Testing and Estimation of Models with Stochastic Trends

Fabio Busetti

The London School of Economics and Political Science

A thesis submitted for the PhD degree, University of London.

December 2000

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The thesis considers time series and econometric models with stochastic trend components. Locally Best Invariant tests for the presence of stochastic trends are constructed and their asymptotic distributions derived. Particular attention is paid to models with structural breaks, as the tests have high power also against alternative hypotheses in which the trends of the series contain a small number of breaks but are otherwise deterministic. Asymptotic critical values of the tests are tabulated for series with a single breakpoint. A modification of the LBI statistic is then proposed, for which the asymptotic distribution depends only on the number of the breaks and not on their location.

Common stochastic trends imply cointegration and thus testing the number of common trends can also be regarded as testing the dimension of the cointegration space. A test for common trends recently proposed in the literature is extended to series which contain structural breaks.

Testing for the presence of a nonstationary seasonal component is then examined. The LBI test, adjusted for serial correlation by means of a nonparametric correction, is extended in various directions and its performance is compared with that of a parametric test.

Representation, estimation and tests of cointegrated structural time series models form the subject of one chapter, where numerous links with the literature on vector autoregressions are established.

Panel data regression models where the individual effects take the form of individual specific random walks are considered in the last chapter. Imposing the constraint of a common signal-to-noise ratio across individuals makes the maximum likelihood estimator computationally feasible also when the number of units in the cross section is large. For these models an average LBI test for stationarity and for the presence of fixed effects is proposed.

Acknowledgements

I wish to express my gratitude to Professor Andrew Harvey for his guidance, encouragement and constructive criticism during the course of this research. I am greatly indebted to him for sharing his views on the whole subject of time series.

I have also benefited from the comments of several people in the Economics Department of the LSE. I would like to specially thank Professor Peter Robinson, Dr. Javier Hidalgo, Liudas Giraitis and my office mate Javier Hualde for the valuable discussions on the issues raised in this work.

Finally I am mostly grateful to my family for their love and encouragement throughout the years.

Financial support from the Italian National Research Council (CNR) and from the LSE is gratefully acknowledged.

Contents

Chapter 1

Introduction

Economic theory often predicts that certain series, or some linear combination of them, ought to be stationary as opposed to having a stochastic trend component. Knowing the order of integration of a series is also a central issue in applied macroeconomics, when the objective is to identify a suitable econometric specification of aggregate behavioural relationships.

The literature has provided numerous statistical procedures to test the null hypothesis of unit root against the alternative of stationarity, starting with the seminal paper by Dickey and Fuller (1979). Less work has been devoted to developing tests where the null hypothesis is stationarity around a deterministic trend; important contributions in this direction are the articles by Nyblom and Makelainen (1983), Kwiatowski et al. (1992), Nyblom and Harvey (2000).

This thesis considers testing for stochastic trends in the latter framework, focusing on the Locally Best Invariant (LBI) tests. Particular attention is paid to models with structural breaks, as the tests have high power also against alternative hypotheses in which the trends of the series contain a small number of breaks but are otherwise deterministic. The asymptotic critical values of the tests in the presence of structural breaks are tabulated for series with a single breakpoint.

The analysis parallels, in some sense, the work of Perron (1989), where the Dickey-Fuller test has been extended to take into account the presence of level shifts and slope changes in the deterministic part of the series. Here we extend

the KPSS test of Kwiatowski et al. (1992) and, in the multivariate framework, the test on the number of common trends of Nyblom and Harvey (2000).

A modification of the LBI statistic is proposed, for which the asymptotic distribution depends only on the number of the breaks and not on their location. This is important, since although it may be feasible to tabulate critical values for a single break at different points in the sample, constructing tables where there are two or more breaks is impractical. The performance of this modified test is shown, via some simulation experiments, to be similar to that of the LBI test. An unconditional test based on the assumption that there is a single break at an unknown position is also advanced, following the line of literature concerned with the endogeneity of the breakpoint in Perron's test.

Testing for the presence of a nonstationary seasonal component is then examined. The LBI test, adjusted for serial correlation by means of a nonparametric correction, is extended in various directions and its performance is compared with that of a parametric test. It is shown that the asymptotic distribution of the test statistic is not affected by the presence of a deterministic trend even if this contains structural breaks, provided that they are correctly modelled by the inclusion of dummy variables. Integrated regressors with nonseasonal unit roots can also be included without having to change the critical values. A modelled deterministic break in the seasonal pattern, however, will affect the distribution in a rather complicated way; for this situation a modified statistic with a known asymptotic distribution is suggested.

This thesis is also concerned with the estimation of time series and econometric models with stochastic trend components. Maximum likelihood estimation of these models can be carried out from their state space representation, using the Kalman filter algorithm to obtain the prediction error decomposition form of the likelihood function. In models with explanatory variables the Kalman filter effectively diagonalizes the covariance matrix of the errors, allowing the computation of the Generalized Least Squares estimator of the regression coefficients.

A structural time series model is set up in terms of orthogonal components which have a direct interpretation, e.g. trend, cycle and seasonal. In a multivariate model the presence of common stochastic trends imply cointegration,

and thus testing the number of common trends can also be regarded as testing the dimension of the cointegration space. Representation, estimation and tests of cointegrated structural time series models form the subject of one chapter, where numerous links with the literature on vector autoregressions are also established.

We finally consider panel data regression models where the individual effects are dynamic, taking the form of individual specific random walks. Imposing the constraint of a common signal-to-noise ratio across individuals makes the maximum likelihood estimator computationally feasible even when the number of units in the cross section is large. An average LBI test for fixed effects is proposed and consistency of the test is showed.

In summary, the thesis proceeds as follows. Chapter 2 is a review of preliminary concepts: Locally Best Invariant tests, state space models, the Kalman filter and GLS estimation of regression models with state space representation of the errors. In chapter 3 we consider testing for a stochastic trend component in univariate series, with particular reference to models with structural breaks; the use of the tests is illustrated using data on US GNP and on the flow of the Nile. Chapter 4 extends the results of the previous chapter to a multivariate setting and it examines testing of the number of stochastic trends. Asymptotic critical values of the tests are tabulated for models with a single breakpoint at a known position and empirical illustrations are given. Chapter 5 considers testing for a nonstationary seasonal component. Numerous examples are offered and some extensions, such as testing for trading day effects and for nonstationarity of groups of seasons, are investigated. In chapter 6 we discuss representation, estimation and tests of cointegrated structural time series models; an empirical example is provided using US macroeconomic data. Finally, chapter 7 deals with estimation and testing of panel data models with individual specific stochastic trends; as an illustration, a Cobb Douglas production function is estimated from a panel of US manufacturing firms.

Chapter 2

Preliminary concepts: the LBI test, state-space models, the Kalman filter

This chapter reviews some concepts that will be used throughout the thesis. The Locally Best Invariant (LBI) test is defined in section 2.1 and the form of the test statistic is derived for the class of problems of interest to us. For these problems the LBI test is also showed to be one-sided Lagrange Multiplier test. The main references for section 2.1 are Lehmann (1959), Ferguson (1967), King and Hillier (1985), Giri (1996). Section 2.2 considers the state space representation of time series. The Kalman filter algorithm and the prediction error decomposition form of the Gaussian likelihood function are given. In a model with explanatory variables (regressors), an efficient estimator of the regression coefficients can be obtained using the Kalman filter to perform the GLS transformation. This is explained in section 2.3. Sections 2.2-2.3 draw primarily on Harvey (1989). An example of a program for computing the maximum likelihood estimator of a regression model with state space error is provided in section 2.4. The program is the basic step for the algorithm that estimates the dynamic panel data model considered in chapter 7.

2.1 The LBI test

The problem of testing for the presence of a stochastic trend is non-standard. In particular, the distributions of the likelihood ratio and the Wald test statistics are unknown. However an optimal test exists, the LBI test, whose distribution can be derived and belongs to a class of well known statistical distributions. This test is the most powerful in a neighbourhood of the null hypothesis within the class of invariant tests. Invariance is with respect to a particular group of transformations. The LBI test for the class of problems of interest to us is also showed to be one-sided Lagrange Multiplier test.

2.1.1 Locally most powerful tests

Let $\mathcal{X}$ be the sample space, let $\mathcal{A}$ be the σ-algebra of subsets of $\mathcal{X}$, and let Θ be the parametric space. Denote by $\mathcal{P}$ the family of probability distributions P_θ on $\mathcal{A}$. We are concerned with the problem of testing $H_0 : \theta \in \Theta_{H_0}$ against $H_1 : \theta \in \Theta_{H_1}$, where Θ_{H_0} and Θ_{H_1} are disjoint subsets of Θ. A (nonrandomized) test is a measurable mapping $\phi : \mathcal{X} \to \{0, 1\}$ such that

$$\phi(x) = \begin{cases} 1 & \text{for } x \in C, \\ 0 & \text{for } x \notin C, \end{cases}$$

where $C \subseteq \mathcal{X}$ is called the Critical Region of the test. The significance level of the test is defined as $\alpha = \sup_{\theta \in H_0} \mathrm{E}_\theta \phi(X)$.

Denote by $\beta_\phi(\theta)$ the power function of the test ϕ, defined as $\beta_\phi(\theta) = \mathrm{E}_\theta \phi(X)$ for $\theta \in \Theta_{H_1}$.

Definition 1 *A test ϕ^* of size α is* uniformly most powerful *for testing H_0 against H_1 if*

$$\beta_{\phi^*}(\theta) \geq \beta_\phi(\theta), \quad \textit{all } \theta \in \Theta_{H_1},$$

for any other test ϕ of the same size. A test ϕ^* of size α is locally most powerful *for testing H_0 against H_1 if there exists an open neighbourhood $\widetilde{\Theta}$ of Θ_{H_0} such that*

$$\beta_{\phi^*}(\theta) \geq \beta_\phi(\theta), \quad \textit{all } \theta \in \widetilde{\Theta} - \Theta_{H_0},$$

for any other test ϕ of the same size.

Denote by $f_\theta(x)$ the probability density function of X. It is known that for the case of simple hypotheses, i.e. when $\Theta = \{\theta_0, \theta_1\}$, the uniformly most powerful test is given by the Neyman-Pearson Lemma.

Theorem 2 *(Neyman-Pearson Lemma). For testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, a uniformly most powerful test of size α has the form*

$$\phi(x) = \begin{cases} 1 & \text{if } f_{\theta_1}(x)/f_{\theta_0}(x) > k, \\ 0 & \text{if } f_{\theta_1}(x)/f_{\theta_0}(x) < k, \end{cases}$$

where $k \geq 0$ and $E_{\theta_0}\phi(X) = \alpha$.

The expression $f_{\theta_1}(x)/f_{\theta_0}(x)$ is the likelihood ratio. Now consider the problem of testing the (composite) hypothesis $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$. The Neyman-Pearson Lemma still provides the uniformly most powerful test for this problem, but only in the case when the distribution of X has a monotone likelihood ratio. For other classes of distributions, however, a uniformly most powerful test may not exist.

Nevertheless, a locally most powerful test can be found simply by letting $\theta_1 \to \theta_0$ in the Neyman-Pearson Lemma. By definition this test will have maximum power in a small enough neighbourhood of θ_0 amongst all tests with the same size. The critical region of the test is obtained from the Taylor expansion of the likelihood ratio,

$$\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} = 1 + (\theta_1 - \theta_0)\frac{\partial}{\partial\theta}\log f_\theta(x)|_{\theta=\theta_0} + o\left(|\theta_1 - \theta_0|\right),$$

assuming differentiability of $f_\theta(x)$. Thus, letting $\theta_1 \to \theta_0$, we have that a locally most powerful test takes the form

$$\phi^*(x) = \begin{cases} 1 & \text{if } \frac{\partial}{\partial\theta}\log f_\theta(x)|_{\theta=\theta_0} > k, \\ 0 & \text{if } \frac{\partial}{\partial\theta}\log f_\theta(x)|_{\theta=\theta_0} < k, \end{cases} \tag{2.1}$$

for some k. In terms of power, let $\beta'_\phi(\theta) = \int \phi(x) \frac{\partial}{\partial\theta} f_\theta(x) dx$ be the slope of the power function. Then for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$, the power function of ϕ^* in (2.1), $\beta_{\phi^*}(\theta)$, has maximum slope at θ_0 among all the tests ϕ for which $\beta_\phi(\theta_0) = \beta_{\phi^*}(\theta_0)$.

As an example, let $y = X\beta + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2\Omega(\theta))$, θ being a parameter constrained to be nonnegative. We want to test $H_0 : \theta = 0$ against $H_1 : \theta > 0$. From evaluating at $\theta = 0$

$$\frac{\partial}{\partial\theta} \log f_\theta(\varepsilon) = -\frac{1}{2} \frac{\partial}{\partial\theta} \log |\Omega(\theta)| - \frac{1}{2\sigma^2} \varepsilon' \frac{\partial}{\partial\theta} \left(\Omega(\theta)^{-1}\right) \varepsilon,$$

we have that the locally most powerful test has a critical region of the form $\varepsilon' H \varepsilon > k$, where $H = -\frac{\partial}{\partial\theta} \left(\Omega(\theta)^{-1}\right)\big|_{\theta=0}$; more details in subsection 2.1.3. Clearly the test statistic depends on the value of β, which is a nuisance parameters in this context. However, using the concept of invariance it is possible to construct tests which are independent of β. Maximizing the local power in this class of tests defines the locally best invariant test for $H_0 : \theta = 0$ against $H_1 : \theta > 0$.

2.1.2 Invariant tests

Let g be a transformation of the sample space $\mathcal{X}$ which is one to one and bimeasurable. Corresponding to g on $\mathcal{X}$ there will be a transformation $\overline{g}$ on the parametric space Θ, defined by

$$P_\theta(A) = P_{\overline{g}\theta}(gA) \text{ for all } A \in \mathcal{A}.$$

In words, if the random variable X with values in $\mathcal{X}$ has probability distribution P_θ, gX is also a random variable with values in $\mathcal{X}$ and has probability distribution $P_{\theta'}$, where $\theta' = g\theta \in \Theta$.

We say that the parametric space remains invariant under the transformation g of the sample space if $\overline{g}\Theta = \Theta$. Also, it can be shown that a set of transformations, each leaving Θ invariant, can always be extended to an algebraic group G of transformations whose members leave Θ invariant; see Giri (1996, p.34). The group operation is the composition of transformations and the unit element of the group is the identity transformation.

Definition 3 *(Invariance of statistical problems) Let G be a group of transformation on $\mathcal{X}$ that leave Θ invariant. The problem of testing $H_0 : \theta \in \Theta_{H_0}$ against $H_1 : \theta \in \Theta_{H_1}$ is said to be invariant with respect to G if*

(i) $P_{\overline{g}\theta}(gA) = P_\theta(A)$, $g \in G$, $A \in \mathcal{A}$, *and*

(ii) $\Theta_{H_0} = \overline{g}\Theta_{H_0}$, $\Theta_{H_1} = \overline{g}\Theta_{H_1}$.

As an example, the problem of testing $H_0 : \mu \leq 0$ against $H_1 : \mu > 0$ from a random sample of n draws from a $N(\mu, \sigma^2)$ is invariant to the group of scale changes: $X_i \to aX_i$, $i = 1, ..., n$, with $a \neq 0$. The induced transformation is $\overline{g}\theta = (a\mu, a^2\sigma^2)$ with obviously $\overline{g}\Theta = \Theta$. It is not difficult to verify that conditions (i) and (ii) of the above definition hold.

If a statistical problem remains invariant under a group of transformations G operating on the sample space, it is then natural to restrict attention to statistical tests ϕ that are also invariant under G, i.e.

$$\phi(x) = \phi(gx), \quad x \in \mathcal{X}, g \in G.$$

Since a transformation g can be interpreted as a change of coordinate, the idea is that a test is invariant if it is independent of the particular coordinate system in which the data are expressed. Furthermore, an invariant test can always be expressed as a function of a special statistic, the maximal invariant. Thus, in practice, attention will be restricted to finding a maximal invariant and its distribution.

Definition 4 *A function $T(x)$ defined on $\mathcal{X}$ is a* maximal invariant *with respect to G if (i) $T(x) = T(gx)$ for all $x \in \mathcal{X}$, $g \in G$, (ii) $T(x_1) = T(x_2)$ implies that there exists a $g \in G$ such that $x_2 = gx_1$.*

In words, the maximal invariant $T(x)$ takes the same value for realizations connected by some transformation g but takes different values for realizations not connected by any g. For example, a maximal invariant for the group of scale changes is given by $T(x_1, ..., x_n) = (x_1/|x|, ..., x_n/|x|)$ where $|x|^2 = \sum_{i=1}^n x_i^2$, i.e. $T(x_1, ..., x_n)$ is a point on the surface of the n-dimensional unit sphere.

Now let $\nu(\theta)$ be a maximal invariant on Θ with respect to $\overline{G}$, the group of induced transformations on Θ. The following theorem contains two important results; see Ferguson (1967).

Theorem 5 *(A) A test $\phi(x)$ is invariant under G if and only if there exists a function h such that $\phi(x) = h(T(x))$. (B) The distribution of $T(X)$ depends on Θ only through $\nu(\theta)$.*

This theorem states that (A) instead of restricting attention to the class of invariant tests, we may restrict attention to the conceptually simpler class of tests which are functions of $T(x)$, for these two classes of tests are equivalent; (B) through invariance not only we reduce the dimension of the sample space to that of the space of the maximal invariant but also we shrink the parametric space.

Returning to the example of testing $H_0 : \mu \leq 0$ against $H_1 : \mu > 0$ from a random sample of n draws from a $N(\mu, \sigma^2)$, a sufficient statistic for this problem is (U, V), where $U = \sqrt{n}\overline{X}$ and $V = \sum_{i=1}^{N}(X_i - \overline{X})^2$. The group of scale changes can be written as $G = \{g_a : g_a(U, V) = (aU, a^2V),\ a \neq 0\}$. It can be easily verified that a maximal invariant for this group is $T(U, V) = U/\sqrt{V/(n-1)}$, which by definition has a noncentral t-distribution with $n - 1$ degrees of freedom and centrality parameter $\delta = \sqrt{n}\mu/\sigma$. Thus the usual one-sided t-test is a uniformly most powerful invariant test. Note that the distribution of $T(U, V)$ depends only on δ.

In the next subsection we will find a locally best invariant test for a problem of direct interest to us. This will be accomplished in two steps: first characterize the class of invariant tests, then find the test that maximizes the local power within this class.

2.1.3 LBI tests on the error covariance matrix in the linear regression model

Consider the Gaussian linear model with non-spherical disturbances

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2\Omega(\theta)), \tag{2.2}$$

where y is $n \times 1$, X is a $n \times k$ matrix of fixed regressors with rank k, β and σ^2 are unknown parameters, and $\Omega(\theta)$ is a symmetric positive definite matrix for all $\theta \geq 0$. Assume without loss of generality that $\Omega(0) = I_n$. The problem is testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$.

The problem is invariant under the group of transformations $G = \{g : g(y) = ay + Xb,\ a > 0,\ b \in \Re^k\}$ as can be seen from definition 3. To find the maximal invariant under G we can proceed in two steps.

First consider the group $G_1 = \{g_1 : g_1(y) = y + Xb,\ b \in \Re^k\}$. An invariant function under G_1 is My, where $M = I_n - X(X'X)^{-1}X'$, since $My = M(y+Xb)$. Clearly My is not a maximal invariant, as the dimension of the null space of M is $k > 0$. To find a maximal invariant, consider the decomposition $M = PP'$ where P is a $n \times (n-k)$ matrix such that $P'P = I_{n-k}$. Then $z = P'My = P'y$ is a maximal invariant under G_1. As $z \sim N(0, \sigma^2 P'\Omega(\theta)P)$, it is not difficult to see that a locally most powerful test invariant under G_1 has a rejection region of the form $y'MHMy > k$, where $H = -\frac{\partial}{\partial\theta}(\Omega(\theta)^{-1})\big|_{\theta=0} = \frac{\partial}{\partial\theta}\Omega(\theta)\big|_{\theta=0}$. Note that My are the OLS residuals from regressing y on X. However this test is not operative as its distribution depends on the unknown parameter σ^2.

Consider now the group $G_2 = \{g_2 : g_2(z) = az,\ a > 0\}$. As we have already seen, a maximal invariant under G_2 is $z/|z|$, i.e. a point on the surface of the unit sphere. Then, since $G = \{g_2 \circ g_1 : g_1 \in G_1,\ g_2 \in G_2\}$, we have that a maximal invariant under G is

$$v = \frac{P'My}{(y'MPP'My)^{\frac{1}{2}}} = \frac{P'y}{(y'My)^{\frac{1}{2}}}.$$

Kariya (1980) and King (1980) show that the probability density function of v is (with $m = n - k$)

$$f_\theta(v)dv = \tfrac{1}{2}\Gamma\left(\tfrac{1}{2}m\right)\pi^{-\frac{m}{2}}\left|P'\Omega(\theta)P\right|^{-\frac{1}{2}}\left(v'\left(P'\Omega(\theta)P\right)^{-1}v\right)^{-\frac{m}{2}}dv, \qquad (2.3)$$

where dv denotes the uniform measure on the surface of the m-dimensional unit sphere.

From applying the Neyman-Pearson Lemma we see that a uniformly most powerful invariant test for $\mathrm{H}_0 : \theta = 0$ against $\mathrm{H}_1 : \theta = \theta_1 > 0$ in general does

not exist as the critical region takes the form $v'\left(P'\Omega(\theta_1)P\right)^{-1}v < k$, that is it depends on θ_1. However a locally best invariant test ϕ^* as defined in (2.1) can be obtained. From evaluating at $\theta = 0$

$$\begin{aligned}\frac{\partial}{\partial\theta}\log f_\theta(v) &= -\tfrac{1}{2}\frac{\partial}{\partial\theta}\log\left|P'\Omega(\theta)P\right| - \tfrac{m}{2}\frac{\partial}{\partial\theta}\log\left(v'\left(P'\Omega(\theta)P\right)^{-1}v\right)\\ &= -\tfrac{1}{2}\frac{\partial}{\partial\theta}\log\left|P'\Omega(\theta)P\right| + \tfrac{m}{2}\frac{v'\left(P'\Omega(\theta)P\right)^{-1}P'\left(\frac{\partial}{\partial\theta}\Omega(\theta)\right)P\left(P'\Omega(\theta)P\right)^{-1}v}{v'\left(P'\Omega(\theta)P\right)^{-1}v},\end{aligned}$$

recalling that $\Omega(0) = I_n$ and using the definition of v, we see that the LBI test has a critical region of the form

$$\frac{y'MHMy}{y'My} > k,$$

where $H = \frac{\partial}{\partial\theta}\Omega(\theta)\big|_{\theta=0}$. Call $e = My$ the OLS residuals from regressing y on X. Then the critical region can also be written as

$$\frac{e'He}{e'e} > k. \tag{2.4}$$

Remark 1 *Kariya (1980) and King (1980) have also showed that (2.4) is the locally best invariant test when the distribution of ε is "elliptically symmetric" under H_1, i.e. when density of ε is of the form $f_\theta(\varepsilon) = \left|\sigma^2\Omega(\theta)\right|^{-1}q\left(\sigma^{-2}\varepsilon'\Omega(\theta)^{-1}\varepsilon\right)$, where q is a function on $[0,\infty)$ satisfying $\int q(x'x)dx = 1$. Gaussianity is a special case; other examples include the multivariate t-distribution, the compound normal, the multivariate Cauchy.*

We now show that the LBI test (2.4) is also one-sided LM test. The log-likelihood for (2.2) is

$$\ell(\beta,\sigma^2,\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log\left|\Omega(\theta)\right| - \frac{1}{2\sigma^2}\left(y - X\beta\right)'\Omega(\theta)^{-1}\left(y - X\beta\right),$$

with

$$\frac{\partial}{\partial\theta}\ell(\beta,\sigma^2,\theta) = -\frac{1}{2}tr\left(\Omega(\theta)\frac{\partial}{\partial\theta}\Omega(\theta)\right) + \frac{1}{2\sigma^2}\left(y - X\beta\right)'\Omega(\theta)^{-1}\left(\frac{\partial}{\partial\theta}\Omega(\theta)\right)\Omega(\theta)^{-1}\left(y - X\beta\right).$$

Under $H_0 : \theta = 0$, the maximum likelihood estimators of β, σ^2, θ are respectively (the OLS estimator) $\widehat{\beta} = (X'X)^{-1}X'y$, $\widehat{\sigma}^2 = n^{-1}e'e$, $\widehat{\theta} = 0$. Thus

$$\frac{\partial}{\partial\theta}\ell(\beta,\sigma^2,\theta)\bigg|_{\beta=\widehat{\beta},\,\sigma^2=\widehat{\sigma}^2,\,\theta=0} = -\frac{1}{2}tr\left(H\right) + \frac{n}{2}\frac{e'He}{e'e}. \tag{2.5}$$

This last expression is also the Lagrange multiplier for maximizing $\ell(\beta, \sigma^2, \theta)$ under the constraint $\theta = 0$. Since θ is positive under the alternative hypothesis, we have that a large positive value of (2.5) provides evidence in favour of H_1. Thus the critical region of the one-sided LM test is given by (2.4).

The following theorem summarizes the results of this subsection.

Theorem 6 *(LBI test; King and Hillier, 1985). For model (2.2) a locally most powerful test of* $H_0 : \theta = 0$ *against* $H_1 : \theta > 0$ *invariant under the group of transformations* $G = \{g : g(y) = ay + Xb, \ a > 0, b \in \Re^k\}$ *is*

$$\phi^*(y) = \begin{cases} 1 & \text{if } \dfrac{e'He}{e'e} > k, \\ \\ 0 & \text{if } \dfrac{e'He}{e'e} < k, \end{cases}$$

where $e = My$ *are the OLS residuals from regressing* y *on* X, $H = \frac{\partial}{\partial\theta}\Omega(\theta)\big|_{\theta=0}$, *and* $k > 0$ *is an appropriate critical value. Further, the test* ϕ^* *is also one sided LM test.*

2.2 State space models and the Kalman filter

Let y_t be a $(n \times 1)$ vector of observations at time t, $t = 1, ..., T$. In a state space model, y_t is generated by the system

$$y_t = Z_t\alpha_t + d_t + \varepsilon_t, \tag{2.6}$$

$$\alpha_t = G_t\alpha_{t-1} + c_t + R_t\eta_t, \tag{2.7}$$

where the ε_t and η_t are zero mean, serially and mutually uncorrelated random vectors with covariance matrices H_t and Q_t, ε_t is n-dimensional and η_t m-dimensional, α_t is the *state vector* of dimension m, and d_t, c_t, Z_t, G_t, R_t are nonstochastic vectors and matrices. The formulation of the model is completed by specifying the initial conditions on the state vector $a_0 = \mathrm{E}(\alpha_0)$, $P_0 = Var(\alpha_0)$.

The Kalman filter is a recursive algorithm for computing the optimal estimator of the state vector at time t based on the information available at time t, i.e. based on all the observations up to time t. The estimator is optimal in the sense that it minimizes the mean square error within the class of the linear estimators. If the model is Gaussian the minimum is among all the estimators.

Let a_t denote the optimal estimator of the state vector and let P_t denote its mean square error. The Kalman filter is given by

$$
\begin{aligned}
a_{t|t-1} &= G_t a_{t-1} + c_t, && (2.8)\\
P_{t|t-1} &= G_t P_{t-1} G_t' + R_t Q_t R_t', && (2.9)\\
a_t &= a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} \left(y_t - Z_t a_{t|t-1} - d_t\right), && (2.10)\\
P_t &= P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}, && (2.11)
\end{aligned}
$$

where $F_t = Z_t P_{t|t-1} Z_t' + H_t$. (2.8)-(2.9) are called predictions equations, (2.10)-(2.11) are called updating equations. The prediction error of y_t is $v_t = y_t - Z_t a_{t|t-1} - d_t$, $t = 1, ..., T$. Its mean square error is F_t, which is also its unconditional variance. The v_t's are often termed innovations since they represent the new information in the latest observation.

The matrices Z_t, G_t, R_t, H_t, Q_t and the vectors d_t, c_t may contain unknown parameters to be estimated on the basis of the observed data. Collect them into a vector ψ. Under Gaussianity the log-likelihood function can be written in the prediction error decomposition form as

$$
\ell(\psi) = -\frac{nT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log |F_t| - \frac{1}{2} \sum_{t=1}^{T} v_t' F_t^{-1} v_t, \qquad (2.12)
$$

which can be maximized using some numerical optimization algorithm.

A time invariant model is one where the matrices Z_t, G_t, R_t, H_t, Q_t are constant over time. For this model the initial conditions for the Kalman filter, a_0 and P_0, are given by the unconditional mean and variance of α_t, provided α_t is a stationary process (in the sense that the eigenvalues of G are inside the unit circle). When α_t is non-stationary its unconditional distribution is not defined, and the initial conditions are obtained assigning a diffuse prior to α_0, i.e. assuming $a_0 = 0$ and $P_0 = \kappa I_m$ with $\kappa \to \infty$. Modifications of the Kalman

filter to deal with this situation are proposed by Ansley and Kohn (1985), de Jong (1991), Koopman (1997). In practice, however, when there are, say, d nonstationary elements in the state vector it is often possible to construct a proper distribution for α_d, and thus the log-likelihood of $(y_{d+1}, ..., y_T)$ conditional on $(y_1, ..., y_d)$ is given by (2.12) with the summations running from $t = d + 1$ instead of $t = 1$. If $y_1, ..., y_d$ are regarded as being fixed, this log-likelihood is an unconditional one.

As an example consider the univariate random walk plus noise model,

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, \qquad & Var(\varepsilon_t) &= \sigma^2, \\ \alpha_t &= \alpha_{t-1} + \eta_t, \qquad & Var(\eta_t) &= q\sigma^2, \end{aligned}$$

where q is the signal-to-noise ratio. The Kalman filter can be written compactly as

$$\begin{aligned} a_t &= (1 - p_t^*)a_{t-1} + p_t^* y_t, \\ p_t^* &= 1 - 1/(p_{t-1}^* + q + 1), \end{aligned}$$

with $p_t^* = P_t/\sigma^2$. Taking y_1 as fixed permits to initialize the filter at $t = 1$ by $a_1 = y_1$, $p_1^* = 1$. Alternatively, starting off at $t = 0$ with $p_0^* = \kappa$ yields $p_1^* \to 1$ and $a_1 \to y_1$ as $\kappa \to \infty$, which shows that the two procedures are equivalent. Furthemore, also the two likelihoods are equivalent as $F_1^{-1} \to 0$ as $\kappa \to \infty$.

2.3 GLS estimation using the Kalman filter

Let y_t, $t = 1, ..., T$, be a scalar time series generated by the model

$$y_t = x_t'\beta + u_t \tag{2.13}$$

$$u_t = Z_t\alpha_t + \varepsilon_t \tag{2.14}$$

$$\alpha_t = G_t\alpha_{t-1} + \eta_t, \tag{2.15}$$

where x_t is a $k \times 1$ vector of nonstochastic regressors, β is a $k \times 1$ vector of unknown parameters and u_t is an error term that has a state space representation as in the previous section. As regards the initial conditions, assume for the

moment that $\mathrm{E}(\alpha_0) = 0$ and that $Var(\alpha_0)$ is bounded. Write (2.13) in matrix notation as

$$Y = X\beta + U,$$

where Y, X, U are obtained by stacking the T observations on y_t, x_t', u_t. Let $\Omega = \mathrm{E}(UU')$. The Generalised Least Squares (GLS) estimator of β is known to be $\widetilde{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y$. Harvey (1989, ch. 3) shows that $\widetilde{\beta}$ can be computed using the Kalman filter to diagonalize Ω^{-1}. The argument proceeds as follows.

It is known that there exists a lower diagonal matrix L with ones on the main diagonal and a diagonal matrix F such that

$$\Omega^{-1} = L'F^{-1}L.$$

This decomposition is effectively performed by the Kalman filter, whose output is a set of serially uncorrelated innovations v_t with variance F_t. Stacking the v_t's into a $T \times 1$ vector V, we have that $V = LU$ with $\mathrm{E}(VV') = diag\,(F_1, ..., F_T) = F$.

Now let $X^* = LX$, $Y^* = LY$, i.e. X^* and Y^* are obtained by running the Kalman Filter on X and Y. Then $\widetilde{\beta}$ can be computed as

$$\widetilde{\beta} = \left(X^{*\prime}F^{-1}X^*\right)^{-1} X^{*\prime}F^{-1}Y^*. \tag{2.16}$$

As regards the maximum likelihood estimation of the full model (i.e. both the regression coefficient β and the state space parameters ψ) we have that

$$\begin{aligned} \ell(\beta, \psi) &= -\frac{T}{2}\log 2\pi - \frac{1}{2}\log|\Omega| - \frac{1}{2}(Y - X\beta)'\,\Omega^{-1}\,(Y - X\beta) \qquad (2.17) \\ &= -\frac{T}{2}\log 2\pi - \frac{1}{2}\log|F| - \frac{1}{2}(Y^* - X^*\beta)'\,F^{-1}\,(Y^* - X^*\beta)\,. \end{aligned}$$

Maximising $\ell(\beta, \psi)$ with respect to β yields the (unfeasible) GLS estimator $\widetilde{\beta}(\psi)$ defined in (2.16), with the concentrated likelihood given by

$$\ell_c(\psi) := \max_{\beta} \ell(\beta, \psi) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log|F| - \frac{1}{2}\widetilde{V}'F^{-1}\widetilde{V}, \tag{2.18}$$

where $\widetilde{V} = Y^* - X^*\widetilde{\beta}(\psi)$.

The assumption of zero mean and bounded variance for α_0 may be relaxed. The case of unbounded variance (diffuse prior) can be handled as suggested in

the previous section, whereas a nonzero mean model can always be reduced to a model with zero mean for α_0 by the inclusion of extra regressors (see Harvey, 1989 p.139).

An algorithm for maximising (2.17) is as follows. Given an initial guess ψ_0 we can compute $\widetilde{\beta}(\psi_0)$ and $\ell_c(\psi_0)$ by running the Kalman filter. The starting value ψ_0 is then updated by some numerical optimization procedure applied to (2.18), so to obtain ψ_1. Having obtained ψ_1, we iterate the previous steps and proceed until convergence of $\ell_c(\psi)$. In the next section we provide a simple program that implements this stepwise optimization algorithm. In particular, it computes the maximum likelihood estimator in a regression model where the errors are random walk plus noise. The program runs under the matrix language Ox 2.0 of Doornik (1998) and it requires the additional package Ssfpack 2.2 of Koopman et al. (1998). This program constitutes the basis for the more complicated maximization of the likelihood in the dynamic panel data models of chapter 6[1] .

2.4 An Ox program for computing the maximum likelihood estimator of a regression model with local level errors

```
/*
   This program computes the maximum likelihood estimator
   of a regression model with local level errors,
   using numerical derivatives in the maximization of the likelihood.
   The program runs under Ox 2.0 with SsfPack 2.2.
   Note:  the two variance parameters of the model are
   sigma_eta, sigma_epsilon.
   Here we construct the likelihood in terms of the transformed
   parameters vp = log(sigmas), since vp must not be constrained
```

[1] The software STAMP 5.0 of Koopman et al. (1995) can easily estimate regressions with local level errors (and also local linear trends with stochastic seasonality and cycles), but it does not work with panel data.

```
to be greater than zero.  Then we obtain the variances by
taking exp(2*vp) or the standard deviations by taking exp(vp).
*/


#include <oxstd.h>
#import <maximize>
#include <packages/ssfpack/ssfpack.h>
likelihood (const vp, const plik, const pvsco, const pmhes);
static decl s_my, s_mx, s_vz, s_betagls, s_varbeta,
s_mphi, s_momega, s_msigma, s_var;


main ()
{
decl vp, ir, lik;
decl data =loadmat(''put-file-name-with-data-here'');
// Insert the file name with data ordered as y, x_1, x_2, ..., x_k
s_my=data[][0]';
s_mx=data[][1:]';
s_mphi=<1;1>; // state space matrices
s_momega=zeros(2,2); // this just defines the dimension of omega
s_msigma=<-1;0>; // diffuse prior on the state


/* Maximization of the Likelihood  */
vp=log(<0.5;1>); // initial values
likelihood (vp,&lik, 0,0); // evaluate likelihood
vp += 0.5*log(s_var); // scale starting values
// Iterative maximization of the log-likelihood
MaxControl (100,10,1);
ir = MaxBFGS(likelihood, &vp, &lik, 0, 1);
// N.B. 1 => NUMERICAL DERIVATIVES
```

```
// 0 => identity matrix as starting hessian
print ( ``\n``,MaxConvergenceMsg(ir),
``\nlog-likelihood = ``, lik*columns(s_my),
``\nvar = ``, s_var,
``\nparameters = ``, vp,
``\nomega = ``, s_momega);
print (``\nBeta_GLS = ``,s_betagls,``\n``,
``\nVariance = ``,s_varbeta,``\n``,
``\n\n t-value = ``, s_betagls./(diagonal(s_varbeta.^0.5)') );
}



likelihood (const vp, const plik, const pvsco, const pmhes)
{       // arguments dictated by MaxBFGS
s_momega[0][0]=exp(2*vp[0][0]);// sigma-square_eta
s_momega[1][1]=exp(2*vp[1][0]);// sigma-square_epsilon
decl mkf, vy, mvx=s_mx, vft, i;
// This will compute KF innovations for y and x
mkf = KalmanFil (s_my, s_mphi, s_momega, s_msigma);
vy = mkf[0][];
vft = mkf[2][]; // this contains INVERSE of mse prediction
for (i=0; i<rows(s_mx); ++i)
{
mvx[i][] = KalmanFil(s_mx[i][], s_mphi, s_momega, s_msigma)[0][];
}
vy=dropc(vy,0); // drop first column (observation) as it
mvx=dropc(mvx,0); // doesn't contain proper KF innovation
vft=dropc(vft,0);
s_betagls = ( ((mvx.*vft)*mvx') ^(-1) ) * ((mvx.*vft)*vy');
s_varbeta = ((mvx.*vft)*mvx') ^(-1);
s_vz = s_my - s_betagls'*s_mx;
SsfLik (plik, &s_var, s_vz, s_mphi, s_momega, s_msigma);
```

```
    // concentrated log-likelihood of our state space model
plik[0] /= columns(s_vz); // loglik scaled by sample size
return 1;
}
```

Chapter 3

Testing for a stochastic trend in univariate time series

This chapter considers tests for the presence of a random walk component (stochastic trend) in a stationary or trend stationary time series and extends them to series which contain structural breaks. The tests fit the LBI framework described in the previous chapter.

Original contribution is provided in analyzing the case of series with structural breaks. This analysis is important because the tests for stochastic trend have high power also against alternative hypotheses in which the trend of the series contains a small number of breaks but is otherwise deterministic. Thus, if structural breaks are known to be present, it is vital to take account of them if a test for a random walk component is to be carried out. This argument parallels Perron (1989) who shows how the augmented Dickey Fuller test needs to be modified in the case of breaking trends. The difference with respect to Perron is that the roles of the null and the alternative hypotheses are reversed, as in our tests the null is (trend) stationarity as opposed to unit root.

We derive the LBI test in the case of structural breaks and we also propose a simple modification of it. The advantage of this modified statistic is that its asymptotic distribution is not dependent on the location of the breakpoint and its form is that of the generalised Cramer-von Mises distribution, with degrees of freedom depending on the number of breakpoints. This is important, since although it may be feasible to tabulate critical values for a single break at different points in the sample, constructing tables where there are two or more

breaks is impractical. The performance of this modified test is shown, via some simulation experiments, to be comparable to that of the LBI test.

An unconditional test, based on the assumption that there is a single break at an unknown point is also examined. The test is obtained by choosing the breakpoint that gives the most favourable result for the null hypothesis of trend stationarity using the LBI statistic, i.e. it parallels the argument of Zivot and Andrews (1992) regarding the endogenization of the breakpoint in the Perron's test.

Then we consider parametric and non-parametric corrections to be applied to our test statistics for handling serial correlation, we highlight some extensions and we provide some empirical examples with data on the flow of the Nile and with US GNP.

In summary, the chapter proceeds as follows. Section 3.1 considers the LBI test for the presence of a random walk component and section 3.2 its modification for the case of structural breaks. The asymptotic distribution is derived and the critical values are tabulated. The modified statistic and the unconditional test are the object of sections 3.3 and 3.4 respectively, with the appropriate critical values being tabulated. The correction for serial correlation is in section 3.5, some extensions in 3.6 and the empirical examples in 3.7. Finally, section 3.8 contains the proofs of this chapter's propositions.

3.1 The LBI test for a random walk component

Consider the model $(t = 1, ..., T)$

$$y_t = x_t'\beta + \mu_t + \varepsilon_t, \tag{3.1}$$

$$\mu_t = \mu_{t-1} + \eta_t, \tag{3.2}$$

$$\varepsilon_t \sim NID(0, \sigma^2), \tag{3.3}$$

$$\eta_t \sim NID(0, \sigma_\eta^2), \tag{3.4}$$

where x_t is a set of nonstochastic regressors (including a constant) with coefficients β, ε_t and η_t are mutually independent, and $\mu_0 = 0$. The notation $NID(0, \sigma^2)$ denotes normally and independently distributed with mean zero

and variance σ^2. If $\sigma_\eta^2 > 0$, μ_t is a random walk component or stochastic trend. The objective is to test for $H_0 : \sigma_\eta^2 = 0$ against $H_1 : \sigma_\eta^2 > 0$, i.e. to test for the presence of a random walk component.

By letting $\theta = \sigma_\eta^2/\sigma^2$, it is clear that (3.1)-(3.4) falls into the class of models of chapter 2, equation (2.2), with $\Omega(\theta) = I_T + \theta H$ and H being a $T \times T$ matrix whose element of position (s,t) is $\min(s,t)$. We will sometimes call θ the *signal-to-noise ratio* and H the *random walk generating matrix.*

Theorem 6 of chapter 2 then provides the locally best invariant (LBI) test for $H_0 : \sigma_\eta^2 = 0$ against $H_1 : \sigma_\eta^2 > 0$, where invariance is with repect to the group of transformations $G = \{g : g(y_t) = ay_t + x_t'b, \ a > 0, b \in \Re^k, t = 1, ..., T\}$. The test is also a one-sided Lagrange Multiplier test.

On dividing by the sample size T, we obtain the statistic

$$\xi = \frac{\sum_{t=1}^{T}(\sum_{s=1}^{t} e_s)^2}{T^2\widehat{\sigma}^2}, \tag{3.5}$$

where the e_t's are the OLS residuals from regressing y_t on x_t and $\widehat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} e_t^2$. Note that (3.5) is obtained because in our model $e'He = \sum_{t=1}^{T}(\sum_{s=1}^{t} e_s)^2$. The LBI test then rejects $H_0 : \sigma_\eta^2 = 0$ when $\xi > k$, with k being an appropriate critical value.

The distribution of ξ depends on the form of the regressors x_t. Nyblom and Makelainen (1983) consider the simple case $x_t = 1$, i.e. a random walk plus noise model. In this case, which we will refer to as NM test, we have that $e_t = y_t - \overline{y}$, where $\overline{y}$ is the sample mean. The asymptotic distribution of ξ under the null hypothesis is found by first observing that the partial sum of deviations from the mean converges weakly to a standard Brownian bridge, that is

$$\sigma^{-1}T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]} e_s \Rightarrow B(\cdot),$$

where $B(r) = W(r) - rW(1)$, $r \in [0,1]$, and $W(\cdot)$ is a standard Wiener process or Brownian motion. Hence, by the continuous mapping theorem,

$$\xi \xrightarrow{d} \int_0^1 B(r)^2 dr, \tag{3.6}$$

since $\widehat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(y_t - \overline{y})^2 \xrightarrow{p} \sigma^2$. This asymptotic distribution is the well-known Cramér-von Mises distribution, as in Anderson and Darling (1952). We will denote it as CvM. Note that it is sufficient for the observations to be independent and identically distributed (and satisfy some moments conditions) to yield the result; see Nabeya and Tanaka (1988).

Nyblom (1986) augments the previous model by including a time trend (or drift), i.e. he considers the case $x_t = (1, t)'$. For this case we have the analogous result that the partial sum of residuals from a first order polynomial regression converges weakly to a second level Brownian bridge denoted $B_2(\cdot)$, where, as in McNeill (1978),

$$B_2(r) = W(r) - rW(1) + 6r\,(1-r)\left\{\tfrac{1}{2}W(1) - \int_0^1 W(s)ds\right\}.$$

Then

$$\xi \xrightarrow{d} \int_0^1 B_2(r)^2 dr\ . \tag{3.7}$$

We will refer to this asymptotic distribution as a second level Cramér-von Mises distribution, and denote it as CvM_2. In the case of any ambiguity we will refer to the distribution in (3.6) as CvM_1.

Percentage points for the Cramér-von Mises distribution have been tabulated by Anderson and Darling (1952), MacNeill (1978), Nyblom and Mäkeläinen (1983), Nyblom (1986) and Kwiatkowski *et al.* (1992).

Kwiatkowski *et al.* (1992) have extended the above tests to allow ε_t to be a weak dependent process (instead of a white noise). The idea is to correct the test statistic (3.5) nonparametrically by replacing $\widehat{\sigma}^2$ with a consistent estimator of the long run variance of ε_t. Then the asymptotic distribution of the corrected statistic will be the same as the one derived under the assumption of white noise disturbances. This corrected version of the test will be denoted as KPSS test. The details on this and other procedures for handling serial correlation in ε_t are contained in section 3.5.

One problem with the NM/KPSS tests is that they are consistent also against the alternative hypothesis of a level shift (and/or slope shift) in the series. This is indirectly shown in an early work of Gardner (1969) who derives the NM

statistic in a Bayesian framework to detect a break in an otherwise i.i.d. series. Later Nyblom (1989) derives (3.5) as an LM statistic to test a general form of parameter constancy in the mean of the series, namely for cases when under the alternative hypothesis the mean is a martingale (which includes both the cases of random walk and single shift at a randomly chosen point). Then Lee et al. (1997) show directly that the KPSS statistic diverges when there is a structural break in a (trend) stationary process. Therefore failing to account for a structural break when testing stationarity of a series is likely to produce evidence of nonstationarity[1] .

To illustrate the problem, we consider the NM test when we assume that the true data generating process is

$$y_t = \mu_0 + \delta w_t + \varepsilon_t$$

where $w_t = 1(t > \lambda T)$, $\lambda \in (0,1)$, and $\varepsilon_t \sim NID(0, \sigma^2)$. That is, the series is a white noise with a level shift at time $\tau = [\lambda T]$. Since

$$y_t - \overline{y} = \delta(w_t - \overline{w}) + \varepsilon_t - \overline{\varepsilon},$$

we have that, for $r \in [0,1]$,

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} (y_t - \overline{y}) = O_p(T^{\frac{1}{2}}).$$

Then, since $T^{-1} \sum_{t=1}^{T} (y_t - \overline{y})^2 = O_p(1)$, it follows that the statistic ξ is $O_p(T)$ and so the NM test is consistent against the alternative hypothesis of a level shift in the series.

In the next section we show how the asymptotic distribution of the LBI test for a random walk component changes in the presence of (correctly modelled) structural breaks.

[1] This corresponds to the Perron's (1989) argument. Indeed Leybourne et al. (1999) have also showed the converse, i.e. that the presence of a break in a I(1) series may lead to spurious rejections of the unit root hypothesis.

3.2 Testing in the presence of structural breaks

Suppose there is a structural break at time $\tau + 1$ and let $\lambda = \tau/T$ denote the fraction of the sample before the breaks occurs. The breakpoint λ is assumed to be exogenous and known.

We consider the model (3.1)-(3.4) under four different specifications of the regressors, which correspond to four cases of structural break. Let

$$x_t^i(\lambda) = \begin{cases} (1, w_t(\lambda))' & i = 1, \\ (1, t, w_t(\lambda), t w_t(\lambda))' & i = 2, \\ (1, t, w_t(\lambda))' & i = 2a, \\ (1, t, (t - \lambda T) w_t(\lambda))' & i = 2b, \end{cases} \tag{3.8}$$

where $w_t(\lambda) = 1(t > \lambda T)$. Case 1 corresponds to a level break with no slope, case 2 to a structural break in both the level and the slope, whereas in case 2a and 2b the break occurs respectively in the level only and in the slope only.

From the previous section we know that the LBI (and one-sided LM) test statistic for $H_0 : \sigma_\eta^2 = 0$ against $H_1 : \sigma_\eta^2 > 0$ takes the form

$$\xi_i(\lambda) = \frac{\sum_{t=1}^{T} \left(\sum_{s=1}^{t} e_s \right)^2}{T^2 \widehat{\sigma}^2}, \quad i = 1, 2, 2a, 2b, \tag{3.9}$$

where the e_t's are the residuals from regressing the observations on the appropriate set of regressors x_t^i, $i = 1, 2, 2a, 2b$. The argument λ denotes that the statistic has been constructed for a specific value of the breakpoint location parameter and that its asymptotic distribution depends on it. Model (3.1)-(3.4) with x_t^i as regressors will be often referred to as model i.

As in the previous section, the limiting distribution can be derived by looking at the asymptotic properties of the process followed by the partial sum of residuals $S_i(r, \lambda) = \sigma^{-1} T^{-\frac{1}{2}} \sum_{s=1}^{[Tr]} e_s$, $r \in [0, 1]$, $i = 1, 2, 2a, 2b$. This will converge to a limiting process -defined on an underlying Wiener process- that will depend on λ and collapse to a (second level) Brownian bridge when $\lambda = 0$ or $\lambda = 1$. The asymptotic distribution of $\xi_i(\lambda)$ is then the integral on the unit interval of the square of this process.

The following proposition states the asymptotic distributions of the $\xi_i(\lambda)$ statistics under the null hypothesis $H_0 : \sigma_\eta^2 = 0$. The proof is given in the last section of the chapter.

Proposition 7 . *Let $\{y_t\}$ be generated under the null hypothesis of model i, $i = 1, 2, 2a, 2b$. Then*

$$\xi_i(\lambda) \xrightarrow{d} \int_0^1 [B_i(r,\lambda)]^2 \, dr,$$

where

$$B_1(r,\lambda) = \begin{cases} W(r) - \frac{r}{\lambda} W(\lambda) & \text{for } r \leq \lambda \\ (W(r) - W(\lambda)) - \frac{r-\lambda}{1-\lambda} (W(1) - W(\lambda)) & \text{for } r > \lambda \end{cases}$$

$$B_2(r,\lambda) = \begin{cases} W(r) - \frac{r}{\lambda} W(\lambda) - \\ \quad -\frac{6}{\lambda^3} r(r-\lambda) \left[\int_0^\lambda r dW(r) - \frac{\lambda}{2} W(\lambda) \right] & \text{for } r \leq \lambda \\ (W(r) - W(\lambda)) - \frac{r-\lambda}{1-\lambda} (W(1) - W(\lambda)) - \\ -\frac{6}{(1-\lambda)^3} (r-1)(r-\lambda) \cdot \\ \quad \cdot \left[\int_\lambda^1 r dW(r) - \frac{1+\lambda}{2} (W(1) - W(\lambda)) \right] & \text{for } r > \lambda \end{cases}$$

$$B_{2a}(r,\lambda) = \begin{cases} W(r) - \frac{r}{\lambda} W(\lambda) \\ \quad -\frac{6}{1-3\lambda+3\lambda^2} r(r-\lambda) \\ \quad \cdot \left[\int_0^1 r dW(r) - \frac{\lambda}{2} W(\lambda) - \frac{1+\lambda}{2} (W(1) - W(\lambda)) \right] & \text{for } r \leq \lambda \\ (W(r) - W(\lambda)) - \frac{r-\lambda}{1-\lambda} (W(1) - W(\lambda)) \\ \quad -\frac{6}{1-3\lambda+3\lambda^2} r(r-\lambda) \\ \quad \cdot \left[\int_0^1 r dW(r) - \frac{\lambda}{2} W(\lambda) - \frac{1+\lambda}{2} (W(1) - W(\lambda)) \right] & \text{for } r > \lambda \end{cases}$$

$$B_{2b}(r,\lambda) = \begin{cases} W(r) - rW(1) - \frac{3}{\lambda^3(1-\lambda)^3} \\ \quad \cdot \left\{ \left(a\frac{r^2}{2} - a\lambda r + \frac{r}{2}\left(a\lambda^2 - b(1-\lambda)^2\right) \right) J_1 \right. \\ \quad \left. + \left(b\frac{r^2}{2} - b\lambda r + \frac{r}{2}\left(b\lambda^2 - c(1-\lambda)^2\right) \right) J_2 \right\} & \text{for } r \leq \lambda \\ \\ W(r) - rW(1) - \frac{3}{\lambda^3(1-\lambda)^3} \\ \quad \cdot \left\{ \left(-a\frac{\lambda^2}{2} + b\frac{r^2-\lambda^2}{2} - b\lambda(r-\lambda) + \frac{r}{2}\left(a\lambda^2 - b(1-\lambda)^2\right) \right) J_1 \right. \\ \quad \left. + \left(-b\frac{\lambda^2}{2} + c\frac{r^2-\lambda^2}{2} - c\lambda(r-\lambda) + \frac{r}{2}\left(b\lambda^2 - c(1-\lambda)^2\right) \right) J_2 \right\} & \text{for } r > \lambda \end{cases}$$

with

$$\begin{aligned} a &= (1-\lambda)^3(1+3\lambda), \\ b &= -3\lambda^2(1-\lambda)^2, \\ c &= \lambda^3(4-3\lambda), \\ J_1 &= \textstyle\int_0^\lambda r dW(r) - \lambda W(\lambda) + \frac{\lambda^2}{2} W(1), \\ J_2 &= \textstyle\int_\lambda^1 r dW(r) - \lambda\left(W(1) - W(\lambda)\right) - \frac{(1-\lambda)^2}{2} W(1). \end{aligned}$$

The upper tail percentage points of the asymptotic distributions above are reported in table 3.1 for different values of λ. As in Kwiatkowski et al (1992), they are obtained by simulating empirical approximations to Brownian motions with samples size of 1000 and 10000 replications. We use the random numbers generator of the matrix programming language OX; see Doornik (1998). The figures for $\lambda \to 0$ or $\lambda \to 1$ correspond to the critical values for the Cramér-von Mises distributions of the previous section. As expected, the percentage points -as functions of λ- are symmetric around $\lambda = 1/2$, which is also the minimum for models 1, 2 and 2b.

TABLE 3.1

Quantiles of the asymptotic distribution of $\xi_i(\lambda)$, $i = 1, 2, 2a, 2b$.

i=1: Break in level, no slope					i=2: Break in level and slope				
	0.900	0.950	0.975	0.990		0.900	0.950	0.975	0.990
λ					λ				
0.01	0.339	0.456	0.559	0.716	0.01	0.119	0.146	0.172	0.213
0.1	0.285	0.378	0.471	0.607	0.1	0.095	0.120	0.145	0.175
0.2	0.225	0.293	0.368	0.478	0.2	0.079	0.097	0.114	0.137
0.3	0.189	0.246	0.302	0.379	0.3	0.064	0.079	0.095	0.112
0.4	0.161	0.204	0.245	0.303	0.4	0.056	0.066	0.076	0.091
0.5	0.150	0.187	0.223	0.264	0.5	0.053	0.062	0.071	0.084
0.6	0.164	0.207	0.251	0.314	0.6	0.056	0.067	0.078	0.092
0.7	0.191	0.242	0.295	0.378	0.7	0.065	0.079	0.095	0.118
0.8	0.231	0.305	0.388	0.484	0.8	0.079	0.095	0.115	0.140
0.9	0.283	0.378	0.484	0.606	0.9	0.097	0.119	0.142	0.173
0.99	0.345	0.463	0.581	0.748	0.99	0.117	0.145	0.174	0.213

i=2a: Break in level only					i=2b: Break in slope only				
λ	0.9	0.95	0.975	0.99	λ	0.9	0.95	0.975	0.99
0.01	0.119	0.146	0.172	0.213	0.01	0.119	0.146	0.176	0.213
0.1	0.096	0.122	0.146	0.177	0.1	0.101	0.126	0.152	0.186
0.2	0.085	0.103	0.122	0.143	0.2	0.088	0.108	0.130	0.155
0.3	0.086	0.105	0.121	0.142	0.3	0.078	0.097	0.114	0.139
0.4	0.097	0.123	0.144	0.178	0.4	0.072	0.086	0.100	0.119
0.5	0.105	0.133	0.162	0.209	0.5	0.070	0.083	0.098	0.116
0.6	0.097	0.121	0.145	0.177	0.6	0.073	0.089	0.104	0.126
0.7	0.085	0.102	0.120	0.144	0.7	0.078	0.096	0.116	0.145
0.8	0.084	0.103	0.123	0.147	0.8	0.087	0.109	0.131	0.161
0.9	0.098	0.120	0.143	0.173	0.9	0.101	0.126	0.151	0.187
0.99	0.117	0.145	0.174	0.213	0.99	0.117	0.146	0.175	0.213

For models 1 and 2 the asymptotic distribution can also be characterized in terms of two independent Cramér-von Mises distributions. To see that this is the case, first notice that we can rewrite the statistic as

$$\xi_i(\lambda) = \lambda^2 \frac{\sum_{t=1}^{\tau}(\sum_{s=1}^{t} e_s)^2}{\tau^2 \widehat{\sigma}^2} + (1-\lambda)^2 \frac{\sum_{t=\tau+1}^{T}(\sum_{s=\tau+1}^{t} e_s)^2}{(T-\tau)^2 \widehat{\sigma}^2}, \quad i = 1, 2, \tag{3.10}$$

since from the OLS orthogonality conditions $\sum_{s=1}^{T} e_s = \sum_{s=\tau+1}^{T} e_s = 0$, and this implies $\sum_{s=1}^{\tau} e_s = 0$. Then it is easy to see that for model 1

$$e_s = \begin{cases} y_s - \overline{y}_1 & \text{for } s \leq \tau \\ y_s - \overline{y}_2 & \text{for } s > \tau \end{cases}$$

where $\overline{y}_1 = \tau^{-1} \sum_{t=1}^{\tau} y_t$ and $\overline{y}_2 = (T-\tau)^{-1} \sum_{s=\tau+1}^{T} y_t$ are the averages in the first and second subsamples respectively. A similar result holds for model 2,

where the two sets of residuals are obtained from regressing on a constant and a time trend. Thus the residuals are independent across subsamples and the following proposition holds.

Proposition 8 . *Let* $\{y_t\}$ *be generated under the null hypothesis of model* i, $i = 1, 2$. *Then*

$$\xi_i(\lambda) \xrightarrow{d} \lambda^2 \int_0^1 [B_i(r)]^2 \, dr + (1-\lambda)^2 \int_0^1 [B_i'(r)]^2 \, dr, \quad i = 1, 2, \tag{3.11}$$

where $B_1(\cdot)$ *and* $B_1'(\cdot)$ *are independent Brownian bridges and* $B_2(\cdot)$ *and* $B_2'(\cdot)$ *are independent second level Brownian bridges. Hence the statistic converges to a weighted average of two random variables with independent Cramér-von Mises distributions.*

This is a very simple way to characterize the asymptotic distribution and it is trivially generalizable to the case of more than one break. Note that if the breaks are equispaced the distribution of the statistic (when multiplied by four) converges to the sum of two random variables with independent Cramér-von Mises distributions. Of course assuming equispaced breaks is not appropriate in general. However, the same additivity property can be obtained after a slight modification of the test statistic, as suggested in the following section. By doing this we can eliminate the dependence on the parameter λ in the asymptotic distribution.

3.3 A modified test

The Cramér-von Mises distribution can be represented by a series expansion of independent $\chi^2(1)$ variables, that is

$$CvM = \int_0^1 B(r)^2 dr = \sum_{j=1}^{\infty} (\pi j)^{-2} \chi_j^2(1). \tag{3.12}$$

The proof follows from the argument in Gikhman and Skorokhod (1975, p. 229-230); see also Nyblom (1989). This allows us to characterise the generalised Cramér-von Mises distribution with k degrees of freedom as

$$CvM(k) = \sum_{j=1}^{\infty} (\pi j)^{-2} \chi_j^2(k), \tag{3.13}$$

with the interesting corrolary that, because of the additive property of chi-square distributions, the sum of k independent random variables with $CvM(1)$ distributions is $CvM(k)$. The same additivity property holds for a second level Cramér-von Mises distribution, $CvM_2(k)$. The series expansion is

$$CvM_2 = \int_0^1 B_2(r)^2 dr = \sum_{j=1}^{\infty} \varphi_j^{-2} \chi_j^2(1),$$

where $B_2(\cdot)$ is a second level Brownian bridge, and φ_j is defined by $\varphi_{2j-1} = 2\pi j$ and φ_{2j} being the root of $\tan(\varphi/2) = \varphi/2$ on $(2\pi j, 2\pi(j+1))$, $j = 1, 2, ...$; see Nyblom (1986).

Bearing the above in mind, we propose the following test statistics for models 1 and 2:

$$\xi_i^* = \frac{\sum_{t=1}^{\tau}(\sum_{s=1}^{t} e_s)^2}{\tau^2 \widehat{\sigma}^2} + \frac{\sum_{t=\tau+1}^{T}(\sum_{s=\tau+1}^{t} e_s)^2}{(T-\tau)^2 \widehat{\sigma}^2}, \qquad i = 1, 2. \tag{3.14}$$

Thus we eliminate the weights in (3.11). The statistics still depend on the location of the breakpoint, but their asymptotic distributions do not since

$$\xi_i^* \xrightarrow{d} \begin{cases} CvM_1(2) & \text{for } i = 1 \\ CvM_2(2) & \text{for } i = 2. \end{cases} \tag{3.15}$$

Not having to consult a table giving the distribution of the test statistic for all the possible values of λ is a big advantage; compare the unit root tests in Perron (1989). Furthermore the test immediately generalises to cases where there are several structural breaks. If there are k breaks at times $\tau_1 = \lambda_1 T < ... < \tau_k = \lambda_k T$ the test statistic is

$$\xi_i^*(k) = \sum_{j=1}^{k+1} \frac{\sum_{t=\tau_{j-1}+1}^{\tau_j} \left(\sum_{s=\tau_{j-1}+1}^{t} e_s \right)^2}{(\tau_j - \tau_{j-1})^2 \widehat{\sigma}^2}, \quad i = 1, 2, \tag{3.16}$$

where $\tau_0 = 0$ and $\tau_{k+1} = T$. The distribution of this statistic converges to a (second-level) generalised Cramér-von Mises distribution with $k+1$ degrees of freedom. The advantage is now even greater since constructing tables for all patterns of k breakpoints would be extremely cumbersome.

The upper tail percentage points for Cramér-von Mises distributions with multiple degrees of freedom are tabulated in Nyblom (1989), Canova and Hansen (1995) and Nyblom and Harvey (2000), the latter being reported in table 3.2 below.

Table 3.2a

$CvM_1(k)$ (no time trend)

k	0.90	0.95	0.99
1	0.347	0.461	0.743
2	0.607	0.748	1.074
3	0.841	1.000	1.359
4	1.063	1.237	1.623

Source: Nyblom and Harvey (2000)

Table 3.2b

$CvM_2(k)$ (time trend)

k	0.90	0.95	0.99
1	0.119	0.149	0.218
2	0.211	0.247	0.329
3	0.296	0.332	0.428
4	0.377	0.423	0.521

Source: Nyblom and Harvey (2000)

How good is this modified test? Table 3.3 compares the LBI test, based on (3.9), and the modified test, (3.14), in terms of size and power by a Monte Carlo experiment. The model with a break in level and slope, model 2, was simulated 5000 times for different values of λ and $q = \sigma_\eta^2/\sigma^2$, the test statistics were computed and the number of rejections was counted for 5% asymptotic critical values obtained from table 3.1b for the LBI test and from table 3.2b for the modified test. For $\lambda = 0.5$ the two tests are the same, except that the critical value of the LBI test is one quarter of the critical value of the modified

test. For other values of λ the size and power are comparable, with the LBI being clearly superior only in the region close to the null hypothesis and for the break point near the beginning or end of the sample.

Table 3.3

Size and power comparison between LBI and simplified test for $i = 2$

	lambda	0.1	0.3	0.5	0.7	0.9
	q					
	0	0.048	0.054	0.048	0.053	0.047
LBI	0.01	0.313	0.221	0.177	0.230	0.325
	0.1	0.884	0.824	0.830	0.819	0.896
	1	0.999	0.999	1	1	0.999
	0	0.050	0.051	0.048	0.047	0.047
Simplified	0.01	0.256	0.194	0.177	0.200	0.260
	0.1	0.852	0.832	0.830	0.831	0.853
	1	0.999	1	1	1	0.999

Note: T=100, #replications=5000.

The above experiment was repeated for a data generating process with two structural breaks, and no slope, with the breaks located in a variety of positions. The 5% asymptotic critical values for the LBI statistic, $\xi_1(\lambda_1, \lambda_2)$, are reported in table 3.4 and then the performance of the modified test, $\xi_1(2)$, is compared with the LBI test in table 3.5. The conclusions are similar to those reached for the case of a single break, with the simplified test having a size close to the nominal and power comparable with the LBI test.

Table 3.4

Asymptotic distribution $\xi_i(\lambda_1, \lambda_2)$, $i = 1$.

λ_1, λ_2	1/8, 1/4	1/8, 3/8	1/8, 5/8	1/4, 1/2	1/3, 2/3
0.90	0.205	0.153	0.124	0.112	0.093
0.95	0.269	0.197	0.151	0.140	0.110
0.99	0.418	0.300	0.218	0.208	0.148

Table 3.5

Size and power comparison between $\xi_1(\lambda_1, \lambda_2)$ and $\xi_1(2)$

	λ_1, λ_2	1/8, 1/4	1/8, 3/8	1/8, 5/8	1/4, 1/2	1/3, 2/3
	q					
	0	0.048	0.047	0.048	0.051	0.051
LBI	0.01	0.775	0.723	0.759	0.708	0.756
	0.1	0.989	0.995	0.995	0.997	0.999
	1	1	1	1	1	1
	0	0.047	0.050	0.043	0.055	0.048
Simplified	0.01	0.748	0.753	0.755	0.747	0.754
	0.1	0.994	0.997	0.998	0.998	0.999
	1	1	1	1	1	1

3.4 Unknown breakpoint

The tests of the last two sections depend on knowing the location of the breakpoints. In some instances one would like to test for trend stationarity under the

assumption that there may be a single break in an unknown position.

For a single structural break at an unknown point, we consider a set of unconditional tests, obtained by following the argument in Zivot and Andrews (1992). The idea is to choose the breakpoint that gives the most favourable result for the null hypothesis of trend stationarity using the $\xi_i(\lambda)$ statistic, that is

$$\widetilde{\xi}_i = \inf_{\lambda \in \Lambda} \xi_i(\lambda), \quad i = 1, 2, 2a, 2b, \tag{3.17}$$

where Λ is a closed subset of the interval (0,1).

The distribution of $\widetilde{\xi}_i$ will depend not only on the location of the true breakpoint, denoted λ_0, but also on the magnitude of the level and/or slope shift, because each $\xi_i(\lambda)$ statistic will depend on the latter when it is computed for a breakpoint different from λ_0. The following assumption on the magnitude of the shift allows us to derive the asymptotic distribution of $\widetilde{\xi}_i$.

Shift assumption. *The magnitude of the shifts decreases to zero with the sample size at a rate faster than $T^{-1/2}$ for the level shifts and at a rate faster than $T^{-3/2}$ for the slope shifts.*

Whether this assumption is a reasonable one is open to question. However, in the literature on breakpoint estimation, Bai (1994, 1997) assumes that the magnitude of the shift shrinks to zero at a rate slower than $T^{-1/2}$ in order to derive the asymptotic distribution of the breakpoint estimator. In our case, the rate is faster. Note that the assumption covers the case of no break actually occuring.

Proposition 9 . *Let $\{y_t\}$ be generated under the null hypothesis of model i, $i = 1, 2, 2a, 2b$. Under the above shift assumption*

$$\widetilde{\xi}_i \xrightarrow{d} \inf_{\lambda \in \Lambda} \int_0^1 [B_i(r, \lambda)]^2 \, dr, \qquad i = 1, 2, 2a, 2b,$$

where $B_i(r, \lambda)$ is defined as in proposition 1.

The proposition is proved in the last section of the chapter. First we prove that, under the shift assumption, the asymptotic distribution of proposition 7 still holds when the location of the breakpoint is wrongly assumed. Then it

is sufficient to apply the continuous mapping theorem as in Zivot and Andrews (1992) to get the result. Note that a co-integration test corresponding to case 1 is proposed by Hao (1996), but he does not apparently make the shift assumption.

Table 3.6 provides the asymptotic critical values for the statistic $\widetilde{\xi}_i$, $i = 1, 2, 2a, 2b$, obtained by simulation for a sample size of T =500 using 5000 replications. Each replication yielded one value of $\xi_i(\lambda)$ from its asymptotic distribution for all possible breakpoints (2 to 499). The minimum of these values was taken as a realization from the distribution of the inf-statistic, $\widetilde{\xi}_i$.

Table 3.6

Distribution of the unconditional test

	0.90	0.95	0.99
i=1	0.071	0.087	0.134
i=2	0.033	0.041	0.054
i=2a	0.071	0.089	0.125
i=2b	0.050	0.060	0.084

Note: simulationwithT=500, #replications=5,000.

3.5 The treatment of serial correlation

The model (3.1)-(3.4) can be written in matrix form as

$$y = X\beta + \mu + \varepsilon, \tag{3.18}$$

where y, X, μ, ε are defined by stacking the T elements of y_t, x_t', $\mu_t - \mu_0$, ε_t respectively, for $t = 1, ..., T$. In this section we drop the assumption of white noise for ε and assume instead

$$Var(\varepsilon) = \sigma^2 V,$$

for some positive definite matrix V, i.e. the disturbance term ε can be serially correlated and/or heteroschedastic.

The test defined by the statistic ξ in (3.5) is now biased, in that the asymptotic distribution of ξ is only proportional to the one corresponding to the case of white noise disturbances, the factor of proportionality being related to the spectrum at frequency zero of ε (or long run variance of ε). Kwiatowski, Phillips, Schmidt and Shin (1992) therefore corrects ξ using a nonparametric estimator of the spectrum at zero. Such (nonparametric) correction is often termed KPSS correction and is described in the first subsection below. Note that the corrected test is no longer LBI.

A different approach to dealing with serial correlation is to compute the LBI test for the model (3.18) using a consistent estimator of V. In fact, after premultiplying (3.18) by $V^{-\frac{1}{2}}$, we can still apply theorem 6 of chapter 2 to obtain the LBI test for the presence of a random walk component. The rejection region now takes the form

$$\frac{\widetilde{e}'V^{-1}HV^{-1}\widetilde{e}}{\widetilde{e}'V^{-1}\widetilde{e}} > k, \tag{3.19}$$

where k is a critical value, H is the random walk generating matrix and $\widetilde{e}$ is the vector of generalised least squares (GLS) residuals

$$\widetilde{e} = \left(I_T - X\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right)y.$$

This approach is parametric, since V needs to be estimated by parametrizing (and estimating) the model under H_1. Leybourne and McCabe (1994) and Harvey and Streibel (1997) follow this approach, that will be described in the second subsection below.

It can be conjectured that the parametric test is bound to be superior to the KPSS test in small samples, provided an appropriate model has been fitted to the data. In effect, simulation evidence in the two papers above shows this to be the case.

3.5.1 Nonparametric correction

By the invariance principle, under regularity conditions[2] the partial sum process of the disturbances weakly converges to a standard Wiener process when stan-

[2] Details on the conditions under which the nonparametric correction works are given in the next chapter, section 4.2.

dardized by the square root of the long run variance σ_L^2, i.e.

$$\sigma_L^{-1} T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \varepsilon_t \Rightarrow W(r), \quad r \in [0,1], \tag{3.20}$$

where the long run variance is defined as

$$\sigma_L^2 = \lim_{T\to\infty} \frac{1}{T} Var\left(\sum_{t=1}^{T} \varepsilon_t\right). \tag{3.21}$$

An estimator for σ_L^2 is given by

$$s^2(m) = \sum_{j=-m}^{m} w(j,m)\widehat{\gamma}(j),$$

where $w(j,m)$ is a weighting function and $\widehat{\gamma}(j)$ is the sample autocovariance of the OLS residuals at lag j,

$$\widehat{\gamma}(j) = T^{-1} \sum_{t=j+1}^{T} e_t e_{t-j}.$$

Here we use $w(j,m) = 1 - |j|/(m+1)$; other options are examined e.g. in Andrews (1991). For consistency it is required that $m \to \infty$ at a rate slower than T; see chapter 4 section 4.2.

Using (3.20) it follows that the LBI test statistic (3.5) can be corrected by replacing $\widehat{\sigma}^2$ with $s^2(m)$ to yield the KPSS statistic

$$\widetilde{\xi}_{KPSS} = \frac{\sum_{t=1}^{T}(\sum_{s=1}^{t} e_s)^2}{T^2 s^2(m)}. \tag{3.22}$$

Clearly the KPSS statistic has the same asymptotic distribution as the corresponding LBI statistic for white noise disturbances; the details are given in the next chapter that considers this same problem in the multivariate case.

3.5.2 Parametric correction

Harvey and Streibel (1997) work with the LBI test (3.19) from a state-space approach. The model (3.18) can be easily put in state-space form and the

Kalman filter and the smoother can be applied to it. In particular, the smoother is an algorithm to compute the optimal estimator of the state based on all the observations; see Harvey (1989).

The numerator of (3.18) is a quadratic form in the Tx1 vector

$$u = V^{-1}\widetilde{e},$$

whose element are sometimes called *smoothing errors*. They are obtained directly by the Kalman filter smoother withouth having to invert the TxT matrix V. An appropriate standardization of (3.19) then leads to the following statistic

$$\widetilde{\xi}_{HS} = \frac{\sum_{t=1}^{T}(\sum_{s=1}^{t} u_s)^2}{T^2\widehat{\gamma}}, \tag{3.23}$$

where $\widehat{\gamma} = T^{-1}\widehat{\sigma}^2 1' V^{-1} 1$ emerges as a by-product of the Kalman smoother calculations; see Harvey and Streibel (1997) for the details. The asymptotic distribution of (3.23) is the same as the one of the corresponding LBI statistic for white noise disturbances. Note that in practice the matrix V depends on unknown parameters that need to be estimated. Harvey and Streibel (1997) suggest to estimate them under H_1, by fitting an appropriate state space model. After estimation, the smoothing errors are obtained and the statistic (3.23) can be computed.

In the ARIMA framework, a parametric correction to the LBI statitic has been proposed by Leybourne and McCabe (1994). They allow for serial correlation by introducing lagged values of the dependent variable, i.e. they consider the model

$$\phi(L)y_t = x_t'\beta + \mu_t + \varepsilon_t, \tag{3.24}$$

$$\mu_t = \mu_{t-1} + \eta_t, \tag{3.25}$$

where $\phi(L)$ is an AR(p) polinomial with roots outside the unit circle. Under $H_1 : \sigma_\eta^2 > 0$, the reduced form of (3.24)-(3.25) is an ARIMA(p,1,1) process. Leybourne and McCabe (1994) estimate the reduced form and construct the LBI statistic (3.5) using the residuals from regressing $\widetilde{\phi}(L)y_t$ on x_t, where $\widetilde{\phi}(L)$

is the estimated AR polinomial obtained from the reduced form. Again, it can be shown that the critical values for this test are the same as the ones for the corresponding LBI test with white noise disturbances.

Finally, it has to be said that the parametric approach to deal with serial correlation in the disturbances requires the extra effort of fitting an adequate model to the data, where adequacy is checked through the usual set of diagnostics. However, such effort is rewarded with higher power and better size in small samples, as showed by the extensive simulation experiments reported in the cited papers.

3.6 Extensions

The presence of seasonal dummies will not affect the asymptotic distributions of the test statistics described so far; this will be proved in the next chapter, section 4.5. If the seasonal pattern evolves according to a nonstationary process with complex unit roots, it can be modelled explicitly as suggested by Harvey and Streibel (1997) or rendered stationary by an appropriate transformation.

Canova and Hansen (1995) developed a procedure, analogous to the KPSS test, for testing against the hypothesis that a series contains a nonstationary seasonal component; this test, and various extension of it, will be the subject of chapter 5.

In the next chapter, section 4.5, we also show that if the model (3.1)-(3.4) is augmented by the inclusion of weakly dependent exogenous regressors, the asymptotic distribution of the test statistics of the previous sections remains unaffected. The asymptotic distributions do change, however, if the regressors are I(1). In this case testing for the presence of a random walk means testing for the null hypothesis of cointegration. This has been considered, in the LBI framework, by Shin (1994), Harris and Inder (1994) and Hao (1996).

An LBI test for a *smooth stochastic trend* has also been proposed in the literature. A smooth trend, or integrated random walk, takes the form

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1}, \\ \beta_t &= \beta_{t-1} + \zeta_t, \end{aligned}$$

$$\zeta_t \sim NID(0, \sigma_\zeta^2),$$

i.e. it is an I(2) process driven by the disturbance ζ_t. Theorem 6 of chapter 2 can be used to derive the LBI statistic for testing $H_0 : \sigma_\zeta^2 = 0$ against $H_1 : \sigma_\zeta^2 > 0$ in a regression model with a smooth trend and white noise disturbances. Nyblom and Harvey (1997) derive the test and tabulate the critical values for the case of $x_t = (1, t)'$; the details are omitted here. However, they also show that this test appears to have little or no power advantage over using the LBI test for the presence of a random walk with drift. The latter one is then advisable in most practical circumstances.

3.7 Empirical examples

Annual data on the volume of the flow of the Nile (in cubic metres $\times 10^8$) is shown in Figure 3.1; see Koopman et al (1995). Fitting a mean and computing the test statistic (3.5) gives a value of $\xi = 2.527$, indicating a clear rejection of the null hypothesis that there is no random walk component; the asymptotic 5% critical value is 0.461. The KPSS test gives the same result with the statistics for $\ell = 3$ and 7 being 1.100 and 0.735 respectively. However, it is known that the first Aswan dam was constructed in 1899 and if a level intervention is included, neither the LBI nor the simplified test rejects the null hypothesis, since $\xi_1(\lambda) = 0.088$ and $\xi_1^* = 0.301$. In fact a simple constant level plus noise model with a break at 1899 provides a good fit to the data; see Harvey, Koopman and Penzer (1998). There is a possible outlier in 1913, but including a dummy variable in the model for this year has little impact on the test statistics. Since the stationary part of the model appears to be white noise, the KPSS correction is unnecessary, but the test statistics for $\ell = 3$ and 7 are 0.074 and 0.096 and so it appears to have little adverse effect.

The unconditional test also does not reject the null hypothesis, the test statistic taking the values 0.058, 0.045 and 0.052 for $\ell = 0$, 3 and 7 respectively. However, it is interesting to note that the break point is located at 1897 rather than 1899.

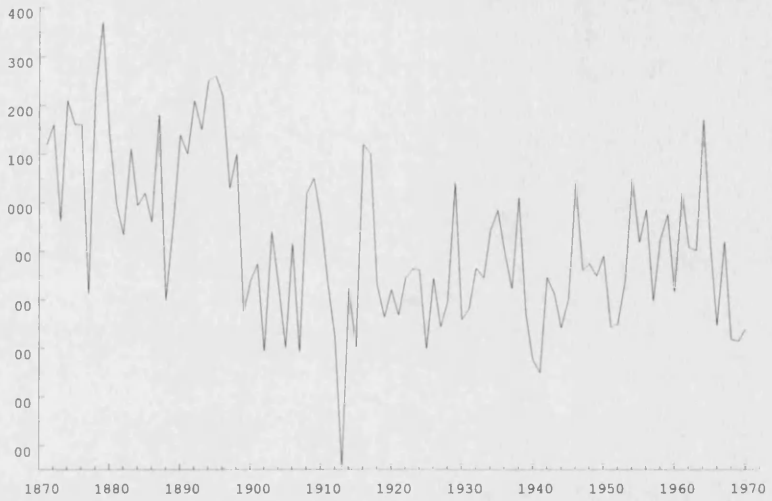

Figure 3.1: Volume of the flow of the Nile (cubic metres $\times 10^8$)

As a second illustration we apply the tests to annual data on US real GNP for the period 1909-1970. These data were used in the well-known article by Nelson and Plosser (1982). On the basis of augmented Dickey-Fuller tests, Nelson and Plosser (1982) did not reject the null hypothesis of a unit root. Subsequently, Perron (1989) observed that a structural break is likely to have occured at the start of the Great Depression, and using his testing procedure he was able to reject the unit root hypothesis in favour of a trend stationary process. Zivot and Andrews (1992) then modified Perron's test by endogenizing the breakpoint, but reached the same conclusion as Perron.

Our own view is that there are a quite a number of places where an argument can be made for the introduction of a break, or a set of breaks, into an economic time series like GNP. Thus we are not dealing with a situation, as in the case of the Nile, where there is a well defined event at a particular point in time which one would expect to give rise to a break in the series. Nevertheless, suppose, following Perron (1989) that we assume there is a known break after 1929 (that is $\tau = 1929$). Table 3.7a shows the results of applying tests with the KPSS correction in the following cases: no break, break in the level, and break in both level and slope. The columns of the table labelled ℓ=0 to ℓ=8 refer to the

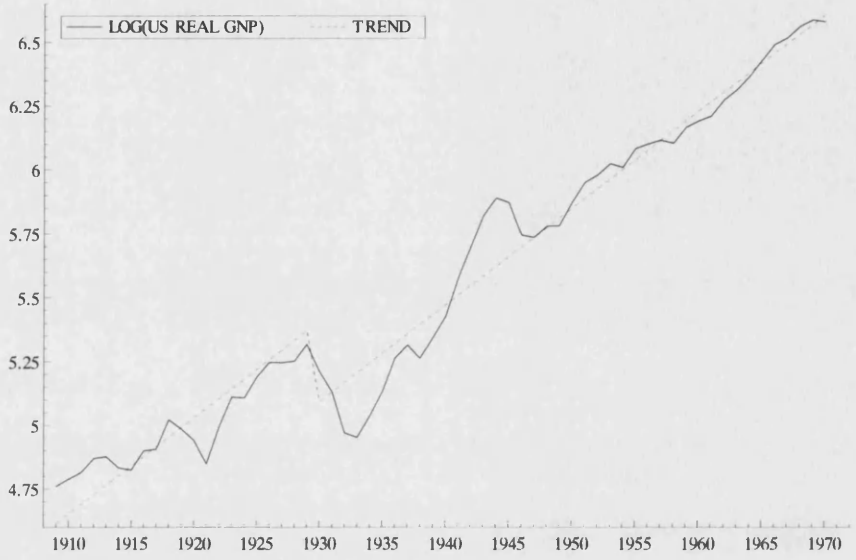

Figure 3.2: US GNP and fitted trend with a break in the level.

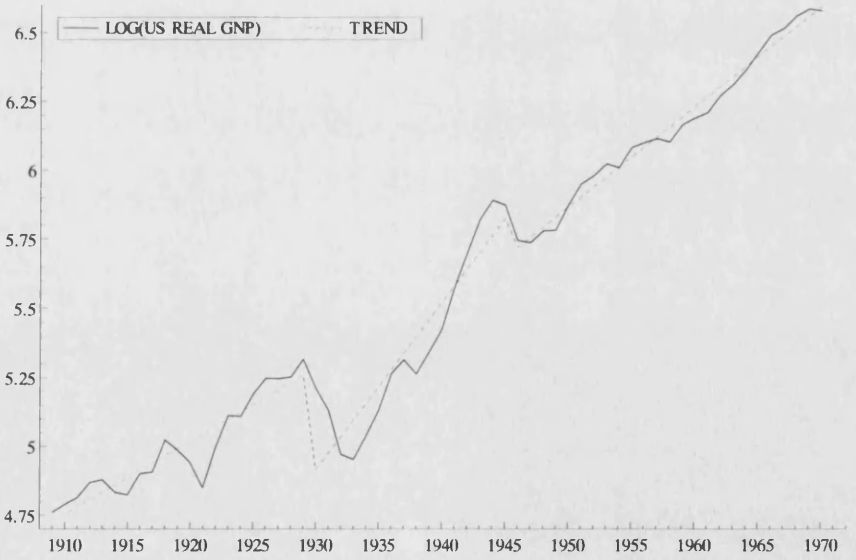

Figure 3.3: US GNP and fitted trend with 2 structural breaks.

lag length in the KPSS correction, while the last two columns report the 5% critical values for $\lambda = 0.3$ and for $\lambda = 0.4$, since the break occurs in between, at $\lambda = 0.34$. The no break case was reported in Kwiatkowski et al. (1992), where it was felt that the evidence favoured the null hypothesis of trend stationarity, though it was noted that the outcome of the test depends on the lag length in the nonparametric correction. Fitting the "Great Crash" model, that is a break in the level as in model 2a, leads to non-rejection of trend stationarity, though the outcome of the test is unclear when we consider a break in both the level and the slope. This finding, then, is in line with Perron (1989). However, the unconditional test, reported in table 3.7b for the case of model 2a, does not lead to a rejection of trend stationarity, though the implied breakpoint is rather misplaced.

Looking at the graph of the series, in figures 3.2 and 3.3, one might equally plausibly assume that there are two breaks: one at the time of the Great Crash and one immediately after the Second World War. Figure 3.3 shows the series with a fitted deterministic trend and breaks in both the level and the slope in 1929 and 1945. Table 3.7c shows results obtained with the simplified test, $\xi_2^*(2)$ in (3.16). These indicate a clear rejection of the trend stationarity hypothesis. Since the breaks have been chosen by examining the data, it could be argued that an unconditional test would be more appropriate. However, this test is redundant if the null hypothesis has already been rejected on the assumption of known breakpoints.

Table 3.7a

Structural break in 1929 ($\lambda = 0.34$)

	l =0	l =1	l =2	l =7	l =8	5% λ=0.3	5% λ=0.4
No break	0.630	0.337	0.242	0.141	0.137	0.149	0.149
Break in level	0.322	0.182	0.138	0.093	0.091	0.105	0.123
Break in level+slope	0.195	0.111	0.086	0.068	0.070	0.079	0.066
Simplified statistic	0.529	0.301	0.232	0.186	0.191	0.247	0.247

Table 3.7b

Unconditional test (structural break in the level)

	l =0	l =1	l =2	l =6	l =7	l =8	5%	1%
Inf-statistic	0.194	0.108	0.081	0.064	0.064	0.066	0.089	0.133
Breakpoint	1926	1926	1926	1920	1920	1920		

Table 3.7c

2 structuralbreaks: 1929 and 1945

	l =0	l =1	l =2	l =6	l =7	l =8	5%	1%
Simplified	0.889	0.552	0.468	0.479	0.501	0.548	0.332	0.428

3.8 Proofs of this chapter's propositions

PROOF OF PROPOSITION 7

To prove the proposition we use the following two lemmas (whose proofs are trivial). The first lemma contains asymptotic approximations of some functions of time, while the second one deals with simple applications of the invariance principle and continuous mapping theorem. Introduce the notations $\sum_1$ for $\sum_{t=1}^{\tau}$,

$\sum_2$ for $\sum_{t=\tau+1}^{T}$, $\bar{t}_1 = \tau^{-1}\sum_1 t$, $\bar{t}_2 = (T-\tau)^{-1}\sum_2 t$ and $\approx$ for asymptotic equivalence.

Lemma 10 *Let $\tau = [\lambda T]$ for $\lambda \in (0,1)$. Then*

- $\sum_1 t \approx \frac{\lambda^2}{2}T^2$,
- $\bar{t}_1 \approx \frac{\lambda}{2}T$,
- $\sum_2 t \approx \frac{1-\lambda^2}{2}T^2$,
- $\bar{t}_2 \approx \frac{1+\lambda}{2}T$,
- $\sum_1 t^2 \approx \frac{\lambda^3}{3}T^3$,
- $\sum_1 (t-\bar{t}_1)^2 \approx \frac{\lambda^3}{12}T^3$,
- $\sum_2 t^2 \approx \frac{1-\lambda^3}{3}T^3$,
- $\sum_2 (t-\bar{t}_2)^2 \approx \frac{(1-\lambda)^3}{12}T^3$,
- $\sum_{s=1}^{t}\left(s-\bar{t}_1\right) \approx \frac{r(r-\lambda)}{2}T^2$ *with* $r = t/T$,
- $\sum_{s=\tau+1}^{t}\left(s-\bar{t}_2\right) \approx \frac{(r-1)(r-\lambda)}{2}T^2$ *with* $r = t/T$.

Lemma 11 *Let $\varepsilon_t \sim NID(0,\sigma^2)$ and $\alpha,\beta \in [0,1]$. Then*

- $\sigma^{-1}T^{-1/2}\sum_{t=[\alpha T]+1}^{[\beta T]}\varepsilon_t \Rightarrow W(\beta) - W(\alpha)$,
- $\sigma^{-1}T^{-3/2}\sum_{t=[\alpha T]+1}^{[\beta T]} t\varepsilon_t \Rightarrow \int_\alpha^\beta r\,dW(r)$,
- $\sigma^{-1}T^{-3/2}\sum_1\left(t-\bar{t}_1\right)\varepsilon_t \Rightarrow \int_0^\lambda r\,dW(r) - \frac{\lambda}{2}W(\lambda)$,
- $\sigma^{-1}T^{-3/2}\sum_2\left(t-\bar{t}_2\right)\varepsilon_t \Rightarrow \int_\lambda^1 r\,dW(r) - \frac{1+\lambda}{2}\left(W(1)-W(\lambda)\right)$,

where $W(.)$ is a standard Wiener process.

Consider model 2 first (the proof for model 1 is a special case). Under $H_0 : \sigma_\eta^2 = 0$, we can reparametrize the model as

$$y_t = \mu_1 d_{1t} + \mu_2 d_{2t} + \beta_1 (d_{1t} t) + \beta_2 (d_{2t} t) + \varepsilon_t,$$

where $d_{1t} = 1(t \leq \tau)$, $d_{2t} = 1(t > \tau)$. Denote by $(\cdot)^*$ the transformed variables after partialing out the effect of the two level dummies d_{1t}, d_{2t} (which corresponds to taking deviations from the subsample averages), i.e. transform the model into

$$y_t^* = \beta_1 (d_{1t} t)^* + \beta_2 (d_{2t} t)^* + \varepsilon_t^*.$$

Let $\mathbf{P}$ be the matrix that projects onto the space of these new regressors, with entries

$$p_{st} = \begin{cases} A_1 (s - \bar{t}_1)(t - \bar{t}_1) & 1 \leq s \leq \tau,\ 1 \leq t \leq \tau, \\ A_2 (s - \bar{t}_2)(t - \bar{t}_2) & \tau < s \leq T,\ \tau < t \leq T, \\ 0 & \text{otherwise,} \end{cases}$$

where $A_1 = \left(\sum_1 (t - \bar{t}_1)^2 \right)^{-1}$, $A_2 = \left(\sum_2 (t - \bar{t}_2)^2 \right)^{-1}$. Thus the regression residuals are

$$e_s = \begin{cases} \varepsilon_s - \bar{\varepsilon}_1 - A_1 \left(s - \bar{t}_1 \right) \sum_1 \left(t - \bar{t}_1 \right) \left(\varepsilon_t - \bar{\varepsilon}_1 \right) & \text{for } s \leq \tau \\ \varepsilon_s - \bar{\varepsilon}_2 - A_2 \left(s - \bar{t}_2 \right) \sum_2 \left(t - \bar{t}_2 \right) \left(\varepsilon_t - \bar{\varepsilon}_2 \right) & \text{for } s > \tau, \end{cases}$$

where the notation $\bar{x}_i$ stands for the average of the variable x in the i-th subsample, $i = 1, 2$.

Then, using the results in the lemmas above, it is easy to see that

$$\sigma^{-1} T^{-\frac{1}{2}} \sum_{s=1}^{[T \cdot]} e_s \Rightarrow B_2 (\cdot, \lambda),$$

where $B_2(r, \lambda)$ is defined in proposition 1. This, the continuous mapping theorem and $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$ imply the proposition.

For model 2a we use the following parametrization under H_0:

$$y_t = \mu_1 d_{1t} + \mu_2 d_{2t} + \beta t + \varepsilon_t.$$

Taking deviations from the subsample averages, we obtain $y_t^* = \beta t^* - \varepsilon_t^*$. The projection onto the space spanned by t^* is now given by the matrix $\mathbf{P}$, with entries

$$p_{st} = \begin{cases} A(s-\bar{t}_1)(t-\bar{t}_1), & 1 \le s \le \tau,\ 1 \le t \le \tau, \\ A\left(s-\bar{t}_1\right)\left(t-\bar{t}_2\right), & 1 \le s \le \tau,\ \tau < t \le T, \\ A\left(s-\bar{t}_2\right)\left(t-\bar{t}_1\right), & \tau < s \le T,\ 1 \le t \le \tau, \\ A(s-\bar{t}_2)(t-\bar{t}_2), & \tau < s \le T,\ \tau \le t \le T, \end{cases}$$

where $A = \left[\sum_1 (t-\bar{t}_1)^2 + \sum_2 (t-\bar{t}_2)^2\right]^{-1}$. This yields the following regression residuals

$$e_s = \begin{cases} \varepsilon_s - \bar{\varepsilon}_1 - A\left(s-\bar{t}_1\right)\left(\sum_1 (t-\bar{t}_1)(\varepsilon_t-\bar{\varepsilon}_1) + \sum_2 (t-\bar{t}_2)(\varepsilon_t-\bar{\varepsilon}_2)\right) & \text{for } s \le \tau \\ \varepsilon_s - \bar{\varepsilon}_2 - A\left(s-\bar{t}_2\right)\left(\sum_1 (t-\bar{t}_1)(\varepsilon_t-\bar{\varepsilon}_1) + \sum_2 (t-\bar{t}_2)(\varepsilon_t-\bar{\varepsilon}_2)\right) & \text{for } s > \tau. \end{cases}$$

Then, using the lemmas above, we have that the partial sum process for these residuals weakly converges to $B_{2a}(\cdot, \lambda)$ and therefore the result follows.

In the case of 2b we parametrize the model as

$$y_t = \mu + \beta_1 z_{1t} + \beta_2 z_{2t} + \varepsilon_t,$$

where $z_{1t} = (t-\tau)1(t \le \tau)$ and $z_{2t} = (t-\tau)1(t > \tau)$. Now, first we take whole sample total averages (i.e. project off the constant) and call z_{1t}^*, z_{2t}^* the resulting transformed variables. Then the projection onto the space of the new regressors is defined by the matrix $\mathbf{P} = \mathbf{Z}^* \left(\mathbf{Z}^{*\prime}\mathbf{Z}^*\right)^{-1} \mathbf{Z}^{*\prime}$, with entries

$$p_{st} \approx \frac{3}{\lambda^3(1-\lambda)^3} T^{-3} \left[\left(a z_{1s}^* + b z_{2s}^*\right) z_{1t}^* + \left(b z_{1s}^* + c z_{2s}^*\right) z_{2t}^*\right],$$

where we have already replaced $\left(\mathbf{Z}^{*\prime}\mathbf{Z}^*\right)^{-1}$ with its asymptotic counterpart

$$\frac{3}{\lambda^3(1-\lambda)^3} T^{-3} \begin{bmatrix} a & b \\ b & c \end{bmatrix},$$

with a, b, c defined in the proposition. Using $\mathbf{P}$ we can again compute the residuals and show that the partial sum process weakly converges to $B_{2b}(\cdot, \lambda)$, and the result follows from the same arguments as in the previous cases.

PROOF OF PROPOSITION 9

Let $\lambda_0 = \tau_0/T$ be the true breakpoint parameter. Consider each of the models $1, 2, 2a, 2b$ under the null hypothesis of no random walk and rewrite them as

$$\begin{aligned} y_t(\lambda_0) &= x_t'\beta + \varepsilon_t \qquad \text{for } t \leq \lambda_0 T, \\ y_t(\lambda_0) &= x_t'\beta + w_t'\delta + \varepsilon_t \quad \text{for } t > \lambda_0 T, \end{aligned} \tag{3.26}$$

where for instance for case [2] $x_t = w_t = (1, t)'$, $\beta = (\mu_1, \beta_1)'$, $\delta = (\delta\mu, \delta\beta)$. We can also rewrite (3.26) as

$$y_t(\lambda_0) = x_t'\beta + w_t(\lambda_0)'\delta + \varepsilon_t, \tag{3.27}$$

where $w_t(\lambda_0) = w_t \cdot 1(t > \lambda_0 T)$.

When we compute the statistic (3.17) we regress $y_t(\lambda_0)$ on x_t and $\mathrm{w}_t(\lambda)$ for all λ. So it is useful to express the true model in terms of $\mathrm{w}_t(\lambda)$ with λ not necessarily equal to λ_0. Consider the case $\lambda > \lambda_0$ for example. Then

$$w_t(\lambda_0) = w_t(\lambda) + \varphi_t(\lambda_0, \lambda), \tag{3.28}$$

where $\varphi_t(\lambda_0, \lambda) = w_t \cdot 1(\lambda_0 T < t \leq \lambda T)$. Therefore we may rewrite (3.27) as

$$\begin{aligned} y_t(\lambda_0) &= x_t'\beta + w_t(\lambda)'\delta + \varepsilon_t^*(\lambda, \lambda_0) && (3.29) \\ &= z_t(\lambda)' \begin{pmatrix} \beta \\ \delta \end{pmatrix} + \varepsilon_t^*(\lambda, \lambda_0), && (3.30) \end{aligned}$$

where $\varepsilon_t^*(\lambda, \lambda_0) = \varepsilon_t + \varphi_t(\lambda, \lambda_0)'\delta$ and $z_t(\lambda) = (x_t',\ w_t(\lambda)')'$. The regression (3.30) corresponds to the sequence (over λ) of models that we estimate to compute the statistic (3.17).

Let $\mathbf{M}_Z(\lambda) = \mathbf{I} - \mathbf{Z}(\lambda)\left(\mathbf{Z}(\lambda)'\mathbf{Z}(\lambda)\right)^{-1}\mathbf{Z}(\lambda)'$. The vector of OLS residuals from regression (3.30) is then

$$\begin{aligned} \mathbf{e}(\lambda, \lambda_0) &= \mathbf{M}_Z(\lambda)\boldsymbol{\varepsilon}^*(\lambda, \lambda_0) && (3.31) \\ &= \mathbf{M}_Z(\lambda)\left(\boldsymbol{\varepsilon} + \boldsymbol{\varphi}(\lambda, \lambda_0)\delta\right) && (3.32) \\ &= \mathbf{e}(\lambda, \lambda) + \mathbf{M}_Z(\lambda)\boldsymbol{\varphi}(\lambda, \lambda_0)\delta. && (3.33) \end{aligned}$$

From this expression we deduce that, under the shift assumption, fitting a model with a misspecified breakpoint does not alter the asymptotic distribution given by proposition 7. In fact $\mathbf{M}_Z(\lambda)\boldsymbol{\varphi}(\lambda,\lambda_0)$ are the residuals from removing the broken trend $z_t(\lambda)$ from the piecewise linear variable $\varphi_t(\lambda,\lambda_0)$. Obviously these residuals are $O(1)$ when $\varphi_t(\lambda,\lambda_0) = 1(\lambda_0 T < t \leq \lambda T)$ and $O(t)$ when $\varphi_t(\lambda,\lambda_0) = t \cdot 1(\lambda_0 T < t \leq \lambda T)$. Then the partial sum of the elements of $\mathbf{M}_Z(\lambda)\boldsymbol{\varphi}(\lambda,\lambda_0)\delta$ is $o(T^{1/2})$ under the shift assumption. Therefore

$$T^{-1/2}\sum_{t=1}^{[T\cdot]} e_t(\lambda,\lambda_0) = T^{-1/2}\sum_{t=1}^{[T\cdot]} e_t(\lambda,\lambda) + o(1) \tag{3.34}$$

$$\Rightarrow B_i(\cdot,\lambda), \tag{3.35}$$

where the process $B_i(\cdot,\lambda)$, $r \in [0,1]$, has been defined in proposition 1 for the cases $i = 1, 2, 2a, 2b$.[3]

Then, since the statistic (3.17) is defined as

$$\widetilde{\xi} = \inf_{\lambda\in\Lambda}\left\{T^{-2}\widehat{\sigma}^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t} e_s(\lambda,\lambda_0)\right)^2\right\}, \tag{3.36}$$

where the subscript i has been dropped, applying the continuous mapping theorem as in Zivot and Andrews (1992) yields the result. This is because the statistic can be expressed as a continuous functional of stochastic processes defined on the underlying innovations, ε_t, where continuity is achieved when Λ is a closed subset of the interval (0,1).

[3] Alternatively we could have obtained the same result by keeping the magnitude of the shift fixed and assuming that $|\lambda - \lambda_0| \to 0$ sufficiently fast as the sample size increases. For example for a break in the level the partial sum of the elements of $\mathbf{M}_Z(\lambda)\boldsymbol{\varphi}(\lambda,\lambda_0)$ is $O\left(|\lambda-\lambda_0|T\right)$, so the assumption $|\lambda - \lambda_0| = o\left(T^{-1/2}\right)$ would be adequate. Incidentally, this shows that the LBI test of section 3 is still valid when we don't get the breakpoint exactly but are close to it. However, in this section we are concerned with λ spanning an arbitrary closed subset of the interval (0,1), so the shift assumption is more appropriate.

Chapter 4

Testing for (common) stochastic trends in multivariate time series

This chapter considers the problem of testing for the presence of stochastic trends in multivariate time series. Here by *stochastic trends* we mean a multivariate random walk component. If the covariance matrix of the disturbance driving the random walk is not of full rank we have the case of *common trends* (which also implies cointegration; see chapter 6).

Section 4.1 provides an LBI test on the error covariance matrix in a Gaussian multivariate regression model. This has been obtained by Nyblom and Harvey (2000), hereafter NH, to test for nonstationarity; it can be viewed as a generalization of theorem 6 of chapter 2. The asymptotic distributions of the NH test statistic, under both the null and the alternative hypotheses, are derived in section 4.2 for a general specification of the deterministic component which allows for structural breaks as a particular case. Serial correlation in the errors is treated nonparametrically as in section 3.5.1.

The case of structural breaks (sometimes referred to as *breaking trends*), not considered by Nyblom and Harvey (2000), is examined thoroughly in section 4.3, where critical values for the nonstationarity test are tabulated across a range of values of the breakpoint location parameter. A modified statistic is also proposed whose asymptotic distribution is independent of the breakpoints location and belongs to the Cramer-von Mises family. These results extend those obtained in the previous chapter to a multivariate setting.

Testing for the presence of a certain number of common trends is considered

in section 4.4. The asymptotic distribution of the test statistic is derived and critical values are tabulated for the case of a model with a single breakpoint. This test can also be interpreted as testing the null hypothesis of a certain dimension of the cointegration space against the alternative hypothesis of a smaller dimension.

Tests on the dimension of the cointegration space have been proposed by Johansen (1988, 1991) and Stock and Watson (1988) in the vector autoregression framework. The main difference with respect to our procedure is in the direction of the alternative hypothesis, as their models are "more nonstationary" under the null than under the alternative. The Johansen's type tests have been extended to the case of structural breaks in Kunimoto (1996) and Inoue (1999). Gregory and Hansen (1996a,b) have instead proposed residual based tests for cointegration under regime shifts, but they are not concerned on testing the number of cointegration relationships.

We also show, in section 4.5, that the asymptotic distributions of the test statistics proposed are unchanged when seasonal dummy variables and/or weakly dependent exogenous regressors are included in the model. Then, in section 4.6, we illustrate the use of the tests with UK macroeconomic data and data on road casualties in Great Britain. Lastly, section 4.7 contains the proofs of this chapter's propositions.

4.1 LBI tests on the error covariance matrix in multivariate regression models

Consider the multivariate Gaussian linear model

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \tag{4.1}$$

where $\mathbf{Y}$ is a TxN matrix of dependent variables, $\mathbf{X}$ is a Txp matrix of fixed regressors, the pxN matrix $\mathbf{B}$ contains the regression coefficients, and $\mathbf{E}$ is a TxN matrix of (zero mean) regression errors. As regards the covariance structure of $\mathbf{E}$, we assume that

$$Cov\left(vec\left(\mathbf{E}\right)\right) \;=\; \mathbf{\Sigma}_{\varepsilon} \otimes \mathbf{\Omega}(\theta), \tag{4.2}$$

$$\mathbf{\Omega}(\theta) \;=\; \mathbf{D}+\theta\mathbf{H}, \tag{4.3}$$

where $\mathbf{\Sigma}_\varepsilon$ is a positive definite $N \times N$ matrix, $\mathbf{D}$ and $\mathbf{H}$ are arbitrary known positive definite $T \times T$ matrices, and θ is a non-negative (scalar) parameter.

Note that if $\mathbf{D}$ is the identity matrix and $\mathbf{H}$ is the random walk generating matrix, then the error term $\mathbf{E}$ is the sum of a white noise and a random walk, with the variance of the disturbance term driving the random walk proportional to the variance of the noise. Such model is "homogeneous" in the terminology of Harvey (1989, ch. 8).

The interest is in testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. The testing problem is invariant under the group of transformation

$$G = \{g : g(\mathbf{Y}) = \mathbf{YP} + \mathbf{XA},\ \mathbf{P} \text{ nonsingular } N\text{x}N \text{ matrix},\ \mathbf{A}\ p\text{x}N \text{ matrix}\}, \tag{4.4}$$

as the conditions of definition 3 of chapter 2 are satisfied.

Proceeding in a way analogous to section 2.1.3 it is possible to obtain the locally most powerful invariant (LBI) test of $H_0 : \theta = 0$ against $H_1 : \theta > 0$. This has been done in Nyblom and Harvey (2000). Here we only report the final result, which is the extension of theorem 6 to a multivariate setting.

Theorem 12 *(Multivariate LBI test; Nyblom and Harvey, 2000). For model (4.1)-(4.3) a locally most powerful test of $H_0 : \theta = 0$ against $H_1 : \theta > 0$ invariant under the group of transformations (4.4) is*

$$\phi^*(\mathbf{Y}) = \begin{cases} 1 & \text{if } tr\left\{\left(\widehat{\mathbf{E}}'\mathbf{D}^{-1}\widehat{\mathbf{E}}\right)^{-1}\left(\widehat{\mathbf{E}}'\mathbf{D}^{-1}\mathbf{H}\mathbf{D}^{-1}\widehat{\mathbf{E}}\right)\right\} > k, \\ 0 & \text{if } tr\left\{\left(\widehat{\mathbf{E}}'\mathbf{D}^{-1}\widehat{\mathbf{E}}\right)^{-1}\left(\widehat{\mathbf{E}}'\mathbf{D}^{-1}\mathbf{H}\mathbf{D}^{-1}\widehat{\mathbf{E}}\right)\right\} < k, \end{cases}$$

where $\widehat{\mathbf{E}} = \left(\mathbf{I}_T - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\mathbf{Y}$ are the OLS residuals from regressing $\mathbf{Y}$ on $\mathbf{X}$ and $k > 0$ is an appropriate critical value. Further, the test ϕ^* is also one sided LM test.

In the next section we apply theorem 12 to test for the presence of stochastic trends in a multivariate model, and we derive the distribution of the test statistic under both the null and the alternative hypothesis. The case of serial correlation in the error term is treated nonparametrically as in section 3.5.1.

4.2 The distribution of the test for stochastic trends

Let $\mathbf{y}_t$ be a vector of N time series. Assume $\mathbf{y}_t$ is generated by the model $(t = 1, ..., T)$

$$\mathbf{y}_t = \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \tag{4.5}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \tag{4.6}$$

$$\boldsymbol{\eta}_t \sim i.i.d.(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \tag{4.7}$$

where $\mathbf{x}_t$ is a p-dimensional vector of $non - stochastic$ regressors (including a constant term), $\boldsymbol{\beta}$ is a $px\ N$ matrix of parameters, $\boldsymbol{\mu}_t$ is a multivariate random walk (stochastic trend) with $\boldsymbol{\mu}_0 = \mathbf{0}$, and for the time being $\boldsymbol{\varepsilon}_t$ is a white noise disturbance term with $Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_\varepsilon$, independent of $\boldsymbol{\eta}_s$ for all t and s. When $\mathbf{x}_t = 1$ for every t, Harvey (1989) refers to (4.5)-(4.7) as the multivariate local level model.

Under the assumption of Gaussianity, theorem 12 provides the LBI test for $\mathrm{H}_0 : \boldsymbol{\Sigma}_\eta = 0$ against $H_1 : \boldsymbol{\Sigma}_\eta = \theta\boldsymbol{\Sigma}_\varepsilon$, $\theta > 0$. After normalizing by T^{-1}, the test has a rejection region of the form

$$tr\left[\mathbf{S}^{-1}\mathbf{C}\right] > k, \tag{4.8}$$

where $\mathbf{C} = T^{-2}\sum_{t=1}^{T}\left[\sum_{s=1}^{t}\mathbf{e}_t\right]\left[\sum_{s=1}^{t}\mathbf{e}_t\right]'$, $\mathbf{S} = T^{-1}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t'$, $\mathbf{e}_t$'s are the OLS residuals from regressing $\mathbf{y}_t$ on $\mathbf{x}_t$ and k is an appropriate critical value. The test is invariant with respect to the group of transformations $\mathbf{y}_t \longmapsto \mathbf{P}\mathbf{y}_t + \mathbf{A}\mathbf{x}_t$, where $\mathbf{P}$ is a nonsingular NxN matrix and $\mathbf{A}$ is an arbitrary Nxp matrix.

Under the null hypothesis the model does not contain any stochastic trend component. Under the alternative hypothesis $\boldsymbol{\Sigma}_\eta$ is proportional to $\boldsymbol{\Sigma}_\varepsilon$, i.e. the model is "homogeneous" in the sense of Harvey (1989, ch. 8). Thus, the test maximizes the local power against homogeneous alternatives. However the test is also consistent against the more general alternative hypothesis $H_1 : rank(\boldsymbol{\Sigma}_\eta) > 0$.

NH concentrate on the cases $\mathbf{x}_t = 1$ and $\mathbf{x}_t = (1, t)'$, i.e. on the null hypothesis of stationarity around respectively a constant level and a linear trend. Here

we consider a more general form for the regressors $\mathbf{x}_t$, which covers in particular the case of breaking trends.

The case of serial dependence in the disturbance term $\boldsymbol{\varepsilon}_t$ can be treated as suggested in section 3.5 relating to the univariate model. In this chapter we only consider the nonparametric correction (or KPSS correction) of the statistic (4.8), which is obtained by replacing $\mathbf{S}$ with a consistent estimator of the long run variance of $\boldsymbol{\varepsilon}_t$. The estimator proposed has the form

$$\widehat{\boldsymbol{\Omega}}(m) = \sum_{\tau=-m}^{m} w(\tau, m)\widehat{\boldsymbol{\Gamma}}(\tau),$$

where $w(\tau, m)$ is a weighting function and

$$\widehat{\boldsymbol{\Gamma}}(\tau) = T^{-1} \sum_{t=\tau+1}^{T} \mathbf{e}_t \mathbf{e}_{t-\tau}'$$

is the sample autocovariance at lag τ. Here we use $w(\tau, m) = 1 - |\tau|/(m+1)$, i.e. the simple Bartlett kernel; other possibilities are examined e.g. in Andrews (1991).

The corrected statistic is then

$$\xi_N = tr\left(\widehat{\boldsymbol{\Omega}}(m)^{-1}\mathbf{C}\right). \tag{4.9}$$

The next proposition gives the asymptotic distribution of ξ_N under the null hypothesis and for an assumption on the regressors $\mathbf{x}_t$ which includes breaking trends as a particular case. In the next section we will analyze the breaking trends case in detail, providing upper tail percentage points from that distribution.

Assumption 4.1. The regressors $\mathbf{x}_t$ are non-stochastic and there exists a scaling matrix $\boldsymbol{\delta}_T$ and a bounded piecewise continuous function $\mathbf{x}(r)$ such that (i) $\boldsymbol{\delta}_T \mathbf{x}_{[Tr]} \to \mathbf{x}(r)$ as $T \to \infty$ uniformly in $r \in [0,1]$, and (ii) $\int_0^1 \mathbf{x}(r)\mathbf{x}(r)'dr$ is positive definite.

Assumption 4.2. The vector process $\{\boldsymbol{\varepsilon}_t\}$ satisfies the following assumptions: (i) $\mathrm{E}(\varepsilon_{jt}) = 0$, $j = 1, ..., N$, $t = 1, ..., T$; (ii) $\sup_t \mathrm{E}|\varepsilon_{jt}|^{2\beta} < \infty$, $j = 1, ..., N$, $\beta > 2$; (iii) $\{\boldsymbol{\varepsilon}_t\}$ is strong mixing with mixing coefficients α_h that satisfy

$\sum_{h=1}^{\infty} \alpha_h^{1-2/\beta} < \infty$; (iv) $\mathbf{\Omega} = \lim T^{-1}\mathrm{E}\left(\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t'\right)$ exists and is positive definite.

Assumption 4.3. $m \to \infty$ as $T \to \infty$ such that $m = o\left(T^{1/4}\right)$.

Proposition 13 *Let $\boldsymbol{y}_t$ be generated by the model (4.5)-(4.7) under assumptions 4.1-4.3. Then under $H_0 : \boldsymbol{\Sigma}_\eta = 0$*

$$\xi_N \xrightarrow{d} \int_0^1 \mathbf{B}^X(r)'\mathbf{B}^X(r)dr, \tag{4.10}$$

where $\mathbf{B}^X(r) = \mathbf{W}(r) - \left(\int_0^1 \mathbf{x}(r)d\mathbf{W}(r)'\right)' \left(\int_0^1 \mathbf{x}(r)\mathbf{x}(r)'dr\right)^{-1} \int_0^r \mathbf{x}(s)ds$, $r \in [0,1]$, *with* $\mathbf{W}(\cdot)$ *being a standard vector Wiener process of dimension* N.

The proof is provided in section 4.7 at the end of the chapter. It extends NH, where only the cases $\mathbf{x}_t = 1$ and $\mathbf{x}_t = (1,t)'$ are considered. In those cases $\mathbf{B}^X(r)$ reduces respectively to a standard Brownian bridge, denoted as $\mathbf{B}^1(r)$, and to a second level Brownian bridge, denoted as $\mathbf{B}^2(r)$, where

$$\mathbf{B}^1(r) = \mathbf{W}(r) - r\mathbf{W}(1), \tag{4.11}$$

$$\mathbf{B}^2(r) = \mathbf{W}(r) - r\mathbf{W}(1) + 6r(1-r)\left\{\tfrac{1}{2}\mathbf{W}(1) - \int_0^1 \mathbf{W}(s)ds\right\}. \tag{4.12}$$

More generally, when $\mathbf{x}_t$ contains all the first h powers of t, i.e. from t^0 to t^{h-1}, $\mathbf{B}^X(r)$ is a (multivariate) h^{th}-level Brownian bridge as in McNeill (1978) and the distribution of $\int_0^1 \mathbf{B}^X(r)'\mathbf{B}^X(r)dr$ is called h^{th}-level Cramér-von Mises distribution with N degrees of freedom. Percentage points when $N = 1$ are tabulated in Anderson and Darling (1952), MacNeill (1978), Nyblom and Mäkeläinen (1983), Nyblom (1986), Nabeya and Tanaka (1988), Kwiatkowski *et al.* (1992); when $N > 1$ percentage points are tabulated in Nyblom (1989), Canova and Hansen (1995), Nyblom and Harvey (2000).

In our case the process $\mathbf{B}^X(r)$ is more general as it includes the case of breaking trends for example. We will call this process generalized Brownian bridge.

Assumption 4.1 follows Phillips and Xiao (1998). Note that it excludes the dummy variables used to model seasonal effects; however in section 4.5 we will

show that adding these dummies does not affect the limiting distribution. Assumption 4.2 permits a fairly general correlation structure for the disturbances ε_t, which can also be heteroscedastic. Assumption 4.2 is sufficient for applying the invariance principle and, together with assumption 4.3, for the consistency of $\widehat{\Omega}(m)$; see Phillips (1987). Note that imposing stronger conditions on ε_t would allow faster rates for m. For example, in the classical spectral theory of stationary process only $m = o(T)$ is required. In practice the rate $m = o(T^{1/2})$ can be satisfactory under both the null and the alternative hypothesis; see Kwiatowski et al. (1992).

The percentage points from the distribution of (4.10) can be used to construct an asymptotically valid test for the null hypothesis of $H_0 : \Sigma_\eta = 0$ against $H_1 : rank(\Sigma_\eta) > 0$. If ε_t is a white noise process, then the test is asymptotically equivalent to the LBI test (4.8).

Under the alternative hypothesis $H_1 : rank(\Sigma_\eta) = K > 0$ the statistic diverges, so the test is consistent. The asymptotic distribution of $(m/T)\,\xi_N$ under H_1 is established by the following proposition.

Proposition 14 *Let $\boldsymbol{y}_t$ be generated by the model (4.5)-(4.7) under assumptions 4.1-4.3. Then under $H_1 : rank(\Sigma_\eta) = K > 0$*

$$\frac{m}{T}\xi_N \xrightarrow{d} tr\left\{\left(\int_0^1 \mathbf{W}^X(s)\mathbf{W}^X(s)'ds\right)^{-1}\int_0^1\left(\int_0^r \mathbf{W}^X(s)ds\right)\left(\int_0^r \mathbf{W}^X(s)ds\right)'dr\right\},$$

where $\mathbf{W}^X(r) = \mathbf{W}(r) - \int_0^1 \mathbf{W}(r)\mathbf{x}(r)'dr\left(\int_0^1 \mathbf{x}(r)\mathbf{x}(r)'dr\right)^{-1}\mathbf{x}(r)$, $r \in [0,1]$, *with* $\mathbf{W}(\cdot)$ *being a standard vector Wiener process of dimension K.*

The proof is provided in section 4.7 at the end of the chapter. The process $\mathbf{W}^X(\cdot)$ is the projection in $L_2[0,1]$ of a Wiener process onto the space orthogonal to the span of $\mathbf{x}(\cdot)$. For $\mathbf{x}_t = 1$ it becomes the "demeaned Wiener process" $\mathbf{W}(r) - \int_0^1 \mathbf{W}(r)dr$, and for $\mathbf{x}_t = (1,t)'$ it becomes the "detrended Wiener process" $\mathbf{W}(r) + (6r-4)\int_0^1 \mathbf{W}(r)dr - (12r-6)\int_0^1 r\mathbf{W}(r)dr$.

The NH test, based on the statistic (4.8)/(4.9) with $\mathbf{x}_t = 1$ or $\mathbf{x}_t = (1,t)'$, is also consistent against the alternative hypothesis of a shift in the deterministic

trend. This follows from the same arguments examined in the previous chapter (relating to the univariate version of this model), see also Gardner (1969), Nyblom (1989), Lee et al. (1997). In the next section we therefore study the case of breaking trends when testing for the presence of a random walk component in a multivariate model. The results collapse to those obtained in the previous chapter when the number of the series N is one.

4.3 Testing in the presence of structural breaks

Suppose that there is a shift in the deterministic trend of the series at time $T_1 = \lambda T$, $\lambda \in (0,1)$. We assume that the breakpoint λ is exogenous and known. Like in the previous chapter, we consider the model (4.5)-(4.7) under four different specifications of the deterministic trend. Let

$$\mathbf{x}_t^i(\lambda) = \begin{cases} (1, w_t(\lambda))' & i = 1, \\ (1, t, w_t(\lambda), t w_t(\lambda))' & i = 2, \\ (1, t, w_t(\lambda))' & i = 2a, \\ (1, t, (t - \lambda T) w_t(\lambda))' & i = 2b, \end{cases} \tag{4.13}$$

where $w_t(\lambda) = 1(t > \lambda T)$. Case 1 corresponds to a level break with no slope, case 2 to a structural break in both the level and the slope, whereas in case 2a and 2b the break occurs respectively in the level only and in the slope only.

Since $\boldsymbol{\delta}_T^i \mathbf{x}_{[Tr]}^i(\lambda) \to \mathbf{x}^i(r;\lambda)$, defined by

$$\mathbf{x}^i(r;\lambda) = \begin{cases} (1, w(r;\lambda))' & i = 1, \\ (1, r, w(r;\lambda), r w(r;\lambda))' & i = 2, \\ (1, r, w(r;\lambda))' & i = 2a, \\ (1, r, (r - \lambda) w(r;\lambda))' & i = 2b, \end{cases} \tag{4.14}$$

with $w(r;\lambda) = 1(r > \lambda)$, assumption 4.1 holds. Therefore we can apply proposition 13. Call $\xi_N^i(\lambda)$ the statistic (4.9) constructed using $\mathbf{x}_t^i(\lambda)$ as regressors,

$i = 1, 2, 2a, 2b$. Then under H_0

$$\xi_N^i(\lambda) \xrightarrow{d} \int_0^1 \mathbf{B}^i(r;\lambda)'\mathbf{B}^i(r;\lambda)dr, \tag{4.15}$$

where the definition of the generalized Brownian bridge $\mathbf{B}^i(r;\lambda)$ is obtained by proposition 7 after replacing the underlying scalar Wiener process with a N-dimensional one.

The upper tail percentage points for the distribution of (4.15) when $\lambda =$ 0.002, 0.1, 0.2, 0.3, 0.4, 0.5 are reported in the first column (labelled $K = 0$) of tables 4.1, 4.2, 4.2a, 4.2b. The values for $\lambda > 0.5$ are not reported since the distribution is symmetric around $\lambda = 0.5$. These percentiles are obtained by simulating the processes $\mathbf{B}^i(r;\lambda)$, $i = 1, 2, 2a, 2b$, using a sample size of 1,000 and 100,000 replications. We use the random number generator of the matrix programming language Ox; see Doornik (1998). Note that the values for $\lambda =$ 0.002 closely agree with the values reported in Nyblom and Harvey (2000), which refer to the case $\lambda = 0$ and were obtained using the series expansion of the distribution.

The statistic $\xi_N^i(\lambda)$ can then be used to construct an asymptotically valid test for the presence of stochastic trends for multivariate time series with structural breaks, using the first column of tables 4.1-4.2b to select the appropriate critical values.

The test based on $\xi_N^i(\lambda)$ has desirable properties, among which being asymptotically equivalent to the LBI test for a Gaussian model with serially independent disturbances. However the critical values depends on the breakpoint parameter λ.

In the following subsection we then propose a modified version of the statistic $\xi_N^i(\lambda)$ for which the asymptotic distribution is independent of λ. This extends to multiple breakpoints and allows us to tabulate critical values across multiple dimensions of the breakpoint location. The following subsection also extends the results of section 3.3 to a multivariate setting.

N	Lambda	K=0			K=1			K=2			K=3			K=4			K=5		
		0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
	λ=0.002	0.349	0.464	0.746															
	λ=0.1	0.284	0.375	0.604															
	λ=0.2	0.231	0.302	0.484															
1	λ=0.3	0.187	0.243	0.380															
	λ=0.4	0.160	0.201	0.303															
	λ=0.5	0.152	0.187	0.271															
	λ=0.002	0.606	0.748	1.078	0.163	0.222	0.356												
	λ=0.1	0.494	0.608	0.876	0.134	0.181	0.323												
	λ=0.2	0.402	0.492	0.704	0.114	0.151	0.266												
2	λ=0.3	0.329	0.398	0.558	0.105	0.136	0.230												
	λ=0.4	0.282	0.334	0.454	0.108	0.137	0.213												
	λ=0.5	0.265	0.309	0.409	0.111	0.140	0.213												
	λ=0.002	0.837	0.999	1.357	0.298	0.381	0.622	0.093	0.120	0.202									
	λ=0.1	0.685	0.815	1.106	0.244	0.313	0.510	0.078	0.100	0.166									
	λ=0.2	0.558	0.660	0.897	0.206	0.262	0.419	0.070	0.088	0.143									
3	λ=0.3	0.459	0.538	0.718	0.190	0.235	0.357	0.072	0.089	0.137									
	λ=0.4	0.395	0.453	0.587	0.194	0.232	0.324	0.075	0.093	0.145									
	λ=0.5	0.371	0.422	0.532	0.198	0.234	0.316	0.075	0.094	0.149									
	λ=0.002	1.057	1.232	1.610	0.422	0.528	0.825	0.169	0.208	0.322	0.082	0.078	0.121						
	λ=0.1	0.868	1.004	1.316	0.348	0.435	0.681	0.141	0.172	0.265	0.053	0.065	0.101						
	λ=0.2	0.708	0.820	1.068	0.295	0.365	0.553	0.127	0.153	0.230	0.051	0.062	0.092						
4	λ=0.3	0.583	0.668	0.859	0.272	0.327	0.473	0.129	0.152	0.219	0.053	0.065	0.096						
	λ=0.4	0.502	0.568	0.712	0.273	0.318	0.426	0.134	0.160	0.228	0.054	0.066	0.100						
	λ=0.5	0.473	0.530	0.651	0.277	0.319	0.411	0.136	0.162	0.231	0.054	0.066	0.100						
	λ=0.002	1.273	1.465	1.871	0.548	0.673	1.012	0.240	0.289	0.431	0.113	0.134	0.193	0.047	0.058	0.084			
	λ=0.1	1.046	1.196	1.527	0.451	0.557	0.830	0.201	0.241	0.362	0.096	0.113	0.161	0.040	0.048	0.070			
	λ=0.2	0.855	0.979	1.239	0.383	0.468	0.681	0.181	0.213	0.314	0.092	0.107	0.148	0.040	0.048	0.068			
5	λ=0.3	0.705	0.797	0.999	0.352	0.418	0.581	0.183	0.212	0.296	0.096	0.112	0.152	0.041	0.049	0.071			
	λ=0.4	0.607	0.679	0.831	0.351	0.402	0.523	0.191	0.222	0.304	0.097	0.114	0.158	0.041	0.050	0.072			
	λ=0.5	0.573	0.635	0.762	0.354	0.400	0.500	0.194	0.227	0.307	0.097	0.114	0.160	0.041	0.049	0.072			
	λ=0.002	1.484	1.686	2.124	0.673	0.823	1.192	0.311	0.369	0.542	0.151	0.188	0.259	0.084	0.097	0.133	0.037	0.044	0.062
	λ=0.1	1.215	1.378	1.727	0.553	0.678	0.981	0.260	0.308	0.451	0.137	0.158	0.218	0.073	0.084	0.113	0.032	0.038	0.053
	λ=0.2	0.996	1.126	1.396	0.470	0.567	0.811	0.235	0.274	0.392	0.132	0.150	0.199	0.073	0.083	0.110	0.033	0.039	0.054
6	λ=0.3	0.824	0.922	1.133	0.433	0.509	0.691	0.236	0.272	0.370	0.137	0.156	0.203	0.075	0.086	0.114	0.033	0.039	0.055
	λ=0.4	0.712	0.787	0.946	0.426	0.484	0.617	0.246	0.283	0.375	0.139	0.159	0.213	0.075	0.086	0.116	0.033	0.039	0.055
	λ=0.5	0.672	0.737	0.871	0.429	0.479	0.589	0.251	0.289	0.378	0.139	0.160	0.217	0.075	0.086	0.116	0.033	0.039	0.055

Figure 4.1: TABLE 4.1

N	Lambda	K=0			K=1			K=2			K=3			K=4			K=5		
		0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
	λ=0.002	0.119	0.148	0.217															
	λ=0.1	0.097	0.121	0.178															
	λ=0.2	0.079	0.097	0.142															
1	λ=0.3	0.065	0.079	0.113															
	λ=0.4	0.056	0.068	0.090															
	λ=0.5	0.052	0.061	0.082															
	λ=0.002	0.210	0.245	0.330	0.084	0.105	0.160												
	λ=0.1	0.172	0.201	0.267	0.069	0.086	0.130												
	λ=0.2	0.140	0.163	0.215	0.057	0.070	0.104												
2	λ=0.3	0.116	0.133	0.174	0.049	0.059	0.086												
	λ=0.4	0.099	0.112	0.142	0.045	0.053	0.073												
	λ=0.5	0.094	0.105	0.131	0.045	0.052	0.070												
	λ=0.002	0.294	0.335	0.427	0.151	0.178	0.245	0.061	0.075	0.113									
	λ=0.1	0.242	0.275	0.349	0.124	0.146	0.200	0.050	0.061	0.092									
	λ=0.2	0.197	0.224	0.281	0.102	0.120	0.162	0.042	0.051	0.075									
3	λ=0.3	0.163	0.184	0.228	0.088	0.102	0.134	0.038	0.046	0.065									
	λ=0.4	0.141	0.157	0.189	0.082	0.092	0.117	0.038	0.044	0.061									
	λ=0.5	0.134	0.147	0.175	0.080	0.090	0.112	0.038	0.044	0.060									
	λ=0.002	0.376	0.422	0.521	0.214	0.246	0.322	0.110	0.129	0.177	0.046	0.056	0.082						
	λ=0.1	0.309	0.345	0.425	0.176	0.201	0.262	0.090	0.106	0.146	0.038	0.046	0.067						
	λ=0.2	0.252	0.281	0.343	0.145	0.166	0.214	0.076	0.089	0.120	0.033	0.039	0.056						
4	λ=0.3	0.209	0.232	0.281	0.126	0.142	0.180	0.069	0.079	0.103	0.031	0.037	0.051						
	λ=0.4	0.182	0.199	0.235	0.116	0.129	0.156	0.068	0.076	0.096	0.031	0.037	0.050						
	λ=0.5	0.173	0.187	0.218	0.114	0.126	0.151	0.068	0.077	0.095	0.032	0.037	0.050						
	λ=0.002	0.455	0.504	0.613	0.274	0.311	0.397	0.156	0.179	0.236	0.083	0.096	0.129	0.036	0.043	0.062			
	λ=0.1	0.373	0.414	0.500	0.226	0.255	0.323	0.129	0.147	0.195	0.069	0.080	0.106	0.030	0.036	0.051			
	λ=0.2	0.306	0.337	0.405	0.187	0.211	0.265	0.109	0.124	0.161	0.060	0.069	0.090	0.027	0.031	0.043			
5	λ=0.3	0.255	0.278	0.330	0.162	0.180	0.223	0.099	0.111	0.140	0.057	0.064	0.082	0.027	0.031	0.041			
	λ=0.4	0.222	0.240	0.278	0.150	0.164	0.196	0.097	0.107	0.130	0.057	0.064	0.080	0.027	0.031	0.042			
	λ=0.5	0.211	0.226	0.259	0.148	0.160	0.187	0.097	0.107	0.128	0.058	0.065	0.081	0.027	0.031	0.042			
	λ=0.002	0.532	0.585	0.700	0.334	0.375	0.469	0.202	0.228	0.293	0.119	0.135	0.175	0.066	0.075	0.099	0.030	0.035	0.048
	λ=0.1	0.438	0.481	0.571	0.275	0.308	0.384	0.166	0.189	0.241	0.099	0.111	0.144	0.055	0.063	0.081	0.025	0.029	0.040
	λ=0.2	0.358	0.392	0.463	0.228	0.254	0.313	0.141	0.158	0.200	0.086	0.096	0.122	0.049	0.055	0.071	0.023	0.026	0.035
6	λ=0.3	0.299	0.325	0.380	0.198	0.218	0.263	0.129	0.143	0.175	0.082	0.090	0.112	0.048	0.054	0.067	0.023	0.026	0.035
	λ=0.4	0.261	0.280	0.321	0.183	0.199	0.234	0.126	0.137	0.163	0.082	0.090	0.109	0.049	0.055	0.067	0.023	0.027	0.035
	λ=0.5	0.248	0.265	0.299	0.180	0.194	0.223	0.125	0.136	0.160	0.082	0.091	0.109	0.049	0.055	0.069	0.023	0.027	0.035

Figure 4.2: TABLE 4.2

N	Lambda	K=0			K=1			K=2			K=3			K=4			K=5		
		0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
	λ=0.002	0.119	0.148	0.217															
	λ=0.1	0.098	0.122	0.179															
	λ=0.2	0.085	0.103	0.148															
1	λ=0.3	0.086	0.103	0.142															
	λ=0.4	0.097	0.120	0.160															
	λ=0.5	0.105	0.134	0.205															
	λ=0.002	0.210	0.245	0.330	0.084	0.105	0.160												
	λ=0.1	0.174	0.204	0.270	0.070	0.087	0.131												
	λ=0.2	0.152	0.175	0.228	0.065	0.079	0.114												
2	λ=0.3	0.151	0.173	0.222	0.068	0.083	0.117												
	λ=0.4	0.172	0.202	0.270	0.069	0.085	0.129												
	λ=0.5	0.188	0.223	0.306	0.067	0.084	0.133												
	λ=0.002	0.294	0.335	0.427	0.151	0.178	0.245	0.081	0.075	0.113									
	λ=0.1	0.245	0.278	0.352	0.127	0.149	0.202	0.052	0.063	0.093									
	λ=0.2	0.214	0.241	0.299	0.117	0.135	0.179	0.051	0.061	0.087									
3	λ=0.3	0.214	0.239	0.292	0.121	0.140	0.182	0.053	0.064	0.092									
	λ=0.4	0.243	0.276	0.352	0.124	0.146	0.200	0.051	0.062	0.092									
	λ=0.5	0.263	0.303	0.395	0.122	0.146	0.212	0.050	0.060	0.089									
	λ=0.002	0.376	0.422	0.521	0.214	0.246	0.322	0.110	0.129	0.177	0.046	0.056	0.082						
	λ=0.1	0.313	0.350	0.430	0.179	0.205	0.266	0.093	0.109	0.149	0.040	0.048	0.068						
	λ=0.2	0.275	0.304	0.367	0.166	0.187	0.237	0.092	0.105	0.139	0.041	0.049	0.068						
4	λ=0.3	0.274	0.301	0.360	0.171	0.193	0.239	0.095	0.109	0.143	0.041	0.049	0.071						
	λ=0.4	0.310	0.348	0.429	0.176	0.202	0.266	0.092	0.107	0.145	0.040	0.048	0.068						
	λ=0.5	0.336	0.380	0.481	0.175	0.205	0.282	0.090	0.104	0.143	0.040	0.047	0.066						
	λ=0.002	0.455	0.504	0.613	0.274	0.312	0.397	0.156	0.179	0.236	0.083	0.096	0.129	0.036	0.043	0.062			
	λ=0.1	0.379	0.419	0.506	0.231	0.260	0.328	0.133	0.152	0.199	0.072	0.083	0.110	0.032	0.038	0.053			
	λ=0.2	0.333	0.364	0.433	0.214	0.238	0.292	0.130	0.146	0.186	0.074	0.084	0.108	0.033	0.039	0.054			
5	λ=0.3	0.333	0.362	0.425	0.220	0.244	0.295	0.134	0.152	0.192	0.074	0.085	0.112	0.033	0.039	0.054			
	λ=0.4	0.376	0.416	0.503	0.227	0.257	0.328	0.131	0.149	0.194	0.072	0.083	0.109	0.033	0.038	0.053			
	λ=0.5	0.407	0.454	0.562	0.227	0.262	0.350	0.126	0.146	0.193	0.072	0.082	0.107	0.032	0.038	0.052			
	λ=0.002	0.532	0.585	0.700	0.334	0.375	0.469	0.202	0.228	0.293	0.119	0.135	0.175	0.066	0.075	0.099	0.030	0.035	0.048
	λ=0.1	0.445	0.488	0.578	0.281	0.314	0.389	0.172	0.194	0.247	0.103	0.118	0.149	0.058	0.066	0.085	0.027	0.031	0.042
	λ=0.2	0.391	0.425	0.497	0.260	0.287	0.347	0.168	0.187	0.230	0.105	0.118	0.146	0.060	0.068	0.087	0.028	0.032	0.044
6	λ=0.3	0.390	0.422	0.489	0.267	0.293	0.347	0.173	0.193	0.237	0.106	0.119	0.152	0.060	0.068	0.088	0.027	0.032	0.043
	λ=0.4	0.441	0.482	0.575	0.277	0.309	0.386	0.169	0.190	0.243	0.103	0.115	0.147	0.059	0.067	0.085	0.027	0.032	0.043
	λ=0.5	0.476	0.527	0.640	0.278	0.320	0.414	0.166	0.187	0.242	0.102	0.114	0.145	0.059	0.067	0.085	0.027	0.032	0.043

Figure 4.3: TABLE 4.2a

N	Lambda	K=0			K=1			K=2			K=3			K=4			K=5		
		0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
	λ=0.002	0.119	0.148	0.218															
	λ=0.1	0.103	0.127	0.186															
	λ=0.2	0.089	0.110	0.161															
1	λ=0.3	0.078	0.096	0.138															
	λ=0.4	0.071	0.086	0.123															
	λ=0.5	0.069	0.083	0.118															
	λ=0.002	0.211	0.248	0.329	0.084	0.106	0.160												
	λ=0.1	0.182	0.213	0.284	0.073	0.090	0.137												
	λ=0.2	0.158	0.183	0.242	0.064	0.079	0.117												
2	λ=0.3	0.139	0.161	0.211	0.058	0.070	0.103												
	λ=0.4	0.127	0.146	0.189	0.055	0.066	0.096												
	λ=0.5	0.123	0.141	0.182	0.054	0.065	0.092												
	λ=0.002	0.295	0.338	0.428	0.151	0.178	0.245	0.081	0.075	0.114									
	λ=0.1	0.255	0.290	0.369	0.131	0.154	0.210	0.053	0.065	0.097									
	λ=0.2	0.222	0.252	0.317	0.115	0.135	0.182	0.047	0.057	0.084									
3	λ=0.3	0.196	0.221	0.277	0.104	0.121	0.162	0.044	0.053	0.076									
	λ=0.4	0.179	0.202	0.249	0.099	0.113	0.149	0.043	0.052	0.073									
	λ=0.5	0.174	0.195	0.239	0.097	0.111	0.148	0.043	0.052	0.073									
	λ=0.002	0.376	0.422	0.522	0.214	0.247	0.322	0.110	0.129	0.178	0.046	0.056	0.082						
	λ=0.1	0.326	0.364	0.451	0.186	0.213	0.278	0.096	0.112	0.155	0.040	0.048	0.070						
	λ=0.2	0.283	0.316	0.389	0.163	0.186	0.240	0.085	0.099	0.134	0.036	0.044	0.062						
4	λ=0.3	0.251	0.279	0.338	0.148	0.168	0.214	0.080	0.092	0.122	0.035	0.042	0.059						
	λ=0.4	0.231	0.255	0.308	0.141	0.158	0.200	0.079	0.090	0.118	0.035	0.042	0.058						
	λ=0.5	0.224	0.246	0.296	0.139	0.155	0.193	0.078	0.089	0.115	0.035	0.042	0.058						
	λ=0.002	0.456	0.505	0.614	0.275	0.312	0.398	0.156	0.180	0.237	0.084	0.097	0.130	0.036	0.043	0.062			
	λ=0.1	0.394	0.437	0.530	0.239	0.270	0.341	0.136	0.156	0.206	0.073	0.084	0.112	0.032	0.038	0.053			
	λ=0.2	0.343	0.378	0.455	0.210	0.237	0.297	0.122	0.139	0.180	0.066	0.076	0.100	0.030	0.035	0.048			
5	λ=0.3	0.305	0.334	0.398	0.191	0.214	0.265	0.115	0.129	0.165	0.064	0.073	0.095	0.029	0.034	0.047			
	λ=0.4	0.280	0.306	0.361	0.181	0.201	0.247	0.112	0.125	0.157	0.064	0.073	0.093	0.029	0.034	0.047			
	λ=0.5	0.272	0.296	0.349	0.179	0.197	0.240	0.112	0.125	0.155	0.064	0.073	0.093	0.030	0.035	0.047			
	λ=0.002	0.534	0.587	0.702	0.334	0.376	0.469	0.202	0.229	0.294	0.119	0.135	0.176	0.066	0.076	0.099	0.030	0.035	0.048
	λ=0.1	0.462	0.507	0.605	0.290	0.325	0.405	0.176	0.198	0.254	0.104	0.118	0.152	0.058	0.067	0.086	0.026	0.031	0.042
	λ=0.2	0.403	0.441	0.521	0.256	0.285	0.353	0.158	0.177	0.225	0.095	0.107	0.136	0.054	0.061	0.078	0.025	0.029	0.039
6	λ=0.3	0.358	0.389	0.457	0.233	0.258	0.316	0.148	0.165	0.206	0.092	0.103	0.128	0.054	0.060	0.076	0.025	0.029	0.039
	λ=0.4	0.330	0.357	0.415	0.221	0.243	0.292	0.145	0.160	0.196	0.092	0.102	0.126	0.054	0.060	0.076	0.025	0.029	0.039
	λ=0.5	0.320	0.346	0.401	0.218	0.239	0.284	0.144	0.159	0.193	0.092	0.102	0.125	0.054	0.060	0.076	0.025	0.029	0.039

Figure 4.4: TABLE 4.2b

4.3.1 A modified statistic

Here we restrict our attention to the cases $i = 1$ and $i = 2$ of the deterministic component $\mathbf{x}_t^i(\lambda)$ in (4.13). Denote by $\mathbf{e}_t^i(\lambda)$ the residuals from the OLS regression of $\mathbf{y}_t$ on $\mathbf{x}_t^i(\lambda)$, $i = 1, 2$. Note that the orthogonality conditions for the residuals allow us to write ($i = 1, 2$)

$$\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\mathbf{e}_s^i(\lambda)\right)\left(\sum_{s=1}^{t}\mathbf{e}_s^i(\lambda)\right)' =$$
$$= \sum_{t=1}^{T_1}\left(\sum_{s=1}^{t}\mathbf{e}_s^i(\lambda)\right)\left(\sum_{s=1}^{t}\mathbf{e}_s^i(\lambda)\right)' + \sum_{t=T_1+1}^{T}\left(\sum_{s=T_1+1}^{t}\mathbf{e}_s^i(\lambda)\right)\left(\sum_{s=T_1+1}^{t}\mathbf{e}_s^i(\lambda)\right)',$$

as the sum of residuals in each of the two subsamples $\{1,...,T_1\}$ and $\{T_1+1, ..., T\}$ is zero. The idea is then essentially to take the sum of the two statistics (4.9) applied to each subsample. In particular, we consider the statistic $\xi_N^{*i}(\lambda)$ defined as

$$\xi_N^{*i}(\lambda) = tr\left[\widehat{\mathbf{\Omega}}(m)^{-1}\left(\mathbf{C}^1(\lambda) + \mathbf{C}^2(\lambda)\right)\right], \qquad i = 1, 2, \tag{4.16}$$

where

$$\mathbf{C}^1(\lambda) = {T_1}^{-2}\sum_{t=1}^{T_1}\left(\sum_{s=1}^{t}\mathbf{e}_s^i(\lambda)\right)\left(\sum_{s=1}^{t}\mathbf{e}_s^i(\lambda)\right)',$$
$$\mathbf{C}^2(\lambda) = (T - T_1)^{-2}\sum_{t=T_1+1}^{T}\left(\sum_{s=T_1+1}^{t}\mathbf{e}_s^i(\lambda)\right)\left(\sum_{s=T_1+1}^{t}\mathbf{e}_s^i(\lambda)\right)',$$

and $\widehat{\mathbf{\Omega}}(m)$ defined as in section 4.2.

Proposition 15 *Let $\boldsymbol{y}_t$ be generated by the model (4.5)-(4.7) under assumptions 4.2-4.3 and with the regressors defined by (4.13). Then under $H_0 : \mathbf{\Sigma}_\eta = 0$*

$$\xi_N^{*i}(\lambda) \xrightarrow{d} \int_0^1 \mathbf{B}^i(r)'\mathbf{B}^i(r)dr, \qquad i = 1, 2, \tag{4.17}$$

where $\mathbf{B}^1(r)$ and $\mathbf{B}^2(r)$ are respectively a standard vector Brownian bridge and a second level standard vector Brownian bridge of dimension $2N$.

The proof is in contained in section 4.7 at the end of the chapter. The random variable to which $\xi_N^{*i}(\lambda)$ converge has then a Cramér-von Mises distribution with

$2N$ degrees of freedom. As said before, percentage points from this distribution are tabulated in Nyblom (1989), Canova and Hansen (1995), Nyblom and Harvey (2000), and also reported in table 3.2 of the previous chapter. They can be used to construct an alternative test for $H_0 : \mathbf{\Sigma}_\eta = 0$ against $H_1 : rank(\mathbf{\Sigma}_\eta) > 0$ in a model with breaking trends. This test coincides with the LBI test (4.15) when $\lambda = 0.5$, since $\xi_N^{*i}(0.5) = 0.25\xi_N^i(0.5)$.[1] For other values of λ the test is of course consistent; furthermore the simulation results in section 3.3, relating to the univariate version of the test, show that it suffers only a small loss in power compared with the LBI test.

The attraction of the test (4.16) is that it can be easily generalized to the case of two or more breaks in the deterministic trend. Let there be two structural breaks at time $T_1 = \lambda_1 T$ and $T_2 = \lambda_2 T$. Then we can base the test for the presence of stochastic trends on the statistic

$$\xi_N^{*i}(\lambda_1, \lambda_2) = tr\left[\widehat{\mathbf{\Omega}}(m)^{-1}\left(\mathbf{C}^1(\lambda_1, \lambda_2) + \mathbf{C}^2(\lambda_1, \lambda_2) + \mathbf{C}^3(\lambda_1, \lambda_2)\right)\right], \quad i = 1, 2,$$

which will be defined by generalizing (4.16) in an obvious way. Its asymptotic distribution is Cramér-von Mises with $3N$ degrees of freedom.

Another case can be covered by the statistic $\xi_N^{*i}(\lambda_1, \lambda_2)$, namely when some of the N series break at the point λ_1 and some at the point λ_2. Then constructing the statistic by forcing all the N series to have two breakpoints gives rise to a valid test, although not efficient. On the other hand, for this situation no alternative procedures seem available in the literature.

4.4 Testing for common trends

In this section we consider the model (4.5)-(4.7) under the null hypothesis $H_0 : rank(\mathbf{\Sigma}_\eta) = K$, with $0 \leq K < N$, which corresponds to nonstationarity with K stochastic trends. The alternative hypothesis is $H_1 : rank(\mathbf{\Sigma}_\eta) > K$.

The existence of K common trends implies the existence of $R = N - K$ cointegration relationships, i.e. there is a RxN (full rank) matrix $\mathbf{A}$ such that

[1] From the results in section 4.7 (see also section 3.3) it follows that for $i = 1, 2$ the asymptotic distribution of $\xi_N^i(\lambda)$ can be also represented as the distribution of a weighted sum of two Cramer-von Mises random variables with N degrees of freedom, with weights given by λ^2 and $(1-\lambda)^2$.

$\mathbf{A}\mathbf{y}_t$ is stationary. If we knew $\mathbf{A}$, we could test H_0 against H_1 by applying the statistic (4.9) to $\mathbf{A}\mathbf{y}_t$. But since $\mathbf{A}$ is unknown, one way of proceeding is to take the minimum of (4.9) over the set of the RxN matrices $\mathbf{A}$. The minimum is given by the sum of the R smallest eigenvalues of $\widehat{\mathbf{\Omega}}(m)^{-1}\mathbf{C}$; see Rao (1973) and Nyblom and Harvey (2000).

The statistic we use is then

$$\xi_{K,N} = \sum_{j=K+1}^{N} \ell_j, \tag{4.18}$$

where $\ell_1 \geq \ell_2 \geq ... \geq \ell_N \geq 0$ are the N ordered eigenvalues of $\widehat{\mathbf{\Omega}}(m)^{-1}\mathbf{C}$. Note that $\xi_{0,N}$ corresponds to the statistic (4.9).

Proposition 16 *Let $\boldsymbol{y}_t$ be generated by the model (4.5)-(4.7) under assumptions 4.1-4.3. Then under $H_0 : rank(\mathbf{\Sigma}_\eta) = K$*

$$\xi_{K,N} \xrightarrow{d} tr\left(\mathbf{C}^*_{22} - \mathbf{C}^{*\prime}_{12}\mathbf{C}^{*-1}_{11}\mathbf{C}^*_{12}\right), \tag{4.19}$$

where the stochastic matrices $\mathbf{C}^*_{ij}$, $i,j = 1,2$, are defined as

$$\mathbf{C}^*_{11} := \int_0^1 \left(\int_0^r \mathbf{W}^X(s)ds\right)\left(\int_0^r \mathbf{W}^X(s)ds\right)' dr,$$

$$\mathbf{C}^*_{12} := \int_0^1 \left(\int_0^r \mathbf{W}^X(s)ds\right)\mathbf{B}^X(r)'dr,$$

$$\mathbf{C}^*_{22} := \int_0^1 \mathbf{B}^X(r)\mathbf{B}^X(r)'dr,$$

where $\mathbf{W}^X(r)$ is a K-dimensional vector process as defined in proposition 14 and $\mathbf{B}^X(r)$ is a R-dimensional vector process as defined in proposition 13.

The proof is in section 4.7 at the end of the chapter. Under $H_1 : rank(\mathbf{\Sigma}_\eta) > K$, the statistic $\xi_{K,N}$ diverges to infinity as it contains at least one eigenvalue that is $O_p(T/m)$; see section 4.7.

For the case of breaking trends, i.e. when the regressors are defined by (4.13), the upper tail percentage points of the distribution of $\xi_{K,N}$ under H_0 are provided in Tables 4.1, 4.2, 4.2a, 4.2b for a range of values of the breakpoint parameter λ. These are obtained simulating the stochastic processes involved using a sample size of 1,000 and 100,000 replications.

These tables can then be used to construct an asymptotically valid test for $H_0 : rank(\Sigma_\eta) = K$ against $H_1 : rank(\Sigma_\eta) > K$ in the breaking trends case. Note that this is also a test on the dimension of the cointegration space, i.e. a test of

$$H_0 : \text{there are } R \text{ cointegration relationships}$$

against

$$H_1 : \text{there less than } R \text{ cointegration relationships.}$$

This is to be contrasted with the Johansen-type tests of Inoue (1999), where under the alternative hypothesis there are more than R cointegration relationships (i.e. under the null hypothesis the model is "more nonstationary"). Then Inoue's tests are also not directly comparable with ours since they are obtained by taking the supremum of Johansen's statistics with respect to the breakpoint location, i.e. they are unconditional to the position of the break. Another difference is that, unlike here, they require fitting a set of statistical models to the data, namely fitting a vector autoregression for each possible breakpoint location.

Unfortunately, for $K > 0$ the test cannot be modified along the lines of the previous section to yield a statistic whose asymptotic distribution is free of λ. Finally note that this model of breaking trends implies that in general the long run equilibrium relation has been subject to a shift at time T_1, as it can well be the case; see Gregory and Hansen (1992a,b) and Hao (1996) for a direct test of cointegration under regime shifts.

4.5 Seasonal effects and weakly dependent exogenous regressors

In this section we will show that augmenting the model (4.5)-(4.7) by including deterministic seasonality and/or weakly dependent exogenous regressors does not affect the asymptotic distributions of the test statistics of the previous sections. For simplicity we only consider the test for nonstationarity of section 4.2.

Replace equation (4.5) by

$$\mathbf{y}_t = \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\gamma}'\mathbf{z}_t + \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \tag{4.20}$$

where $\mathbf{z}_t$ are additional regressors and $\boldsymbol{\gamma}$ the corresponding coefficients.

Assumption 4.4. Either [A] or [B] below holds.

[A] $\mathbf{z}_t$ is a zero mean second order stationary process such that:

[A1] $\sum_{t=1}^{\infty} \|\mathrm{E}(\mathbf{z}_t\mathbf{z}_1')\| < \infty$, [A2] $\mathrm{E}(\mathbf{z}_t\boldsymbol{\varepsilon}_t') = \mathbf{0}$,

[A3] $\lim T^{-1}\mathrm{E}\left(\sum_{t=1}^{T} \mathbf{z}_t\varepsilon_{it}\right)\left(\sum_{t=1}^{T} \mathbf{z}_t\varepsilon_{it}\right)' < \infty$, $i = 1, ..., N$.

[B] $\mathbf{z}_t = (z_{1t}, ..., z_{s-1,t})'$ is a set of $s-1$ deterministic seasonal dummy variables, defined by

$$z_{ht} = \begin{cases} 1 & t = h + ns, \\ 0 & t \neq h + ns, \\ -1 & t = ns, \end{cases} \tag{4.21}$$

for $n = 0, 1, 2, ...$ Furthermore the function $\mathbf{x}(r)$ of assumption 4.1 is of bounded variation, i.e. there exists $M < \infty$ such that for every finite partition of the unit interval $0 = r_0 < r_1 < ... < r_n = 1$, $\sum_{i=1}^{n} \|\mathbf{x}(r_i) - \mathbf{x}(r_{i-1})\| < M$.

In assumption 4.4[A] $\mathbf{z}_t$ is a weakly dependent process (in second order sense) since 4.4[A1] implies a finite spectrum at the origin. 4[A2] is an exogeneity condition. The zero mean assumption is innocuous since an intercept is included among the other regressors $\mathbf{x}_t$.

The seasonal dummies of assumption 4[B] sum to zero over s periods, so they represent the relative seasonal deviation from the common mean. A reparametrization of the seasonal effect could be used. The bounded variation condition is of course satisfied in the breaking trends case.

Let the statistic ξ_N of section 4.2, equation (4.9), be constructed using the residuals from regressing $\mathbf{y}_t$ on $(\mathbf{x}_t', \mathbf{z}_t')'$. Then the following proposition holds.

Proposition 17 *Let $\mathbf{y}_t$ be generated by the model (4.20),(4.6)-(4.7) under assumptions 4.1-4.4. Then, under $H_0 : \boldsymbol{\Sigma}_\eta = 0$, the asymptotic distribution of ξ_N is the one defined in proposition 13.*

The proof is in section 4.7 at the end of the chapter. Proposition 17 implies that we can use the tests of sections 4.2, 4.3 to test for the presence of stochastic trends also in models containing a deterministic seasonal component and/or weakly dependent regressors. The same result carries over to the common trend test of section 4.4. The proof is not provided but follows similar lines.

When the seasonality is stochastic and possibly there are seasonal unit roots, the strategy advocated by Harvey and Streibel (1997) can be applied. The idea is described in section 3.5.2 as a way to account for serial correlation in ϵ_t, but it can be extended to deal with stochastic seasonality (by introducing it in the state space model to be estimated).

The case of I(1) regressors $\mathbf{z}_t$ is examined in Choi and Ahn (1995). They consider testing for stationarity of the errors in multiple equations with integrated variables. In their framework stationarity of the errors corresponds to cointegration between regressands and regressors.

4.6 Examples

The use of the tests is illustrated with two examples, one using UK macroeconomic data and the other using data on road casualties in Great Britain. The last dataset was used by Harvey and Durbin (1986) to study the effect of the introduction of the seat belt law and is provided with the program STAMP 5.0 of Koopman et al. (1995).

Figures 4.5-4.7 plot the logarithms of UK gross domestic product, consumption and investment for the period 1960-1990, together with a fitted deterministic trend. The data are quarterly, seasonal adjusted and at constant 1990 prices. The source is the Central Statistical Office.

We have assumed that there was an exogenous structural break around 1979-1980; the plot of the series, the investment in particular, seems to sustain the assumption. The exogeneity of the break may be due to the following two reasons: first the second oil shock of 1979-1980 may have caused a drop in the series and secondly 1979 coincides with the election of a right wing government with Margaret Thatcher as Prime Minister which may have determined the

Figure 4.5: UK GDP with fitted trend

Figure 4.6: UK Consumption and fitted trend

Figure 4.7: UK Investment and fitted trend

higher growth rates of investment thereafter. Thus we have chosen to fit a broken linear trend to the data, with the break occurring in the second quarter of 1980.[2]

We consider the case of a break in both level and slope and we apply the common trend test of section 4.4 to the trivariate series of (log) gdp, consumption and investment. The results are displayed in table 4.3, for various values of the lag length parameter m. The breakpoint is at the second quarter of 1980, corresponding to a value for λ of 0.66. Note that the statistic for $K = 0$ is equivalent to the nonstationarity test statistic of section 4.3, case 2. The critical values for the case of no break are taken from Nyblom and Harvey (2000), those for the case of structural break are taken from table 4.2, N=3 and λ=0.3.

From standard macroeconomic arguments, we would expect the existence of one stochastic trend among the variables, probably representing the effect of technological progress. However table 4.3 shows that the null hypothesis of one common trend ($K = 1$) is rejected at 5% significance level when we don't take into account the structural break of 1979-1980. Indeed, even the hypothesis of two trends seems to be rejected. On the contrary, fitting a broken deterministic

[2] Of course there might have been other events that, on a priori grounds, could have determined a break in the series. However, since the purpose of this section is only to illustrate the use of the tests, we don't go deep into the issue of exogeneity of the breakpoint.

trend to the data results in the acceptance of $H_0 : K = 1$ at 10% level of significance. The result is obtained for $m \geq 4$; note also the little adverse effect of the KPSS correction on the value of the statistic as m grows.

Table 4.3

Common trend test for UK quarterly series of Y,C,I (1960-1990)

		$m=1$	$m=2$	$m=4$	$m=8$	$m=11$	**10%**	**5%**
	K=0	2.043	1.409	0.896	0.558	0.457	0.296	0.332
No break	K=1	0.845	0.584	0.374	0.238	0.199	0.151	0.180
	K=2	0.180	0.129	0.087	0.062	0.057	0.061	0.075
	K=0	0.694	0.494	0.331	0.228	0.207	0.163	0.184
Break 80.2	K=1	0.159	0.116	0.082	0.066	0.074	0.088	0.102
($\lambda = 0.66$)	K=2	0.038	0.030	0.024	0.025	0.032	0.038	0.046

Now consider the series of the logarithm of front and rear passengers killed or seriously injured (KSI) in road accidents, that are displayed in figure 4.8. The data are monthly, not seasonally adjusted, and cover the period January 1969 to December 1984. These data were used by Harvey and Durbin (1986) to assess the effect of the seat belt law, which made compulsory the wearing of seat belts for the front seat passengers after January 31 1983. Clearly, there is an exogenous structural break in the series of front seat passengers but not in the series of rear passengers. As explained in section 4.3, it is possible to apply our tests by forcing both series to break at February 1983.

Harvey and Durbin (1986) show that a reasonable univariate time series model for the KSI series is the simple random walk plus noise and seasonal component, with seasonality being fixed. Given the nature of the data, it would seem plausible that in a multivariate model the random walk component has dimension 1.

Table 4.4 shows the results of applying the common trend test of section 4.4

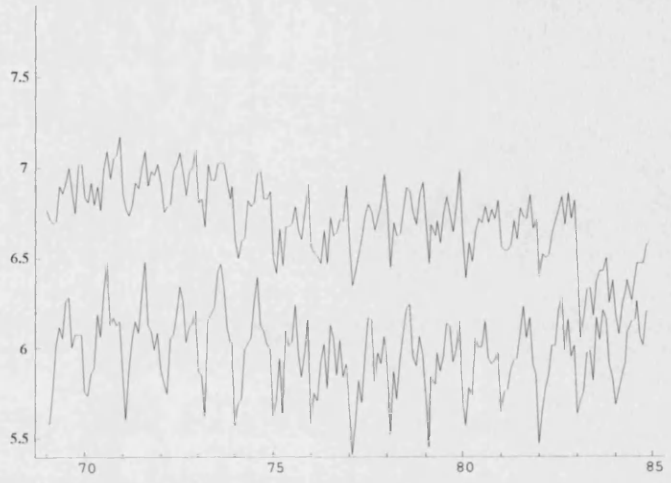

Figure 4.8: Front and rear passengers KSI, 1969-1984.

to the bivariate series of front and rear passengers. A slope component is not included. The breakpoint parameter λ takes the value 0.88, corresponding to February 1983. The statistic is computed including as regressors also the set of 11 dummy variables to account for seasonality. As explained in section 4.5 the asymptotic distribution is not affected by that.

If we do not take into account of the break (first two rows of the table), we end up rejecting the null hypothesis of one common trend, $K = 1$, at 5% level of significance even for very large values of the lag truncation parameter m ($m = 14$ corresponds to the square root of the sample size). On the other hand, forcing both series to break at 1983.2 results into accepting the hypothesis of one common trend even for $m = 1$. Note that since each series can be modelled as a univariate random walk plus noise, considering small values of m seems appropriate. Finally, the modified test of nonstationarity (fifth row of the table) confirms the finding that $K > 0$.

Table 4.4

Results of the test for KSI front and rear passengers (1969-1984)

		$m=0$	$m=1$	$m=2$	$m=5$	$m=14$	**10%**	**5%**
No break	K=0	13.002	7.210	5.081	2.785	1.535	0.596	0.746
	K=1	1.121	0.855	0.694	0.454	0.274	0.156	0.212
Break 83.2	K=0	7.992	4.640	3.339	1.889	0.881	0.494	0.608
($\lambda = 0.88$)	K=1	0.184	0.171	0.161	0.139	0.107	0.134	0.181
Modif. test		10.667	6.255	4.537	2.612	1.257	0.607	0.748

4.7 Proofs of this chapter's propositions

PROOF OF PROPOSITION 13

Under H_0 and assumptions 4.1-4.3, $\widehat{\mathbf{\Omega}}(m) \xrightarrow{p} \mathbf{\Omega}$. By assumption 4.2, the long run variance of the disturbances $\mathbf{\Omega}$ is finite and of full rank. Then there exists a nonsingular matrix $\mathbf{P}$ such that $\mathbf{P\Omega P}' = \mathbf{I}$ and $\mathbf{P\Sigma}_\eta\mathbf{P}' = diag(q_1, q_2, ..., q_N)$, where the q_j's are the N roots of $|\mathbf{\Sigma}_\eta - q\mathbf{\Omega}| = 0$; see Rao (1973, p.41).

Since the test statistic ξ_N is invariant to premultiplying the observations $\mathbf{y}_t$ by an arbitrary nonsingular NxN matrix $\mathbf{P}$, without loss of generality we can restrict to the case of $\mathbf{\Omega} = \mathbf{I}$ and $\mathbf{\Sigma}_\eta = diag(q_1, q_2, ..., q_N)$.

Then $\xi_N = tr\left((\mathbf{I}+o_p(1))^{-1}\mathbf{C}\right)$.

Since $\mathbf{e}_t = \boldsymbol{\varepsilon}_t - \sum_{t=1}^{T}\boldsymbol{\varepsilon}_t\mathbf{x}_t'\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{x}_t$, it follows that under assumptions 4.1-4.2

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Tr]}\mathbf{e}_t \Rightarrow \mathbf{W}(r) - \left(\int_0^1 \mathbf{x}(r)d\mathbf{W}(r)'\right)'\left(\int_0^1 \mathbf{x}(r)\mathbf{x}(r)'dr\right)^{-1}\int_0^r \mathbf{x}(s)ds,\ r \in [0,1], \tag{4.22}$$

where $\mathbf{W}(r)$ is a standard vector Wiener process of dimension N. We denote the process at the right hand side of (4.22) as $\mathbf{B}^X(r)$, and we call it generalized Brownian bridge. Thus, by the continuous mapping theorem and using the definition of $\mathbf{C}$, $\xi_N \xrightarrow{d} \int_0^1 \mathbf{B}^X(r)'\mathbf{B}^X(r)dr$.

PROOF OF PROPOSITION 14

Let assumptions 4.1-4.3 hold. Without loss of generality we again restrict to the case of $\mathbf{\Omega} = \mathbf{I}$ and $\mathbf{\Sigma}_\eta = diag(q_1, ..., q_N)$. The hypothesis $H_1 : rank(\mathbf{\Sigma}_\eta) = K$ can be equivalently formulated as $H_1 : q_j > 0$ for $j = 1, ..., K$ and $q_j = 0$ for $j = K+1, ..., N$.

Consider the OLS residuals for the j-equation under H_1,

$$e_{jt} = (\mu_{jt} + \varepsilon_{jt}) - \mathbf{x}_t' \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t (\mu_{jt} + \varepsilon_{jt}), \qquad j = 1, ..., K,$$

and

$$e_{jt} = \varepsilon_{jt} - \mathbf{x}_t' \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_{jt}, \qquad j = K+1, ..., N,$$

since $\mu_{jt} = 0$ for $j = K+1, ..., N$, $t = 1, ..., T$.

Then, the following weak convergence results hold:

$$T^{-1/2} e_{j,[Tr]} \Rightarrow q_j W_j^X(r), \qquad j = 1, ..., K, \tag{4.23}$$

$$T^{-3/2} \sum_{t=1}^{[Tr]} e_{jt} \Rightarrow q_j \int_0^r W_j^X(s)ds, \qquad j = 1, ..., K, \tag{4.24}$$

$$T^{-1/2} \sum_{t=1}^{[Tr]} e_{jt} \Rightarrow B_j^X(r), \qquad j = K+1, ..., N, \tag{4.25}$$

where the processes $W_j^X(r)$ and $B_j^X(r)$, $r \in [0, 1]$, are the one-dimensional version of the processes defined in the statements of proposition 14 and 13, and are uncorrelated across j (and with each other). The results (4.23)-(4.25) are obtained applying the invariance principle to the partial sums of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ and using assumption 4.1 of non-stochastic regressors.

Denote by c_{jh}, $\widehat{\omega}_{jh}$ the (j,h)-elements of the matrices $\mathbf{C}$, $\widehat{\mathbf{\Omega}}(m)$. Then we have the following further asymptotic results:

$$T^{-2}c_{jh} \xrightarrow{d} (q_j q_h)^{1/2} \int_0^1 \left(\int_0^r W_j^X(s)ds \int_0^r W_h^X(s)ds \right) dr, \qquad j,h \leq K, \tag{4.26}$$

$$T^{-1}c_{jh} \xrightarrow{d} q_j^{1/2} \int_0^1 \left(\int_0^r W_j^X(s)ds \right) B_h^X(r)dr, \qquad j \leq K, h > K, \tag{4.27}$$

$$c_{jh} \xrightarrow{d} \int_0^1 B_j^X(r)B_h^X(r)dr, \qquad j,h > K, \tag{4.28}$$

$$(mT)^{-1}\widehat{\omega}_{jh} \xrightarrow{d} (q_j q_h)^{1/2} \int_0^1 W_j^X(r)W_h^X(r)dr, \qquad j,h \leq K, \tag{4.29}$$

$$\widehat{\omega}_{jh} = O_p(m), \qquad j \leq K, h > K, \tag{4.30}$$

$$\widehat{\omega}_{jh} \xrightarrow{p} 1(j = h), \qquad j,h > K. \tag{4.31}$$

(4.26)-(4.28) come directly from application of the continuous mapping theorem, (4.29) corresponds to equation (23) of Kwiatowski et al. (1992), (4.31) holds because of the consistency of the long run variance estimator, and (4.30) because $T^{-1}\sum_{t=\tau+1}^{T} e_{tj}e_{t-\tau,h} = O_p(1)$ uniformly in $\tau = -m, -m+1, ..., m$.

Our test statistic is defined as the trace of $\widehat{\mathbf{\Omega}}(m)^{-1}\mathbf{C}$, i.e. the sum of its eigenvalues. Let $\ell_1 \geq \ell_2 \geq ... \geq \ell_N \geq 0$ be the N ordered eigenvalues of $\widehat{\mathbf{\Omega}}(m)^{-1}\mathbf{C}$. Using the results (4.26)-(4.31) we will show that, under $H_1 : rank(\mathbf{\Sigma}_\eta) = K$, K eigenvalues are $O_p(T/m)$ and $N - K$ eigenvalues are $O_p(1)$. Thus asymptotically the distribution of the statistic coincides with the distribution of the sum of those K asymptotically bigger eigenvalues; see also Nyblom and Harvey (2000).

Partition $\widehat{\mathbf{\Omega}}(m)$ and $\mathbf{C}$ as

$$\widehat{\mathbf{\Omega}}(m) = \begin{bmatrix} \widehat{\mathbf{\Omega}}_{11} & \widehat{\mathbf{\Omega}}_{12} \\ \widehat{\mathbf{\Omega}}_{12'} & \widehat{\mathbf{\Omega}}_{22} \end{bmatrix} \text{ and } \mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12'} & \mathbf{C}_{22} \end{bmatrix},$$

where $\widehat{\mathbf{\Omega}}_{11}$ and $\mathbf{C}_{11}$ are KxK. The eigenvalues of $\widehat{\mathbf{\Omega}}(m)^{-1}\mathbf{C}$ solves the determinantal equation

$$0 = \left| \mathbf{C} - \ell_j \widehat{\mathbf{\Omega}}(m) \right|$$

$$= \left|\mathbf{C}_{11} - \ell_j \widehat{\mathbf{\Omega}}_{11}\right| \times \left|\mathbf{C}_{22} - \ell_j \widehat{\mathbf{\Omega}}_{22} - \left(\mathbf{C}_{12} - \ell_j \widehat{\mathbf{\Omega}}_{12}\right)' \left(\mathbf{C}_{11} - \ell_j \widehat{\mathbf{\Omega}}_{11}\right)^{-1} \left(\mathbf{C}_{12} - \ell_j \widehat{\mathbf{\Omega}}_{12}\right)\right|,$$

see Rao (1973, p.32). Then using (4.26)-(4.31) we see that the roots of the first determinant are $O_p(T/m)$, whereas the roots of the second determinant are $O_p(1)$ and asymptotically equivalent to the eigenvalues of $(\mathbf{C}_{22} - \mathbf{C}_{12}'\mathbf{C}_{11}^{-1}\mathbf{C}_{12})$; see Nyblom and Harvey (2000) for further details.

Therefore $(m/T)\,\xi_N \xrightarrow{d} tr\left(\mathbf{\Omega}_{11}^{*-1}\mathbf{C}_{11}^{*}\right)$, where

$$\mathbf{C}_{11}^{*} := \int_0^1 \left(\int_0^r \mathbf{W}^X(s)ds\right)\left(\int_0^r \mathbf{W}^X(s)ds\right)' dr,$$

$$\mathbf{\Omega}_{11}^{*} := \int_0^1 \mathbf{W}^X(r)\mathbf{W}^X(r)'dr,$$

with $\mathbf{W}^X(r)$ being the K-dimensional vector process defined in the statement of the proposition.

PROOF OF PROPOSITION 15

Let $\mathbf{z}_t = 1$ for the case $i = 1$ and $\mathbf{z}_t = (1, t)'$ for the case $i = 2$, and drop the superscript i for simplicity. The model can be equivalently parametrized using as regressors $\mathbf{x}_t(\lambda) = (\mathbf{x}_{1t}(\lambda)', \mathbf{x}_{2t}(\lambda)')'$, where $\mathbf{x}_{1t}(\lambda) = \mathbf{z}_t \cdot 1\,(t \leq \lambda T)$, $\mathbf{x}_{2t} = \mathbf{z}_t \cdot 1\,(t > \lambda T)$. Then the model is orthogonal and to obtain the OLS residuals we can consider the two subsamples $\{1,...,T_1\}$,$\{T_1+1,...,T\}$ separately and in each of them regress $\mathbf{y}_t$ on $\mathbf{z}_t$.

Again, without loss of generality and under assumption 4.2, we can restrict to the case of $\mathbf{\Omega} = \mathbf{I}$ and $\mathbf{\Sigma}_\eta = diag(q_1, q_2, ..., q_N)$. Then, under assumptions 4.2-4.3, we can write $\xi_N^{i*}(\lambda) = tr\left[(\mathbf{I} + op(1))^{-1}\left(\mathbf{C}^1(\lambda) + \mathbf{C}^2(\lambda)\right)\right]$. Under H_0 the invariance principle can be applied in each subsample, yielding

$$T_1^{-1/2} \sum_{t=1}^{[T_1 r]} \mathbf{e}_t(\lambda) \Rightarrow \mathbf{B}_1^i(r),$$

and

$$(T - T_1)^{-1/2} \sum_{t=1}^{[(T-T_1)r]} \mathbf{e}_{t+T_1}(\lambda) \Rightarrow \mathbf{B}_2^i(r), \qquad j = 1, 2, ..., N,$$

where $\mathbf{B}_1^i(r)$, $\mathbf{B}_2^i(r)$ are N-dimensional independent standard vector Brownian bridges for $i = 1$ and second level Brownian bridges for $i = 2$. Independence holds because the two Brownian bridges are the limit of partial sum processes containing non-overlapping subsets of disturbances.

By the continuous mapping theorem it then follows that $\xi_N^{i*}(\lambda)$ converges to the sum of two independent random variables, each with a Cramér-von Mises distribution with N degrees of freedom. The limiting distribution is then Cramér-von Mises with $2N$ degrees of freedom; see section 3.3 for details on the additivity property of Cramér-von Mises random variables.

PROOF OF PROPOSITION 16

From the proof of proposition 14 we know that under assumptions 4.1-4.3 and under $rank(\mathbf{\Sigma}_\eta) = K$, the R smallest eigenvalues of $\widehat{\mathbf{\Omega}}^{-1}\mathbf{C}$ are asymptotically equivalent to the eigenvalues of $\mathbf{C}_{22} - \mathbf{C}_{12}'\mathbf{C}_{11}^{-1}\mathbf{C}_{12}$. Then, using the results (4.26)-(4.28),

$$\begin{aligned} T^{-2}\mathbf{C}_{11} &\xrightarrow{d} \mathbf{Q}^{\frac{1}{2}}\mathbf{C}_{11}^*\mathbf{Q}^{\frac{1}{2}}, \\ T^{-1}\mathbf{C}_{12} &\xrightarrow{d} \mathbf{Q}^{\frac{1}{2}}\mathbf{C}_{12}^*, \\ \mathbf{C}_{22} &\xrightarrow{d} \mathbf{C}_{22}^*, \end{aligned}$$

where $\mathbf{C}_{ij}^*$, $i = 1, 2$, are defined in the statement of proposition 16 and $\mathbf{Q}^{\frac{1}{2}} = diag\left(q_1^{\frac{1}{2}}, ..., q_K^{\frac{1}{2}}\right)$. Thus, by the continuous mapping theorem,

$$\xi_{K,N} \xrightarrow{d} tr\left(\mathbf{C}_{22}^* - \mathbf{C}_{12}^{*\prime}\mathbf{C}_{11}^{*-1}\mathbf{C}_{12}^*\right).$$

PROOF OF PROPOSITION 17

Without loss of generality we assume $\boldsymbol{\delta}_T = \mathbf{I}$ in assumption 4.1 and $\mathbf{\Omega} = \mathbf{I}$ in assumption 4.2. The standardized partial sums of OLS residuals can be written as

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Tr]}\mathbf{e}_t = T^{-\frac{1}{2}}\sum_{t=1}^{[Tr]}\boldsymbol{\epsilon}_t - T^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'T^{-1}\sum_{t=1}^{[Tr]}\mathbf{x}_t - T^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'T^{-1}\sum_{t=1}^{[Tr]}\mathbf{z}_t, \quad r \in [0,1],$$

where

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} \\ \widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t' & \sum_{t=1}^{T} \mathbf{x}_t\mathbf{z}_t' \\ \sum_{t=1}^{T} \mathbf{z}_t\mathbf{x}_t' & \sum_{t=1}^{T} \mathbf{z}_t\mathbf{z}_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} \mathbf{x}_t\boldsymbol{\varepsilon}_t' \\ \sum_{t=1}^{T} \mathbf{z}_t\boldsymbol{\varepsilon}_t' \end{pmatrix}.$$

If we show that

$$T^{-1}\sum_{t=1}^{T} \mathbf{x}_t\mathbf{z}_t' \xrightarrow{p} \mathbf{0} \tag{4.32}$$

and

$$T^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'T^{-1}\sum_{t=1}^{[Tr]} \mathbf{z}_t \xrightarrow{p} \mathbf{0} \text{ uniformly in } r \in [0,1], \tag{4.33}$$

then $T^{-\frac{1}{2}}\sum_{t=1}^{[Tr]} \mathbf{e}_t \Rightarrow \mathbf{B}^X(r)$, the generalized Brownian bridge defined in proposition 13, and thus proposition 17 follows by applying the continuous mapping theorem.

Consider assumption 4[A] first. Condition (4.32) above holds because for each z_{it} element of $\mathbf{z}_t$ and x_{jt} element of $\mathbf{x}_t$ we have

$$\mathrm{E}\left|T^{-1}\sum_{t=1}^{T} z_{it}x_{jt}\right|^2 \leq \max_{t,s} |x_{jt}x_{js}|\, T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T} |\mathrm{E}(z_{it}z_{is})| \to 0$$

by assumption 4[A1] and by recalling that without loss of generality x_{it} is assumed bounded throughout this proof. Then from assumption 4[A] it follows that $T^{-1}\sum_{t=1}^{T} \mathbf{z}_t\mathbf{z}_t' = O_p(1)$, $T^{-\frac{1}{2}}\sum_{t=1}^{T} \mathbf{z}_t\boldsymbol{\varepsilon}_t' = O_p(1)$ and $T^{-1}\sum_{t=1}^{[Tr]} \mathbf{z}_t = o_p(1)$ uniformly in $r \in [0,1]$; thus condition (4.33) holds too.

Now consider assumption 4[B]. To check condition (4.32) first note that $\sum_{t=1}^{T} z_{ht} = 1(T \neq ns,\ n = 1,2,...)$, for each $h = 1,...,s-1$. Then, using the summation by parts argument,

$$\begin{aligned} \left\|T^{-1}\sum_{t=1}^{T} \mathbf{x}_t z_{ht}\right\| &\leq \left\|T^{-1}\sum_{t=1}^{T-1} (\mathbf{x}_{t+1}-\mathbf{x}_t)\sum_{j=1}^{t} \mathbf{z}_{hj}\right\| + \left\|T^{-1}\mathbf{x}_T\sum_{t=1}^{T} z_{ht}\right\| \\ &\leq T^{-1}\sum_{t=1}^{T-1} \|\mathbf{x}_{t+1}-\mathbf{x}_t\| + T^{-1}\|\mathbf{x}_T\| \\ &\to 0 \quad \text{as } T \to \infty \end{aligned}$$

by the assumption of bounded variation. Condition (4.33) holds since $T^{-1}\sum_{t=1}^{T} \mathbf{z}_t\mathbf{z}_t' \to s^{-1}\boldsymbol{\Sigma}$, where $[\boldsymbol{\Sigma}]_{hl} = 1 + 1(h=l)$, and clearly $T^{-\frac{3}{2}}\sum_{t=1}^{T} \mathbf{z}_t\boldsymbol{\varepsilon}_t' = o_p(1)$.

Chapter 5

Testing for stochastic seasonality

This chapter considers testing for the presence of a nonstationary seasonal component. The LBI test is obtained from theorem 6 of chapter 2. It essentially corresponds to the test proposed by Canova and Hansen (1995), hereafter CH test, where serial correlation is accounted for nonparametrically as in section 3.5.1. The aim of this chapter is to interpret the CH test in different ways so as to give more insight into its construction and application, to extend it in various directions and to compare its performance with that of a parametric test.

Section 5.1 shows how the correction for serial correlation can be set up in terms of the spectrum at seasonal frequencies and how the dummy variable form of the test can be used to test the stability of the relationship between different seasons.

Contrary to what is stated in Canova and Hansen (1995, p 238), it is shown in section 5.2 that the asymptotic distribution of the test statistic is not affected by the presence of a deterministic trend. Furthermore regressors with unit roots can be included provided they do not have seasonal unit roots. It is also shown that breaks in the trend leave the asymptotic distribution unaffected if they are correctly modelled by the inclusion of dummy variables. However, a modelled deterministic break in the seasonal pattern will affect the distribution. Section 5.3 shows how to construct a modified test statistic for stochastic seasonality, the asymptotic distribution of which is independent of the breakpoint location; the test extends the modification to the LBI test at frequency zero proposed in chapter 3.

Section 5.4 sets out the seasonal models which are normally used in a structural time series framework and shows how to construct parametric versions of the CH tests for these models. The evidence in Leybourne and McCabe (1994) and Harvey and Streibel (1997) suggests that a parametric approach will usually give tests with a higher power and more reliable size. Using autoregressive models, as in Caner (1998), is not appealing since a large number of lags will typically be needed to approximate a series with a slowly changing seasonal component; see Taylor (2000).

The parametric and nonparametric tests, with the breakpoint modification, are illustrated by an application to a quarterly series on UK marriages in section 5.5. The point about this example is that there is an identifiable break in the seasonal pattern because of a known change in policy. The modified test is trying to assess if it is necessary to allow for stochastic seasonality once the deterministic break has been accounted for by dummy variables.

Section 5.6 looks at testing for nonstationarity of a group of seasons. This issue is not addressed in CH. We set up the test and then suggest a class of partly periodic models for which such tests might make sense. The series on Italian industrial production is used as an illustration in section 5.7.

Section 5.8 briefly discusses corresponding results concerning the implications of trends and breaks for the test of Hylleberg, Engle, Granger and Yoo (1990).

Section 5.9 suggests a general test for seasonality. This takes the same form as CH, except that seasonal dummies are not fitted. The asymptotic distribution is given and an example presented.

Section 5.10 suggests a Canova-Hansen type test for trading day effects and section 5.11concludes.

5.1 The Canova-Hansen test

The Canova-Hansen test is based on the Gaussian unobserved components model

$$y_t = \mu + \gamma_t + \varepsilon_t, \qquad t = 1, ..., T \tag{5.1}$$

where μ is a constant, γ_t is a seasonal component of the form

$$\gamma_t = \mathbf{z}_t' \boldsymbol{\gamma}_t \tag{5.2}$$

where $\mathbf{z}_t$ and $\boldsymbol{\gamma}_t$ are vectors and ε_t is a linear stationary process, though it can be generalised to have heteroscedastic innovations.

Let s be the number of seasons. In the *trigonometric* form of the test $\mathbf{z}_t = \left(\mathbf{z}_{1t}', \mathbf{z}_{2t}', ..., z_{s/2,t}\right)'$ is an $(s-1) \times 1$ vector defined by $\mathbf{z}_{jt} = (\cos \lambda_j t, \sin \lambda_j t)'$ for $j < s/2$ and, when s is even, $z_{s/2,t} = \cos \lambda_{s/2} t$, where $\lambda_j = 2\pi j/s$, $j = 1, ..., [s/2]$, are the seasonal frequencies.

In the *dummy variable* form of the test, $\mathbf{z}_t$ is an $s \times 1$ vector with a one in the position corresponding to the current month and zeroes elsewhere.

When pre-multiplied by a full rank matrix $\mathbf{A}'$, with $k \leq s-1$ rows, the vector $\boldsymbol{\gamma}_t$ is assumed to follow a multivariate random walk under the alternative hypothesis, that is

$$\mathbf{A}'\boldsymbol{\gamma}_t = \mathbf{A}'\boldsymbol{\gamma}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\omega), \tag{5.3}$$

where the rank of Σ_ω is positive. The initial value, $\boldsymbol{\gamma}_0$, is fixed. The null hypothesis is that $\boldsymbol{\Sigma}_\omega = \mathbf{0}$ so the seasonal component specified by $\mathbf{A}'\boldsymbol{\gamma}_t$ is deterministic.

A complete test for nonstationary stochastic seasonality in the trigonometric formulation is obtained by setting $\mathbf{A}$ equal to the identity matrix. A test for nonstationarity at a particular seasonal frequency, or frequencies, is obtained by letting $\mathbf{A}$ select the appropriate elements in $\boldsymbol{\gamma}_t$.

In the dummy variable formulation of the complete test, $\mathbf{A}$ is an $s \times (s-1)$ matrix with a one and a minus one in each column and zeroes elsewhere. For example, with $s = 4$,

$$\mathbf{A}' = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \tag{5.4}$$

Thus each season is contrasted with the one before or after to see if they are co-integrated. This is not the only way of setting up the selection matrix. For example, Canova and Hansen (1995, p243) set $\mathbf{A}' = [\mathbf{I}_{s-1} \quad -\mathbf{i}_{s-1}]$, so each of the

first $s-1$ seasons is compared with season s. Section 5.7 explores situations in which contrasting fewer than the $s-1$ pairs of seasons is a sensible option. Note that testing for nonstationarity of one particular season, or group of seasons, is only appropriate within the framework of some kind of periodic model in which the observations in each month, or groups of months, are modelled separately. Similarly testing for nonstationarity of all seasons by setting $\mathbf{A}$ to an $s \times s$ identity matrix, is not usually appropriate since, as CH point out, it is normally desirable to separate out the trend and seasonals for purposes of modelling and testing.

CH show that the complete trigonometric and dummy variable tests are identical.

The simplest form of the CH tests is when ε_t is white noise. The locallly best invariant (LBI) test, which is also a (one-sided) LM test, for the presence of a stochastic trigonometric component at any one of the seasonal frequencies, λ_j, apart from the one at π, is

$$\omega_j = 2T^{-2}\widehat{\sigma}^{-2}\sum_{t=1}^{T}\left[\left(\sum_{i=1}^{t} e_i \cos\lambda_j i\right)^2 + \left(\sum_{i=1}^{t} e_i \sin\lambda_j i\right)^2\right], \quad j = 1,...,[(s-1)/2], \tag{5.5}$$

where $\widehat{\sigma}^2$ is the sample variance of the OLS residuals, $e_t, t = 1,..,T$, from a regression on the seasonal sines and cosines, $\mathbf{z}_t$, and a constant; (5.5) is obtained from theorem 6 of chapter 2 by assuming that under the alternative hypothesis the coefficients of the seasonal cycle at frequency λ_j follow a random walk[1] .

Canova and Hansen (1995) show that, under the null hypothesis, the asymptotic distribution of this statistic is generalized Cramér-von Mises with two degrees of freedom; see the next section. If s is even, the statistic at frequency

[1] The argument proceeds as follows. Let $z_t = \exp(-i\lambda t)$ be the seasonal cycle at frequency λ. Using the notation of chapter 2, the variance covariance matrix of the model is $\Omega(\theta) = \sigma^2(\mathrm{I}+\theta H(\lambda))$ where $H(\lambda)$ is a $T \times T$ matrix whose (t,s)-th element is $\exp(-i\lambda(t-s))\min(t,s)$. Then the LBI statistic is $e'H(\lambda)e/e'e$, where e are the OLS residuals. After taking the Choleski square root of $H(\lambda)$, the numerator of the statistic can be written as $\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\exp(-i\lambda s)e_s\right)^2$, from which the expression for ω_j given above is obtained.

π is

$$\omega_{s/2} = T^{-2}\widehat{\sigma}^{-2}\sum_{t=1}^{T}\left(\sum_{i=1}^{t} e_i(-1)^i\right)^2,$$

and this has an asymptotic distribution which is Cramér-von Mises with one degree of freedom . A joint test for the presence of stochastic trigonometric components at all seasonal frequencies is obtained by summing the individual test statistics, that is

$$\omega = \sum_{j=1}^{[s/2]} \omega_j$$

This statistic has an asymptotic distribution which is generalized Cramér-von Mises with $s-1$ degrees of freedom, denoted $CvM(s-1)$.

Canova and Hansen show how the above tests can be generalized to handle serial correlation and heteroscedasticity by making a KPSS-like correction; see the discussion in chapter 3. The test statistic is

$$\omega_A(m) = \frac{1}{T^2} tr\left(\left(\mathbf{A}'\widehat{\mathbf{\Omega}}\mathbf{A}\right)^{-1}\mathbf{A}'\sum_{t=1}^{T}\mathbf{S}_t\mathbf{S}_t'\mathbf{A}\right), \tag{5.6}$$

where $\mathbf{S}_t = \sum_{i=1}^{t}\mathbf{z}_i e_i$ and $\widehat{\mathbf{\Omega}}$ is a consistent estimator of the "long run variance" of $\mathbf{z}_t\varepsilon_t$, that is

$$\mathbf{\Omega} = \lim T^{-1}\mathrm{E}(\sum \mathbf{z}_t\varepsilon_t)(\sum \mathbf{z}_t\varepsilon_t)'.$$

A simple option for $\widehat{\mathbf{\Omega}}$ is

$$\widehat{\mathbf{\Omega}} = \sum_{\tau=-m}^{m} w(\tau, m)\widehat{\mathbf{\Gamma}}(\tau), \tag{5.7}$$

where $w(\tau, m)$ is a weighting function or kernel, such as $w(\tau, m) = 1-|\tau|/(m+1)$, and

$$\widehat{\mathbf{\Gamma}}(\tau) = T^{-1}\sum_{t=\tau+1}^{T}\mathbf{z}_t e_t e_{t-\tau}\mathbf{z}_{t-\tau}'$$

is the sample autocovariance of $\mathbf{z}_t e_t$ at lag τ. Consistency requires $m \to \infty$ at a slower rate than T; CH suggest $m^5/T = O(1)$. In their applications to quarterly post-war US macroeconomic time series, they take m to be four. Their Monte Carlo experiments use $m = 3$ and 5 for sample sizes 50 and 150 respectively.

They also include a lagged dependent variable in the model, though for the reasons given in Taylor (2000) we do not follow this route.

If there is no heteroscedasticity, the covariance matrix of what is now the stationary part of the model is diagonalised by the trigonometric terms. The need to correct for serial correlation can now be seen in terms of the spectrum at seasonal frequencies rather than at zero. (It is interesting to note that the partial sums in (5.5) are proportional to the periodogram ordinates, $I_t(\lambda_j) = (2\pi t)^{-1}\left|\sum_{s=1}^{t} e_s \exp(-i\lambda_j s)\right|^2$, for sample sizes $t = 1,..,T$, except insofar as a particular frequency, λ_j, is expressed as $2\pi k/t$, with k not necessarily an integer). Thus a valid test for nonstationarity at a seasonal frequency could be formulated as

$$\omega_j = 2T^{-2}\sum_{t=1}^{T}\left[\left(\sum_{i=1}^{t} e_i \cos\lambda_j i\right)^2 + \left(\sum_{i=1}^{t} e_i \sin\lambda_j i\right)^2\right] / \widehat{g}(\lambda_j),\ j = 1,...,[(s-1)/2], \tag{5.8}$$

and similarly for $\omega_{s/2}$, where $\widehat{g}(\lambda_j)$ is a nonparametric estimator of the spectral generating function (sgf) of the stationary part of the model, ε_t, at frequency λ_j. For example

$$\widehat{g}(\lambda_j) = \sum_{\tau=-m}^{m} w(\tau, m)\widehat{\gamma}(\tau)\cos\lambda_j\tau,$$

where $\widehat{\gamma}(\tau)$ is the sample autocovariance at lag τ.

The series in question may require differencing to make it satisfy the conditions for the CH test to be asymptotically valid. From the formulation of the dummy variable test based on (5.4), it is clear that carrying out the test using the residuals from a regression in first differences requires that $\mathbf{A}'\mathbf{S}_t$ be simply replaced by an $(s-1)\times 1$ vector in which each element is a partial sum of residuals in one of the seasons. One season is omitted. Apart from asymptotically negligible end effects, the tests are the same irrespective of which season is dropped and are also the same as applying the test to first differences in the way suggested by CH for levels. If no season is omitted, the test is for stochastic seasonality and a stochastic slope. Note that in many cases the first difference regression will simply amount to the estimation of seasonal means.

5.2 Deterministic trends, trend-breaks and integrated regressors

Suppose that (5.1) now includes two sets of regressors, one deterministic with a constant term, denoted as $\mathbf{x}_{1t}$, and the other stochastic and integrated of order one, denoted as $\mathbf{x}_{2t}$; that is consider the model

$$\begin{aligned} y_t &= \mathbf{x}_{1t}'\boldsymbol{\beta}_1 + \mathbf{x}_{2t}'\boldsymbol{\beta}_2 + \gamma_t + \varepsilon_t, \quad t = 1, 2, ..., T \qquad (5.9) \\ (1-L)\mathbf{x}_{2t} &= \boldsymbol{\eta}_t, \end{aligned}$$

where γ_t is as in (5.2), $\mathbf{u}_t = (\varepsilon_t, \boldsymbol{\eta}_t')'$ satisfy the linear process assumptions of Phillips and Solo (1992) and $\mathbf{x}_{2,0} = 0$. For convenience, as in CH we take $\mathbf{z}_t$ to be a set of $s-1$ trigonometric terms; as noted in the previous section, a parametrization of the seasonal effects with seasonal dummies yields the same result.

For the deterministic regressors $\mathbf{x}_{1t}$ we assume that

$$T^{-1}\sum_{t=1}^{T}\mathbf{D}_T^{-1}\mathbf{x}_{1t}\mathbf{x}_{1t}'\mathbf{D}_T^{-1} \to \mathbf{Q}_x, \text{ and } T^{-1}\sum_{t=1}^{T}\mathbf{D}_T^{-1}\mathbf{x}_{1t}\mathbf{z}_t' \to \mathbf{0},$$

where $\mathbf{D}_T$ is a (diagonal) scaling matrix and $\mathbf{Q}_x$ is a positive definite matrix. Note that $\mathbf{x}_{1t}$ may include polynomial trends and structural breaks. For example if the deterministic regressors are $\mathbf{x}_{1t} = (1, t, d_t(\alpha))$, where $d_t(\alpha)$ is a dummy variable equal to 1, for $t > \alpha T$ with $0 < \alpha < 1$, we correspondingly have $\mathbf{D}_T = diag(1, T, 1)$.

Contrary to what is stated by Canova and Hansen (1995) the asymptotic distribution of $\omega_A(m)$ is independent of the regressors $(\mathbf{x}_{1t}, \mathbf{x}_{2t})$, as long as they satisfy the assumptions above. This is shown below. The fact that a deterministic linear trend makes no difference to the distribution was demonstrated in Harvey and Streibel (1997) for the case of $s = 2$.

Let $\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\gamma}_0$ be the OLS estimators of $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \gamma_0$ under H_0. First we show that $\widehat{\boldsymbol{\beta}}_1$,$\widehat{\boldsymbol{\beta}}_2$ are asymptotically uncorrelated with $\widehat{\gamma}_0$. Write the OLS estimators

as

$$\begin{pmatrix} T^{\frac{1}{2}}\mathbf{D}_T(\widehat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1) \\ T(\widehat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2) \\ T^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}}_0-\boldsymbol{\gamma}_0) \end{pmatrix} = \mathbf{A}^{-1}\mathbf{b},$$

where

$$\mathbf{A} = \begin{pmatrix} T^{-1}\sum \mathbf{D}_T^{-1}\mathbf{x}_{1t}\mathbf{x}_{1t}'\mathbf{D}_T^{-1} & T^{-\frac{3}{2}}\sum \mathbf{D}_T^{-1}\mathbf{x}_{1t}\mathbf{x}_{2t}' & T^{-1}\sum \mathbf{D}_T^{-1}\mathbf{x}_{1t}\mathbf{z}_t' \\ T^{-\frac{3}{2}}\sum \mathbf{x}_{2t}\mathbf{x}_{1t}'\mathbf{D}_T^{-1} & T^{-2}\sum \mathbf{x}_{2t}\mathbf{x}_{2t}' & T^{-\frac{3}{2}}\sum \mathbf{x}_{2t}\mathbf{z}_t' \\ T^{-1}\sum \mathbf{z}_t\mathbf{x}_{1t}'\mathbf{D}_T^{-1} & T^{-\frac{3}{2}}\sum \mathbf{z}_t\mathbf{x}_{2t}' & T^{-1}\sum \mathbf{z}_t\mathbf{z}_t' \end{pmatrix},$$

$$\mathbf{b} = \begin{pmatrix} T^{-\frac{1}{2}}\sum \mathbf{D}_T^{-1}\mathbf{x}_{1t}\varepsilon_t \\ T^{-1}\sum \mathbf{x}_{2t}\varepsilon_t \\ T^{-\frac{1}{2}}\sum \mathbf{z}_t\varepsilon_t \end{pmatrix}.$$

We have to show that A_{13} and A_{23} are $o_p(1)$, whereas A_{11}, A_{12}, A_{22}, A_{33} and all the elements of $\mathbf{b}$ are $O_p(1)$. The result for A_{11}, A_{13}, A_{33}, b_1, b_3 follows either from assumption or from standard results on regression. The required orders of magnitude for A_{22} and b_2 also follow from standard results on unit root processes, see for example Hamilton (1994). The order for A_{12} is obtained immediately from the Cauchy-Schwartz inequality and the one for A_{23} can be obtained using summation by parts; in fact, for $h = 1, ..., s-1$, we have

$$\begin{aligned} \sum_{t=1}^{T} \mathbf{x}_{2t} z_{ht} &= -\sum_{t=1}^{T-1} (\mathbf{x}_{2,t+1} - \mathbf{x}_{2t}) \sum_{j=1}^{t} z_{hj} + \mathbf{x}_{2T} \sum_{t=1}^{T} z_{ht} \\ &= O_p(T^{\frac{1}{2}}), \end{aligned}$$

since $\mathbf{x}_{2T} = O_p(T^{\frac{1}{2}})$, $\left|\sum_{j=1}^{t} z_{hj}\right|$ is bounded and

$$\mathrm{E}\left(\sum_{t=2}^{T} \boldsymbol{\eta}_t^*\right)\left(\sum_{t=2}^{T} \boldsymbol{\eta}_t^*\right)' = O(T),$$

where $\boldsymbol{\eta}_t^* \equiv -\boldsymbol{\eta}_t \sum_{j=1}^{t-1} z_{hj}$.

Now write the OLS residuals as

$$e_t = \varepsilon_t - \mathbf{x}_{1t}'(\widehat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1) - \mathbf{x}_{2t}'(\widehat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2) - \mathbf{z}_t'(\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0), \qquad t = 1, ..., T,$$

and consider the (normalised) partial sum process $\mathbf{S}_T(r) = T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \mathbf{z}_t e_t$, $r \in [0,1]$. Then we have that

$$\begin{aligned}
\mathbf{S}_T(r) &= T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \mathbf{z}_t \varepsilon_t - T^{-1} \sum_{t=1}^{[Tr]} \mathbf{z}_t \mathbf{x}_{1t}' \mathbf{D}_T^{-1} \mathbf{D}_T T^{\frac{1}{2}} (\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \\
&\quad - T^{-\frac{3}{2}} \sum_{t=1}^{[Tr]} \mathbf{z}_t \mathbf{x}_{2t}' T (\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) - T^{-1} \sum_{t=1}^{[Tr]} \mathbf{z}_t \mathbf{z}_t' T^{\frac{1}{2}} (\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0) \\
&= T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \mathbf{z}_t \varepsilon_t - T^{-1} \sum_{t=1}^{[Tr]} \mathbf{z}_t \mathbf{z}_t' \left(T^{-1} \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' \right)^{-1} T^{-\frac{1}{2}} \sum_{t=1}^{T} \mathbf{z}_t \varepsilon_t + o_p(1) \\
&= T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \mathbf{z}_t \varepsilon_t - r T^{-\frac{1}{2}} \sum_{t=1}^{T} \mathbf{z}_t \varepsilon_t + o_p(1)
\end{aligned}$$

as $T \to \infty$. Then using the functional central limit theorem of Chan and Wei (1988), and the continuous mapping theorem, we have that

$$\mathbf{S}_T(r) \Rightarrow \Omega^{\frac{1}{2}} \mathbf{B}_{s-1}(r), \quad r \in [0,1], \tag{5.10}$$

where $\mathbf{B}_{s-1}(r)$ is a $s-1$ dimensional standard Brownian bridge, Ω is the long run variance of $\mathbf{z}_t \varepsilon_t$ and $\Rightarrow$ denotes weak convergence. By the continuous mapping theorem it then follows that under H_0 the CH statistic for the complete test converges to a Cramer von Mises distribution with $s-1$ degrees of freedom, i.e.

$$\omega_I(m) \xrightarrow{d} \int_0^1 \mathbf{B}_{s-1}(r)' \mathbf{B}_{s-1}(r) dr \equiv CvM(s-1).$$

Clearly if the test is carried out with a selection matrix $\mathbf{A}'$ with $k \leq s-1$ linearly independent rows, the corresponding distribution for $\omega_A(m)$ is $CvM(k)$.

Thus the inclusion of deterministic trends and stochastic integrated regressors doesn't affect the asymptotic distribution of the CH statistic. However, the inclusion of seasonal slopes, as used, for example, by Proietti (1998), will affect the distribution, just as a time trend affects the distribution of KPSS.

Note that the presence of cross correlation between ε_t and $\boldsymbol{\eta}_t$ is not important for our testing problem; in particular, if we are not interested in making inference on $\boldsymbol{\beta}_2$, there is no need to replace the OLS by, say, the fully modified least squares procedure.

The result would hold also for a process $\mathbf{x}_{2t}$ which is integrated of order one at any frequency, ϕ, as long as it is not one of the seasonal frequencies $\lambda_j = 2\pi j/s$, $j = 1, ..., [s/2]$. Indeed the distribution of $T^{-\frac{3}{2}} \sum \mathbf{x}_{2t} e^{i\phi t}$ has been established by Chan and Wei (1988, lemma 3.3.7); this implies that one cannot have $A_{23} = o_p(1)$ when $\mathbf{x}_{2t}$ is integrated at λ_j.

5.3 Deterministic breaks in the seasonal pattern

Consider now a break in the seasonal pattern at a known time $[\alpha T]$, $\alpha \in [0, 1]$, that is replace (5.9) and (5.2) with

$$y_t = \mathbf{x}_{1t}\boldsymbol{\beta}_1 + \gamma_t + \varepsilon_t, \tag{5.11}$$

$$\gamma_t = \mathbf{z}_t'\boldsymbol{\gamma}_t + d_t(\alpha)\mathbf{z}_t'\boldsymbol{\theta} \tag{5.12}$$

where as before $d_t(\alpha) = 1(t > \alpha T)$ and, for brevity, we have not included stochastic regressors. The model implies that after time αT the seasonal effect has changed from $\boldsymbol{\gamma}_t$ into $\boldsymbol{\gamma}_t + \boldsymbol{\theta}$. In this case the distribution of the CH statistic $\omega_I(m)$, constructed by regressing y_t on $\mathbf{x}_{1t}$, $\mathbf{z}_t$, $d_t(\alpha)\mathbf{z}_t$ is no longer $CvM(s-1)$ but will depend in a rather complicated way on the breakpoint parameter α. However a slightly different statistic can be considered, the asymptotic distribution of which is $CvM(2s-2)$ under H_0. This extends the modification to the LBI test at frequency zero suggested in chapters 3 and 4.

Let $\mathbf{S}_{1t} = (1 - d_t(\alpha))\mathbf{S}_t$, $\mathbf{S}_{2t} = d_t(\alpha)\mathbf{S}_t$, where $\mathbf{S}_t = \sum_{i=1}^{t} \mathbf{z}_i e_i$ and the e_t are the OLS residuals from regressing y_t on $\mathbf{x}_{1t}$, $\mathbf{z}_t$, $d_t(\alpha)\mathbf{z}_t$; note that $\mathbf{S}_{[\alpha T]} = 0$ by the OLS orthogonality conditions. The modified statistic is then

$$\omega_A^*(m) = tr\left(\left(\mathbf{A}'\widehat{\Omega}\mathbf{A}\right)^{-1}\mathbf{A}'\widehat{\mathbf{C}}\mathbf{A}\right),$$

where

$$\widehat{\mathbf{C}} = [\alpha T]^{-2}\sum_{t=1}^{T}\mathbf{S}_{1t}\mathbf{S}_{1t}' + (T - [\alpha T])^{-2}\sum_{t=1}^{T}\mathbf{S}_{2t}\mathbf{S}_{2t}'$$

and $\widehat{\Omega}$ is defined as in section 5.1.

Note that $\omega_A^*(m) = 0.25\omega_A$ when $\alpha = 0.5$ so when the breakpoint is in the middle of the sample the LBI test and the test defined by $\omega_A^*(m)$ are the same. For $\alpha \neq 0.5$ this modified test is consistent, although it is no longer LBI. However, for the corresponding statistic at frequency zero, it has been shown in chapter 3 that the loss of power is not great.

As regards the asymptotic distribution of $\omega_A^*(m)$ under H_0, we have

$$\begin{aligned}(\alpha T)^{-\frac{1}{2}}\mathbf{S}_{1,[\alpha rT]} &\Rightarrow \mathbf{\Omega}^{\frac{1}{2}}\mathbf{B}_{s-1}(r), \qquad r \in [0,1],\\ ((1-\alpha)T)^{-\frac{1}{2}}\mathbf{S}_{2,[\alpha T+(1-\alpha)rT]} &\Rightarrow \mathbf{\Omega}^{\frac{1}{2}}\mathbf{B}^*_{s-1}(r), \qquad r \in [0,1],\end{aligned}$$

where $\mathbf{B}_{s-1}(r)$, $\mathbf{B}^*_{s-1}(r)$ are independent $s-1$ dimensional standard Brownian bridges. By the continuous mapping theorem and the additivity property of the Cramer-von Mises distributions, it follows that

$$\omega_I^*(m) \xrightarrow{d} \int_0^1 \mathbf{B}_{2s-2}(r)'\mathbf{B}_{2s-2}(r)dr \equiv CvM(2s-2),$$

so the asymptotic distribution is independent of the breakpoint location α. With a selection matrix , $\mathbf{A}$, convergence is to $CvM(2k)$ where $k \leq s-1$ is the rank of $\mathbf{A}$.

5.4 Parametric tests based on structural time series models

The model underlying the trigonometric CH test is one in which the seasonal regressors have coefficients which follow random walks under the alternative. The covariance matrix of the disturbances driving the random walks is scalar. For a frequency, λ_j, less than π, the covariance matrix is $\sigma_\omega^2\mathbf{I}_2$, while at π there is a single disturbance with variance $\sigma_\omega^2/2$. This model can be transformed to the one in Harvey (1989, ch 2) except that there the variance of the disturbance driving the component at frequency π is not divided by two. Proietti (2000) gives reasons as to why division by two is preferable.

In the dummy variable model of nonstationary seasonality the $s \times 1$ vector γ_t in (5.2) is a multivariate random walk with the condition that the seasonal

effects sum to zero, that is $\mathbf{i}'\boldsymbol{\gamma}_t = \mathbf{0}$, enforced by a covariance matrix for the disturbances given by $\sigma^2_{\omega *}(\mathbf{I} - s^{-1}\mathbf{i}\mathbf{i}')$. Proietti (2000) shows that this model is equivalent to the trigonometric model with $\sigma^2_{\omega *} = (s/2)\sigma\omega^2$. It is therefore not surprising that the LBI test statistics for the two models are the same.

These above two seasonal specifications are normally used to characterise the seasonal component in a structural time series model. We now consider how to set up parametric tests in this framework. As stated in the introduction, parametric tests have the attraction of avoiding the somewhat arbitrary decisions about lag length selection made in connection with (3.21) and are likely to exhibit a more reliable size and higher power; see Harvey and Streibel (1997) and section 3.5.2 in chapter 3 for a discussion on parametric corrections for the KPSS test.

Suppose that ε_t in (5.1) is replaced by υ_t, a linear stationary process, possibly consisting of more than one component. If $\boldsymbol{\upsilon}$ is a $T \times 1$ vector with υ_t in the $t - th$ position,

$$Var\,(\boldsymbol{\upsilon}) = \mathbf{V} = \sigma^2_* \mathbf{V}_*, \tag{5.13}$$

where σ^2_* is a variance associated with some disturbance in υ_t. If $\mathbf{V}_*$ is known, it follows from King and Hillier (1985), that the LBI test is of the form (5.5) except that the OLS residuals are replaced by the elements of the $T \times 1$ vector , $\mathbf{V}_*^{-1}\widetilde{\mathbf{e}}$, where $\widetilde{\mathbf{e}}$ is the vector of generalised least squares (GLS) residuals. If υ_t contains a white noise component with variance σ^2_* then it is straightforward to show that $\mathbf{V}_*^{-1}\widetilde{\mathbf{e}}$ is equal to the smoothed estimator of the vector of the white noise series. More generally, when multiplied by σ_*^{-2} it becomes the vector of what de Jong and Penzer (1998) call smoothing errors, $\mathbf{u}$. The smoothing errors are produced as a by-product of the smoother applied to the state space form of the model.

With $\mathbf{V}_*$ known, an exact test can be carried out using numerical inversion to construct critical values or probability values. However, $\mathbf{V}_*$ will normally depend on unknown parameters, so there are good reasons for wishing to use a statistic with a known asymptotic distribution. If the test statistic is formed from smoothing errors, it is necessary to take account of their serial correlation.

Following a similar argument to the one used to give (5.8), the denominator needs an estimator of the sgf of $\mathbf{V}^{-1}\boldsymbol{v}$. This sgf is equal to $1/g_v(L)$. The estimator is formed using maximum likelihood (ML) estimators of the unknown parameters under the alternative hypothesis, but the smoothing errors are obtained by setting σ_ω^2 to zero. The parametric test statistic corresponding to (5.8) is therefore

$$\omega_j = 2T^{-2}\widetilde{g}_v(\lambda_j)\sum_{t=1}^{T}\left[\left(\sum_{i=1}^{t} u_i \cos\lambda_j i\right)^2 + \left(\sum_{i=1}^{t} u_i \sin\lambda_j i\right)^2\right], \quad j = 1, ..., [(s-1)/2]. \tag{5.14}$$

The test statistic has the same asymptotic distribution as (5.8), namely $CvM(2)$.

The seasonal and stationary components are usually combined with a trend, so that, for example

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \qquad t = 1, ..., T \tag{5.15}$$

where μ_t is a random walk with drift

$$\mu_t = \mu_{t-1} + \beta + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2).$$

More generally, the slope may be stochastic as well. The asymptotic distribution of the CH statistic is unaffected by the presence of a trend, irrespective of whether it be deterministic or stochastic. The sgf of $\mathbf{V}^{-1}\boldsymbol{v}$ still exists since the smoothing errors are stationary, though when a stochastic trend is included in the model they will not be (strictly) invertible. However, the noninvertibility only affects the zero frequency and the 'quasi' sgf of the observations can be inverted at λ_j. Thus for the model in (5.15)

$$g_v(\lambda) = \left[\frac{\sigma_\eta^2 + 2(1-\cos\lambda)\sigma_\varepsilon^2}{2(1-\cos\lambda)}\right], \quad 0 < \lambda \leq \pi \tag{5.16}$$

If instead of the smoothing errors the smoothed estimator of an irregular component, ε_t, is used, the above correction factor must be divided by σ_ε^4; see Harvey and Streibel (1997).

The corrections to allow for the serial correlation in the smoothing errors can be avoided by constructing an asymptotically equivalent test set up in terms of

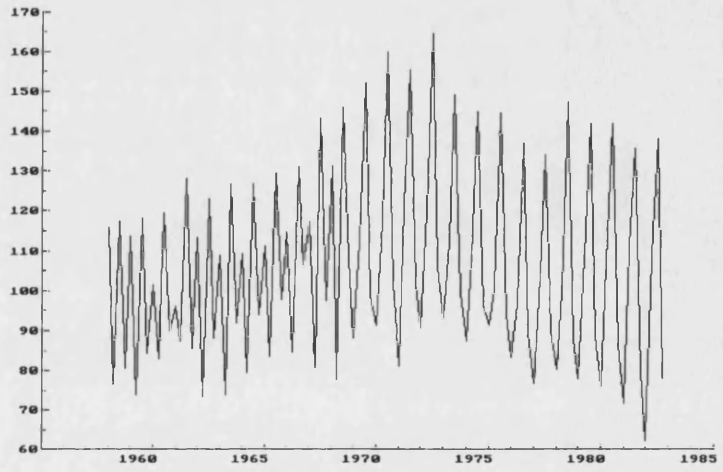

Figure 5.1: Number of Marriages in UK

the T innovations calculated by treating nonstationary and deterministic components as having fixed initial conditions. This requires running the Kalman filter starting with smoothed estimates of the initial conditions. The statistic is of the form (5.5) with any nuisance parameters again being estimated under the alternative.

5.5 Example: UK marriages

The quarterly series of marriages registered in the UK from 1958q1 to 1982q4 was extracted from various issues of the UK Monthly Digest of Statistics. It is shown in figure 5.1. The nonparametric CH test statistic calculated from first differences (with seasonal means subtracted) is 6.79, 4.18, 2.74 and 2.11 for lags of 0, 4, 8 and 12 respectively. This leads to a rejection of the null hypothesis as the 5% critical value for the $CvM(3)$ distribution is 1.00. Formula (5.8) was used; the original CH statistic, (5.6), gave somewhat smaller values, namely 3.47, 1.78, 1.20 and 0.96.

Estimating (5.15) with a random walk trend using the STAMP program of

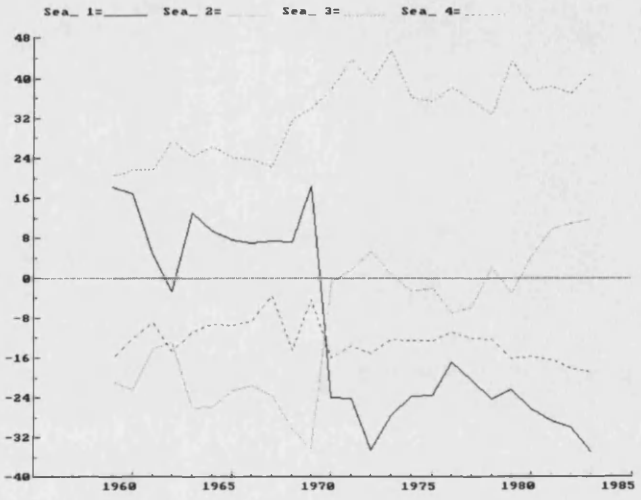

Figure 5.2: Extracted individual seasonal components

Koopman *et al* (1995) gives

$$\widetilde{\sigma}\varepsilon = 0.00 \qquad \widetilde{\sigma}\eta = 1.61 \qquad \widetilde{\sigma}\omega = 2.69$$

with a standard error (the standard deviation of the innovations), s, of 7.91. The parametric test statistic, constructed as $\omega_1 + \omega_2$ in (5.14), with correction factor as in (5.16), is 9.81 which is a much firmer rejection of the null hypothesis than was given by the nonparametric test. The reason for the rejection can be seen from figure 5.1: there appears to be a break in the seasonal pattern at the beginning of 1969. This is very clear from the plot of the individual seasons in figure 5.2 where it can be seen that there was a switch from winter marriages to marriages in the spring quarter. This happened because of a change in the tax law. Up to the end of 1968 couples were allowed to claim the married persons tax allowance retrospectively for the entire year in which they married. As the tax year begins in April this arrangement provided an incentive to marry in the first quarter of the calendar year, rather than in the spring. The abolition of this rule led to a marked decrease in the number of weddings in quarter one and a compensating rise in quarter two.

Adding a set of three seasonal break dummy variables, starting in the first quarter of 1969, to take account of a complete change in the seasonal pattern

leads to the following estimates of the parameters:

$$\widetilde{\sigma}_\varepsilon = 2.42 \qquad \widetilde{\sigma}_\eta = 1.59 \qquad \widetilde{\sigma}_\omega = 1.36$$

with

$$Q(9,7) = 12.54 \qquad H(32) = 0.77 \quad N = 0.22 \qquad s = 5.66$$

where $Q(P, f)$ is the Box-Ljung statistic based on P residual autocorrelations but with f degrees of freedom, H is a heteroscedasticity statistic and N is the Bowman-Shenton normality test; see Koopman et al. (1995). The $t-$statistics for the seasonal break dummies are -8.33, 7.58 and 2.09 respectively. There is a big reduction in the estimate of the seasonal parameter, $\sigma\omega$, which no longer needs to be such as to allow the stochastic seasonal model to accommodate the change, and the equation standard error, s, has fallen considerably.

The modified nonparametric CH test statistics carried out on the residuals obtained from regressing first differences on seasonal means and the seasonal break dummies are 2.71, 2.06, 1.69 and 1.58 for $m = 0, 4, 8$ and 12 respectively for (5.8) and 2.18, 1.80, 1.57 and 1.50 for (5.6). Thus for $m = 4$ and 8 the null of a constant seasonal pattern is not rejected at the 5% level of significance since the critical value for $CvM(6)$ is 1.69. The corresponding parametric test statistic, calculated from the estimated irregular component, is 9.59 indicating that there is still stochastic seasonality. This is backed up by the fact that estimating the model with a fixed seasonal gives a significant Box-Ljung statistic of $Q(9, 8) = 22.38$ while the fourth order residual autocorrelation, $r(4)$, is 0.33.

Although it is difficult to make firm statements about power on the basis of the above results the contrast between the parametric and nonparametric tests is striking. Furthermore, the statistic in (5.6) seems to be smaller than the one based on (5.8). Estimating the model under the null hypothesis and making the correction in formula (5.8) analytically as $\widehat{g}(\lambda_j) = \sigma_\eta^2 + 2(1 - \cos\lambda)\sigma_\varepsilon^2$ gave a test statistic of 6.19 with no break and 2.26 with a break. Both figures are consistent with the nonparametric correction for moderate lag length. On the other hand if the estimated parameters under the alternative are used, the test statistic in the modelled break case is 9.39 which is not far below the figure of 9.59 reported in the previous paragraph.

5.6 Tests on groups of seasons

Tests to be applied to a group of seasons can be set up by constructing $\mathbf{A}'$ matrices with fewer rows and with seasons not necessarily contrasted with adjacent ones. If we are contrasting two seasons m periods apart then the $\mathbf{A}'$ matrix is a vector such as $\mathbf{A}' = (1\ 0\ 0\ -1\ 0)$, while if the observations have been differenced, the corresponding vector is $\mathbf{A}' = (1\ 1\ 1\ 0\ 0)$. This is just a reflection of the identity $\Delta_m = (1 + L + ... + L^{m-1})\Delta$. The $\mathbf{A}'$ matrix for testing the stability of a group of s_1 seasons has $(s_1 - 1)$ rows constructed in this way.

What kind of model might a test on a group of seasons imply? One possibility is a *partially periodic* one in which different models are assumed for two groups of seasons. Suppose there are s_1 seasons in the first group and s_2 in the second and $s_1 + s_2 = s$. Then

$$y_t = \mu_t^{(k)} + \gamma_t^{(k)} + \varepsilon_t^{(k)}, \qquad k = 1, 2, \tag{5.17}$$

where the seasonal, $\gamma_t^{(k)}$, is modelled by a set of s_k, $k = 1, 2$ time-varying dummies which embody the zero sum restriction over the group. If a group has only one season then the seasonal component is not needed. The trends can be assumed to have slopes for more generality so that

$$\begin{aligned} \mu_t^{(k)} &= \mu_{t-1}^{(k)} + \beta_{t-1}^{(k)} + \eta_t^{(k)}, \qquad Var(\eta_t^{(k)}) = \sigma_{\eta,k}^2, \qquad k = 1, 2. \\ \beta_t^{(k)} &= \beta_{t-1}^{(k)} + \zeta_t^{(k)}, \qquad Var(\zeta_t^{(k)}) = \sigma_{\zeta,k}^2. \end{aligned}$$

More generally, the use of the state space form allows parameters and components common to both groups. For example in the present model the irregular terms can have the same variance. In any case the point is that one may wish to test the hypothesis that the seasonals in $\gamma_t^{(1)}$ and/or $\gamma_t^{(2)}$ are deterministic.

For tests of groups of seasons, making the correction to the smoothing error statistics is difficult and so if parametric tests are to be used the innovations tests seem like the best option.

Note that if there are as many groups as seasons, then each season has its own trend, which is independent of the others. A joint test for nonstationarity of all the series is obtained by applying the dummy variable test setting $\mathbf{A}$ to

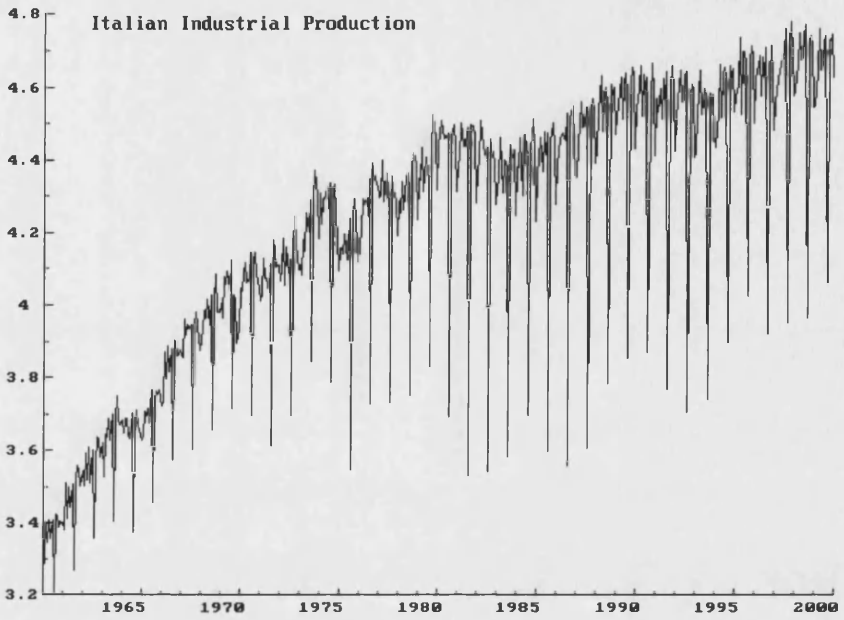

Figure 5.3: Index of Italian Industrial Production

an $s \times s$ identity matrix and keeping the observations in levels. The series may have stationary components in common. However, if the series for each season are mutually independent then the model is a purely periodic one.

5.7 Example: Italian industrial production

The logarithm of the index of Italian industrial production from January 1960 until December 1999 is shown in figure 6.3. The month of August, the traditional holiday period, behaves very differently from the other months, with what is conceivably a different trend. This suggests a partly periodic model of the kind set out in the previous section. Within this framework one may wish to test that the seasonal pattern is constant if August is excluded. (Since August is in a group by itself the only relevant test would be for the presence of a stochastic trend).

There are a number of ways of carrying out the test. If the trend is a random

walk with drift, the nonparametric test can be carried out on the differenced observations with seasonal means subtracted. The $\mathbf{A}'$ matrix is formed by taking an $\mathbf{I}_{12}$ matrix and deleting the two rows which would pick up data on August (the differenced August and September figures both depend on August).

If the trend contains a stochastic slope so that second differencing is appropriate, the nonparametric test appears to run into problems as August affects three observations in each year.

The parametric test requires fitting a partly periodic model. If this is done, the $\mathbf{A}'$ matrix is a 10×12 matrix of the type shown in (5.4) except that there is no row corresponding to August or September.

Applying the full CH test to first differences gives $\omega_A(10) = 4.60$, which rejects the null at the 5% level, since the critical value is 2.75. If $\mathbf{A} = \mathbf{I}_{12}$, the test is for stochastic seasonality and a stochastic slope and the test statistic is $\omega_A(10) = 4.74$. Dropping the 8th and 9th rows of the $\mathbf{I}_{12}$ matrix, so as to exclude August, gives $\omega_A(10) = 4.32$, as compared with a 5% critical value of 2.54, so the partly periodic model which excludes August still appears to display stochastic seasonality.

5.8 Seasonal unit root tests

The test of Hylleberg et al. (1990) - HEGY- is related to CH in that it is testing the null of a nonstationary seasonal against the alternative of a stationary seasonal; thus it parallels the relationship between (augmented) Dickey-Fuller test and KPSS. As with the CH test the distribution changes if seasonal trends are included in the model; see Smith and Taylor (1998).

Using arguments similar to those of Perron (1989), it can be conjectured that the performance of the HEGY test is affected by the presence of structural breaks in the seasonal pattern. Indeed, Smith and Otero (1997) have shown, via simulation experiments, that a change in the deterministic seasonal pattern results in a corresponding power loss for the seasonal unit root tests. They then tabulate new critical values for the HEGY tests which account for structural breaks. Unlike our tests of sections 3-4, their critical values depend on the

location of the breakpoint.

The asymptotic distributions of the HEGY statistics in the presence of neglected structural breaks in the seasonal pattern have been derived by Lopes and Montanes (1998) as functions of the break magnitude, the breakpoint location and the variance of the innovation. They show that all the relevant HEGY statistics asymptotically diverge, although in finite samples the probability of rejecting the unit root null hypothesis may become very low. Most of the simulation results of Smith and Otero (1997) are justified analytically, and the role of the break magnitude in obtaining those results is clarified

The papers by Franses and Vogelsang (1998) and Balcombe (1999) consider a single breakpoint at an unknown position and propose inf-type tests for seasonal unit roots that extend the framework of Perron and Vogelsang (1992) and Zivot and Andrews (1992) respectively. The advantage of these unconditional tests is that they cover the case of endogenous breakpoints. However, when a change in the seasonal pattern is caused by some exogenous event (as with the policy change affecting UK marriages) the unconditional testing approach is likely to favour the null hypothesis of unit roots.

If the unobserved components model is a good approximation for a time series, then the alternative hypothesis in the HEGY test corresponds to seasonality being deterministic. There is then a paradox in that there are more parameters under the null. Further, the autoregressive approximation may be very poor. Suppose, for simplicity, that the model is a random walk plus seasonal plus irregular with just two seasons. Then

$$\Delta_2 y_t = (1 - L^2)y_t = (1 + L)\eta_t + (1 - L)\omega_t + (1 - L^2)\varepsilon_t \qquad (5.18)$$

If either the level or seasonal variance is small relative to the irregular variance, the right hand side of the above expression is close to the invertibility boundary. When there are more than two seasons there are more potentially deterministic terms to worry about.

5.9 Test of Seasonality

The CH test takes the null to be deterministic seasonality. Sometimes we may wish to test whether there is any seasonality at all. One strategy is to carry out the CH test and if this does not reject, a test of significance of the seasonal coefficients is carried out. Such a test, based on a fitted structural time series model is implemented in the STAMP package. However, it has the disadvantage that it will indicate no seasonal effects in a situation where seasonality has become less pronounced over time. This is precisely the kind of behaviour noted by Canova and Hansen (1995, p 24-50) in their analysis of US macroeconomic series. Another strategy would be to test for nonstationary seasonality using partial sums computed without fitting seasonal dummies. Such a test will have power against deterministic seasonality[2] and it has the advantage that two tests are replaced by one. A modification of the nonparametric test would be to fit seasonal dummies when calculating the denominator of the test statistic. This makes no difference to the asymptotic distribution under the null hypothesis and it could increase power. The parametric test is best carried out by estimating the nuisance parameters from the general model with a stochastic seasonal fitted. However, using estimates from a model with deterministic seasonal, or indeed no seasonal at all, would still be valid. Proceeding in this way may be attractive with, say, weekly data where fitting a stochastic seasonal can be somewhat complex and time consuming.

If the CH test statistic is formed without fitting seasonal dummies, its asymptotic distribution under the null will be a function of Brownian motion rather than of a Brownian bridge. It is still of the Cramer-von Mises family. The 5% critical values for one, two and three degrees of freedom - the last appropriate for a full test on quarterly data - are 1.65, 2.63 and 3.46 respectively; see the tables in Nyblom (1989) and Hobijn and Franses (2000). The 5% critical value for eleven degrees of freedom is 9.03. For the reasons given in section 4, the asymptotic distribution is unaffected by the inclusion of a constant or a constant

[2] Following a similar line of argument to the one used by Hobijn and Franses (2000, appendix) in a somewhat different context, it can be shown that the test will be consistent against deterministic seasonality.

and a time trend.

As an example we consider the logarithm of the ratio of the first decile to the median for quarterly wages of US males from 1979q1 to 1999q3; see Harvey and Bernstein (2000) for further details. While there are clear deterministic seasonal patterns in the deciles, these almost cancel in the log ratio leaving a weak seasonal which changes over time. The question is whether this seasonal is any sense significant. Fitting a random walk plus seasonal plus noise model gives a chi square statistic for the seasonals at the end of the series, 1.83, which is not significant, but this may be unreliable if seasonality becomes less pronounced over time. If the seasonal component is omitted, the fourth order residual autocorrelation, $r(4)$, takes a value of 0.24 which is rather high. Unfortunately the distribution of this statistic is not known. However, the $Q(8,7)$ statistic is 10.70 which is not significant. Our parametric seasonal test is 2.60 so it doesn't quite reject at the 10% significance level for which the critical value is 2.83. This backs up the evidence from the other statistics which is that there is some indication of a weak seasonal pattern.

The nonparametric statistics 0.42, 0.48 and 0.49 for $m = 4, 8$ and 12 so again it appears to have very low power.

As a final point, note that the test can be applied to a seasonally adjusted series to see if the adjustment has been effective. This assumes that the adjustment has been done by means of moving averages, rather than by regressing on seasonal dummies. If dummies have been used, then the test statistics have the usual CH distribution.

5.10 Testing for the presence of trading day effects

Cleveland and Devlin (1980) showed that peaks at certain frequencies in the estimated spectra of monthly time series indicate the presence of trading day effects. Specifically there is a peak at a frequency of $0.348\times 2\pi$ radians, with the possibility of subsidiary peaks at $0.432\times 2\pi$ and $0.304\times 2\pi$ radians. An option in the output of the X-12-ARIMA program provides a comparison of the estimates of

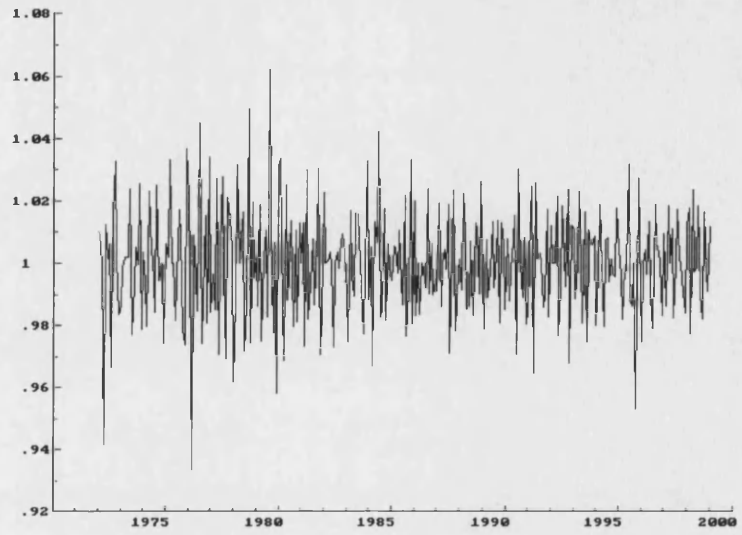

Figure 5.4: Extreme value irregular component

these frequencies with the adjacent frequencies; see Soukup and Findlay (2000). However, there is no formal test. One possibility, suggested by the methods of this paper, is to carry out a CH test at the relevent frequency or a joint test at all three frequencies. Assuming that no (deterministic) trading day model has been fitted, the statistic will be as in (5.8) or in a corresponding parametric form, and the asymptotic distribution is as in the previous sub-section with the 5% critical value being 2.63 for a test at a single frequency and 5.68 for a test at all three frequencies.

As an example we took the extreme value irregular component, obtained from X12-ARIMA, of series s0b56ym, *U.S. Retail Sales of Children's, Family, and Miscellaneous Apparel,* as supplied by the Bureau of the Census; it is depicted in figure 5.4. Since the process followed by the this irregular component cannot be derived, it was decided to use the nonparametric test. The CH test statistic with ten lags for the single main frequency was 7.03. For all three frequencies it was 8.21. Both give a clear rejection of the null hypothesis that there is no trading day effect. In effect, the plot of the periodogram in figure 5.5 clearly shows a high peak at the main trading days frequency.

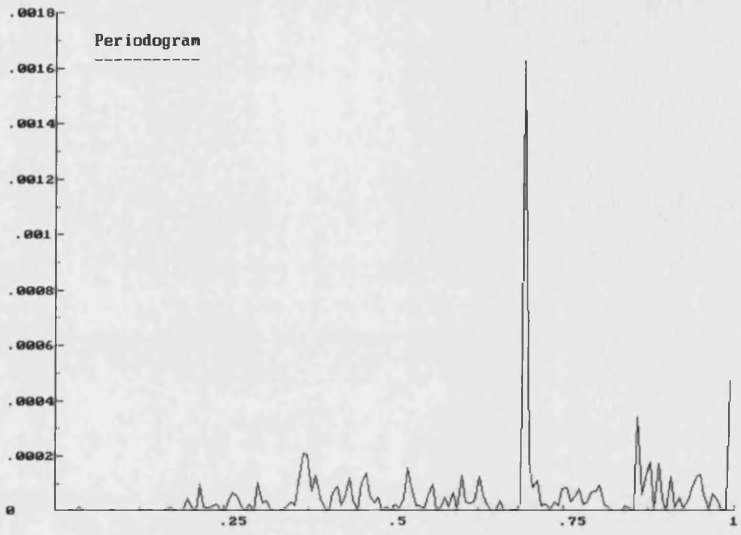

Figure 5.5: Periodogram of the irregular component

5.11 Concluding remarks

The example of UK marriages highlights the weakness of the nonparametric aspect of the CH test, namely the different inferences drawn with different choices of lag length. Asymptotic theory offers little guidance to the choice of m in small samples and Lee (1996) has demonstrated that, for the KPSS test, optimal bandwidth selection procedures such as those in Andrews and Monahan (1992) tend to lead to a loss of power. Indeed our examples indicate that the parametric tests based on fitted structural time series models may be considerably more powerful than nonparametric tests though this is something which needs to be investigated further by a Monte Carlo study. Only when dealing with potentially complex models, such as those which are partially periodic, does a nonparametric approach becomes relatively more appealing.

The suggested test for overall seasonality appears to be effective and again is best carried out parametrically.

Chapter 6

Cointegration in structural time series models

A structural time series model is set up in terms of unobserved components, such as trends, and cycles, which have a direct interpretation; see Harvey (1989), Kitagawa and Gersch (1996) and Young (1984). In a multivariate structural model we have cointegration when the number of stochastic trends is less than the number of the series (*common trends*).

This chapter deals with representation, estimation and test of cointegrated structural time series models. Identification of the cointegrating vectors is discussed and order and rank conditions for identification are given. A number of links with the literature on vector autoregressions are established. In particular, it is showed that a cointegrated structural time series model can be represented as a vector autoregression, and it is also explained how to compute the VAR coefficients using the filtering algorithm given in Koopman and Harvey (1999).

The tests developed in chapters 3 and 4 can be naturally applied to make inference on the dimension of the cointegration space and to test for prespecified cointegration vectors. Maximum likelihood estimation of cointegrated structural time series models is done by putting the models in state space form and computing the prediction error decomposition form of the likelihood function, using the Kalman filter to obtain the innovations. Efficient estimation requires incorporating the common trend restrictions. An empirical example is provided with the series of US GNP and Investment.

In summary the chapter proceeds as follows. Section 6.1 reviews the concept

of cointegration, the VAR/MA/ECM representations of cointegrated systems and the Granger Representation Theorem. In section 6.2 representation and identification of cointegrated system in the structural time series framework are considered and links with the VAR representation are established. Estimation and test of cointegrated structural time series are discussed in section 6.3 and an empirical example is provided in section 6.4.

6.1 Cointegration and the Granger representation theorem

We say that a vector process $\mathbf{x}_t$ is *integrated of order d*, denoted $I(d)$, if $\Delta^d \mathbf{x}_t$ is a covariance stationary process with positive definite spectrum at frequency zero; slightly different definitions of integrated processes are possible, see Davidson (1999). The vector process $\mathbf{x}_t$ is said to be *cointegrated of order* d, g, denoted $CI(d, g)$, if (i) each component of $\mathbf{x}_t$ is $I(d)$ and (ii) there exists a nonzero vector $\mathbf{b}$ such that $\mathbf{b}'\mathbf{x}_t$ is $I(d-g)$, where $d \geq g > 0$. The vector $\mathbf{b}$ is called cointegrating vector. In the following we will restrict attention to $CI(1,1)$ processes.

Consider an N-dimensional process $\mathbf{x}_t$ that in first differences has the moving average representation

$$\Delta \mathbf{x}_t = \boldsymbol{\delta} + C(L)\boldsymbol{\varepsilon}_t, \tag{6.1}$$

where $\boldsymbol{\varepsilon}_t$ is a N-dimensional *i.i.d.* process with zero mean and positive definite variance, δ is a drift and the matrix polynomial in the lag operator $C(L) \equiv \sum_{j=0}^{\infty} C_j L^j$ satisfies $\sum_{j=0}^{\infty} \|C_j\| < \infty$. From (6.1) it follows that $\mathbf{x}_t$ is cointegrated of order 1,1 if $C(1)$ is not of full rank. Further, the number of linearly independent cointegrating vectors is given by $R = N - rank(C(1))$.

In fact, first note the identity

$$C(L) = C(1) + (1 - L)C^*(L), \tag{6.2}$$

where $C^*(L)$ is defined by $C_i^* = \sum_{j=i+1}^{\infty} C_j$, $i = 0, 1, ...$ Then integrate (6.1) to get

$$\mathbf{x}_t = \mathbf{x}_0^* + \boldsymbol{\delta} t + C(1) \sum_{s=1}^{t} \boldsymbol{\varepsilon}_s + C^*(L)\boldsymbol{\varepsilon}_t, \tag{6.3}$$

with $\mathbf{x}_0^* \equiv \mathbf{x}_0 - C^*(L)\boldsymbol{\varepsilon}_0$. Clearly, the nonstationary part $C(1)\sum_{s=1}^{t}\boldsymbol{\varepsilon}_s$ is annihilated only if there exist nonzero vectors $\mathbf{b}$ such that $\mathbf{b}'C(1) = 0$. This, in turn, is true if and only if $C(1)$ is not of full rank. In particular, call S the left null space of $C(1)$, i.e.

$$S \equiv \{\mathbf{b}\colon\ \mathbf{b}'C(1) = 0\}\,.$$

Then S is the cointegration space, i.e. all the elements $\mathbf{b}$ in this space make the linear combination $\mathbf{b}'\mathbf{x}_t$ stationary. By a well known result of linear algebra the dimension of S is $R = N - rank(C(1))$, and thus R is the maximum number of linearly independent cointegrating relations. Usually we want to collect R of these linearly independent vectors into a $(N\times R)$ matrix $\mathbf{B}=[\mathbf{b}_1\,\mathbf{b}_2\,...\,\mathbf{b}_R]$.

For any $\mathbf{b}\in S$ we have the following cointegrating relation

$$\mathbf{b}'\mathbf{x}_t = \mathbf{b}'\mathbf{x}_0^* + \mathbf{b}'\boldsymbol{\delta}t + \mathbf{b}'C^*(L)\boldsymbol{\varepsilon}_t, \tag{6.4}$$

i.e. $\mathbf{b}'\mathbf{x}_t$ is stationary around the linear trend $\mathbf{b}'\boldsymbol{\delta}t$. If $\boldsymbol{\delta}\in S^{\perp}$, where $S^{\perp}$ is the space orthogonal to S, the equilibrium relation (6.4) does not contain a time trend.

If there were not cointegration, the MA representation (6.1) could be inverted to yield the autoregressive representation of the process in *first differences.* With cointegration this is no longer possible for the presence of unit roots in the matrix polynomial $C(L)$.[1] However it is possible to construct the autoregressive representation in the *levels* of the process, inverting (1) as showed in Engle and Yoo (1991) by factorizing $C(L)$ appropriately. This inversion is the essence of the Granger representation theorem, which establishes an isomorphism between the MA and AR representations of a cointegrated system. Further, by a simple reparametrization, the system can be put in error correction form. The theorem is stated below for the case of a driftless process.

Theorem 18 *(Granger Representation Theorem). Let $\boldsymbol{\varepsilon}_t$ be a N-dimensional i.i.d. process with zero mean and positive definite variance. Let $C(z)$ be a $N\times N$ matrix polynomial such that*

(a) $|C(z)| = 0$ has no roots $|z| < 1$,

(b) $|C(z)| = 0$ has $N - R$ roots $z = 1$ or equivalently rank$(C(1))= N - R$,

[1] The characteristic equation $|C(z)| = 0$ has the solution $z = 1$ if $C(1)$ is not of full rank.

(c) $C(1)\mathbf{A} = 0$, $\mathbf{B}'C(1) = 0$ *for two* $N \times R$ *matrices* $\mathbf{A}$, $\mathbf{B}$ *with rank* R. *Then a process* $\boldsymbol{x}_t$ *has the MA representation*

$$\Delta \mathbf{x}_t = C(L)\boldsymbol{\varepsilon}_t, \qquad (MA)$$

if and only if it has the AR representation

$$\Pi(L)\mathbf{x}_t = \boldsymbol{\varepsilon}_t, \qquad (AR)$$

where the matrix polynomial $\Pi(z)$ *satisfies*
(a)' $|\Pi(z)| = 0$ *has no roots* $|z| < 1$,
(b)' $|\Pi(z)| = 0$ *has* R *roots* $z = 1$ *or equivalently rank*$(\Pi(1))$$= R$,
(c)' $\Pi(1) = \mathbf{AB}'$.

The theorem is proved in Engle and Granger (1987), Engle and Yoo (1991) and Johansen (1991). Now define the polynomial matrix $\Gamma(L)$ by

$$\Pi(L) \equiv \Pi(1)L + (1 - L)\Gamma(L),$$

i.e. $\Gamma_0 \equiv \Pi_0$, $\Gamma_i \equiv -\sum_{j=i+1}^{\infty} \Pi_i$, $i = 1, 2, ...$ Then the process $\mathbf{x}_t$ of the theorem has the further representation

$$\Gamma(L)\Delta \mathbf{x}_t = \mathbf{AB}'\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, \qquad (EC)$$

called error correction form.

The matrix $\mathbf{B}$ contains R linearly independent cointegrating vectors, i.e. $\mathbf{B}'\mathbf{x}_t$ is $I(0)$ and represents the equilibrium relationships between the variables. The error correction representation (EC) allows the modeling of the process as deviation from that equilibrium, incorporating both the short run and the long run in the system. The matrix $\mathbf{A}$ contains the adjustment coefficients and, together with the autoregressive parameters Γ_i's, represents the short run dynamics.

The importance of the theorem is that the representations (AR) and (EC) are easy to handle and estimate, e.g. if we assume Gaussianity for the innovations. Then it is possible to construct tests for cointegration by looking at (the rank of) the long run matrix $\Pi(1)$, and in case of positive answer incorporate this restriction into the error correction form obtaining a fully efficient maximum likelihood estimator (Johansen 1988, 1991 and 1992).

In the case of a drift, a linear trend of the form $\mathbf{d}_0 + \mathbf{d}_1 t$ must be added in the right hand side of (EC). The restriction $C(1)\mathbf{d}_1 = \mathbf{0}$ excludes a quadratic trend in the levels of the process while the restriction $\mathbf{d}_1 = \mathbf{0}$ excludes a linear trend in the cointegration relation; see also Pesaran and Shin (1994).

Consider now the problem of identification. Any basis for the R-dimensional space S can be taken as a valid set of cointegrating vectors, i.e. it can describe the equilibrium relationships of the system. The non uniqueness of the cointegrating vectors is evident in that, for every $R \times R$ non singular matrix $\mathbf{Q}$, if $\mathbf{B}'\mathbf{x}_t$ is stationary so is $\mathbf{Q}\mathbf{B}'\mathbf{x}_t$. In other words, when we estimate the (EC) representation of the system, the set of cointegrating vectors $\mathbf{B}$ and the adjustment coefficients $\mathbf{A}$ are not identified, since they are observationally equivalent to $\widetilde{\mathbf{A}} \equiv \mathbf{A}\mathbf{Q}^{-1}$ and $\widetilde{\mathbf{B}} \equiv \mathbf{B}\mathbf{Q}'$ for every non singular $\mathbf{Q}$.[2]

Identification thus requires placing constraints on $\mathbf{B}$, in order to work out the R^2 elements of the matrix $\mathbf{Q}$. At a first stage, these constraints can simply be restrictions implied by some normalization rule, as in the case of Johansen's maximum estimate of $\mathbf{B}$ obtained via reduced rank regression (see Johansen, 1988). However, in general these will not produce interpretable or economically meaningful equilibrium relationships. Thus, in a second stage economic theory should supply help, providing the necessary overidentifying restrictions on the cointegration space to work out a unique set of cointegrating vectors. This identification problem, which is relevant for empirical studies, is thoroughly analyzed by Pesaran and Shin (1995).

6.2 Representation of cointegrated structural time series

A structural time series model is set up in terms of orthogonal components which have a direct interpretation, e.g. trend, cycle, seasonal, etc. Harvey (1989) gives a full account of this modelling strategy. Here we assume for simplicity that the N-dimensional process $\mathbf{x}_t$ is made up of only two components,

$$\mathbf{x}_t = \boldsymbol{\mu}_t + \mathbf{u}_t, \tag{6.5}$$

[2] Only the long run matrix $\Pi(1)$ is identified, as $\Pi(1) = \mathbf{A}\mathbf{B}' = \widetilde{\mathbf{A}}\widetilde{\mathbf{B}}'$.

where $\boldsymbol{\mu}_t$ is a "trend" (long run component), and $\mathbf{u}_t$ is a "cycle" (short run component).

We assume that the cycle $\mathbf{u}_t$ is a $I(0)$ process as defined in the previous section, i.e. it is a covariance stationary process with positive definite spectrum at frequency zero.

In the structural time series framework, the trend $\boldsymbol{\mu}_t$ is often modeled as a *local linear trend* (see Harvey, 1989)

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\delta}_{t-1} + \boldsymbol{\eta}_t, \tag{6.6}$$

$$\boldsymbol{\delta}_t = \boldsymbol{\delta}_{t-1} + \boldsymbol{\zeta}_t, \tag{6.7}$$

where $(\boldsymbol{\eta}_t', \boldsymbol{\zeta}_t')'$ is *i.i.d.* with zero mean and variance $diag\,(\Sigma\eta, \Sigma_\zeta)$. In this case, if Σ_ζ is of full rank, $\boldsymbol{\mu}_t$ is an $I(2)$ process.

Since in this chapter we are mainly interested in $CI(1,1)$ processes, it will be assumed below that $\Sigma_\zeta = \mathbf{0}$, i.e. that the slope $\boldsymbol{\delta}_t$ is fixed and equal to $\boldsymbol{\delta}$. Then the trend is a random walk with drift and the structural time series model (6.5) can be compared to the models considered in the previous section.

6.2.1 Common trends and cointegration

Cointegration is directly connected with the existence of common trends. Let K be the rank of Σ_η, $0 < K < N$. Since there exists a $(N \times K)$ matrix $\boldsymbol{\Theta}$ such that $\Sigma_\eta = \boldsymbol{\Theta}\boldsymbol{\Theta}'$, we can reformulate the model in terms of only K *orthogonal trends*. In particular we can write $\boldsymbol{\eta}_t = \boldsymbol{\Theta}\boldsymbol{\eta}_t^+$, with $\boldsymbol{\eta}_t^+ \backsim i.i.d.\,(0, I_K)$, and the model becomes

$$\mathbf{x}_t = \boldsymbol{\Theta}\boldsymbol{\mu}_t^+ + \boldsymbol{\mu}^* + \boldsymbol{\delta}t + \mathbf{u}_t \tag{6.8}$$

$$\boldsymbol{\mu}_t^+ = \boldsymbol{\mu}_{t-1}^+ + \boldsymbol{\eta}_t^+ \tag{6.9}$$

where $\boldsymbol{\mu}^* \equiv \begin{pmatrix} 0' & \overline{\mu}' \end{pmatrix}'$ with the $N-K$ dimensional vector $\overline{\mu}$ obtained from the initial conditions on the trend.

$\boldsymbol{\Theta}$ is called the matrix of *factor loadings*. Note that $\boldsymbol{\Theta}$ is not unique, since for any $(K \times K)$ orthogonal matrix $\mathbf{H}$, $\widetilde{\boldsymbol{\Theta}} \equiv \boldsymbol{\Theta}\mathbf{H}$ is another square root of Σ_η. Therefore identification of the loadings requires placing $K(K-1)/2$ constraints

on $\boldsymbol{\Theta}$. The simplest way to enforce identification is to assign a lower triangular structure to $\boldsymbol{\Theta}$, that is assuming $[\boldsymbol{\Theta}]_{ij} = 0$ for $j > i$. Of course, this identification scheme of the loadings, which implies a causal chain in the trends, is arbitrary. However, after estimating a model set up in that way, it is possible to look for an appropriate rotation $\mathbf{H}$ of the loadings that has an economic interpretation (as done in Harvey et al. 1995).

From the formulation (6.8)-(6.9), it is easy to see that the process $\mathbf{x}_t$ is $CI(1,1)$, with the cointegration space being the left null space of $\boldsymbol{\Theta}$,

$$S = \{\mathbf{b} : \ \mathbf{b}'\boldsymbol{\Theta} = \mathbf{0}\}.$$

This space has dimension $R = N - K$, i.e. there exist R linearly independent cointegrating vectors $\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_R$ that we can collect into a matrix $\mathbf{B} \equiv (\mathbf{b}_1 \, \mathbf{b}_2 \, ... \, \mathbf{b}_R)$. Note that S is independent of the rotation of the trends, since $\mathbf{0} = \mathbf{b}'\boldsymbol{\Theta} = \mathbf{b}'(\boldsymbol{\Theta}\mathbf{H})$ for every rotation $\mathbf{H}$.

As in the previous section, if $\boldsymbol{\delta} \in S^{\perp}$ a linear trend is excluded from the cointegrating relations. But since $S^{\perp}$ is also the range of $\boldsymbol{\Theta}$, when $\boldsymbol{\delta} \in S^{\perp}$ there exists a K-dimensional vector $\boldsymbol{\delta}^{+}$ such that $\boldsymbol{\delta} = \boldsymbol{\Theta}\boldsymbol{\delta}^{+}$. Then the slope can be incorporated into the common trend $\boldsymbol{\mu}_t^{+}$, adding $\boldsymbol{\delta}^{+}$ in the right hand side of equation (6.9) and removing $\boldsymbol{\delta}t$ from (6.8).

6.2.2 Identification of the cointegrating vectors

Let the dimension of the cointegration space be R and let $\mathbf{B}$ be an $N \times R$ matrix that contains R linearly independent cointegrating vectors. Then, as we have seen before, $\mathbf{B}$ satisfies

$$\mathbf{B}'\boldsymbol{\Theta} = \mathbf{0}, \tag{6.10}$$

where $\boldsymbol{\Theta}$ is a $N \times K$ matrix of factor loadings, with $K = N - R$, which can be efficiently estimated in the structural time series framework; see section 6.3.

As in section 6.1, identification of $\mathbf{B}$ (i.e. identification of a unique set of cointegrating vectors) requires placing additional restrictions on the elements of $\mathbf{B}$ other than (6.10). As for a simultaneous equations model, we can obtain an order condition and a rank condition for identification of $\mathbf{B}$.

Since (6.10) defines a system of KR equations in NR unknowns, the order condition for identification requires imposing $p \geq R^2$ extra restrictions. Let these be linear and denoted as

$$\boldsymbol{\alpha}'\mathbf{W} = \boldsymbol{\omega}', \tag{6.11}$$

where $\boldsymbol{\alpha} = vec(\mathbf{B})$, $\mathbf{W}$ is a known $NR \times p$ matrix and $\boldsymbol{\omega}$ is a known $p \times 1$ vector. Note that (6.11) has to include R inhomogeneous restrictions, e.g. obtained by normalizing one element in each cointegrating vector.

Then, by the usual argument on the existence of a unique solution in systems of linear equations, we have that the rank condition for identification is

$$rank\,(\boldsymbol{\Psi}) = NR, \tag{6.12}$$

where $\boldsymbol{\Psi} = (I_R \otimes \boldsymbol{\Theta}, \mathbf{W})$.

The identifying restrictions (6.11) should have economic interpretation as opposed to being arbitrary. However, arbitrary identification schemes are often employed in the literature. For example, we could think of partitioning our cointegrated vector process $\mathbf{x}_t$ into K trending variables $\mathbf{x}_{1t}$ and R variables $\mathbf{x}_{2t}$ cointegrated with the previous ones. If we correspondingly partition $\boldsymbol{\Theta}$ into $(\boldsymbol{\Theta}_1', \boldsymbol{\Theta}_2')'$, where $\boldsymbol{\Theta}_1$ is $K \times K$ and $\boldsymbol{\Theta}_2$ is $R \times K$, we have that the matrix $\mathbf{B}$ of cointegrating vectors becomes

$$\mathbf{B} = (\boldsymbol{\beta}', I_R)',$$

with $\boldsymbol{\beta} = -\boldsymbol{\Theta}_2\boldsymbol{\Theta}_1^{-1}$. The resulting system is sometimes called Phillips triangular system, as it corresponds to the model analyzed by Phillips (1991, 1994); see also Harvey and Koopman (1997). In this case, the R^2 restrictions (6.11) are obtained by setting the $R \times R$ lower submatrix of $\mathbf{B}$ equal to the I_R.

6.2.3 VAR representation

The following proposition extends theorem 18, the Granger Representation Theorem, by showing that a structural time series model with common trends can be represented as cointegrated VAR (and as error correction model).

Proposition 19 *Let* $\mathbf{x}_t$ *be generated by the model (6.5)-(6.7) with* $\Sigma\zeta = 0$ *and* $rank(\Sigma\eta) = K$, $0 < K < N$, *where the short run component* $\boldsymbol{u}_t$ *is a I(0) process with positive definite spectrum everywhere. Then* $\boldsymbol{x}_t$ *can be represented as a cointegrated VAR.*

The proof amounts to showing that $\Delta\mathbf{x}_t$ has a MA representation that satisfy conditions (a) and (b) of theorem 18[3] . This is most easily done by considering the spectrum of $\Delta\mathbf{x}_t$, which is given by

$$F_{\Delta x}(\lambda) = (2\pi)^{-1}\Sigma_\eta + 2(1 - \cos\lambda)F_u(\lambda), \qquad -\pi < \lambda \leq \pi,$$

where $F_u(\lambda)$ is the spectrum of $\mathbf{u}_t$.

Clearly, $rank(F_{\Delta x}(0)) = K$, thus condition (b) is satisfied. Then, since we have assumed $F_u(\lambda)$ positive definite everywhere, we have that $F_{\Delta x}(\lambda)$ is positive definite for $\lambda \neq 0$. This is equivalent to saying that, except for the K unit roots of condition (b) above, all the other roots of the matrix polynomial in the MA representation of $\Delta\mathbf{x}_t$ are outside the unit circle, and so condition (a) is satisfied too.

Thus we have established that a cointegrated structural time series can be represented as a vector autoregression with reduced rank long run multiplier (and with in principles an infinite number of lags). However, note that estimating a finite order VAR can be a bad approximation if the true model is of the type (6.5)-(6.7), as the reduced form of these models is often close to the region of noninvertibility; see Harvey (1989). This could happen, for example, if the "size" of $\Sigma\eta$ is small or if there are comovements also at business cycle frequencies.

After having established that, under mild assumptions, a cointegrated structural time series model admits a VAR representation, the next step is to derive the VAR coefficients from the parameters of the structural model. Koopman and Harvey (1999) have proposed a numerical algorithm that can be adopted, as it is valid for any state space model. Their result is outlined in the following section.

[3] To check condition (c) is not necessary as it is implied by (b).

6.2.4 Computing the VAR coefficients

Koopman and Harvey (1999) have obtained algorithms to compute the observation weights for signal extraction and filtering in a state space model like

$$\mathbf{x}_t = Z_t\alpha_t + d_t + \varepsilon_t, \tag{6.13}$$

$$\alpha_t = G_t\alpha_{t-1} + c_t + R_t\eta_t; \tag{6.14}$$

see section 2.2 for the details on the state space representation of time series.

The filtered estimator of the state vector in the next time period, that is the estimator of α_t based on information available at time $t-1$, can be written as

$$a_{t|t-1} = \sum_{j=1}^{t-1} w_j \mathbf{x}_j,$$

where the w_j's are appropriated weights, that can be computed using the algorithm in Koopman and Harvey (1999). Then the one-step ahead predictor of $\mathbf{x}_t$ becomes

$$\widehat{\mathbf{x}}_{t|t-1} = \sum_{j=1}^{t-1} w_j^* \mathbf{x}_j,$$

where $w_j^* = Z_t w_j$.

Since $\mathbf{x}_t = \widehat{\mathbf{x}}_{t|t-1} + \mathbf{v}_t$, with $\mathbf{v}_t$ being the one-step ahead prediction error, the weights w_j^* are the coefficients of the autoregressive representation for a time invariant model when the filter is in steady state. Using the notation of section 6.1 we have

$$\mathbf{x}_t = \sum_{k=1}^{t-1} \Pi_k \mathbf{x}_{t-k} + \mathbf{v}_t, \qquad Var(\mathbf{v}_t) = \overline{F},$$

where $\Pi_k = w_{t-k}^*$. Clearly, for the multivariate model with common trends of section 6.2.1 it will hold that $|\mathrm{I} - \sum_k \Pi_k| = 0$; see Koopman and Harvey (1999) for a numerical example.

6.3 Estimation and test

The common trends test of chapter 4, section 4.4, can be naturally applied to a structural time series model of the type (6.5)-(6.7) to make inference on the

number of cointegrating relations. In a model with fixed slope, the statistic $\xi_{K,N}$ (4.18), constructed using $\mathbf{z}_t = (1, t)'$ as regressors, tests the null hypothesis that the cointegration space has dimension $R = N - K$ against the alternative hypothesis that the dimension is lower; note the direction of the alternative hypothesis, opposite with respect to Johansen's tests. In practice, one can compute $\xi_{0,N}$, $\xi_{1,N}$, ... sequentially, stopping at the first non-rejection (although, strictly speaking, this sequential procedure would be affected by pretesting bias).

A structural time series model which incorporates common trends restrictions can then be estimated. Under Gaussianity, maximum likelihood estimation of a model like (6.5)-(6.7) can be carried out by putting it in state space form and using the (multivariate) Kalman filter to evaluate the Gaussian likelihood function in the time domain; see section 2.2 [4].

In particular, the state space form for the common trend model (6.8)-(6.9), when $\mathbf{u}_t$ is a white noise, can be obtained by setting

$$Z_t = \begin{pmatrix} \Theta_1 & \mathbf{0} \\ \Theta_2 & \mathbf{I}_R \end{pmatrix}, \; c_t = \begin{pmatrix} \boldsymbol{\delta} \\ 0 \end{pmatrix}, \; R_t = \begin{pmatrix} \mathbf{I}_K \\ \mathbf{0} \end{pmatrix}, \; G_t = \mathbf{I}_N, \; d_t = \mathbf{0}$$

in (6.13)-(6.14), with the $N \times K$ matrix of factor loadings Θ partitioned as $\Theta = (\Theta_1', \Theta_2')'$, where Θ_1 is $K \times K$ and Θ_2 is $R \times K$.

Note that in this formulation $\boldsymbol{\delta}$ is treated as an unknown parameter whereas the initial condition $\overline{\boldsymbol{\mu}}$ is included in the state vector. An alternative approach is to treat also $\overline{\boldsymbol{\mu}}$ as an unknown parameter, to be estimated jointly with $\boldsymbol{\delta}$ by the GLS algorithm of section 2.3.

Given Θ, an asymptotically efficient estimator for $\boldsymbol{\delta}$, alternative to the GLS, is given by

$$\widetilde{\boldsymbol{\delta}}(\Theta) = (\Theta'\Theta)^{-1}\Theta'\overline{\Delta \mathbf{x}_t},$$

where $\overline{\Delta \mathbf{x}_t} = (\mathbf{x}_T - \mathbf{x}_1)/(T-1)$. This is obtained from concentrating $\boldsymbol{\delta}$ out of the spectral likelihood function for the observations in first differences, and

[4] The menu-driven program Stamp 5.0 by Koopman et al. (1995) is designed to estimate most of the standard (multivariate) structural time series models, also with common components and regression effects. Alternatively, more flexibility can be achieved by using the matrix programming language Ox 2.1 together with the set of routines for state space models implemented in Ssfpack 2.2; see Doornik (1998) and Koopman et al. (1998).

intuitively from the consideration that only the contribution at frequency zero matters for the estimation of $\boldsymbol{\delta}$; see Harvey (1989, p.453) for the expression of the spectral likelihood and the derivation of the estimator above.

A number of diagnostic tests is available to assess the adequacy of the structural time series model fitted to the data. These are constructed from the one step ahead prediction errors (or "residuals") in a similar way as in the Box and Jenkins methodology; see Harvey (1989 ch. 5) and Harvey and Koopman (1992).

As discussed in section 6.2.2, an important empirical issue is to identify the cointegrating relations, or -in other words- to test for restrictions on the cointegration space. A simple nonparametric test for a pre-specified cointegration vector $\mathbf{b}$ is given by the KPSS test (3.22) applied to $\mathbf{b}'\mathbf{x}_t$. Similarly, the nonstationarity test of section 4.2 can be used when multiple cointegrating relations are to be tested simultaneously.

Restrictions on the factor loadings matrix $\boldsymbol{\Theta}$ that do not change its rank (overidentifying restrictions) can be tested using a Likelihood Ratio test. If the model is correctly specified, under the null hypothesis the statistic $-2(l_r - l_u)$, where l_r, l_u are respectively the restricted and unrestricted log-likelihood functions, is asymptotically χ^2 with degrees of freedom equal to the number of restrictions (provided they do not change the rank of $\boldsymbol{\Theta}$).

Note that restrictions on $\boldsymbol{\Theta}$ imply restrictions on the cointegration matrix $\mathbf{B}$ via the relation (6.10). Thus the LR approach can also be interpreted as a test for overidentifying restrictions on $\mathbf{B}$. As an example, consider a model with three variables and one trend, thus with $\boldsymbol{\Theta} = (1, \theta_1, \theta_2)'$ and

$$\mathbf{B}' = \begin{pmatrix} 1 & b_{12} & b_{13} \\ 1 & b_{22} & b_{23} \end{pmatrix}.$$

Here the constraint $\theta_1 = \theta_2 = \theta$ corresponds to the cross equation restriction

$$b_{12} + b_{13} = b_{22} + b_{23}. \tag{6.15}$$

This can then be tested using the LR test described above. Note also that adding another restriction, e.g. $b_{12} = 0$, to (6.15) is sufficient for identification.

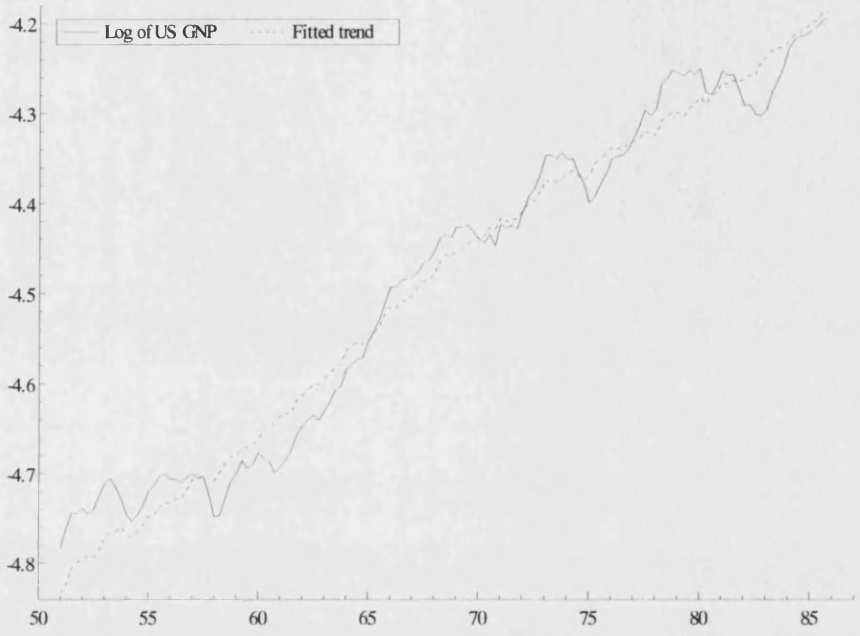

Figure 6.1: US Real GNP and estimated stochastic trend

6.4 An example of cointegrated structural time series model

As an illustrative example, we consider a bivariate model for (the logarithm of) US GNP and Investment. The data are quarterly, seasonally adjusted and refer to the period 1951-1985. They are part of a dataset which is provided with STAMP 5.0.

Figures 6.1-6.2 show the plot of the series. A cyclical component appears quite evident, particularly in the investment series. Harvey and Jaeger (1993) fitted a plausible univariate structural model to the data, yielding a stochastic cycle with estimated period of about 20 quarters (5 years). This seems to be a good characterization of US business cycle for the years covered by the sample.

Economic theory and numerous empirical studies suggest that the series of GNP and Investment should be nonstationary and cointegrated, possibly with a cointegration vector equal to $(1 \ -1)'$. It is also plausible that the cyclical

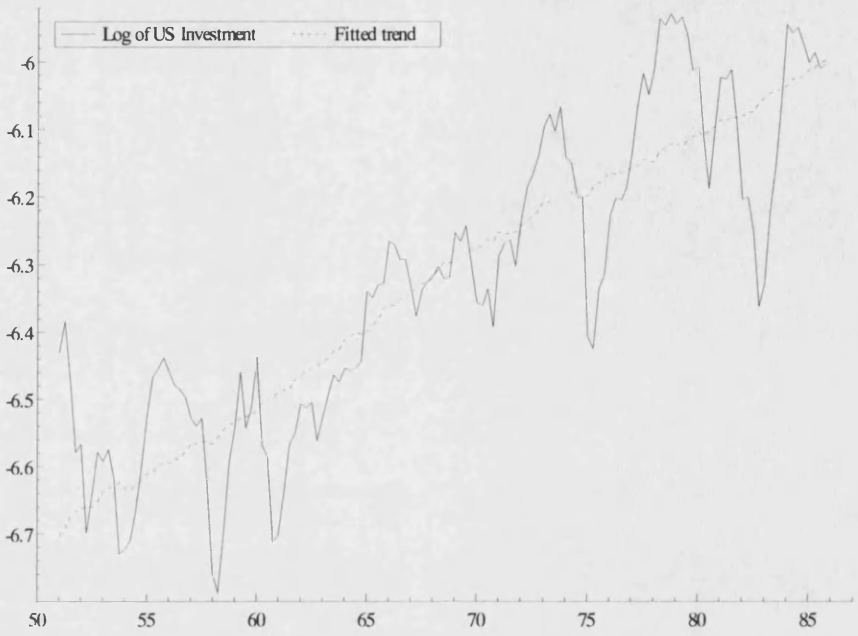

Figure 6.2:

components of the series have similar characteristics in terms of periodicity, as in the definition of "similar cycles" of Koopman et al. (1995), where the stochastic cycles are constrained to share the same autocorrelation function.

The table below reports the outcome of the common trend test of chapter 4, section 4.4, for a range of values of the lag truncation parameter m. The 10% and 5% critical values, taken from Nyblom and Harvey (2000), are for trending series (and no structural breaks). The null hypothesis of trend stationarity ($K = 0$) is rejected at 5% significance level for $m \leq 6$ and at 10% level for $m \leq 9$, while the hypothesis $K = 1$ is not rejected even for $m = 2$. Thus the evidence suggests that the series are nonstationary with a common trend, or CI(1,1). The same test applied to the first differences of the data does not provide evidence of I(2) behaviour.

$\xi_{K,2}$ test	$m=0$	$m=2$	$m=6$	$m=9$	$m=10$	**10%**	**5%**
$K=0$	1.504	0.532	0.263	0.212	0.203	0.211	0.247
$K=1$	0.177	0.073	0.051	0.053	0.055	0.085	0.105

The first model that we fit to the data is then the simple common trend with fixed slope and similar cycles (plus an irregular component). This corresponds to (6.8)-(6.9) with $\mathbf{u}_t$ being the sum of a similar cycles component and a white noise. We call it Model 1.

The estimation results for Model 1 are not fully reported to save space, but the picture is the following. The model seems to fit quite well the Investment series but not so well the GNP series, for which the residuals (one step ahead prediction errors) appear to be serially correlated. Indeed, they fail the Box-Ljung diagnostic test. The estimated cycle has period of 4.97 years and closely agrees with the results of Harvey and Jaeger (1993). The estimated matrix of factor loadings is $\widehat{\Theta}=(1,\ 1.34)'$, which corresponds to the cointegrating vector $\widehat{\mathbf{b}}=(1,\ -0.75)$. Relaxing the common trend restriction does not lead to any substantial improvement.

The problem with the previous specification seems to be that the GNP series behaves somewhat differently from the investment series at the business cycle frequencies, and this cannot be fully captured by the similar cycles restriction. One possible solution is to add a stationary VAR(1) component, to model the short run in a richer way. This also allows a dynamic interaction between GNP and investment, which is ruled out by the common trend plus similar cycles specification.

In the tables below, in the columns labelled Model 2, are contained the most important estimation results and diagnostics for this specification, i.e. common trend + similar cycles + stationary VAR(1) + irregular component. As can be seen, the model provides a good fit to the data and successfully passes the main diagnostic tests; see Koopman et al. (1995) for full explanations on how to interpret the results. A nice feature is the estimate of the factor loading matrix, $\widehat{\Theta}=(1, 0.97)'$, in line with the prior from economic theory. However the

estimated stochastic cycle component presents a different periodicity from what expected (3.18 years against around 5 years).

	Model 2		Model 3	
DIAGNOSTICS	**GNP**	**INV.**	**GNP**	**INV.**
Std. Error ($*10^{-2}$)	0.974	5.102	0.970	5.173
Normality	0.935	8.966	0.611	10.12
H(46)	0.887	0.827	0.973	0.867
r(1)	0.099	-0.031	0.094	0.070
r(12)	0.026	-0.076	-0.063	-0.094
Durbin-Watson	1.783	2.036	1.807	1.820
Box-Ljung: Q(12,6)	13.13	9.771	11.99	11.51
Rd^2	0.144	0.208	0.150	0.186
Log-likelihood	1100.67		1099.84	

	Model 2		Model 3	
EST. RESULTS	**GNP**	**INV.**	**GNP**	**INV.**
Std. Dev. Irr. ($*10^{-3}$)	0.000	8.704	0.000	12.06
Std. Dev. Level ($*10^{-3}$)	6.407	6.241	5.836	5.836
Std. Dev. Cycle ($*10^{-3}$)	5.193	35.78	6.033	42.38
Std. Dev. AR(1) ($*10^{-3}$)	2.132	23.62	2.624	15.45
Slope Coeff. (% yearly)	1.780	1.980	1.860	2.027
Factor Loadings Θ	$(1,0.97)'$		$(1, 1)'$ restricted	
Cycle Period (years)	3.183		4.93 restricted	
Cycle Rho	0.867		0.886	
AR(1) Coefficients	$\begin{pmatrix} 0.568 & 0.113 \\ -1.641 & 1.285 \end{pmatrix}$		$\begin{pmatrix} 0.644 & 0.196 \\ -0.488 & 1.207 \end{pmatrix}$	

The presence of a cointegration vector $\mathbf{b} = (1, -1)'$ can be tested nonparametrically via a KPSS test on the series $\log(INV) - \log(GNP)$. As expected after the estimation of the loadings, the restriction is easily accepted in both cases of computing the statistic with and without time trend. The results are in the next table, with critical values taken from Kwiatowski et al. (1992).

KPSS test	m =0	m =1	m =2	m =7	m =10	**10%**	**5%**
Without slope	0.719	0.396	0.291	0.191	0.193	0.347	0.463
With slope	0.187	0.104	0.077	0.052	0.054	0.119	0.146

The restriction can also be tested by a LR test on the loadings, i.e. by testing $\theta_2 = 1$, as we have fitted a succesful parametric model to the data (and the restriction does not change the rank of the matrix of loadings). Clearly, the restriction is not rejected: the LR statistic is virtually zero.

The cointegration restriction $\mathbf{b} = (1, -1)'$ is then imposed on Model 2, together with a restriction on the period of the stochastic cycle which is set to 4.93 years[5] . The main diagnostics and estimation results for this restricted model, which we call Model 3, are reported in the tables above in the appropriate columns. As can be seen, the main effect of the restrictions is to increase the volatility of the cycle (with respect to Model 2) and smooth out the trend. The goodness of fit measures for the GNP series have improved, the downside being some loss of fit for the Investment series. By comparing the likelihood functions, it is also clear that the restricted model is not rejected.

The estimated components for Model 3, obtained using all the observations (smoothed estimates), are depicted in figures 6.1-6.3: the extracted trends are showed in figures 6.1-6.2 and the short run components, stochastic cycle + AR(1), in figure 6.3. Note the higher volatility of Investment.

A direct examination of the multivariate AR(1) component is also interesting. First note that the roots of the AR polynomial are complex conjugate, yielding the pseudo cyclical behaviour depicted in figure 6.4. The graph, then, seems to suggest that the Investment series leads the GNP one, which is also plausible by economic arguments. Indeed, the cross correlation coefficient between the AR(1) component of GNP and that of Investment lagged twice is 0.990, while for lags of Investment equal to 0, 1, 3 the correlation coefficient is 0.937, 0.975, 0.986 respectively.

[5] This figure is the estimate period of the cycle for Investment in a univariate model.

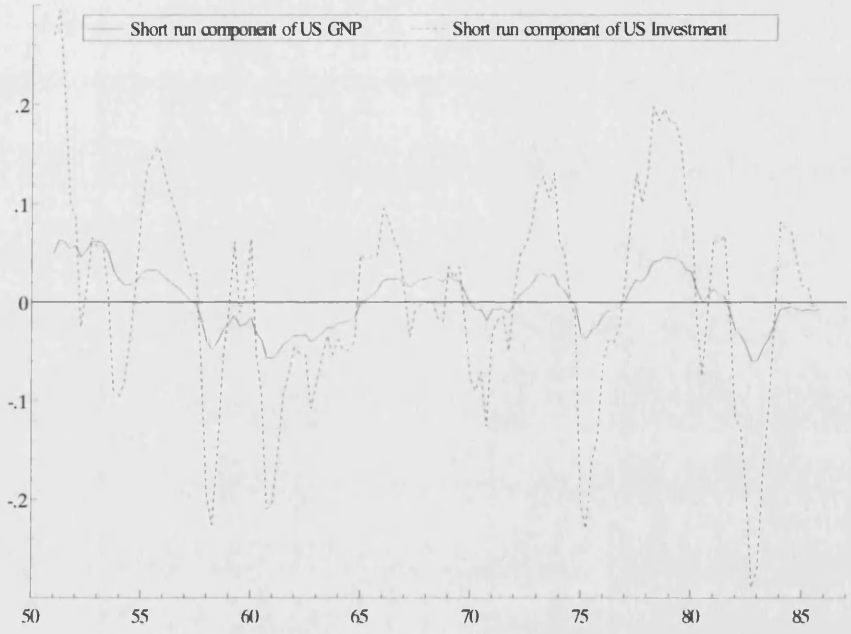

Figure 6.3: Cyclical Components

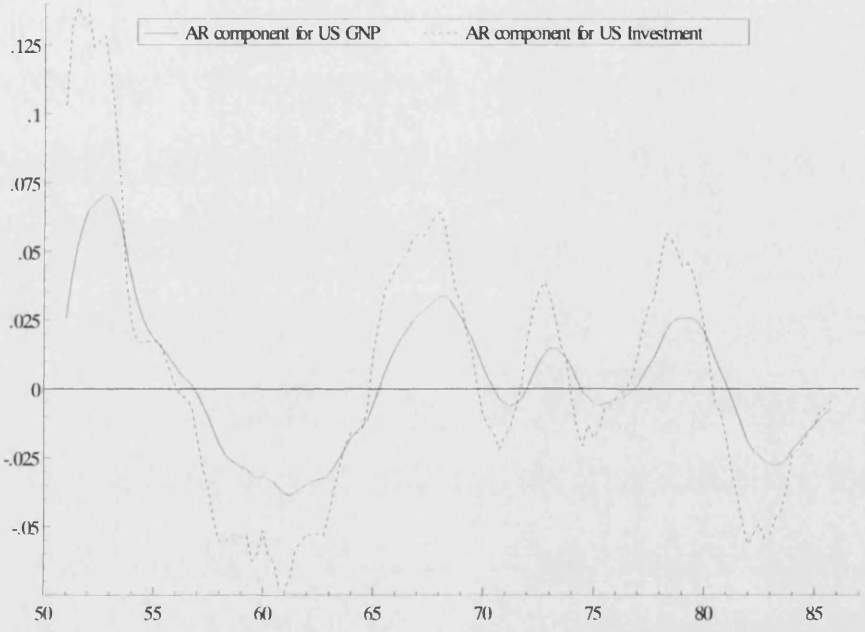

Figure 6.4: AR(1) Components

Chapter 7

Estimation and tests of certain dynamic panel data models

In this chapter we consider panel data models where the individual effects are modeled as individual specific random walks. The likelihood function of the models is written in the prediction error decomposition form as explained in chapter 2. Imposing the constraint of a common signal-to-noise ratio across individuals makes the maximum likelihood estimator computationally feasible also when the number of units in the cross section is large. The reason is that the same univariate Kalman filter applies to each equation in turn and numerical optimization is with respect to a single parameter only. Testing for exogeneity of the regressors and estimation with few time observations are discussed. An average LBI test for fixed effects is proposed, closely related to the cointegration test of Mc-Coskey and Kao (1998). Finally, as an example, we estimate and test a Cobb-Douglas production function from a panel of US manifacturing firms.

7.1 Introduction

Most of the literature on dynamic panel data models is centered on autoregressive models with time invariant individual effects. An important issue in that framework is that most of the commonly used estimators (OLS, GLS, LSDV) are not consistent when the time dimension of the panel is small; see e.g. Nickell (1981), Hsiao (1986). Proposed solutions include writing the model in first differences and estimating it by the Generalized Methods of Moments, using past

values of the series to instrument the lagged endogenous variable. However, if -as often is the case- the series are very persistent the instruments are only weakly correlated, and thus the properties of the GMM estimator are generally poor; see -inter alia- Arellano and Bond (1991), Ahn and Schmidt (1995, 1997), Blundell and Bond (1998, 1999) for the application of the GMM technique to panel data and Blundell et al. (2000) for a discussion of the issue of weak instruments in this context.

The panel data models considered in this chapter assume, on the other hand, a specific time varying structure for the individual effect, namely that of a random walk, but other processes could be considered with minor modifications of the arguments. Giving this structure to the individual effect can be sensible in many instances, for example in the case of a production function where it can be identified with the firm's (unobserved) technical progress or the Solow residual.

The random walks representing the individual effects are constrained to share the same signal-to-noise ratio across individuals, similarly to the autoregressive case where the same coefficient applies for the lagged endogenous variables. Further heterogeneity is permitted by leaving unrestricted the variances of the noise.

This chapter has then some points in common with some recent literature that extends well known time series results, especially on unit roots and cointegration, to the case of panel data, see e.g. Im et al. (1997), Kao (1999), Phillips and Moon (2000). As in that literature, we tipically assume a large time dimension of the panel, although we also discuss the case of few time periods. Indeed, testing for fixed effects in our framework is closely related to testing for cointegration, as done in McCoskey and Kao (1998).

The models we consider can be estimated using state-space techniques to obtain a consistent and fully efficient estimator under the assumption of strict exogeneity of the regressors. The likelihood function of the model is written in the prediction error decomposition form using the Kalman filter algorithm to obtain the innovations, as explained in chapter 2. Since the same univariate Kalman filter is applied to each equation in turn and the Kalman filter recursions depend on a single parameter, evaluating and maximizing this likelihood

function is computationally feasible even when the number of individuals in the panel is large. Indeed, numerical optimization is required only for one scalar parameter, the common signal-to-noise ratio. This is the subject of sections 7.2-7.3.

In section 7.4 we propose an Hausman-type statistic for testing exogeneity of the regressors, whereas section 7.5 discusses estimation with few observations along the time dimension of the panel. The baseline model is then extended in section 7.6 by including an unobserved component common to all individuals, or common trend.

Section 7.7 considers testing for stationarity and for the presence of fixed effects. We propose an *Average LBI test,* constructed by taking the average of the univariate LBI statistics across individuals. The test is closely related to the panel unit root test of Hadri (1998) and the panel cointegration test of McCoskey and Kao (1998). Here we prove consistency of the test when only a fraction of the individuals have an unobserved random walk component. Finally, as an illustration, in section 7.8 we estimate a Cobb-Douglas production function for a panel of US manufacturing firms and we apply the tests described.

7.2 The baseline model

Let $(y_{it}, \mathbf{x}'_{it})$, $i = 1, ..., N$, $t = 1, ..., T$, be a set of observations across N individuals and for T time periods, with y_{it} scalar and $\mathbf{x}_{it}$ a $(p \times 1)$ vector. We consider the linear panel data regression model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}, \tag{7.1}$$

where the regressors $\mathbf{x}_{it}$ are strictly exogenous, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of regression coefficients and u_{it} is an error term having the following two components structure

$$u_{it} = \mu_{it} + \varepsilon_{it}, \tag{7.2}$$

$$\mu_{it} = \mu_{i,t-1} + \eta_{it}, \tag{7.3}$$

$$\mu_{i0} \sim N(0,\kappa), \quad \kappa \to \infty, \tag{7.4}$$

$$(\varepsilon_{it},\eta_{it})'|\,X \sim NID\left(0,\text{diag}(\sigma_i^2, q\sigma_i^2)\right),\ q \geq 0. \tag{7.5}$$

The two components in the error term are an individual specific stochastic trend and an irregular disturbance respectively. The disturbances $(\varepsilon_{it},\eta_{it})$ are assumed orthogonal across both time and individuals.

The individual effect is μ_{it}. It is time varying and takes the form of a random walk (stochastic trend). We impose the restriction that the *signal-to-noise ratio* q is the same across individuals[1] . The variances of the irregular components are, in general, left unrestricted, thus allowing for further heterogeneity in the panel.

The random walk is initialized with a diffuse prior (7.4). Alternatively we could treat μ_{i0} as a parameter to estimate; however, Shephard and Harvey (1990) and Shephard (1993) have shown that the diffuse prior assumption leads to better properties of the maximum likelihood estimator of q when the true value is zero. Note that if $q = 0$ and μ_{i0} is fixed we have the standard fixed effect model, whereas if $q = 0$ and κ is finite we have the standard random effect model, see e.g. Hsiao (1986).

Apart from including lagged values for the regressors, the dynamics of the model are the individual specific dynamics μ_{it} only; also no correlation is allowed across individuals. In section 7.4 the model will be extended by adding a time varying component common to all individuals ("common trend"), by which richer dynamics and cross section correlation are obtained.

A model similar to (7.1)-(7.5), but without regressors, was proposed by Marshall (1992) in the context of modelling cross sections of time series, and its extension to panel data was advocated by Harvey and Koopman (1996).

The interest lies in making inference on the vector of coefficients $\boldsymbol{\beta}$ and the signal-to-noise ratio q.

Model (7.1)-(7.5) can be written in matrix notation as

$$Y_i = X_i\boldsymbol{\beta} + U_i, \quad i = 1,...,N, \tag{7.6}$$

[1] Other processes for the individual effects could be considered (as long as constrained to share the same autocorrelation function) with minor modifications of the arguments below.

where $Y_i = (y_{i1}, ..., y_{iT})'$, $X_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{iT})'$, $U_i = (u_{i1}, ..., u_{iT})'$.

Let Ω_i =E$(U_i U_i' | X_i)$, $i = 1, ..., N$. In our model (7.1)-(7.5) $\Omega_i = \sigma_i^2(\mathrm{I}_T + qH) + \kappa \mathbf{1}_T \mathbf{1}_T'$, where H is the random walk generating matrix, i.e. a TxT matrix whose element of position (t, s) is $\min(t, s)$, and $\mathbf{1}_T$ is a $T \times 1$ vector of ones. The (unfeasible) Generalized Least Squares estimator of $\boldsymbol{\beta}$ is known to be

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{i=1}^{N} X_i' \Omega_i^{-1} X_i \right)^{-1} \sum_{i=1}^{N} X_i' \Omega_i^{-1} Y_i. \tag{7.7}$$

This is also the formula of the Maximum Likelihood Estimator of $\boldsymbol{\beta}$ when the Ω_i's are known. When the Ω_i's are not known it is possible to compute the joint likelihood of $(\boldsymbol{\beta}, \Omega_1, ..., \Omega_N)$ via the prediction error decomposition and maximize it numerically, as explained in section 2.3 for the case of a univariate model ($N = 1$). Although not generally true, for model (7.1)-(7.5) this approach is computationally feasible even for a large number of cross section units N, since the same Kalman filter can be applied to each equation in turn. This is explained in the next section. The resulting estimator is asymptotically equivalent (as $T \to \infty$) to the unfeasible GLS (7.7).

The model is extended in section 7.4 by including a trend component common to all individuals, showing that only few complications arise. In particular the common trend can be washed out by means of a "Within transformation" giving rise to a transformed model formally identical to (7.1)-(7.5). Alternatively, the common trend can be modelled explicitly and included in the likelihood function.

7.3 Estimation of the model

Consider $U_i = Y_i - X_i \boldsymbol{\beta}$ from (7.6) with E$(U_i U_i' | X_i) = \Omega_i$. It is known that there exists a lower diagonal matrix L with ones on the main diagonal and a diagonal matrix F such that

$$\Omega_i^{-1} = \sigma_i^{-2} L' F^{-1} L, \qquad i = 1, ..., N.$$

As explained in section 2.3, the above decomposition is performed by the Kalman filter. This allows to write the likelihood function in terms of innovations that are

not serially correlated but heteroschedastic ("prediction error decomposition") and the MLE/GLS estimator of $\boldsymbol{\beta}$ will be essentially a weighted least squares estimator.

Assuming a diffuse prior for the individual effect, $\mu_{i0} \sim N(0, \kappa)$ with $\kappa \to \infty$, is equivalent to maximizing the likelihood conditional on the first observations or also equivalent to dropping the first observations and obtain a proper prior for μ_{i1}, namely $\mu_{i1} \sim N(y_{i1} - \mathbf{x}'_{i1}\boldsymbol{\beta}, \sigma_i^2)$; see Harvey (1989). More precisely, given the prediction error decomposition form of the likelihood for the i-th individual,

$$\log p\left(Y_i | X_i\right) = \sum_{t=2}^{T} \log p\left(y_{it} | \mathbf{x}_{it}, \Im_{t-1}\right) + \log p\left(y_{i1} | \mathbf{x}_1, \Im_0\right),$$

where $\Im_t$ denotes the information at time t, we have that, except for constants, as $\kappa \to \infty$

$$\log p\left(Y_i | X_i\right) + \frac{1}{2} \log \kappa \to \sum_{t=2}^{T} \log p\left(y_{it} | \mathbf{x}_{it}, \Im_{t-1}\right);$$

see deJong (1988), Shephard and Harvey (1990) and the discussion at the end of section 2.2. The right hand side of the previous expression can be computed running the standard Kalman Filter with a "large κ approximation"; this is what we have done in this paper. In practice, assuming a diffuse prior implies dropping the first Kalman Filter innovation, as its variance would diverge. Alternatively, to avoid the approximation, modified Kalman Filter recursions have been proposed by deJong (1988) and Koopman (1997).

Thus, let X_i^*, Y_i^* be the $(T-1)$ dimensional vectors obtained from dropping the first element of LX_i, LY_i respectively and let F^* be the $(T-1) \times (T-1)$ matrix obtained from dropping the first row and column of F. Note that X_i^*, Y_i^* and F^* depend only on q.

Since the Kalman filter recursions runs independently of σ_i^2 (which is just a scale parameter), in our model the same univariate Kalman filter applies to each individual series. The advantage is that we don't need a multivariate filter, which would be computationally very intensive if the number of the series N is large.

The log-likelihood for (7.6) can then be written as

$$\ell_i = -\frac{T-1}{2}\log 2\pi - \frac{T-1}{2}\log\sigma_i^2 - \frac{1}{2}\log|F^*| - \frac{1}{2\sigma_i^2}(Y_i^* - X_i^*\beta)' F^{*-1}(Y_i^* - X_i^*\beta), \tag{7.8}$$

where $\ell_i = \sum_{t=2}^{T}\log p(y_{it}|\,\mathbf{x}_{it}, \Im_{t-1})$. Denote by $\ell = \ell(\beta, q, \sigma_1^2, ..., \sigma_N^2)$ the log-likelihood of the full model (7.1)-(7.5). Since the observations are independent across individuals,

$$\begin{aligned}\ell &= \sum_{i=1}^{N}\ell_i \\ &= c - \frac{T-1}{2}\sum_{i=1}^{N}\log\sigma_i^2 - \frac{N}{2}\log|F^*| - \frac{1}{2}\sum_{i=1}^{N}\sigma_i^{-2}(Y_i^* - X_i^*\beta)' F^{*-1}(Y_i^* - X_i^*\beta),\end{aligned}$$

with $c = -\frac{N(T-1)}{2}\log 2\pi$. The goal is to maximize ℓ with respect to β, q, σ_i^2, $i = 1, ..., N$.

Suppose first that we know the true value of β. Then to maximize ℓ we can first concentrate out each σ_i^2,

$$\widehat{\sigma}_i^2(\beta, q) = \frac{1}{T-1}(Y_i^* - X_i^*\beta)' F^{*-1}(Y_i^* - X_i^*\beta), \qquad i = 1, ..., N, \tag{7.9}$$

and then maximize with respect to q the resulting concentrated log-likelihood,

$$\widehat{q}(\beta) = \arg\sup_{q\geq 0}\ell_c(q, \beta), \tag{7.10}$$

where

$$\ell_c(q, \beta) = c' - \frac{T-1}{2}\sum_{i=1}^{N}\log\widehat{\sigma}_i^2(\beta, q) - \frac{N}{2}\log|F^*|,$$

with $c' = -\frac{N(T-1)}{2}(\log 2\pi + 1)$. Numerical optimization is needed only with respect to the scalar parameter q.

When β is not known, the previous procedure needs to be iterated starting from an initial value $\widehat{\beta}^{(0)}$ to be updated by

$$\widehat{\beta}^{(j+1)} = \left(\sum_{i=1}^{N}\frac{\widehat{X}_i^{*\prime}\widehat{F}^{*-1}\widehat{X}_i^*}{\widehat{\sigma}_i^2\left(\widehat{\beta}^{(j)}, \widehat{q}\left(\widehat{\beta}^{(j)}\right)\right)}\right)^{-1}\sum_{i=1}^{N}\frac{\widehat{X}_i^{*\prime}\widehat{F}^{*-1}\widehat{Y}_i^*}{\widehat{\sigma}_i^2\left(\widehat{\beta}^{(j)}, \widehat{q}\left(\widehat{\beta}^{(j)}\right)\right)}, \qquad j = 1, 2, ... \tag{7.11}$$

where $\widehat{X}_i^*$, $\widehat{F}^*$, $\widehat{Y}_i^*$ are obtained running the Kalman filter with $\widehat{q}\left(\widehat{\boldsymbol{\beta}}^{(j)}\right)$. Note that (7.11) is obtained from the "likelihood equations" (first order conditions) for $\boldsymbol{\beta}$ and it corresponds to the scoring algorithm. As usual, if the starting value $\widehat{\boldsymbol{\beta}}^{(0)}$ is a consistent estimator for $\boldsymbol{\beta}$, e.g. it is the first difference estimator (7.19) of section 4, then one iteration is sufficient for asymptotic efficiency (as $T \to \infty$). In practice, however, one iterates the previous steps J times, where J is such that $\widehat{\boldsymbol{\beta}}^{(J)} - \widehat{\boldsymbol{\beta}}^{(J-1)}$ is less than some ε. In the following, we will refer to $\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{(J)}$, $\widetilde{q} = \widehat{q}\left(\widetilde{\boldsymbol{\beta}}\right)$, $\widetilde{\sigma}_i^2 = \widehat{\sigma}_i^2\left(\widetilde{\boldsymbol{\beta}},\widetilde{q}\right)$ as the maximum likelihood estimators of $\boldsymbol{\beta}$, q, σ_i^2, $i = 1, ..., N$.

Now let $\boldsymbol{\psi} = (q, \sigma_1^2, ..., \sigma_N^2)'$ and assume that the true value of every element of $\boldsymbol{\psi}$ is strictly greater than zero, i.e. it is in the interior of parameter space.[2]

For any fixed N, the asymptotic distribution (as $T \to \infty$) of the maximum likelihood estimator is given by

$$\sqrt{T}\begin{pmatrix} \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2^{-1} \end{pmatrix}\right), \tag{7.12}$$

where

$$\begin{aligned} \mathbf{I}_1 &= \lim_{T\to\infty} T^{-1}\mathrm{E}\left(\frac{\partial^2 \ell}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right) \\ &= \lim_{T\to\infty} \mathrm{E}\left(\sum_{i=1}^{N} \frac{X_i^{*\prime} F^{*-1} X_i^*}{\sigma_i^2 T}\right), \end{aligned}$$

and

$$\mathbf{I}_2 = \lim_{T\to\infty} T^{-1}\mathrm{E}\left(\frac{\partial^2 \ell}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right).$$

An expression for $\mathbf{I}_2$ in terms of the output of the Kalman filter may be obtained using the results of Harvey (1989, p.142-143). As $N \to \infty$, (in the sense of sequential asymptotics) the asymptotic distribution of $\boldsymbol{\beta}$ is obtained after

[2] If one of the σ_i^2 is zero, the asymptotic distribution of the corresponding maximum likelihood estimator is half-normal instead of normal; see Harvey (1989, p.212). The case $q = 0$ is more complicated due to the fact that the reduced form of the model is strictly non invertible. Shephard and Harvey (1990) and Shephard (1993) derive the probability that the maximum likelihood estimator of q is equal to zero in the univariate random walk plus noise model under different assumptions on the initial conditions.

premultiplying by $\sqrt{N}$, i.e.

$$\sqrt{NT}\left(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0},\bar{\mathbf{I}}_1^{-1}\right), \tag{7.13}$$

with $\bar{\mathbf{I}}_1 = \lim_{N\to\infty} N^{-1}\mathbf{I}_1$.

The asymptotic results above, (7.12)/(7.13), hold under regularity assumptions on the process for the regressors $\mathbf{x}_{it}$. Note that the regressors are allowed to be integrated of order one as running the Kalman filter corresponding to a random walk plus noise removes a unit root at frequency zero.

Testing hypotheses on the regression coefficients $\boldsymbol{\beta}$ can be done straightforwardly in the usual framework of Wald, LR and LM tests. For example, a Wald statistic for $H_0 : R\boldsymbol{\beta} = r$ where R is a known $m \times k$ matrix with rank m and r is a known $m \times 1$ vector is

$$W = NT\left(R\widetilde{\boldsymbol{\beta}}-r\right)'\left(R\widetilde{\mathbf{V}}R'\right)^{-1}\left(R\widetilde{\boldsymbol{\beta}}-r\right), \tag{7.14}$$

where

$$\widetilde{\mathbf{V}} = \left(\sum_{i=1}^{N}\frac{\widetilde{X}_i^{*\prime}\widetilde{F}^{*-1}\widetilde{X}_i^{*}}{\widetilde{\sigma}_i^2 NT}\right)^{-1}, \tag{7.15}$$

and $\widetilde{X}_i^*$, $\widetilde{F}_i^*$ denote the output of the Kalman filter run with $\widetilde{\sigma}_i^2$, $\widetilde{q}$. Under H_0, $W \xrightarrow{d} \chi^2(m)$ as $T \to \infty$.

It is also possible to test equality of the variances in a straightforward manner. A Likelihood ratio test for $H_0 : \sigma_i^2 = \sigma^2 > 0$ for all i is given by the statistic

$$LR = -2(\ell_R - \ell_U),$$

where ℓ_R is the log-likelihood for the restricted model and ℓ_U the one for the unrestricted model. Assuming $q > 0$, we have that, as $T \to \infty$ and for fixed N, $LR \xrightarrow{d} \chi^2(N-1)$. Using the fact that a $\chi^2(k)$ random variable is the sum of k independent $\chi^2(1)$, by the Lindberg-Levi Central Limit Theorem we also have that, as $N \to \infty$ as well,

$$\frac{1}{\sqrt{3(N-1)}}\left(LR-(N-1)\right) \xrightarrow{d} N(0,1).$$

7.4 Testing exogeneity of the regressors

If the regressors are correlated with the innovations of the random walks representing the individual effects (failure of strict exogeneity), our proposed estimator of $\boldsymbol{\beta}$ (7.11) will not be consistent. However, there is an estimator which is consistent in more general situations than (7.11): the OLS in first differences.

Consider the first differences of model (7.1)-(7.5), i.e. ($i = 1, ..., N$, $t = 2, ..., T$)

$$\Delta y_{it} = \Delta \mathbf{x}'_{it}\boldsymbol{\beta} + v_{it}, \tag{7.16}$$

$$v_{it} = \eta_{it} + \Delta\varepsilon_{it}. \tag{7.17}$$

Clearly, under the assumption

$$\mathrm{E}(v_{it} | \Delta \mathbf{x}_{it}) = 0, \qquad \text{all } i, t, \tag{7.18}$$

the OLS in first differences $\widehat{\boldsymbol{\beta}}_{FD}$ is consistent, where

$$\widehat{\boldsymbol{\beta}}_{FD} = \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \mathbf{x}_{it} \Delta \mathbf{x}'_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \mathbf{x}_{it} \Delta y_{it} \right). \tag{7.19}$$

The exogeneity assumption (7.18) is not very strong, as for instance it only requires the *changes* in the regressors to be *contemporaneously* uncorrelated with the innovations driving the individual effect.

On the other hand, consistency of the GLS estimator for model (7.16)-(7.17) requires the much stronger assumption $\mathrm{E}(v_{it} | \Delta \mathbf{x}_{jt}) = 0$ for all i, j, t, s, i.e. strict exogeneity of the regressors.

Since, as $T \to \infty$, the GLS of the model in first difference is equivalent to the MLE/GLS of the model in levels $\widetilde{\boldsymbol{\beta}}$ (7.11), we can construct an Hausman type test for

$$H_0 : \mathrm{E}(v_{it} | \Delta \mathbf{x}_{js}) = 0 \qquad \text{for all } i, j, t, s$$

against

$$H_1 : \mathrm{E}(v_{it} | \Delta \mathbf{x}_{js}) = 0 \qquad t = s.$$

Under H_0 $\widetilde{\boldsymbol{\beta}}$ and the OLS first difference estimator $\widehat{\boldsymbol{\beta}}_{FD}$ are consistent and $\widetilde{\boldsymbol{\beta}}$ is efficient, whereas under H_1 only $\widehat{\boldsymbol{\beta}}_{FD}$ is consistent. Thus under H_0 we have

$$h = NT\left(\widehat{\boldsymbol{\beta}}_{FD} - \widetilde{\boldsymbol{\beta}}\right)'\left(\widehat{\mathbf{V}}_{FD} - \widetilde{\mathbf{V}}\right)^{-1}\left(\widehat{\boldsymbol{\beta}}_{FD} - \widetilde{\boldsymbol{\beta}}\right) \xrightarrow{d} \chi^2(p), \tag{7.20}$$

where $\widetilde{\mathbf{V}}$, $\widehat{\mathbf{V}}_{FD}$ are estimators of the asymptotic variances of $\widetilde{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\beta}}_{FD}$ respectively and p is the number of regressors. For the asymptotic variance of $\widetilde{\boldsymbol{\beta}}$ we can use the estimator (7.15) given in the previous section, whereas for the variance of $\widehat{\boldsymbol{\beta}}_{FD}$ we cannot use the formula for the variance of the OLS estimator as the error term in model (7.16)-(7.17) is not white noise. In fact we have that, as $(T, N) \to \infty$ sequentially,

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}_{FD} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{FD}), \tag{7.21}$$

where

$$\begin{aligned}
\mathbf{V}_{FD} &= \lim_{N\to\infty}\lim_{T\to\infty} NT\mathrm{E}\left(\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right), \\
\mathbf{A} &= \sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\mathbf{x}_{it}\Delta\mathbf{x}_{it}', \\
\mathbf{B} &= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T}\sum_{s=2}^{T} v_{it}v_{js}\Delta\mathbf{x}_{it}\Delta\mathbf{x}_{js}'.
\end{aligned}$$

Having assumed independence across i, a consistent estimator of $\mathbf{V}_{FD}$ is then

$$\widehat{\mathbf{V}}_{FD} = NT\mathbf{A}^{-1}\left\{\sum_{i=1}^{N}\sum_{\tau=-m}^{m} w(\tau, m)\sum_{t=\tau+2}^{T}\Delta\mathbf{x}_{it}\widehat{v}_{it}\widehat{v}_{i,t-\tau}\Delta\mathbf{x}_{i,t-\tau}'\right\}\mathbf{A}^{-1},$$

where $\widehat{v}_{it}$'s are the OLS residuals from model (7.16)-(7.17), $w(\tau, m)$ is a weighting function such as $w(\tau, m) = 1 - |\tau|/(m+1)$ and $m = o(T^{\alpha})$ for some $\alpha \in (0, 1]$. In the literature $\widehat{\mathbf{V}}_{FD}$ is often referred to as an *heteroskedasticity and autocorrelation consistent estimator of the variance of* $\widehat{\boldsymbol{\beta}}_{FD}$; see e.g. Andrews (1991).

Note that h is only a partial test for exogeneity as not rejecting H_0 does not automatically imply evidence of strict exogeneity; indeed, the case of contemporaneous correlation between changes in the regressors and the innovations of the individual effects is not covered.

7.5 Estimation with few time periods

As in our model the number of parameters (individual specific variances of the noise) grows at the same rate as N, when the number of time periods T is fixed we potentially have the so-called incidental parameter problem[3] , as in Neyman and Scott (1948), which is likely to affect the consistency of the maximum likelihood estimators of q and $\boldsymbol{\beta}$. Of course in the special case of same variance of the noise across individuals, i.e. in a model where $\sigma_i^2 = \sigma^2$ for all i, the resulting maximum likelihood estimator of $(\boldsymbol{\beta}, q, \sigma^2)'$ is consistent for each $T \geq 3$. Note that the first time period is "lost" because of the diffuse prior and then two more periods for each individual are needed to estimate q.

With heterogeneity in the noise, it then does not appear possible to estimate q consistently when the time dimension of the panel is short. However, the first difference estimator of the regression coefficient $\widehat{\boldsymbol{\beta}}_{FD}$ (7.19) will be consistent in any case. In fact, for $T \geq 2$, under the usual regularity conditions we have that

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{FD} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{FD}^*),$$

where

$$\mathbf{V}_{FD}^* = \lim_{N\to\infty} N\mathrm{E}\left(\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right).$$

Having assumed independence across i, a consistent estimator of $\mathbf{V}_{FD}^*$ is

$$\widehat{\mathbf{V}}_{FD}^* = N\mathbf{A}^{-1}\left\{\sum_{i=1}^{N}\sum_{\tau=-(T-2)}^{T-2}\sum_{t=\tau+2}^{T}\Delta\mathbf{x}_{it}\widehat{v}_{it}\widehat{v}_{i,t-\tau}\Delta\mathbf{x}_{i,t-\tau}'\right\}\mathbf{A}^{-1}, \tag{7.22}$$

which corresponds to White (1980) heteroschedasticity consistent estimator.

7.6 Inclusion of a common trend component

Replace equation (7.2) by

$$u_{it} = \mu_{it} + \varepsilon_{it} + \tau_t, \tag{7.23}$$

[3] Lancaster (2000) provides a review of the incidental parameter problem in econometrics.

where τ_t is a (time-varying) component common to all individuals, or "common trend".

One approach to deal with a common trend component is to wash it out by means of a "Within transformation". The transformed model can then be written as

$$
\begin{aligned}
y_{it}^{*} &= \mathbf{x}_{it}^{*\prime}\boldsymbol{\beta} + u_{it}^{*}, && (7.24)\\
u_{it}^{*} &= \mu_{it}^{*} + \varepsilon_{it}^{*}, && (7.25)\\
\mu_{it}^{*} &= \mu_{i,t-1}^{*} + \eta_{it}^{*}, && (7.26)\\
\mu_{i0}^{*} &\sim N(0,\kappa),\ \kappa \to \infty, && (7.27)\\
(\varepsilon_{it}^{*}, \eta_{it}^{*})' &\sim NID\left(0,\mathrm{diag}(\sigma_{*i}^{2}, q\sigma_{*i}^{2})\right),\ q \geq 0, && (7.28)
\end{aligned}
$$

where for a variable z_{it} we have let $z_{it}^{*} = z_{it} - \overline{z}_t$ and $\overline{z}_t = N^{-1}\sum_{i=1}^{N} z_{it}$. The model is now the same as before with $\sigma_{*i}^{2} = \sigma_i^2(1-N^{-1})$. No further complications arise except that we have not modeled the common trend.

Another approach is to specify a form for the common trend τ_t and maximize the likelihood of the full model. For example, consider the model defined by (7.1), (7.23), (7.3)-(7.5) with $\sigma_i^2 = \sigma^2$ for all i, and assume

$$
\begin{aligned}
\tau_t &= \mu_t^{+} + \varepsilon_t^{+}, && (7.29)\\
\mu_t^{+} &= \mu_{t-1}^{+} + \eta_t^{+}, && (7.30)\\
\mu_0^{+} &\sim N(0,\kappa),\ \kappa \to \infty, && (7.31)\\
(\varepsilon_t^{+}, \eta_t^{+})' &\sim NID\left(0,\mathrm{diag}(\sigma_{+}^{2}, q_{+}\sigma_{+}^{2})\right),\ q_{+} \geq 0, && (7.32)
\end{aligned}
$$

where the disturbances $(\varepsilon_t^{+}, \eta_t^{+})$ are also orthogonal to $(\varepsilon_{it}, \eta_{it})$, $i = 1, ..., N$, for all leads and lags. This model, without regressors, was proposed by Marshall (1992) to model small cross sections of time series and then was advocated by Harvey and Koopman (1996) also for the case of panel data. It can be handled by assuming

$$
N^{-1}\sum_{i=1}^{N}\mu_{it} = N^{-1}\sum_{i=1}^{N}\varepsilon_{it} = 0 \qquad (7.33)
$$

to achieve identification of the stochastic trends and considering the following

transformation ($t = 1, ..., T$):

$$y_{it}^* = \mathbf{x}_{it}^{*\prime}\boldsymbol{\beta} + u_{it}^*, \quad i = 1, ..., N-1, \tag{7.34}$$

$$\overline{y}_t = \overline{\mathbf{x}}_t'\boldsymbol{\beta} + \tau_t. \tag{7.35}$$

To justify (7.34)-(7.35), we need further notation. Let $\mathbf{y}_t$ and $\mathbf{X}_t$ be respectively the $(N \times 1)$ vector and the $(N \times p)$ matrix containing the observations for all the individuals at time t, $t = 1, ..., T$. Then we can rewrite the model as

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \tag{7.36}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \tag{7.37}$$

$$\boldsymbol{\mu}_0 \sim N(\mathbf{0}, \kappa\mathbf{I}_N),\ \kappa \to \infty, \tag{7.38}$$

$$(\boldsymbol{\varepsilon}_t', \boldsymbol{\eta}_t')' \sim NID\ (0, \mathrm{diag}\,(\boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\Sigma}_\eta)), \tag{7.39}$$

where $\boldsymbol{\Sigma}_\varepsilon = \sigma_+^2\mathbf{1}_N\mathbf{1}_N' + \sigma^2(\mathrm{I} - N^{-1}\mathbf{1}_N\mathbf{1}_N')$, $\boldsymbol{\Sigma}_\eta = q_+\sigma_+^2\mathbf{1}_N\mathbf{1}_N' + q\sigma^2(\mathrm{I} - N^{-1}\mathbf{1}_N\mathbf{1}_N')$. Note that the i-th components of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\mu}_t$ are respectively $\varepsilon_{it} | \, N^{-1}\sum_{i=1}^N \varepsilon_{it} = 0$ and $\mu_{it} | \sum_{i=1}^N \mu_{it} = 0$; in other words $\boldsymbol{\Sigma}_\varepsilon$, $\boldsymbol{\Sigma}_\eta$ have been obtained by using the restriction (7.33). Consider the $(N \times N)$ matrix $\mathbf{H}$ with entries

$$[\mathbf{H}]_{ij} = \begin{cases} 1 - 1/N & \text{for } i = j < N, \\ 1/N & \text{for } i = N, \\ -1/N & \text{otherwise,} \end{cases}$$

and determinant equal to $1/N$. Premultiplying (7.36) by $\mathbf{H}$ yields (7.34)-(7.35).

The attractiveness of (7.34)-(7.35) is that, apart from $\boldsymbol{\beta}$, the two equations are independent. Thus the log-likelihood is just the sum of the log-likelihoods corresponding to each of the two equations. Sharing the same $\boldsymbol{\beta}$ doesn't pose particular problems in the estimation process, since it can be concentrated out.

Thus, denoting by $X_i^*(q)$, $Y_i^*(q)$, $F^*(q)$ the output of running the Kalman filter on the deviations from the time averages for the i-th individual and by $\overline{X}_i^*(q_+)$, $\overline{Y}_i^*(q_+)$, $\overline{F}^*(q_+)$ the output of running the Kalman filter on the averages, we have that the log-likelihood for (7.34)-(7.35) can be written as

$$\ell = \ell_1 + \ell_2,$$

where

$$\begin{aligned}
\ell_1 &= c - \frac{(N-1)(T-1)}{2}\log\sigma^2 - \frac{N-1}{2}\log|F^*(q)| - \\
&\quad -\frac{1}{2\sigma^2}\sum_{i=1}^{N-1}\left(Y_i^*(q) - X_i^*(q)\boldsymbol{\beta}\right)' F^*(q)^{-1}\left(Y_i^*(q) - X_i^*(q)\boldsymbol{\beta}\right), \\
\ell_2 &= c_+ - \frac{T-1}{2}\log\sigma_+^2 - \frac{1}{2}\log\left|\overline{F}^*(q)\right| - \\
&\quad -\frac{1}{2\sigma_+^2}\left(\overline{Y}^*(q_+) - \overline{X}^*(q_+)\boldsymbol{\beta}\right)'\overline{F}^*(q_+)^{-1}\left(\overline{Y}^* - \overline{X}^*(q_+)\boldsymbol{\beta}\right).
\end{aligned}$$

Proceeding as in section 7.3 we obtain

$$\begin{aligned}
\widehat{\sigma}^2(\boldsymbol{\beta},q,q_+) &= \frac{1}{(N-1)(T-1)}\sum_{i=1}^{N-1}\left(Y_i^*(q) - X_i^*(q)\boldsymbol{\beta}\right)' F^*(q)^{-1}\left(Y_i^*(q) - X_i^*(q)\boldsymbol{\beta}\right), \\
\widehat{\sigma}_+^2(\boldsymbol{\beta},q,q_+) &= \frac{1}{(T-1)}\left(\overline{Y}^*(q_+) - \overline{X}^*(q_+)\boldsymbol{\beta}\right)'\overline{F}^*(q_+)^{-1}\left(\overline{Y}^*(q_+) - \overline{X}^*(q_+)\boldsymbol{\beta}\right), \\
\widehat{\boldsymbol{\beta}}(q,q_+) &= \left(\sum_{i=1}^{N-1}\frac{X_i^*(q)'F^*(q)^{-1}X_i^*(q)}{\widehat{\sigma}^2\left(\widehat{\boldsymbol{\beta}},q,q_+\right)} + \frac{\overline{X}^*(q)'\overline{F}^*(q)^{-1}\overline{X}^*(q)}{\widehat{\sigma}_+^2\left(\widehat{\boldsymbol{\beta}},q,q_+\right)}\right)^{-1} \cdot \\
&\quad \cdot\left(\sum_{i=1}^{N-1}\frac{X_i^*(q)'F^*(q)^{-1}Y_i^*(q)}{\widehat{\sigma}^2\left(\widehat{\boldsymbol{\beta}},q,q_+\right)} + \frac{\overline{X}^*(q)'\overline{F}^*(q)^{-1}\overline{Y}^*(q)}{\widehat{\sigma}_+^2\left(\widehat{\boldsymbol{\beta}},q,q_+\right)}\right).
\end{aligned}$$

Thus the maximum likelihood estimator of (q, q_+) is

$$(\widetilde{q},\widetilde{q}_+) = \arg\sup_{q,q_+\geq 0}\ell_c(q,q_+),$$

where

$$\begin{aligned}
\ell_c(q,q_+) &= c^* - \frac{T-1}{2}\log\widehat{\sigma}_+^2\left(\widehat{\boldsymbol{\beta}},q,q_+\right) - \frac{1}{2}\log\left|\overline{F}^*(q_+)\right| - \\
&\quad -\frac{T-1}{2}\sum_{i=1}^{N-1}\log\widehat{\sigma}^2\left(\widehat{\boldsymbol{\beta}},q,q_+\right) - \frac{N-1}{2}\log|F^*(q)|.
\end{aligned}$$

Note that, as $N \to \infty$, the maximum likelihood estimator of $\boldsymbol{\beta}$ will be asymptotically equivalent to the estimator obtained in section 7.3 applied to the deviations from the averages, so modelling the common trend provides no advantage asymptotically.

7.7 Testing stationarity and fixed effects

This section is concerned with testing for the presence of fixed effects in dynamic panel data (or testing for stationarity if explanatory variables are not included in the model).

We consider the following model, slightly different from the one in section 7.1:

$$
\begin{aligned}
y_{it} &= \mathbf{z}_{it}'\boldsymbol{\gamma}_i + u_{it}, && (7.40)\\
u_{it} &= \mu_{it} + \varepsilon_{it}, && (7.41)\\
\mu_{it} &= \mu_{i,t-1} + \eta_{it} && (7.42)\\
\mu_{i0} &= 0, && (7.43)\\
(\varepsilon_{it}, \eta_{it})' &\sim NID\left(0, \text{diag}(\sigma_i^2, q_i\sigma_i^2)\right),\ q_i \geq 0, && (7.44)
\end{aligned}
$$

where the regressors $\mathbf{z}_{it}$ and corresponding coefficients $\boldsymbol{\gamma}_i$ can be partitioned as $\mathbf{z}_{it} = (1, \mathbf{x}_{it}')'$, $\boldsymbol{\gamma}_i = (\alpha_i, \boldsymbol{\beta}_i')'$. In this model the α_i's are the individual specific intercepts and that both the regression coefficients $\boldsymbol{\beta}_i$ and the signal-to-noise ratios q_i are individual specific. As in section 7.1, we also assume independence across individuals.

We are interested in testing

$$H_0 : q_i = 0, \text{ for all } i,$$

against

$$H_1 : q_i > 0,\ i = 1, ..., N_1,\ \ q_i = 0,\ i = N_1 + 1, ..., N.$$

Note that the alternative hypothesis allows fixed effects for some of the individuals and random effects (in the form of random walks) for others. We immediately anticipate that consistency of the test requires $\lim_{N\to\infty} N_1/N > 0$.

Consider first each individual $i = 1, ..., N$ in turn. As showed in chapter 3,

section 3.1, the LBI test statistic for $H_0 : q_i = 0$ against $H_1 : q_i > 0$ is given by

$$\xi_i = \frac{\sum_{t=1}^{T} \left(\sum_{s=1}^{t} \widehat{u}_{is} \right)^2}{T \sum_{t=1}^{T} \widehat{u}_{it}^2}, \tag{7.45}$$

where the $\widehat{u}_{it}$'s are the OLS residuals from regressing y_{it} on $\mathbf{z}_{it}$.

The test we propose for our panel data model (7.40)-(7.44) is obtained by taking the average of the N individual LBI statistics ξ_i, i.e. we consider the statistic

$$\overline{\xi} = N^{-1} \sum_{i=1}^{N} \xi_i, \tag{7.46}$$

and the test will reject H_0 when $\overline{\xi} > k$, where k is an appropriate critical value. We call this test *Average LBI test.* It is constructed in an analogous manner to the Average LM test for unit roots proposed by Im, Pesaran, Shin (1997).

If the error terms ε_{it} are not white noise (and the regressors are exogenous), we can apply the KPSS correction to $\overline{\xi}$ by replacing $\sum_{t=1}^{T} \widehat{u}_{it}^2$ with $T\widehat{\omega}_i$, where $\widehat{\omega}_i$ is a consistent estimator of the long run variance of ε_{it}; see section 3.5. The limiting distribution of the resulting statistic will be unchanged.

In the two subsections below we derive the limiting distribution of $\overline{\xi}$ under H_0 first for a model without explanatory variables (test for stationarity) and then with explanatory variables (test for fixed effects). In both cases we have asymptotic normality. Then in the third subsection below we show the consistency of the test.

Our tests are closely related to the tests of Hadri (1998) and McCoskey and Kao (1998), who consider an LBI test for panel data models where the same variance for the noise and the same signal-to-noise ratio across individuals are assumed[4] . The LBI statistic for their model can be derived using theorem 6 of chapter 2 and it is asymptotically equivalent to our statistic $\overline{\xi}$. Then they also propose to use $\overline{\xi}$ to account for heterogeneity in the noise, although they do not prove consistency of the test when only a fraction of the individuals have random effects (in the form of a random walk component).

[4] McCoskey and Kao (1998) interpret the test as a test for cointegration in panel data.

7.7.1 The distribution of the test statistic for stationarity

Let $\mathbf{z}_{it} = 1$ and $\boldsymbol{\gamma}_i = \alpha_i$ in (7.40), i.e. consider a model without explanatory variables. For each $i = 1, ..., N$ we know from chapter 3 that

$$\xi_i \xrightarrow{d} V_i \equiv \int_0^1 B(r)^2 dr,$$

where $B(r)$ is a standard Brownian bridge and thus V_i is a random variable with a Cramer-von Mises distribution. Further, V_i is independent of V_j for all $i \neq j$.

Then it suffices to know the first two moments of V_i to apply the Lindberg-Levi central limit theorem and obtain the asymptotic distribution of $\overline{\xi}$.

From the series expansion of V_i (see section 3.3),

$$V_i \equiv \sum_{j=1}^{\infty} (\pi j)^{-2} \chi_j^2(1),$$

we easily obtain $\mathrm{E}(V_i) = 1/6$, $Var(V_i) = 1/45$. Therefore we can conclude that, as $T \to \infty$ followed by $N \to \infty$,

$$\sqrt{N}\left(\frac{\overline{\xi} - a}{b}\right) \xrightarrow{d} N(0,1), \tag{7.47}$$

where $a = 1/6$, $b = \sqrt{1/45}$.

If an individual specific slope is included, i.e. $\mathbf{z}_{it} = (1, t)'$ and $\boldsymbol{\gamma}_i = (\alpha_i, \delta_i)'$, we still get (7.47) but with $a = 1/15$, $b = \sqrt{11/6300}$. This follows more easily from expanding the characteristic function of V_i, see Hadri (1998).

7.7.2 The distribution of the test statistic for fixed effects

When we include regressors in the model, i.e. when $\mathbf{z}_{it} = (1, \mathbf{x}_{it}')'$ and $\boldsymbol{\gamma}_i = (\alpha_i, \boldsymbol{\beta}_i')'$, in principles the asymptotic distribution of the statistic depends on the marginal process for the regressors. We then consider two cases.

CASE 1: Stationary regressors.

If the regressors are stationary and exogenous, using proposition 17 of chapter 4, we obtain that for each $i = 1, ..., N$ under H_0

$$T^{-\frac{1}{2}} \sum_{s=1}^{t} \widehat{u}_{is} = T^{-\frac{1}{2}} \sum_{s=1}^{t} (\varepsilon_{is} - \overline{\varepsilon}_i) + o_p(1),$$

where $\overline{\varepsilon}_i = T^{-1}\sum_{t=1}^{T}\varepsilon_{is}$. Thus, the partial sum of residuals converges to a standard Brownian bridge and ξ_i converges to a random variable with a Cramer-von Mises distribution. Then the result of the previous subsection applies and the asymptotic distribution of $\overline{\xi}$ does not depend on the marginal process of the regressors.

Note that we can test if the regressors are stationary by applying this same average LBI test to $\mathbf{x}_{it}$.

CASE 2: Integrated regressors.

Shin (1994) and Harris and Inder (1994) have considered the LBI statistic ξ_i in a (univariate) model with integrated regressors. Their aim was to test for the presence of cointegration. Since they have obtained the asymptotic distribution of ξ_i under H_0, here we can use their result directly and simply apply the central limit theorem to get the distribution of $\overline{\xi}$.[5]

For each $i = 1, ..., N$ consider the following marginal process for the k-dimensional vector of regressors

$$\mathbf{x}_{it} = \mathbf{x}_{i,t-1} + \zeta_{it},$$

with $\mathbf{x}_{i0}$ fixed. Let $\mathbf{v}_{it} = (\varepsilon_{it}, \zeta_{it}')'$ and assume that, for each $i = 1, ..., N$, $\mathbf{v}_{it}$ satisfies a Functional Central Limit Theorem[6] ,

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Tr]}\mathbf{v}_{it} \Rightarrow \mathbf{\Omega}_i^{\frac{1}{2}}\mathbf{W}_i(r), \quad r \in [0,1],$$

where $\mathbf{W}_i(r)$ is a $(k+1)$-dimensional standard Wiener process and $\mathbf{\Omega}_i$ is the long run variance of $\mathbf{v}_{it}$. Further, as in the previous sections, we assume $\mathbf{v}_{it}$ independent across i.

Partition $\mathbf{\Omega}_i$ and $\mathbf{W}_i(r)$ as

$$\mathbf{\Omega}_i = \begin{pmatrix} \omega_{i,11} & \boldsymbol{\omega}_{i,12} \\ \boldsymbol{\omega}_{i,12}' & \mathbf{\Omega}_{i,22} \end{pmatrix},$$

[5] The resulting distribution corresponds to the one obtained by McCoskey and Kao (1998).
[6] A set of assumptions for a Functional Central Limit Theorem is contained in Assumption 4.2 of chapter 4; see also Phillips and Durlauf (1986).

$$\mathbf{W}_i(r) = \begin{pmatrix} W_{i,1}(r) \\ \mathbf{W}_{i,2}(r) \end{pmatrix},$$

where $\omega_{i,11}$ and $W_{i,1}(r)$ are scalars. Note that if ε_{it} is a white noise then $\omega_{i,11}$ is its variance (otherwise it is the long run variance, proportional to the spectrum at frequency zero).

Consider first the case $\boldsymbol{\omega}_{i,12} = 0$, which corresponds to exogeneity of the regressors.

Under $H_0 : q_i = 0$, Harris and Inder (1994) show that

$$\xi_i \xrightarrow{d} V_i \equiv \int_0^1 B_{i,X}(r)^2 dr,$$

where (dropping the subscripts i)

$$B_X(r) = W_1(r) - P^{-1}Qr - \left(\int_0^1 W_2(s)'ds\right)\left(\mathbf{S} - \mathbf{R}P^{-1}Q\right),$$

with

$$\begin{aligned}
P &= 1 - \int_0^1 W_2(r)'dr \left(\int_0^1 W_2(r)W_2(r)'dr\right)^{-1} \int_0^1 W_2(r)dr, \\
Q &= W_1(1) - \int_0^1 W_2(r)'dr \left(\int_0^1 W_2(r)W_2(r)'dr\right)^{-1} \int_0^1 W_2(r)dW_1(r), \\
\mathbf{R} &= \left(\int_0^1 W_2(r)W_2(r)'dr\right)^{-1} \int_0^1 W_2(r)dr, \\
\mathbf{S} &= \left(\int_0^1 W_2(r)W_2(r)'dr\right)^{-1} \int_0^1 W_2(r)dW_1(r).
\end{aligned}$$

Let $a_p = \mathrm{E}(V_i)$, $b_p = \sqrt{Var\,(V_i)}$, where the subscript p indicates dependence on the number of regressors in the model. Then, as $T \to \infty$ followed by $N \to \infty$, we have

$$\sqrt{N}\left(\frac{\bar{\xi} - a_p}{b_p}\right) \xrightarrow{d} N(0,1). \tag{7.48}$$

In this case the quantities a_p, b_p cannot be computed analytically but need to be simulated. This has been done in McCoskey and Kao (1998); their result is reproduced in the table below for $p = 1, 2, 3, 4, 5$.

	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
Mean a_p	0.1162	0.0850	0.0658	0.0533	0.0440
Std. Dev. b_p	0.1044	0.0741	0.0529	0.0400	0.0300

The same distribution also applies for the case when ε_{it} are not white noises and we have used the KPSS correction to $\bar{\xi}$ by replacing $T^{-1}\sum_{t=1}^{T}\widehat{u}_{it}^2$ with consistent estimators of the long run variances of ε_{it}, $i=1,...,N$.

Finally, if the regressors are endogenous, i.e. $\omega_{i,12}\neq 0$, we can correct each ξ_i (and thus our test statistic $\bar{\xi}$) using the Fully Modified Least Squares of Phillips and Hansen (1990). The corrected statistic will have the same asymptotic distribution as (7.48); see Harris and Inder (1994) for details on how to construct the corrected statistic.

7.7.3 Consistency of the test

Under the alternative hypothesis

$$H_1 : q_i > 0 \text{ for } 0 \leq i \leq N_1,\ q_i = 0 \text{ for } N_1 < i \leq N,$$

we know either from chapter 3 or from Shin (1994) that, as $T \to \infty$,

$$T^{-1}\xi_i \xrightarrow{d} M_i, \quad 1 \leq i \leq N_1,$$

for some random variable M_i with $\mathrm{E}(M_i) = \mu > 0$, and, from the previous subsection, that

$$\xi_i \xrightarrow{d} V_i, \qquad N_1 < i \leq N.$$

Thus, for fixed N and as $T \to \infty$,

$$T^{-1}\bar{\xi} = \frac{1}{N}\sum_{i=1}^{N_1} T^{-1}\xi_i + o_p(1).$$

Therefore, as $N \to \infty$ with $\lim_{N\to\infty} N_1/N = \lambda > 0$, using a Law of Large Number we have that

$$T^{-1}\bar{\xi} \xrightarrow{p} \lambda\mu; \tag{7.49}$$

clearly, as under H_1 the statistic diverges, the test is consistent (in the sense of sequential asymptotics).

7.8 Illustration: estimating a production function from a panel of US manufacturing firms

In this section we use model (7.1)-(7.5) to estimate a Cobb-Douglas production function for a balanced panel of 404 R&D manifacturing companies observed for 12 years, 1978-1989. These data are a subset of a dataset constructed by Bronwyn Hall (and available at her homepage at Berkeley University). Similar data have been used by Blundell and Bond (1999) to estimate a production function in the autoregressive framework. Further details on the data constructions are in Mairesse and Hall (1996).

The dependent variables is sales, as a proxy for output, and the regressors are capital stock and employment, measured at the end of the firm's accounting year. The data are in logarithms and a time component common to all individuals has been subtracted from the outset, as suggested in section 7.6.

Let n_{it}, k_{it} be respectively the labour and the capital input for firm i at time t. We have considered two specifications of the production function, corresponding to (7.1)-(7.5) with $\mathbf{x}_{it} = (n_{it}, k_{it})'$ and $\mathbf{x}_{it} = (n_{it}, n_{i,t-1}, k_{it}, k_{i,t-1})'$ respectively. The two models, labelled model 1 and model 2, have been estimated by MLE with heterogeneous variance of the noise (i.e. σ_i^2 unrestricted across firms), by MLE with homogeneous variance of the noise, and by OLS in first differences. The estimation results are reported in the tables below, where within brackets are contained either the t-statistics for the regression coefficients or the p-values for the tests. Note the starred coefficients in the tables, which are averages across individuals.

From the t-statistics of model 2 we immediately see that we cannot exclude lagged values of the regressors, so model 1 is misspecified. We have also tried to use more lags, without statistical significance. So model 2 seems the appropriate one.

	MODEL 1		
	Heterog.	**Homog.**	**First Diff.**
β_{n_t} - labour	0.59 (51.8)	0.56 (40.5)	0.52 (17.3)
β_{k_t} - capital	0.22 (18.7)	0.25 (16.5)	0.25 (9.99)
Signal-noise ratio q	2.5826	0.1974	-
Std. Dev. of noise	0.0357*	0.0507	-
Log-lik. / $N(T-1)$	1.9131	1.6089	-
S.E. prediction	0.1044*	0.1212	0.1226
$r(1)$	-0.0596*	-0.0173*	-0.1175
t-test for CRS	-18.7 (0.00)	-15.4 (0.00)	-9.53 (0.00)
LR test for $\sigma_i^2 = \sigma^2$	66.17 (0.00)	-	-
Hausman test	7.52 (0.023)	-	-

Consider model 2 with heterogeneous variances. The long run elasticities of labour and capital have been estimated to 0.79 and 0.11 respectively, summing up to 0.9. Clearly, the hypothesis of constant return to scale (CRS) is easily rejected. In particular, the unobserved individual effect (random walk) plays an important role in explaining the behaviour of y_{it}; in average, 9% of the change in y_{it} is explained by the evolution of the individual effect (the standard deviation σ_η being estimated as 0.0334*2.6697=0.0902). Note also that the null hypothesis of homogeneous variances of the noise is strongly rejected. The goodness of fit measure is the average (across individuals) standard error of prediction, which is around 10%. Note also that, although not reported in the table, the test for fixed effects of section 7.5 strongly rejects, as expected from having estimated a high signal to noise ratio q.

Unfortunately, there are some problems with this model. One problem can be the short sample, which is likely to bias the estimators in a model with heterogeneous variances. In fact there are some discrepancies with the estimates

of the OLS in first differences (which in turn are very similar to the model with homogeneous variances, but this should be the case as the signal-to-noise ratio is high). The other problem is that it fails the Hausman test, i.e. the regressors appear to be correlated with the innovations in the random walk (which would make the estimators inconsistent). Failing the Hausman test has, however, also an economic interpretation; namely that the correlation between the labour input and the "productivity shocks" could signal the existence of learning by doing in the industry.

	MODEL 2		
	Heterog.	**Homog.**	**First Diff.**
β_{n_t} - labour	0.57 (48.6)	0.51 (36.1)	0.51 (18.3)
$\beta_{n_{t-1}}$ - lagged labour	0.22 (19.0)	0.24 (17.2)	0.25 (9.82)
β_{k_t} - capital	0.16 (12.8)	0.21 (12.3)	0.20 (6.45)
$\beta_{k_{t-1}}$ - lagged capital	-0.05 (-3.81)	-0.08 (-5.08)	-0.08 (3.39)
Signal-noise ratio q	2.6997	2.0124	-
Std. Dev. of noise	0.0334*	0.0485	-
Log-lik. / $N(T-2)$	1.9464	1.6385	-
S.E. prediction	0.1011*	0.1177	0.1191
$r(1)$	-0.0705*	-0.0347*	-0.1371
t-test for CRS	-7.63 (0.00)	-7.53 (0.00)	-4.81 (0.00)
LR test for $\sigma_i^2 = \sigma^2$	59.95 (0.00)	-	-
Hausman test	23.3 (0.00)	-	-

REFERENCES

Ahn, S.C. and P: Schmidt (1995), Efficient estimation of models for dynamic panel data, Journal of Econometrics 68, 5-28.

Ahn, S.C. and P: Schmidt (1997), Efficient estimation of dynamic panel data models: alternative assumptions and simplified estimation, Journal of Econometrics 76, 309-321.

Anderson, T.W. and D.A. Darling (1952), Asymptotic theory of certain 'Goodness of Fit' criteria based on stochastic processes, Annals of Mathematical Statistics 23, 193-212.

Andrews, D.J. (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica, 59, 817-858.

Andrews, D.W.K. and J.C. Monahan (1992). An improved heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 60, 953-66.

Ansley, C.F. and R. Kohn (1985), Estimation, filtering and smoothing in state space models with incompletely specified initial conditions, Annals of Statistics 13, 1286-1316.

Arellano, M. and S. Bond (1991), Some tests of specification for panel data; Monte Carlo evidence and an application to employment equation, Review of Economic Studies 58, 277-297.

Bai, J. (1994), Least squares estimation of a shift in linear processes, Journal of Time Series Analysis 15, 453-472.

Bai, J. (1997), Estimation of a change point in multiple regression models, Review of Economics and Statistics, 551-563.

Balcombe, K. (1999), Seasonal unit root tests with structural breaks in deterministic seasonality, Oxford Bulletin of Economics and Statistics 61, 569-595.

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

Blundell, R. and S. Bond (1998), Initial conditions and moment restrictions in dynamic panel data models, Journal of econometrics 87, 115-143.

Blundell, R. and S. Bond (1999), GMM estimation with persistent panel data: an application to production function estimation, IFS Working Paper W99/4.

Blundell, R., Bond, S. and F. Windmeijer (1999), Estimation of dynamic panel data models: improving on the performance of the standard GMM estimators, IFS Working Paper W00/12.

Box, G.E.P. and G.M. Jenkins (1976), Time series analysis, forecasting and control, San Francisco, Holden-Day.

Caner, M. (1998), A locally optimal seasonal unit root test. Journal of Business and Economic Statistics, 16, 349-56.

Canova, F. and B.E. Hansen (1995), Are seasonal patterns constant over time? A test for seasonal stability, Journal of Business and Economic Statistics, 2, 292-349.

Chan, N.H. and C.Z. Wei (1988), Limiting distributions of least squares estimates of unstable autoregressive processes, Annals of Statistics 16, 367-401.

Choi, I. and B.C. Ahn (1995), Testing for cointegration in systems of equations, Econometric Theory 11, 952-983.

Cleveland, W.S. and S.J Devlin (1980), Calendar effects in monthly time series: detection by spectrum and graphical methods. Journal of the American Statistical Association, 75, 487-96.

Davidson (1999), When is a time series I(0)? Evaluating the memory properties of nonlinear dynamic models, mimeo.

de Jong, P. (1988), The likelihood for a state space model, Biometrika 75, 165-169.

de Jong, P. (1991), The diffuse Kalman filter, Annals of Statistics 19, 1073-1082.

de Jong, P. and J. Penzer (1998), Diagnosing shocks in time series. Journal of the American Statistical Association, 93 :796-806.

Dickey, D.A. and W.A. Fuller (1979), Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427-431.

Dickey, D.A. and W.A. Fuller (1981), Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica, 1057-1072.

Doornik, J.A. (1998), Object-oriented matrix programming using Ox 2.0, Timberlake Consultants Press, New York.

Engle, R.F. and C.W.J. Granger (1987), Co-integration and error correction: representation, estimation and testing, Econometrica 55, 251-276.

Engle, R.F. and B.S. Yoo (1991), Cointegrated economic time series: an overview with new results, in Engle, R.F. and C.W.J. Granger (eds.), Long Run Economic Relationships: Readings in Cointegration. Advanced Text in Econometrics, Oxford University Press, Oxford, 237-266.

Ferguson, T.S. (1969), Mathematical Statistics: a decision theoretic approach, Academic Press, New York.

Franses, P.H. and T.J. Vogelsang (1998), On seasonal cycles, unit roots, and mean shifts, The Review of Economics and Statistics 80, 231-240.

Gardner, L.A. (1969), On detecting changes in the mean of normal variates, Annals of Mathematical Statistics 40, 116-126.

Gikhman, I.I. and A.V. Skorokhod (1975), The theory of stochastic processes (translated from the Russian by S. Kotz). Berlin: Springer-Verlag.

Giri, N.C. (1996), Multivariate statistical analysis, Dekker, New York.

Gregory, A.W. and B.E. Hansen (1996a), Residual based tests for cointegration in models with regime shifts, Journal of Econometrics 70, 99-126.

Gregory, A.W. and B.E. Hansen (1996b), Tests for cointegration in models with regime and trend shifts, Oxford Bulletin of Economics and Statistics 58, 555-560.

Hao, K. (1996), Testing for structural change in cointegrated regression models: some comparisons and generalizations, Econometric Reviews 15, 401-429.

Hadri, J. (1998), Testing for stationarity in heterogeneous panel data, mimeo.

Harris, D. and B. Inder (1994), A test of the null hypothesis of cointegration, in Nonstationary Time Series Analysis and Cointegration, Hargreaves, Colin P., ed., Oxford University Press, New York.

Harvey, A.C. (1989), Forecasting, structural time series models and the Kalman filter, Cambridge, Cambridge University Press.

Harvey, A.C. and J. Durbin (1986), The effect of seat belt legislation on British road casualties: a case study in structural time series modelling, Journal of the Royal Statistical Society, Series A 149, 187-227.

Harvey, A.C. and A. Jaeger (1993), Detrending, stylized facts and the business cycle, Journal of Applied Econometrics 8, 231-247.

Harvey, A.C. and J. Bernstein (2000), Measurement and testing of inequality from time series of deciles with an application to US wages. Mimeo.

Harvey, A.C. and S.J. Koopman (1992), Diagnostic checking of unobserved components time series models, Journal of Business and Economics Statistics 10, 377-389.

Harvey, A.C. and S.J. Koopman (1996), Structural time series models in medicine, Statistical Methods in Medical Research, 23-49.

Harvey, A.C. and S.J. Koopman (1997), Multivariate structural time series models, in System Dynamics in Economic and Financial Models, C. Heij, J.M. Schumacher, B. Hanzon and C. Praagman editors, New York, John Wiley and Sons.

Harvey, A.C., Koopman, S.J. and J. Penzer (1998), Messy time series, In T.B. Folby and R.C. Hill (eds.), Advances in Econometrics, Volume 13, JAI Press, New York.

Harvey, A.C., Ruiz, E. and N. Shepard (1994), Multivariate stochastic variance models, Review of Economic Studies 61, 247-264.

Harvey, A.C. and M. Streibel (1997), Testing for nonstationary unobserved components, mimeo.

Hobijn, B. and P.H.Franses (2000), Asymptotically perfect and relative convergence of productivity, Journal of Applied Econometrics, 15, 59-81.

Hsiao, C. (1986), Analysis of panel data, Cambridge University Press, Cambridge.

Hylleberg, S., Engle, R.F., Granger, C.W.J. and B.S. Yoo (1990), Seasonal integration and cointegration, Journal of Econometrics 99, 215-238.

Im, K., Pesaran, H. and Y. Shin (1997), Testing for unit roots in heterogeneous panels, mimeo.

Inoue, A. (1999), Test of cointegrating rank with a trend-break, Journal of Econometrics 90, 215-237.

Johansen, S. (1988), Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12, 131-154.

Johansen, S. (1991), Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive model, Econometrica 59, 1551-1580.

Johansen, S. (1995), Likelihood-based inference in cointegrated vector autoregressive models, Oxford University Press, Oxford.

Kao, C. (1999), Spurious regression and residual-based tests for cointegration in panel data, Journal of Econometrics 90, 1-44.

Kariya, T. (1980), Locally robust tests for serial correlation in least squares regression, Annals of Statistics 6, 1065-1070.

King, M.L. (1980), Robust tests for spherical symmetry and their application to least squares regressions, Annals of Statistics 8, 1265-1271.

King, M.L. and G.H. Hillier (1985), Locally best invariant tests of the error covariance matrix in the linear regression model, Journal of the Royal Statistical Society, Series B 47, 98-102.

Kitagawa, G., and W Gersch. *Smoothness priors analysis of time series.* Berlin: Springer-Verlag (1996).

Koopman, S.J. (1997), Exact initial Kalman filtering and smoothing for nonstationary time series models, Journal of the American Statistical Association 92, 1630-1638.

Koopman, S.J. and A.C. Harvey (1999), Computing observation weights for signal extraction and filtering, mimeo.

Koopman, S.J., Shepard, N. and J.A. Doornik (1998), Statistical algorithms for models in state space using SsfPack 2.2, Econometrics Journal 1, 1-55.

Koopman, S.J., Harvey, A.C., Doornik J.A. and N. Shephard (1995), STAMP 5.0 Structural time series analyser, modeller and predictor. London: Chapman and Hall.

Kunimoto, N. (1996), Structural changes, unit roots, cointegration hypotheses with an application to macro time series. Kinyu Kenkyu, Institute of Monetary and Economic Studies, Bank of Japan 15 (5), 1-43.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Y. Shin (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?, Journal of Econometrics 44, 159-178.

Lancaster, T. (2000), The incidental parameter problem since 1948, Journal of Econometrics 95, 391-413.

Lee, J. (1996). On the power of stationarity tests using optimal bandwidth estimates. Economics Letters, **51,** 131-7.

Lee, D. and P. Schmidt (1996), On the power of the KPSS test against fractionally integrated alternatives. Journal of Econometrics 73, 285-302.

Lee, J., Huang, C.J. and Y. Shin (1997), On stationary tests in the presence of structural breaks, Economics Letters 55, 165-172.

Lehmann, E.L. (1959), Testing statistical hypotheses, New York, Wiley.

Leybourne, S.J. and B.P.M. McCabe (1994), A consistent test for a unit root, Journal of Business and Economic Statistics 12, 157-166.

Leybourne, S.J., Mills, T.C. and P. Newbold (1999), Spurious rejections by Dickey-Fuller tests in the presence of a break under the null, Journal of Econometrics 87, 191-203.

Lopes, A.S. and A. Montanes (1998), The asymptotic and finite sample behaviour of seasonal unit roots tests under the presence of a break in the deterministic seasonal pattern, mimeo.

MacNeill, I. (1978), Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times, Annals of Statistics 6, 422-433.

Mairesse, J. and B.H. Hall (1996), Estimating the productivity of research and development in French and US manufacturing firms: an exploration of simultaneity issues with GMM methods, in Wagner, K. and B. Van Ark (eds.), International productivity differences and their explanations, Elsevier Science, 285-315.

McCoskey, S. and C. Kao (1998), A residual based test of the null of cointegration in panel data, Econometric Review 17, 57-84.

Marshall, P. (1992), Estimating time-dependent means in dynamic models for cross-section of time series, Empirical Economics 17, 25-33.

Nabeya, S. and K. Tanaka (1988), Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative, Annals of Statistics 16, 218-235.

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

Neyman, J. and E.L. Scott (1948), Consistent estimates based on partially consistent observations, Econometrica 16, 1-32.

Nickell, S. (1981), Biases in dynamic models with fixed effects, Econometrica 49, 1417-1426.

Nyblom, J. (1986), Testing for deterministic linear trend in time series, Journal of the American Statistical Association 81, 545-549.

Nyblom, J. (1989), Testing for the constancy of parameters over time, Journal of the American Statistical Association 84, 223-230.

Nyblom, J. and A.C. Harvey (1997), Tests against smooth stochastic trends, mimeo.

Nyblom, J. and A.C. Harvey (2000), Tests of common trends, Econometric Theory 16, forthcoming.

Nyblom, J. and T. Mäkeläinen (1983), Comparison of tests for the presence of random walk coefficients in a simple linear model, Journal of the American Statistical Association 78, 856-864.

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

Perron, P. and T.J. Vogelsang (1992a), Nonstationarity and level shifts with an application to Purchasing Power Parity, Journal of Business and Economic Statistics 10, 301-320.

Perron, P. and T.J. Vogelsang (1992b), Testing for a unit root in a time series with a changing mean: corrections and extensions, Journal of Business and Economic Statistics 10, 467-470.

Pesaran, M.H. and Y. Shin (1995), Long run structural modeling, University of Cambridge Department of Applied Economics Working Paper.

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

Phillips, P.C.B. (1991), Optimal inference in cointegrated systems, Econometrica 59, 283-306.

Phillips, P.C.B. (1994), Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models, Econometrica 62, 73-93.

Phillips, P.C.B and S.N. Durlauf (1986), Multiple time series regressions with integrated processes, Review of Economic Studies 53, 473-95.

Phillips, P.C.B. and B. Hansen (1990), Statistical inference in instrumental variables regression with I(1) processes, Review of Economic Studies 53, 473-495.

Phillips, P.C.B. and H.R. Moon (2000), Linear regression limit theory for nonstationary panel data, Econometrica, forthcoming.

Phillips, P.C.B. and S. Ouliaris (1990), Asymptotic properties of residual based tests for cointegration, Econometrica 58, 165-194.

Phillips, P.C.B. and P. Perron (1988), Testing for a unit root in time series regression, Biometrika 75, 335-346.

Phillips, P.C.B and V. Solo (1992), Asymptotics for linear processes, Annals of Statistics 20, 971-1001.

Phillips, P.C.B. and Z. Xiao (1998), A primer on unit root testing, mimeo.

Proietti, T. (1998), Seasonal heteroscedasticity and trends, Journal of Forecasting, 17, 1-17.

Proietti, T. (2000). Comparing seasonal components for structural time series models, International Journal of Forecasting, 16, 247-60.

Rao, C.R. (1973), Linear statistical inference and its applications, 2nd ed., Wiley, New York.

Said, S. and D.A. Dickey (1984), Testing for unit roots in autoregressive moving average models of unknown order, Biometrika 71, 599-608.

Shephard, N.G. (1993), Maximum likelihood estimator of regression models with stochastic trend components, Journal of the American Statistical Association 88, 590-595.

Shephard, N.G. and A.C. Harvey (1990), On the probability of estimating a deterministic component in the local level model, Journal of Time Series Analisys 11, 339-347.

Shin, Y. (1994), A residual based test of the null of cointegration against the alternative of no cointegration, Econometric Theory 10, 91-115.

Smith, J. and J. Otero (1997), Structural breaks and seasonal integration, Economics Letters 56, 13-19.

Smith, R.J. and A.M.R. Taylor (1998), Additional critical values and asymptotic representations for seasonal unit root tests, Journal of Econometrics 85, 269-288.

Soukup, R.J. and D.F. Findlay (2000), On the spectrum diagnostics used by X-12- ARIMA to indicate the presence of trading day effects after modeling or adjustment, @ 1999 Proceedings of the Business and Economic Statistics Section, American Statistical Association, 144-149.

Taylor, A.M.R.(2000), Locally optimal tests against seasonal unit roots. University of Birmingham discussion paper, 99-12

White (1980), An heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48, 817-838.

Young, P., Recursive estimation and time series analysis. Berlin: Springer-verlag (1984).

Zivot, E. and 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.