February 26, 1999

Testing for Unit Roots in Time Series with Level Shifts*

Pentti Saikkonen and Helmut Lütkepohl†

Department of Statistics
University of Helsinki
P.O. Box 54
SF-00014 University of Helsinki
FINLAND
Tel.: +358-9-1918867
Fax: +358-9-1918872
Email: saikkone@valt.helsinki.fi

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

Abstract

Tests for unit roots in univariate time series with level shifts are proposed and investigated. The level shift is assumed to occur at a known time. It may be a simple one-time shift which can be captured by a dummy variable or it may have a more general form which can be modeled by some general nonlinear transition function. There may also be more than one shift point and there may be other deterministic terms such as a linear trend term or seasonal components. It is proposed to estimate the deterministic parts of the series in a first step by a generalized least squares procedure, subtract the estimated deterministic terms from the series and apply standard unit root tests to the residuals. It is shown that the tests have known asymptotic distributions under the null hypothesis of a unit root and nearly optimal asymptotic power under local alternatives. The procedure is applied to German macroeconomic time series which have a level shift in 1990 where the reunification took place.

JEL classification: C22, C12

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

---

*We are grateful to Christian Müller for performing the computations. The Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213) provided financial support. Part of this research was done while the first author was visiting the Humboldt University in Berlin.

†Corresponding author.

1 Introduction

Today it is common practice in time series econometrics to investigate the trending properties of the variables of interest at an early stage of an analysis. In particular, testing for unit roots is done routinely to check the possibility of stochastic trends in the data generation process (DGP). Such preliminary investigations are of central importance because their outcome determines to some extent which models and inference procedures are suitable in the subsequent analysis. Unfortunately, the usual tests for unit roots are beset with problems. In particular, they are unreliable if structural shifts have occurred during the sample period (see, e.g., Perron (1989)). Since many time series of interest in applied work have quite obvious shifts in their levels the problem is of considerable importance and it is not surprising that it has received substantial attention in the literature (see, e.g., Perron (1990), Perron & Vogelsang (1992), Rappoport & Reichlin (1989), Zivot & Andrews (1992), Banerjee, Lumsdaine & Stock (1992), Amsler & Lee (1995), Ghysels & Perron (1996), Leybourne, Newbold & Vougas (1998), Montañés & Reyes (1998)). Different assumptions regarding the DGP have been made in this context. For instance, the break point may be known or unknown, it may be a shift in the level of a series or it may be a break in the deterministic trend component.

In this study we will assume that the change point is known and we will allow for very general types of shifts which include a number of shift functions that have been proposed in the literature so far. The shift function is set up as a general nonlinear function which depends on unknown parameters. The assumption of a known break point may be regarded as restrictive in some cases. However, there are also many situations where it is quite realistic. For instance, in many German macroeconomic time series there is a shift in 1990 when the German reunification took place. Examples will be given in Sec. 5.

The idea underlying our tests is to estimate and remove the deterministic part of the DGP first and then to apply well-known tests for unit roots to the adjusted data. The deterministic part may include a linear trend term and seasonal components in addition to a quite general nonlinear function representing the shift in the mean of the DGP. Our approach generalizes results of Amsler & Lee (1995) who consider more special shift functions. The resulting unit root tests have distributions under the null hypothesis which are well-known from the unit root literature. Critical values are therefore readily available. A similar approach was also suggested by Leybourne, Newbold & Vougas (1998) who do not assume prior knowledge of the break date and propose to remove the deterministic parts by a least squares (LS) procedure. The disadvantage of their proposal is, however, that the asymptotic distribution of the resulting unit root tests needs to be evaluated by simulation methods on an individual basis, whereas in our approach the asymptotic null distribution of the test statistic does not depend on individual properties of the DGP or the deterministic part. Moreover, our tests are asymptotically nearly optimal under local alternatives in the same way as in Elliott, Rothenberg & Stock (1996).

The structure of the paper is as follows. In the next section the general model is presented and some special cases are discussed in detail. Section 3 considers estimation of the nuisance parameters of the DGP and the tests for unit roots are presented in Section 4. Empirical examples are given in Section 5 and conclusions follow in Section 6. Proofs are deferred to the appendix.

The following general notation is used. The lag and differencing operators are denoted by L and Δ, respectively, that is, for a time series variable y_t we define $Ly_t = y_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$. The symbol $I(d)$ is used to denote a process which is integrated of order d, that is, it is stationary or asymptotically stationary after differencing d times while it is still nonstationary after differencing just $d-1$ times. 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 the matrix A. Moreover, $\|\cdot\|$ and $\|\cdot\|_1$ denote the Euclidean norm and the operator norm, respectively (see, e.g., Lütkepohl (1996) for definitions and properties). GLS is used to abbreviate generalized least squares and sup and inf are short for supremum and infimum, respectively. The n-dimensional Euclidean space is denoted by $\mathbf{R}^n$.

2 A General Model and some Special Cases

We consider a model of the general form

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

where the scalar μ, the $(m \times 1)$ vector θ and the $(k \times 1)$ vector γ are unknown parameters and $g_t(\theta)$ is a $(k \times 1)$ vector of deterministic sequences depending on the parameters θ. The quantity x_t represents an unobservable stochastic error term which is assumed to have a finite order autoregressive (AR) representation of order p,

$$a(L)x_t = \varepsilon_t, \tag{2.2}$$

where $a(L) = 1 - a_1 L - \cdots - a_p L^p$ is a polynomial in the lag operator and $\varepsilon_t \sim iid(0, \sigma^2)$. Assumptions for the initial values will be discussed later. The essential requirement is that they must be independent of the sample size T.

We are interested in testing the null hypothesis that x_t is $I(1)$ against the alternative that it is $I(0)$. Therefore, we assume that the lag polynomial $a(L)$ can be factored as

$$a(L) = (1 - \rho L)b(L) \tag{2.3}$$

where $b(L) = 1 - b_1 L - \cdots - b_{p-1} L^{p-1}$ has all its zeros outside the unit circle if $p > 1$, while $-1 < \rho \leq 1$. Although the parameter space of ρ is restricted to the interval $(-1, 1]$ this will not be taken into account in subsequent estimation and testing procedures.

With respect to the function $g_t(\theta)$ 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,

$$g_t(\theta) = [1 : f_t(\theta)']' \tag{2.4}$$

where $f_t(\theta)$ is a $(k-1)$-dimensional deterministic sequence to be described below. The reason why the trend term has not been included in the function $g_t(\theta)$ is that treating it separately is convenient later on. For illustrative purposes we give examples of possible sequences $f_t(\theta)$ in the following.

A simple version of a function $f_t(\theta)$ that has been considered in the literature (see, e.g., Amsler & Lee (1995)) is one which represents a single shift in the mean,

$$f_t(\theta) = d_{1t} := \begin{cases} 0, & t < T_1 \\ 1, & t \geq T_1 \end{cases} \tag{2.5}$$

that is, d_{1t} is a shift dummy variable and we assume that T_1 is known. An easy extension of this model would be to allow for more than one shift and/or include impulse dummy variables in addition.

Although assuming a shift in the mean at some time point may be reasonable occasionally one may sometimes wish to consider models in which the effect of the dummies is gradual or smoother than in (2.5) (see Leybourne, Newbold & Vougas (1998) for a discussion of reasons). One possibility to achieve this is to define

$$f_t(\theta) = \psi_t(\theta) := \begin{cases} 0, & t < T_1 \\ 1 - \exp\{-\theta(t - T_1)\}, & t \geq T_1 \end{cases} \tag{2.6}$$

or

$$f_t(\theta) = \begin{cases} 0, & t < T_1 \\ \exp\{-\theta(t - T_1)\}, & t \geq T_1 \end{cases} \tag{2.7}$$

with $\theta > 0$ an unknown parameter. Both of these functions generate smooth transitions of the mean and they could be combined as two components of the function $g_t(\cdot)$ in which case the parameters may differ, of course. Similar ideas have been used in modeling the transition of regression equations in smooth transition regression models (e.g., Granger & Teräsvirta (1993), Lin & Teräsvirta (1994)). Of course, these are just examples of various possibilities one might consider. They are related to the cumulative distribution function and the density function of the exponential distribution. In the same way one may consider other density functions or distribution functions.

Another possibility to model smooth effects of dummies is to follow the approach used in intervention analysis (see Box & Tiao (1975) and Franses & Haldrup (1994) for a recent application to unit root testing). In this context we may consider a shift function

$$\frac{\delta(L)}{\varphi(L)} d_{1t},$$

where d_{1t} is a step dummy as defined in (2.5), $\delta(L) = \delta_0 - \delta_1 L - \cdots - \delta_q L^q$ and $\varphi(L) = 1 - \varphi_1 L - \cdots - \varphi_r L^r$ are lag polynomials such that the zeros of $\varphi(L)$ lie outside the unit circle. This latter condition guarantees that the interpretation of the dummy is basically the same as in (2.5). Indeed, if unit roots were allowed in $\varphi(L)$ so that $\varphi(1) = 0$, the effect of the step dummy would essentially change the slope parameter μ at T_1 whereas in the present paper we are interested in modeling level shifts. In terms of the basic model (2.1)/(2.4) we can write the shift function as

$$f_t(\theta) = \left[\frac{d_{1t}}{\varphi(L)} : \cdots : \frac{d_{1,t-q}}{\varphi(L)}\right]', \tag{2.8}$$

where the components of θ are given by the unknown coefficients of $\varphi(L)$.

A simple special case of (2.8) is obtained by choosing $\delta(L) = \delta_0$ and $\varphi(L) = 1 - \theta L$, where $0 \leq \theta < 1$ is a reasonable additional assumption. The model obtained in this way is actually very close to (2.6) or (2.7) the main difference being that the sequences in (2.6) and (2.7) are bounded between zero and one while $(1 - \theta L)^{-1} d_{1t}$ takes values larger than one. To put this another way, the parameter θ in (2.6) or (2.7) affects only the shape of the sequence $f_t(\theta)$ while θ in (2.8) affects both the shape and the size of the shift function.

The parameters μ and γ in the model (2.1) are supposed to be completely unrestricted although the case where $\mu = 0$ a priori will be discussed. Conditions required for the parameters θ and the sequence $f_t(\theta)$ are collected in the following set of assumptions.

Assumption 1

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

(b) For each $t = 1, 2, \ldots,$ $f_t(\theta)$ is a continuous function of θ and

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

where $f_0(\theta) = 0$.

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

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

where we define $\Delta g_1(\theta) = [1 : f_1(\theta)']'$. □

Thus, we restrict the parameter space of θ to be compact. This is a standard assumption in nonlinear estimation and testing problems. The same is true for the continuity requirement in Assumption 1(b). Assuming that the parameter space Θ is defined in a suitable way the summability condition in Assumption 1(b) holds in the applications we have in mind and in that sense it is not restrictive. To understand why the summability condition in

Assumption 1(b), as well as the condition in Assumption 1(c), is formulated for differences of the sequences $f_t(\theta)$ and $g_t(\theta)$, recall that our intention is to study unit root testing. Therefore we shall consider estimation of the parameters μ, θ and γ under the null hypothesis that the error process in (2.1) contains a unit root. Efficient estimation then requires that the variables in (2.1) are differenced, which explains why differences appear in Assumption 1. To see the meaning of the condition in Assumption 1(c), suppose first that the value of the parameter θ is known and that the parameters μ and γ are estimated by applying LS to the differenced model, which is optimal under the null hypothesis when $p = 1$. Then Assumption 1(c) guarantees that the regressors $\Delta g_t(\theta)$ in this LS estimation are linearly independent for T large enough. When the value of θ is known there is of course no need to include the infimum in the condition of Assumption 1(c). That, however, is needed when the value of θ is not known and has to be estimated. Since consistent estimation of θ is not possible we have to impose an assumption which guarantees that the above mentioned linear independence of regressors holds whatever the value of θ. This is achieved by Assumption 1(c). Consistent estimation of θ, as well as γ, is not possible because, by Assumption 1(b), the variation of (the differenced) regressors does not increase as $T \to \infty$.

Since $f_t(\theta) = \Delta f_1(\theta) + \cdots + \Delta f_t(\theta)$ it follows from Assumption 1(b) that the sequence $f_t(\theta)$ and hence $g_t(\theta)$ is bounded uniformly in θ and t. Assumption 1(b) also implies that the series in Assumption 1(c) converges uniformly in θ and that the limit is a continuous function of θ. Thus, Assumptions 1(b) and (c) could also be formulated by replacing the finite series by corresponding infinite series. An advantage of the present formulation is that it also applies when the sequence $f_t(\theta)$ and hence $g_t(\theta)$ depends on T. We have not made this feature explicit because it is not needed in the present application of Assumption 1. This dependence on T is obtained, for instance, if asymptotic results are derived under the assumption that T_1/T or $T - T_1$ is constant.

Finally, note that Assumption 1 implies that, for each value of θ, the sequence $g_t(\theta)$ defines a slowly evolving trend if the terminology in Condition B of Elliott, Rothenberg & Stock (1996) is used. Our conditions are stronger than those assumed by these authors, however. Although it might be possible to weaken Assumption 1 we will not pursue this matter because in its present form Assumption 1 is convenient and applies to the previously discussed example models. Overall the model (2.1) and Assumption 1 provide a general (parametric) framework for testing for a unit root in the context of slowly evolving trends.

To illustrate the implications of Assumption 1 it may be helpful to consider what it implies in terms of the example models (2.5) - (2.8). First, for (2.5) the assumption is obviously satisfied. Note that $f_t(\theta)$ in (2.5) actually does not depend on any parameter and, hence, Assumption 1(a) is trivially satisfied here.

Next consider the function $f_t(\theta)$ specified in (2.6). To meet the compactness requirement of Assumption 1(a) we have to assume that $0 < d_- \leq \theta \leq d^- < \infty$. Assuming an upper bound does not appear to be very serious because one can choose d^- such that e^{-d^-} is very close to zero so that, for $\theta \geq d^-$, the sequence $f_t(\theta)$ behaves essentially like the dummy variable d_{1t}. It is also clear that a lower bound condition, $\theta \geq d_-$, is necessary because when the value of θ gets small the slope of the sequence $f_t(\theta)$ decreases and in the limit where $\theta = 0$ we have $f_t(\theta) = 0$ for all t and hence no shift. Obviously, this case has to be excluded. Now consider Assumption 1(b). We have

$$\Delta f_t(\theta) = \begin{cases} 0, & t \leq T_1 \\ \{\exp(\theta) - 1\} \exp\{-\theta(t - T_1)\}, & t > T_1 \end{cases} .$$

From this expression it can be seen that the summability condition of Assumption 1(b) holds while the continuity requirement is obvious. Note that here it is not necessary to restrict the values of θ (except $\theta \geq 0$). As to Assumption 1(c), the above expression of $\Delta f_t(\theta)$ shows that the sum of squares of these variables has a positive limit and, when $0 < d_- \leq \theta$ is assumed, this holds uniformly for all θ. It is similarly clear that $\Delta f_t(\theta)$ and the constant term cannot be (asymptotically) linearly dependent so that Assumption 1(c) holds. A similar discussion also can be given for (2.7).

Finally, consider the function in (2.8). Since unit roots in $\varphi(L)$ are to be avoided the compactness requirement of Assumption 1(a) is met by assuming that the zeros of $\varphi(L)$ are outside the unit circle and are, hence, bounded away from the unit circle, that is, $\varphi(L) \neq 0$ for $|L| \leq 1 + \epsilon$ for some (small) $\epsilon > 0$. This assumption also implies that the summability condition of Assumption 1(b) holds while the continuity condition therein is obviously satisfied. Since the condition of Assumption 1(c) is also straightforward to verify we can conclude that the function in (2.8) fits our general framework.

Given the generality of our shift term, the model (2.1) is quite flexible. For some time series it is still not general enough, however. In particular, if seasonal time series are considered one may want to include seasonal dummy variables in addition to the deterministic parts in (2.1). In this case we may simply use a model

$$y_t = \sum_{i=1}^{q} \mu_i s_{it} + \mu t + g_t(\theta)'\gamma + x_t, \qquad t = 1, 2, \ldots, \tag{2.9}$$

where the μ_i are scalar parameters and the s_{it} $(i = 1, \ldots, q)$ represent seasonal dummy variables. For instance, for quarterly data, s_{it} assumes the value 1 if t is associated with the ith quarter and zero otherwise. For quarterly data we use $q = 3$ seasonal dummies because an intercept term is included in $g_t(\theta)$. For convenience we focus on the model (2.1) in the following theoretical analysis because adding seasonal dummies has no impact on the asymptotic properties of our test statistics but only complicates the notation. Occasionally we will comment on the changes necessary for including seasonal dummies because they are used in the empirical examples in Section 5.

3 Estimation of Nuisance Parameters

In the next section we shall develop a test procedure for the unit root hypothesis $\rho = 1$ in the context of the general model (2.1). This test procedure requires suitable estimators for the nuisance parameters μ, θ and γ. Our approach for estimating these parameters is similar to that in Elliott, Rothenberg & Stock (1996) and Hwang & Schmidt (1996). These authors used GLS estimators of the trend parameters to detrend the observed series. Then the unit root hypothesis is tested on the trend adjusted series. Unlike in the analogous multivariate case considered by Saikkonen & Lütkepohl (1997) our GLS estimation does not necessarily assume validity of the null hypothesis but is based on appropriate local alternatives to be specified by the analyst. Thus, suppose that the error process x_t defined by (2.2) and (2.3) is near integrated so that

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

where $c \leq 0$ is a fixed real number. Then the generating process of x_t can be written as

$$\Delta x_t = \frac{c}{T} x_{t-1} + b(L)^{-1}\varepsilon_t, \qquad t = 1, 2, \ldots \tag{3.2}$$

For simplicity we make the initial value assumption $x_0 = 0$ although our asymptotic results also hold under more general conditions (cf. Elliott et al. (1996), where the implications of initial value assumptions are also discussed). It follows from the stated assumptions that

$$T^{-1/2}x_{[sT]} \xrightarrow{d} \omega B_c(s), \tag{3.3}$$

where $\omega = \sigma/b(1)$ and $B_c(s) = \int_0^s \exp\{c(s-u)\}dB_0(u)$ with $B_0(u)$ a standard Brownian motion (cf. Elliott et al. (1996)).

Our GLS estimation assumes employing an empirical counterpart of the parameter c. This means that we shall replace c by a chosen value $\bar{c}$ and act as if $\bar{c} = c$ would hold. The choice of $\bar{c}$ will be discussed later. Now, if $\bar{\rho}_T = 1 + \frac{\bar{c}}{T}$, the idea is to first transform the variables in (2.1) by the filter $1 - \bar{\rho}_T L$. For convenience we will use matrix notation and define

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

$$Z_1 = [1 : (2 - \bar{\rho}_T) : \cdots : (T - \bar{\rho}_T(T-1))]'$$

and

$$Z_2(\theta) = [g_1(\theta) : (g_2(\theta) - \bar{\rho}_T g_1(\theta)) : \cdots : (g_T(\theta) - \bar{\rho}_T g_{T-1}(\theta))]'.$$

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

$$Y = Z(\theta)\phi + U, \tag{3.4}$$

where $Z(\theta) = [Z_1 : Z_2(\theta)]$, $\phi = [\mu : \gamma']'$ and $U = [u_1 : \cdots : u_T]'$ is an error term such that $u_t = x_t - \bar{\rho}_T x_{t-1}$. It follows from the definitions that

$$u_t = b(L)^{-1}\varepsilon_t + T^{-1}(c - \bar{c})x_{t-1} \stackrel{def}{=} u_t^{(0)} + T^{-1}(c - \bar{c})x_{t-1}. \tag{3.5}$$

The second term on the r.h.s. of this equation is asymptotically negligible because, as a consequence of (3.3), $T^{-1}\max_{1\le t\le T}|x_t| = O_p(T^{-1/2})$. Thus, we shall consider a nonlinear GLS estimation of (3.4) by proceeding in the same way as in the case $c = 0$ or under the null hypothesis. The reason why we still do not assume $\bar{c} = 0$ is that choosing $\bar{c} < 0$ yields more powerful tests (see Elliott et al. (1996)). This means that our GLS estimation is based on the covariance matrix resulting from the first term on the r.h.s. of (3.5). Hence, defining $U^{(0)} = [u_1^{(0)} : \cdots : u_T^{(0)}]'$, we shall consider the covariance matrix of $U^{(0)}$ or, more conveniently, the matrix $\Sigma(b) = \sigma^{-2}\mathrm{Cov}(U^{(0)})$, where $b = [b_1 : \cdots : b_{p-1}]'$. Our GLS estimators are thus obtained by minimizing the generalized sum of squares function

$$Q_T(\phi, \theta, b) = (Y - Z(\theta)\phi)'\Sigma(b)^{-1}(Y - Z(\theta)\phi). \tag{3.6}$$

Note that in this estimation method an 'arbitrary' initial value assumption is only made for x_0 but not for $x_t, t < 0$.

The following technical assumption is helpful when asymptotic properties of the above GLS estimator are studied.

Assumption 2.

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. □

Thus, we restrict the roots of the lag polynomial $b(L)$ in the same way as for the lag polynomial $\varphi(L)$ in (2.8) to meet Assumption 1. Assumption 2 implies that the parameter space for b is compact. It simplifies proofs and is therefore attractive. For this reason similar assumptions have also been quite common in the statistical analysis of stationary ARMA models. Although it is not necessary to specify a value of ϵ a priori in practice, it may be useful to check the location of the roots of the estimate of $b(L)$. If roots very close to the unit circle are found the original model specification may not be appropriate and unit root tests based on it may not be on firm grounds. In particular, if $b(L)$ has a near unit root our null hypothesis means that we have a process which is nearly $I(2)$ and this feature would be useful to take into account in the analysis.

It is shown in the appendix that when Assumptions 1 and 2 hold, GLS estimators obtained by minimizing the function $Q_T(\phi, \theta, b)$ exist for all T large enough. We shall demonstrate here that the same result holds for all values of T provided the matrix $Z(\theta)$ is of full column rank for all $\theta \in \Theta$. First observe that this condition implies that, for any fixed values of θ and b, the (ordinary) GLS estimator of ϕ, denoted by $\hat{\phi}(\theta, b)$, obviously exists. By Assumption 1(b), $Z(\theta)$ is a continuous function of θ while the continuity of $\Sigma(b)$ in b is well-known. This implies that $\hat{\phi}(\theta, b)$ is continuous in (θ, b) and from its definition one obtains, for any values of θ and b,

$$Q_T(\phi, \theta, b) \geq Q_T(\hat{\phi}(\theta, b), \theta, b) \geq \inf_{\theta, b} Q_T(\hat{\phi}(\theta, b), \theta, b). \tag{3.7}$$

The continuity of $\hat{\phi}(\theta, b)$ implies that $Q_T(\hat{\phi}(\theta, b), \theta, b)$ is continuous in (θ, b) so that the infimum in (3.7) is attained at $\theta = \hat{\theta}$ and $b = \hat{b}$, say, if the parameter spaces of θ and b are compact. This, however, follows from Assumptions 1(a) and 2. Thus, $\hat{\phi} = \hat{\phi}(\hat{\theta}, \hat{b})$, $\hat{\theta}$ and

$\hat{b}$ are nonlinear GLS estimators of the parameters ϕ, θ and b, respectively. The additional assumption made about the rank of the matrix $Z(\theta)$ to obtain this result is natural and not restrictive. It is easily seen to hold in the special cases discussed in the previous section. Its asymptotic counterpart is the condition in Assumption 1(c).

The above discussion implies that we can write

$$\hat{\phi} = (Z(\hat{\theta})'\Sigma(\hat{b})^{-1}Z(\hat{\theta}))^{-1}Z(\hat{\theta})'\Sigma(\hat{b})^{-1}Y. \tag{3.8}$$

Of course, the computation of $\hat{\phi}$ still requires iterative methods. However, if preliminary estimators of θ and b are available they can be used on the r.h.s. of (3.8) in place of θ and b, respectively, to yield a feasible GLS estimator of θ. This idea is implicit in some of the procedures to be discussed below.

If $Z(\theta)$ is independent of θ, like in (2.5), the above GLS estimation is simple because we have a linear regression model with AR($p-1$) errors. If computationally simple alternatives are desired one can then also consider conventional two-step estimators or even estimate ϕ by LS. The asymptotic properties of our test procedures are the same even if these estimators are employed. However, in finite samples it may be worthwhile to use proper (nonlinear) GLS estimators which are still very simple.

When $Z(\theta)$ is not independent of θ the situation is more complicated although usually still quite feasible. When the value of θ is fixed we have the situation discussed above so that a grid search over the values of θ may provide a convenient estimation procedure when θ is scalar or possibly even when it is two-dimensional but takes values in a reasonably small set. Since consistent estimation of θ is not possible (see below) and since it may often be sufficient to obtain a relatively rough estimate of a smoothness parameter like the one in (2.6) or (2.7), a fairly coarse grid may suffice. If grid search is not used one can apply one of the available nonlinear estimation algorithms (see, e.g., Judge et al. (1985, Appendix B) or Seber & Wild (1989, Chapters 13 and 14)).

Asymptotic properties of the above nonlinear GLS estimators are described in the following lemma which is proven in the Appendix where also other proofs are given. The estimator $\hat{\phi}$ is partitioned as $\hat{\phi} = [\hat{\mu} : \hat{\gamma}']'$ conformably with the partition of ϕ. The lemma assumes local alternatives specified by (3.1) so that the null hypothesis is obtained by setting $c = 0$.

Lemma 1.

Suppose that Assumptions 1 and 2 hold and also that the matrix $Z(\theta)$ is of full column rank for all $T \geq k+1$ and all $\theta \in \Theta$. Then,

$$\hat{\theta} = \theta + O_p(1), \tag{3.9}$$

$$\hat{\gamma} = \gamma + O_p(1), \tag{3.10}$$

$$\hat{b} \xrightarrow{p} b \tag{3.11}$$

and

$$T^{1/2}(\hat{\mu} - \mu) \xrightarrow{d} \omega \left(\lambda B_c(1) - 3(1-\lambda) \int_0^1 s B_c(s) ds \right), \tag{3.12}$$

where $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$. □

We have included the condition for the rank of the matrix $Z(\theta)$ in Lemma 1 because it is plausible and simplifies the exposition. It is seen in the proof that, as a consequence of Assumption 1(c), this condition always holds for T large enough. Lemma 1 shows that the estimators $\hat{b}$ and $\hat{\mu}$ are consistent but $\hat{\theta}$ and $\hat{\gamma}$ are not. These latter estimators are only bounded in probability. For $\hat{\theta}$ this is, of course, trivial because the parameter space of θ is compact by assumption. However, for $\hat{\gamma}$ the situation is different because the parameter space of γ is totally unrestricted. Since Assumption 1(b) implies that $g_t(\theta) - \bar{\rho}_T\, g_{t-1}(\theta) \approx \Delta g_t(\theta) \to 0$ as $t \to \infty$ the inconsistency of the estimators $\hat{\gamma}$ and $\hat{\theta}$ is expected (for more details, see Seber & Wild (1989, p. 565/566) and Wu (1981)). The limiting distribution obtained for the estimator $\hat{\mu}$ in (3.12) agrees with that obtained by Elliott et al. (1996) in a model with $g_t(\theta) = 1$.

The following example may be helpful for seeing more clearly how the procedure works and why, for instance, $\hat{\gamma}$ is not consistent in general. Consider the function in (2.5) which implies a $g_t(\theta)$ independent of θ and $\gamma = (\gamma_1, \gamma_2)'$ is just the coefficient vector associated with the constant and the step dummy d_{1t}. In this case

$$Z(\theta) = \begin{bmatrix} 1 & 1 & 0 \\ 2-\bar{\rho}_T & 1-\bar{\rho}_T & 0 \\ \vdots & \vdots & \vdots \\ T_1 - \bar{\rho}_T(T_1-1) & 1-\bar{\rho}_T & 1 \\ T_1+1-\bar{\rho}_T T_1 & 1-\bar{\rho}_T & 1-\bar{\rho}_T \\ \vdots & \vdots & \vdots \\ T-\bar{\rho}_T(T-1) & 1-\bar{\rho}_T & 1-\bar{\rho}_T \end{bmatrix}$$

and computing estimators is very easy for $p = 1$. For higher order processes an iterated GLS method may be used, for instance, where μ and γ are first estimated by LS from (3.4). Then an estimator for b is determined from the residuals again by LS. This estimator is used in setting up $\Sigma(\hat{b})$ and in obtaining second round estimators of γ by replacing $\Sigma(b)$ in (3.6) by $\Sigma(\hat{b})$. The procedure may be repeated until convergence or it may be stopped after a small number of iterations. Since γ_1 is estimated separately from the first T_1 observations only, it is clear that the estimator does not improve if T_1 is fixed and T increases. Note that from observation $T_1 + 1$ onwards the sample contains information on the sum $\gamma_1 + \gamma_2$ only and not on γ_1 and γ_2 separately.

We close this section by noting that the case where the model does not contain a linear trend term can be handled in a straightforward way. Then the trend is simply dropped from (2.1) and the above estimation procedure is modified accordingly. The results in Lemma 1 for b, θ and γ continue to hold in this case, as the derivations in the appendix show. A similar comment applies if seasonal dummy variables are added to the model. In that case appropriate columns for the seasonal dummies have to be added to the matrix Z. Clearly, the associated parameter estimates are consistent. It is argued in the Appendix that including seasonal dummies has no impact on the asymptotic properties of the other estimators.

4 Testing Procedures

Once the nuisance parameters in (2.1) have been estimated one can form the residual series $\hat{x}_t = y_t - \hat{\mu}t - g_t(\hat{\theta})'\hat{\gamma}$ and use it to obtain unit root tests. There are several possibilities in this respect. For instance, Elliott et al. (1996) consider Dickey-Fuller (DF) tests. We shall only give a detailed discussion of one approach and briefly mention some other possibilities.

Consider the auxiliary regression model

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

where $\hat{x}_0 = 0$. In the previous section it was seen that if $\hat{x}_t$ is replaced by x_t the covariance matrix of the error term in (4.1) is $\sigma^2 \Sigma(b)$. Since the parameter b is estimated to obtain $\hat{x}_t$ it seems reasonable to use this estimator also here and base a unit root test on (4.1) with ρ estimated by feasible GLS with weight matrix $\Sigma(\hat{b})^{-1}$. Thus, if $\hat{X} = [\hat{x}_1 : \cdots : \hat{x}_T]'$ and $\hat{X}_{-1} = [0 : \hat{x}_1 : \cdots : \hat{x}_{T-1}]'$ we introduce the estimator

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

We also need an estimator of the error variance σ^2. Based on the GLS estimation of (4.1) we use

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

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

$$\tau = (\hat{X}'_{-1} \Sigma(\hat{b})^{-1} \hat{X}_{-1})^{1/2} (\hat{\rho} - 1)/\hat{\sigma}. \tag{4.4}$$

The limiting distribution of this test statistic is given in the following theorem which again assumes the local alternatives defined in (3.1).

Theorem 1.

Suppose the assumptions of Lemma 1 hold. Then,

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

where

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

□

The limiting distribution in Theorem 1 is the same which Elliott et al. (1996) obtained for their t-statistic in a model whose deterministic part only contained a mean value and linear trend term. The limiting null distribution, obtained by setting $c = 0$, is free of unknown nuisance parameters but depends on the quantity $\bar{c}$. Elliott at al. (1996) suggest using $\bar{c} = -13.5$ and give some critical values for this choice in their Table I.C (see their paper for a motivation of this choice and further discussion). Since our alternative is $I(0)$, small values of τ are critical. Elliott et al. (1996) show that with the above choice of $\bar{c}$ the asymptotic local power of their t-test is nearly optimal for all values of c. From their results and Theorem 1 we can conclude that this is also the case for our test. Hence, substantial gains in local power may be possible relative to other tests.

It may be worth noting that to avoid the initial value assumption $\hat{x}_0 = 0$ one could consider (4.1) for $t = 2, \ldots, T$ and modify $\hat{X}$, $\hat{X}_{-1}$ and $\Sigma(\hat{b})$ accordingly. The given formulation has been used to avoid redefining $\Sigma(\hat{b})$.

In the same way as in Elliott et al. (1996) we could 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 their DF t-statistic appeared somewhat better than those of the point optimal tests. Their DF t-statistic is not similar to our τ but is based on a regression of $\hat{x}_t$ on $\hat{x}_{t-1}, \ldots, \hat{x}_{t-p}$, $t = p+1, \ldots, T$. This approach could also be used here to obtain a test statistic with the same limiting distribution as τ.

Finally, note that if we have the a priori restriction $\mu = 0$ the above test remains the same except that in this case $\bar{c} = -7$ is recommended and the limiting null distribution is then the same as in an AR(p) model without any deterministic terms. Power gains can be considerable compared to tests whose properties depend on deterministic terms as in Elliott et al. (1996). It may also be worth noting that seasonal dummies may be included without affecting the limiting distribution of our test statistic as is shown in the Appendix.

5 Examples

To illustrate the use of the tests presented in the foregoing we consider three German time series with obvious shifts at the time of the German reunification. In particular, we will investigate the unit root properties of quarterly real GNP (1975(1) - 1996(4)), money stock M1 (1960(1) - 1997(1)) and M3 (1972(1) - 1996(4)). None of the series is seasonally adjusted.*

*Data sources: GNP – quarterly, seasonally unadjusted data, 1975(1) - 1990(2) West Germany, 1990(3) - 1996(4) all of Germany, Deutsches Institut für Wirtschaftsforschung, Volkswirtschaftliche Gesamtrechnung.

The logarithms of the three variables are plotted in Figures 1 - 3 together with some other functions and series which will be discussed later. In the figures it is seen that the three series all have seasonal patterns and clear shifts in 1990 where the German unification occurred.† Seasonal dummies are included in the models to take care of the seasonal components and the shifts in 1990 are dealt with by including a shift dummy as in (2.5) or alternatively by using the transition functions in (2.6) and (2.8) with $q = r = 1$. Thus, we consider the following 3 versions of the shift function $f_t(\theta)$:

$$f_t^{(1)}(\theta) = d_{1t}, \quad f_t^{(2)}(\theta) = \psi_t(\theta) \quad \text{and} \quad f_t^{(3)}(\theta) = \left[\frac{d_{1,t}}{1-\theta L} : \frac{d_{1,t-1}}{1-\theta L}\right]'.$$

A smooth transition to a new level is at least a possibility for the series under consideration because the East German economy entered into a transition process which changed the economy in a fundamental way. Since $f_t^{(2)}(\theta)$ and $f_t^{(3)}(\theta)$ contain a single parameter only, estimation of θ is done by nonlinear GLS with a grid search over the relevant part of the space of θ.

For comparison purposes we also performed regular augmented Dickey-Fuller (ADF) tests with a linear trend. Perron (1989) showed that these tests may have low power if there is a level shift in the time series considered. The results of all the tests are given in Table 1 together with critical values. The lag lengths are chosen such that residual autocorrelation is largely eliminated, that is, models with increasing lag lengths were fitted until the residual autocorrelation was insignificant. The orders used in the tests are also shown in Table 1.‡ We will now discuss the test results in detail in conjunction with the estimation results for the shift functions.

In addition to the graphs of the series the estimated shift functions and the series adjusted for deterministic terms are also depicted in the figures. In particular, $\hat{x}_t^{(i)}$ denotes the adjusted series obtained by subtracting the intercept, seasonal dummies, the trend and the shift function based on $f_t^{(i)}(\hat{\theta})$, $i = 0, 1, 2, 3$, where $f_t^{(0)} \equiv 0$, that is, the shift is ignored for

---

M1 – quarterly, seasonally unadjusted data, 1960(1) - 1990(3) West Germany, 1990(4) - 1997(1) all of Germany, OECD.

M3 – quarterly, seasonally unadjusted data, 1972(1) - 1990(2) West Germany, 1990(3) - 1996(4) all of Germany, Monatsbericht der Deutschen Bundesbank.

†Due to the specific definitions of the data the shift occurs in the third quarter in GNP and M3 and in the fourth quarter of 1990 in M1.

‡Using AR order 4 in all tests we obtained qualitatively similar results.

Table 1. Unit Root Tests

Variable	AR order	ADF test: value of test statistic	ADF test: critical values* 5%	ADF test: critical values* 10%	τ test: $f_t^{(1)}(\theta)$	τ test: $f_t^{(2)}(\theta)$	τ test: $f_t^{(3)}(\theta)$	τ test: critical values** 5%	τ test: critical values** 10%
log GNP	5	−2.52	−3.46	−3.16	−1.80	−1.80	−2.19		
log M1	6	−1.82	−3.44	−3.15	−2.61	−2.36	−2.43	−2.89	−2.57
log M3	6	−2.18	−3.46	−3.15	−0.80	−0.80	−1.15		

* Source: MacKinnon (1991). ** Source: Elliott et al. (1996, Table I.C, $T = \infty$).

$i = 0$. For log GNP the estimated shifts based on $f_t^{(1)}$ and $f_t^{(2)}(\hat{\theta})$ are similar which is also reflected in the adjusted $\hat{x}_t^{(i)}$ $(i = 1, 2)$. Whereas $\hat{x}_t^{(0)}$ has a clear shift in 1990 this is not the case for $\hat{x}_t^{(i)}$ $(i = 1, 2)$. The shift based on $f_t^{(3)}(\hat{\theta})$ is quite different from the previous ones. After a steep increase in 1990 it declines towards zero and, hence, the shift slowly disappears. For German GNP this outcome is quite plausible assuming that the situation in all of Germany slowly approaches the preunification situation in West Germany. The shift functions based on $f_t^{(1)}$ and $f_t^{(2)}(\hat{\theta})$ cannot reflect this kind of behavior because they are not sufficiently flexible. Thus, in this case for modeling the shift allowing for some flexibility may be advantageous. The adjusted series $\hat{x}_t^{(3)}$ also does not display a clear shift in 1990 and, hence, the shift my be captured adequately by $f_t^{(3)}(\hat{\theta})$ as well. Despite the shift in the series and despite the differences in capturing the shift the ADF and τ tests all reach the same conclusions. They do not reject a unit root in log GNP. Thus the tests confirm that the choice of shift function is not critical in this case.

Looking at Figure 2, the situation is seen to be a bit different for log M1. In this case a step dummy ($f_t^{(1)}$) results in a smaller shift than the other two shift functions. For $f_t^{(2)}(\hat{\theta})$ and $f_t^{(3)}(\hat{\theta})$ also quite steep shifts are obtained with a short adjustment period. At a 5% significance level all tests indicate a unit root in log M1 (see Table 1). However, the value of the τ test corresponding to $f_t^{(1)}$ is significant at the 10% level. Hence, in this case, not being able to reject the unit root hypothesis may just be a reflection of insufficient power of unit root tests in the presence of a shift in the deterministic component. Alternatively, the step dummy may be too restrictive in this case to capture the actual shift in the series and, hence, the tests based on $f_t^{(2)}(\hat{\theta})$ and $f_t^{(3)}(\hat{\theta})$ may be more reliable. In any case the evidence against a unit root in log M1 is not very strong.

The estimated shift functions for log M3 are displayed in Figure 3. For $f_t^{(1)}$ and $f_t^{(2)}(\hat{\theta})$ a one-time shift of very similar size is obtained. As for log GNP the shift based on $f_t^{(3)}(\hat{\theta})$ is quite different. After the jump in 1990 it slowly tends back towards zero. Again, this kind of shift is not unreasonable if there is a transition towards the preunification situation in West Germany. Despite the differences in the shift functions the test results are again robust and unanimously point to a unit root in log M3. Thus, overall our results confirm unit roots in log GNP and log M3 even if deterministic shifts are allowed for whereas the evidence for a unit root in log M1 is less clear in this case.

6 Conclusions

In this study we have proposed new tests for unit roots in univariate time series with a shift in the mean. The timing of the shift is assumed to be known and the form of the shift may be of a very general type ranging from a simple one-time step to a longer term smooth adjustment to a new level. Also there may be more than one shift and there may be further deterministic terms such as a linear trend and seasonal components. It is proposed to estimate the deterministic part of the series first by a GLS procedure. The estimated deterministic part is then subtracted from the original series and a unit root test is performed on the residual series. Although there are various different tests that can be used in the second step of the procedure we have focused on Dickey-Fuller type tests as proposed by Elliott et al. (1996). The asymptotic distribution under the null of a unit root is nonstandard but critical values are available in the literature. We have illustrated the tests using German macroeconomic time series which have a level shift in 1990 where the German reunification occurred.

Appendix. Proofs

A.1 Proof of Lemma 1

Using the definitions of the previous sections we first observe that

$$Z_1 = \begin{bmatrix} 1 \\ 1-\frac{\bar{c}}{T} \\ \vdots \\ 1-\frac{\bar{c}(T-1)}{T} \end{bmatrix} \quad \text{and} \quad Z_2(\theta) = \begin{bmatrix} g_1(\theta)' \\ \Delta g_2(\theta)' - \frac{\bar{c}}{T} g_1(\theta)' \\ \vdots \\ \Delta g_T(\theta)' - \frac{\bar{c}}{T} g_{T-1}(\theta)' \end{bmatrix}.$$

From this expression of Z_1 it is straightforward to check that

$$\begin{aligned} T^{-1} Z_1' Z_1 &= 1 - \bar{c} + \tfrac{\bar{c}^2}{3} + O(T^{-1}) \\ &\overset{def}{=} h(\bar{c}) + O(T^{-1}). \end{aligned} \tag{A.1}$$

Recall from Section 2 that the sequence $g_t(\theta)$ is bounded uniformly over θ and t. Thus, using the above expression of $Z_2(\theta)$ and Assumption 1(b) we find that

$$T^{-1/2} Z_1' Z_2(\theta) = O(T^{-1/2}) \tag{A.2}$$

and

$$Z_2(\theta)' Z_2(\theta) = \sum_{t=1}^{T} \Delta g_t(\theta) \Delta g_t(\theta)' + O(T^{-1}) \tag{A.3}$$

uniformly in θ. Combining $(A.1) - (A.3)$ and denoting $D_{1T} = \text{diag}[T^{1/2} : I_k]$ yields

$$\begin{aligned} D_{1T}^{-1} Z(\theta)' Z(\theta) D_{1T}^{-1} &= \text{diag}\left[h(\bar{c}) : \textstyle\sum_{t=1}^{T} \Delta g_t(\theta) \Delta g_t(\theta)'\right] + O(T^{-1/2}) \\ &\overset{def}{=} M_T(\theta) + O(T^{-1/2}) \end{aligned} \tag{A.4}$$

uniformly in θ. We note in passing that $(A.4)$ implies that the matrix $Z(\theta)$ is of full column rank for all θ and all T large enough because, by Assumption 1(c), the matrix $M_T(\theta)$ is positive definite for all θ and all T large enough.

Next note that, by Assumption 2, the spectral density function of the stationary process $u_t^{(0)} = b(L)^{-1}\varepsilon_t$ is bounded and bounded away from zero uniformly over the permissible space of b. This implies that there exist numbers $\underline{K}$ and $\bar{K}$ such that

$$0 < \underline{K} \le \lambda_{min}\left(\Sigma(b)\right) \le \lambda_{max}\left(\Sigma(b)\right) \le \bar{K} < \infty \tag{A.5}$$

(cf. Elliott et al. (1996), proof of Lemma A.1). From ($A.4$), ($A.5$) and the continuity of eigenvalues we thus find that

$$\begin{aligned}\lambda_{min}\left(D_{1T}^{-1}Z(\hat{\theta})'\Sigma(\hat{b})^{-1}Z(\hat{\theta})D_{1T}^{-1}\right) &\geq \bar{K}^{-1}\lambda_{min}\left(D_{1T}^{-1}Z(\hat{\theta})'Z(\hat{\theta})D_{1T}^{-1}\right)\\ &= \bar{K}^{-1}\lambda_{min}(M_T(\hat{\theta}))+o(1).\end{aligned}$$

Since $\lambda_{min}(M_T(\hat{\theta}))\geq\epsilon>0$ for $T\geq T_*$ by Assumption 1(c) it follows from the above that

$$\left\|\left(D_{1T}^{-1}Z(\hat{\theta})'\Sigma(\hat{b})^{-1}Z(\hat{\theta})D_{1T}^{-1}\right)^{-1}\right\|_1=O(1), \qquad (A.6)$$

where $\|\cdot\|_1$ signifies the operator norm of a matrix.

Next note that

$$Y=Z(\hat{\theta})\phi+\xi,$$

where $\xi=U+(Z_2(\theta)-Z_2(\hat{\theta}))\gamma$ with U as in (3.4). Note also that ξ is not a function of the parameters θ and γ because here θ and γ signify true parameter values. From this and (3.8) one obtains

$$\hat{\phi}-\phi=\left(Z(\hat{\theta})'\Sigma(\hat{b})^{-1}Z(\hat{\theta})\right)^{-1}Z(\hat{\theta})'\Sigma(\hat{b})^{-1}\xi$$

which in conjunction with ($A.6$) and the norm inequality $\|AB\|\leq\|A\|_1\|B\|$ implies

$$\|D_{1T}(\hat{\phi}-\phi)\|\leq O(1)\left\|D_{1T}^{-1}Z(\hat{\theta})'\Sigma(\hat{b})^{-1}\xi\right\|.$$

Hence, if we show that

$$D_{1T}^{-1}Z(\hat{\theta})'\Sigma(\hat{b})^{-1}\xi=O_p(1), \qquad (A.7)$$

we can conclude

$$D_{1T}(\hat{\phi}-\phi)=O_p(1) \qquad (A.8)$$

which proves (3.10).

To justify ($A.7$), let $Z_{(p-1)}(\hat{\theta})$ be the $((p-1)\times(k+1))$ matrix containing the first $p-1$ rows of $Z(\hat{\theta})$ and let $\xi_{(p-1)}$ be the $p-1$ vector containing the first $p-1$ components of ξ. Furthermore, let $\Sigma_{(p-1)}(\hat{b})$ be the $((p-1)\times(p-1))$ dimensional counterpart of $\Sigma(\hat{b})$ and define $\hat{b}(L)=1-\hat{b}_1L-\cdots-\hat{b}_{p-1}L^{p-1}$. Then we can write

$$D_{1T}^{-1}Z(\hat{\theta})'\Sigma(\hat{b})^{-1}\xi=D_{1T}^{-1}Z_{(p-1)}(\hat{\theta})'\Sigma_{(p-1)}(\hat{b})^{-1}\xi_{(p-1)}+D_{1T}^{-1}\sum_{t=p}^{T}[\hat{b}(L)Z_t(\hat{\theta})][\hat{b}(L)\xi_t], \qquad (A.9)$$

where $Z_t(\hat{\theta})$ $((k+1)\times 1)$ is the tth row of the matrix $Z(\hat{\theta})$ and ξ_t is the tth component of the vector ξ. By the definitions and our previous derivations it is clear that the first term on the right hand side of $(A.9)$ is of order $O_p(1)$ and its first component is actually of order $O_p(T^{-1/2})$.

To analyse the second term on the right hand side of $(A.9)$, let ζ_t denote the tth component of the vector $(Z_2(\theta) - Z_2(\hat{\theta}))\gamma$ so that $\xi_t = u_t + \zeta_t$ with u_t as in (3.5). It follows from Assumption 1(b) that the sequence ζ_t is absolutely summable, while Assumption 2 implies that the coefficients of the polynomial $b(L)$ belong to a bounded set. Thus, using these facts, the expressions of Z_1 and $Z_2(\theta)$ given at the beginning of the proof, the definition of $Z_t(\hat{\theta})$, and Assumption 1(b) we find that

$$D_{1T}^{-1}\sum_{t=p}^{T}[\hat{b}(L)Z_t(\hat{\theta})][\hat{b}(L)\zeta_t] = \begin{bmatrix} O_p(T^{-1/2}) \\ O_p(1) \end{bmatrix}, \tag{$A.10$}$$

where the partition is after the first component. Now we can conclude that the second term on the right hand side of $(A.9)$ is of order $O_p(1)$ if

$$D_{1T}^{-1}\sum_{t=p}^{T}[\hat{b}(L)Z_t(\hat{\theta})][\hat{b}(L)u_t] = O_p(1) \tag{$A.11$}$$

or if

$$T^{-1/2}\sum_{t=p}^{T} Z_{1,t-i}u_{t-j} = O_p(1) \tag{$A.12$}$$

and

$$\sum_{t=p}^{T} Z_{2,t-i}(\hat{\theta})u_{t-j} = O_p(1), \tag{$A.13$}$$

where $i, j = 1, \ldots, p-1$ and the partition $Z_t(\hat{\theta}) = [Z_{1t} : Z_{2t}(\hat{\theta})']'$ has been used. Since $Z_{1t} = 1 - \frac{\bar{c}(t-1)}{T}$ it follows from (3.5) and well-known properties of stationary and near integrated processes that $(A.12)$ holds. To justify $(A.13)$, recall from (3.5) that $u_t = u_t^{(0)} + (c-\bar{c})T^{-1}x_{t-1}$ with $u_t^{(0)} = b(L)^{-1}\varepsilon_t$ and notice that the left hand side of $(A.13)$ is dominated by

$$\sum_{t=p}^{T}\sup_{\theta\in\Theta}\|Z_{2,t-i}(\theta)\||u_{t-j}^{(0)}| + |c-\bar{c}|T^{-1}\max_{1\leq t\leq T}|x_t|\sum_{t=p}^{T}\sup_{\theta\in\Theta}\|Z_{2,t-i}(\theta)\|.$$

Since $Z_{2t}(\theta) = \Delta g_t(\theta) - T^{-1}\bar{c}g_{t-1}(\theta)$ it follows from Assumption 1(b) that $\sup_{\theta\in\Theta}\|Z_{2t}(\theta)\|$ is summable. Thus, the first term in the last expression is of order $O_p(1)$ because $E|u_{t-j}^{(0)}|$ is a finite constant while the second term is of order $O_p(T^{-1/2})$ because $T^{-1}\max_{1\leq t\leq T}|x_t| = O_p(T^{-1/2})$, as noticed below (2.5). Hence, we can conclude that $(A.13)$ also holds and, furthermore, that the second term on the right hand side of $(A.9)$ is of order $O_p(1)$. As a whole, we have thus established $(A.7)$ and thereby $(A.8)$ as well. As already noticed, this proves (3.10) while (3.9) holds by the assumed compactness of the parameter space Θ. To complete the proof, we still need to show (3.11) and (3.12).

To prove (3.11), that is the consistency of $\hat{b}$, it will be useful to let b_0, θ_0 and $\phi_0 = [\mu_0 : \gamma_0']'$ stand for the true values of the indicated parameters. We also introduce the notation

$$\begin{aligned} r(\theta, \phi) &= Z(\theta)\phi - Z(\theta_0)\phi_0 \\ &= Z_1(\mu - \mu_0) + Z_2(\theta)\gamma - Z_2(\theta_0)\gamma_0. \end{aligned}$$

Thus, since $U = Y - Z(\theta_0)\phi_0$, we have $Y - Z(\theta)\phi = U - r(\theta, \phi)$ and furthermore

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

In the same way as in $(A.9)$ we can write

$$T^{-1}Q_{1T}(b) = T^{-1}U_{(p-1)}'\Sigma_{(p-1)}(b)^{-1}U_{(p-1)} + T^{-1}\sum_{t=p}^{T}[b(L)u_t]^2, \qquad (A.14)$$

where the vector $U_{(p-1)}$ contains the first $p-1$ components of U. Using $(A.5)$, (3.5), the fact that the coefficients of $b(L)$ belong to a bounded set, and well-known properties of stationary and near integrated processes we can conclude from $(A.14)$ that

$$T^{-1}Q_{1T}(b) \xrightarrow{p} \bar{Q}_1(b), \qquad (A.15)$$

where the convergence is uniform in b and the right hand side equals the variance of the stationary process $b(L)b_0(L)^{-1}\varepsilon_t$ with $b_0(L)$ defined in terms of b_0. It is also well-known that $\bar{Q}_1(b)$ is continuous and that $\bar{Q}_1(b) \geq \bar{Q}_1(b_0)$ with equality if and only if $b = b_0$.

It will be shown later that

$$T^{-1}Q_{iT}(\hat{\phi}, \hat{\theta}, \hat{b}) = o_p(1), \quad i = 2, 3. \qquad (A.16)$$

Assuming this for the moment, one obtains

$$\begin{aligned} T^{-1}Q_T(\phi_0, \theta_0, b_0) &\geq T^{-1}Q_T(\hat{\phi}, \hat{\theta}, \hat{b}) \\ &= T^{-1}Q_{1T}(\hat{b}) + o_p(1), \end{aligned}$$

where the first relation is based on the definitions of the estimators $\hat{\phi}_T, \hat{\theta}_T$ and $\hat{b}_T$ and the second one on $(A.16)$. Since $Q_T(\phi_0, \theta_0, b_0) = Q_{1T}(b_0)$ the above inequality and $(A.15)$ give $\bar{Q}_1(\hat{b}) - \bar{Q}_1(b_0) \leq o_p(1)$ and, since $\bar{Q}_1(b)$ is uniquely minimized at $b = b_0$, the consistency of $\hat{b}$ follows.

To show $(A.16)$ consider the case $i = 3$. Since $(A.8)$ implies $\hat{\mu} - \mu_0 = O_p(T^{-1/2})$ we have by using first $(A.5)$ and then $(A.1)$,

$$(\hat{\mu} - \mu_0)^2 Z_1' \Sigma(\hat{b})^{-1} Z_1 \leq \underline{K}^{-1} (\hat{\mu} - \mu_0)^2 \|Z_1\|^2 = O_p(1).$$

Similarly,

$$\begin{aligned}
&(Z_2(\hat{\theta})\hat{\gamma} - Z_2(\theta_0)\gamma_0)' \Sigma(\hat{b})^{-1} (Z_2(\hat{\theta})\hat{\gamma} - Z_2(\theta_0)\gamma_0) \\
&\leq \underline{K}^{-1} \|Z_2(\hat{\theta})\hat{\gamma} - Z_2(\theta_0)\gamma_0\|^2 \\
&\leq 2\underline{K}^{-1} \|Z_2(\hat{\theta})\|^2 \|\hat{\gamma}\|^2 + 2\underline{K}^{-1} \|Z_2(\theta_0)\|^2 \|\gamma_0\|^2 \\
&= O_p(1),
\end{aligned}$$

where the equality is a straightforward consequence of $(A.3)$, Assumption 1(b) and the result $\hat{\gamma} = O_p(1)$ obtained from $(A.8)$. Now, to see that $(A.16)$ holds for $i = 3$, recall the definition of $Q_{3T}(\phi, \theta, b)$ and use the latter expression of $r(\theta, \phi)$ in conjunction with the above results and the Cauchy-Schwarz inequality. Moreover, it can be deduced from this, the fact that $T^{-1} Q_{1T}(\hat{b}) = O_p(1)$ obtained from $(A.15)$ and the Cauchy-Schwarz inequality that $(A.16)$ holds for $i = 2$. Thus, we have proved the consistency of $\hat{b}$.

Finally, we have to demonstrate (3.12). Treating $\Sigma(\hat{b})^{-1}$ in the same way as in $(A.9)$ and $(A.14)$ and using arguments similar to those for $(A.2)$ it can be seen that

$$T^{-1/2} Z_1' \Sigma(\hat{b})^{-1} Z_2(\hat{\theta}) = O_p(T^{-1/2}).$$

Hence,

$$D_{1T}^{-1} Z(\hat{\theta})' \Sigma(\hat{b})^{-1} Z(\hat{\theta}) D_{1T}^{-1} = \text{diag}[T^{-1} Z_1' \Sigma(\hat{b})^{-1} Z_1 : Z_2(\hat{\theta})' \Sigma(\hat{b})^{-1} Z_2(\hat{\theta})] + O_p(T^{-1/2}).$$

From this, $(A.6)$, and Lemma A.2 of Saikkonen & Lütkepohl (1996) it follows that a similar equality also holds for the corresponding inverses, which together with $(A.7)$ implies

$$\begin{aligned}
T^{1/2}(\hat{\mu} - \mu) &= (T^{-1} Z_1' \Sigma(\hat{b})^{-1} Z_1)^{-1} T^{-1/2} Z_1' \Sigma(\hat{b})^{-1} \xi + o_p(1) \\
&= (T^{-1} Z_1' \Sigma(\hat{b})^{-1} Z_1)^{-1} T^{-1/2} Z_1' \Sigma(\hat{b})^{-1} U + o_p(1).
\end{aligned}$$

Here the latter equality follows from the analysis given for $(A.9)$ (see in particular $(A.10)$) and the fact that the inverse is bounded by $(A.6)$. In the last expression we can treat the inverse in the same way as in $(A.9)$ and $(A.14)$, use the consistency of the estimator $\hat{b}$ and arguments used earlier in the proof to conclude that $\hat{b}$ can be replaced by the true parameter value. The arguments given in the proof of Lemma A.4 of Elliott et al. (1996) then imply that $\Sigma(\hat{b})^{-1}$ can further be replaced by $\omega^{-2}I_T$ and that the limiting distribution of $\hat{\mu}$ is the same as stated on p. 835 of that paper. This completes the proof of Lemma 1.

To see how seasonal dummies affect the result of Lemma 1, let Z_3 be the matrix containing the values of the seasonal dummies corresponding to $y_1, \ldots, y_T$ transformed by the filter $1 - \bar{\rho}_T L$. Assume that the seasonal dummies are linearly independent and also that the constant term is linearly independent of the seasonal dummies. Then $T^{-1}Z_3'\Sigma(b)^{-1}Z_3$ converges to a positive definite limit while $T^{-1}Z_3'\Sigma(b)^{-1}Z_1 = o(1)$ and $T^{-1/2}Z_3'\Sigma(b)^{-1}Z_2(\theta) = o(1)$ uniformly in b and θ. These last facts can be established by using arguments similar to those in the proof of Lemma 1. Since the argument used in $(A.9)$ can also be used to show that $T^{-1/2}Z_3'\Sigma(b)^{-1}\xi = O_p(1)$ uniformly in b it follows that the estimation of the coefficients of the seasonal dummies is asymptotically orthogonal to the estimation of other regression coefficients so that the coefficient estimators related to the seasonal dummies are consistent and the results of Lemma 1 still hold in the stated form.

A.2 Proof of Theorem 1

First observe that

$$\hat{x}_t = x_t - (\hat{\mu} - \mu)t - g_t(\hat{\theta})'\hat{\gamma} + g_t(\theta)'\gamma. \tag{A.17}$$

Since $g_t(\theta)$ is bounded uniformly over θ and t, it follows from this, (3.3) and Lemma 1 that

$$T^{-1/2}\hat{x}_{[Ts]} \xrightarrow{d} \omega G_c(s;\bar{c}). \tag{A.18}$$

We also note that from $(A.17)$, Lemma 1 and Assumption 1 it is straightforward to conclude that, for $i, j = 0, \ldots, p-1$,

$$\begin{aligned} T^{-1}\textstyle\sum_{t=p}^{T}\Delta\hat{x}_{t-i}\Delta\hat{x}_{t-j} &= T^{-1}\textstyle\sum_{t=p}^{T}\Delta x_{t-i}\Delta x_{t-j} + o_p(1) \\ &= T^{-1}\textstyle\sum_{t=p}^{T} u_{t-i}^{(0)}u_{t-j}^{(0)} + o_p(1). \end{aligned} \tag{A.19}$$

Now, treating the inverse $\Sigma(\hat{b})^{-1}$ in the same way as in $(A.9)$ and $(A.14)$ we find that

$$\begin{aligned} T^{-2}\hat{X}_{-1}'\Sigma(\hat{b})^{-1}\hat{X}_{-1} &= T^{-2}\textstyle\sum_{t=p}^{T}[\hat{b}(L)\hat{x}_{t-1}]^2 + o_p(1) \\ &= \hat{b}(1)^2T^{-2}\textstyle\sum_{t=p}^{T}\hat{x}_{t-1}^2 + o_p(1) \\ &\xrightarrow{d} \sigma^2\textstyle\int_0^1 G_c(s;\bar{c})^2ds. \end{aligned} \tag{A.20}$$

Here the second equality is a simple consequence of $(A.18)$, $(A.19)$ and the representation $\hat{b}(L) = \hat{b}(1) + \hat{b}_*(L)\Delta$. The last relation follows from the consistency of $\hat{b}$, $(A.18)$ and the continuous mapping theorem. In the same way we also have

$$\begin{aligned} T^{-1}\hat{X}'_{-1}\Sigma(\hat{b})^{-1}(\hat{X}-\hat{X}_{-1}) &= T^{-1}\textstyle\sum_{t=p}^{T}[\hat{b}(L)\hat{x}_{t-1}][\hat{b}(L)\Delta\hat{x}_t]+o_p(1) \\ &= \tfrac{1}{2}T^{-1}[\hat{b}(L)\hat{x}_t]^2-\tfrac{1}{2}T^{-1}\textstyle\sum_{t=p}^{T}[\hat{b}(L)\Delta\hat{x}_t]^2+o_p(1) \\ &\xrightarrow{d} \tfrac{1}{2}\sigma^2 G_c(1;\bar{c})^2-\tfrac{1}{2}\sigma^2. \end{aligned} \qquad (A.21)$$

Here the second relation is due to a simple algebraic identity (cf. Phillips (1987), Equations ($A2$) and ($A3$)), while the third one can be obtained from ($A.18$), ($A.19$) and arguments used above. These arguments in conjunction with the result $\hat{\rho}=1+O_p(T^{-1})$ obtained from ($A.20$) and ($A.21$) also imply that $\hat{\sigma}^2=\sigma^2+o_p(1)$. The stated result follows from this fact, ($A.20$), ($A.21$) and the continuous mapping theorem. Thereby the proof is complete.

Now suppose that seasonal dummies are included in the model. Then, according to what was said above about parameter estimation in this context it is clear that the counterpart of the residual series $\hat{x}_t$ obtained in this case still satisfies ($A.18$) and, furthermore, that the resulting test statistic has the same limiting distribution as in the model where no seasonal dummies are included.

References

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

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

Box, G.E.P. & G.C. Tiao (1975), Intervention analysis with applications to economic and environmental problems, *Journal of the American Statistical Association*, 70, 70 - 79.

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

Franses, P.H. & N. Haldrup (1994), The effects of additive outliers on tests for unit roots and cointegration, *Journal of Business & Economic Statistics*, 12, 471 - 478.

Ghysels, E. & P. Perron (1996), The effects of linear filters on dynamic time series with structural change, *Journal of Econometrics*, 70, 69 - 97.

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

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

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl & T.C. Lee (1985), *The Theory and Practice of Econometrics*, Second Edition, New York: John Wiley.

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. (1996), *Handbook of Matrices*, Chichester: John Wiley.

MacKinnon, J.G. (1991), Critical values for cointegration tests, in R.F. Engle & C.W.J. Granger (eds.), *Long-run Economic Relationships*, Oxford: Oxford University Press, 267 - 276.

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

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

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

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

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

Rappoport, P. & L. Reichlin (1989), Segmented trends and non-stationary time series, *Economic Journal*, 99, 168 - 177.

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

Saikkonen, P. & H. Lütkepohl (1997), Trend adjustment prior to testing for the cointegrating rank of a VAR process, Discussion Paper 84, SFB 373, Humboldt-Universität zu Berlin.

Seber, G.A.F. & C.J. Wild (1989), *Nonlinear Regression*, New York: John Wiley.

Wu (1981), Asymptotic theory of nonlinear least squares estimation, *Annals of Statistics*, 9, 501 - 513.

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

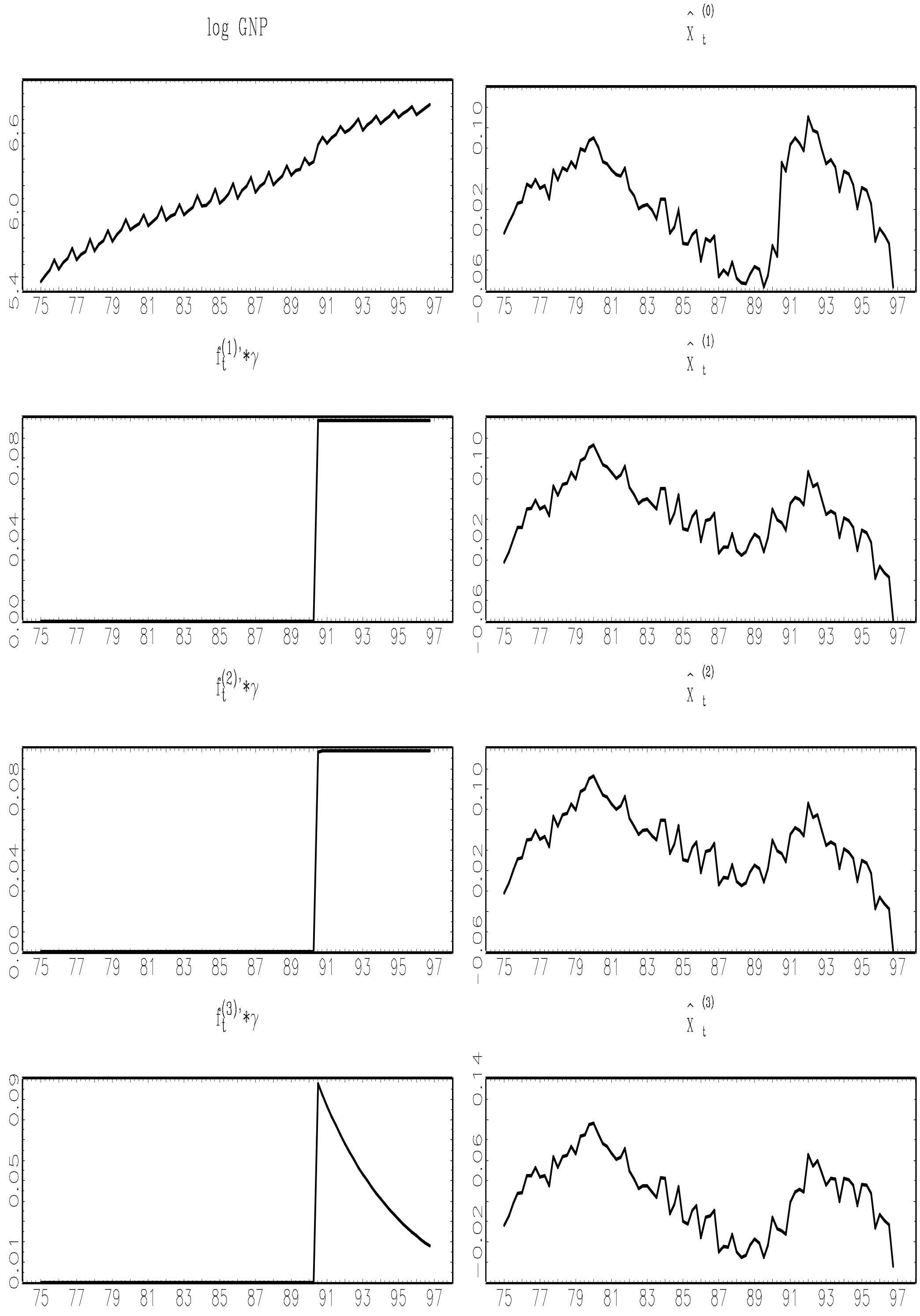

Figure 1: Plots of log GNP, shift functions and corresponding adjusted series.

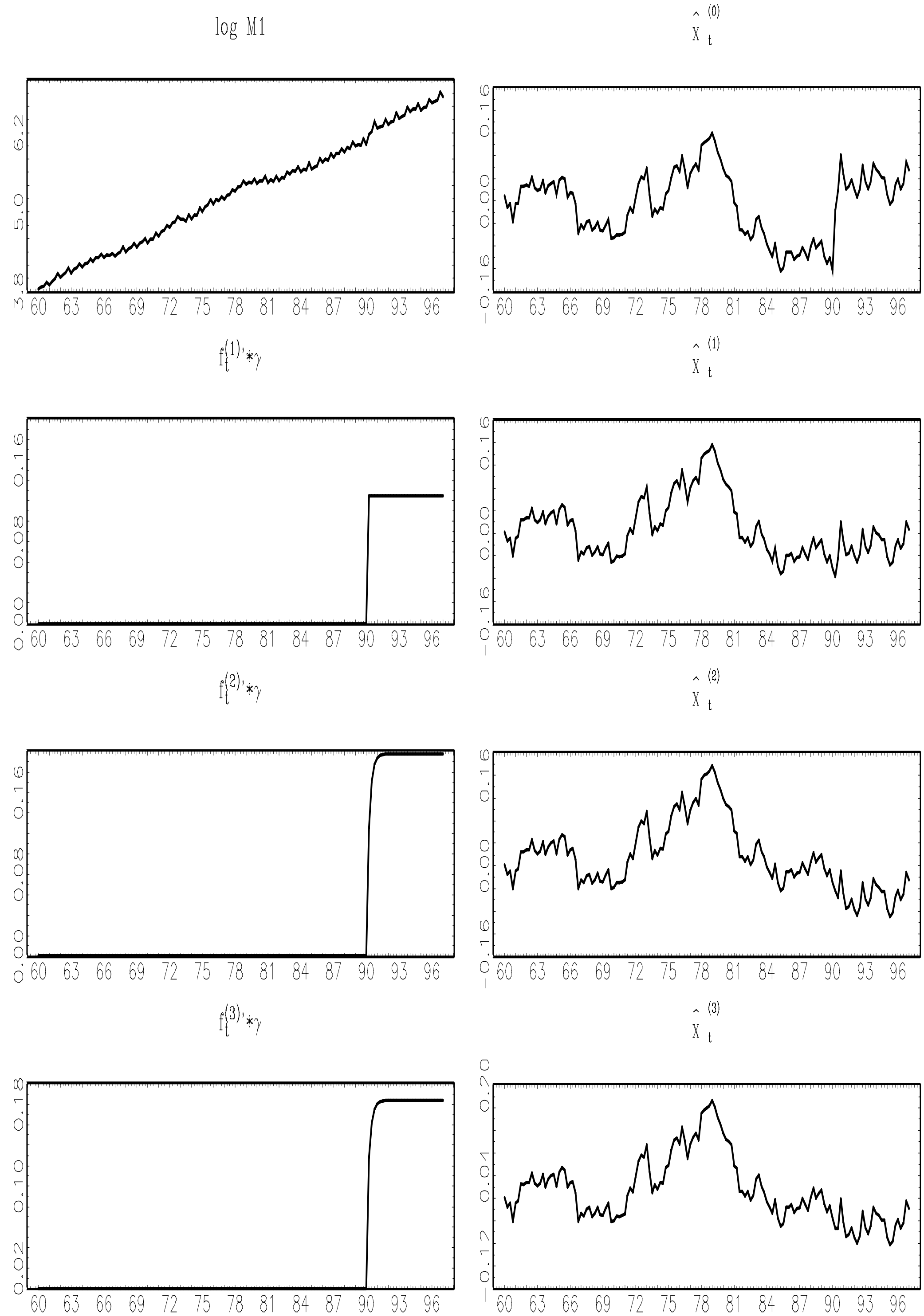

Figure 2: Plots of log M1, shift functions and corresponding adjusted series.

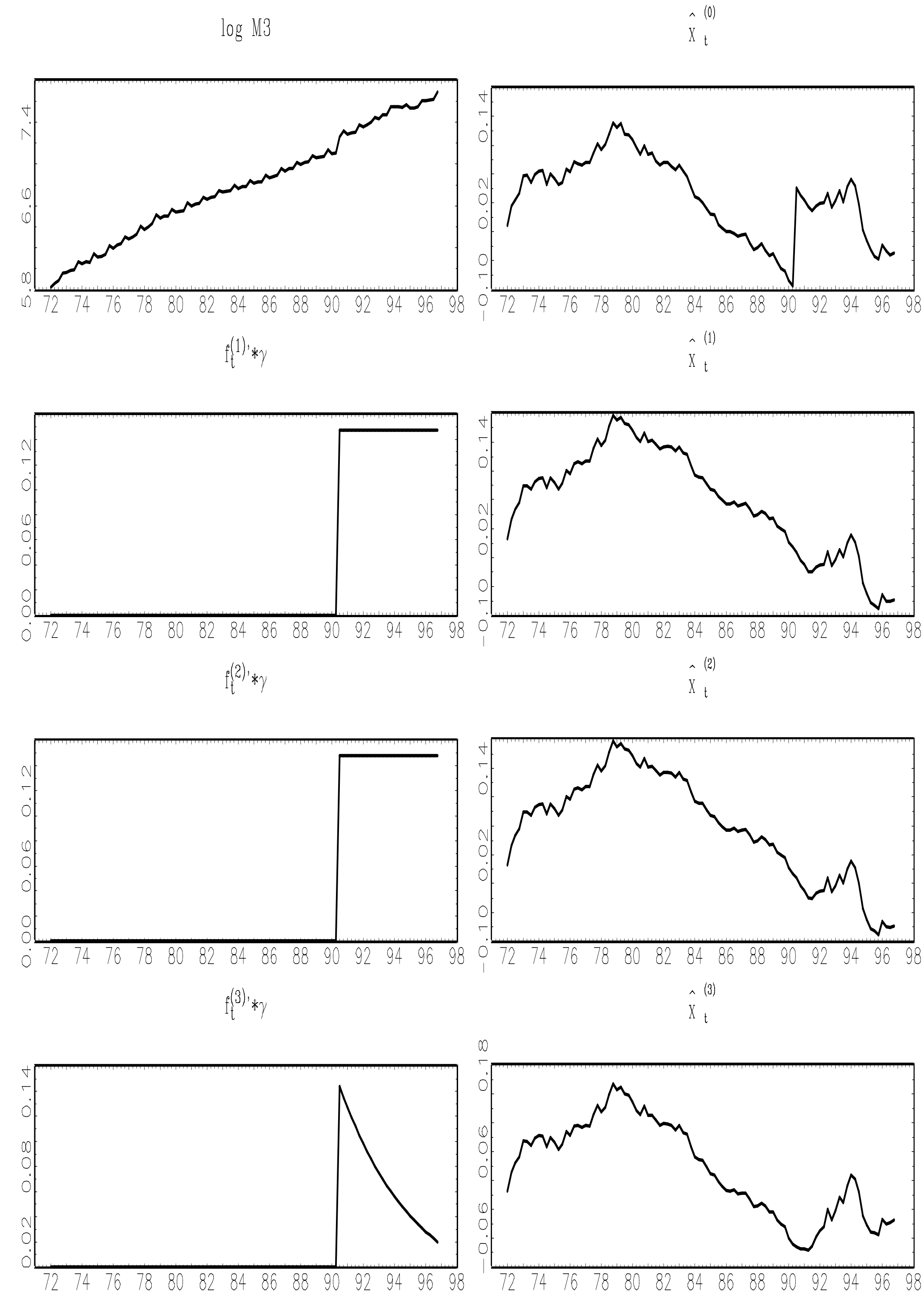

Figure 3: Plots of log M3, shift functions and corresponding adjusted series.