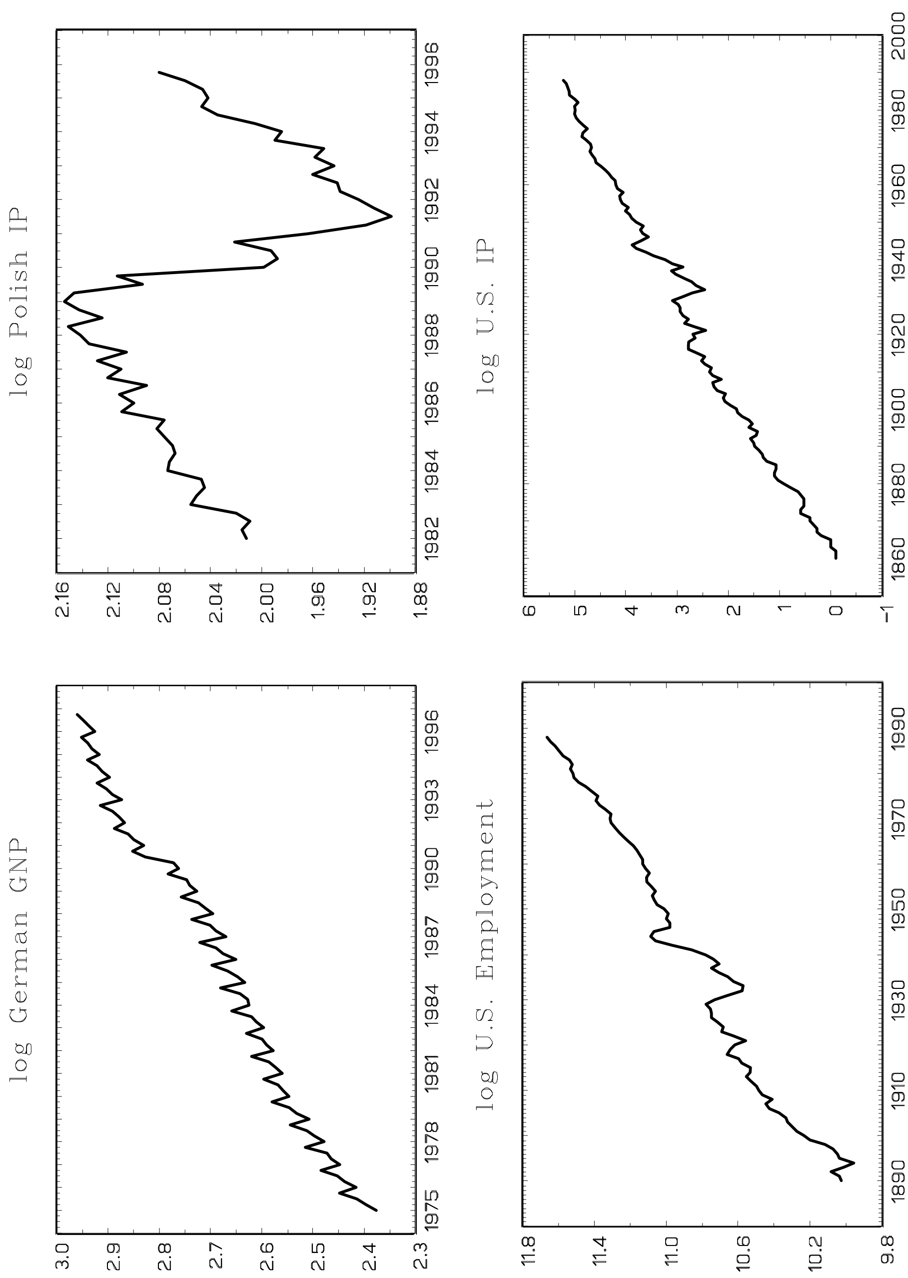

Figure 1: Example time series