Econometric Theory, **18**, 2002, 313–348. Printed in the United States of America.
DOI: 10.1017.S0266466602182053

TESTING FOR A UNIT ROOT IN A TIME SERIES WITH A LEVEL SHIFT AT UNKNOWN TIME

PENTTI SAIKKONEN
University of Helsinki

HELMUT LÜTKEPOHL
European University Institute, Florence
and
Humboldt University

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 the same limiting distributions as for the case of a known break date. Simulations are performed to investigate the small sample properties of the tests, and empirical examples are discussed to illustrate the procedure.

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 and stochastic shocks are transitory. Usually tests are carried out 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.

We are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract ERBFMRXCT980213) for financial support. The first author also acknowledges financial support by the Yrjö Jahnsson Foundation and the Alexander von Humboldt Foundation under a Humboldt research award. Part of this research was done while he was visiting the Humboldt University in Berlin. Comments by two anonymous referees resulted in a number of improvements to the paper. We thank Christian Müller and Markku Lanne for helping with the computations. Address correspondence to: Pentti Saikkonen, Department of Statistics, University of Helsinki, P.O. Box 54, FIN-00014 University of Helsinki, Finland; e-mail: saikkone@valt.helsinki.fi.

 0266-4666/02 $9.50

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 that has occurred at a known date. Examples of articles where this assumption was made are Perron (1989, 1990), Park and Sung (1994), Saikkonen and Lütkepohl (2001), and Lütkepohl, Müller, and Saikkonen (2001). Another part of the literature assumes that the break date is unknown to the investigator and it may be a random event that 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 and Vogelsang (1992), Zivot and Andrews (1992), Banerjee, Lumsdaine, and Stock (1992), and also Leybourne, Newbold, and 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 and Vogelsang, 1992; Zivot and Andrews, 1992; Banerjee et al., 1992). For instance, the former two articles propose to estimate the break date such that the unit root test becomes least favorable to the null hypothesis of a unit root and consider the asymptotic distribution theory of the resulting test statistic. Leybourne et al. (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. They do not show asymptotic properties of their tests so that it is not clear which critical values are appropriate for a specific time series under consideration.

In the present study we will use an approach similar to that of Amsler and Lee (1995). More precisely, we propose estimating all nuisance parameters of the process in a first step in such a way that the limiting distributions of the subsequent unit root tests do not depend on the estimator of the break date. Our approach differs from that of Amsler and 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 and 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 et al. (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 and Lee.

Our procedure for estimating the nuisance parameters extends the approach of Elliott, Rothenberg, and Stock (1996), who propose to estimate the parameters of the deterministic term under local alternatives. Elliott et al. show that their procedure leads to unit root tests with nearly optimal local power. In developing our theoretical results we therefore use this approach for models that allow for a level shift in addition to other deterministic terms.

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 Saikkonen and Lütkepohl (2001) and Lütkepohl et al. (2001) 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. Small sample simulation results are given in Section 5, and empirical examples are discussed in Section 6. Conclusions are contained in Section 7. 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 $\xrightarrow{p}$ and $\xrightarrow{d}$ 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 abbreviates generalized least squares. The m-dimensional Euclidean space is denoted by $\mathbf{R}^m$.

2. THE MODEL FRAMEWORK

Saikkonen and Lütkepohl (2001) and Lütkepohl et al. (2001) 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 Saikkonen and Lütkepohl has the form

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

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 that 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 that 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 $p > 1$, whereas $-1 < \rho \leq 1$. A unit root is present, of course, if $\rho = 1$. There are different possible assumptions that could be made regarding the initial conditions of x_t without affecting the asymptotic derivations. The essential requirement is that the initial conditions must be independent of the sample size T. For convenience we assume, however, that $x_0 = 0$ and $x_t = \rho x_{t-1} + b(L)^{-1}\varepsilon_t$

for $t = 1,2,\ldots$. Using this generation mechanism, there is of course no need for specifying further initial values. 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 subsequently.

The model considered by Lütkepohl et al. (2001) has the form

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

where

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

is an AR process of order one with initial value $v_0 = 0$ again assumed for convenience. The other notation is as before. Obviously, if $\rho = 1$, then v_t and, hence, y_t has a unit root. Initial values of y_t are assumed to be fixed observed presample values.

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. For this shift function the difference between the two models (2.1) and (2.5) is particularly easy to see. For both models the shift function results in a permanent level shift. For the first model the new level is reached immediately in period τ, whereas in model (2.5) there is a gradual transition to the new level. The precise form of the transition depends on the AR operator $b(L)$. In the unit root literature this model is sometimes called an innovational outlier model, whereas (2.1) is referred to as an additive outlier model. Further examples of sequences $f_{t\tau}(\theta)$ will be discussed later.

In Saikkonen and Lütkepohl (2001) and Lütkepohl et al. (2001) it is assumed that the shift point τ is known a priori. In the following discussion 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. Some of the assumptions and results are discussed only briefly. More details on these issues may be found in Saikkonen and Lütkepohl (2001) and Lütkepohl et al. (2001).

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 $\mathbf{R}^m$, 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 exist a real number $\epsilon > 0$ and an integer T_* such that, for all $T \geq T_*$,

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

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

The assumption of a compact parameter space Θ is standard in nonlinear estimation and testing problems. Instead of assuming that the space of τ is the whole set $\{2,\ldots,T-1\}$, as supposed in the preceding special case, 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.

Because $f_{t\tau}(\theta) = \Delta f_{1\tau}(\theta) + \cdots + \Delta f_{t\tau}(\theta)$ Assumption A(b) ensures that $f_{t\tau}(\theta)$ and $g_{t\tau}(\theta)$ are bounded uniformly in t, τ, and θ. Moreover, it guarantees that, if estimation is done under the unit root null hypothesis, the level shift does not affect the limiting distribution of the test statistic, as we will see in the following sections. Estimation under the null hypothesis effectively amounts to using differenced data. Hence, if for example a simple shift dummy variable is considered as shift function, differencing it results in an impulse dummy that has the value one in one period only and is zero otherwise. Therefore it does not affect the asymptotic results as in the approach of Amsler and Lee (1995). The idea of our assumption is to provide a condition that implies the same result in the case of much more general shift functions.

Assumption A(c) guarantees that estimation of nuisance parameters is possible for the differenced model, that is, under the null hypothesis. It ensures that the regressor matrix corresponding to the parameter vector γ is of full column rank for sufficiently large sample size T.

In the terminology of condition B of Elliott et al. (1996), Assumption A implies that $g_{t\tau}(\theta)$ is a slowly evolving trend. It is easy to see that Assumption A is satisfied for the shift dummy defined in (2.7). Because 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, if the parameter space is defined in a suitable way, the assumption holds for a sequence

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

where $\theta > 0$ is an unknown parameter. Another example sequence 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. Yet another possible choice of a shift sequence is 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 and Teräsvirta, 1994, where also alternative specifications are discussed). Leybourne et al. (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 at once due to market inefficiencies, for example. Hence, it is important to allow for the more general nonlinear shifts in the present context.

In Section 6, we consider, for instance, the series of U.S. industrial production, which is likely to have a downward shift at the time of the Great Crash in 1929. Although the crash happened at a known time, it was only the starting point of the related adjustment processes. 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 discussion but will 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, and so on, 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 and Andrews, 1992; Banerjee et al., 1992). We prefer the preceding formulation with integer valued shift date parameter τ because, for our purposes, it offers some advantages in our mathematical derivations. It is therefore used in the following discussion. 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 Section 4.

3. ESTIMATION OF NUISANCE PARAMETERS

In the following discussion we assume that τ is any value from N_T. Notice, however, that the chosen value of τ is not necessarily the true break date. If the value of τ is fixed, the GLS estimation methods considered in Saikkonen and Lütkepohl (2001) and Lütkepohl et al. (2001) can be readily modified for the present context. We will begin with model (2.1).

3.1. Estimation of Model (2.1)

Following Elliott et al. (1996), it is assumed that the true error process x_t is near-integrated so that the parameter ρ 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 $\bar{\rho}_T = 1 + (\bar{c}/T)$ 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,

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

For any given value of τ the parameters θ and ϕ and also 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 $U^{(0)} = [u_1^{(0)} : \cdots : u_T^{(0)}]'$. We denote by $\hat{\phi}_\tau$, $\hat{\theta}_\tau$, and $\hat{b}_\tau$ the resulting estimators of ϕ, θ, and b, respectively. It is shown in the Appendix that these estimators exist for any given value of τ and for all T large enough. Thus we can here simply assume that these estimators exist for all $T \geq k + 1$, which is the case if the matrix $Z_\tau(\theta)$ is of full column rank for all $\theta \in \Theta$ and all $\tau \in N_T$ (see the Appendix). For simplicity, this latter assumption is also made in the following lemma, which 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 use the least squares (LS) estimator obtained by minimizing the function $Q_{T\tau}(\hat{\phi}_\tau, \hat{\theta}_\tau, \hat{b}_\tau)$ with respect to τ. 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. The lemma assumes the local alternatives specified in (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 that the parameter space of b is such 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 ρ is given by (3.1) and 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,*

$$\sup_{\tau \in N_T} \|\hat{\theta}_\tau - \theta\| = O_p(1), \tag{3.5}$$

$$\sup_{\tau \in N_T} \|\hat{\gamma}_\tau - \gamma\| = o_p(T^\eta), \quad \textit{for any } \eta \textit{ with } \quad \frac{1}{\alpha} < \eta \leq \frac{1}{2}, \tag{3.6}$$

$$\sup_{\tau \in N_T} \|\hat{b}_\tau - b\| \xrightarrow{p} 0 \tag{3.7}$$

and

$$\sup_{\tau \in N_T} \|T^{1/2}(\hat{\mu}_\tau - \mu) - \hat{U}_T\| \xrightarrow{p} 0, \tag{3.8}$$

where

$$\hat{U}_T = (T^{-1}Z_1'\Sigma(b)^{-1}Z_1)^{-1}T^{-1/2}Z_1'\Sigma(b)^{-1}U$$

$$\xrightarrow{d} \omega\left(\lambda B_c(1) + 3(1-\lambda)\int_0^1 sB_c(s)\,ds\right) \tag{3.9}$$

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

The lemma 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. Because $\alpha > 2$, 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 (3.6) because it is needed to prove (3.7) and (3.8). It is also the worst rate of divergence that still suffices for the development of the next section. If the value of τ is assumed known a considerable improvement is obtained in (3.6) because then the right-hand side (r.h.s.) can be replaced by $O_p(1)$ (Saikkonen and Lütkepohl, 2001, Lemma 1). However, in (3.7) and (3.8) 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 behavior 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 previously, and will be seen in the next section.

The results in (3.5), (3.6), and (3.8) remain true if $\Sigma(b)$ is replaced by an identity matrix in (3.4), that is, if an LS estimator is used instead of the GLS estimator. Although such a procedure would be even simpler than our GLS procedure, we have treated the GLS approach here because it is often not much more difficult than LS estimation computationally. Moreover, it may result in better small sample properties. Our unit root tests maintain their asymptotic properties even if a simple LS estimator is used for the nuisance parameters because they are based on the results in Lemma 3.1.

To gain intuition for the preceding discussion, consider model (2.1) with $f_{t\tau}$ a shift dummy 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 aspects of the observed process that are only minor. Despite this remark, ignoring these aspects can have serious consequences on unit root testing in finite samples.

A similar result is obtained by Amsler and 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 and Lee.

3.2. Estimation of Model (2.5)

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

$$Y = W_\tau(\theta)\beta + \mathcal{E}, \tag{3.10}$$

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

$$e_t = \varepsilon_t + T^{-1}(c - \bar{c})v_{t-1}. \tag{3.11}$$

In this case the estimators are obtained by minimizing

$$S_{T\tau}(\theta, \beta) = (Y - W_\tau(\theta)\beta)'(Y - W_\tau(\theta)\beta). \tag{3.12}$$

In the same way as in the case of the objective function $Q_{T\tau}(\phi, \theta, b)$, it can be shown that, for any chosen value of τ and any T large enough, 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 (see the Appendix). 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 discussion regarding the estimation of τ in the case of model (2.1) 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. For simplicity we assume in the lemma that the matrix $W_\tau(\theta)$ is of full column rank for all values of θ and all $\tau \in N_T$. This assumption ensures that the estimators $\tilde{\theta}_\tau$ and $\tilde{\beta}_\tau$ exist for all $T \geq k+1$ (see the Appendix).

LEMMA 3.2. *Suppose that Assumption A holds, ρ is given by (3.1), and 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 $\alpha > 4$,*

$$\sup_{\tau \in N_T} \|\tilde{\theta}_\tau - \theta\| = O_p(1), \tag{3.13}$$

$$\sup_{\tau \in N_T} \|\tilde{\gamma}_\tau - \gamma\| = o_p(T^\eta), \quad \textit{for any } \eta \textit{ with } \quad \frac{1}{\alpha} < \eta \leq \frac{1}{4}, \tag{3.14}$$

$$\sup_{\tau \in N_T} \|\tilde{b}_\tau - b\| \xrightarrow{p} 0, \tag{3.15}$$

and

$$\sup_{\tau \in N_T} \|T^{1/2}(\tilde{\mu}_\tau - \tilde{b}_\tau(1)\mu/b(1)) - \tilde{U}_T\| \xrightarrow{p} 0, \tag{3.16}$$

where

$$\tilde{U}_T = (T^{-1}Z_1'Z_1)^{-1}T^{-1/2}Z_1'\mathcal{E} \xrightarrow{d} \sigma\left(\lambda B_c(1) + 3(1-\lambda)\int_0^1 sB_c(s)\,ds\right). \tag{3.17}$$

Compared with Lemma 3.1 we now need a stronger moment condition for the error term ε_t. Consequently, the rate of divergence that is always obtainable in (3.14) is $o_p(T^{1/4})$. We have again made this rate of divergence an upper bound because it is needed to prove (3.15) and (3.16).

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. Analogously, feasible counterparts of $\tilde{\beta}_\tau = [\tilde{b}_\tau' : \tilde{\phi}_\tau']'$ and $\tilde{\theta}_\tau$ are denoted by $\tilde{\beta}_{\hat{\tau}} = [\tilde{b}_{\hat{\tau}}' : \tilde{\phi}_{\hat{\tau}}']'$ and $\tilde{\theta}_{\hat{\tau}}$. 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 estimator $\hat{\tau}$.

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 $\tilde{v}_t = \tilde{b}_{\hat{\tau}}(L)y_t - \tilde{\mu}_{\hat{\tau}} t - g_{t\hat{\tau}}(\tilde{\theta}_{\hat{\tau}})'\tilde{\gamma}_{\hat{\tau}}$ can be computed and used to obtain unit root tests. Here $\tilde{b}_{\hat{\tau}}(L)$ is defined in terms of $\tilde{b}_{\hat{\tau}}$ in an obvious way. There are several possible unit root tests that can be used. In the following discussion we will only present Dickey–Fuller type tests but note that other tests can be set up in an analogous manner. We will provide the limiting distributions of our tests under local alternatives, and we will also consider their (global) consistency against fixed alternatives.

4.1. A Test Based on Model (2.1)

First consider model (2.1) and the auxiliary regression model

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

Similarly to Saikkonen and Lütkepohl (2001) 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 Saikkonen and Lütkepohl (2001) in the case where the shift date is known a priori. It is the one that Elliott et al. (1996) obtained for their t-statistic in a model with mean and linear trend and without a shift term. Critical values for our test are therefore available from Elliott et al. (1996, Table I.C) 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. Obviously, the unit root null hypothesis is rejected for small values of $\mathcal{T}_1$. 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.

Theorem 4.1 implies that the test has more local power if the distance of the local alternative from the null hypothesis increases and the asymptotic power tends to unity as $c \to -\infty$. Unfortunately, the (global) consistency of the test against fixed alternatives is still not easy to show because in our present framework the break date is unknown and may be specified incorrectly and the model is a nonlinear one. Despite these complications the consistency of the test based on $\mathcal{T}_1$ can be established. It follows from the next theorem.

THEOREM 4.2. *Suppose the assumptions of Lemma 3.1 hold except that the value of the parameter ρ is fixed and satisfies $-1 < \rho < 1$. Then, $T^{\eta-1/2}\mathcal{T}_1$ diverges in probability to $-\infty$.*

The consistency of the test based on $\mathcal{T}_1$ follows from this result because $1/\alpha < \eta \leq \frac{1}{2}$ by assumption. The rate of divergence of $\mathcal{T}_1$ improves when α increases or, in other words, when more moments exist for the white noise error process. From the proof of the theorem it can be seen that the rate of divergence is $O_p(T^{1/2})$ in the case where the break date is known. Similarly, this rate applies if N_T is assumed to be a bounded set. This suggests that power gains may be expected if one can use a priori information about the break date. This result is in line with intuition, of course.

Instead of using the statistic $\mathcal{T}_1$, which is based on GLS estimation, an augmented Dickey–Fuller (ADF) test based on

$$\Delta\hat{x}_t = \alpha\hat{x}_{t-1} + \sum_{j=1}^{p-1} \pi_j \Delta\hat{x}_{t-j} + e_t, \quad t = p+1,\ldots,T, \tag{4.5}$$

can be used. Here $\alpha = (1-\rho)b(1)$ and a unit root test with the same asymptotic distribution as $\mathcal{T}_1$ can be based on the usual t-statistic for the null hypothesis $\alpha = 0$ obtained from least squares estimation of the auxiliary model (4.5). This approach is used by Elliott et al. (1996) in their framework. Our GLS approach is more in line with our basic model, however, and it is equally easy to apply in the present context because an estimate of the covariance matrix $\Sigma(b)$ is available from the previous steps of the analysis. As in Elliott et al. (1996) we could also derive point optimal tests. These tests would be based on the statistics $\hat{\sigma}^2(1)$ and $\hat{\sigma}^2(\bar{\rho}_T)$ defined by replacing $\hat{\rho}$ in (4.3) by unity and $\bar{\rho}_T$, respectively. According to the simulation results of Elliott et al. (1996) the overall properties of the ADF t-statistic appeared somewhat better than those of the point optimal tests. We therefore use our DF type statistic $\mathcal{T}_1$ in the following discussion.

4.2. A Test Based on Model (2.5)

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

$$\tilde{v}_t = \rho\tilde{v}_{t-1} + e_t^*, \qquad t = 2,\dots,T. \tag{4.6}$$

We define the estimator

$$\tilde{\rho} = \left(\sum_{t=2}^{T} \tilde{v}_{t-1}^2\right)^{-1} \sum_{t=2}^{T} \tilde{v}_{t-1}\tilde{v}_t, \tag{4.7}$$

the associated error variance estimator

$$\tilde{\sigma}^2 = (T-1)^{-1} \sum_{t=2}^{T} (\tilde{v}_t - \tilde{\rho}\tilde{v}_{t-1})^2, \tag{4.8}$$

and the test statistic

$$\mathcal{T}_2 = \left(\sum_{t=2}^{T} \tilde{v}_{t-1}^2\right)^{1/2} (\tilde{\rho} - 1)/\tilde{\sigma}. \tag{4.9}$$

For this test statistic we have the following theorem.

THEOREM 4.3. *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 $\mathcal{T}_1$ in Theorem 4.1.*

The discussion given for the test statistic $\mathcal{T}_1$ in the foregoing exposition applies here also with obvious modifications. In particular, we can also prove consistency of the test based on $\mathcal{T}_2$ against fixed alternatives. The result follows from the next theorem by noting that $\eta \leq \frac{1}{4}$.

THEOREM 4.4. *Suppose the assumptions of Lemma 3.2 hold except that the value of the parameter ρ is fixed and satisfies $-1 < \rho < 1$. Then, $T^{2\eta-1/2}\mathcal{T}_2$ diverges in probability to $-\infty$.*

Finally, note that both tests can also be used with the a priori restriction $\mu = 0$. The tests remain the same except for the limiting distribution, which is then the same as in the case without any deterministic terms. In the following discussion we denote the resulting test statistics corresponding to $\mathcal{T}_1$ and $\mathcal{T}_2$ by $\mathcal{T}_1^0$ and $\mathcal{T}_2^0$, respectively. 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 distributions of our test statistics. In the next section we consider the small sample properties of our tests.

5. MONTE CARLO STUDY

We have performed a small simulation study to explore the finite sample properties of the unit root tests based on the following two DGP's:

$$y_t = \gamma d_{1t} + x_t, \qquad (1 - b_1 L)(1 - \rho L)x_t = \varepsilon_t, \qquad t = 1,\dots,T, \tag{5.1}$$

and

$$(1 - b_1 L)y_t = \gamma d_{1t} + v_t, \qquad v_t = \rho v_{t-1} + \varepsilon_t, \qquad t = 1,\dots,T, \tag{5.2}$$

with $\varepsilon_t \sim iid\,N(0,1)$, $b_1 = 0.5,-0.5$, $\rho = 1,0.9,0.8$, $T = 100$, and $\gamma = 3$. In the simulations we also generated 100 presample values that were discarded except that they were used in the estimations underlying model (2.5). We also performed simulations with larger samples of size $T = 200$ and other values of the size of the break γ. The results did not affect our main conclusions. Therefore they are not presented in the following discussion.

The first process (5.1) is a special case of the model (2.1) with an abrupt shift at time τ so that the $\mathcal{T}_1$ tests are the appropriate tests, whereas in general the model underlying the $\mathcal{T}_2$ tests can only approximate the DGP (5.1). By applying these tests also we hope to get some indication of the flexibility of the framework and of the consequences of using an approximate model. The DGP (5.2) is a special case of (2.5) and generates a smooth shift in the deterministic term. The $\mathcal{T}_2$ tests are designed for this process, whereas the $\mathcal{T}_1$ tests are approximations only. To capture the smooth transition from one regime to another the $\mathcal{T}_1$ tests have to be combined with a smooth shift function. For both types of tests we use the shift functions $f_{t\tau}^{(1)} = d_{t\tau}$,

$$f_{t\tau}^{(2)}(\theta) = \begin{cases} 0, & t < \tau \\ 1 - \exp\{-\theta(t - \tau + 1)\}, & t \geq \tau \end{cases} \quad \text{and}$$

$$f_{t\tau}^{(3)}(\theta) = \left[\frac{f_{t\tau}^{(1)}}{1 - \theta L} : \frac{f_{t-1,\tau}^{(1)}}{1 - \theta L} \right]'. \tag{5.3}$$

The latter two can generate smooth shifts at time τ. For large values of θ, $f_{t\tau}^{(2)}(\theta)$ will also generate an abrupt shift, and the same is true for $f_{t\tau}^{(3)}(\theta)$ if θ and the second component of γ are close to zero. Thus, the last two shift functions can also well approximate the abrupt shift in DGP (5.1). Although there is no linear trend term in the DGP's we allow for such a term in computing the test statistics $\mathcal{T}_1$ and $\mathcal{T}_2$. In other words, when we consider these tests we assume that the absence of a linear trend is not known a priori.

In Table 1 empirical test sizes are shown for break date estimates $\hat{\tau} = 50$ and $\hat{\tau} = 55$.[1] The first break date is the correct value, of course, whereas $\hat{\tau} = 55$ assumes a misspecified break date. We have picked a quite substantial misspecification of the break date to investigate the implications of using a poor estimate. Notice that in the present case the size of the break is such that choosing the correct break date by visual inspection may be possible. In Table 1 it is

TABLE 1. Empirical sizes of tests, $T = 100$, $\tau = 50$, $\gamma = 3$, nominal significance level 5%

			$\hat{\tau} = 50$				$\hat{\tau} = 55$			
			DGP (5.1)		DGP (5.2)		DGP (5.1)		DGP (5.2)	
Test	$\bar{c}$	Shift function	$b_1 = 0.5$	-0.5	$b_1 = 0.5$	-0.5	$b_1 = 0.5$	-0.5	$b_1 = 0.5$	-0.5
$\mathcal{T}_1^0$	0	$f_t^{(1)}$	0.039	0.056	0.070	0.040	0.032	0.045	0.070	0.050
		$f_t^{(2)}$	0.043	0.058	0.080	0.060	0.033	0.053	0.050	0.050
		$f_t^{(3)}$	0.047	0.062	0.050	0.060	0.033	0.041	0.050	0.040
	-7	$f_t^{(1)}$	0.077	0.157	0.090	0.150	0.066	0.144	0.080	0.200
		$f_t^{(2)}$	0.141	0.263	0.162	0.246	0.165	0.312	0.185	0.313
		$f_t^{(3)}$	0.195	0.304	0.190	0.230	0.169	0.289	0.230	0.330
$\mathcal{T}_2^0$	0	$f_t^{(1)}$	0.023	0.050	0.050	0.060	0.030	0.034	0.050	0.060
		$f_t^{(2)}$	0.025	0.049	0.040	0.060	0.030	0.038	0.030	0.040
		$f_t^{(3)}$	0.035	0.044	0.070	0.060	0.030	0.034	0.070	0.040
	-7	$f_t^{(1)}$	0.191	0.130	0.240	0.090	0.235	0.103	0.230	0.130
		$f_t^{(2)}$	0.299	0.195	0.301	0.183	0.373	0.267	0.392	0.250
		$f_t^{(3)}$	0.364	0.243	0.360	0.240	0.367	0.249	0.400	0.340
$\mathcal{T}_1$	0	$f_t^{(1)}$	0.022	0.060	0.020	0.030	0.016	0.042	0.020	0.070
		$f_t^{(2)}$	0.025	0.063	0.050	0.040	0.030	0.050	0.050	0.090
		$f_t^{(3)}$	0.031	0.072	0.050	0.080	0.028	0.052	0.050	0.090
	-13.5	$f_t^{(1)}$	0.085	0.154	0.130	0.140	0.071	0.126	0.140	0.150
		$f_t^{(2)}$	0.237	0.288	0.271	0.311	0.277	0.310	0.308	0.351
		$f_t^{(3)}$	0.158	0.239	0.190	0.260	0.144	0.198	0.200	0.240
$\mathcal{T}_2$	0	$f_t^{(1)}$	0.016	0.055	0.040	0.060	0.016	0.042	0.040	0.070
		$f_t^{(2)}$	0.025	0.049	0.010	0.060	0.027	0.045	0.030	0.060
		$f_t^{(3)}$	0.019	0.043	0.030	0.080	0.024	0.045	0.030	0.060
	-13.5	$f_t^{(1)}$	0.155	0.174	0.210	0.110	0.238	0.129	0.260	0.120
		$f_t^{(2)}$	0.362	0.262	0.369	0.246	0.567	0.350	0.568	0.344
		$f_t^{(3)}$	0.301	0.204	0.350	0.240	0.337	0.224	0.340	0.230

Notes: Critical values for $\bar{c} = 0$ are -1.96 for $\mathcal{T}_1^0$ and $\mathcal{T}_2^0$; -2.62 for $\mathcal{T}_1$ and $\mathcal{T}_2$. Critical value for $\bar{c} = -7$ is -1.96. Critical value for $\bar{c} = -13.5$ is -2.85.

seen, however, that even a substantial misspecification of the break date does not have a very large impact on the sizes of the tests if $\bar{c} = 0$. A substantial effect is observed, however, for nonzero $\bar{c}$. Obviously, for nonzero $\bar{c}$ all tests have size problems in some situations even if the break date is specified correctly. In some cases these distortions are so large that the tests become useless for practical purposes if estimation is done under local alternatives. On the other hand, if the nuisance parameters are estimated under the unit root hypothesis ($\bar{c} = 0$) all tests have reasonable empirical sizes or are conservative for both processes and all parameter values, even if the break date is estimated incorrectly. Although the sizes improve for larger sample sizes T, we still found considerable distortions for nonzero $\bar{c}$ in simulations with $T = 200$. Therefore we recommend estimating the nuisance parameters under the null hypothesis because controlling the size is a minimum requirement for a test.

Power results are presented in Table 2. An obvious finding from that table is that a poorly estimated break date results in a substantial decline in power. Again, this result is obtained for all tests, DGP's, and shift functions. We have also performed simulations with different values of $\hat{\tau}$ and found that the power tends to decline with increasingly distorted shift estimate, as one would expect. Moreover, using different values of γ we found that the power declines with increasing size of the shift. This result is in line with findings for other unit root tests, for example, by Perron (1989), who observed that a shift in a time series may substantially reduce the power of Dickey–Fuller type tests.

Another observation from Table 2 is that the power is larger if a linear trend term can be excluded a priori, that is, the $\mathcal{T}^0$ tests tend to have more power than the tests that include a linear trend term. Again these results were to be expected. Moreover, the $\mathcal{T}_1$ tests tend to be more powerful than the $\mathcal{T}_2$ tests. Surprisingly, this also holds for DGP (5.2), which can only be approximated by the model underlying the $\mathcal{T}_1$ tests. In other words, using a test that is especially designed for a specific DGP is not necessarily preferable to using an approximate test. Based on these limited simulations, the $\mathcal{T}_1$ tests are therefore our preferred choice for use in applied work. Examples are considered in the next section.

6. ILLUSTRATIONS

To illustrate how our testing procedures work in practice we use two time series that have been considered in previous studies on unit root tests in the presence of structural shifts. These series are annual U.S. Employment (1860–1988) and U.S. Industrial Production (1890–1988) from the well-known Nelson and Plosser (1982) data set extended as in Kleibergen and Hoek (1999).[2] The variables are in logs. The unit root properties of similar series were analyzed by Perron (1989), Zivot and Andrews (1992), and Amsler and Lee (1995) among others. The series are plotted in Figure 1, where they are seen to have a shift 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 and Andrews, 1992). Moreover, as a result of the necessary adjustments in the economy the shift may be a smooth one. Therefore we use the same three shift functions as in Section 5.

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. 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 are used. Of course, because in the present cases some prior information on the possible ranges of τ

TABLE 2. Empirical rejection frequencies of tests, $T = 100$, $\tau = 50$, $\bar{c} = 0$, $\gamma = 3$, nominal significance level 5%

			DGP (5.1)						DGP (5.2)					
			$b_1 = 0.5$			$b_1 = -0.5$			$b_1 = 0.5$			$b_1 = -0.5$		
Test	$\hat{\tau}$	Shift function	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8	$\rho = 1$	0.9	0.8
$\mathcal{T}_1^0$	50	$f_t^{(1)}$	0.039	0.289	0.533	0.056	0.385	0.664	0.070	0.220	0.330	0.052	0.384	0.659
		$f_t^{(2)}$	0.043	0.253	0.469	0.058	0.354	0.618	0.041	0.230	0.406	0.061	0.359	0.629
		$f_t^{(3)}$	0.047	0.215	0.377	0.062	0.323	0.568	0.050	0.170	0.300	0.065	0.337	0.586
	55	$f_t^{(1)}$	0.032	0.234	0.417	0.045	0.144	0.181	0.040	0.120	0.160	0.050	0.246	0.404
		$f_t^{(2)}$	0.033	0.203	0.360	0.053	0.173	0.252	0.034	0.147	0.226	0.064	0.242	0.407
		$f_t^{(3)}$	0.033	0.186	0.358	0.041	0.147	0.209	0.020	0.110	0.220	0.049	0.216	0.375
$\mathcal{T}_2^0$	50	$f_t^{(1)}$	0.023	0.184	0.399	0.050	0.330	0.517	0.050	0.210	0.370	0.045	0.331	0.600
		$f_t^{(2)}$	0.025	0.160	0.330	0.049	0.263	0.495	0.025	0.189	0.374	0.045	0.271	0.501
		$f_t^{(3)}$	0.035	0.186	0.342	0.044	0.255	0.493	0.070	0.160	0.330	0.050	0.246	0.458
	55	$f_t^{(1)}$	0.030	0.206	0.353	0.034	0.116	0.124	0.040	0.080	0.100	0.037	0.194	0.313
		$f_t^{(2)}$	0.030	0.159	0.306	0.038	0.137	0.171	0.036	0.140	0.186	0.055	0.209	0.306
		$f_t^{(3)}$	0.030	0.159	0.277	0.034	0.141	0.157	0.020	0.120	0.190	0.046	0.190	0.293
$\mathcal{T}_1$	50	$f_t^{(1)}$	0.022	0.106	0.317	0.060	0.249	0.606	0.020	0.080	0.330	0.060	0.256	0.615
		$f_t^{(2)}$	0.025	0.105	0.314	0.063	0.259	0.602	0.027	0.109	0.297	0.062	0.266	0.612
		$f_t^{(3)}$	0.031	0.106	0.292	0.072	0.255	0.576	0.050	0.100	0.250	0.067	0.248	0.574
	55	$f_t^{(1)}$	0.016	0.089	0.279	0.042	0.153	0.360	0.030	0.060	0.210	0.055	0.199	0.499
		$f_t^{(2)}$	0.030	0.093	0.257	0.050	0.150	0.326	0.031	0.086	0.199	0.061	0.202	0.458
		$f_t^{(3)}$	0.028	0.092	0.243	0.052	0.151	0.312	0.070	0.090	0.210	0.062	0.203	0.456
$\mathcal{T}_2$	50	$f_t^{(1)}$	0.016	0.069	0.225	0.055	0.198	0.479	0.040	0.100	0.300	0.045	0.331	0.600
		$f_t^{(2)}$	0.015	0.059	0.198	0.049	0.177	0.450	0.014	0.075	0.224	0.039	0.157	0.438
		$f_t^{(3)}$	0.019	0.075	0.221	0.043	0.178	0.439	0.030	0.070	0.230	0.049	0.180	0.438
	55	$f_t^{(1)}$	0.016	0.067	0.217	0.042	0.114	0.267	0.040	0.040	0.140	0.045	0.148	0.360
		$f_t^{(2)}$	0.027	0.071	0.183	0.045	0.109	0.231	0.022	0.048	0.134	0.055	0.148	0.321
		$f_t^{(3)}$	0.024	0.059	0.183	0.045	0.116	0.230	0.050	0.060	0.170	0.049	0.140	0.316

Note: Critical values for $\bar{c} = 0$ are -1.96 for $\mathcal{T}_1^0$ and $\mathcal{T}_2^0$; -2.62 for $\mathcal{T}_1$ and $\mathcal{T}_2$.

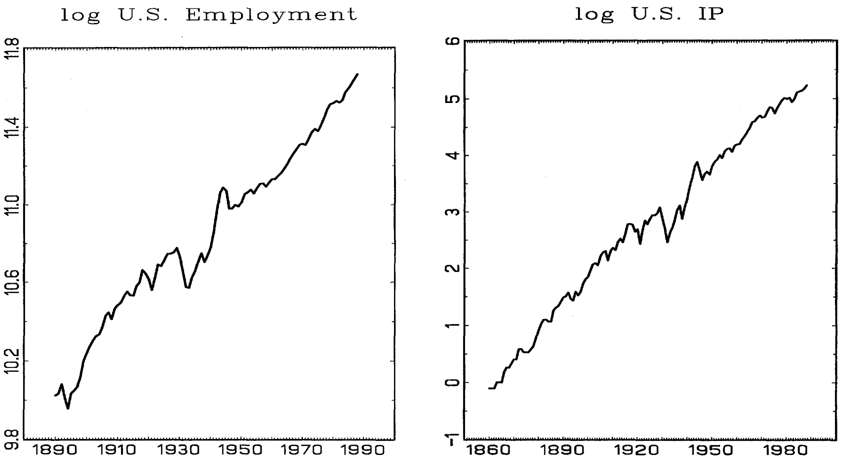

FIGURE 1. Example time series.

is available, it may be useful to restrict the range of permissible τ values in the estimation procedure.

There are also other possibilities for estimating τ that could be considered. We will confine the analysis to the two options "visual inspection" and "minimization of the objective function" in the following discussion, however. In all cases we include a linear trend. The value of $\bar{c}$ is fixed at zero because of the size distortions found in the previous section for nonzero values of $\bar{c}$. We use the same lag order p that has been used in previous studies. Perron (1989) argues that the two 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 and Andrews (1992) also reject the unit root hypothesis for log Industrial Production (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 and 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 and Andrews (1992).

The graphs of the two series in Figure 1 indicate that there was a shift after the Great Crash in 1929. Because of our definition of the shift date we therefore specify $\hat{\tau} = 1930$ as our visual inspection estimator. Note that our definition of the shift date is slightly different than in some other literature. As a result of this difference the shift year 1929 in Perron (1989) and Zivot and Andrews (1992) corresponds to our year 1930. The test results are presented in Tables 3 and 4. Given the low power of tests allowing for a linear trend that was observed in the simulations, using a 10% significance level may be reason-

TABLE 3. Unit root tests for U.S. log Employment

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

Note: Critical values are −2.62 (5%), −2.33 (10%).

able in this case, and, hence, we provide the corresponding critical values in the notes for Tables 3 and 4.

If $\mathcal{T}_1$ is used in conjunction with $f^{(1)}$ and $f^{(2)}$, a unit root can be rejected for the Employment series at a 10% significance level, whereas $\mathcal{T}_2$ favors the null hypothesis. Of course, this result may reflect the lower small sample power of $\mathcal{T}_2$ that was observed in the simulations reported in the previous section. 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 those of Zivot and Andrews (1992) in that we find some weak evidence against a unit root in log Employment.

TABLE 4. Unit root tests for U.S. log IP

Estimation method for shift date	$p-1$	Shift function	Test statistic (shift date) $\mathcal{T}_1$ $(\hat{\tau})$	$\mathcal{T}_2$ $(\hat{\tau})$
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, $1884 \leq \tau \leq 1978$	8	$f^{(1)}$	−1.52 (1921)	−1.79 (1932)
		$f^{(2)}$	−1.52 (1921)	−1.78 (1930)
		$f^{(3)}$	−1.46 (1921)	−1.69 (1931)

Note: Critical values are −2.62 (5%), −2.33 (10%).

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) and Zivot and Andrews (1992) and to some extent also with those of Amsler and 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 $\mathcal{T}_2$ framework results in shift dates ranging from 1930 to 1932, that is, they are close to the Great Crash.

7. CONCLUSIONS

In this study we have shown that unit root tests can be constructed that 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 and Andrews, 1992). In our approach it does not matter whether we condition on the shift date or treat it as endogenous.

In a small Monte Carlo simulation study it is found that estimating the nuisance parameters under local alternatives as recommended elsewhere in the unit root literature may lead to substantial size distortions in the presence of level shifts. Therefore we recommend estimation under the unit root null hypothesis. It is also found that in this case the test sizes are not very sensitive to choosing a poor estimate of the break date, whereas it can have a substantial impact on the test power. More precisely, a loss in power may result from using a poor estimate of the break date. Empirical examples are discussed to illustrate how the tests work in practice.

There are a number of possible directions for extensions of our study. Because the main objective of this study is to present our theoretical approach to treating an unknown shift date in unit root tests, we have only done a small Monte Carlo study to explore the finite sample properties of our tests. In future research it may be of interest to do a more extensive small sample investigation. In practice 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. Finally, it may also be of interest to allow for even greater flexibility by considering a break in the slope of the linear trend term. Unfortunately, such an extension is not a trivial one because a change in the slope is likely to have an impact on the limiting distributions of the test statistics and cannot be handled in our framework in a straightforward manner.

NOTES

1. The critical values are simulated with a GAUSS program as follows: series $x_t = x_{t-1} + \varepsilon_t$ $(t = 1,2,\dots,1{,}000)$, $x_0 = 0$, $\varepsilon_t \sim iid N(0,1)$ are generated and trend (mean) adjusted as $\hat{x}_t = x_t - \hat{\mu}_0 - \hat{\mu}t$ $[\hat{x}_t = x_t - \hat{\mu}_0]$, where $\hat{\mu}_0$ and $\hat{\mu}$ $[\hat{\mu}_0]$ are obtained from a regression $(1 - \bar{\rho}_T L)x_t = \mu_0 z_{0t} + \mu(t - \bar{\rho}_T(t-1)) + error_t$ $[(1 - \bar{\rho}_T L)x_t = \mu_0 z_{0t} + error_t]$ with $z_{0t} = 1$ for $t = 1$ and $(1 - \bar{\rho}_T)$ 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.
2. We thank Frank Kleibergen for providing the data.

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 and Economic Statistics* 10, 271–287.

Christiano, L.J. (1992) Searching for a break in GNP. *Journal of Business and 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.

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

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.

Park, J.Y. & J. Sung (1994) Testing for unit roots in models with structural change. *Econometric Theory* 10, 917–936.

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

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

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

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

Saikkonen, P. & H. Lütkepohl (2001) Testing for unit roots in time series with level shifts. *Allgemeines Statistisches Archiv* 85, 1–25.
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 and Economic Statistics* 10, 251–270.

APPENDIX: PROOFS

In the proofs we make extensive use of results from Saikkonen and Lütkepohl (2001) and Lütkepohl et al. (2001) where possible. Thereby we hope to minimize repetition of derivations and make the main line of arguments more transparent. At the same time the major differences between the known break date case and the presently considered unknown break date situation can be seen.

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

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

where $\hat{\xi}_\tau = U + (Z_{2\tau_o}(\theta_o) - Z_{2\tau}(\hat{\theta}_\tau))\gamma_o$. Assuming that the matrix $Z_\tau(\theta)$ is of full column rank for all $\theta \in \Theta$ and all $\tau \in N_T$ we can repeat the argument used in Section 3 of Saikkonen and Lütkepohl (2001) and conclude that the estimators $\hat{\phi}_\tau$, $\hat{\theta}_\tau$, and $\hat{b}_\tau$ exist. Moreover,

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

From this and (A.1) it follows that

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

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

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

It will sometimes be convenient to denote by Z_{1t} the tth component of Z_1 and by $Z_{2t\tau}(\theta)'$ the tth row of $Z_{2\tau}(\theta)$. Because $\|f_{t\tau}(\theta)\| \leq \|\Delta f_{1\tau}(\theta)\| + \cdots + \|\Delta f_{t\tau}(\theta)\|$ 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

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

This result will be used to show that

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

To justify this, proceed in the same way as in (A.9) and (A.14) of Saikkonen and Lütkepohl (2001) and use the previously 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 $\hat{b}_\tau$. Because the roots of $b(L)$ are bounded away from the unit circle by assumption, the estimators $\hat{b}_{j\tau}$, $j = 1,\dots,p-1$, belong to a bounded set for all τ 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 that 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 Saikkonen and Lütkepohl (2001), 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 note that $\lambda_{\max}(\Sigma(b)) \leq \bar{K} < \infty$ so that

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

Here the second inequality follows from (A.6) 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

$$T^{-1}Z_1'\Sigma(\hat{b}_\tau)^{-1}Z_1 \geq \bar{K}^{-1}T^{-1}Z_1'Z_1$$
$$= \bar{K}^{-1}(1-\bar{c}+\bar{c}^2/3)+O(T^{-1}) \qquad \textbf{(A.8)}$$

that the same is true for the upper left-hand block. Here the equality is obtained from (A.1) of Saikkonen and Lütkepohl (2001). Now, using (A.4), (A.7), and (A.8) in conjunction with Lemma A.2 of Saikkonen and Lütkepohl (1996) one obtains

$$(D_{1T}^{-1}Z_\tau(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1}Z_\tau(\hat{\theta}_\tau)D_{1T}^{-1})^{-1}$$
$$= \operatorname{diag}[(T^{-1}Z_1'\Sigma(\hat{b}_\tau)^{-1}Z_1)^{-1} : (Z_{2\tau}(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1}Z_{2\tau}(\hat{\theta}_\tau))^{-1}] + O_p(T^{-1/2}) \qquad \textbf{(A.9)}$$

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 $O_p(1)$. Note also that from the given derivations it is straightforward to check that (A.9) holds with $O_p(T^{-1/2})$ replaced by $O(T^{-1/2})$ even if $\hat{\theta}_\tau$ and $\hat{b}_\tau$ are replaced by any parameter values θ and b in their respective parameter spaces. This makes clear that the matrix $Z_\tau(\theta)$ is of full column rank for all $\theta \in \Theta$, $\tau \in N_T$ and all T large enough.

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

$$T^{-1/2}Z_1'\Sigma(\hat{b}_\tau)^{-1}\hat{\xi}_\tau = T^{-1/2}Z_1'\Sigma(\hat{b}_\tau)^{-1}U + O_p(T^{-1/2}) = O_p(1) \qquad \textbf{(A.10)}$$

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 Saikkonen and Lütkepohl (2001). 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,

$$Z_{2\tau}(\hat{\theta}_\tau)'\Sigma(\hat{b}_\tau)^{-1}\hat{\xi}_\tau = \sum_{t=p}^{T}[\hat{b}_\tau(L)Z_{2t\tau}(\hat{\theta}_\tau)][\hat{b}_\tau(L)\hat{\xi}_{t\tau}] + O_p(1)$$
$$= \sum_{t=p}^{T}[\hat{b}_\tau(L)Z_{2t\tau}(\hat{\theta}_\tau)][\hat{b}_\tau(L)u_t] + O_p(1), \qquad \textbf{(A.11)}$$

where $\hat{\xi}_{t\tau}$ and u_t are tth components of the vectors $\hat{\xi}_\tau$ and U, respectively. These equalities can be justified by using the argument in (A.9) of Saikkonen and Lütkepohl (2001) and the fact that

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

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

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

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 $T^{-1}\max_{1\le t\le T}|x_{t-1}| = O_p(T^{-1/2})$, which holds because $T^{-1/2}x_{[Ts]}$ satisfies the invariance principle. As for $u_t^{(0)}$, we have $E|u_t^{(0)}|^\alpha < \infty$ by (2.3) so that

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

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

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

Thus, we can conclude that

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

which, combined with (A.12), shows that the r.h.s. of (A.13) is of order $o_p(T^\eta)$. Because 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. Because $\alpha > 2$ (see (2.3)) we can proceed by assuming that $\eta \le \frac{1}{2}$. Hence, using (A.9), (A.10), and (A.15) we find from (A.2) that

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

uniformly in τ. This proves (3.6), whereas (3.5) is obvious because the parameter space Θ is compact by assumption.

Next we shall prove (3.7). We introduce the notation

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

Because $U = Y - Z_{\tau_o}(\theta_o)\phi_o$, we have $Y - Z_\tau(\theta)\phi = U - r_\tau(\theta,\phi)$ and, furthermore,

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

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 $\hat{b}_\tau$. Because $Q_{1T}(b)$ is the same as its counterpart in Saikkonen and Lütkepohl (2001) we have $T^{-1}Q_{1T}(b) \xrightarrow{p} \bar{Q}_1(b)$, where the convergence is uniform in b and $\bar{Q}_1(b)$ equals the variance of the stationary process $b(L)b_o(L)^{-1}\varepsilon_t$ with $b_o(L)$ defined in terms of b_o. As is well known, $\bar{Q}_1(b)$ is continuous and $\bar{Q}_1(b) \geq \bar{Q}_1(b_o)$ with equality if and only if $b = b_o$. Thus, because $Q_{T\tau}(\phi_o,\theta_o,b_o) = Q_{1T}(b_o)$ we can conclude from the preceding 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 (3.7) follows from this, denote $\hat{b}_\tau = \hat{b}_{T\tau}$ and suppose that (3.7) does not hold. This means that we can find a subsequence $\hat{b}_{T_j\tau}$, say, such that, for some $\epsilon > 0$ and $\vartheta > 0$,

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

for all j. This implies that for some $\tau_j \in N_{T_j}$ we have $P\{\|\hat{b}_{T_j\tau_j} - b_o\| \geq \epsilon\} \geq \vartheta$ for all j. However, because $\bar{Q}_1(b)$ is continuous and uniquely minimized at $b = b_o$ this implies that we can find some $\epsilon^* > 0$ such that

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

This is a contradiction to (A.18).

Thus, to complete the proof of (3.7) we have to justify (A.17). Because $\lambda_{\min}(\Sigma(b)) \geq \underline{K} > 0$, we have 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,

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

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 preceding 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 (3.7).

To complete the proof of the lemma, we still have to establish (3.8) and (3.9). The arguments used to obtain (A.16) readily show that (3.8) holds if $b(= b_o)$ 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_o is of order $o_p(1)$ uniformly in τ. Because we have proved (3.7) this follows with standard arguments used in GLS estimation. Thereby (3.8) is established, whereas (3.9) is obtained from the proof of (3.12) in Saikkonen and Lütkepohl (2001). This completes the proof of Lemma 3.1. ∎

Proof of Lemma 3.2.

We shall assume that $v_t = 0$ for $t \leq 0$. First note that

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

where $\tilde{\xi}_\tau = \mathcal{E} + (Z_{2\tau_o}(\theta_o) - Z_{2\tau}(\tilde{\theta}_\tau))\gamma_o$ (see (3.10)). The existence of the estimators $\tilde{\theta}_\tau$ and $\tilde{\beta}_\tau' = [\tilde{b}_\tau' : \tilde{\phi}_\tau']'$ can be established using the arguments from Lütkepohl et al. (2001). These estimators are related by the equation

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

and, according to what was said with respect to the matrix $Z_\tau(\theta)$ after (A.9), it is clear that the same conclusions are obtained for all T large enough even if the stated assumption regarding the rank of the matrix $W_\tau(\theta)$ is not made. Thus, using (A.19) we can write

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

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

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

uniformly in τ. Notice that here and also subsequently $\eta \leq \frac{1}{4}$ is assumed. To justify (A.21), note first that in the latter term on the l.h.s. the first inverse is of order $O_p(1)$ by the discussion given in the proof of Lemma A.1 of Lütkepohl et al. (2001), whereas (A.6) and Assumption A(c) imply that the same is true for the second inverse uniformly in τ. Thus, because (A.3) also holds in the present context we can conclude from the inversion formula of a partitioned matrix and Lemma A.2 of Saikkonen and Lütkepohl (1996) that we only need to show that

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

To justify this, we solve the difference equation (2.5) and use the solution to obtain a representation for the difference $y_t - \bar{\rho}_T y_{t-1}$ whose lagged values define the columns of the matrix V. The solution of (2.5) is discussed in the proof of Lemma A.1 of Lütkepohl et al. (2001), and from equation (A.11) of that paper we find that

$$y_t - \bar{\rho}_T y_{t-1} = Z_{1t}\mu_* + k_{t\tau} - \bar{\rho}_T k_{t-1,\tau} + u_t, \qquad t = 1,2,\ldots, \tag{A.23}$$

where $\mu_* = \mu_o/b_o(1)$, $k_{0\tau} = 0$, and $k_{t\tau}$ $(t > 0)$ is a sequence that depends on the initial values of the difference equation (2.5) and the sequence $g_{t\tau}(\theta_o)$. An important point to note is that in the subsequent analysis we can treat the sequence $k_{t\tau}$ in the same way as $g_{t\tau}(\theta)$. In particular, the differences $\Delta k_{t\tau}$ form an absolutely summable sequence in the same way as $\Delta f_{t\tau}(\theta)$ in Assumption A(b). A final point to notice regarding equation (A.23) is that the process u_t satisfies equation (3.3) except for transient effects caused by differences in initial values that have no effect on asymptotic results and can therefore be ignored.

Now we can show that a typical column of the matrix $T^{-1/2}Z_{2\tau}(\theta)'V$ satisfies

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

uniformly in θ and τ. Here the first equation is obtained by using (A.23), (A.3), and the properties of the sequence $k_{t\tau}$ discussed earlier. The second follows by combining (A.12), (A.14), and an obvious modification of (A.13). Thus, we have established (A.22) and thereby (A.21) also.

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

$$T^{-1/2}Z_1'\tilde{\xi}_\tau = T^{-1/2}Z_1'\mathcal{E} + O_p(T^{-1/2}) \tag{A.24}$$

and

$$T^{-1/2}V'\tilde{\xi}_\tau = T^{-1/2}V'\mathcal{E} + o_p(T^{\eta-1/2}) \tag{A.25}$$

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

$$T^{-1/2}V_1'\tilde{\xi}_\tau = T^{-1/2}V_1'\mathcal{E} + o_p(T^{\eta-1/2}) \tag{A.26}$$

uniformly in τ. Let e_t denote the tth component of $\mathcal{E}$ and note that $e_t = \varepsilon_t + T^{-1}(c - \bar{c})v_{t-1}$. Similarly to (A.14) we therefore have $\max_{1\le t\le T}|e_t| = o_p(T^\eta)$. Using this fact, the definition of $\tilde{\xi}_\tau$, and (A.12) one thus obtains

$$Z_{2\tau}(\tilde{\theta}_\tau)'\tilde{\xi}_\tau = Z_{2\tau}(\tilde{\theta}_\tau)'\mathcal{E} + O_p(1) = o_p(T^\eta) \tag{A.27}$$

uniformly in τ (cf. (A.13)). From (A.26) and (A.27) we find that

$$D_T^{-1}W_\tau(\tilde{\theta}_\tau)'\tilde{\xi}_\tau = \begin{bmatrix} T^{-1/2}V_1'\mathcal{E} + o_p(T^{\eta-1/2}) \\ o_p(T^\eta) \end{bmatrix} \tag{A.28}$$

uniformly in τ. Because $T^{-1/2}V_1'\mathcal{E} = O_p(1)$ (see the proof of Lemma A.1 of Lütkepohl et al., 2001) we can combine (A.20), (A.21), and (A.28) and obtain

$$T^{1/2}\begin{bmatrix} \tilde{b}_\tau - b_o \\ \tilde{\mu}_\tau - \mu_o \end{bmatrix} = (T^{-1}V_1'V_1)^{-1}T^{-1/2}V_1'\mathcal{E} + o_p(T^{2\eta-1/2}) \tag{A.29}$$

and

$$\tilde{\gamma}_\tau - \gamma_o = o_p(T^\eta) \tag{A.30}$$

uniformly in τ. The latter result implies (3.14), whereas (3.13) holds trivially by the assumed compactness of the parameter space Θ. To prove (3.15)–(3.17), define the $(p \times p)$ matrix

$$B = \begin{bmatrix} I_{p-1} & 0 \\ \mu_* \mathbf{1}_{p-1}' & 1 \end{bmatrix},$$

where $\mathbf{1}_{p-1}$ is a $((p-1) \times 1)$ vector of ones. Then, a premultiplication of equation (A.29) by B yields

$$T^{1/2}\begin{bmatrix} \tilde{b}_\tau - b_o \\ \tilde{\mu}_\tau - \tilde{b}_\tau(1)\mu_* \end{bmatrix} = (T^{-1}B^{-1\prime}V_1'V_1B^{-1})^{-1}T^{-1/2}B^{-1\prime}V_1'\mathcal{E} + o_p(T^{2\eta-1/2}), \tag{A.31}$$

where the term $o_p(T^{2\eta-1/2}) = o_p(1)$ is uniform in τ. Now, partition the matrix that is inverted on the r.h.s. conformably with B and note that a typical element of the off-diagonal block is given by

$$T^{-1}\sum_{t=1}^{T}(y_{t-i} - \bar{\rho}_T y_{t-i-1} - Z_{1t}\mu_*)Z_{1t} = O_p(T^{-1/2}), \tag{A.32}$$

where the equality is a straightforward consequence of (A.23) and the properties of the sequence $k_{t\tau}$ and the process u_t therein. Thus, Lemma A.2 of Saikkonen and Lütkepohl (1996) implies that the error of treating the off-diagonal blocks of the matrix $T^{-1}B^{-1\prime}V_1'V_1B^{-1}$ as zeros is of order $O_p(T^{-1/2})$. This fact and arguments used to justify (A.19) and (A.20) of Lütkepohl et al. (2001) imply (3.15)–(3.17). This completes the proof of Lemma 3.2. ∎

Proof of Theorem 4.1.

First note that

$$\hat{x}_t = x_t - (\hat{\mu}_{\hat{\tau}} - \mu_o)t - g_{t\hat{\tau}}(\hat{\theta}_{\hat{\tau}})'\hat{\gamma}_{\hat{\tau}} + g_{t\tau_o}(\theta_o)'\gamma_o. \tag{A.33}$$

Recall that $\max_{1\le t\le T}\|g_{t\tau}(\theta)\|$ is bounded uniformly in θ, τ, and T. From this fact and (3.6) of Lemma 3.1 it follows that

$$\max_{1\le t\le T}\|g_{t\hat{\tau}}(\hat{\theta}_{\hat{\tau}})'\hat{\gamma}_{\hat{\tau}}\| \le \max_{1\le t\le T}\|g_{t\hat{\tau}}(\hat{\theta}_{\hat{\tau}})\|(\|\hat{\gamma}_{\hat{\tau}} - \gamma_o\| + \|\gamma_o\|) = o_p(T^{1/2}). \tag{A.34}$$

Thus, we can conclude from (A.33) that

$$
\begin{aligned}
T^{-1/2}\hat{x}_{[Ts]} &= T^{-1/2}x_{[Ts]} - T^{1/2}(\hat{\mu}_{\hat{\tau}} - \mu_o)\frac{[Ts]}{T} + o_p(1) \\
&= T^{-1/2}x_{[Ts]} - \hat{U}_T \frac{[Ts]}{T} + o_p(1) \\
&\xrightarrow{d} \omega G_c(s;\bar{c}).
\end{aligned} \tag{A.35}
$$

Here the latter equality is based on (3.8) and the weak convergence is obtained from (3.9) and the argument used to obtain (A.18) of Saikkonen and Lütkepohl (2001).

Next note that $\Delta x_t = T^{-1}cx_{t-1} + b(L)^{-1}\varepsilon_t$ by (2.2). Using this, (A.34), Lemma 3.1, and Assumption A(b) it is not difficult to conclude from (A.33) that

$$
\begin{aligned}
T^{-1}\sum_{t=p}^{T}\Delta\hat{x}_{t-i}\Delta\hat{x}_{t-j} &= T^{-1}\sum_{t=p}^{T}\Delta x_{t-i}\Delta x_{t-j} + o_p(1) \\
&= \sum_{t=p}^{T} u_{t-i}^{(0)}u_{t-j}^{(0)} + o_p(1) \qquad (i,j = 0,\ldots,p-1),
\end{aligned} \tag{A.36}
$$

where again $u_t^{(0)} = b_o(L)^{-1}\varepsilon_t$. Thus, (A.35), (A.36), and the consistency of the estimator $\hat{b}_{\hat{\tau}}$ (see (3.7)) imply that we can repeat the argument used to obtain (A.20) and (A.21) of Saikkonen and Lütkepohl (2001). Hence, we have

$$
T^{-2}\hat{X}_{-1}'\Sigma(\hat{b}_{\hat{\tau}})^{-1}\hat{X}_{-1} \xrightarrow{d} \sigma^2\int_0^1 G_c(s;\bar{c})^2 ds \tag{A.37}
$$

and

$$
T^{-1}\hat{X}_{-1}'\Sigma(\hat{b}_{\hat{\tau}})^{-1}(\hat{X} - \hat{X}_{-1}) \xrightarrow{d} \frac{1}{2}\sigma^2 G_c(s;\bar{c})^2 - \frac{1}{2}\sigma^2. \tag{A.38}
$$

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 Saikkonen and Lütkepohl (2001). ■

Proof of Theorem 4.2.

Instead of (3.3) we now have

$$
u_t = x_t - \bar{\rho}_T x_{t-1} = \Delta x_t - \frac{\bar{c}}{T}x_{t-1}. \tag{A.39}
$$

In addition to this identity the following proof makes use of the fact that the properties of the matrices Z_1 and $Z_{2\tau}(\theta)$ are the same as under local alternatives and the estimator $\hat{b}_\tau$ (and $\hat{\theta}_\tau$) is bounded for all τ. These facts imply, for example, that (A.9), (A.11), and the first equation of (A.10) still hold. We shall also make use of approximations similar to those in (A.5) and (A.11). These approximations are always easy to justify and will therefore be employed without further notice (cf. (A.9) and (A.14) of Saikkonen and Lütkepohl, 2001).

We shall next demonstrate that

$$\sup_{\tau \in N_T} \|\hat{\mu}_\tau - \mu_o\| = o_p(T^{\eta-1}) \tag{A.40}$$

and

$$\sup_{\tau \in N_T} \|\hat{\gamma}_\tau - \gamma_o\| = o_p(T^{\eta}), \tag{A.41}$$

where $1/\alpha < \eta \leq \frac{1}{2}$, as in Lemma 3.1. To this end, we first show that the last term on the r.h.s. of (A.10) can be replaced by $O_p(T^{-1/2})$. Because we already noticed that the first equality in (A.10) still holds, it suffices to consider

$$T^{-1/2} Z_1' \Sigma(\hat{b}_\tau)^{-1} U = T^{-1/2} \sum_{t=p}^{T} [\hat{b}_\tau(L) Z_{1t}][\hat{b}_\tau(L) u_t] + O_p(T^{-1/2}), \tag{A.42}$$

where the term $O_p(T^{-1/2})$ is uniform in τ. Now write $\hat{b}_\tau(L) = \hat{b}_\tau(1) + \Delta \hat{b}_\tau^*(L)$ and observe that

$$\hat{b}_\tau(L) Z_{1t} = \hat{b}_\tau(1) Z_{1t} - \frac{\bar{c}}{T} \hat{b}_\tau^*(1), \qquad t = p, \ldots, T.$$

Using this identity in conjunction with (A.39) and the definition of Z_{1t}, it can be seen that the first term on the r.h.s. of (A.42) can be replaced by $O_p(T^{-1/2})$, so that we have shown that

$$T^{-1/2} Z_1' \Sigma(\hat{b}_\tau)^{-1} \hat{\xi}_\tau = O_p(T^{-1/2}) \tag{A.43}$$

uniformly in τ.

Next, using (A.39) and the fact that now x_t is stationary, we can repeat the arguments leading to (A.14) and see that this result also holds under the alternative. Consequently, we still have (A.15), whereas (A.9) was noticed to hold previously. Thus, we can use (A.9), (A.43), and (A.15) to conclude from (A.2) that instead of (A.16) we now have

$$D_{1T}(\hat{\phi}_\tau - \phi_o) = \begin{bmatrix} T^{1/2}(\hat{\mu}_\tau - \mu_o) \\ \hat{\gamma}_\tau - \gamma_o \end{bmatrix} = \begin{bmatrix} o_p(T^{\eta-1/2}) \\ o_p(T^{\eta}) \end{bmatrix}$$

uniformly in τ. This gives (A.40) and (A.41).

For $\hat{b}_\tau$ we now have

$$\sup_{\tau \in N_T} \|\hat{b}_\tau - a\| = o_p(1), \tag{A.44}$$

where $a = [a_1, \ldots, a_p]'$ is obtained by minimizing the function

$$E(\Delta x_t - b_1 \Delta x_{t-1} - \cdots - b_{p-1} \Delta x_{t-p+1})^2 = E[b(L) \Delta x_t]^2$$

with respect to $b_1, \ldots, b_{p-1}$. As is well known, this function is continuous and has a unique minimum. One can justify (A.44) in a straightforward manner by making appropriate changes to the proof given for (3.7). Thus, consider the decomposition $Q_{T\tau}(\phi, \theta, b) = Q_{1T}(b) + Q_{2T\tau}(\phi, \theta, b) + Q_{3T\tau}(\phi, \theta, b)$ defined in the proof of Lemma

3.1. Using (A.40) and (A.41) it is not difficult to check from the proof of Lemma 3.1 that (A.17) holds even under the alternative and further that (A.44) follows if we have $T^{-1}Q_{1T}(b) \xrightarrow{p} E[b(L)\Delta x_t]^2$ uniformly in b. To see this last point, notice that

$$\begin{aligned}
T^{-1}Q_{1T}(b) &= T^{-1}U'\Sigma(b)^{-1}U \\
&= T^{-1}\sum_{t=p}^{T}[b(L)u_t]^2 + o_p(1) \\
&= T^{-1}\sum_{t=p}^{T}[b(L)\Delta x_t]^2 + o_p(1) \\
&= E[b(L)\Delta x_t]^2 + o_p(1),
\end{aligned}$$

where the terms $o_p(1)$ are uniform in b. Here the last two equations are based on the stationarity of the process x_t and equation (A.39). Thus, we have justified (A.44).

The next step is to note that

$$T^{-1}\sum_{t=p}^{T}\hat{x}_{t-i}\hat{x}_{t-j} = T^{-1}\sum_{t=p}^{T}x_{t-i}x_{t-j} + o_p(T^{2\eta}), \qquad i,j = 0,\ldots,p-1. \tag{A.45}$$

This can be justified by using equation (A.33) in conjunction with (A.40), (A.41), and the boundedness of the sequence $g_{tT}(\theta)$ discussed preceding (A.3). Details are straightforward but rather tedious and will be omitted. Note, however, that the second term on the r.h.s. of (A.45) is due to the second sample moments of $(\hat{\mu}_T - \mu_o)t$ and $g_{t\hat{\tau}}(\hat{\theta}_{\hat{\tau}})'\hat{\gamma}_{\hat{\tau}}$ and their lagged values. When $\hat{x}_{t-j}$ on the l.h.s. of (A.45) is replaced by $\Delta\hat{x}_{t-j}$, Assumption A(b) can be used instead of the mere boundedness of the sequence $g_{tT}(\theta)$ so that then arguments otherwise similar to those used for (A.45) yield

$$T^{-1}\sum_{t=p}^{T}\hat{x}_{t-i}\Delta\hat{x}_{t-j} = T^{-1}\sum_{t=p}^{T}x_{t-i}\Delta x_{t-j} + o_p(1), \qquad i,j = 0,\ldots,p-1. \tag{A.46}$$

Now consider the test statistic $\mathcal{T}_1$. Writing $a(L) = 1 - a_1L - \cdots - a_{p-1}L^{p-1}$ we first note that

$$\begin{aligned}
T^{-1}\hat{X}'_{-1}\Sigma(\hat{b}_{\hat{\tau}})^{-1}\hat{X}_{-1} &= T^{-1}\sum_{t=p}^{T}[\hat{b}_{\hat{\tau}}(L)\hat{x}_{t-1}]^2 + o_p(1) \\
&= T^{-1}\sum_{t=p}^{T}[a(L)x_{t-1}]^2 + o_p(T^{2\eta}) \\
&= E[a(L)x_{t-1}]^2 + o_p(T^{2\eta}).
\end{aligned} \tag{A.47}$$

Here the second equation is a straightforward consequence of (A.44) and (A.45), whereas the third equation is due to a weak law of large numbers. Using (A.46) instead of (A.45) we similarly find that

$$\begin{aligned}
T^{-1}\hat{X}'_{-1}\Sigma(\hat{b}_{\hat{\tau}})^{-1}(\hat{X} - \hat{X}_{-1}) &= E[a(L)x_{t-1}][a(L)\Delta x_t] + o_p(1) \\
&\overset{def}{=} \kappa + o_p(1),
\end{aligned} \tag{A.48}$$

where, by the Cauchy–Schwarz inequality, $\kappa < 0$. From (A.47) and (A.48) it follows that $\hat{\rho} - 1 = O_p(1)$ and furthermore that $\hat{\sigma}^2 = O_p(1)$. This latter result can be justified by writing

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

and observing that (A.46) also holds with $\hat{x}_{t-i}$ and x_{t-i} replaced by their differences. Thus, using this fact in conjunction with (A.47) and (A.48), we find that

$$\mathcal{T}_1 = \frac{T^{-1/2}\hat{X}'_{-1}\Sigma(\hat{b}_{\hat{\tau}})^{-1}(\hat{X} - \hat{X}_{-1})}{(T^{-1}\hat{X}_{-1}\Sigma(\hat{b}_{\hat{\tau}})^{-1}\hat{X}_{-1})^{1/2}\hat{\sigma}} = \frac{T^{1/2}(\kappa + o_p(1))}{(O_p(1) + o_p(T^{2\eta}))^{1/2}}. \tag{A.49}$$

Hence, $T^{\eta - 1/2}\mathcal{T}_1 \to -\infty$ in probability as $T \to \infty$ because $\kappa < 0$. Thereby we have established Theorem 4.2. ■

Proof of Theorem 4.3.

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 Lütkepohl et al. (2001) that

$$\tilde{v}_t = v_t - (\tilde{\mu}_{\hat{\tau}} - \tilde{b}_{\hat{\tau}}(1)b_o(1)^{-1}\mu_o)t + \tilde{b}_*(1)b_o(1)^{-1}\mu_o$$
$$+ \tilde{b}_{\hat{\tau}}(L)k_{tT} + (\tilde{b}_T(1) - b_o(1))x_t + (\tilde{b}_{*\hat{\tau}}(1) - b_{*o}(L))\Delta x_t - g_{t\hat{\tau}}(\tilde{\theta}_{\hat{\tau}})'\tilde{\gamma}_{\hat{\tau}},$$

where k_{tT} is as in (A.23). Thus, this equality, Lemma 3.2, an analog of (A.34), and arguments similar to those in the proofs of Theorem 4.1 and Theorem 1 of Lütkepohl et al. (2001) yield

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

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

Proof of Theorem 4.4.

In the same way as in the proof of Theorem 4.2, the first step is to obtain asymptotic properties of the nuisance parameter estimators. This can be done by making appropriate changes to the proof of Lemma 3.2. The assumption $v_t = 0$, $t \leq 0$, used in that proof will also be made here. We begin by considering (A.21). Because under the alternative the properties of the matrices Z_1 and $Z_{2T}(\theta)$ are the same as under local alternatives this amounts to showing that (A.22) still holds. This can be done in exactly the same way as under local alternatives by using the representation (A.23) and the fact that the process

u_t therein satisfies (A.39) except for differences due to initial values. Because (A.24)–(A.26) are based on (A.3) and (A.22) and both of these hold under the alternative we also have (A.24)–(A.26). To see that (A.27) still holds, notice that

$$e_t = \Delta v_t - \frac{\bar{c}}{T} v_{t-1}, \tag{A.50}$$

where v_t is asymptotically stationary by assumption. Thus, $\max_{1 \le t \le T} |e_t| = o_p(T^{\eta})$ by the argument used to show (A.14), and therefore (A.27) follows in the same way as in the proof of Lemma 3.2.

Next consider a typical component of the vector $T^{-1/2} V' \mathcal{E}$, which is

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{T} (y_{t-i} - \tilde{\rho}_T y_{t-i-1}) e_t &= T^{-1/2} \sum_{t=p}^{T} u_{t-i} e_t + O_p(T^{-1/2}) \\ &= O_p(T^{1/2}) \qquad (0 \le i \le p-1). \end{aligned}$$

Here the first equality can be justified by using the representation (A.23) and (A.50). The second one follows from (A.39) and (A.50) and the asymptotic stationarity of the processes x_t and e_t. We also have to consider

$$T^{-1/2} Z_1' \mathcal{E} = T^{-1/2} \sum_{t=1}^{T} Z_{1t} e_t = O_p(T^{-1/2}),$$

where the latter equality is a straightforward consequence of the definition of Z_1, equation (A.50), and the asymptotic stationarity of the process v_t. The preceding two results can now be combined with (A.24), (A.25), and (A.27) to show that

$$D_T^{-1} W_\tau(\tilde{\theta}_\tau)' \tilde{\xi}_\tau = \begin{bmatrix} T^{-1/2} V' \mathcal{E} + o_p(T^{\eta - 1/2}) \\ T^{-1/2} Z_1' \mathcal{E} + O_p(T^{-1/2}) \\ Z_{2\tau}(\tilde{\theta}_\tau)' \mathcal{E} + O_p(1) \end{bmatrix} = \begin{bmatrix} O_p(T^{1/2}) \\ O_p(T^{-1/2}) \\ o_p(T^{\eta}) \end{bmatrix} \tag{A.51}$$

uniformly in τ. It follows from this, (A.20), and (A.21) that, uniformly in τ,

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

Using the definitions we find from this result that

$$\sup_{\tau \in N_T} \| \tilde{\gamma}_\tau - \gamma_o \| = o_p(T^{\eta}), \tag{A.52}$$

which is the desired result for the estimator $\tilde{\gamma}_\tau$. As for the estimators $\tilde{b}_\tau$ and $\tilde{\mu}_\tau$, a similar argument combined with a multiplication by the transformation matrix B introduced in the proof of Lemma 3.2 shows that

$$\begin{bmatrix} \tilde{b}_\tau - b_o \\ \tilde{\mu}_\tau - \tilde{b}_\tau(1) \mu_* \end{bmatrix} = (T^{-1} B^{-1\prime} V_1' V_1 B^{-1})^{-1} T^{-1} B^{-1\prime} V_1' \mathcal{E} + o_p(T^{2\eta - 1}), \tag{A.53}$$

where the term $o_p(T^{2\eta-1}) = o_p(1)$ is uniform in τ. As in (A.32) we now consider a typical element of the off-diagonal block of the matrix that is inverted on the r.h.s. and obtain

$$T^{-1}\sum_{t=1}^{T}(y_{t-i} - \bar{\rho}_T y_{t-i-1} - Z_{1t}\mu_*)Z_{1t} = T^{-1}\sum_{t=1}^{T}u_{t-i}Z_{1t} + O_p(T^{-1}) = O_p(T^{-1}).$$

Here we have used the representation (A.23) and the fact that the process u_t satisfies (A.39) so that the latter equality is justified by the same argument that was used for (A.43). Thus, it follows that the error of treating the off-diagonal blocks of the matrix $T^{-1}B^{-1\prime}V_1'V_1B^{-1}$ in (A.53) as zeros is of order $O_p(T^{-1})$. This result and (A.51) imply that

$$\sup_{\tau\in N_T}\|\tilde{\mu}_\tau - \tilde{b}_\tau(1)\mu_*\| = o_p(T^{2\eta-1}). \tag{A.54}$$

Using (A.23), it can be shown from (A.53) and arguments similar to those employed in the foregoing discussion that the estimator $\tilde{b}_\tau$ converges in probability and uniformly in τ to a fixed constant. Because the explicit expression of this constant is not relevant for us we simply denote it by a^*. Hence, we have

$$\sup_{\tau\in N_T}\|\tilde{b}_\tau - a^*\| = o_p(1). \tag{A.55}$$

Using (A.52), (A.54), and (A.55) in conjunction with the representation of $\tilde{v}_t$ given in the proof of Theorem 4.3 we can now show that

$$T^{-1}\sum_{t=2}^{T}\tilde{v}_{t-1}^2 = T^{-1}\sum_{t=2}^{T}v_{t-1}^2 + o_p(T^{4\eta}) \tag{A.56}$$

and

$$T^{-1}\sum_{t=2}^{T}\tilde{v}_{t-1}\Delta\tilde{v}_t = T^{-1}\sum_{t=2}^{T}v_{t-1}\Delta v_{t-1} + o_p(1). \tag{A.57}$$

Details are straightforward but tedious and will be omitted. We note, however, that the error term in (A.56) is due to the second sample moment of the second term in the representation of $\tilde{v}_t$, which dominates the error but is asymptotically negligible in (A.57). Using (A.56) and (A.57) it is straightforward to analyze the test statistic $\mathcal{T}_2$ in the same way as test statistic $\mathcal{T}_1$ in (A.49) and obtain the stated assertion. ■