UNIVERSITY OF SOUTHAMPTON

Bootstrap Inference in Cointegrated VAR Models

by

Alessandra Canepa

A dissertation submitted in the fulfillment of the requirements for the Degree of Doctor of philosophy

DEPARTMENT OF ECONOMICS

September 2002

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCE
ECONOMICS

Doctor of Philosophy

BOOTSTRAP INFERENCE IN COINTEGRATED VAR MODELS

by Alessandra Canepa

The main problem discussed in this work may be described as the lack of coherence between the test statistics and their reference distribution. In small sample the approximations of the first order asymptotic theory are often quite inaccurate. As a result the empirical and nominal probabilities that a test rejects a correct hypothesis can be very different when critical values based on first-order approximation are used. This may lead one to reject too many null hypotheses when they are actually true. In principle there are two ways of solving this problem; either for a given reference distribution to correct the test statistic or for a given test statistic to correct the reference distribution. In Chapter 2 of this thesis we consider Johansen's likelihood ratio and Wald tests for linear restrictions on cointegrating space and we compare analytical corrections to the test statistics such as the ones suggested by Podivinsky and Psaradakis with a numerical approximation of the distribution function obtained using computer intensive methods such as the bootstrap. In Chapter 3 we approximate the finite sample expectation of the likelihood ratio test using the bootstrap and we compare the finite sample properties of the asymptotic, the bootstrap, and the bootstrap Bartlett corrected likelihood ratio tests. Furthermore, we propose bootstrapping the Bartlett corrected likelihood ratio test, using the Bartlett correction proposed by Johansen (1999). In Chapter 4 we provide an empirical application to illustrate the usefulness of the bootstrap test using real data in place of the simulated ones.

Acknowledgments

While working on this thesis I benefit from the advise, constructive criticism, and encouragement of a number of people. In particular, I would like to thank my supervisor Ray O'Brien, for insightful and constructive advise. His guidance made the time I spent at Southampton University a very enjoyable part of my life. I am also grateful to Mathias Hoffmann not only for supervising me while I was working on Chapter 4, but also for being a good friend.

I am grateful to the members of the committee, for their willingness to read this manuscript. I would like to express my appreciation to the participants of various conferences and workshops I attended while I was working on my thesis. In particular, I would like to thank: F. Bravo, B. Nielsen, J. Podivinsky, and the other participants to the 8^{th} World Congress of the Econometric Society for useful suggestions on a preliminary version of Chapter 2 of this thesis; the participants to the Harvard University Workshop on Monte Carlo Simulations for useful comments on Chapter 3; C. Flinn, R. Marchionatti, G. Mizon, and the other participants to the seminars I gave at Southampton and Turin University for useful suggestions on Chapter 4.

I would like to thank the ESRC for sponsoring this thesis, and P. Stoneman for creating the opportunity to take special leave while I was working at Warwick University.

Furthermore, I wish to express my sincere gratitude to a number of friends, listed below, who helped and encourage me along the way: Grazia Rapisarda who has always been a first class friend, Surjinder Johal and Zheng Wang who have especially contributed to my fitness, and Gianluca Grimalda for being an excellent concert mate (without his constructive criticism attending concerts was not the same fun). Finally, I am deeply grateful to Enrico Gambino for his support and his love.

Coventry, 7 September 2002

Alessandra Canepa

Contents

Preface

Empirical econometric models are usually accompanied by diagnostics which serve to support the models' statistical adequacy. Both the diagnostic checking and the ultimate inference are based on test statistics. The basic quantities needed in hypothesis testing are: (i) the critical value that provides the desired significance level of the test, (ii) the power properties of the test, (iii) some knowledge of the sample size required to achieve a given power. In addition, any inferential procedure has limited robustness when the assumptions on which it is based are violated, so one wishes to know to what extent the test is robust against departures from the assumptions under which it is derived.

To calculate (i)-(iii) simple approximate formulas are usually obtained using central limit theory. However, in order to work well the first order asymptotic approximation requires that the asymptotic distribution is an accurate approximation to the finite sample distribution. Unfortunately, particularly for time series models, this is not generally the case.

In the past various correction methods were proposed to improve upon the first-order asymptotic approximation. One strand of the literature addresses this problem by proposing corrections to the test statistic in use. A case in point is the Bartlett correction. The idea behind the Bartlett correction is to adjust the test statistic by its expectation. By doing this we improve the fit of the asymptotic distribution. Another strand of the literature focuses on replacing the critical values of the limit distribution with values that will generate an actual test size closer to the nominal one. Here, one of the techniques suggested

in the literature is the bias-correction method. Loosely speaking this method involves approximating the moments of the distributions using asymptotic expansions. For example, White (1961) obtained higher-order moment expansions for the first two moments of the $AR(1)$ parameter estimator. The higher-order moment expansions can be used to correct the possible bias, which lead to better finite sample inference by adjusting the finite sample distribution. However, bias-correction methods only adjust the centre of the distribution, so that if the finite sample distribution exhibits substantial skewness this method will not result in accurate inference. The Bartlett correction can be regarded as adjusting the first moment.

In contrast to the bias correction method, Edgeworth corrections are based on a series expansion of the whole distribution function. However, the Edgeworth expansion requires the estimation of higher-order moments of the underlying population distribution. These moments are often difficult to estimate accurately so that the Edgeworth expansion is less frequently used by practitioners.

The bootstrap is a computer intensive technique that involves resampling one's data or a model estimated from the data. It can be shown (see for example Hall (1992)) that in many cases the bootstrap delivers an automatic approximation to the Edgeworth expansion, so that it can be considered as a numerical approximation to analytical calculations of one-term Edgeworth expansions. The advantage is, of course, that it does not involve the same tedious calculations.

In this thesis we investigate the small sample behaviour and the robustness of the bootstrap inference procedure in cointegrated vector autoregressive models (VAR). In-

ference in cointegrated systems has been a major topic of debate since Johansen (1988) showed that in principle standard asymptotic methods can be applied to certain classes of estimators of the coefficients. However, several simulation studies suggest that the performance of asymptotic tests on the cointegration coefficients may be rather poor. The bootstrap may deal with finite sample inaccuracies in two ways. Firstly, the bootstrap takes the effects of the small sample into account by replacing the nuisance parameters by consistent estimators in the finite sample distribution, whereas the asymptotic approximation replaces the nuisance parameters by consistent estimators in the asymptotic distribution. Secondly, by using the empirical distribution function in place of some specific parametric distribution the non-parametric bootstrap is able, for example, to mimic possible skewness of the finite sample distribution (e.g. to take the non-normality of the finite sample distribution into account). Throughout the thesis, bootstrap accuracy in small samples is mainly investigated through classical Monte Carlo studies, although an empirical application is provided to illustrate the performance of the bootstrap with some real data.

The thesis is organized as follows. Chapter 1 contains a brief survey of bootstrap inference procedures in econometric models. The survey is far from being comprehensive, but it provides the theoretical background for the subsequent chapters. Chapters 2-4 contain the main body of the research. Finally, Chapter 5 summarizes the main results of this thesis.

Chapter 1 starts with an introduction of the bootstrap principles. After discussing the first order asymptotic validity of the bootstrap, the theory underlying the bootstrap's ability to provide asymptotic refinements is considered in some detail. Then bootstrap hypothesis testing is discussed. Finally, we consider applications of the bootstrap procedures in

the context of the regression model, and the distinction between the residual based bootstrap and the block bootstrap is introduced. The chapter ends with an appendix where the Glivenko-Cantelli Theorem is stated.

The purpose of Chapter 2 is twofold. Firstly, we use bootstrap hypothesis testing as a way to reduce the size distortion of the tests for linear restrictions on the cointegrating space, and we compare analytical corrections to the test statistics such as the ones suggested by Podivinsky (1992) and Psaradakis (1994) with a numerical approximation of the distribution function obtained using the bootstrap. Secondly, we consider the Johansen likelihood ratio and Wald test statistics as well as the small sample corrected version of these tests, and we explore the robustness of the inference procedure in a situation where we allow for potential over-fitting and under-fitting of the number of cointegrating vectors included in the restricted model.

Chapter 3 is closely related to Chapter 2. Again we consider Johansen's likelihood ratio tests for linear restriction on cointegrating space and we propose that the Bartlett adjustment factor be computed using the bootstrap. Further, we consider bootstrapping Johansen's Bartlett corrected likelihood ratio test. Since the Bartlett correction can be regarded as an analytical approximation to the bootstrap test (see for example Beran (1988)), bootstrapping the Bartlett corrected LR test amounts to a sort of double bootstrap procedure which may lead to higher order asymptotic refinements. However, the performance of the Johansen Bartlett correction crucially depends on the parameters of the model so that the potential of the bootstrap test to provide second order asymptotic refinements is obvi-

ously affected. For this reason we undertake a response surface analysis. Finally, the power properties of the bootstrap tests are evaluated.

Chapter 4 is a self contained chapter where an empirical application of the bootstrap test is undertaken using real data instead of the simulated ones. The idea is to analyse the effects of macroeconomic shocks on unemployment, and in particular the effects of shifts in labour supply and labour demand on the rise of European unemployment. The econometric model considered is a structural VAR with cointegrated constraints. This model has its roots in the Beveridge-Nelson (1981) decomposition of univariete time series since it involves a linear decomposition of a VAR into stationary and non-stationary parts. In particular, rewriting the VAR in the $VECM$ form and inverting it we find a MA representation which is the sum of the initial values, an $I(1)$ component and an $I(0)$ component. This MA representation is a natural starting point for the impulse response analysis.

Finally, the thesis concludes with a summary of the main results in Chapter 5. In addition some directions for possible research are given.

Chapter 1
An Introduction to the Bootstrap

1.1 Introduction

The purpose of this chapter is to illustrate the usefulness and the limitations of the bootstrap by providing a brief overview of the literature.

After the seminal paper by Efron (1979) a lot of work has been done, and the literature on this subject is now enormous. For this reason after having introduced the general conditions under which the bootstrap provide a consistent estimator of the statistic under study we will focus on the branch of the bootstrap literature which is more closely related to the areas of application in time series analysis.

The general idea on which bootstrapping is based is to use the single data set to design a sort of Monte Carlo experiment in which the data themselves are used to generate an approximation to the distribution of the statistics in which we are interested. However, as Veall (1998) suggests there are two main stages in the development of bootstrap theory. The first stage is related to its introduction by Efron (1979) as a computer-based method for evaluating the accuracy of a statistic by using the bootstrap algorithm for estimating standard errors or confidence intervals. This procedure can be useful when the finite-sample distribution of the statistics we are analysing is not known or a good asymptotic approximation is not available. The second stage of the bootstrap literature concerns the case where asymptotic analytic tools are available but in which bootstrap refinements are used to im-

prove finite-sample performance. Good references in this sense are Horowitz (1994) where the bootstrap method is applied to the information matrix tests. For an excellent discussion based on the Edgeworth expansion see Hall (1992). These and other studies have found that bootstrap provides a higher-order asymptotic approximation to critical values based on "smooth" statistics. This means that for bootstrap-based critical values the size distortion (that is the difference between the nominal level and its actual rejection probability) decreases more rapidly with increasing the sample size than if the critical values obtained from first-order asymptotic theory are used.

The outline of this chapter is the following. Section 1.2 introduce the bootstrap principle. In Section 1.3 the first higher-order asymptotic validity of the bootstrap is indicated in the i.i.d. setting. In Section 1.4 we will illustrate the higher order refinements provided by the bootstrap for pivotal statistics. In Section 1.5, we consider application of the bootstrap method to test statistics. Finally in Section 1.6 we will go through some recent developments of the bootstrap method in time series analysis.

1.2 The bootstrap principle

The bootstrap is a method for estimating the distribution of an estimator or test statistic by resampling the data. It amounts to treating the data as if they were the population for the purpose of evaluating the distribution of interest. Under conditions that hold in a variety of applications, the bootstrap provides an approximation to distribution of the statistics under study that is at least as accurate as the approximation of first-order asymptotic distribution theory (Horowitz (1999)).

The basic bootstrap procedure in the simplest setting works as follows. Let F denote some distribution function and suppose we are interested in a real-valued parameter $\theta \in \Theta$ which can be written as a functional of F[1]. Let $X_1, X_2, ..., X_n$ denote a sample of n i.i.d. random variables having common distribution function F. Since θ is unknown we seek to obtain information about θ from the sample $X_1, X_2, ..., X_n$. That is, we are interested in the relationship between the population parameter and the sample. The bootstrap provides a method for estimating the distribution of $\hat{\theta}$ or a feature of F such as a moment or a quantile, by replacing the unknown distribution of F with a known estimator. Let $\hat{F}_n$ denote the estimator of F. Two possible choices of $\hat{F}_n$ are:

(1) The empirical distribution function [2] (EDF) of the data:

$$\hat{F}_n = \frac{1}{n} \sum_{i=1}^{n} I\left(X_i \leq x\right)$$

where I is the indicator function. The case where $\hat{F}_n$ correspond to the empirical distribution was discussed by Efron (1979) and it is usually referred to as non-parametric bootstrap.

(2) A parametric estimator of F. In this case we assume that the random variables X_i have a particular distribution function, for example the normal. In the literature this is defined as the parametric bootstrap.

Practical application of the bootstrap technique requires the generation of bootstrap samples or resamples (i.e. samples obtained by independently sampling with replacement from the empirical distribution). Regardless of the choice of $\hat{F}_n$, usually the bootstrap esti-

---

[1] A functional is simply a mapping that assigns a real value to a function. Most commonly used parameters of distribution functions can be expressed as functional of the distribution, including the mean, the variance, the skewness, and the kurtosis of the distribution.

[2] The EDF is defined to be the cumulative distribution function (CDF) of a random variable which takes value $X_1, X_2, ..., X_n$ each with probability mass $1/n$.

mator of the functional of F cannot be evaluated analytically. However, it can be estimated with arbitrary accuracy by carrying out a Monte Carlo simulation in which the random samples are drawn from $\hat{F}_n$. Thus, the bootstrap is usually implemented by Monte Carlo simulation. The procedure used for evaluating the bootstrap by Monte Carlo is straightforward:

Step 1) Generate a bootstrap sample of size n, by sampling the distribution corresponding to $\hat{F}_n$ randomly.

Step 2) Compute θ^*, the boostrap stastistic.

Step 3) Repeat steps 1 and 2 k times to compute the empirical probability of the event $\theta^* \leq \hat{\theta}$. In this way we obtain the proportion of the repetitions in which the event $\theta^* \leq \hat{\theta}$ occurs.

Under certain conditions described in the next section, the basic bootstrap procedure described above works very well in extremely general settings.

1.3 Consistency of the Bootstrap

Asymptotic validity for the bootstrap requires that as the sample size n increases the bootstrap distribution will be close to the actual distribution of the root[3] under study. In the literature the asymptotic validity of the non-parametric bootstrap has been proved by showing that the Mallows distance between the finite sample distribution and the bootstrap distri-

[3] The relationship between the parameter of interest and the sample is often referred as the root $R_n(\theta, \hat{\theta})$. A root is a function of both the population and the sample value (e.g. $R_n = (\theta - \hat{\theta})$).

bution converges to zero as the number of observations goes to infinity (see for example Bickel and Freedman (1981) or Basawa (1991b)).

Let Γ_p^k the set of distribution functions G in $\mathbb{R}^k$, such that $\int \|X\|^p dG < \infty$, where $\|\cdot\|$ is the Euclidean norm. For $G, H \in \Gamma_p^k$, the Mallows distance[4] of order p between the distributions G and H is defined as

$$\rho_p(G,H) = \inf\left\{E\|X-Y\|^p : X \backsim G, Y \backsim H\right\}^{1/p},$$

where $F(x,y)$ is the joint distribution of (X,Y), $F(x,\infty) = G$ is the marginal distribution of X, and $F(\infty,y) = H$ is the marginal distribution of Y. Weak convergence of a sequence of distributions in the Mallows metric implies convergence of the corresponding sequences of first p-moments. A detailed description of this metric is discussed in Section 8 of Bickel and Freedman (1981).

Before, discussing the conditions under which the bootstrap distribution of a statistic is consistent, it may be useful to consider an example.

Example 1 *The distribution of the sample average.*

Let $X_1, ..., X_n$ be $i.i.d.$ random variables with common distribution F, with $\mu = E(X_i)$, $\sigma^2 = Var(X_i)$. Assume $0 < \mu < \infty$, and $0 < \sigma^2 < \infty$. Define $\hat{F}_n$ as the empirical CDF of X_i, $\mu_n = n^{-1}\sum_{i=1}^{n} X_i$, $s^2 = n^{-1}\sum_{i=1}^{n}(X_i - \mu_n)^2$. Finally, define $X_1^*, ..., X_n^*$ to be

[4] Let E be any set. Let $\rho(x,y)$ be a function defined on the set $E \times E$ of all ordered pairs (x,y) of members of E, and satisfying the following conditions:
(i) $\rho(x,y)$ is a finite real number for every pair (x,y) of $E \times E$;
(ii) $\rho(x,y) = 0$ if and only if $x = y$;
(iii) $\rho(x,z) \leq \rho(x,y) + \rho(x,z)$, where x, y, z are three elements of E.
Such function $\rho(x,y)$ is a metric space on E. That is, is a mapping of $E \times E$ into $\mathbb{R}$. A set E is called a metric space, and the function $\rho(x,y)$ the distance from the point x to y.

the resampled data with common distribution $\hat{F}_n$, $\bar{X}_n^* = n^{-1}\sum_{i=1}^{n} X_i^*$. Let G_n and G_n^* denote the distribution of the roots $\sqrt{n}(\mu_n - \mu)$ and $\sqrt{n}\left(\bar{X}_n^* - \mu_n\right)$ respectively. Define ρ_2 as the Mallows metric of order two. The Glivenko-Cantelli (see the Appendix of this Chapter) theorem and the strong low of large numbers implies that the condition

$$d_2\left(\hat{F}_n, F\right) \to 0$$

is satisfied. Using the properties of d_2 , (see Bickel and Freedman (1981)), it can be shown that

$$\rho_2\left(G_n^*, G_n\right)^2 \leq \rho_2\left(\hat{F}_n, F\right)^2,$$

which shows that in the non-parametric bootstrap the distance between the bootstrap distribution and the finite sample distribution can be bounded between the EDF and the underlying distribution function. ■

In the literature Bickel and Freedman (1981) were the first to show the conditions under which the bootstrap distribution of a statistic is consistent in $i.i.d.$ contexts. They list three conditions for the bootstrap distribution to be consistent. The first is weak convergence of the statistic $X_i \backsim G$ for all distribution G in a neighborhood of the true distribution F. The second is uniform weak convergence over distributions G in a neighborhood of the true distribution F. The third is continuity of the mapping from the underlying distribution G to the asymptotic distribution of the statistic. As an example we consider the consistency of the bootstrap for von Mises functionals.

Let $X_1, ..., X_n$ be a vector of random variables. Consider

$$\frac{\sqrt{n}\left(g\left(n^{-1}S_n\right) - g\left(\mu\right)\right)}{v\left(n^{-1}T_n\right)} \tag{1.3.1}$$

where $g : \mathbb{R}^k \to \mathbb{R}$, $v : \mathbb{R}^l \to \mathbb{R}$

$$\begin{aligned} S_n &= \sum_{i=1}^{n} h(X_i) \\ T_n &= \sum_{i=1}^{n} r(X_i) \end{aligned}$$

where $h : \mathbb{R}^p \to \mathbb{R}^k$, $r : \mathbb{R}^p \to \mathbb{R}^l$, and

$$\begin{aligned} \mu &= Eh(X_1), \\ v &= Er(X_1). \end{aligned}$$

Applying the mean value theorem the numerator of $(1.3.1)$ is

$$\sqrt{n}\left(g\left(n^{-1}S_n\right) - g(\mu)\right) = \overset{\circ}{g}(\mu)\sqrt{n}\left(n^{-1}S_n - \mu\right) + o_p(1)$$

where $\overset{\circ}{g} = \Delta_\mu g$ is the Jacobian. Consider now the case where g is a functional

$$g : \mathcal{F} \to \mathbb{R},$$

where $\mathcal{F}$ is a convex set of probability measures on $\mathbb{R}^m$ including all point masses and F. Define

$$\begin{aligned} g(F) &= \int x dF(x), \\ g(F_n) &= \int x dF_n(x), \\ g(F_n^*) &= \int x dF_n^*(x). \end{aligned}$$

Let us first define a von Mises functionals. Suppose that g is Gateaux differentiable[5] at F with derivative $\overset{\circ}{g}(F)$ representable as an integral

$$g(G) = g(F) + \overset{\circ}{g}(F)(G - F) + o(1), \tag{1.3.2}$$

and

$$\overset{\circ}{g}(F)(G - F) = \frac{\partial}{\partial \epsilon} g(F + \epsilon(G - F))|_{\epsilon=0} = \int \psi(x, F)\, dG(x),$$

where necessarily

$$\int \psi(x, F)\, dF(x) = 0,$$

and substituting the integral representation of the Gateaux differential into (1.3.2)

$$g(G) = g(F) + \int \psi(x, F)\, dG(x) + o(|G - F|).$$

If g satisfy this properties we call it a von Mises functional.

Asymptotic normality results imply that $\sqrt{n}(g(F_n) - g(F))$ and

$$\sqrt{n} \int \psi(x, F)\, d(F - F_n)$$

are distributed as $N\left(0, \int \psi^2(x, F)\, dF\right)$ since

$$g(F_n) - g(F) = \overset{\circ}{g}_F(F_n - F) + \Delta_n(F_n, F),$$

---

[5] Let Y be a linear space, and let $J : D \subseteq Y \to R$ be a functional with domain D in Y. If y and v are elements of Y such that for all ε in a neighborhood of 0, $y + \varepsilon v$ is an element of D, then

$$\delta J(y; v) \overset{def}{=} \lim_{\varepsilon \to 0} \frac{J(y + \varepsilon v) - J(y)}{\varepsilon},$$

is called the Gateaux Differential (or Gateaux Variation) of J at y in the direction of v. Note, that

$$\delta J(y; v) = \frac{\partial}{\partial \varepsilon} J(y + \varepsilon v)|_{\varepsilon=0},$$

if this derivative w.r.t. $\varepsilon = 0$.

where

$$\Delta_n(F_n, F) = o_p(g_F(F_n - F))$$

Let G_n the empirical distribution of $X_1^*, ..., X_n^*$. Consider

$$\sqrt{n}(g(G_n) - g(F_n)) = \sqrt{n}\,\overset{\circ}{g}_{F_n}(G_n - F_n) + \sqrt{n}\Delta_n(G_n, F_n),$$

the bootstrap works if

$$\sqrt{n}\Delta_n(G_n, F_n) = o_p(1)$$

so that the conditional distribution of

$$\sqrt{n}\,\overset{\circ}{g}_{F_n}(G_n - F_n) = \sqrt{n}\int \psi(x, F_n)\, dG_n = \sqrt{n}\left(n^{-1}\sum_{i=1}^{n}\psi(X_i^*, F_n)\right). \qquad (1.3.3)$$

Indeed, it can be shown that

$$\sqrt{n}\left(n^{-1}\sum_{i=1}^{n}\psi(X_i^*, F_n)\right) \xrightarrow{d} N\left(0, \int \psi^2(x, F)\, dF(x)\right)$$

as it turns out that this is also the limit distribution of $\sqrt{n}(g(F_n) - g(F))$.

Sufficient conditions to ensure the bootstrap works are

1) $0 < \int \psi^2(x, F)\, dF(x) < \infty$

2) $\int (\psi(x, F_n) - \psi(x, F))^2 \to 0$ a.s.

1.4 Asymptotic refinements

In this section we explain why the bootstrap provides an improved approximation to the finite sample distribution of an asymptotically pivotal statistic. Recall that a statistic is pivotal if its limiting distribution does not depend on unknown quantities (see example 2

in this section). The advantage of bootstrapping pivotal roots can be explained by means of the Edgeworth expansion. In this section we firstly introduce the Edgeworth expansion, and secondly the bootstrap will be expressed and interpreted in terms of the Edgeworth expansion.

1.4.1 The bootstrap and the Edgeworth expansion

Let $X_i\,(i=1,...,n)$ be i.i.d. with common distribution function F having mean μ and variance σ^2. Let μ the parameter of interest. By the central limit theorem (CLT) we know that

$$K_n(x) = P\left[\sqrt{n}\left(\bar{X}-\mu\right)/\sigma \leq x\right] \rightarrow \Phi(x).$$

The CLT gives us useful information about the distribution of sums of random variables when little is known about the individual terms. However, it does not provide any insight on the rate of the error made by the first-order normal approximation.

If X_i has sufficient moments and F is non-lattice[6], the Edgeworth expansion of the distribution function of the root $\sqrt{n}\left(\bar{X}-\mu\right)/\sigma$ is given by

$$\begin{aligned} P\left[\sqrt{n}\left(\bar{X}-\mu\right)/\sigma \leq x\right] &= \Phi + n^{-1/2}p_1(x,F)\,\phi(x) + ... \\ &\quad + n^{-j/2}p_j(x,F)\,\phi(x) + o\left(n^{-j/2}\right), \end{aligned}$$

---

[6] A lattice distribution is one in which X_i takes values on an evenly spaced grid of points. More formally, a random variable X_i has a lattice distribution if there are constants c and $h(>0)$ such that the lattice $[c+kh : k=0,\pm1,...]$ supports the distribution of X_i.

Suppose $X_i \backsim F$ have characteristic function $\chi(t) = E[\exp(itX)]$, where $i=\sqrt{-1}$. Then a necessary and sufficient condition for X_i to have a non lattice distribution is that $\lim \sup_{|t|\rightarrow\infty} |\chi(t)| < 1$. This is called the Cramer condition.

where $\phi(x) = (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x^2\right)$ is the standard normal density function, and

$$\Phi(x) = \int_{-\infty}^{x} \phi(u)\, du,$$

is the standard normal distribution function. The function p_j is a polynomial of degree no more than $3j-1$ and is odd for even j, even for odd j. In particular, we have in the above case (of normalised mean)

$$p_1(x, F) = -\frac{1}{6} k_3 \left(x^2 - 1\right),$$

where $k_3 = E\left[(X-\mu)^3\right]/\sigma^3$. The population value k_3 is referred to as the skewness. Since $\Phi(x)$ is the standard normal distribution function, $n^{-1/2} p_1(x, F)\, \phi(x)$ is considered to be the first term of the Edgeworth expansion, and it corrects the basic Normal distribution from the main effect of skewness. The distribution of an asymptotically pivotal root $R_n = \sqrt{n}\left(\hat{\theta} - \theta\right)/\sigma^2$ admits the following expansion

$$\begin{aligned} P\left[\sqrt{n}\left(\bar{X} - \mu\right)/\hat{\sigma} \leq x\right] &= \Phi + n^{-1/2} q_1(x, F)\, \phi(x) + ... \\ &\quad + n^{-j/2} q_j(x, F)\, \phi(x) + o\left(n^{-j/2}\right), \end{aligned}$$

where the polynomial q_j is of degree of no more than $3j-1$ and is odd for even j, even for odd j. Under appropriate conditions the expansion can be developed to any desired order in principle. However, we will consider only first-order expansion.

Consider now the Edgeworth expansion for more general statistics. Consider the smooth function model as analysed by Hall (1992). Let X_i $(i = 1, ..., n)$ be i.i.d. random k-vectors variable with distribution function F with mean μ and finite second moments. Let $g : \mathbb{R}^k \to \mathbb{R}$ be continuously differentiable in a neighborhood of $g(\mu)$ with $\nabla g(\mu) \neq 0$, where ∇g denote the first order derivative of a function g in $\mathbb{R}^k$. The parameter of interest

is $\theta = g(\mu)$, which can be estimated by $\hat{\theta} = g(\bar{X})$ with $\bar{X} = n^{-1}\sum_{1}^{n} X_i$. Let σ^2 denote the asymptotic variance of $\sqrt{n}\left(\hat{\theta} - \theta\right)$ and $\hat{\sigma}^2$ an estimator of σ^2. Suppose we are interested in the expansion of the following statistic

$$T = n^{1/2}\frac{\left(\hat{\theta} - \theta\right)}{\hat{\sigma}^2},$$

Under sufficient regularity conditions, the distribution function of T may be expanded as

$$G(x) = P(T \leq x) = \Phi(x) + n^{-1/2}q(x)\phi(x) + O\left(n^{-1}\right), \tag{1.4.1}$$

where q is an even quadratic polynomial and Φ, ϕ are the Standard Normal distribution and the Standard Normal density, respectively. Hall (1992) shows that the bootstrap estimate G admits an analogous expansion.

$$\hat{G}(x) = P(T^* \leq x) = \Phi(x) + n^{-1/2}\hat{q}(x)\phi(x) + O\left(n^{-1}\right), \tag{1.4.2}$$

where T^* is the bootstrap version of T, and the polynomial $\hat{q}$ is obtained from q by replacing unknowns, such as skewness, by bootstrap estimates. Note that the coefficients of $\hat{q}$ depends on the estimated moments of X up to the third order. By the central limit theorem, these coefficients can be estimated $\sqrt{n}$ consistently, so that $\hat{q} = q + O_p\left(n^{-1/2}\right)$. Subtracting equation (1.4.1) from (1.4.2) we get

$$P^*\left[T_n^* \leq x\right] - P\left[T_n \leq x\right] = O_p\left(n^{-1}\right).$$

Therefore, the bootstrap approximation to G is in error by only n^{-1} whereas the asymptotic normal approximation is of order $O_p\left(n^{-1/2}\right)$. It is important to stress on the

importance of asymptotically pivotal statistics. Recall that a statistic T_n is (asymptotically) pivotal if its (asymptotic) distribution does not depend on any unknown parameters.

Example 2 *A pivotal statistic*

Consider the distribution of $T_n = \sqrt{n}\left(\bar{X} - \mu\right)$ where $X_i \backsim N\left(\mu, \sigma^2\right)$. The finite sample distribution of T_n is given by

$$G_n = P\left[\sqrt{n}\left(\bar{X} - \mu\right) \leq x\right] = \Phi\left(\frac{x}{\sigma}\right),$$

where Φ denotes the distribution function of $Z \backsim N\left(0, 1\right)$. Therefore, the finite sample distribution G_n depends on the nuisance parameter σ^2. However, the studentized root

$$T_n = \sqrt{n}\left(\bar{X} - \mu\right)/s_n,$$

where $s_n^2 = (n-1)^{-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$ has finite sample distribution given by

$$G_n = P\left[\sqrt{n}\left(\bar{X} - \mu\right)/s_n \leq x\right] = \mathcal{T}_{(n-1)}\left(x\right),$$

where $\mathcal{T}_{(n-1)}$ denotes the Student t-distribution with $n-1$ degree of freedom. Therefore, under normality assumption the studentized root $T_n = \sqrt{n}\left(\bar{X} - \mu\right)/s_n$ is a pivotal statistic.■

The bootstrap provides higher-order approximation only for pivotal roots. If the distribution of statistics are not pivotal the bootstrap may still be applied but it does not provide higher-order approximation to their distribution. To see why consider the non pivotal statistic

$$U = n^{1/2}\left(\hat{\theta} - \theta\right),$$

the analogues of (1.4.1) and (1.4.2) in this case are

$$\begin{aligned} H(x) &= P(U \leq x) = \Phi\left(\frac{x}{\sigma}\right) + n^{-1/2} p(\frac{x}{\sigma})\phi(\frac{x}{\sigma}) + O\left(n^{-1}\right), \\ \hat{H}^*(x) &= P(U^* \leq x) = \Phi\left(\frac{x}{\hat{\sigma}^*}\right) + n^{-1/2} \hat{p}(\frac{x}{\hat{\sigma}^*})\phi(\frac{x}{\hat{\sigma}^*}) + O\left(n^{-1}\right), \end{aligned}$$

respectively, where p is a polynomial, $\hat{p}$ is obtained from p on replacing unknowns by their bootstrap estimates, σ^2 equals the asymptotic variance of U, $\hat{\sigma}^{*2}$ is the bootstrap estimate of σ^2, U^* is the bootstrap analog of U. Again $\hat{p} - p = O_p\left(n^{-1/2}\right)$, and also $\hat{\sigma}^* - \sigma = O_p\left(n^{-1/2}\right)$. Therefore,

$$\hat{H}^*(x) - H(x) = \Phi(x/\hat{\sigma}^*) - \Phi(x/\sigma) + O_p\left(n^{-1}\right),$$

Considering the distribution of H, the first term of its Edgeworth expansion depends on unknown parameter, typically, it will have the form $N(0, \sigma^2)$. Hence, the Edgeworth expansion of $\hat{H}^*$ will have leading term $N(0, \sigma^2)$. It follows, that the error of the bootstrap will be controlled by the error of $\hat{\sigma}^2$ which is usually of order $O\left(n^{-1/2}\right)$. This is of the same order as the error of the standard normal distribution.

To summarize, the bootstrap provide an improvement upon the first order approximation because the approximation error of the bootstrap distribution for asymptotically pivotal statistics is of order $O\left(n^{-1}\right)$ and not $O\left(n^{-1/2}\right)$. Of course, the bootstrap cannot be expected to perform well when the Edgeworth expansion provides a poor approximation to the distribution of interest. A case in point is the instrumental-variables estimation with poorly correlated instruments and regressors (see Hillier (1985)).

1.4.2 The bootstrap asymptotic minimax property

We now consider the accuracy of bootstrapping statistics based on the minimax error measures. We first consider non-studentized, and then we extend the results to studentized statistics.

Let $X_1, X_2, ..., X_n$ be independent identically distributed random variables with unknown distribution function F. Define $\left\{\hat{T}_n, n \geq 1\right\}$ as statistics based on $X_1, X_2, ..., X_n$, (e.g. $\hat{T}_n = g\left(\hat{F}_n\right)$, $\hat{F}_n$ as the empirical CDF of $X_1, X_2, ..., X_n$, and $T_n(F)$ as the asymptotic centering for $\hat{T}_n$. So that $\{T_n(F)\}$ is a sequence of random numbers such that

$$H_{n,F}(x) = P_F\left[\sqrt{n}\left(\hat{T}_n - T_n(F)\right) \leq x\right] \longrightarrow \Phi\left(\frac{x}{\sigma}\right),$$

where Φ indicate the CDF of a Normal with mean zero, and σ is a scaling factor[7]. This distribution can be approximated using the non-parametric bootstrap distribution

$$H_n^*(x) = H_{n,\hat{F}_n}(x) = P_F^*\left[\sqrt{n}\left(\hat{T}_n^* - T_n\left(\hat{F}_n\right)\right) \leq x\right],$$

where the symbol "$*$" signals the bootstrap framework as before.

We are interested in how well H_n^* approximates $H_{n,F}$. Beran (1982) has shown that $\hat{H}_n^*$ is asymptotically minimax, (i.e. minimizes the maximum risk over a neighborhood). The normal approximation $\Phi(x/\sigma)$ is asymptotically minimax if and only if the distribution of the root $\sqrt{n}\left(\hat{\theta} - \theta\right)$ has no skewness and bias of order $o(n^{-1})$. Hence, although the bootstrap has the same convergence rate as the normal approximation in the non-pivotal case, the bootstrap can be superior in term of the minimax criterion. Beran also shows that the distribution based on the first order Edgeworth expansion is asymptotically minimax

[7] Note: Defining $T_n(F)$ as a sequence and rather then just a constant allows us to accommodate for the most general case.

implying that the bootstrap is asymptotically equivalent to the first order Edgeworth expansion. However, Beran's definition of asymptotic optimality involves computing the error $\hat{H}_n - H_{n,F}$ using a smooth distribution function v before defining the risk of $H_{n,F}$, where $\hat{H}_n$ is an estimate of $H_{n,F}$. Hence, Beran's proof makes use of the existence of a uniform one term Edgeworth expansion for $v * H_{n,F}$. By contrast, Singh and Babu (1990) reach the same result proving again the existence of one term Edgeworth expansion, but they remove this artificial smoothing and obtain the required uniform one term Edgeworth expansion for $H_{n,F}$ itself. For this reason we will consider Singh and Babu work in more details.

As anticipated, the bootstrap has an asymptotic minimax property. This means that the bootstrap achieves the smallest risk in the worst case asymptotically. What do we mean by "risk" in the "worst case"? In order to proceed we restrict our attention to continuous distributions.

Let $\delta > 0$ and $K > 1$ be given. Let $\mathcal{F}$ denote the class of non-lattice distribution $G \in \mathbb{R}$ satisfying the inequality

$$\int |x|^{4+\delta} dG(x) \leq K,$$

so that K is an upper bound for the fourth moment.

Singh and Babu define a risk function and consider the sample mean of n independent observations of $X_1, X_2, ..., X_n$ from the distribution G, and they show that the bootstrap estimator of the sample mean attends this lower bound.

So, for $G \in \mathcal{F}$, let

$$\mu_G = \int x dG(x),$$

$$\sigma_G^2 = \int (x - \mu_G)^2 \, dG(x),$$
$$H_{n,G} = P\left[\sqrt{n}\left(\bar{X} - \mu_G\right) \leq x\right],$$

where $\bar{X} = n^{-1} \sum_{i=1}^{n} x$. The risk function is defined under a monotonically increasing loss function

$$u : [0, \infty) \to [0, \infty)$$

and is defined by

$$R_n\left(\hat{H}_n, G\right) = E_G\left\{u\left(\sqrt{n}\|\hat{H}_n - H_{n,G}\|\right)\right\}, \tag{1.4.2.1}$$

where $\|\cdot\|$ stands for the supnorm[8]. So, the expected loss under G is proportional to u which is a monotonically increasing function of the maximum distance between $\hat{H}_n$ and $H_{n,G}$ multiplied by the scalar $\sqrt{n}$ which prevent the argument of u collapsing to zero.

Theorem 1 below gives an asymptotic lower bound for R_n. To find a lower bound we need to define a neighborhood of the distribution function F, this requires some additional notation. Define

$$B_{n,c,F} = \left\{G \in \mathcal{F} : \|F - G\| \leq c/\sqrt{n}\right\},$$

as the Kolmorogov's ball of radius $c/\sqrt{n}$,

$$a = \left(\mu_{4,F} - \sigma_F^4\right)^{1/2} \left(8\pi\sigma_F^4\right),$$

---

[8] The Euclidean norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\|x\| = \left\|(x_1, ..., x_n)'\right\| = \sqrt{\sum_{i=1}^{n} x_i^2},$$

so, the supnorm is given by

$$\|F \cdot G\| = \sup_x |F(x) - G(x)|.$$

and

$$R(F) = \int u(a|z|\varphi(z)\,dz).$$

Theorem 1 *(cf. Singh and Babu). For* $F \in \mathcal{F}$,

$$lim_{c\to\infty}\ liminf_{n\to\infty}\ inf_{\hat{H}_n}\ sup_{n,c,F}\ R_n\left(\hat{H}_n, G\right) \geq R(F),$$

where $\inf_{\hat{H}_n}$ *denotes the infimum over all estimator* $\hat{H}_n$ *of* $H_{n,G}$ *and* $\sup_{n,c,F}$ *denotes the supremum over all distribution functions* G *in* $B_{n,c,F}$.

The formal proof of this theorem is beyond our purposes, and we refer the interested reader to the original article. Nevertheless, we will give an heuristic explanation of the meaning of this theorem.

"Decode" for theorem 1 :

1) Consider

$$\sup_{n,c,F} R_n\left(\hat{H}_n, G\right) \geq R(F),$$

since F is fixed, we can focus our attention on sup of n, c and vary G in the Kolmorogov's ball. The worst G in $B_{n,c,F}$ has risk at least as bad as $R(F)$.

2) Consider

$$\inf_{\hat{H}_n}\ \sup_{G\in B_{n,c,F}} R_n\left(\hat{H}_n, G\right) \geq R(F).$$

Let now treat n, c as given, and pick $\hat{H}_n$ to minimise the worst risk that can occur for a given $\hat{H}_n, G$ combination. Let now consider what happens in the limit

3)

$$\liminf_{n\to\infty}\ \inf_{\hat{H}_n}\ \sup_{G\in B_{n,c,F}}\ R_n\left(\hat{H}_n, G\right) \geq R(F),$$

so, for all n sufficiently large we can never get a risk better than $R(F)$.

4)

$$\lim_{c\to\infty} \liminf_{n\to\infty} \inf_{\hat{H}_n} \sup_{n,c,F} R_n\left(\hat{H}_n, G\right) \geq R(F).$$

Recall that F is fixed. As the neighborhood around F expands $(i.e. \lim_{c\to\infty})$ we get closer and closer to a situation in which for all n sufficiently large $(i.e. \liminf_{n\to\infty})$ the best estimator $\hat{H}_n$ for the neighborhood $\left(i.e. \inf_{\hat{H}_n}\right)$ can never do better than a risk $R(F)$.

Singh and Babu show that the lower bound $R(F)$ is attained for the bootstrap estimate $\hat{H}_n^*$ of $H_{n,F}$ as well as for H_n^E, the empirical one-term Edgeworth expansion given by

$$H_n^E = \Phi(x/s_n) - \frac{1}{6\sqrt{n}}\left(\hat{\mu}_{3,n}\right)\left(x^2-1\right)\varphi(x/s_n) \qquad (1.4.2.2)$$

where s_n is the sample standard deviation and $\hat{\mu}_{3,n}$ is the third central sample moment.

The results established above for the sample mean can be extended to functions of multivariate statistics, corresponding studentized statistics and ratio estimators when the auxiliary variable is lattice. We just consider studentized roots.

Suppose $Z_1, ..., Z_n$ be i.i.d. random $k \times 1$ vectors with distribution function F. Let $g : \mathbb{R}^k \to \mathbb{R}$ with $y \in C^3\left(\mathbb{R}^k\right)$ continuously differentiable of order three on $\mathbb{R}^k$. Let $\mu = E(z_i)$ and $\widetilde{T}_n = \sqrt{n}\left[g(\bar{z}) - g(\mu)\right]$. In the simplest case of the studentized mean

$$T_n = \sqrt{n}\left[\bar{X}_n - \mu_x\right]/S_n,$$

where $\mu_x = E(X_i)$ we can write T_n in the form

$$\sqrt{n}\left[g\left(\bar{Z}_n\right) - g\left(E\left(Z_i\right)\right)\right],$$

where $Z_i = \left[X_i - \mu_x, (X_i - \mu_x)^2\right]$ and $g(z_1, z_2) = z_1/(z_2 - z_1^2)^{1/2}$. In fact, the normalised distribution $\left(g\left(\bar{Z}_n\right)\right)$ is given by

$$\begin{aligned}
g\left(\bar{Z}_n\right) &= g\left(n^{-1}\sum_{i=1}^{n}(X_i - \mu_x), n^{-1}\sum_{i=1}^{n}(X_i - \mu_x)^2\right) \\
&= \frac{n^{-1}\sum_{i=1}^{n}(X_i - \mu_x)}{\left(n^{-1}\sum_{i=1}^{n}(X_i - \mu_x)^2 - \left(n^{-1}\sum_{i=1}^{n}(X_i - \mu_x)\right)^2\right)^{1/2}} \\
&= \frac{n^{-1}\sum_{i=1}^{n}(X_i - \mu_x)}{\left(n^{-1}\sum_{i=1}^{n}X_i^2 - 2\mu_x\bar{X}_n + \mu_x^2 - \left(\bar{X}_n^2 - 2\mu_x\bar{X}_n + \mu_x^2\right)\right)^{1/2}} \\
&= \frac{\bar{X}_n - \mu_x}{\left(n^{-1}\sum_{i=1}^{n}X_i^2 - \bar{X}_n^2\right)^{1/2}} \\
&= \frac{\bar{X}_n - \mu_x}{S_n^2} = g\left(\bar{Z}_n\right).
\end{aligned}$$

Define the distribution of interest as

$$H_{n,F}(x) = P_F\left[\sqrt{n}\left(\tilde{T}_n - T_n(F) \leq x\right)\right].$$

We have already seen in the previous section that by the central limit theorem

$$H_{n,F_n}(x) = P_F\left[\sqrt{n}\left(\tilde{T}_n - T_n(F) \leq x\right)\right] \to \Phi(x),$$

this implies

$$H_n^*(x) \to H_{n,\hat{F}_n}(x),$$

almost surely. We want to know how good is the bootstrap $H_n^*(x)$ as approximation to $H_{n,F_n}(x)$. Let $\delta \geq 0$, $k > 1$ be fixed, and $\mathcal{F}_k$ be the class of k-variate strongly non-lattice

(i.e. continuous in all dimensions) distributions on $\mathbb{R}^k$ which satisfy

$$\int \|z\|^{4+\delta} dG(z) \leq k,$$

where $\|z\| = (z^2)^{1/2}$. Babu and Singh (1984) have shown that for studentized statistics, if enough moments are assumed, typically

$$n \left\| H_{n,F_n} - H_n^* \right\| = O_p(1).$$

Therefore, the risk function (1.4.2.1) tends to zero. Thus, in this case to get meaningful results the risk function has to be redefined as

$$R_n\left(\hat{H}_n, G\right) = E_G\left\{u\left(n\|\hat{H}_n - H_{n,G}\|\right)\right\},$$

(i.e. Considering n in place of $\sqrt{n}$ prevents a degenerate argument in $R_n\left(\hat{H}_n, G\right)$). To obtain the optimality properties we need to consider two-term Edgeworth expansion instead of the one-term Edgeworth expansion like (1.4.2.2).

To summarise, if there are sufficient conditions to ensure a two-term empirical Edgeworth expansion (H_n^E), then both H_n^* and H_n^E achieve the lower bound, hence are asymptotically optimal.

1.5 Bootstrapping test statistics

In this section we consider applications of the bootstrap method to test statistics. Suppose we want to test the null hypothesis H_0 using the asymptotically pivotal statistic T_n. Consider a symmetrical smooth two tailed test of H_0. This test rejects H_0 at the α level if $|T_n| > z_{n,\alpha/2}$, where $z_{n,\alpha/2}$ is the exact $1-\alpha/2$ quantile of the distribution of T_n. The

critical value solves the equation

$$G_n\left(z_{n,\alpha/2}, F\right) - G_n\left(-z_{n,\alpha/2}, F\right) = 1 - \alpha. \tag{1.5.1}$$

Since F is unknown the finite-sample critical value cannot be obtained unless T_n is exactly pivotal. First order asymptotic approximation replaces the unknown distribution G_n with the known function G (i.e. the asymptotic distribution of T_n). According to the standard asymptotic theory, the critical value approximates the exact finite sample critical value with an error whose size is of order $O\left(n^{-1}\right)$.

The bootstrap provides an alternative approximation to the finite-sample distribution of the statistic T_n. In other words, the bootstrap replaces F with F_n. Thus the bootstrap critical value solve the equation

$$G_n\left(z^*_{n,\alpha/2}, F_n\right) - G_n\left(-z^*_{n,\alpha/2}, F_n\right) = 1 - \alpha. \tag{1.5.2}$$

The distribution of (1.5.2) usually cannot be solved analytically. However, $z^*_{n,\alpha/2}$ can be estimated with the desired accuracy by Monte Carlo.

To evaluate the accuracy of the bootstrap critical value $z^*_{n,\alpha/2}$ we need to consider again the Edgeworth expansion. We have seen in the previous section that the error of the bootstrap approximation to a one-sided[9] distribution function is

$$G_n\left(\tau, F_n\right) - G_n\left(\tau, F\right) = O_p\left(n^{-1}\right),$$

almost surely uniformly over τ. Consider now the error for a symmetrical distribution function. In this case, for asymptotically pivotal roots, the accuracy of the bootstrap is then as

[9] For one-side distribution function we mean $\Pr(|T_n| \leq z_{n,\alpha})$ when the statistic is symmetrically distributed about 0.

for a one-sided distribution. Consider the distribution function

$$P\left(|T_n| \leq \tau\right) = G_n\left(\tau, F\right) - G_n\left(-\tau, F\right).$$

Let Φ denote the Standard Normal distribution function. Then it follows from the asymmetry of the polynomials q in their first argument that

$$\begin{aligned} G_n\left(\tau, F\right) - G_n\left(-\tau, F\right) &= \left[G\left(\tau, F\right) - G\left(-\tau, F\right)\right] + \\ &\quad + \frac{2}{n} q\left(\tau, F\right) + O\left(n^{-2}\right) \\ &= 2\Phi\left(\tau\right) - 1 + \frac{2}{n} q\left(\tau, F\right) + O\left(n^{-2}\right). \end{aligned} \tag{1.5.3}$$

Similarly, from (1.5.2) it follows that

$$\begin{aligned} G_n\left(\tau, F_n\right) - G_n\left(-\tau, F_n\right) &= \left[G\left(\tau, F_n\right) - G\left(-\tau, F_n\right)\right] + \\ &\quad + \frac{2}{n} q\left(\tau, F_n\right) + O\left(n^{-2}\right) \\ &= 2\Phi\left(\tau\right) - 1 + \frac{2}{n} q\left(\tau, F_n\right) + O\left(n^{-2}\right), \end{aligned} \tag{1.5.4}$$

almost surely. The remainder terms (1.5.3) and (1.5.4) are $O\left(n^{-2}\right)$. Now subtracting (1.5.3) from (1.5.4) and using the fact that $F_n - F = O\left(n^{-1/2}\right)$ almost surely to obtain

$$\begin{aligned} \left[G_n\left(\tau, F_n\right) - G_n\left(-\tau, F_n\right)\right] - \left[G_n\left(\tau, F\right) - G_n\left(-\tau, F\right)\right] &= \\ &= \frac{2}{n}\left[q\left(\tau, F_n\right) - q\left(\tau, F_n\right)\right] + O\left(n^{-2}\right) \\ &= O\left(n^{-3/2}\right) \end{aligned}$$

almost surely if T_n is asymptotically pivotal. Thus the error made by the bootstrap approximation in this symmetrical distribution function is $O\left(n^{-3/2}\right)$ compared to the error of $O\left(n^{-1}\right)$ made by the first-order asymptotic approximation.

In the same way, to evaluate the accuracy of the bootstrap critical value $z^*_{n,\alpha/2}$ as an estimator of the exact finite-sample critical value $z_{n,\alpha/2}$, combine (1.4.1) and (1.5.1) to obtain

$$2\Phi\left(z_{n,\alpha/2}\right) - 1 + \frac{2}{n}q\left(z_{n,\alpha/2}, F\right) = 1 - \alpha + O\left(n^{-2}\right). \tag{1.5.5}$$

Similarly, combining (1.4.2) and (1.5.2) yields

$$2\Phi\left(z^*_{n,\alpha/2}\right) - 1 + \frac{2}{n}q\left(z^*_{n,\alpha/2}, F_n\right) = 1 - \alpha + O\left(n^{-2}\right). \tag{1.5.6}$$

almost surely. Hall (p. 111) shows that equation (1.5.5) and (1.5.6) can be solved to yield Cornish-Fisher expansions for $z_{n,\alpha/2}$ and $z^*_{n,\alpha/2}$. The results are

$$z_{n,\alpha/2} = z_{\infty,\alpha/2} - \frac{1}{n}\frac{q\left(z_{\infty,\alpha/2}, F\right)}{\phi\left(z_{\infty,\alpha/2}\right)}, \tag{1.5.7}$$

and

$$z^*_{n,\alpha/2} = z_{\infty,\alpha/2} - \frac{1}{n}\frac{q\left(z_{\infty,\alpha/2}, F_n\right)}{\phi\left(z_{\infty,\alpha/2}\right)}, \tag{1.5.8}$$

almost surely. It follows from (1.5.7) and (1.5.8) that

$$z^*_{n,\alpha/2} = z_{n,\alpha/2} + O\left(n^{-3/2}\right),$$

almost surely. Therefore, the bootstrap critical value for a two tailed test is more accurate then the asymptotic critical value, the error of the second being $O\left(n^{-1}\right)$ and the error of the first $O\left(n^{-3/2}\right)$.

1.6 Bootstrap methods for time series

In recent years there have been many developments in bootstrapping time series (see for example Li and Maddala (1997) and Berkowitz and Kilian (2000) for excellent surveys). In the literature, two main approaches to implement the bootstrap in dependent data setting have been proposed. The first approach models the dependent process as one that is driven by i.i.d. disturbances, which allows one to use the resampling scheme of the residual-based bootstrap in linear regression models. The other way is to resample blocks of adjacent observation instead of individual residuals. We will consider them in turn.

1.6.1 Residual based resampling

Before considering the applications of the bootstrap method for time series models it may be useful to consider the residual-based bootstrap for a standard regression model. Consider the model

$$y_i = X_i'\beta + \varepsilon_i \tag{1.6.1.1}$$

where ε_i are i.i.d. $\varepsilon_i \backsim F_n(0, \sigma_i^2)$, β is a k-vector of parameters. The vector β may be estimated using the OLS method. Let

$$\hat{\beta} = (X'X)^{-1} X'y$$

where $X = (x_1, ..., x_n)'$, $y = (y_1, ..., y_n)$, and $\hat{\beta}$ be the OLS estimator of β. The bootstrap can be used to conduct inference on the distribution of the estimator vector $\hat{\beta}$. We distinguish between the non-parametric bootstrap and the parametric bootstrap according to the assumptions we make on the distribution F_n of the residuals. For the non-parametric

bootstrap we make no assumptions about F_n, but we use the empirical distribution of the bootstrap residuals $\varepsilon_t^* = y_t - X_t'\hat{\beta}$ to estimate F_n. We know that $\hat{F}_n$ converge to F_n by the Glivenko Cantelli theorem. By contrast, in the parametric bootstrap we make an assumption on the parametric family for $\hat{F}_n$, for example we may assume that $\hat{F}_n \backsim N(0, \sigma^2)$. No matter the assumptions on $\hat{F}_n$, we then proceed as follows:

Step 1) Generate a random sample of $\varepsilon_1^*, ...\varepsilon_n^*$.

Step 2) Calculate $y_i^* = X'\beta + \varepsilon_i^*$.

Step 3) Calculate the OLS estimate β^* using the y_t^*.

Step 4) Repeat Steps (1)-(3) B times.

This procedure gives B bootstrapped estimates of β which can be used for example to evaluate the accuracy of the estimator $\hat{\beta}$. Freedman (1981) has shown that for the residual based resampling described originally by Efron (1979) the bootstrap distribution of $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)$ is strongly consistent for the distribution of $\sqrt{n}\left(\hat{\beta} - \beta\right)$. Moreover, it appears that the residual based bootstrap works under weaker conditions than those which are necessary for the classical normal approximation (see Mammen (1993)).

Example 3 *Non-Parametric bootstrap for the OLS coefficients in a simple regression model using simulated data.*

Consider the model for the univariete case of (1.6.1.1)

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

where ε_i $(i = 1, ..., 50)$ are i.i.d. $\varepsilon_i \backsim F(0, \sigma_i^2)$. In Table 1.1 we present a simple example of the residual-based bootstrap for this model using the simulated data. In the second column we report the parameters of the data generation process (DGP), in the third column the OLS estimates and the corresponding standard errors, in the fourth the mean of the $\hat{\beta}^*$ and their standard deviations, and the t-statistic in the last column. All simulations were carried out using the matrix programming language GAUSS Version 3.2.32. The random numbers were generated by the function $rndns$. For each sample we calculated the $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{S}^2$, then we generated $B = 10,000$ bootstrap samples according to the algorithm given above.

Table 1.1.

	GDP	OLS	B_Mean	$t - Stat$
β_1	2	1.95431 (0.2735)	1.9534 (0.2631)	7.1440
β_2	1	0.9833 (0.0181)	0.9833 (0.0175)	54.229
S^2	1	0.8552	0.7867 (0.1228)	–

As we can see from Figure 1.1 the distribution of the $\hat{\beta}_i^*$ are reasonably closed to that of the $\hat{\beta}_i$, and they are normally distributed. Unfortunately, the same cannot be said for the estimate of the variance since it is clearly not distributed as χ^2 with 48 degree of freedom.

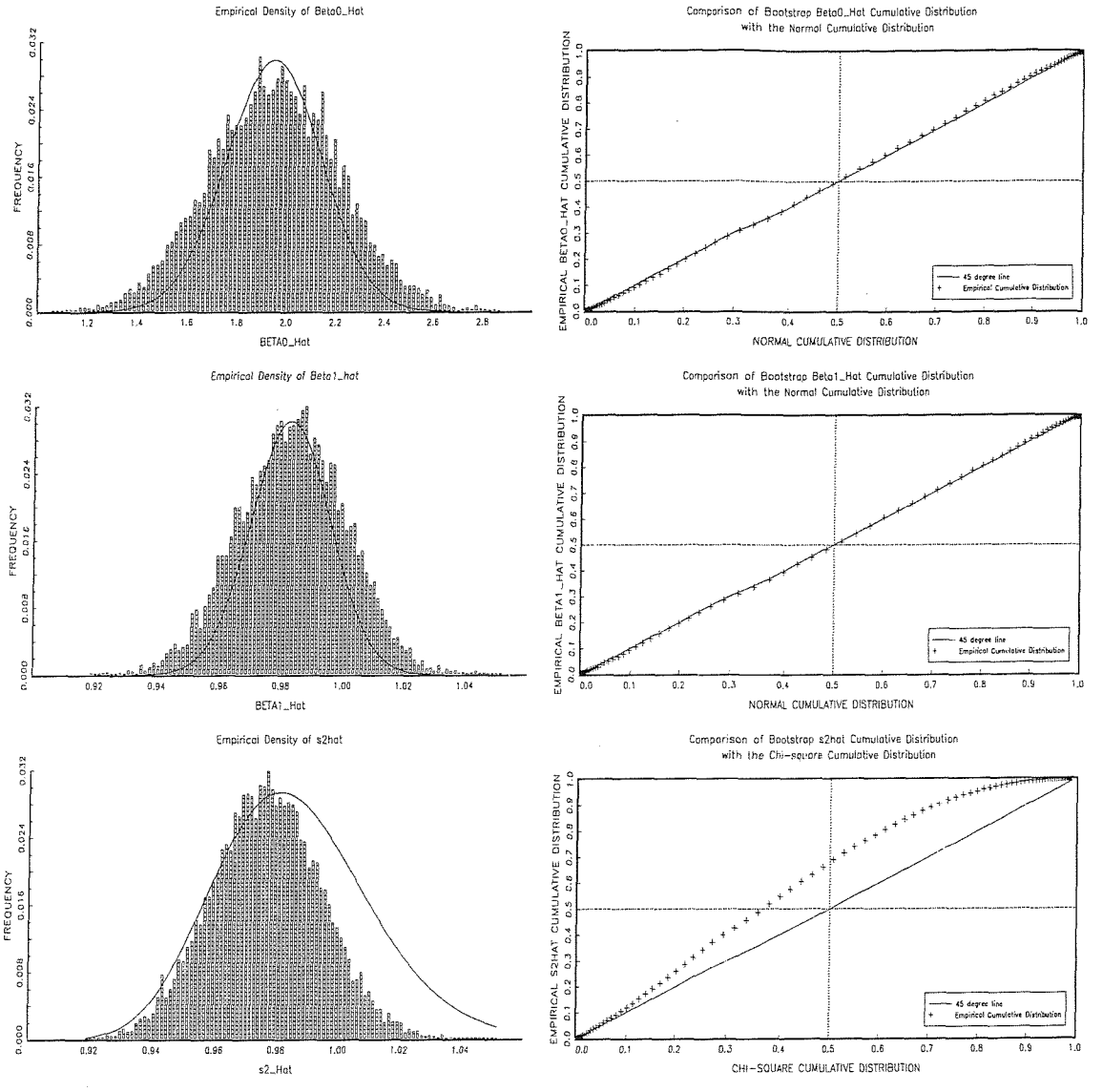

Figure 1.1. Approximation to OLS density in linear model.

Note that Efron's (1979) original bootstrap algorithm was designed for data which are independent and identically distributed. If the data come from a heterogeneous population the bootstrap will fail to give good results. An heuristic explanation for this failure is that if the data display heteroskedasticity or serial correlation a randomly resampled set of data will not preserve these properties, and as a result statistics calculated from the resampled data will not be consistent. This problem is particularly important when we consider time-dependent data.

Early application of the bootstrap algorithm to time dependent data assumed that the underlying process follows a stationary $AR(1)$ model. Consider the model

$$y_t = \rho y_{t-1} + \varepsilon_t, \qquad y_0 = 0, \tag{1.6.1.2}$$

where ε_t is i.i.d. with common distribution $F\left(0, \sigma^2\right)$ and $0 < \sigma^2 < \infty$. For this model the least square estimator $\hat{\rho}$ of ρ is given by

$$\hat{\rho} = \left(\sum_{t=1}^{n} y_t y_{t-1}\right)\left(\sum_{t=1}^{n} y_{t-1}^2\right)^{-1}.$$

The consistency of $\hat{\rho}$ was already established in the 50s'. However, the limit distribution of $\hat{\rho}$ is different for the three possible cases: stationary, unit root, and explosive. Indeed, if $|\rho| < 1$, then

$$T\left(\hat{\rho} - \rho\right) \xrightarrow{d} N\left(\rho, 1 - \rho^2\right),$$

if $|\rho| = 1$

$$T\left(\hat{\rho} - 1\right) \xrightarrow{d} \frac{\int W\left(r\right) dW\left(r\right)}{\int W\left(r\right)^2 dr};$$

where $\int$ is a shorthand for $\int_0^1$ and $W(r)$ denotes the univariate Wiener process on $[0, 1]$. In this case the limiting distribution of $\hat{\rho}$ is neither normal nor symmetric since it is negatively

skewed. Furthermore, $\hat{\rho}$ is super consistent since it converges to its true value at rate $1/T$ instead of the usual rate $1/\sqrt{T}$.

The asymptotic validity of the bootstrap estimator corresponding to ρ for the stationary case (i.e. $|\hat{\rho}| < 1$) was established by Bose (1988). Using the Edgeworth expansion for sum of dependent random variables, Bose proved that the bootstrap distribution is second order accurate for the distribution of the studentized root $R_n = (\hat{\rho} - \rho)/\hat{\sigma}_\rho^2$,(where $\hat{\sigma}_\rho^2$ denotes the usual OLS estimator of the variance of $\hat{\rho}$); this property extends to $AR(p)$ models with unknown mean.

Basawa *et al.* (1989) considered the case $|\rho| > 1$ and they established the validity of the bootstrap even for explosive processes.

The consistency of the bootstrap estimator of the distribution of the slope coefficient or studentized slope coefficient in a simple unit-root model has been investigated by Basawa *et al.* (1991a, 1991b), Datta (1996), and Ferretti and Romo (1996).

Basawa *et al.* (1991a) consider the consistency of the t-statistic for ρ in the special case $\varepsilon_t \backsim N(0,1)$ given by

$$t_n = \left(\sum_{t=1}^{n} y_{t-1}^2 \right)^{1/2} (\hat{\rho} - \rho)$$

and they investigate the distribution of

$$t_n^* = \left(\sum_{t=1}^{n} \left(y_{t-1}^* \right)^2 \right)^{1/2} (\hat{\rho}^* - \rho)$$

obtained by replacing y_i with y_i^* in (1.6.1.2). They show that the null limit distribution of t_n^* conditional on $(y_1, ..., y_n)$ is not the same as that of t_n , so that the bootstrap distribution function $P_n^* (t_n^* < \tau)$ does not consistently estimate the population distribution function

$P(t_n < \tau)$. The problem arises from the discontinuity in the asymptotic distribution of $\hat{\rho}$ at $\rho = 1$, since as we have seen above this is different according to $\rho \gtrless 1$. It follows that the third condition given in Section 1.3 is not satisfied if the DGP under consideration includes ones with and without $\rho = 1$. However, Basawa *et al.* (1991b) show that it is possible to overcome this problem by specifying that $\rho = 1$ when constructing ε_i^*, therefore removing the source of discontinuity.

For the model specified in (1.6.1.2) consider the null hypothesis $H_0 : \rho = 1$. A natural test statistic for testing H_0 is given by

$$z_n = n(\hat{\rho} - 1).$$

The bootstrap sample $\{y_t^*, t = 1, ..., n\}$ is generated recursively from the estimated model

$$y_t^* = y_{t-1}^* + \varepsilon_t^*, \tag{1.6.1.3}$$

where $y_0^* = 0$. The centered residuals are $\hat{\varepsilon}_t = y_t - y_{t-1} - \bar{\varepsilon}$, where $\bar{\varepsilon} = n^{-1} \sum_{i=1}^{n} (y_t - y_{t-1})$ [10]. The bootstrap analogue of the statistic z_n is $z_n^* = n(\rho^* - 1)$ where

$$\hat{\rho}^* = \left(\sum_{t=1}^{n} y_t^* y_{t-1}^* \right) \left(\sum_{t=1}^{n} \left(y_{t-1}^* \right)^2 \right)^{-1}.$$

Basawa *et al.* (1991b) derive the null limit distribution of z_n^* conditional on $(y_1, ..., y_n)$, and show that if H_0 is true, then $|P_n^*(z_n^* < z) - P(z_n < z)| = o_p(1)$ uniformly over z.

Datta (1996) established that the discontinuity problem can be overcome without restriction. Consider the model (1.6.1.2) with the additional assumption that the second

[10] Note: If an intercept is not included in the model, the residuals $\hat{\varepsilon}_t$ must be recentered prior to resampling to ensure that their bootstrap population mean is zero.

moment is bounded (i.e. $E\left(|\varepsilon_t|^{2+\delta} < \infty \text{ for some } \delta > 0\right)$. Let

$$\hat{\varepsilon}_t = y_t - \hat{\beta} y_{t-1} - T^{-1} \sum_{t=1}^{n} \left(y_t - \hat{\beta} y_{t-1}\right)$$

denote the centered residuals from the estimated model, and let $\hat{\varepsilon}_t^*$ be a random sample of $\hat{\varepsilon}_t$ for some $m < n$. The random sample is generated by the recursion

$$y_t^* = \hat{\beta} y_{t-1}^* + \varepsilon_t^*$$

but with $i = 1, ..., m$ instead of $i = 1, ..., n$. Define the bootstrap version of t_n by

$$t_m^* = \left[\sum_{i=1}^{m} \left(y_{t-1}^*\right)\right]^{1/2} \left(\hat{\beta}_m^* - \beta\right).$$

Datta proved that if $\left[m \left(\log \log n\right)^2\right] / n \to 0$ as $n \to \infty$, then

$$\left|P_m^* \left(t_m^* < z\right) - P_n \left(t_n < z\right)\right| = o\left(1\right),$$

almost surely as $n \to \infty$ uniformly over z for any $\beta \in (-\infty, \infty)$.

The asymptotic validity of the bootstrap has been established also for the stationary autoregressive moving average of order (p, q) ($ARMA(p, q)$). The $ARMA(p, q)$ is given by model

$$\rho\left(L\right) y_t = \varphi\left(L\right) \varepsilon_t$$

where $\rho\left(L\right) = 1 - \rho_1 L - ... - \rho_q L^p$, $\varphi\left(L\right) = 1 - \varphi_1 L - ... - \varphi_q L^q$, and L is the lag operator defined by the property $L y_t = y_{t-1}$. Let $\theta = \left(\rho_1, ..., \rho_p, \varphi_1, ..., \varphi_q\right)$ denote the $(p + q)$-vector of unknown parameters. Assume that the process $\{y_t\}$ is stationary as well as invertible. If $\hat{\theta}$ denotes an M-estimator of θ, then Kreiss and Franke (1992) have shown that the Mallows distance between the bootstrap distribution of $\sqrt{T}\left(\hat{\theta}^* - \hat{\theta}\right)$ and the finite-sample distribution of $\sqrt{T}\left(\hat{\theta} - \theta\right)$ converge to zero in probability. The bootstrap estimator

$\hat{\theta}^*$ is based on the resampling scheme

$$\hat{\rho}(L) y_t^* = \hat{\varphi}(L) \varepsilon_t^*,$$

where $\hat{\rho}(L) = 1 - \hat{\rho}_1 L - ... - \hat{\rho}_q L^p$, $\hat{\varphi}(L) = 1 - \hat{\varphi}_1 L - ... - \hat{\varphi}_q L^q$, and ε_t^* is sampled from the set of centered $ARMA$ residuals. In $MA(q)$ models Bose (1990) proved that the bootstrap is second-order correct if the bootstrap is based on a studentized root.

Li and Xiao (2001) establish the validity of the residual based bootstrap for the regression model with $I(1)$ explanatory variables and $I(0)$ errors,

$$y_t = \beta' x_t + u_t.$$

y_t is scalar, but x_t can be a VAR, and u_t can be autocorrelated as well, with appropriate adjustments to the bootstrapping procedure.

1.6.2 Resampling blocks of data

Application of the residual based bootstrap is straightforward if the underlying distribution follows an i.i.d. stationary process. However, if the structure of serial correlation is not tractable or is misspecified, the residual based methods will give inconsistent estimates. To take the dependency into account other approaches which do not require fitting the data into a parametric form have been developed to deal with dependent data. These procedures are called the "moving block bootstrap" because they involve resampling blocks rather then individual data. Blocking methods may involve either non-overlapping blocks (see Carlstein (1986)), or overlapping blocks (see Künsch (1989) and Liu and Singh (1992)). The non-overlapping blocks method divide the data of t observations into blocks of length l and

select b of these blocks by resampling with replacement all the possible blocks. In the overlapping blocks method there are $t-l+1$ blocks. The blocks are $L_k = \{x_k, x_{k+1}, ..., x_{k+l-1}\}$ for $k = 1, 2, ..., (t-l+1)$.

The idea that underlies the block resampling scheme is that if the blocks are long enough, then enough of the original dependence is preserved in the resampled series. Therefore, the bootstrapped statistics, say $\hat{T}^*$, under consideration will have approximately the same distribution as the value T calculated from the replicates of the original series. Of course, this approximation will be best if the dependence is weak and the blocks are as long as possible, so that the dependence is preserved more faithfully. However, the drawback of the moving block bootstrap is that pseudo time series generated by the moving block method is not stationary even if the original series are is stationary. For this reason Politis and Romano (1994) suggest the stationary bootstrap.

1.6.3 The stationary bootstrap resampling scheme

The main idea behind the stationary bootstrap is to draw a sequence of blocks of random length, where the length of each block has a geometric distribution. Let Y_t denote the vector of m consecutive observations from $\{y_t\}$ starting at y_{t-m+1}. To describe the algorithm let

$$B_{t,l} = \{Y_t, Y_{t+1}, ..., Y_{t+l-1}\},$$

be the block of l observations starting at Y_t. If the index $j > T$, then Y_j is defined as Y_k (i.e. $Y_j = Y_k$), where $k = k(\mathrm{mod}\, T)$ and $Y_0 = Y_T$. Let p be a fixed number in $[0, 1]$. Independent of the data, let $L_1, L_2, ...$ be a sequence of $i.i.d.$ random variables having the geometric distribution, so that the $P\{L_i = m\} = (1-p)^{m-1} p$ for $m \geq 1$. Independent

of the data and the L_i , let $T_1, T_2, ...$ be a sequence of $i.i.d.$ random variables that have the discrete uniform distribution on $\{1, ..., T\}$. Now the bootstrap sample $Y_1^*, ..., Y_T^*$ is generated in the following way. Sample a sequence of blocks of random length

$$B_1^* = B_{T_1,L_1}, B_2^* = B_{T_2,L_2}, ...,$$

the first L_1 observations in the bootstrap sample $Y_1^*, ..., Y_T^*$ are determined by the first block B_1^* of observations containing $Y_{T_1}, ..., Y_{T_1+L-1}$, and the next L_2 observations of Y_t^* are determined by the second block B_2^*, and so forth. This process is stopped once T observations has been generated. Once that the bootstrap sample has been generated, compute the statistic $\hat{T}_n^*$ as usual. Politis and Romano prove the first-order consistency of the stationary bootstrap (see also Lahiri (1999).

1.6.4 The Sieve bootstrap for linear processes

An alternative model-based resampling procedure is the sieve bootstrap. The sieve bootstrap has been studied by Kreiss (1992) Bühlmann (1997) and Paparoditis (1996). The main idea behind the sieve bootstrap is approximating the general linear process by a finite autoregressive process of order increasing with the sample size, and resampling from the approximated autoregressions. Consider the following DGP

$$Y_i - \mu = \sum_{j-1}^{\infty} \zeta_j \left(Y_{i-j} - \mu\right) + \varepsilon_i, \qquad (1.6.4.1)$$

where $\mu = E(Y_i)$ for all i, the process ε_i consist of $i.i.d.$ random variables, and Y_i may be a scalar or a vector process. Assume that $\sum_{j-1}^{\infty} \zeta_j < \infty$ and that the roots of the power series

$$1 - \sum_{j-1}^{\infty} \zeta_j z^j$$

are outside the unit circle. Suppose we approximate (1.6.4.1) by an $AR(p)$ model in which $p = p(n)$ increases with increasing sample size. Let $\{a_{nj} \text{ for } j = 1, ..., p\}$ denote the least squares estimates of the coefficients of the approximating process, and let ε_{nj} denote the centered residuals. The sieve bootstrap consists of generating bootstrap samples to the process

$$Y_i^* - m = \sum_{j-1}^{\infty} a_{nj} \left(Y_{i-j}^* - m \right) + \varepsilon_i^*,$$

where $m = n^{-1} \sum_{i=1}^{n} Y_i$ and ε_i^* are the bootstrap residuals. In the literature Bühlmann (1997) have given conditions under which this procedure consistently estimates the distributions of several statistics, while Choi and Hall (2000) investigate the ability of the sieve bootstrap to provide asymptotic refinements.

1.7 Concluding remarks

The purpose of this chapter has been to illustrate how the bootstrap method works and its properties. It appears that the bootstrap is a general method for estimating the statistical accuracy of an estimator or test statistic by resampling the data. Particular emphasis is given to the bootstrap's ability to improve upon first order approximation. The theoretical explanation of the bootstrap's ability to provide asymptotic refinements is based on the

Edgeworth expansion. Indeed, the bootstrap provides asymptotic refinement because it amounts to a one-term Edgeworth expansion, and the bootstrap will not necessarily perform well when an Edgeworth expansion provides a poor approximation of the distribution of interest.

The bootstrap principle extends easily to a variety of statistical models and there is also a growing number of studies investigating the usefulness of bootstrap methods for small sample inference in cointegrating regression models. The possibility of using the bootstrap for improving the accuracy of the inference in cointegrating regressions was proposed for example by Li and Maddala (1997). In their paper the authors consider situations where the errors driving the cointegrated relationships are autocorrelated and they compare the small sample properties of the standard bootstrap, the moving block and the stationary bootstrap. Their simulation experiments show that bootstrap methods can be successfully employed to reduce the size distortion of the tests statistics. More recently, Psaradakis (2001) suggests using a sieve bootstrap procedure based on resampling residual from an autoregressive approximation to the innovation process driving the cointegrated system.

1.8 Appendix : The Glivenko-Cantelli Theorem

The theoretical justification for the use of the empirical CDF is provided by the Glivenko-Cantelli Theorem. This theorem asserts that $\hat{F}_n$ converges uniformly[11] to the true distribution F

Theorem 2 *As* $n \to \infty$

$$\sup_{-\infty<x<\infty} \left|\hat{F}_n(x) - F(x)\right| \to 0.$$

almost surely .

This theorem is of fundamental importance in probability and the proof can be found in several books. Here we follow Zaman (1996).

Proof. For a fixed x,say $x = x_0$, $\hat{F}_n(x_0)$ is the average of i.i.d Binomial random variable and $P\left(\lim_{n\to\infty} d\left(\hat{F}_n(x_0), F(x_0)\right) = 0\right) = 1$ by the weak low of large numbers. Now consider a partition P such that $x_0, x_1, ..., x_{99}, x_{100}$ so that $F(x_0) = 0$, $F(x_1) = 0.01, F(x_2) = 0.02, ..., F(x_{99}) = 0.99$, $F(x_{100}) = 1$. We allow for $x_0 = -\infty$ and $x_{100} = +\infty$. It follows that $P\left(\lim_{n\to\infty} d\left(\hat{F}_n(x_j), F(x_j)\right) = 0\right) = 1$ for $j = 1, 2, ..., 100$. We

---

[11] It is important to distinguish between uniform convergence and pointwise convergence.
A sequence of functions $(f_n(x))$ is said to converge uniformely to a function $f(x)$ for x belonging to some set A if

$$\forall \varepsilon > 0, \exists N, \forall x \in A, \forall n \geq N, \sup |f_n(x) - f(x)| < \varepsilon$$

A sequence of functions $(f_n(x))$ is said to converge pointwise to a function $f(x)$ for x belonging to some a set A if for each value of x in A $(f_n(x))$, considered as a sequence of real numbers, converges to the real number $f(x)$.

$$\forall \varepsilon > 0, \forall x \in A, \exists N, \forall n \geq N, |f_n(x) - f(x)| < \varepsilon$$

The crucial difference between these definitions relates to the order of quantification. In the second case the value of N can vary with x as well as ε, whereas in the first case N must exist independent of x.

have that $\forall\, x, \exists\, x_j$ such that $x_{i-j} \leq x \leq x_j$, so that using the triangular inequality

$$\begin{aligned} d\left(F(x), \hat{F}_n(x)\right) &\leq \left|F(x) - F(x_j) + F(x_j) - \hat{F}_n(x_j) + \hat{F}_n(x_j) - \hat{F}_n(x)\right| \\ &\leq |F(x) - F(x_j)| + \left|F(x_j) - \hat{F}_n(x_j)\right| + \left|\hat{F}_n(x_j) - \hat{F}_n(x)\right| \end{aligned}$$

The first term is less than or equal to 0.01, the second is convergent to zero, and the third can be bounded as:

$$\left|\widehat{F}_n(x) - \widehat{F}_n(x_j)\right| = \widehat{F}_n(x_j) - \widehat{F}_n(x) \leq \widehat{F}_n(x_j) - \widehat{F}_n(x_{j-1}) \xrightarrow{a.s.} F(x_j) - F(x_{j-1}) = 0.01.$$

Thus with probability 1 as $n \to \infty$, $\left|\widehat{F}_n(x) - F(x)\right| < 0.03$. However, this remains true for all positive real numbers, however large. We thus achieve uniform convergence with probability one. ■

To see why the assumption on the form of convergence (i.e. uniform or pointwise) is important consider estimating

$$F(a) = P(X \leq a)$$

for some fixed a, then the CDF

$$\hat{F}_n(a) = \frac{Number\ of\ X_i \leq a}{n}.$$

The number Y of X's $\leq a$ has the binomial distribution

$$Y : b(p, n) \quad \text{with } p = F(a),$$

with

$$E\left(\frac{Y}{n}\right) = p \text{ and } Var\left(\frac{Y}{n}\right) = \frac{np}{n} \quad (q = 1 - p).$$

It follows that $\hat{F}_n(a)$ is a consistent estimator of $F(a)$ for each fixed a, since

$$E\left(\hat{F}_n(a)\right) = F(a),$$

and

$$Var\left(\hat{F}_n(a)\right) = \frac{1}{n}F(a)\left[1 - F(a)\right].$$

In addition, it follows by the De Moivre's theorem that

$$\sqrt{n}\left[\hat{F}_n(a) - F(a)\right] \xrightarrow{d} N\left(0, F(a)\left[1 - F(a)\right]\right).$$

However, the Glivenko-Cantelli theorem implies a much stronger consistency property. Indeed, it asserts that if the difference between $\hat{F}_n(x)$ and $F(x)$ is considered not only for a fixed x but simultaneously for all x, then the supremum of the distance between $\hat{F}_n(x)$ and $F(x)$ tends in probability to zero as $n \to \infty$.

Chapter 2
Small Sample Corrections for Tests of Hypotheses on the Cointegrating Vectors

2.1 Introduction

The first procedure for testing cointegrating relationships was proposed by Engle and Granger (1987). After their seminal paper cointegration became an extremely intensive field of research, and in the literature many alternatives to their procedure have been developed.

Among them the Johansen (1988, 1995) and Johansen and Juselius (1990) procedure for estimation and testing of cointegrating relationships is widely used in applied econometric research. This method applies the maximum likelihood procedure to a multivariate vector autoregressive model written in the error correction form. Maximizing the Gaussian likelihood function leads via reduced rank regression to the analysis of eigenvalues and eigenvectors. To test for linear restrictions on the cointegrating vectors and their weights Johansen (1988) and Johansen and Juselius (1990) proposed likelihood ratio and Wald tests. However, the asymptotic χ^2 distributed tests are quite heavily affected by the sample size. The problem of the poor approximation of the asymptotic distribution to the finite sample distribution may be described as one of lacking coherence between the statistic and its reference distribution.

In principle there are two ways of solving this problem, either for a given reference distribution correct the test statistic so that the finite sample distribution is closer to the asymptotic distribution, or for a given test statistic correct the reference distribution. Podivinsky (1992) and Psaradakis (1994) followed the first route and they proposed an alternative approximate F-type test, and a small sample adjustment for LR criterion and the $Wald$ test, respectively. The second route involves replacing the critical values of the limit distribution with transformations of critical values obtained from the Edgeworth expansions of the distribution function. Unfortunately, this approach is analytically rather demanding. In this sense estimating critical values using simulated-based method is a plausible numerical alternative.

The purpose of this chapter is twofold. Firstly, we use bootstrap hypothesis testing as a way to reduce the size distortion of the tests for linear restrictions on the cointegrating space. Secondly, we consider the Johansen LR and Wald test statistics as well as the small sample corrected version of these tests, and we explore the robustness of the inference procedure in a situation where we allow for potential over-fitting and under-fitting of the number of cointegrating vectors included in the restricted model.

The outline of this chapter is the following. Section 2.2 briefly introduces the Johansen maximum likelihood estimation and, in particular the likelihood ratio and Wald tests for linear restrictions as well as Podivinsky's (1992) F-type test and Psaradakis's (1994) corrected LR and Wald tests. Section 2.3 describes the bootstrap test. Finally, in Section 2.4 describes the Monte Carlo experimental design and some simulation results are reported.

2.2 Johansen's Maximum Likelihood Procedure

Johansen considers a general vector autoregression in error correction form,

$$\Delta Y_t = \mu + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{k-1} \Delta Y_{t-k+1} + \Pi Y_{t-k} + \epsilon_t, \tag{2.2.1}$$

where Y_t , and ϵ_t are $(p \times 1)$ vectors, and Γ_1 through Γ_k are $(p \times p)$ matrices of coefficients. $\Delta Y_t = Y_t - Y_{t-1}$. $\epsilon_t \backsim NID(0, \Sigma)$.We specialise to the case $k = 1$, so

$$\Delta Y_t = \mu + \Pi Y_{t-1} + \epsilon_t, \tag{2.2.2}$$

The matrix Π determines whether or not, and to what extent, the system (2.2.2) is cointegrated.

We assume first that the eigenvalues of $I + \Pi$ lie on or inside the unit circle. Suppose that Π has rank r. If $r = 0$, and thus Π is a null matrix, Y_t is a vector of random walks related only through the covariances of their innovations ϵ_t . If $r = p$, Y_t is stationary. If $0 < r < p$ (2.2.2) can be interpreted as an error correction model. The hypothesis of r cointegrating vectors β can be written as:

$$H_0 : \Pi = \alpha\beta',$$

where α and β are $(p \times r)$ matrices. The rows of β' can be interpreted as the distinct cointegrating vectors of Y_t (i.e. such that the linear combinations $\beta' Y_t$ are $I(0)$) and the elements of α represent the weights of each of these r cointegrating relations in the p component equations (2.2.2).

Johansen (1988) shows that maximising the likelihood function involves solving the eigenvalue problem

$$\left|\lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k}\right| = 0,$$

to give p ordered eigenvalues $\hat{\lambda}_1 > \ldots > \hat{\lambda}_p > 0$ and corresponding eigenvectors $\hat{V} = [\hat{v}_1 \ldots \hat{v}_p]$ normalised such that $\hat{V}'S_{kk}\hat{V} = I$. The matrices $S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R'_{jt}$, $i,j = 0,k$, where R_{0t} and R_{kt} , are the residuals obtained by regressing ΔY_t and Y_{t-k} on, in general, $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}$, D_t and 1. In our case the S_{ij} are just mean adjusted moment matrices. A basis for the space spanned by the cointegrating vectors is estimated by $\hat{\beta} = [\hat{v}_1 \ldots \hat{v}_r]$. The corresponding estimate of α is given by $\hat{\alpha}\left(\hat{\beta}\right) = S_{0k}\hat{\beta}$.

A test for the number r of cointegrating vectors can be based on the p eigenvalues $\hat{\lambda}_1 > \ldots > \hat{\lambda}_p > 0$. Johansen (1988) derives a likelihood ratio(LR) test of the hypothesis that there are at most r cointegrating vectors by testing that the $(p-r)$ smallest eigenvalues $\lambda_{r+1}, \ldots, \lambda_p$ are zero against the assumption that $\lambda_i \geq 0$ for $i = 1, \ldots, p$. The LR test statistic for this is known as the trace test, defined as

$$LR(trace)_r = -T\sum_{i=r+1}^{p} \ln\left(1 - \hat{\lambda}_i\right).$$

In addition, the maximum eigenvalue test statistic is given by

$$LR(\max)_r = -T\ln\left(1 - \hat{\lambda}_{r+1}\right),$$

and can be used to test the null $H_0(r) : rank(\Pi) = r$ against the alternative $H_1(r+1) : rank(\Pi) = r+1$.

Johansen (1988) shows that the asymptotic distribution of the LR_r trace test is

$$tr\left(\int_0^1 dBB'\left[\int_0^1 BB'du\right]^{-1}\int_0^1 BdB'\right),$$

where $B(u)$ is an $(p-r)$-dimensional Brownian motion with covariance matrix I. He tabulates simulated values of selected percentiles of this asymptotic distribution for a range of values of $(p-1) = 1,2,3,4,5$. These tabulated values serve for testing $r = 0, r \leq 1, \ldots, r \leq (p-1)$ when p ranges from 2 to 5.

The value of r chosen using the LR tests determines the matrices $\hat{\alpha}$ and $\hat{\beta}$: both are $(p \times r)$. It is then possible to test linear restrictions upon the elements of α and β .

Now we can briefly outline the proposed tests for linear restrictions on the cointegrating vectors. Under the hypothesis $H_0 : \Pi = \alpha\beta'$, the maximised value of the concentrated likelihood function satisfies

$$\hat{L}^{-2/T} = |S_{00}|\prod_{i=1}^{r}\left(1-\hat{\lambda}_i\right),$$

where S_{00} and $\hat{\lambda}_i$ were defined earlier. Johansen and Juselius (1990) use this to develop LR tests of linear restrictions on the matrices α and β . Here we will consider only the case $\beta = H\varphi$.

To understand how this test is derived, recall that only the ranges of the columns of α and β are identified. If we set $\alpha^* = \alpha B'$ and $\beta^* = \beta B^{-1}$ than $\alpha^*\beta^{*\prime} = \alpha\beta' = \Pi$. Therefore, α and β are identified only up to a non-singular transformation $B(r \times r)$. Now, what enters the model is $\beta' y_{t-k}$, r linear combinations of the p elements in Y_{t-k} . Restricting

$$\underset{(p\times r)}{\beta} = \underset{(p\times s)}{H}\ \underset{(s\times r)}{\varphi}$$

implies that $\beta' y_{t-k} = \varphi' H' y_{t-k}$ and if Y_k is a matrix whose t^{th} row is y'_{t-k} , the column space of $Y_k H$ is now s dimensional.

The maximised value of the concentrated likelihood function subject to the restriction is

$$\tilde{L}^{-2/T} = |S_{00}| \, \Pi_{i=1}^{r} \left(1 - \tilde{\lambda}_i\right)$$

where $\tilde{\lambda}_1 > \ldots > \tilde{\lambda}_s$ are the $s > r$ eigenvalues obtained from solving

$$\left|\lambda H' S_{kk} H - H' S_{k0} S_{00}^{-1} S_{0k} H\right| = 0.$$

The LR test of $\beta = H\varphi$ can be obtained from the concentrated likelihood functions above, and is

$$LR(\beta) = -2 \ln\left(\tilde{L}/\hat{L}\right) = T \sum_{i=1}^{r} \ln\left[\left(1 - \tilde{\lambda}_i\right) / \left(1 - \widehat{\lambda}_i\right)\right].$$

Johansen (1995), p. 104-5 and 192-3 shows that $LR(\beta)$ is asymptotically $\chi^2(r(p-s))$ under $H_0 : \beta = H\varphi$.

Alternatively, Johansen and Juselius (1990) propose a Wald test. Consider the following null hypothesis $H_0 : K'\beta = 0$ where K is an $(p \times (p-s))$ matrix of full rank, then the W statistic for testing H_0 is:

$$W(\beta) = Ttr\left(\left[K'\hat{\beta}\left(\hat{\Lambda}^{-1} - I_r\right)^{-1}\hat{\beta}' K\right]\left[K'\hat{V}_* \hat{V}_*' K\right]^{-1}\right) \tag{2.2.3}$$

where $\hat{\Lambda} = diag\left(\hat{\lambda}_1, \ldots, \hat{\lambda}_r\right)$ and $\hat{V}_* = [\hat{v}_{r+1}, \ldots, \hat{v}_p]$. Since the limiting distribution of $\hat{\beta}$ is a Gaussian mixture, $W(\beta)$ is asymptotically distributed as $\chi^2(r(p-s))$ under the hypothesis $K'\beta = 0$ respectively.

It may help to relate the two forms of the restrictions. Given

$$\underset{(p\times r)}{\beta} = \underset{(p\times s)}{H} \; \underset{(s\times r)}{\varphi} ,$$

we order the rows of β so that

$$H = \begin{matrix} s \\ (p-s) \end{matrix} \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$$

has H_1 of full rank. So partitioning conformably

$$\begin{matrix} s \\ (p-s) \end{matrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = H\varphi = \begin{bmatrix} H_1\varphi \\ H_2\varphi \end{bmatrix}.$$

Then

$$\beta_1 = H_1\varphi \Rightarrow \varphi = H_1^{-1}\beta_1.$$

Substituting in $\beta_2 = H_2\varphi$

$$\beta_2 = H_2H_1^{-1}\beta_1.$$

Hence $\beta = H\varphi$ implies

$$\begin{bmatrix} \underset{((p-s)\times s)}{-H_2} & \underset{(s\times s)}{H_1^{-1}} & I_{(p-s)} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underset{((p-s)\times r)}{0}$$

This is one way of obtaining

$$K'_{((p-s)\times r)}\beta_{(p\times r)} = 0,$$

i.e. $(p-s)$ common linear restriction on the columns of β.

Similarly, given an arbitrary K, and $K'\beta = 0$ we can write

$$\begin{bmatrix} \underset{((p-s)\times s)}{K_1} & \underset{((p-s)\times(p-s))}{K_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = 0,$$

hence, if we order the rows of β so that K_2 is invertible then

$$K_1\beta_1 + K_2\beta_2 = 0 \Rightarrow -K_1\beta_1 = K_2\beta_2 \Rightarrow \beta_2 = -K_2^{-1}K_1\beta_1.$$

Thus

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} I \\ \underset{((p-s)\times s)}{-K_2^{-1}K_1} \end{bmatrix} \underset{(s\times r)}{\beta_1} = \begin{bmatrix} I \\ -K_2^{-1}K_1 \end{bmatrix} \underset{(s\times s)}{H_1H_1^{-1}\beta_1}$$

$$= \begin{bmatrix} H_1 \\ -K_2 K_1^{-1} H_1 \end{bmatrix} \varphi = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \varphi = H\varphi.$$

Thus to move from $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$ to K', we have to set

$$K' = \left[-K_2 H_2 H_1^{-1} : K_2\right],$$

where K_2 is an arbitrary non singular $(p-s) \times (p-s)$. In the same way, to move from $K' = \begin{bmatrix} K_1 & K_2 \end{bmatrix}$ to H_1 we have to set $H = \begin{bmatrix} H_1 \\ -K_2^{-1} K_1 H_1 \end{bmatrix}$, where H_1 is an arbitrary non singular $(s \times s)$ matrix.

2.2.1 Podivinsky and Psaradakis corrections to the tests for linear restrictions: "A variation on a theme"

The Johansen (1988) simulated critical values are based on asymptotic results, and may not be appropriate when used with relatively small sample sizes. In the literature a lot of work has been done on the procedure for inference in cointegrated systems. Among others Podivinsky (1992) and Psaradakis (1994) investigated the adequacy of these asymptotic critical values in moderately sized samples.

They consider a simple DGP with limited number of lags, and just one cointegrating vector. Their simulation analysis indicate that the asymptotic χ^2 distributed LR tests are quite heavily affected by the size of the sample. Accordingly, they proposed small sample adjustments respectively for an F-type test and for the LR criterion and the Wald test.

First, consider the Podivinsky (1992) approximate F-type test. If again we denote estimation under the null by tilde, and unrestricted estimation by a circumflex, and

$$\widehat{S} = \Pi_{i=1}^{r} \left(1 - \widehat{\lambda}_i\right),$$

$$\widetilde{S} = \Pi_{i=1}^{r} \left(1 - \widetilde{\lambda}_i\right),$$

then the F-type statistics for testing the linear restriction hypothesis $\beta = H\varphi$ is

$$F(\beta) = \frac{\left(\tilde{S} - \widehat{S}\right) / (r(p-s))}{\widehat{S} / (T-l)}$$

where l is the number of parameters estimated subject to the maintained hypothesis $\Pi = \alpha\beta'$. In our case $l = 2pr - r^2 + p$, when estimating α, β, and μ. Then $F(\beta)$ is approximately distributed as $F(r(p-s), T-l)$.

Psaradakis (1993) proposes the application of LR and Wald tests adjusted by certain correction factors. Letting

$$C(\beta) = (l/p) + (1/2)[p - r(p-s)/p + 1],$$

the modified statistics are defined as

$$\begin{aligned} LR_c(\beta) &= LR(\beta)[T - (l/p)]/T, \\ LR_a(\beta) &= LR(\beta)[T - C(\beta)]/T, \\ W_c(\beta) &= W(\beta)[T - (l/p)]/T, \end{aligned}$$

where the $LR_c(\beta)$ and $W_c(\beta)$ are obtained by replacing T by $T - (l/p)$ in standard likelihood ratio and Wald tests.

Monte Carlo evidence in Podivinsky (1992) indicates that the application of the modified F-type test is worthwhile, since improvement are shown with respect to the size properties of LR tests proposed by Johansen (1988). These results are mainly confirmed by Psaradakis (1994), but in addition he shows that the small-sample behavior of LR statistics may be improved by the use of simple scale corrections as indicated above.

More recent work in the literature points out that the problem of size distortion can be substantial when more complex DGP are considered (i.e. when $r > 1$, and more lags and seasonal dummy are inserted), see for instance Fachin (1997), Gredenhoff and Jacobson (1998). One reason is that for the asymptotic theory to be valid it is necessary that rejection probabilities do not depend on the DGP, which is not usually the case in small samples. As a result, the true and the nominal probabilities that a test rejects a correct H_0 can be very different when the p-value is obtained from the asymptotic distribution of the test statistic. Since the bootstrap distribution is able to mimic possible skewness of the finite sample distribution it may account for deviations of the actual distribution from the χ^2 distribution.

2.3 The bootstrap test

As seen above, the LR and W test proposed by Johansen (1988) and Johansen and Juselius (1990) enable a researcher to test for linear restrictions on β after having accepted cointegration among variables and Podivinsky (1992) and Psaradakis (1994) propose small sample adjustment for these tests and for an $F-$type test. In this section, (i) we investigate the size distortion of these tests in finite sample, (ii) we analyse the robustness of the Johansen inference procedure to misspecification in the number of cointegrating relationships, (iii) we apply the bootstrap method to the LR and Wald tests above (iv) we evaluate the robustness of the bootstrap tests. The evaluation of points (i)-(iv) is via Monte Carlo simulation experiments.

The model estimated is a $VAR(1)$ defined by

$$\Delta y_t = \Pi y_{t-1} + \mu + \epsilon_t, \tag{2.3.1}$$

where y_t, and y_{t-1} are (4×1) vectors, μ is a vector of intercepts and $\epsilon_t \approx i.i.d.N(0, I)$

When testing for linear restrictions on cointegrating vectors, the true DGP is not known. Since the null model, and consequently the DGP is unknown, the estimated DGP is used. In our case the estimated error correction model is

$$\Delta y_t = \tilde{\alpha}\tilde{\beta}' y_{t-1} + \hat{\mu} + \hat{\varepsilon}_t \tag{2.3.2}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are the restricted estimates.

The non-parametric bootstrap[12] involves approximating the finite sample distribution of the $\widehat{LR}, \widehat{W}, \widehat{F}-$type tests by drawing several B bootstrap realizations $\left\{\widehat{LR_i}^*\right\}, \left\{\widehat{W_i}^*\right\}$,or $\left\{\widehat{F_i}^*\right\}$ for $i = 1, 2, ..., B$ bootstrap samples $\left\{\left(\Delta y^*, y^*_{t-1}\right)_i\right\}$. In order to do this we re-sample the residuals $(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_t)$ from (2.3.2). Denote the bootstrap sample $(\varepsilon_1^*, ..., \varepsilon_t^*)$. The bootstrap algorithm can be summarised as follows:

1) Estimate the error correction model given by (2.3.2) and compute $\widehat{LR}$, $\widehat{W}, \widehat{F}-$type as described in Section 2.2.

2) Re-sample the residual from $(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T)$ independently with replacement to obtain a bootstrap sample $(\varepsilon_1^*, ..., \varepsilon_T^*)$. Generate the bootstrap sample $(y_1^*, ..., y_T^*)$ recursively from $y_0 = 0$ and $(\varepsilon_1^*, ..., \varepsilon_t^*)$ using the estimated restricted model

[12] Note the stationary bootstrap is often used in time series analysis in place of the ordinary bootstrap. One reason is that the stationary bootstrap is able to preserve the correlation structure of the residuals. However, in our case we are primarily interested in whether the bootstrap is able to reduce the size distortion with respect to the inference based on first-order asymptotic critical values (Osterwald-Lenum (1992)), so we abstract from the more complicated issue of correcting dynamic specification of the underlying VAR. Moreover, Van Giersbergen (1996) results show that the ordinary bootstrap has better power properties with respect to the stationary bootstrap.

$$\Delta y_t = \tilde{\alpha}\tilde{\beta}' y_{t-1}^* + \hat{\mu} + \varepsilon_t^*$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ denote the restricted estimates under the null hypothesis $\beta_1 = 0$.

3) Compute the bootstrap replication of $\left\{\widehat{LR}^*\right\}$, $\left\{\widehat{W}^*\right\}$,or $\left\{\widehat{F}^*\right\}$, using $(y_1^*, ..., y_t^*)$

4) Repeat steps 2-4 B times. Defining the bootstrap $p-$values function by the quantity

$$p^*\left(\hat{\theta}\right) = B^{-1}\sum_{i=1}^{B} I\left(\theta^* \geq \hat{\theta}\right) \tag{2.3.3}$$

where $i = 1, ...B$, θ is the test statistic considered, and $I(\cdot)$ is the indicator function that equals one if the inequality is satisfied and zero otherwise.

6) Reject the null hypothesis if the selected significance level exceeds $p^*\left(\hat{\theta}\right)$.

As seen before $LR(\beta)$ and $W(\beta)$ are asymptotically pivotal since they asymptotically distributed χ^2. Therefore, the we may expect refinements of order T^{-1}.

2.3.1 Design of the Monte Carlo experiments

In order to keep an high degree of experimental control the DGP used are simple $VAR(1)$ processes with small dimension. We consider three different DGP, the first is given by: $DGP1$:

$$\begin{aligned} \Delta y_{1t} &= \epsilon_{1t}, \\ \Delta y_{2t} &= \epsilon_{2t}, \end{aligned}$$

where $\epsilon_t = \left[\begin{array}{cc} \epsilon'_{1t} & \epsilon'_{2t} \end{array} \right]' \approx i.i.d.N\left(0, \sum\right)$, y_{2t}, y_{1t} are (2×1) vectors and $\sum$ is a (4×4) matrix. The variance-covariance matrix of the disturbances is set to a unit matrix throughout. So, we have four unrelated random walks and $r = 0$.

The second DGP is given by $DGP2$:

$$\begin{aligned} \Delta y_{1t} &= \epsilon_{1t}, \\ \Delta y_{2t} &= \epsilon_{2t}, \\ \Delta y_{3t} &= \epsilon_{3t}, \\ y_{4t} &= \beta_{23} y_{2,t-1} + \beta_{33} y_{3,t-1} + \beta_{43} y_{4,t-1} + \epsilon_{4t}, \end{aligned}$$

with $\beta_{23}, \beta_{33,}\beta_{43} < 1$, and $\epsilon_t = \left[\begin{array}{cccc} \epsilon_{1t} & \epsilon_{2t} & \epsilon_{3t} & \epsilon_{4t} \end{array} \right]' \approx i.i.d.N\left(0, I\right)$. So that we have one cointegrating vector $\left[\begin{array}{cccc} 0 & \beta_{23} & \beta_{33} & \beta_{43} - 1 \end{array} \right]'$.

The third is given by $DGP3$:

$$\begin{aligned} \Delta y_{1t} &= \epsilon_{1t}, \\ \Delta y_{2t} &= \epsilon_{2t}, \\ y_{3t} &= \beta_{22} y_{2t-1} + \beta_{32} y_{3t-1} + \beta_{42} y_{4t-1} + \epsilon_{3t}, \\ y_{4t} &= \beta_{23} y_{2t-1} + \beta_{33} y_{3t-1} + \beta_{43} y_{4t-1} + \epsilon_{4t}, \end{aligned}$$

with $\epsilon_t = \left[\begin{array}{cccc} \epsilon_{1t}, & \epsilon_{2t} & \epsilon_{2t}, & \epsilon_{3t} \end{array} \right]' \approx i.i.d.N\left(0, I\right)$. So that we have two cointegrating vectors.

Two possible situations are investigated:

a) The model is correctly specified:

-DGP is $DGP2$ and in model estimated $r = 1$

-DGP is $DGP3$ and in model estimated $r = 2$

b) The number of cointegrating vectors is over-fitted or under-fitted:

-DGP is $DGP1$ but we are assuming $r = 1$

-DGP is $DGP2$ but we are assuming $r = 2$

-DGP is $DGP3$ but we are assuming $r = 1$

All simulations were carried out on 400MHz Pentium PC using the matrix programming language GAUSS Version 3.2.32. The random numbers were generated by the function $rndns$. For each sample we calculated the six tests considered above in a VAR(1) model with intercept and we generated $B = 400$ bootstrap samples according to the algorithm given in the previous section. Then the bootstrap is evaluated by Monte Carlo, and each Monte Carlo experiment is based on 1,000 replications. Obviously, the level of accuracy of the experiment could be improved using a larger number of bootstrap replications and a larger number Monte Carlo replicates, (a 95% confidence interval around a 5% nominal size is [3.6-6.4] for 1,000 replicates). However, 1,000 replications with B=400, T =800, uses 3.2×10^8 random deviates of the 4×10^9 distinct deviates available from $rndns$. For the non-bootstrapped tests, 100,000 Monte Carlo replications were used. The random number generator was restarted for each T value.

According to Davidson and MacKinnon (1996b), in some situations $B = 400$ is the smallest number of replications that guarantees a reasonable trade off between the gains in power and computational costs. However, increasing the number of bootstrap replications involves increasing computational costs, consequently it is necessary to reduce them to a number that minimizes the loss of power. To explore the sensitivity of the estimated size to the number of bootstrap replications we made a pilot experiment for

$B \in \{100, 200, 400, 600, 800, 1200\}$ (the results are reported in Appendix of this chapter) and this simulation confirms that $B = 400$ is adequate for our purposes.

2.4 Monte Carlo Results

In this section we report the result of the Monte Carlo experiment. We firstly consider the size properties of the test statistics and then we present the results for the power.

2.4.1 The probability of the type I error

In Table 2.1-2.5 we report the results of the Monte Carlo experiment with respect to the sizes of the tests. The notation is the following: T is the sample size, LR is the uncorrected likelihood ratio test; LC_c and LR_a are the likelihood ratio tests adjusted by Psaradakis's (1994) correction factors; W and W_c are respectively the uncorrected and corrected Wald tests; F is the F-type test proposed by Podivinsky (1992). Therefore, from column 2 to column 7 we report the Monte Carlo estimated sizes, and column 8 and 9 report the bootstrap corrected likelihood ratio and the bootstrap Wald tests.

The first thing it is important to note is that the empirical sizes of $BootLR$ are equal to those for $BootF$, the bootstrap corrected F statistic, as the F statistic is a one to one function of the LR statistic. Hence, the columns of $BootF$ have been omitted.

Monte Carlo evidence in Table 2.1 and 2.2 confirms Psaradakis (1993) and the Podivinsky's (1992) results in the case where the number of cointegrating vectors is correctly specified and this is particularly true for the F-type. We find the poorest performance for both the W and W_c versions of the Wald statistic. For the W_c test the actual significance

level is much higher then the 5% nominal level, and as a consequence the true null hypothesis will be rejected too often. A reason for this may be the non-invariance property of the Wald test. The invariance property states that the decision reached by the hypothesis testing procedure should remain unchanged under transformation of the parameters. So, the Wald statistic varies with the parametrisation of the null hypothesis being tested and its numerical value can vary greatly according to the specification of H_0 that is being used. As a result, the finite sample level of the Wald test can be greatly different from the nominal level, and using the asymptotic distribution of the Wald statistic can be misleading. In this sense the bootstrap provides a better approximation to the finite sample distribution than first order asymptotic theory and therefore smaller size distortion.

The overall impression is that when the number of cointegrating relationships is correctly specified the size distortion asymptotically vanishes, but the asymptotic theory is uniformly satisfactory only for $T \geq 150$. For smaller sample sizes the only tests that provide nearly exact α level is Podivinsky's (1992) F-type test, $BootLR$ and $BootW$.

Table 2.1. Sizes for tests of $\beta_{1,1} = 0$ assuming correct cointegrating rank of 1.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.100	0.091	0.083	0.186	0.174	0.061	0.046	0.050
75	0.080	0.075	0.067	0.129	0.123	0.057	0.055	0.050
100	0.071	0.067	0.064	0.104	0.099	0.055	0.050	0.049
150	0.064	0.061	0.059	0.083	0.081	0.054	0.049	0.048
200	0.061	0.059	0.058	0.075	0.072	0.053	0.050	0.048
400	0.054	0.054	0.053	0.060	0.060	0.051	0.044	0.043
800	0.051	0.051	0.050	0.054	0.054	0.050	0.052	0.050

$DGP2 : \beta_{23} = 0.5, \beta_{33} = 0.4, \beta_{43} = 0.1.$

Table 2.2. Sizes for tests of $[\beta_{11}, \beta_{21}] = [0, 0]$ assuming correct cointegrating rank of 2.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.100	0.091	0.082	0.171	0.160	0.060	0.045	0.047
75	0.080	0.075	0.070	0.125	0.118	0.057	0.062	0.061
100	0.074	0.070	0.066	0.105	0.101	0.057	0.055	0.059
150	0.067	0.064	0.062	0.087	0.084	0.056	0.048	0.051
200	0.062	0.603	0.059	0.077	0.075	0.054	0.054	0.059
400	0.057	0.056	0.055	0.064	0.063	0.053	0.049	0.052
800	0.053	0.052	0.052	0.056	0.056	0.051	0.058	0.058

$DGP3 : \beta_{23}, \beta_{33}, \beta_{43}$ as in Table 2.1, $\beta_{22} = 0, \beta_{32} = 0.9, \beta_{42} = 0.1.$

Table 2.3. Probability of rejecting $\beta_{1,1} = 0$ when true but assuming $r = 1$ when $r = 0$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.412	0.394	0.379	0.681	0.673	0.327	0.132	0.191
75	0.405	0.393	0.383	0.674	0.668	0.351	0.124	0.176
100	0.406	0.398	0.390	0.672	0.668	0.367	0.152	0.215
150	0.401	0.395	0.390	0.670	0.667	0.375	0.138	0.172
200	0.399	0.395	0.391	0.669	0.667	0.380	0.138	0.205
400	0.398	0.396	0.394	0.666	0.665	0.388	0.122	0.204
800	0.397	0.396	0.396	0.664	0.664	0.393	0.134	0.211

$DGP1$

Table 2.4. Probability of rejecting $\beta_{1,1} = 0$ when true, but assuming $r = 1$ when $r = 2$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.099	0.082	0.073	0.189	0.165	0.034	0.045	0.044
75	0.079	0.068	0.064	0.130	0.117	0.040	0.061	0.062
100	0.070	0.063	0.059	0.105	0.096	0.042	0.061	0.057
150	0.063	0.058	0.056	0.084	0.079	0.045	0.051	0.050
200	0.060	0.057	0.055	0.076	0.072	0.048	0.052	0.047
400	0.053	0.052	0.051	0.061	0.059	0.048	0.042	0.043
800	0.052	0.051	0.051	0.055	0.054	0.049	0.056	0.054

$DGP3$, as Table 2.2.

Table 2.5. Probability of rejecting $[\beta_{11}, \beta_{21}] = [0, 0]$ when true, but assuming $r = 2$ when $r = 1$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.375	0.337	0.318	0.706	0.691	0.207	0.103	0.179
75	0.357	0.332	0.319	0.689	0.679	0.252	0.108	0.154
100	0.351	0.333	0.323	0.681	0.673	0.274	0.100	0.163
150	0.342	0.330	0.324	0.677	0.671	0.293	0.100	0.158
200	0.338	0.329	0.324	0.673	0.669	0.301	0.090	0.165
400	0.334	0.330	0.328	0.668	0.666	0.317	0.098	0.168
800	0.332	0.330	0.329	0.666	0.665	0.323	0.123	0.162

$DGP2$, as Table 2.1.

Table 2.6. Probability of rejecting $[\beta_{11}, \beta_{21}] = [0, 0]$ when true, but assuming $r = 2$ when $r = 0$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.573	0.532	0.510	0.880	0.872	0.371	0.14	0.203
75	0.564	0.535	0.520	0.878	0.872	0.438	0.128	0.190
100	0.565	0.543	0.532	0.877	0.872	0473	0.152	0.206
150	0.559	0.545	0.538	0.874	0.871	0.500	0.117	0.171
200	0.556	0.546	0.541	0.875	0.873	0.513	0.145	0.201
400	0.553	0.548	0.545	0.875	0.874	0.533	0.145	0.176
800	0.553	0.551	0.549	0.874	0.873	0.543	0.145	0.196

$DGP1$

Tables 2.3-2.6 report the Monte Carlo sizes for the tests considered in situations where the number of cointegration vectors is over-fitted or under-fitted. As we can see from Table 2.4 the size distortion when underfitting is not very different, either for magnitude or direction, from the size distortion when the model is correctly specified. This suggests that the difference between the nominal and the empirical size is more likely to be due to finite sample effects than to misspecification. In fact, in both cases the size of the tests depends on the sample size and on the many parameters of the model, and in both cases this dependence asymptotically vanishes (even though the adjustment is quite slow). Therefore, as long as the asymptotic theory works, the bootstrap works. By contrast, when overfitting the asymptotic theory does not help. Indeed, the size distortion caused by overfitting is so large that it calls into question the use of the tests. A size greater than 0.5 implies that one is more often wrong than right when using the test.

One explanation of the excessive size of the Wald test when overfitting the cointegrating rank is as follows. In the case of a single constraint, $\beta_{11} = 0$, $K' = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$,with $r = 1$ assumed, one can write the Wald test in the form

$$\widehat{W} = \widehat{\beta}_{11}^2 \widehat{\sigma}_1^2 / \sum_{j=2}^{p} \widehat{v}_{1j}^2$$

where

$$V = \begin{bmatrix} \widehat{\beta}_1 & \widehat{v}_2 & \cdots & \widehat{v}_p \end{bmatrix}$$

$\widehat{\beta}_{11}$ is the first element of $\widehat{\beta}_1$and

$$\sigma_1^2 = \widehat{\lambda}_1/(1 - \widehat{\lambda}_1).$$

If $r = 0$, the properties of $\widehat{\beta}_{11}^2$ change: it becomes an $O_p(T^{-1/2})$ estimator of 0, rather than $O_p(T^{-1})$ under $H_0 : K'\beta = 0$. Thus $\widehat{\beta}_{11}$ is more variable, and $\widehat{\beta}_{11}^2$ on average larger, if $r = 0$. While the other two terms in $\widehat{W}$ also change, both being on average smaller when $r = 0$, in simulations it seems that the effect on $\widehat{\beta}_{11}^2$ dominates by an order of magnitude.

Considering the general case of $\widehat{W}$ as defined in (2.2.3) , this intuition suggests that overfitting can be regarded as misclassifying the columns of V. If one assumes that the rank of Π is $r + 1$ when it is r, one erroneously regards $\widehat{v}_{r+1}$ as $\widehat{\beta}_{r+1}$ and includes it in the 'numerator' of $\widehat{W}$ rather than the 'denominator'. As it is $O_p(T^{-1/2})$ rather than $O_p(T^{-1})$, and its 'square' enters $\widehat{W}$, this shifts the distribution of $\widehat{W}$ to the right.

This only explains the behavior of the likelihood ratio indirectly insofar as $\widehat{W}$ and $\widehat{LR}$ are correlated. Turning to the bootstrap tests, when overfitting their size is around 10%, and does not converge to the correct value. Why does bootstrapping fail?

In the correctly specified model $\beta' y_{t-1}$ and $\alpha\beta' y_{t-1}$are stationary. If we overfit, we include in $\widehat{\beta}' y_{t-1}$ linear combinations of y_{t-1} which are not stationary, and when generating Δy_t^* from the resampled residuals $\Delta y_t - \widehat{\alpha}\widehat{\beta}' y_{t-1}$ both the residuals and Δy_t^* will be $I(1)$. Thus bootstrapping fails. The size is not as distorted as the non-bootstrapped tests, but there is no reason to think the power properties will be desirable.

One might try to recover the situation by using the parametric bootstrap. If one does so, the residuals are replaced by independent and identically distributed Normal vectors, and Δy_t^* has the properties implied by the cointegrating rank r assumed and $\beta_1 = 0$. However, the test statistic, $\widehat{W}$ or $\widehat{LR}$, being compared with this bootstrap distribution is

calculated from data with a smaller r, and the equivalent of Table 2.5 for the parametric bootstrap shows sizes from 23% to 63%.

When underfitting, as in Table 2.4, the performance of 5 out of the 6 tests is much better, the exception being the F test.

Table 2.7. Parametric bootstrap: Probability of rejecting $\beta_1 = [\beta_{11}, \beta_{21}]' = 0$ when true, but assuming $r = 2$ when $r = 1$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.392	0.357	0.333	0.728	0.713	0.221	0.226	0.467
75	0.370	0.345	0.335	0.698	0.682	0.265	0.238	0.489
100	0.357	0.335	0.324	0.669	0.664	0.269	0.228	0.507
150	0.313	0.307	0.297	0.671	0.668	0.265	0.232	0.557
200	0.348	0.338	0.332	0.681	0.672	0.300	0.259	0.567
400	0.315	0.310	0.304	0.655	0.653	0.292	0.268	0.586
800	0.338	0.336	0.333	0.675	0.674	0.327	0.304	0.631

$DGP2$ as Table 2.1.

2.4.2 The probability of the type II error

The power is defined as the probability of rejecting the null hypothesis (H_0) when it is false. Therefore, to evaluate the power of the test statistics considered in the previous section, it is necessary to evaluate the behavior of the tests when the null hypothesis being tested in false. In our case, the model tested is given (2.3.1), but the data are generated by simulation from two different models

a) The DGP is given by

$$\Delta y_{1t} = \epsilon_{1t},$$

$$\Delta y_{2t} = \epsilon_{2t},$$

$$\begin{aligned}
\Delta y_{3t} &= \epsilon_{3t}, \\
y_{4t} &= 0.1y_{1,t-1} + 0.5y_{2,t-1} + 0.4y_{3,t-1} + 0.1y_{4,t-1} + \epsilon_{4t},
\end{aligned}$$

with $\epsilon_t = \left[\begin{array}{cccc} \epsilon_{1t} & \epsilon_{2t} & \epsilon_{3t} & \epsilon_{4t} \end{array} \right]' \approx i.i.d.N(0, I)$.

b) The DGP is given by

$$\begin{aligned}
\Delta y_{1t} &= \epsilon_{1t}, \\
\Delta y_{2t} &= \epsilon_{2t}, \\
y_{3t} &= 0.9y_{3t-1} + 0.1y_{4t-1} + \epsilon_{3t}, \\
y_{4t} &= 0.1y_{1,,t-1} + 0.5y_{2,t-1} + 0.4y_{3,t-1} + 0.1y_{4,t-1} + \epsilon_{4t},
\end{aligned}$$

with $\epsilon_t = \left[\begin{array}{cccc} \epsilon_{1t} & \epsilon_{2t} & \epsilon_{3t} & \epsilon_{4t} \end{array} \right]' \approx i.i.d.N(0, I)$. Therefore, H_0 is false because $\beta_{11} \neq 0$ in $DGP1$, and $\left[\begin{array}{cc} \beta_{11} & \beta_{21} \end{array} \right] \neq 0$ in $DGP2$.

In Figure 2.1-2.2 we report the rejection frequencies of the test statistics considered. As expected the Wald test has higher 'power', being the test with the highest size distortion. The rejection frequencies based on the small sample corrected version of Wald and LR tests are quite close to rejection frequencies based on the asymptotic critical values. The bootstrap test performs well for the LR test since the 'power' loss in relatively low. By contrast, the 'power' loss of the bootstrapped Wald is more substantial as the 'power' curve for the $BootW$ is uniformly lower than for the asymptotic Wald tests.

Turning to the overfitted and the underfitted model, we can see in Figure 2.3-2.4 that the power properties of the test statistics are quite different. When underfitting all the test statistics considered present a substantial loss in power, but the power curve of the bootstrap test mimic the ones of the reference asymptotic tests. By contrast, when

overfitting the asymptotic tests have much higher 'power' then the bootstrap tests, but this just reflects their greater inflation in size.

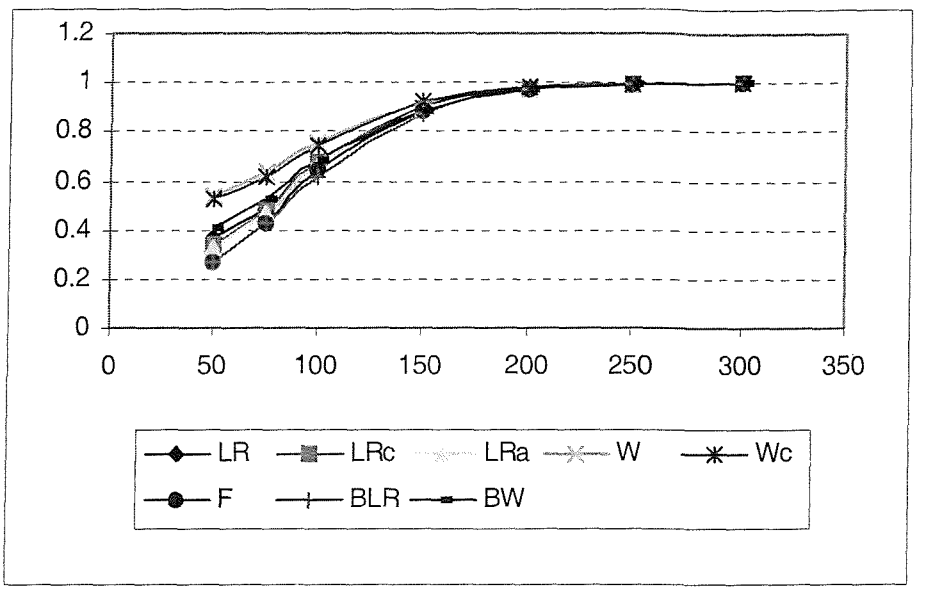

Figure 2.1. Power for tests of $\beta_{1,1} = 0$ assuming correct rank of r =1.

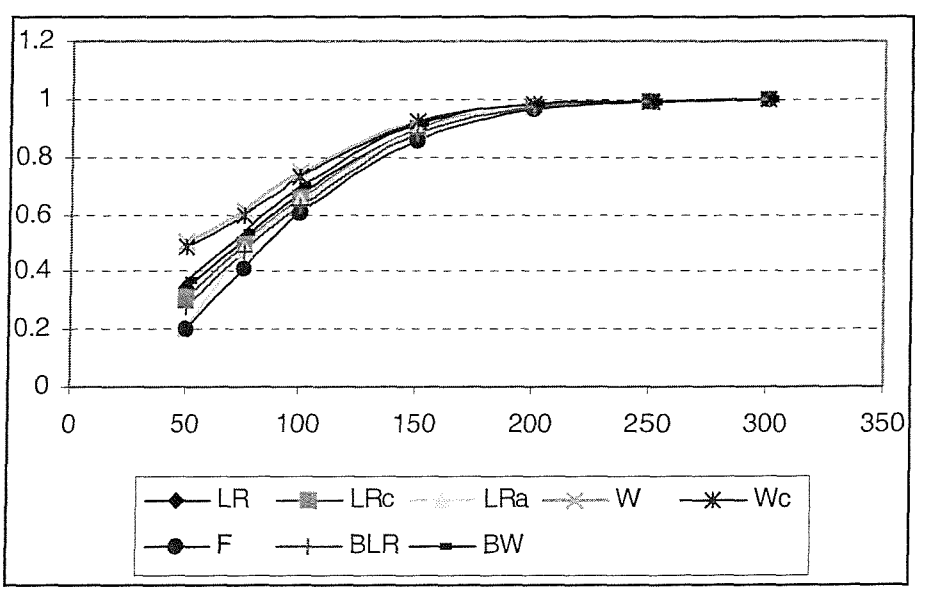

Figure 2.2. Power for tests of $[\beta_{11}, \beta_{21}] = [0, 0]$ assuming correct rank of $r = 2$.

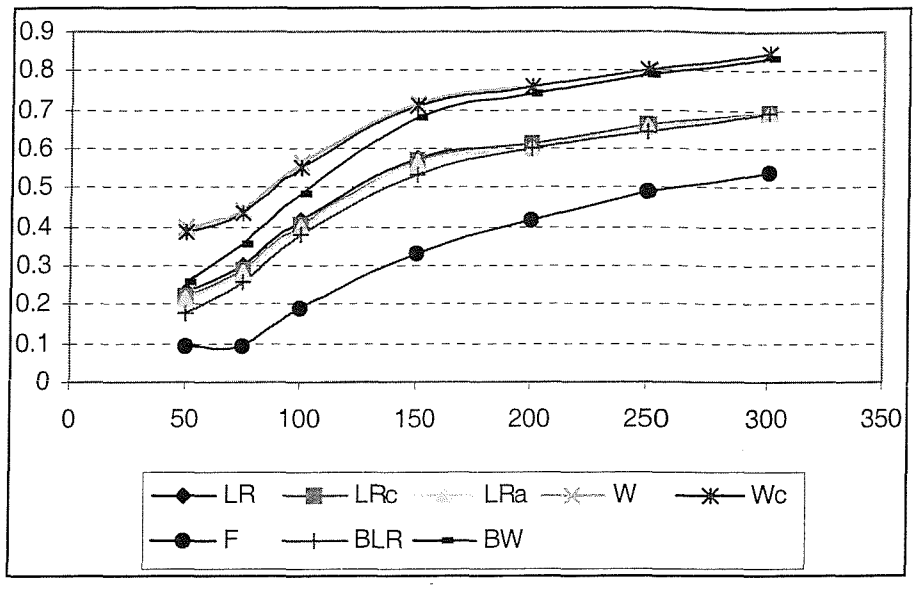

Figure 2.3. Probability of rejecting $\beta_{1,1} = 0$ when false, and assuming $r = 1$ when $r = 2$.

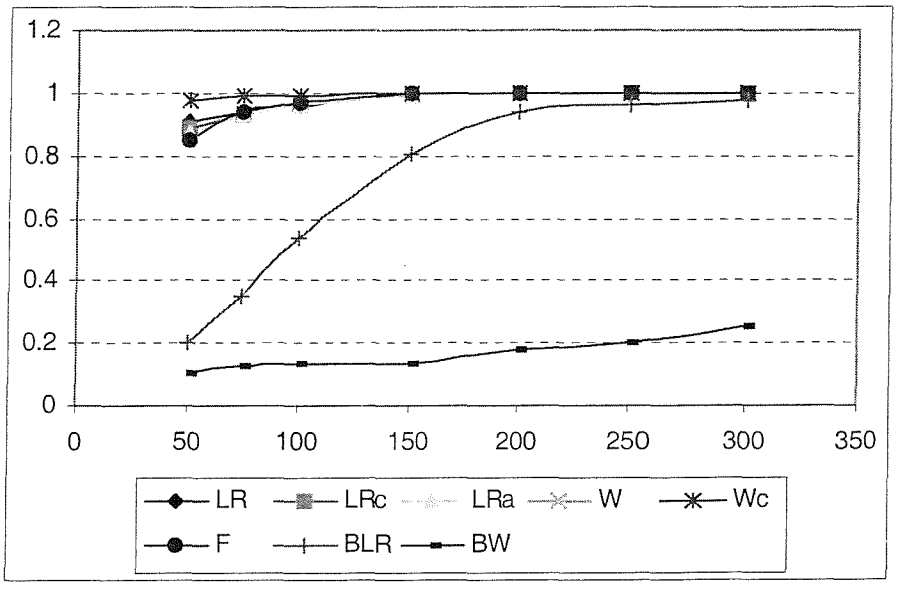

Figure 2.4. Probability of rejecting $[\beta_{11}, \beta_{21}] = [0, 0]$ when false, and assuming $r = 2$ when $r = 1$.

2.5 Concluding remarks

In this chapter we consider the small sample properties of Johansen LR and Wald tests for linear restrictions on cointegrating space, as well as the small sample corrected versions of these tests proposed by Psaradakis (1994) and Podivinsky (1992). In addition, we analyse the sensitivity of the LR, Wald, and F-type to misspecification on the number of the cointegrating vectors, and both the cases of over-fitting and under-fitting have been considered.

The Monte Carlo evaluation of the bootstrap tests show that when the model is correctly specified the resampling procedure provides empirical sizes which are much closer to the nominal size, and this is particular true the Wald test. Furthermore, the bootstrap tests seem to have good power properties. Although our Monte Carlo design is limited, the results suggest that the bootstrap provides a good alternative to procedures relying on first order approximations or small sample corrected tests.

When the number of cointegrated vectors is misspecified the overall picture changes completely. Indeed, our Monte Carlo results show that overspecifing the number of cointegrating vectors leads to considerable size distortion of the tests for linear restrictions on the cointegrating space, whereas underspecifing leads to severe power loss. In this case the bootstrap does not perform well, since in the first case the residuals are not stationary, whereas in the second case the residuals are correlated.

From the practitioner point of view, we may suggest that if there is any uncertainty about the cointegrating rank r, tests on β should be conducted under different assumptions about r. If the conclusions change when r is increased, especially if the bootstrap test

results start to diverge from the those of the asymptotic tests, then only the results for smaller r should be relied upon. This is in contrast to the suggestion in Podivinsky (1998), that "possible overspecification of the number of variables in a model has less serious consequences" (than underspecification): we argue that overestimating cointegrating rank seriously biases tests on β.

2.6 Appendix A: Computation details

The implementation of Johansen's cointegrating tests was not conducted using Johansen original algebra, but using QR and singular value decomposition as employed in O'Brien (1996). In this appendix we show how Johansen cointegration analysis can be rewritten in term of QR decomposition [13].

For ease of notation we report here the model in Section 3

$$\Delta y_t = \Pi y_{t-1} + \mu + \epsilon_t$$

where y_t and is y_{t-1} are (4×1), α_i is a vector of intercepts, and $\boldsymbol{\epsilon}_t \approx N(0, I)$. The $VAR(1)$ model can be rewritten as

$$w_t' = \left[\mu', y_{t-1}', \Delta y_t'\right], \tag{2.6.1}$$

which forms the $t-$th row of the matrix W. Then a QR decomposition of the matrix W yields a Cholesky factorisation[14] R such that such that $R'R = W'W$. We partition

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

where R_{11} has 1 rows and columns, while $R_{22}$and R_{33} are each $(p \times p)$.

Using the Cholesky factorisation we can estimate $\hat{\Pi}$ in a reasonably straightforward way. First note that if $R'R = W'W$, $W = [W_A, W_B]$, and R is conformably partitioned into$\begin{bmatrix} R_{AA} & R_{AB} \\ 0 & R_{BB} \end{bmatrix}$ then,

$$R_{AA}'R_{AA} = W_A'W_A, \tag{2.6.2}$$

$$R_{AA}'R_{AB} = W_A'W_B, \tag{2.6.3}$$

---

[13] For further details see O'Brien (1996). This method is also discussed in Doornik and O'Brien (2002).

[14] If A is a positive definite $(m \times m)$ matrix there exists a lower triangular matrix P such that or $A = P'P$. The decomposition $A = P'P$ is called a Cholesky decomposition.

and

$$R'_{AB}R_{AB} + R'_{BB}R_{BB} = W'_A W_B. \tag{2.6.4}$$

Thus from (2.6.3) solving for R_{AB}

$$R_{AB} = (R_{AA})^{-1} W'_A W_B, \tag{2.6.5}$$

solving (2.6.4) for $R'_{BB}R_{BB}$ and substituting (2.6.5) in (2.6.4) we get

$$\begin{aligned} R'_{BB}R_{BB} &= W'_B W_B - R'_{AB}R_{AB} \qquad (2.6.6) \\ &= W'_B W_B - \left[(R_{AA})^{-1} W'_A W_B\right]' \left[(R_{AA})^{-1} W'_A W_B\right] \\ &= W'_B W_B - W'_B W_A \left(R'_{AA}R_{AA}\right)^{-1} W'_A W_B \\ &= W'_B W_B - W'_B W_A \left(W'_A W_A\right)^{-1} W'_A W_B. \end{aligned}$$

Identifying R_{AA} with R_{11}, and $\begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix}$ with R_{BB}, and conformably partitioning $W = \begin{bmatrix} W_1 & W_2 & W_3 \end{bmatrix}$ so that W_2 and W_3 each have p columns, we can rewrite (2.6.4) as

$$\begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix}' \begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix} = [W_2, W_3]' \left[I - W_1 \left(W'_1 W_1\right)^{-1} W'_1\right] [W_2, W_3]$$

which in Johansen's notation is the product moment matrix

$$T \begin{bmatrix} S_{kk} & S_{k0} \\ S_{0k} & S_{00} \end{bmatrix}.$$

Thus,

$$R'_{22}R_{22} = TS_{kk}, R'_{22}R_{23} = TS_{k0}, R'_{23}R_{23} + R'_{33}R_{33} = TS_{00}. \tag{2.6.7}$$

Using the latent root of $\hat{\lambda}_i$ and the latent vector ε_i of $C^{-1}S_{k0}S_{00}^{-1}S_{0k}\left(C'\right)^{-1}$, where $CC' = S_{kk}$, then defining $E = [e_1 e_2 ... e_r]$ and $S = \begin{bmatrix} I_r \\ O \end{bmatrix}$,we have

$$\hat{\beta} = \left(C'\right)^{-1} E S_r$$

and

$$\alpha\left(\beta\right) = S_{0k}\beta\left(\beta' S_{kk}\beta\right)^{-1}$$

this gives us

$$\left(\hat{\beta}' S_{kk}\hat{\beta}\right) = S_r' E' C^{-1} S_{kk}\left(C'\right)^{-1} E S_r = S_r' E' I_k E S_r = S_r' S_r = I_r$$

and

$$\alpha\left(\hat{\beta}\right) = S_{0k}\hat{\beta}.$$

Identifying $\sqrt{T}C$ with R_{22}' from equation (2.6.7)

$$\begin{aligned} C^{-1}S_{k0}S_{00}^{-1}S_{0k}\left(C'\right)^{-1} &= \left(R_{22}'\right)^{-1} R_{22}' R_{23}\left(R_{23}' R_{23} + R_{33}' R_{33}\right)^{-1} R_{23}' R_{22} R_{22}^{-1} \quad (2.6.8) \\ &= R_{23}\left(R_{23}' R_{23} + R_{33}' R_{33}\right)^{-1} R_{23}' \\ &= I - \left[I + R_{23}\left(R_{33}' R_{33}\right)^{-1} R_{23}'\right]^{-1} \end{aligned}$$

Using the singular value decomposition, let

$$R_{23}R_{33}^{-1} = U\Sigma_R V$$

where $U'U = I_k = V'V$ and Σ_R is diagonal with the singular values σ_i of $R_{23}R_{33}^{-1}$ as its diagonal elements. Thus

$$R_{23}R_{33}^{-1}\left(R_{23}R_{33}^{-1}\right)' = U\Sigma_R^2 U'$$

and

$$R_{23}\left(R_{33}'R_{33}\right)^{-1}R_{23}'(U')^{-1}=U\Sigma_R^2$$

so if u is a column of U, and σ the corresponding diagonal element of Σ_R,

$$R_{23}\left(R_{33}'R_{33}\right)^{-1}R_{23}'u=\sigma^2u$$

so that σ^2, u are the eigenvalues and the eigenvectors of $R_{23}\left(R_{33}'R_{33}\right)^{-1}R_{23}'$ respectively. Thus rearranging equation (2.6.8) we have

$$\left\{I-\left(I+R_{23}\left(R_{33}R_{33}^{-1}\right)^{-1}R_{23}'\right)^{-1}\right\}u=\left\{1-\left(1+\sigma^2\right)^{-1}\right\}u$$

so that u is a latent vector and $\left\{1-\left(1+\sigma^2\right)^{-1}\right\}u=\frac{\sigma^2}{1+\sigma^2}$ a latent root of $R_{23}\left(R_{33}'R_{33}\right)^{-1}R_{23}'$. Thus the Johansen required quantities are $\hat{\lambda}_i=\frac{\sigma_i^2}{1+\sigma_i^2}$, and $\hat{\beta}=(C')^{-1}ES_r=\sqrt{T}R_{22}^{-1}US_r$,with

$$\hat{\alpha}=\alpha\left(\hat{\beta}\right)=S_{0k}=\sqrt{T}\left(R_{22}'R_{23}\right)'\hat{\beta}=1/\sqrt{T}R_{23}'US_r.$$

Moreover, for the LR likelihood test of $H_0:\beta=H\varphi$, where $\beta=H\varphi$ is a set of restrictions, with $H(p\times s)$, we can again use a QR decomposition. First, adapting equation (2.6.1) we have:

$$\underset{(2p\times(s+p))}{\begin{bmatrix}R_{22}H & R_{23}\\ 0 & R_{33}\end{bmatrix}}$$

so that

$$\begin{bmatrix}R_{22}H & R_{23}\\ 0 & R_{33}\end{bmatrix}'\begin{bmatrix}R_{22}H & R_{23}\\ 0 & R_{33}\end{bmatrix}=T\begin{bmatrix}H'S_{kk}H & H'S_{k0}\\ S_{0k}H & S_{00}\end{bmatrix}\tag{2.6.9}$$

then we can perform a QR decomposition of this matrix to produces, $R_\beta=\begin{bmatrix}R_{\beta_{22}} & R_{\beta_{23}}\\ 0 & R_{\beta_{33}}\end{bmatrix}$ where $R_{\beta_{22}}$ is $(s\times s)$, and $R_{\beta_{33}}$ is $(p\times p)$. Then if we replaces R_{22}, R_{23} and R_{33} in our initial analysis with $R_{\beta_{22}}$, $R_{\beta_{23}}$, and $R_{\beta_{33}}$ this will yield $\lambda_i, \hat{\varphi}_0$, $\hat{\beta}=H\hat{\varphi}_0$ and $\widehat{\alpha}$. Tests on α are handled in a similar way.

2.7 Appendix B: Supplementary simulations

In this appendix we report the results of some supplementary simulations. In Table 2.1A-2.2A we reports the p-values for increasing number of bootstrap replications. From these tables we can see that for the bootstrapped likelihood ratio $(BootLR)$, $B = 400$ is only slightly improved on by $B = 800$. The results for the bootstrapped $Wald$ test, $BootW$, also suggest $B = 400$ is a reasonable compromise.

Table 2.1A. Sizes (%) for $BootLR$ tests of $\beta_{11} = 0$ assuming correct cointegrating rank of $r = 1$ and $N = 1000$[15].

T \ B	100	200	400	600	800	1000	1200
50	5.1	4.8	4.7	5.1	5.0	5.2	5.7
75	4.9	6.5*	5.1	5.1	4.7	4.9	6.0
100	5.6	5.2	5.0	6.3	4.7	4.3	5.1
150	5.2	4.7	5.6	4.3	4.7	4.5	5.2
200	4.3	4.3	5.3	4.4	5.0	5.7	5.5

Table 2.2A. Sizes (%) for $BootW$ test of $[\beta_{11}, \beta_{21}] = 0$ assuming correct cointegrating rank of $r = 2$ and $N = 1000$.

T \ B	100	200	400	600	800	1000	1200
50	6.9*	6.0	5.6	5.2	5.6	5.5	6.1
75	5.0	6.3	5.6	5.2	4.9	5.4	6.0
100	5.7	5.2	5.2	6.3	5.1	5.1	4.6
150	4.9	5.3	6.0	4.6	4.8	5.4	5.2
200	4.6	4.6	5.4	4.6	5.1	5.6	5.8

[15] Monte Carlo precision $\pm 1.35\%$; values marked * are significantly different from the nominal size of 5% when testing at a 5% level of significance. Time required, 18.5 hours (400 MHz Pentium).

Chapter 3
Bootstrap-Bartlett Adjustment in Cointegrated VAR Models

3.1 Introduction

In the previous chapter we have considered the small sample properties of Johansen LR and Wald tests for linear restrictions on cointegrating space. Our Monte Carlo experiments revealed that the accuracy of Johansen tests is heavily affected by the sample size and also that the small-sample corrected versions of these tests are quite useful in reducing the size distortion problem. Another simple technique to obtain more accurate small-sample inference for the LR tests was already introduced in 1937 by Bartlett. The basic idea behind the Bartlett correction is to adjust the test statistics so that its finite sample distribution is closer to its asymptotic distribution. In i.i.d. situation the Bartlett correction has been useful for solving size distortion problem. However, calculating the Bartlett correction usually involves calculating an asymptotic expansion of the expected value of the test statistics. Because of the complicated form of the LR tests this can be rather painful; the bootstrap may save us from tedious calculations.

In this chapter we consider Johansen's likelihood ratio tests for linear restriction on cointegrating space and we propose that the Bartlett adjustment factor be computed using the bootstrap. In the literature this approach was first suggested by Rocke (1989), and

Rayner (1990) showed that the bootstrap-Bartlett correction applied to a seemingly unrelated model regression provides accuracy to order $O(T^{-3/2})$.

The Bartlett correction for LR test yields a test which is asymptotically consistent with an error in the rejection probability of order $O(T^{-3/2})$. Therefore, bootstrapping the Bartlett corrected LR test may amount to a one term Edgeworth expansion of the distribution function of the Bartlett corrected likelihood ratio test. This procedure may yield a level of the error in rejection probability of order $O(T^{-2})$, so considerably smaller than the conventional first order approximation (see Beran (1988)).

One possible drawback of this approach is that the performance of Johansen tests using Bartlett correction crucially depends on the parameters of the model. Johansen's (2000) simulation results show that for some parameter values the correction factor is a useful improvement, whereas there are parameter points close to the boundary of the parameter space where the correction does not work well. If this is the case, the ability of the bootstrap to provide second order asymptotic refinements is obviously affected. For this reason it is important to study the dependence of the finite sample size distortion on the parameters. In order to do that we undertake a response surface analysis.

It is well known that the Bartlett correction factor is designed to bring the actual size of asymptotic tests close to their respective nominal sizes, but it may lead to a loss in power. Therefore, it is important to evaluate the power properties of the tests. In our case, the analysis of the power reveals that the procedures have power.

The plan of the chapter is as follows. In Section 3.2 we introduce the model considered and the Bartlett correction. In section 3.3 the consistency of bootstrap Bartlett correc-

tion is considered. In Section 3.4, we explain how the Monte Carlo experiment has been designed and some simulation results are reported. In Section 3.5 we report the simulation results for the bootstrap Bartlett corrected LR test as well as the result of the response surface analysis.

3.2 Model and definitions

Consider the p-dimensional VAR model

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_t + \Phi d_t + \varepsilon_t. \tag{3.2.1}$$

where ε_t are i.i.d. $N(0, \Omega)$. The initial conditions are fixed, the matrices α and β are $(p \times r)$ and Γ_i for $i = 1, ..., k-1$ are $(p \times p)$, Φ the vector d_t contains deterministic terms and Φ their correspondent coefficients, $\Delta Y_t = Y_t - Y_{t-1}$.

Once the cointegrating rank has been established we can test for linear restrictions on cointegrating space. Let Θ_0 be the parameter space under H_0 and Θ_1 the parameter space under H_1. Let $G_{\theta,T}$ be the Johansen's (1988) likelihood ratio test statistic for H_0 against H_1, given by

$$G_{\theta,T} = -2\log\left(H_0|H_1\right) = T\sum_{i=1}^{r} \ln\left[\left(1-\tilde{\lambda}_i\right) / \left(1-\hat{\lambda}_i\right)\right].$$

where $\hat{\lambda}_i$ and $\tilde{\lambda}_i$ are the eigenvalues found as solutions to the eigenvalue problem implied by the maximum likelihood estimation of the restricted and unrestricted models.

Let $g_{\theta,T}$ and g_∞ be the 95% quantiles of the finite sample distribution and the asymptotic distribution of $G_{\theta,T}$ respectively. Let $E\left(G_{\theta,T}\right)$ and $E_\infty\left(G_{\theta,T}\right)$ be the corresponding

expectations. Johansen and Juselius (1990) show that the likelihood ratio test for cointegration has the correct size asymptotically, for example

$$\sup_{\theta \in \Theta_r} P_{\theta,T}\left(G_{\theta,T} > g_\infty\right) \to 5\% \quad \text{as } T \to \infty. \tag{3.2.2}$$

However, this asymptotic result does not give information about the size of the test for finite samples. Many simulation studies have revealed that this test is oversized.

A simple technique to obtain accurate small sample correction for the likelihood ratio test was introduced by Bartlett (1937). His idea was that instead of looking directly at $G_{\theta,T}$ which as $T \to \infty$ tends to G_∞, we focus on the distribution of $\frac{G_{\theta,T}}{E\left(G_{\theta,T}\right)}$. In other words, he suggested the approximation

$$g_{\theta,T} \approx E\left(G_{\theta,T}\right) \frac{g_\infty}{E\left(G_\infty\right)}.$$

Because of the complicated form of the typical LR test statistics, it is difficult to find the expected value, but it may be easier to find a series expansion such as

$$E\left(G_{\theta,T}\right) = E\left(G_\infty\right) + \frac{R}{T} + O\left(T^{-2}\right), \tag{3.2.3}$$

where R is a known constant. Hence the approximation becomes

$$g_{\theta,T} \approx \left(E\left(G_\infty\right) + \frac{R}{T}\right) \frac{g_\infty}{E\left(G_\infty\right)}.$$

This is called the Bartlett correction. In the i.i.d. situation the Bartlett correction has been useful for solving size problem. Lawley (1956) proved that in the context of i.i.d. variables, the same correction improves not only the mean but also all moments. Unfortunately, no similar theorem has been proved in the case of $I(1)$ variables. However,

results in Johansen (1999, 2000) show that the Bartlett correction can be useful in the cointegration model to solve the size distortion problem.

3.3 Bootstrapping the Bartlett correction

This section provides an heuristic explanation of the consistency of the bootstrap-Bartlett corrected LR test (G_{bB} henceforth).

Let the p-dimensional process Y_t be generated by a Gaussian $VAR(k)$ model. In equilibrium-correction form, we may write this as

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_t + \Phi d_t + \epsilon_t. \tag{3.3.1}$$

where $\epsilon_t \backsim N(0, \Omega)$, Π and Γ_i are $p \times p$ parameter matrices, the vector d_t contains deterministic terms and Φ their correspondent coefficients, $\Delta Y_t = Y_t - Y_{t-1}$. The characteristic polynomial associated with model (3.3.1) is given by

$$\Psi(z) = (1-z)I - \Pi z - \sum_{i=1}^{k-1} \Gamma_i (1-z) z^i,$$

with determinant $|\Psi(z)|$. When the rank of Π equals r (for $r < p$), there exist $p \times r$ matrices α and β such that $\Pi = \alpha\beta'$ so that the process (3.3.1) can be written in the vector correction form given in (3.2.1).

Define $\alpha_\perp$ and $\beta_\perp$ as the $p \times (p-r)$ matrices, such that $\alpha'\alpha_\perp = 0$ and $\beta'\beta_\perp = 0$. If the restriction $\Pi = \alpha\beta'$ holds and $\Psi(z)$ has $p - r$ roots equal to 1 and all the other roots outside the unit circle from Granger's representation theorem (see Appendix A in Chapter

4) the process (3.3.1) has the representation

$$Y_t = C\sum_{i=1}^{t}(\Phi d_i + \epsilon_i) + C(L)\left(\Phi d_t + \epsilon_t\right) + \tilde{Y}_0,$$

where $C(z) = \sum_{i=0}^{\infty} C_i z^i$, $\tilde{Y}_0 = Y_0 + \sum_{i=0}^{\infty} C_i \epsilon_0$, and

$$C = \beta_\perp \left(\alpha_\perp' \Gamma \beta_\perp\right)^{-1} \alpha_\perp'.$$

with $\Gamma = I - \Gamma_1 - ... - \Gamma_{k-1}$. Hence, the process (3.3.1) has a Wold vector moving average representation which contains: (i) an $I(1)$ component given by

$$Y_t^p = C\sum_{i=1}^{t}(\epsilon_i + \Phi d_i) = \beta_\perp \left(\alpha_\perp' \Gamma \beta_\perp\right)^{-1} \left(\alpha_\perp' \Phi \sum_{i=1}^{t} d_t + \sum_{i=1}^{t} \alpha_\perp' \epsilon_t\right),$$

where $\left(\alpha_\perp' \sum_{i=1}^{t} \Phi d_t + \sum_{i=1}^{t} \alpha_\perp' \epsilon_t\right)$ represent the $(p-r)$ common trends along with their coefficients $\beta_\perp \left(\alpha_\perp' \Gamma \beta_\perp\right)^{-1}$, (ii) an $I(0)$ component given by

$$Y_t^s = C(L)\sum_{i=1}^{t}(\Phi d_t + \epsilon_t);$$

and an initial values denoted by $\bar{Y}_0$. Cointegration implies that $\beta' Y_t^p = 0$ (i.e. the cointegrating vectors act as a detrending model) so that the process $\beta' Y_t$ is stationary.

Turning to the non-parametric bootstrap, let $\hat{F}_\epsilon$ denotes the empirical distribution function of the residuals, and $\hat{\alpha}$, $\hat{\beta}$, $\hat{\Gamma}_i$, $\hat{\Phi}$ the maximum likelihood estimators of α, β, Γ, Φ under the null hypothesis. Since the resampling scheme imposes r unit roots and all the other roots lie outside the unit circle, the process generated by the resampling scheme

$$\Delta Y_t^* = \hat{\alpha}\hat{\beta}' Y_{t-1}^* + \sum_{i=1}^{k-1} \hat{\Gamma}_i \Delta Y_t^* + \hat{\Phi} d_t + \epsilon_t^*$$

where $\epsilon_t^* \backsim \hat{F}_\epsilon$, differs from the DGP in that ϵ_t^* replaces ϵ_t and the parameters are estimated. A problem with this bootstrapping scheme is the potential non-stationarity of $\hat{\beta}' Y_{t-1}^*$. If we

take $(\hat{\beta} - \beta) = O_p(T^{-1})$, then

$$(\hat{\beta} - \beta)'Y_t = O_p(T^{-1})O_p(T^{1/2}) = O_p(T^{-1/2})$$

and we make an $O_p(T^{-1/2})$ error in an $O_p(1)$ quantity $\beta'Y_t$. Starting the data construction from the observed Y_0, the error is propagated as we construct ΔY_t and $Y_t, t = 1, .., T$. If ΔY_t^+ is the bootstrap data which would be obtained if $\widehat{\beta}$ were replaced by β, then $\Delta Y_t^+ - \Delta Y_t^*$ involves $(\alpha\beta')^{t-1}(\hat{\beta} - \beta)'Y_0$. Unfortunately, the convergence of this error depends on the moduli of the eigenvalues of $\alpha\beta'$, and not of $(I + \alpha\beta')$. Thus for stationarity of ΔY_t and $\beta'Y_t$ we require real eigenvalues of $\alpha\beta' \in (-2, 0)$. However, from the bootstrapping point of view, values in $(-2, -1]$ may be problematical.

Even if one could show that ΔY_t^* were stationary, the method used by Caner and Hansen (2001), relying on Hansen (1996), would not suffice, as the likelihood ratio test depends on eigenvalues which are functions of the moments of the levels of Y_t^*, not ΔY_t^*. One may need to establish directly the properties of the bootstrapped test statistics. This problem does not seem to have been solved in the literature, and hence the validity of the bootstrap in this situation can be empirically illustrated but not proven under general conditions.

We may consider using the parametric in place of the non-parametric bootstrap. Consider model (3.2.1), for all $p \times p$ matrices L of full rank, the transformation $X = LY$ leaves the statistic invariant, so that equation (3.2.1) becomes

$$\Delta X_t = \tilde{\alpha}\tilde{\beta}' X_{t-1} + \sum_{i=1}^{k-1} \tilde{\Gamma}_i \Delta X_t + \tilde{\Phi} d_t + \tilde{\epsilon}_t$$

where $\tilde{\alpha} = L\alpha$, $\tilde{\beta} = L^{-1}\beta$, $\tilde{\Gamma}_i = L\Gamma_i$, $\tilde{\Phi} = L\Phi$, and $\tilde{\epsilon}_t \backsim N(0, L\Omega L')$. We can now chose $L = \Omega^{-1/2}$ such that $\tilde{\epsilon}_t \backsim N(0, I)$. In this case, the consistency of our resampling procedure follows from the fact that we are resampling from a $N(0, I)$ distribution.

Turning now to the Bartlett correction, the test G_T rejects H_0 at the α level if $|G_T| > g_{T,\alpha}$, where $g_{T,\alpha}$, the exact, finite sample, α level critical value, is the $1-\alpha$ quantile of the distribution of G_T. Since G_T is asymptotically pivotal first-order asymptotic approximation implies that $g_\infty - g_{T,\alpha} = O(T^{-1})$[16], where g_∞ is the asymptotic critical value. Consider

[16] Note: assuming that the conditions under which the Bartlett correction corrects the LR test are satisfied (see Lawley (1956)), under weak regularity conditions, we have

$$g_\infty - g_{T,\alpha} = O(T^{-3/2}).$$

now $E_\theta G_T$, since G_T is a non-negative random variable, its mean may be written as[17]

$$E_\theta G_T = \int_0^\infty \Pr(G_T > g_\infty) dg.$$

In view of the fact that the bootstrap p-value is first order asymptotically correct, the average value $E^*_\theta G_T$ of the empirical bootstrap distribution may be represented by

$$E^*_\theta G_T = \int_0^\infty \Pr(G^*_T > g_\infty) dg.$$

The Mallows distance between the bootstrap distribution of

$$\sqrt{T}\left(E^*_{\hat\theta} G_T - E_{\hat\theta} G_T\right),$$

and the finite-sample distribution

$$\sqrt{T}\left(E_{\hat\theta} G_T - E_\theta G_T\right),$$

converge to zero in probability (see Bickel and Freedman (1981)).

---

[17] We are using here the following lemma. For a non-negative continuous random variable Y with probability density f_Y

$$E[Y] = \int_0^\infty P\{Y > y\}\, dy.$$

To see this note that

$$\int_0^\infty P\{Y > y\}\, dy = \int_0^\infty \int_y^\infty f_Y(x) dx dy,$$

where we have used the fact that

$$P\{Y > y\} = \int_y^\infty f_Y(x)\, dx.$$

Interchanging the order of integration in the preceding equation yields

$$\begin{aligned}\int_0^\infty P\{Y > y\}\, dy &= \int_0^\infty \left(\int_0^x dy\right) f_Y(x) dx \\ &= \int_0^\infty x f_Y(x) dx \\ &= E(Y).\end{aligned}$$

3.3.1 The bootstrap experiment

In the previous section, the asymptotic validity of the bootstrap procedure was investigated. The asymptotic validity, however, is only a prerequisite. In this subsection some Monte Carlo experiments are performed to shed light on the small sample behavior of the bootstrap procedure considered. In particular, we are interested in studying the performance of the simulated Bartlett correction in reducing the error in the rejection probability[18] of the tests considered and compare it with the bootstrap tests. Let g_{bB} and g_b be, respectively, the Bartlett corrected quantiles and the bootstrapped quantiles of $G_{\theta,T}$. The probability to be studied are

$$\begin{aligned} \phi_{bB} &= P_{\hat{\theta},T}\left(G_{\hat{\theta},T} > g_{bB}\right), \\ \phi_b &= P_{\hat{\theta},T}\left(G_{\hat{\theta},T} > g_b\right). \end{aligned}$$

These need to be compared with the probability

$$\phi_a = P_{\hat{\theta},T}\left(G_{\hat{\theta},T} > g_\infty\right),$$

where g_∞ are the quantiles of the asymptotic distribution of the test (G_∞).

The model we consider is a simpler version of equation (3.2.1). This is given by

$$\Delta y_t = \alpha\beta' y_{t-1} + \mu + \epsilon_t, \tag{3.3.1}$$

[18] By error rejection probability or size distortion we mean the difference between the actual probability that the test rejects when the true value of θ lie in Θ_0 and the nominal size.

where y_t, and y_{t-1} are (4×1) vectors, μ is a vector of intercepts and $\epsilon_t \approx i.i.d.N(0, I)$. So that the estimated error correction model is

$$\Delta y_t = \widehat{\alpha}\widehat{\beta}' y_{t-1} + \hat{\mu} + \varepsilon_t \tag{3.3.2}$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ are the restricted estimates.

The non-parametric bootstrap involves approximating the finite sample distribution of $G_{\theta,T}$, by drawing B bootstrap realizations $\left\{G^*_{\hat{\theta}_i,T}\right\}$, for $i = 1, 2, ..., B$ bootstrap samples $\left\{\left(\Delta y^*, y^*_{t-1}\right)_i\right\}$. In order to do this we re-sample the residuals $(\varepsilon_1, ..., \varepsilon_t)$ from (3.3.2). Denote the bootstrap sample $(\varepsilon^*_1, ..., \varepsilon^*_t)$. The algorithm to calculate the bootstrap test can be summarised as follows:

1) Estimate the error correction model given by (3.3.2) and compute $\left\{G_{i,\hat{\theta},T}\right\}$.

2) Re-sample the residual from $(\varepsilon_1, ..., \varepsilon_T)$ independently with replacement to obtain a bootstrap sample $(\varepsilon^*_1, ..., \varepsilon^*_T)$. Generate the bootstrap sample $(y^*_1, ..., y^*_T)$ recursively from $y_0 = 0$ and $(\varepsilon^*_1, ..., \varepsilon^*_t)$ using the estimated restricted model

$$\Delta y_t = \tilde{\alpha}\tilde{\beta}' y^*_{t-1} + \hat{\mu} + \varepsilon^*_t$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ denote the restricted estimates under the null hypothesis $\beta_1 = 0$.

3) Compute the bootstrap replication of $\left\{G^*_{i,\hat{\theta},T}\right\}$, using $(y^*_1, ..., y^*_t)$.

4) Repeat steps 3-4 B times to get $\left\{G^*_{1,\hat{\theta},T}\right\}, ..., \left\{G^*_{B,\hat{\theta},T}\right\}$.

5) Defining the bootstrap $p-$values function by the quantity

$$\hat{\phi}_b = B^{-1} \sum_{i=1}^{B} I\left(G^*_{i,\hat{\theta},T} \geq \hat{g}_b\right) \tag{3.3.3}$$

where $i = 1, ...B$, and $I(\cdot)$ is the indicator function that equals one if the inequality is satisfied and zero otherwise. Reject the null hypothesis if the selected significance level exceeds $\hat{g}_b$.

To calculate the bootstrap Bartlett corrected likelihood ratio test simply repeat steps 1-4 and average the observed $\left\{G^*_{1,\hat{\theta},T}\right\}, ..., \left\{G^*_{B,\hat{\theta},T}\right\}$ to get an estimate of $E(G_{\theta,T})$. The quantity

$$G^B_{\hat{\theta},T} = \frac{r(p-s)G_{\hat{\theta},T}}{B^{-1}\left(\sum_{i=1}^{B-1} G^*_{i,\hat{\theta},T}\right)}$$

is the Bartlett corrected value of the likelihood ratio statistic. Defining the bootstrap p-value function by the quantity

$$\hat{\phi}_{bB} = T^{-1}\sum_{i=1}^{B} I\left(G^B_{\hat{\theta}_i,T,i} \geq g_{bB}\right), \tag{3.3.4}$$

reject the null hypothesis if the selected significance level exceeds $\hat{g}_{bB}$.

3.3.2 The Monte Carlo design

In order to evaluate the size accuracy and power the bootstrap procedures described above we have undertaken a Monte Carlo experiments. The DGP used are very similar of the ones used in the previous chapter (i.e. a $VAR(1)$ process with just one or two cointegrated vectors), doing this allow us to compare our results with the ones given in the previous chapter. The first DGP is given by

$DGP = DGP1:$

$$\begin{aligned} \Delta y_{1t} &= \varepsilon_{1t}, \\ \Delta y_{2t} &= \varepsilon_{2t}, \end{aligned}$$

$$\begin{aligned} \Delta y_{3t} &= \varepsilon_{3t}, \\ y_{4t} &= \beta_{33} y_{3,t-1} + \beta_{43} y_{4,t-1} + \varepsilon_{4t}, \end{aligned}$$

with where $\beta_{33} = 0.1$, $\beta_{43} = 0.1$, and $\epsilon_t = \left[\begin{array}{cccc} \varepsilon_{1t} & \varepsilon_{2t} & \varepsilon_{3t} & \varepsilon_{4t} \end{array} \right]' \approx i.i.d.N(0, \mathrm{I})$. So that the variance-covariance matrix of the disturbances is set to a unit matrix throughout. The cointegrating vector is $\left[\begin{array}{cccc} 0 & 0 & \beta_{33} & \beta_{43} - 1 \end{array} \right]'$.

The second DGP is given by:

$$\begin{aligned} \Delta y_{1t} &= \varepsilon_{1t}, \\ \Delta y_{2t} &= \varepsilon_{2t}, \\ y_{3t} &= \beta_{22} y_{2t-1} + \beta_{32} y_{3t-1} + \beta_{42} y_{4t-1} + \varepsilon_{3t}, \\ y_{4t} &= \beta_{23} y_{2t-1} + \beta_{33} y_{3t-1} + \beta_{43} y_{4t-1} + \varepsilon_{4t}, \end{aligned}$$

where $\beta_{22} = 0.9, \beta_{32} = 0.1, \beta_{23} = 0.4, \beta_{33} = 0.5, \beta_{43} = 0.1$, and $\epsilon_t = \left[\begin{array}{cccc} \varepsilon_{1t}, & \varepsilon_{2t} & \varepsilon_{3t} & \varepsilon_{4t} \end{array} \right]' \approx i.i.d.N(0, \mathrm{I})$. So that we have two cointegrating vectors.

The simulations for the bootstrap and the bootstrap Bartlett corrected test were carried out using 1,000 replications of B=400 bootstrap replications, while for the non-bootstrapped test we used 100,000 Monte Carlo replications.

3.3.3 The Monte Carlo results

The Monte Carlo results can be summarized as follows. The first thing to note in Table 3.1-3.2 is that inference based on first order asymptotic critical values is again markedly inaccurate. When $T = 50$, the empirical size can be almost four times as large as the nominal size. Although the inference improves when the sample size increases, the size

distortion of the test is still noticeable when $T = 200$. In general, the higher the precision of the test required (i.e. 10% vs 1% nominal size) the higher the probability of falsely rejecting a true null hypotheses. With respect to the asymptotic inference, the bootstrap has better performance. For a nominal level of 10% or 5% the empirical sizes of the bootstrap tests are only marginally different from the nominal level; this remain true regardless the sample size considered.

Turning to the bootstrap Bartlett corrected empirical sizes we can see that again they are much closer to the nominal sizes than the first order asymptotic critical values. However, the ordinary bootstrap test seems to perform slightly better. This is particularly true for a sample size $T \leq 100$. The only exception to this trend is for the nominal significance level of 1% where for G_{bB} the size distortion is smaller than for G_b.

Table 3.1. Sizes for tests of $\beta_{11} = 0$ (DGP=DGP1)[19].

T		$\hat{G}_a$			$\hat{G}_b$			$\hat{G}_{bB}$	
	10%	5%	1%	10%	5%	1%	10%	5%	1%
50	0.216	0.133	0.045	0.106	0.050	0.005	0.136	0.073	0.015
75	0.169	0.099	0.027	0.101	0.047	0.007	0.121	0.063	0.008
100	0.143	0.081	0.020	0.091	0.047	0.005	0.106	0.054	0.009
150	0.132	0.073	0.019	0.103	0.048	0.006	0.107	0.054	0.007
200	0.120	0.064	0.015	0.094	0.052	0.013	0.102	0.055	0.013

Table 3.2. Sizes for tests of $[\beta_{11}, \beta_{21}] = [0, 0]$ (DGP=DGP2).

T		$\hat{G}_a$			$\hat{G}_{\hat{b}}$			$\hat{G}_{bB}$	
	10%	5%	1%	10%	5%	1%	10%	5%	1%
50	0.198	0.103	0.037	0.128	0.067	0.016	0.129	0.068	0.015
75	0.163	0.093	0.025	0.131	0.072	0.015	0.130	0.071	0.015
100	0.142	0.084	0.019	0.112	0.067	0.014	0.117	0.066	0.014
150	0.127	0.075	0.014	0.113	0.064	0.012	0.118	0.063	0.011
200	0.115	0.059	0.015	0.093	0.054	0.009	0.093	0.053	0.010

[19] Note: $\hat{G}_a$, $\hat{G}_b$, $\hat{G}_{bB}$ are the asymptotic sizes, the empirical sizes of bootstrap test, the empirical sizes of bootstrap-Bartlett adjusted LR test, respectively.

Considering now the power of the test, the evaluation of the power has been has been carried out by generating the data under the following alternatives:

a) $DGP = DGP1$:

$$H_{A,1} : \beta = \begin{bmatrix} 0.1 & 0 & 0.1 & 0.1 \end{bmatrix}',$$
$$H_{A,2} : \beta = \begin{bmatrix} 0.4 & 0 & 0.1 & 0.1 \end{bmatrix}',$$
$$H_{A,3} : \beta = \begin{bmatrix} 0.6 & 0 & 0.1 & 0.1 \end{bmatrix}',$$

b) $DGP = DGP2$:

$$H_{A,1} : \beta = \begin{bmatrix} 0 & 0.9 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.5 & 0.1 \end{bmatrix}',$$
$$H_{A,2} : \beta = \begin{bmatrix} 0 & 0.9 & 0.1 & 0.1 \\ 0.4 & 0.4 & 0.5 & 0.1 \end{bmatrix}',$$
$$H_{A,3} : \beta = \begin{bmatrix} 0 & 0.9 & 0.1 & 0.1 \\ 0.6 & 0.4 & 0.5 & 0.1 \end{bmatrix}'.$$

Tables 3.3-3.4 report Monte Carlo estimated power for the likelihood ratio test, its bootstrap analogue, and the bootstrap Bartlett corrected test. From these tables we can see that, in general, as expected, the power increases with the sample sizes and the distance between the null and the alternative. The power for the larger sample size $T = 200$ is reasonable irrespective of which alternative we use. Note, however, that in the empirical literature sample sizes of $T \geq 150$ are rarely available to practitioners.

Turning to the comparison of the power among the different procedures, in general, we found that the power of G_b, and the power of G_{bB} are almost as good as the asymptotic

power. The only exception being for $T = 50$ under the alternative H_{A1}, which is the worst possible scenario. In this situation the Bartlett corrected test seems to have higher power than the test based on the ordinary bootstrap[20].

The results concerning the power of the bootstrap test are consistent with the theoretical result by Davidson and MacKinnon (1996b). As far as the power of the Bartlett adjusted LR test is concerned, the theory is less conclusive. Indeed, it is well known that the size adjusted test statistics are characterized by a loss in power. In spite of this Cox and Reid (1987) show that the uncorrected statistic and its Bartlett corrected version have the same local power to an order $T^{-1/2}$. However, their result concerns a stationary $AR(1)$ process; the question whether or not it holds in the case of cointegrated processes is still open.

[20] Note that these 'power' comparisons are not size corrected, and perhaps should be more accurately described as 'rejection frequency' comparison.

Table 3.3. Power for G_a, G_b, G_{bB} a nominal level of 5% under different alternative hypotheses $(r = 1)$.

H_A/T		50	75	100	150	200
$H_{A,1} : \beta = 0.1$	G_a	0.363	0.490	0.677	0.899	0.976
	G_b	0.213	0.379	0.617	0.870	0.969
	G_{bB}	0.273	0.420	0.642	0.884	0.966
$H_{A,2} : \beta = 0.4$	G_a	0.924	0.992	1.00	1.00	1.00
	G_b	0.821	0.982	0.999	1.00	1.00
	G_{bB}	0.895	0.991	1.00	1.00	1.00
$H_{A,3} : \beta = 0.6$	G_a	0.984	1.00	1.00	1.00	1.00
	G_b	0.921	0.998	1.00	1.00	1.00
	G_{bB}	0.964	0.999	1.00	1.00	1.00

Table 3.4. Power for G_a, G_b, G_{bB} a nominal level of 5% under different alternative hypotheses $(r = 2)$.

H_A/T		50	75	100	150	200
$H_{A,1} : \beta = 0.1$	$\hat{G}_a$	0.343	0.515	0.673	0.900	0.976
	$\hat{G}_b$	0.273	0.468	0.632	0.885	0.972
	$\hat{G}_{bB}$	0.277	0.466	0.642	0.894	0.973
$H_{A,2} : \beta = 0.4$	$\hat{G}_a$	0.932	0.993	1.00	1.00	1.00
	$\hat{G}_b$	0.897	0.986	1.00	1.00	1.00
	$\hat{G}_{bB}$	0.903	0.989	1.00	1.00	1.00
$H_{A,3} : \beta = 0.6$	$\hat{G}_a$	0.980	0.993	1.00	1.00	1.00
	$\hat{G}_b$	0.967	1.00	1.00	1.00	1.00
	$\hat{G}_{bB}$	0.903	0.989	1.00	1.00	1.00

3.4 Bootstrapping the Bartlett adjusted LR test

In the previous section we have estimated the Bartlett correction using the bootstrap, and we have seen that both the bootstrapped LR test ($\hat{G}_b$) and the bootstrap estimated Bartlett adjusted LR test ($\hat{G}_{bB}$) are able to reduce the size distortion of the test without involving substantial loss in power. Further refinements, however, may be obtained by bootstrapping the Bartlett adjusted likelihood ratio test, the adjustment being estimated using the model parameters. This idea rests on the fact that under regularity conditions the Bartlett correction for LR test yields a test which is asymptotically of correct size with an error in the rejection probability of order $O(T^{-3/2})$. Bootstrapping the Bartlett corrected likelihood ratio test amounts to a one term Edgeworth expansion of the distribution function of the Bartlett corrected likelihood ratio test. This procedure may yield a level of the error in rejection probability of order $O(T^{-2})$, considerably smaller than the conventional first order approximation. The conditions under which the Bartlett correction corrects the LR test for the first and higher moments are given in Lawley (1956). These are some continuity assumptions on the likelihood and its derivatives, together with the assumption that the second derivatives of the likelihood with respect to the parameters are of order T as $T \to \infty$. Unfortunately, no similar theorem has been proved in the case of $I(1)$ variables. However, an analytical calculation of the Bartlett correction for the LR test for linear restrictions on cointegrating vectors is given in Johansen (1999). The correction factor proposed by Johansen depends on the parameters under the null hypothesis, so that in practise the estimated parameters have to be used in order to calculate the Bartlett correction. In our case, for the hypothesis $\beta = H\varphi$ and the $DGP = DGP1$ with $r = 1$ and a constant term

Johansen's correction factor is given by

$$\frac{E\left[-2\log LR|\alpha'_{\perp}\varepsilon\right]}{r\left(p-s\right)} = 1+\frac{1}{T}\frac{3\left(p+1\right)}{2} - \frac{\hat{\alpha}'\hat{\beta}\left[\left(2+\hat{\alpha}'\hat{\beta}\right)p+4\left(1+\hat{\alpha}'\hat{\beta}\right)\right]}{\hat{\beta}'\widehat{\Omega}\hat{\beta}\hat{\alpha}\widehat{\Omega}^{-1}\hat{\alpha}}. \quad (3.4.1)$$

In Table 3.5 we report the Monte Carlo results for the bootstrapped LR test scaled by the factor given in equation (3.4.1)[21]. The empirical sizes for the bootstrapped LR test are labelled as $\hat{G}_{bJB}$.

Table 3.5. Sizes (5%) for G_{JBb} test of $\beta_1 = 0$

T	$\hat{G}_a$	$\hat{G}_b$	$\hat{G}_{bB}$	$\hat{G}_{bJB}$
50	0.129	0.050	0.073	0.047
75	0.097	0.047	0.063	0.045
100	0.076	0.047	0.054	0.041
150	0.066	0.048	0.054	0.045
200	0.069	0.052	0.055	0.045

To facilitate the comparison with the previous procedures we report some of the results given in Table 3.1. Recall that we have labelled $\hat{G}_a$, $\hat{G}_b$, $\hat{G}_{bB}$ the asymptotic sizes, the empirical sizes of bootstrap test, the empirical sizes of bootstrap-Bartlett adjusted LR test, respectively. From Table 3.5 we can see that the inference based on the bootstrapped procedure is quite accurate; no matter the sample size the error in the rejection probability of $\hat{G}_{bJB}$ is only marginal.

3.4.1 The dependence of the size on the parameters: a response surface analysis

Calculating the Johansen's Bartlett correction factor given in (3.4.1) for the hypothesis $\beta = H\varphi$ for the DGP with $r = 1$ gives

[21] The design of the bootstrap experiment follows closely the procedure described in the previous section for the bootstrap test, but in this case we correct the test statistic by the correction factor given in (3.4.1) before applying the resampling procedure.

$$\frac{E\left[-2\log LR|\alpha'_{\perp}\varepsilon\right]}{r\left(p-s\right)}=1+\frac{1}{T}\left[8-\frac{14-9\beta_{43}}{\beta_{33}^{2}+\beta_{43}^{2}}\right],$$

for $\alpha'=\left[\begin{array}{cccc}0 & 0 & 0 & 1\end{array}\right]$ and $\beta'=\left[\begin{array}{cccc}0 & 0 & \beta_{33} & \beta_{43}\end{array}\right]$.

It can be seen that in our case the correction factor depends only on the parameters and T. In order to evaluate the sensitivity of the empirical sizes to the parameter values of the DGP we have undertaken a response surface analysis. In doing that we are able to analyse not only the finite sample properties of the correction factor, but also the effects of the parameters on the distribution of the likelihood ratio test statistic as the parameters enter into the distribution function of the test through the functions $\alpha'\beta$ and $\alpha'\Omega^{-1}\alpha\beta'\Omega\beta$ of equation (3.4.1) (cf. Johansen (2000)). In our case the matrix $\alpha'\beta$ reduces to the scalar $\alpha'\beta=\beta_{43}$ and

$$\alpha'\Omega^{-1}\alpha\beta'\Omega\beta=\beta_{33}^{2}+\beta_{43}^{2},$$

The parameter space in the Monte Carlo experiment has been chosen in order to preserve the stability of the system[22], and is given by

$$\psi\in\Psi=\left\{\begin{array}{c}\beta_{43}\in-1.9,-1.5,-0.9,-0.5,-0.3,-0.2-0.1\\ \beta_{33}\in0.1,0.2,0.3,0.4,0.9,1.5,2\end{array}\right\}, \tag{3.4.2}$$

[22] Calculating the characteristic polynomial we have:

$$\begin{aligned}A\left(z\right) &= I\left(1-z\right)-\alpha\beta'z= \\ &= \left(\begin{array}{cccc}1-z & 0 & 0 & 0\\ 0 & 1-z & 0 & 0\\ 0 & 0 & 1-z & 0\\ 0 & -\beta_{23}z & -\beta_{33}z & 1-\left(1+\beta_{43}\right)z\end{array}\right),\end{aligned}$$

such that $\left|A\left(z\right)\right|=\left(1-z\right)^{3}\left(1-\left(1+\beta_{43}\right)z\right)=0$ if and only if

$$z=\left\{\begin{array}{ll}1 \text{ or } 1/\left(1+\beta_{43}\right), & \text{if } \beta_{43}\neq-1\\ 1, & \text{if } \beta_{43}=-1\end{array}\right\}$$

Therefore, if β_{43} is in the interval $(-2,0]$, then the process Y_t is $I(1)$. (In the case $\beta_{43}=0$, the process is a pure $I(1)$ process which does not cointegrate. For $\beta_{43}<-2$ or $\beta_{43}>0$ the process Y_t is explosive).

and

$$T \in \{50, 75, 100, 150\}.$$

This give us a number of $7 \times 7 \times 4 = 196$ Monte Carlo experiments with $10,000$ replications each.

To summarize the results of the Monte Carlo experiments we use a 3D plot. In figure 3.1A-D we report how the empirical sizes change as a function of the parameter values keeping the sample size fixed to 50, 75, 100, 150, respectively. On the vertical axis we report the empirical sizes corresponding to each pair of (β_{43}, β_{33}). On the horizontal axis the values assumed by β_{43} and β_{33} given in (3.4.2).

From Figure 3.1A-D, it appears that there are points in the parameter space where the χ^2 approximation works relatively well. This is true for example for

$$\psi \in \{-2 \leq \beta_{43} \leq -1.5 \cup 0.5 \leq \beta_{33} \leq 9\}$$

in Figure 3.4A, where $T = 50$. The implication of this is that there are points of the parameter space where the Bartlett correction or the bootstrap test are less needed, since the usual first order approximation gives good results. On the other side, from Figure 3.4A-D it appears that there are points of the parameter space where the size distortion of the test does not vanish, even for $T = 150$ (e.g. $\psi \in \{\beta_{43} = -0.1 \cup \beta_{33} = 0.1\}$). The overall impression is that the size distortion of the test greatly depends on the values of the parameters other than the sample size.

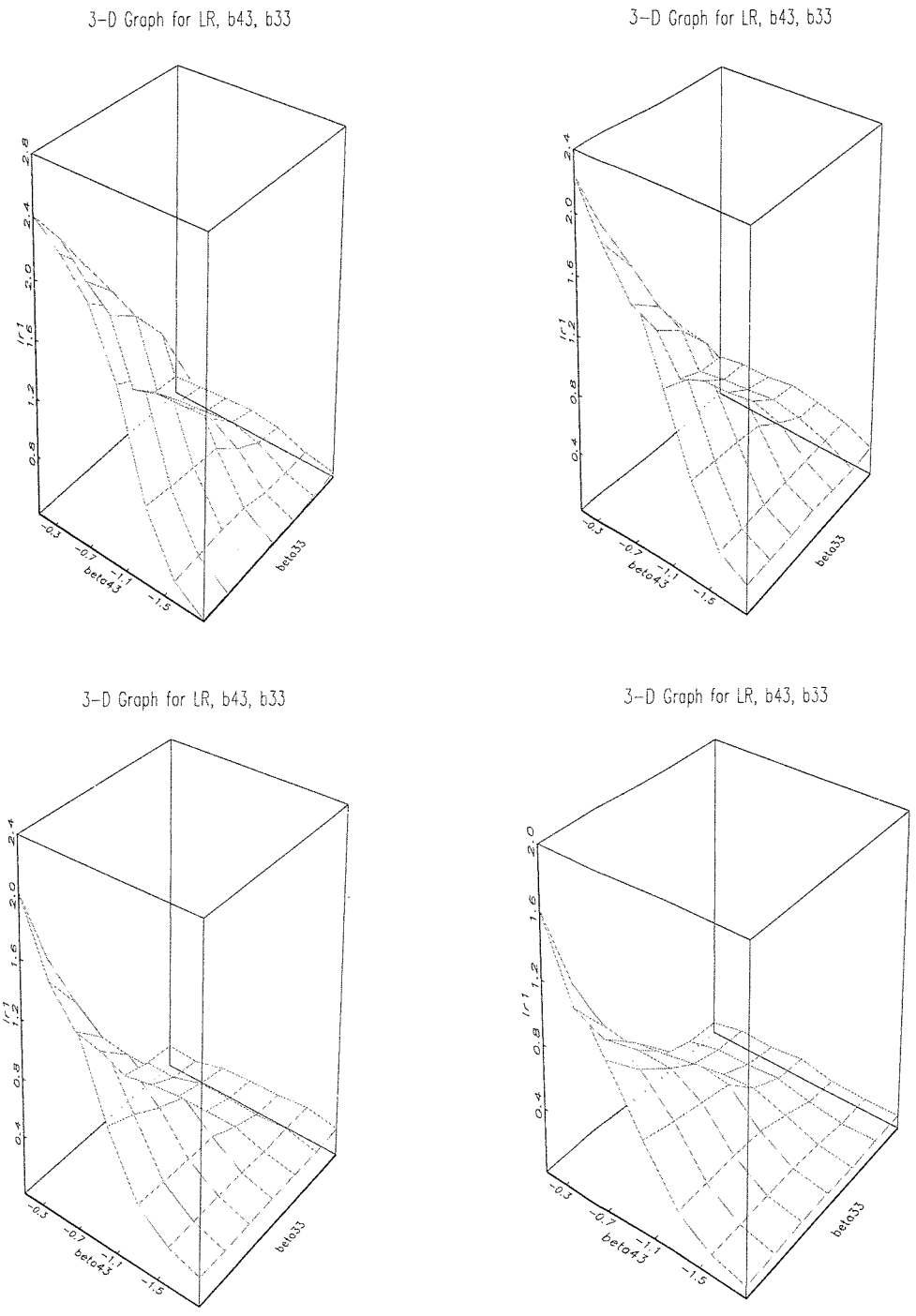

Figure 3.1. Sizes (5%) for G_a for different values of β_{33} and β_{43} . $T = 50, 75, 100, 150$ in 3.4A-D respectively.

The response surface function has been estimated using the logistic transformation [23]

$$\ln\left(\frac{\hat{G}_a}{1-\hat{G}_a}\right) = M(\psi, T) + \varepsilon_{t,}.$$

where $\varepsilon_t \backsim D(0, [1-\widehat{LR}_a]/N)$, and N is the sample size of the Monte Carlo experiment.

After several attempts, we find that the equation that best describes the relations between the parameter values and the empirical size of the likelihood ratio test is

$$\begin{aligned}\hat{G}_a &= \hat{\alpha} + \hat{\gamma}_1 T^{-1} + \hat{\gamma}_2 T^{-1}\beta_{33}^2 + T^{-1}\beta_{43}^2 + \hat{\gamma}_4 T^{-1}(\beta_{33}\beta_{43}) + \\ &+\hat{\gamma}_5 T^{-1}\left(\beta_{43}^2\beta_{33}\right) + \hat{\gamma}_6 T^{-1}\beta_{43}\beta_{33}^2 + \\ &+\hat{\gamma}_7 T^{-1}\left(\beta_{43}\beta_{33}^3\right) + \hat{\gamma}_8 T^{-1}\left(\beta_{43}\left(1-\beta_{43}^2\right)^{-2}\right) + \\ &+\hat{\gamma}_9 T^{-2}\beta_{33} + \varepsilon_t.\end{aligned}$$

In terms of the influence of the parameters and T we can see in Table 3.6 that after the intercept, the most significant coefficient is $\hat{\gamma}_2$, the coefficient of T^{-1}, which is negative.

---

[23] Cox's (1970) linear logistic models of binary data are natural bases for developing response surfaces of estimated finite sample probabilities. Consider a binary response, denoting as M the number of replications in a particular monte Carlo experiment, S the number of "successes" (i.e. the number of replications for which the value of the test lies in the critical region), ϕ (for $0 < \phi < 1$) the finite sample probability of the test lying in the critical region, and $s = S/M$ is the finite sample rejection frequency. Letting

$$\Phi = \frac{S(M-S)}{M-1},$$

$$\tau(\varsigma) = \Phi^{1/2}\log\left[\frac{\varsigma}{1-\varsigma}\right], \text{ for } 0<\varsigma<1,$$

and for $(2M)^{-1} < \varsigma < 1-(2M)^{-1}$:

$$\tau^*(\varsigma) = \Phi^{1/2}\log\left[\frac{\varsigma-(2M)^{-1}}{1-\varsigma-(2M)^{-1}}\right],$$

it can be shown that

$$\tau^*(s) - \tau(\phi) \xrightarrow{D} N(0,1).$$

Table 3.6. Estimated coefficients standard errors, Wald test, two-tailed p-values.

	$Coeff.$	$Std.Er.$	Z	$P > \|Z\|$
$\hat{\gamma}_1$	92.99	3.062	30.36	0.00
$\hat{\gamma}_2$	-205.36	9.936	-20.67	0.00
$\hat{\gamma}_3$	1465.30	169.98	8.61	0.00
$\hat{\gamma}_4$	-2954.99	321.61	-9.19	0.00
$\hat{\gamma}_5$	-80975.3	9574.02	-8.45	0.00
$\hat{\gamma}_6$	-7.132	0.377	-18.89	0.00
$\hat{\gamma}_7$	-120880.1	11245.1	-10.75	0.00
$\hat{\gamma}_8$	2.21	0.51	4.30	0.00
$\hat{\gamma}_9$	8330.44	656.20	12.69	0.00
$\hat{\alpha}$	5.07	0.41	12.37	0.00

$(1/df)$Deviance: 394.7; $(1/df)$ Pearson: 398.3;

3.5 Concluding remarks

In this chapter we consider computer intensive methods for inference on cointegrating vectors in maximum likelihood cointegration analysis. The first part of this chapter focuses on the finite sample behavior of the asymptotic, bootstrap, and bootstrap Bartlett corrected likelihood ratio tests for testing linear restriction on the cointegrating space. The Monte Carlo results show that asymptotic χ^2 based inference can be quite inaccurate in small sample applications. By contrast the bootstrap and the Bartlett corrected LR tests delivers remarkably accurate inference for the restrictions considered. Furthermore, the comparison of the power among different procedure reveals that the power of the bootstrap, and bootstrap Bartlett corrected likelihood is almost as good as the asymptotic power, although in some situations the bootstrap Bartlett corrected LR test seems to have higher power than the bootstrap test.

In the second part of this chapter we propose bootstrapping the Bartlett corrected likelihood ratio test, but in this case the Bartlett correction is calculated analytically using the correction factor proposed by Johansen (1999). According to theoretical arguments proposed by Beran (1988) this procedure may produce an error of rejecting probability of order $O(T^{-2})$. The simulation results reveals that this procedure works remarkably well. However, the response surface analysis reveals that the size distortion of the test heavily depends on the parameter values. There are regions of the parameter space were the usual asymptotic χ^2 approximation works reasonably well, whereas there are parameters points close to the boundary where the distribution of the LR test is very sensitive to the parameter

values. In this case the first order approximation is quite inaccurate, as is the Bartlett corrected LR test.

The general conclusion is that both the bootstrap hypothesis testing and the Bartlett correction the LR test are useful devices for robust inference in the context considered in this chapter, but of course, further theoretical work is needed to confirm the simulation results.

3.6 Appendix : Supplementary simulations

In this appendix we report the results of some supplementary simulations. In Table 3.1A-3.4A we reports the p-values for increasing number of bootstrap replications.

Table 3.1A. Sizes for tests of $\beta_{11} = 0, r = 1$ and $N = 1000$ for ϕ_{bB}.

$T\backslash B$	100	200	400	600	800
50	0.065	0.067	0.073	0.08	0.078
75	0.074	0.059	0.063	0.084	0.073
100	0.05	0.057	0.054	0.058	0.059
150	0.057	0.060	0.054	0.058	0.064
200	0.043	0.055	0.055	0.059	0.051

Table 3.2A. Sizes for tests of $[\beta_{11}, \beta_{21}] = 0$ rank of $r = 2$ and $N = 1000$ for ϕ_{bB}.

$T\backslash B$	100	200	400	600	800
50	0.051	0.048	0.047	0.051	0.050
75	0.049	0.065	0.051	0.051	0.047
100	0.056	0.052	0.050	0.063	0.047
150	0.052	0.047	0.056	0.043	0.047
200	0.043	0.043	0.053	0.044	0.050

Table 3.3A. Sizes for tests of $\beta_1 = 0, r = 1$ and $N = 1000$ for ϕ_b.

$T\backslash B$	100	200	400	600	800
50	0.076	0.079	0.068	0.081	0.083
75	0.076	0.064	0.071	0.077	0.070
100	0.049	0.064	0.066	0.054	0.064
150	0.056	0.065	0.063	0.055	0.054
200	0.049	0.053	0.053	0.054	0.056

Table 3.4A. Sizes for tests of $\beta_1 = 0$ rank of $r = 2$ and $N = 1000$ for ϕ_b.

$T\backslash B$	100	200	400	600	800
50	0.061	0.076	0.067	0.080	0.081
75	0.075	0.059	0.072	0.072	0.071
100	0.047	0.058	0.067	0.057	0.063
150	0.054	0.061	0.064	0.055	0.057
200	0.040	0.052	0.054	0.055	0.055

Chapter 4
Macroeconomic Shocks and Unemployment

4.1 Introduction

In the previous chapters we have considered the small sample performances of the bootstrap tests using mainly simulated data. In this chapter we consider real data instead. The main issues addressed in this chapter are the relationships between macroeconomic shocks and unemployment. In particular, the questions we try to shed light on are: (i) What is the relative importance of shifts in labour supply and labour demand in the rise of unemployment? (ii) What explains the asymmetry of unemployment rate across countries? To investigate these issues we analyse the joint dynamic behavior of three key variables: the profit rate, the real interest rate, and real wages. In the literature there is a wide consensus (see for example Bean (1994)) that the cause of the rise in unemployment in Europe during the 1970s' has been a large adverse shift in the wage-setting relation. Specifically, there was a widely documented slowdown in the rate of the total factor productivity growth in the European countries during the seventies, and a failure of real wages to adjust to a slowdown in productivity growth was one of the causes of the rise in unemployment rate. On the other side, though it would be generally agreed that an increase of the price markup reduces the labour demand and increases unemployment, relatively few authors have investigated this possibility. Blanchard (1997) argues that during the 1980s' shift in the labour supply in the European countries were substituted by shifts in the labour demand. The ex-

planations he gives are: (i) A shift in the distribution of rents from workers to firms due to the fact that firms have steadily increased their markups in goods markets starting from the early 1980s. (ii) A technological bias against labour, that is at given factor prices firms have been adopting technologies that use less labour and more capital, thus decreasing the labour demand.

Blanchard's hypothesis is quite challenging, and particularly interesting for the policy implication that it involves. However, in the recent literature the relationships among capital accumulation, real interest rate and unemployment remains relatively obscure, particularly from the quantitative point of view. With this work we try to fill the gap. Although based on a general economic framework, the contribution of this chapter is more methodological. Our task is to provide an empirically valid description of the interrelations in actual economic data in accordance with the economic theory. We thus impose theoretical restrictions as long as these do not conflict with the empirical evidence.

The model we study is a cointegrated VAR. This model allows us to distinguish between the effects of transitory and permanent shocks to unemployment. Inference about the cointegrating rank is, once again, carried out using the Johansen procedure (1988, 1995). In order to investigate the robustness of our inference we have carried out a simulation study using non-parametric bootstrap.

The structure of the chapter is as follows. In Section 4.2, a simple economic model is presented. In section 4.3 we briefly summarize the econometric model. Section 4.4, reports the empirical results.

4.2 The model

In this section we introduce the model that will be estimated in Section 4.3. The structure of model is described below. As far as the notation is concerned we use the lower case to indicate that the variable is expressed in logarithms and the upper case when the variable is expressed in levels.

Production function, Price-setting relation,and unemployment

Consider a production function characterized by constant return to scale

$$y_t - n_t = \gamma_1 (k_t - n_t) + a_t \tag{4.2.1}$$

with y: output, k: capital, n: labour, and a_t stands for technological progress which follows the stochastic process given by

$$a_t = a_{t-1} + \varepsilon_{t,a},$$

where $\varepsilon_{t,a}$ is a stationary error term. Assuming that firms maximise profit, the price-setting relation is given by

$$w_t = (y_t - n_t) + \mu_t,$$

where w_t = real wage. Hence, the wage is a function of the marginal product of labour and a mark-up (μ_t) [24] on the labour costs. From equation $(4.2.1)$, we can write

$$w_t = \gamma_1 (k_t - n_t) + a_t + \mu_t,$$

[24] The markup is defined as the ratio of the marginal product to the real wage.

and solving for n_t we obtain an expression for the labour demand given by

$$n_t = -\gamma_1^{-1} w_t + k_t + \gamma_1^{-1} a_t + \gamma_1^{-1} \mu_t.$$

In the short run the capital is fixed, so that labour demand is decreasing in the wages, while the ratio of labour to capital is decreasing in the wages and increasing in the mark-up on the labour cost

$$n_t - k_t = -\gamma_1^{-1} \left(w_t - \mu_t + a_t \right).$$

In the long run firms adjust the factor proportions, and capital accumulation depends on the firms' profit.

Defining the profit per unit of capital as

$$\frac{\Pi_t}{K_t} = f\left(\frac{N_t}{K_t}\right) - w_t\left(\frac{N_t}{K_t}\right),$$

we can express the profit rate as

$$\pi_t - k_t = -\eta_2 w_t + \mu_t$$

where μ_t follows a $I(1)$ stochastic process. If $\mu_t = 0$ this is simply the factor price frontier relation implied by the production function (4.2.1). In the long-run, for a given interest rate, the zero profit condition implies that the wages must be such as to generate a profit equal to the user cost of capital, so that

$$\frac{\Pi_t}{K_t}\left(W_t, \mu_t\right) = \rho_t + r_t. \tag{4.2.2}$$

where ρ_t and r_t stand for the depreciation rate and the real interest rate, respectively. Therefore, the long run labour demand is horizontal.

From the short run and the long run labour demand, in order to consider the relationship between unemployment and factor prices, we write a general expression for unemployment given by (4.2.3). Note in (4.2.3) the variable a_t has been omitted since under the assumption of Hicks-neutral production function the ratio of the marginal products remains unchanged for a given capital to labour ratio. Therefore, u_t does not depend on a_t.

Inverting the price-setting relation we can derive an expression for unemployment given by

$$l_t - n_t = \gamma_2 w_t + \eta_1 uc_t + \theta_{t,u}, \tag{4.2.3}$$

where l_t stands for labour force, and

$$\theta_{t,u} = \gamma_5 \theta_{t-1,u} + \varepsilon_{t,u},$$

where $|\gamma_5| \leq 1$, and $\varepsilon_{t,u}$ is a stationary process. Assuming that a linear combination of $(l_t - n_t)$ and w_t is stationary, and that the real interest rate is stationary, we may consider two cases[25]:

$$\left\{ \begin{array}{l} \gamma_5 = 1 \Rightarrow \theta_{t,w} \backsim I(1) \Rightarrow u_t \backsim I(1) \\ \gamma_5 < 1 \Rightarrow \theta_{t,w} \backsim I(0) \Rightarrow u_t \backsim I(0) \end{array} \right\},$$

The stochastic process $\theta_{t,u}$ may reflect changes in the equilibrium level of unemployment caused by an increase in the markup of prices over wages. If $\gamma_5 \rightarrow 0$ then the equilibrium level of unemployment does not change, since it depends only on a pure labour demand shock $(\varepsilon_{t,u})$. For $\gamma_5 \rightarrow 1$ the equilibrium unemployment is not stationary.

---

[25] These hypothesis will be subject to testing in Section 4.

The wage-setting relation

The wage-setting relation or pseudo-labour supply, may be thought as the relation between the wage set in bargaining between firms and workers, and labour market conditions. In the literature (see Layard *et al.* (1991)) this function has been expressed as a relationship between wages, unemployment and labour productivity. In particular,

$$w_t = -\gamma_3 (l_t - n_t) + \gamma_4 (y_t - n_t) + \varepsilon_{t,w}, \tag{4.2.4}$$

where $\varepsilon_{t,w}$ is an $I(0)$ process, and γ_3, γ_4 are the elasticity of the real wage with respect to unemployment and productivity, respectively. In order to concentrate on the relationships between the wages and unemployment we rewrite equation $(4.2.4)$ as

$$w_t = -\gamma_3 [(l_t - n_t)] + \gamma_4 y_t - n_t + \bar{\varpi}_{t,w}, \tag{4.2.5}$$

where

$$\bar{\varpi}_{t,w} = ((y_t - n_t)^e - (y_t - n_t)),$$

and $\varpi_{t,w}$ is assumed $I(0)$.

If $\varpi_{t,w} \backsim I(0)$, $\gamma_4 = 1$, $\mu_t = 0$, and $\gamma_3 > 0$, then we are in the "competitive framework" where the wage is equal to the marginal product of labour (i.e. the wage share is stationary) and unemployment is white noise. In equation $(4.2.5)$, $(y_t - n_t)^e$ indicate the productivity level perceived by workers, rather then the actual one.

The stochastic process $\bar{\varpi}_{t,w}$ may reflect changes in the equilibrium level of unemployment caused by the mismatch between the perceived and the actual productivity growth.

The idea behind this formulation of the wage setting relation is that "hysteresis" in the labour market may be caused by a failure of wages to adjust to a shock to productivity. The assumption that workers have imperfect foresight captures some important aspects of the data. It may be useful to illustrate our hypotheses about non stationary in unemployment with a picture.

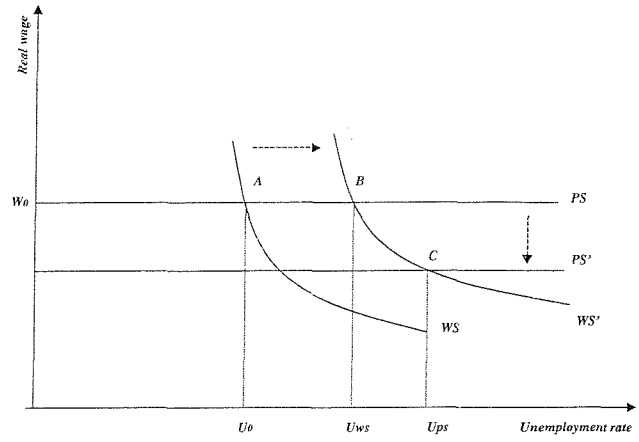

Figure 4.1. The relation between real wage and unemployment.

Figure 4.1 relates the unemployment rate and the real wage. The wage setting relation is denoted as WS , and the price setting relation is denoted as PS. A positive shock to $\bar{\omega}_t$ shifts the wage setting relation from WS to WS' and the equilibrium unemployment moves from A to B, while real wage does not change. Thus a positive shock to $\bar{\omega}_t$ does not show up in higher wages in the long run, since higher unemployment forces wages back to their initial level.

On the other side increases in μ_t shift the PS curve down to PS', moving the equilibrium from A to C, and thus leading also to an increase in the natural rate of unemployment. Whether this movement is permanent or not depends on the properties of the stochastic process in μ_t.

Under the assumptions above there are two common stochastic trends among employment, real wages, and profit rate arising from productivity growth and a markup trend. A third possible trend is derived from equation (4.2.3) . This implies that unemployment can have a permanent component and a serially correlated transitory component. These components may be interpreted as "structural" and "cyclical" unemployment respectively.

One important feature of this model is that the endogenous variables are driven by unit root processes. This property motivate the interpretation of the model in terms of cointegration.

4.3 The Econometric model

In this section we briefly discuss the econometric model we use to estimate the model described in the previous section. We refer the reader to Stock and Watson (1988), Johansen (1995), King *et al.* (1987), and Warne (1993) for more rigorous treatments.

The common trends model involves a linear decomposition of a VAR into stationary and non-stationary parts. Rewriting the VAR in the $VECM$ form and inverting it we find a MA representation which is the sum of the initial values, an $I(1)$ component and an $I(0)$ component. This formulation has its roots in the Beveridge-Nelson (1981) decomposition of univariate time series.

Consider the model

$$\begin{aligned} x_t &= x_0 + A\tau_t + \Phi(L)\, v_t \\ \tau_t &= \mu + \tau_{t-1} + \varphi_t \end{aligned} \tag{4.3.1}$$

where x_0 is an $(n \times 1)$ vector of constants, τ_t is an $(k \times 1)$ vector of random walks with drift μ, and innovations φ_t, L is the lag operator, $\Phi(L)$ is an $(n \times n)$ matrix of lag polynomials, and v_t is an $(n \times 1)$ vector of serially uncorrelated innovations, with mean zero and covariance matrix $\sum$.

The matrix A is called the loading matrix, and gives the impact of the trend τ_t on x_t. By recursive substitution in (4.3.1) the model can be rewritten as

$$\begin{aligned} x_t^s &= x_0 + \Phi(L)\, v_t \\ x_t^p &= A\left[\tau_0 + \mu t + \sum_{j=1}^{t} \varphi_j\right] \\ x_t &= x_t^s + x_t^p \end{aligned} \tag{4.3.2}$$

So the trend component $A\tau_t$ is driven by the impulses to the random walks φ_j and the propagation mechanisms in A. The deviations from the trend are a product of the impulses v_t and the propagation mechanism $\Phi(L)$.

As Stock and Watson (1988) point out cointegration implies the number of trends, k to be less then the number of variables, n. That is, there are exactly $r = n - k$ linearly independent vectors which are orthogonal to the columns of the loading matrix A so that there exists an $n \times r$ matrix β such that the vector

$$\beta' x_t = \beta' x_0 + \beta' A\tau_t + \beta' \Phi(L)\, v_t$$

is stationary (i.e. $\beta' A = 0$).

To see how the common trends model can be estimated let us consider

$$\Gamma(L)\,\Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t. \tag{4.3.3}$$

The $VECM$ can be inverted to yield a Beveridge-Nelson-Stock-Watson representation in term of reduced form disturbances

$$x_t = x_0 + C(1)\xi_t + C^*(L)\,\varepsilon_t$$

where $C^*(L)$ is a stationary moving average representation, $\xi_t = \rho + \xi_{t-1} + \varepsilon_t$, and

$$C^*(L) = \sum_{j=0}^{\infty} C_j^* L^j.$$

Therefore, equation (4.3.3) can be rewritten

$$\begin{aligned} x_t^s &= x_0 + C^*(L)\varepsilon_t \\ x_t^p &= C(1)\left[\xi_0 + \rho t + \sum_{j=1}^{t} \varepsilon_t\right] \end{aligned} \tag{4.3.4}$$

Combining (4.3.2) and (4.3.4) we find that

$$\begin{aligned} C(1)\,\rho &= A\mu \\ C(1)\,\varepsilon_t &= A\psi_t \end{aligned} \tag{4.3.5}$$

and hence (assuming that $E(\psi_t \psi_t') = I$) that

$$C(1)\sum C(1)' = AA'.$$

To estimate the loading matrix A we need to know $C(1)$ and $\sum$. The covariance matrix can be consistently estimated from (4.3.3). However, to obtain an estimate of $C(1)$ we need to invert the $VECM$ representation. To do it, we follow Warne (1993). Define

M as a non-singular matrix given by

$$\underset{(n\times n)}{M} = \begin{bmatrix} \underset{(n\times k)}{S_k'} & \underset{(n\times r)}{\beta} \end{bmatrix}'.$$

Furthermore, $\underset{(n\times n)}{\alpha^*} = \begin{bmatrix} 0 & \alpha \end{bmatrix}$, and the polynomial matrices $D(L)$ and $D_\perp(L)$ defined by

$$D(L) = \begin{bmatrix} I_k & 0 \\ 0 & (1-L)\, I_r \end{bmatrix},$$

and

$$D_\perp(L) = \begin{bmatrix} (1-L)\, I_k & 0 \\ 0 & I_r \end{bmatrix}.$$

Under hypotheses of cointegration there exist a restricted VAR $(RVAR)$ representation of the form

$$B(L)\, y_t = \theta + \theta^* D_t + \eta_t, \tag{4.3.6}$$

where $B(L) = M\left[A^*(L)\, M^{-1} D(L) + \alpha^* L\right]$ and for $y_t = D_\perp(L) M x_t$ the process $\{x_t\}$ is stationary.

The following relationship holds between the $RVAR$ and the unrestricted VAR model:

$$\begin{aligned} \theta &= M\rho, \\ \theta^* &= M\rho^*, \\ \eta_t &= M\varepsilon_t, \\ A(L) &= M^{-1} B(L)\, D_\perp(L) M. \end{aligned}$$

Since $|B(L)| = 0$ has all solutions outside the unit circle and $D_\perp$ has rank r , $A(1)$ has rank r.

To summarize we use Johansen's (1995) procedure to estimate the matrix polynomial M and $D_{\perp}(L)$. From here we can construct the vector $\{y_t\}$. To choose the first k elements of y_t we set $S_k = \beta'_{\perp}$ where $\beta'_{\perp}\beta = 0$. Consistent estimates of the parameters in (4.3.6) can be obtained from Gaussian likelihood estimation of y_t on a constant and p lags.

The next step is to estimate the loading matrix of the common trends parameters, (i.e. the matrix A defined above).

Let us rewrite A as $A_0\pi$, where A_0 is a known and π is a lower triangular matrix of unknown parameters. Then the estimate of A and π can be constructed from the estimates of $C(1)$ and $\sum = E(\varepsilon_t\varepsilon'_t)$. To determine π, recall that,

$$x_t^p = C(1)\sum_{j=1}^{t}\varepsilon_j = A\tau_t.$$

Combining this with the assumption that $E(\varphi_t\varphi'_t)$ (which is a $(k \times k)$ identity matrix), it follows that

$$C(1)\sum C(1)' = A_0\pi\pi' A_0.$$

Therefore, given $C(1)$ and $\sum$, π can be estimated using a Cholesky factor of

$$\pi\pi' = (A_0A_0)^{-1}A_0'C(1)\sum C(1)'A_0(A_0A_0)^{-1} \tag{4.3.7}$$

The right hand side of (4.3.7) is a $(k \times k)$ positive definite and symmetric matrix with $k(k+1)/2$. So, if A_0 is known we can solve $k(k+1)/2$ independent equations in $\pi\pi'$. However, in order to uniquely identify π, in addition to the requirement $\beta' A_0 = 0$, we need $k(k-1)/2$ extra restrictions.

4.3.1 Permanent vs transitory shocks

The usual way of analysing how a VAR system reacts to various impulses is using impulse response functions and variance decomposition. The impulse response function shows the shape of the dynamic response of the variables to an innovation in the permanent component. By contrast, the variance decomposition gives us the relative importance of the response to a typical innovation in determining the short run evolution of the variables.

The moving average representation of (4.3.3) is a natural starting point for impulse response analysis and variance decomposition. This is given by

$$\Delta x_t = \delta + C(L)\varepsilon_t. \tag{4.3.8}$$

We write

$$\varphi_t = \left[\begin{array}{cc} \psi_t' & v_t' \end{array} \right]',$$

and assume that $E\left[\begin{array}{cc} \varphi_t & \varphi_t' \end{array} \right]$ is diagonal. From (4.3.5) we know that

$$\psi_t = (A'A)^{-1} A'C(1)\varepsilon_t. \tag{4.3.9}$$

Defining

$$F_k = (A'A)^{-1} A'C(1)$$

and $F = \left[\begin{array}{cc} F_k' & F_r' \end{array} \right]$, we can express (4.3.9) as

$$\psi_t = F\varepsilon_t,$$

so that the moving average model (4.3.8) becomes

$$\begin{aligned} \Delta x_t &= \delta + C(L)F^{-1}\varphi_t \\ &= \delta + R(L)\varphi_t \end{aligned} \tag{4.3.10}$$

with $R(L) = \sum_{i=0}^{\infty} R_i L^i$. The representation(4.3.10) can be used to calculate the impulse response function and decomposition of variance with respect to the innovation ψ_t and v_t.

4.4 Empirical Evidence

In this section we present the result from the estimation on the models described in the previous section. The empirical evaluation of the model is carried out by considering the following variables

$$x_t = \left[\begin{array}{ccccc} y_t - n_t & w_t & \pi_t - k_t & u_t & uc_t \end{array} \right]'$$

for France, UK, Germany, Italy, US, Japan and Canada.

The data-set consists of quarterly observations from the $OECD$ data-set[26] ((1970-1998) for France, and (1960-1998) for US and Canada, on real GDP (Y_t) , total employment in hours (N_t), gross fixed capital formation (INV_t), capital stock at constant prices (K_COST), and the nominal interest rate on long-run government securities (IR_NOM).

From these series the variables of the estimated models are constructed as follows: the unemployment series, (L_t/N_t), is equal to $\ln[1/(1-u_t/100)]$, where u_t is the percentage unemployment rate for the economy. Note that expressing the unemployment in this way implicitly give us data on the labour force (L_t) in hours. The real-wage series (W_t) is calculated from the average nominal wages per hour (which include pay roll tax rate) deflated by the GDP deflator. The real interest rate (R_IRATE) is calculated as IR_NOM less the inflation rate. From annual K_COST and quarterly INV, using the

[26] The data relating to employment (in hours of work) are obtained from International Labour Office (ILO): Bulletin of Labour Statistics.

perpetual inventory method, we calculate the capital stock (K_t) . The user cost of capital is calculated as $(R\_IRATE + DEP\_RATE)$ for DEP_RATE = depreciation ("scrapping") from the $OECD$ data-set. The profit rate $(\pi_t - k_t)$ is calculated as the log of the profit divided by K_t.

Prior to cointegration analysis, we need to establish the lag order for the VAR model. We have considered lag lengths ranging from 1 to 6, together with the univariate tests for the analysis of the residuals (vector Portmanteau statistic for residual correlation, LM test for autocorrelated residual, $ARCH$ effects, test for normality) and multivariate versions of these tests. The misspecification analysis suggests that a reasonable choice of the lags length is 4 for Canada, Japan,UK, US, Germany,France, and 5 for Italy. We have also considered the Akaike (1969) and Hannan and Quinn (1979) information criteria. The result was that any model with lower lag order, although supported by these information criteria failed to pass the misspecification tests.

4.4.1 The deterministic variables

In modelling the relationship between capital accumulation, factor prices, and unemployment it is important to keep in mind that many economies in Europe have experienced significant changes in their economic structure during the last thirty years. For example, UK and Italy have undergone substantial liberalisation of their labour markets. As a result there has been a shift in the structure of the aggregate labour sector. As Marcellino and Mizon (1999) point out the 1980s in UK correspond to a period of major changes in labour markets legislation which aimed at substantially increasing labour flexibility. Some of the

institutional changes are: a decrease in unemployment benefits, the weakening of union power, and the possibility for employers not to contract with unions. To take this into account we allow for a step dummy not restricted to lie in the cointegration space which take value one up to 1979:2. This date coincides with the election of the Thatcher Conservative Government in May 1979.

As far as Italy is concerned, in the 80s labour mobility increased due to amendments of the Wage Supplementation Fund (Cassa Integrazione Guadagni). On the other hand, at the beginning of this period a tighter monetary policy was introduced to control inflation, and the Banca d'Italia was concerned with the defence of the lira within the ERM. For this reason we introduce a step dummy which takes value one from 1981:2 to 1981:4 when there was a sharp drop in inflation and an increase in unemployment. In addition we introduce a step dummy for the period 1980:4 to1982:1 to capture the effect of recession induced by tight monetary policy. Finally, we include an impulse dummy for 1992:2 , since this corresponds to an important change in the measurement of unemployment .

To account for German reunification we insert a step dummy which takes the value one up to 1 July 1990, the date of reunification. The German economic and monetary union is associated with an important shock for the labour market. Indeed, after the reunification East-German wages adjusted rather quickly towards West-German wages despite the low labour productivity in the East. Wage differentials together with high unemployment in the East led to migration of workers from East to West Germany.

Finally, in order to account for outliers we include a number of impulse dummy variables in all the countries considered. The first group of dummies refers to the first oil shock in the 70s, and the second to the second oil shock in the 80s.

4.4.2 Determining the cointegrating rank

As specified in Section 4.3 to estimate the model we use the full information maximum likelihood approach introduced by Johansen (1988). As additional information we look at the dynamics of the VAR model. In particular, we consider the roots of the companion matrix, since this provides us additional information of how many $(n-r)$ roots are on the unit circle, and thus on the number of r cointegrating relations. In Table 4.1-4.7 we report the tests for cointegrating rank calculated for the different countries considered. The various hypotheses to be tested from no cointegration (i.e. $r=0$ or alternatively $n-r=5$) to increasing the number of cointegration vectors, are presented in the first column. The eigenvalues are presented in the second column, ordered from the highest to the lowest. Next come the $\lambda_{\max}$ statistics which test whether $r=0$ against $r=1$, or $r=1$ against $r=2$, etc. The λ_{trace} test is given in column six for the null that $r=q$ $(q=1,2,...,n-1)$. Using the asymptotic reference distribution in Osterwald and Lenum (1992), on the basis of the rank tests it is possible to accept that there are three cointegrating vectors for France, US, Canada, Japan and only two for Germany, UK , and Italy .

Table 4.1. Test for cointegrating rank using France data (1970:1-1998:1).

$H_0 : r$	k	$\hat{\lambda}$	$-T\log(1-\hat{\lambda}_{r+1})$	$\lambda_{\max R}$	$-T\sum\log(1-\hat{\lambda}_i)$	λ_{traceR}
0	5	0.583	92.8**	70.91**	194.2**	148.4**
1	4	0.409	55.73**	42.59**	101.4**	77.51**
2	3	0.225	27.63**	21.12	45.7**	34.92*
3	2	0.128	14.49	11.07	18.07	13.81
4	1	0.033	3.579	2.735	3.579	2.735

Table 4.2. Test for cointegrating rank using Germany data (1970:1-1998:1).

$H_0 : r$	k	$\hat{\lambda}$	$-T\log(1-\hat{\lambda}_{r+1})$	$\lambda_{\max R}$	$-T\sum\log(1-\hat{\lambda}_i)$	λ_{traceR}
0	5	0.5087	74.63**	56.86**	143.3**	109.2**
1	4	0.3755	49.44**	37.67**	68.67*	52.32*
2	3	0.1178	13.16	10.03	19.23	14.65
3	2	0.0476	5.061	3.856	6.067	4.623
4	1	0.009	1.006	0.766	1.006	0.766

Table 4.3. Test for cointegrating rank using UK data (1971:1-1998:1).

$H_0 : r$	k	$\hat{\lambda}$	$-T\log(1-\hat{\lambda}_{r+1})$	$\lambda_{\max R}$	$-T\sum\log(1-\hat{\lambda}_i)$	λ_{traceR}
0	5	0.4945	75.72**	56.03**	151.5**	124.2**
1	4	0.2547	32.63*	26.75	75.75**	62.1**
2	3	0.2096	26.11	21.4	43.13	35.36
3	2	0.1329	15.82	12.97	17.02	13.95
4	1	0.0107	1.194	0.979	1.194	0.979

Table 4.4. Test for cointegrating rank using Italy data (1971:1-1998:1).

$H_0 : r$	k	$\hat{\lambda}$	$-T\log(1-\hat{\lambda}_{r+1})$	$\lambda_{\max R}$	$-T\sum\log(1-\hat{\lambda}_i)$	λ_{traceR}
0	5	0.4139	57.16**	43.8**	115.6**	88.57**
1	4	0.3132	40.22**	30.82**	58.41**	44.76**
2	3	0.1157	13.16	10.09	18.19	13.94
3	2	0.0456	4.989	3.824	5.013	3.856
4	1	0.0004	0.042	0.032	0.042	0.031

Table 4.5. Test for cointegrating rank using US data (1960:1-1998:1).

$H_0 : r$	k	$\hat{\lambda}$	$-T\log(1-\hat{\lambda}_{r+1})$	$\lambda_{\max R}$	$-T\sum\log(1-\hat{\lambda}_i)$	λ_{traceR}
0	5	0.3675	67.81**	58.65**	158.9**	137.4**
1	4	0.2565	43.88**	37.95**	91.07**	78.76**
2	3	0.2088	34.67**	29.99**	47.19**	40.81**
3	2	0.0555	8.452	7.31	12.52	10.83
4	1	0.027	4.066*	3.516	4.066*	3.516

Table 4.6. Test for cointegrating rank using Canada data (1960:1-1998:1).

$H_0 : r$	k	$\hat{\lambda}$	$-T\log(1-\hat{\lambda}_{r+1})$	$\lambda_{\max R}$	$-T\sum\log(1-\hat{\lambda}_i)$	λ_{traceR}
0	5	0.4671	90.64**	78.05**	202.1**	174.1**
1	4	0.3824	69.41**	59.77**	111.5**	96**
2	3	0.1389	21.54	18.54	42.08**	36.24*
3	2	0.1190	18.26*	15.72*	20.55*	17.69
4	1	0.015	2.292	1.074	2.292	1.974

Table 4.7. Test for cointegrating rank using Japan data (1971:1-1998:1).

$H_0 : r$	k	$\hat{\lambda}$	$-T\log(1-\hat{\lambda}_{r+1})$	$\lambda_{\max R}$	$-T\sum\log(1-\hat{\lambda}_i)$	λ_{traceR}
0	5	0.3456	45.8**	37.32**	122.2**	99.57**
1	4	0.2963	37.95**	30.92*	76.4**	62.25**
2	3	0.2041	24.66*	20.09	38.45**	31.33*
3	2	0.0811	9.135	7.443	13.79	11.24
4	1	0.042	4.657*	3.795*	4.657*	3.795*

** Denotes rejection at the 10% significance level; * Denotes rejection at the 5% significance level.

The cointegration analysis was conducted using PcFiml for Windows. However, it is well known that χ^2 asymptotic distribution is potentially a poor approximation in small sample applications. Although, PcFiml by default calculates Reimers (1992) adjusted trace and λ_{max} statistics[27], it is not clear whether this is the preferred correction (see Doornik and Hendry (1994)) [28]. Since our theoretical model in Section 4.2 does not prespecify the exact number of cointegrating relations the rank test is crucially important in our analysis. For this reason, it is important to investigate the robustness of our results. To do it we augment the asymptotic inference with a simulation study based on non-parametric bootstrapping (see Appendix B of this chapter). In Table 4.8 we report the results of our experiment. The p-value we report are based on $B = 400$ replications.

Table 4.8. Bootstrapped p-values for $\lambda_{\max}$ and λ_{trace} tests (% values).

	Test Stat.	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
France	$\lambda_{\max}$	5.00	8.25	10.75	30.25	56.75
	λ_{trace}	5.00	5.75	4.50	32.75	56.75
Germany	$\lambda_{\max}$	5.00	5.25	48.50	73.25	-
	λ_{trace}	5.00	9.25	56.25	73.25	-
UK	$\lambda_{\max}$	5.00	3.25	31.25	77.75	82.25
	λ_{trace}	5.00	3.00	42.5	85.25	82.25
Italy	$\lambda_{\max}$	8.75	19.0	86.75	89.0	40.25
	λ_{trace}	8.5	29.0	18.81	6.27	0.81
US	$\lambda_{\max}$	5.00	5.25	5.75	35.75	2.5
	λ_{trace}	5.00	5.00	5.25	11.5	2.5
Canada	$\lambda_{\max}$	5.00	5.00	24.0	10.5	76.75
	λ_{trace}	5.00	5.00	7.0	16.0	76.75
Japan	$\lambda_{\max}$	5.00	5.00	26.25	21.75	18.25
	λ_{trace}	5.00	5.00	6.25	16.75	18.25

[27] These are given in column 4 and 7 of table 4.1-4.7.

[28] Reimers's correction factor allows us to take account of the number of parameters to be estimated in the model by making an adjustment for the degree of freedom. This is done by replacing the sample size (T) in the original Johansen's (1988) trace and λ_{max} statistics with $(T - np)$, where n is the dimension of the VAR, and p are the number of lags included in the estimated model.

Comparing Table 4.1-4.7 with Table 4.8 we can see that the conclusions about the cointegrating rank are quite in agreement with Reimers's small-sample corrected tests, while Johansen tests in some case (e.g. Canada) tend to overestimate the number of cointegrating vectors. Moreover, from Table 4.8 we can see that in general the p-values of bootstrapped $\lambda_{\max}$ are higher then the p-values of the λ_{trace} test. In the literature it has been established (see Cheung and Lai (1993)) that the trace test is more robust then the $\lambda_{\max}$ test. For all these reasons we decided: (i) to rely more on the bootstrap test, (ii) to place greater weight on the trace test than the $\lambda_{\max}$ test. Following this criteria we are able to accept $r = 3$ for US, Canada (at 7%), for France (at 4.5%), and Japan (at 6.25%), and $r = 2$ Germany (at 9.25%), and UK (at 3%). The conclusion about the cointegrating rank change for Italy (where $r = 1$ at 8.5%). However, the moduli of the three largest roots in the companion matrix are 0.9893×1 and 0.9566×2, are close to unity suggesting that $k = 3$, so we decided to proceed with $r = 2$.

4.4.3 Testing for linear restrictions on the cointegrating space

From Table 4.1-4.7, all we know is how many cointegration vectors span the cointegration space. However, any linear combination of the stationary vectors is also a stationary vector, as a consequence the estimates for any column in $\hat{\beta}$ are not necessarily unique. Therefore, the next step is to impose linear restrictions on the cointegrating space, and then test if the columns of $\hat{\beta}$ are identified.

In the case where $r = 3$, we test if the following vectors are contained in the estimated cointegration space

$$\hat{\beta}_{1.1} = [\ *\ \ 1\ \ 0\ \ *\ \ 0\ ]',$$
$$\hat{\beta}_{1.2} = [\ 0\ \ *\ \ 1\ \ 0\ \ *\ ]',$$
$$\hat{\beta}_{1.3} = [\ 0\ \ *\ \ 0\ \ 1\ \ *\ ]'.$$

The symbol " * " is used to indicate the parameters which are left unrestricted. In term of the theoretical labour market model, we identify the first as the wage setting relation given by equation (4.2.5), the second with equation (4.2.2), and the third with equation (4.2.3).

As a next step we additionally require equal coefficients with opposite sign on the real wage and productivity (i.e. we test for $\gamma_3 = 1$ in equation (4.2.5), and we test for homogeneity between the profit rate and user cost of capital in equation (4.2.2), (i.e. we test for $(\pi_t - k_t) - uc_t = 0)$.

In the case $r = 2$, we test for

$$\hat{\beta}_{1.1} = [\ *\ \ 1\ \ 0\ \ *\ \ 0\ ]',$$
$$\hat{\beta}_{1.2} = [\ 0\ \ *\ \ 1\ \ 0\ \ *\ ]',$$

and then for homogeneity restriction between the profit rate and user cost of capital as above[29].

---

[29] In addition we check for the hypotheses of stationarity of unemployment by testing if

$$\hat{\beta}_{2.1} = [\ *\ \ 1\ \ 0\ \ 0\ \ 0\ ]',$$

Table 4.9. The estimated structural long-run relations.

France	w_t =01.2727 + $(y_t - n_t)$ − 3.6958 u_t (0.252) (0.427)	
	$\pi_t - k_t$ =2.5049+0.73291 $w_t - uc_t$ (0.329) (0.086)	
	u_t =3.6015−0.92778w_t−0.984 uc_t (0.100) (0.102) (0.355)	$\chi^2(2) = 5.27$ $p = 0.0716$
Germany	$\pi_t - k_t$ =−0.07 w_t− 2.956 uc_t (0.524) (0.40)	
	w_t = − 0.6735 u_t (0.0136)	$\chi^2(1) = 0.034$ $p = 0.8423$
UK	w_t =0.3682 $(y_t - n_t)$ − 4.6322 u (0.044) (2.0979)	
	$\pi_t - k_t$ =−0.020617w_t−0.82493 uc_t (0.0250) (0.108)	$\chi^2(2) = 1.45$ $p = 0.4837$
Italy	w_t =−0.266 $(y_t - n_t)$ + 6.028 u_t (0.2907) (1.7928)	
	$\pi_t - k_t$ =−1.5372w_t − 1.880 uc_t (0.5032) (0.5662)	$\chi^2(2) = 4.72$ $p = 0.095$
US	$w_t = (y_t - n_t)$ − 16.352 u_t (1.5836)	
	$\pi_t - k_t$ = − 0.2468 w_t− 0.001 uc_t (0.2409) (0.0003)	
	u_t =0.309 w_t+ 0.0029uc_t (0.0367) (0.0004)	$\chi^2(1) = 0.182$ $p = 0.6699$
Canada	w_t =−2.371−0.1273 $(y_t - n_t) - u_t$ (0.0598) (0.0156)	
	$\pi_t - k_t$ =−6.064+ 5.236 $w_t - uc_t$ (1.005) (0.3593)	
	u_t =−0.01949w_t − 3.4122uc_t (0.0168) (0.2322)	$\chi^2(2) = 3.88$ $p = 0.1434$
Japan	w_t = − 0.1409 $(y_t - n_t)$ − 31.21 u_t (0.00745) (2.3245)	
	$\pi_t - k_t$ =−6.305 $w_t + uc_t$ (0.2538)	
	u_t =−0.074 w_t− 0.042 uc_t (0.0073) (0.043)	$\chi^2(1) = 1.024$ $p = 0.3116$

$$\hat{\beta}_{2.2} = [\ 0\ \ *\ \ 1\ \ 0\ \ *\ ]',$$
$$\hat{\beta}_{2.3} = [\ 0\ \ 1\ \ 0\ \ -1\ \ 0\ ]'.$$

are contained in the cointegrating space. This hypothesis has been rejected for all countries except for US and Japan with $\chi^2(3) = 7.1874$, $(p = 0.0662)$, and $\chi^2(3) = 6.4248$, $(p = 0.094)$, respectively.

In Table 4.9, we report the estimated coefficients of the equations described in Section 4.2, χ^2 statistic, and the correspondent p-value. The first equation, determines how unemployment and productivity affect wages. The long-run elasticity of wages with respect to unemployment is lower for the European countries, US, and Canada than for Japan. This fact seems to reflect differences in the structure of the labour market in particular for Japan. However, the estimated elasticity for Japan may be affected by the starting point of the time series we have considered. Indeed, the wage flexibility in Japan increased after the second oil shock, therefore, it may be interesting to extend the sample to a period before the 1965 to see how the estimate of γ_3 is affected.

The response of wages to unemployment has played a crucial role on the explanation of European unemployment. For this reason equation (4.2.5) has been estimated by many authors using different econometric methods. For example, Layard *et al.* (1991) report values of γ_3 ranging from 0.53 for the UK to 41 for Japan. Alogoskoufis and Manning (1988) obtain similar results. For the Scandinavian countries, Jacobson *et al.* (1997) using the common trends model find $\gamma_3 = 1.2$ for Denmark and $\gamma_3 = 9.67$ for Norway. It is important to keep in mind that since all the variables are endogenous a stationary relation between unemployment and real wages does not mean that a rise in the real wages causes a rise in unemployment [30]

A separate discussion is required for Germany. In Table 4.9 we report the estimated parameters for a four dimensional VAR model which does not include the variable $(y_t - n_t)$. The reason is that although on the base of the rank test we were able to accept

[30] The likelihood ratio test rejects the weak exogeneity hypothesis for any variable of the model.

two cointegrating vectors, the restrictions described above are rejected by the test for linear restrictions on the $\hat{\beta}$. One explanation for this result may be the poor quality of our data set relating to this country[31]. Indeed, Carstensen and Hansen (2000) using seasonally adjusted data of the DIW database for the West Germany, find a stationary relation between unemployment and the wage wedge, and they estimate $\gamma_3 = 1.824$.

The second equilibrium equation indicates, except for Japan, a negative relationship between profit rate and user costs of capital. The estimated parameter for the user cost of capital can be interpreted as the elasticity of investment with respect to the shadow price of capital. In general, the lower the cost of adjustment, the faster the difference between profit and user cost is translated into capital accumulation or decumulation. From Table 4.9, we find the estimated elasticity is particularly low for US.

The third equation relates unemployment to factor prices. Unemployment would be expected to rise with increases in real wages, since this translate to higher cost for firms. However, an increase of w_t may also increase the labour force and therefore increase unemployment in the presence of nominal rigidities. This may explain the positive sign for US.

The long run elasticity of unemployment with respect to user cost of capital is particularly large for Canada and surprisingly low for US, where the "wage channel" is more important. Since the coefficient of variation of the depreciation rate along the time series considered in our sample is low, it is plausible to assume that the pattern of the user cost of

[31] The definition of "earning per hour" in the *ILO* publication has been changed several times in the estimation. Moreover, the data for this variable relate to the Fed. Rep. of Germany before 3.10.1990, and include East Germany afterward.

capital over the time is mainly determined by the behavior. of the real interest rate. According to Blanchard (1998) the real interest rate affects unemployment through a reduction of investment and consequent capital decumulation, while Fitoussi and Phelps (1988) insist on the effect of the real interest rate on the markup chosen by imperfectly competitive firms[32]. In our model we are not able to distinguish between these two channels; nevertheless both types of shocks have a positive effect on unemployment.

Figures 4.2-4.8 present the $\beta' X_t$-vectors adjusted for the short run dynamics. An inspection of these graphs reveals that these vectors are stationary. As additional check of the adequacy of the model we plotted the recursive estimates of the first r non-zero eigenvalues[33]. Generally, for the model we consider there is no evidence of parameter instability due to, for example the failure to account for structural breaks.

[32] According to Blanchard (1997) an increase in the interest rate leads firms to increase their mark-up of price over cost: to do this they decrease the real wage to workers. Increased unemployment is needed for workers to accept this lower real wage.

[33] To save space we omit these figures.

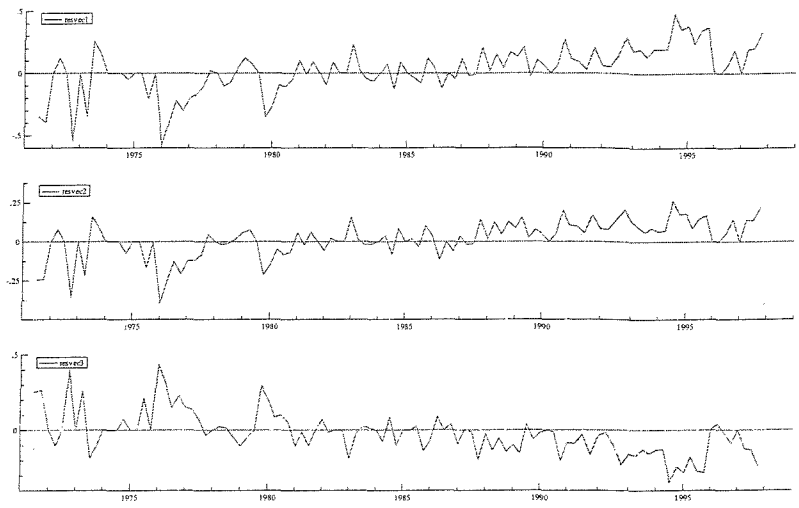

Figure 4.2. Estimated restricted cointegrating relations for France.

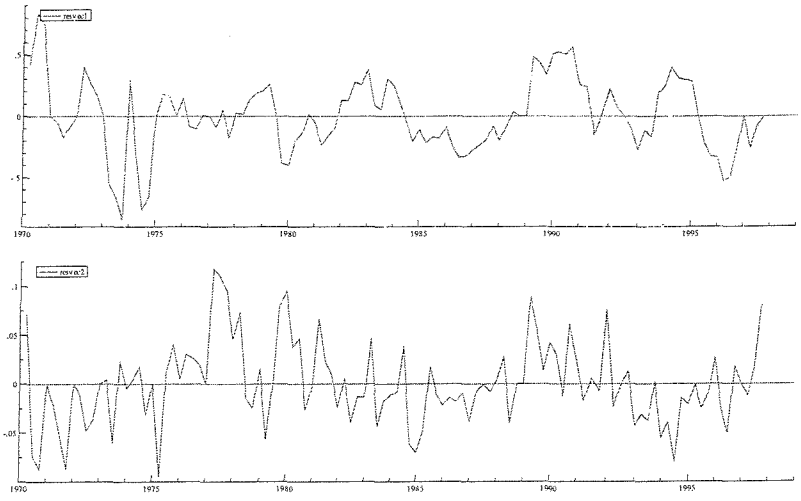

Figure 4.3. Estimated restricted cointegrating relations for UK.

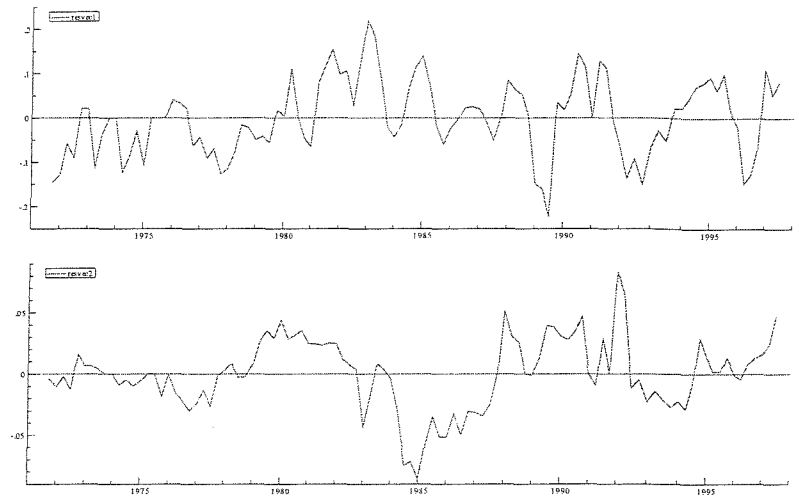

Figure 4.4. Estimated restricted cointegrating relations for Germany.

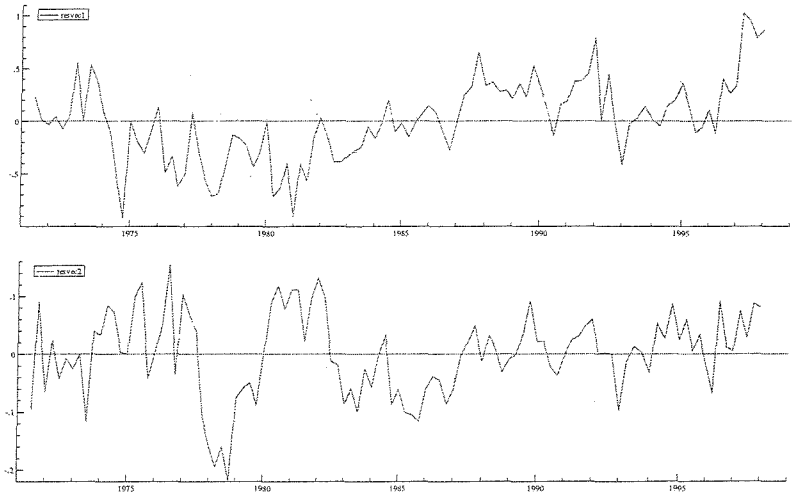

Figure 4.5. Estimated restricted cointegrating relations for Italy.

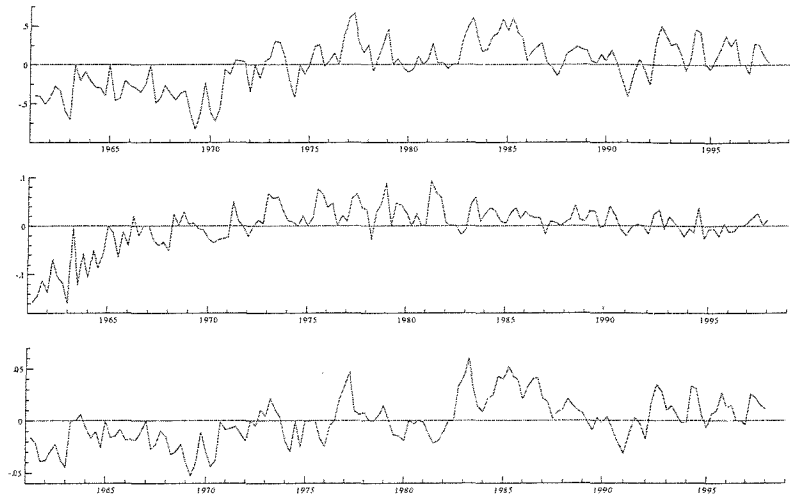

Figure 4.6. Estimated restricted cointegrating relations for US.

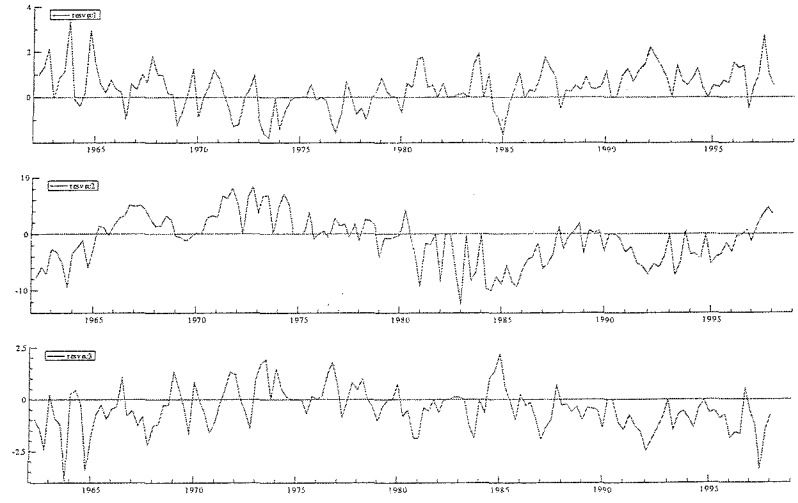

Figure 4.7. Estimated restricted cointegrating relations for Canada.

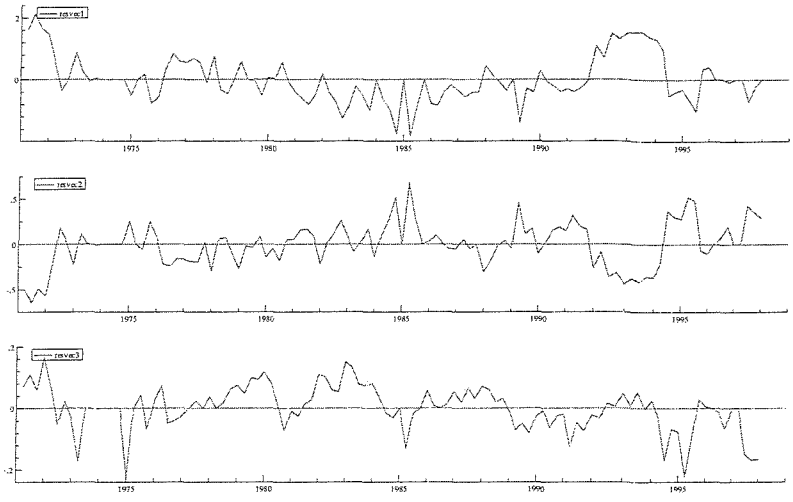

Figure 4.8. Estimated restricted cointegrating relations for Japan.

4.4.4 Common trends analysis

Using the Johansen procedure for the five variable VAR model we found a cointegration rank of three for UK, France, US, Canada, Japan and two for Italy, Germany.

As we have seen in Section 4.2, the structural shocks are identified as a productivity shock, and a stochastic shock to the profit rate. The transitory shocks are a shift in the labour demand and shock to the user cost of capital. The equilibrium unemployment innovation is identified as long-run shock in the case of three common trends, and short-run shock in the case where the cointegrating rank is three.

To identify the common trends model in the first case (i.e. when $n - r = 2$) we need to impose $k(k-1)/2 = 1$ further restrictions. To identify the structural VAR in the latter case (i.e. the long-run shocks for $n - r = 3$) we need $k(k-1)/2 = 3$ extra restrictions.

The restriction we impose is that the markup trend has no long-run effects on the labour productivity. The assumption is exactly identifying in the case of two trends. If there are three trends we add two extra assumptions by imposing a zero restriction on the coefficients a_{13} and a_{23} of the loading matrix A described in equation (4.3.1) of Section 4.3, that is we impose that the unemployment innovations do not affect labour productivity and the profit per unit of capital in the long-run.

Table 4.10-4.16 reports the estimated loading matrix of the common trends. The standard errors of the estimated coefficients are given in parenthesis [34]. The interpretation of the coefficient a_{ij} of the loading matrix A is that a_{ij} measures the long-run effects on the i^{th} endogenous variable from a unit shock to the j^{th} trend innovation.

Since the model in Section 4.2 does not holds for Germany we report the estimated parameters for the five variable VAR described above, but we omit the productivity term. Note that in this case the restriction for $a_{12} = 0$ is exactly identifying.

Considering the first common trend coefficients, we can see that a one standard-deviation increase in technology increases the labour productivity around 5% in Japan, 3% in Canada, while in US only 1%. In the long run both wages and unemployment are also noticeably affected by technological progress.

The second trend captures the effect of a unit shock to the markup. Considering the second column of the A matrix, we can see that a one standard deviation increase in markup increases the profit rate, this is particularly true for Japan and Italy. However, the effects on unemployment are more controversial. If μ_t is stationary, according to economic

[34] The asymptotic distribution of the common trends coefficients is Gaussian.

theory an increase of the labour markup leads in the short and medium run to a downward movement in the labour demand, since an increase of the markup acts like a tax on labour. As a consequence employment falls and unemployment rises. A rise in unemployment leads to a decrease in real wages. On the other hand lower wages lead to an increase in the profit rate, and because of the excess of profit new entry of firms, and capital accumulation. Therefore, employment rises as a result of either capital accumulation or a larger number of firms. This process continues until the profit rate has returned to its original value, as the real wage.

To what extent can economic theory explain our empirical results? It seems plausible to consider wage behavior as the main protagonist of our story. We can see from Table 4.12 and Table 4.14 that in UK and US a shock to the markup leads to an increase in the wages and a decrease inunemployment while the same is not true for the other European countries. One interpretation of this result is that for US and UK the "capital accumulation effect" induced by the increase of the profit rate is more dominating than the "tax effect". This seems to support the idea that wage rigidity matters. Indeed, if wages do not adjust to unemployment, then markup shocks lead firms to decrease the ratio of labour to capital so that the profit rate remains unchanged in the long run. Therefore, both higher capital and a lower labour to capital ratio lead to a decrease in employment and consequently higher unemployment. Our results support the Blanchard (1997) hypothesis that the upward trend on markup (and consequently in the profit rate) in most "continental European countries" is able to explain, at least in part, the evolution of the unemployment rate in the 80s.

Finally, for Italy and Germany from Table 4.11 and 4.13 we can see that a unit shock to unemployment innovation decreases wages by about 4% in Italy and 3% in Germany.

Table 4.10. Estimated common trends for France, 1968:1,1998:1[35].

$$\begin{bmatrix} y_t - n_t \\ \pi_t - k_t \\ u_t \\ w_t \\ uc_t \end{bmatrix} = x_0 + \begin{bmatrix} \underset{(0.013)}{0.039} & \underset{(-)}{0} \\ \underset{(0.0007)}{0.017} & \underset{(0.0002)}{0.0012} \\ \underset{(0.0005)}{0.0014} & \underset{(0.0001)}{-0.0006} \\ \underset{(0.001)}{-0.217} & \underset{(0.005)}{-0.0024} \\ \underset{(0.0014)}{0.0028} & \underset{(0.0006)}{-0.0029} \end{bmatrix} \begin{bmatrix} \hat{\tau}_{y,t} \\ \hat{\tau}_{\pi,t} \end{bmatrix} + \Phi(L)v_t$$

Table 4.11. Estimated common trends for Germany, 1970:1,1998:1.

$$\begin{bmatrix} \pi_t - k_t \\ u_t \\ w_t \\ uc_t \end{bmatrix} = x_0 + \begin{bmatrix} \underset{(0.06)}{0.0316} & \underset{(-)}{0} \\ \underset{(0.033)}{0.004} & \underset{(0.1428)}{-0.1101} \\ \underset{(0.029)}{-0.0353} & \underset{(0.062)}{0.0125} \\ \underset{(0.0157)}{0.0092} & \underset{(0.0023)}{-0.0017} \end{bmatrix} \begin{bmatrix} \hat{\tau}_{\pi,t} \\ \hat{\tau}_{u,t} \end{bmatrix} + \Phi(L)v_t$$

Table 4.12. Estimated common trends for UK, 1968:1,1998:1.

$$\begin{bmatrix} y_t - n_t \\ \pi_t - k_t \\ u_t \\ w_t \\ uc_t \end{bmatrix} = x_0 + \begin{bmatrix} \underset{(0.068)}{0.046} & \underset{(-)}{0} & \underset{(-)}{0} \\ \underset{(0.067)}{0.045} & \underset{(0.00305)}{0.0072} & \underset{(-)}{0} \\ \underset{(0.02)}{-0.016} & \underset{(0.00158)}{-0.0021} & \underset{(0.00045)}{0.0023} \\ \underset{(0.013)}{-0.093} & \underset{(0.0073)}{0.0097} & \underset{(0.00211)}{-0.0104} \\ \underset{(0.039)}{0.025} & \underset{(0.0038)}{0.0091} & \underset{(0.00007)}{0.0033} \end{bmatrix} \begin{bmatrix} \hat{\tau}_{y,t} \\ \hat{\tau}_{\pi,t} \\ \hat{\tau}_{u,t} \end{bmatrix} + \Phi(L)v_t$$

[35] Note: The estimated asymptotic standard errors are reported within parentheses.

Table 4.13. Estimated common trends for Italy, 1970:1,1998:1.

$$\begin{bmatrix} y_t - n_t \\ \pi_t - k_t \\ u_t \\ w_t \\ uc_t \end{bmatrix} = x_0 + \begin{bmatrix} 0.0715 & 0 & 0 \\ (0.64) & (-) & (-) \\ -0.1256 & 0.02 & 0 \\ (0.1512) & (0.019) & (-) \\ -0.01 & 0.002 & 0.0024 \\ (0.095) & (0.002) & (0.0004) \\ 0.0459 & 0.0067 & -0.004 \\ (0.4211) & (0.0057) & (0.0007) \\ 0.0551 & 0.001 & 0.006 \\ (0.509) & (0.072) & (0.001) \end{bmatrix} \begin{bmatrix} \hat{\tau}_{y,t} \\ \hat{\tau}_{\pi,t} \\ \hat{\tau}_{u,t} \end{bmatrix} + \Phi(L)v_t$$

Table 4.14. Estimated common trends for US, 1960:1,1998:1.

$$\begin{bmatrix} y_t - n_t \\ \pi_t - k_t \\ u_t \\ w_t \\ uc_t \end{bmatrix} = x_0 + \begin{bmatrix} 0.0117 & 0 \\ (0.0056) & (-) \\ 0.0008 & 0.00027 \\ (0.00044) & (0.00005) \\ -0.00045 & 0.00014 \\ (0.00021) & (0.00002) \\ 0.00441 & 0.00226 \\ (0.0027) & (0.00041) \\ -0.3157 & -0.2882 \\ (0.2461) & (0.05187) \end{bmatrix} \begin{bmatrix} \hat{\tau}_{y,t} \\ \hat{\tau}_{\pi,t} \end{bmatrix} + \Phi(L)v_t$$

Table 4.15. Estimated common trends for Canada, 1960:1,1998:1.

$$\begin{bmatrix} y_t - n_t \\ \pi_t - k_t \\ u_t \\ w_t \\ uc_t \end{bmatrix} = x_0 + \begin{bmatrix} 0.037 & 0 \\ (0.07) & (-) \\ 0.0227 & 0.0024 \\ (0.042) & (0.0007) \\ -0.0034 & 0.0016 \\ (0.007) & (0.0004) \\ -0.0082 & -0.0016 \\ (0.017) & (0.0004) \\ -0.02 & -0.011 \\ (0.004) & (0.002) \end{bmatrix} \begin{bmatrix} \hat{\tau}_{y,t} \\ \hat{\tau}_{\pi,t} \end{bmatrix} + \Phi(L)v_t$$

Table 4.16. Estimated common trends for Japan, 1965:1,1998:1.

$$\begin{bmatrix} y_t - n_t \\ \pi_t - k_t \\ u_t \\ w_t \\ uc_t \end{bmatrix} = x_0 + \begin{bmatrix} 0.0522 & 0 \\ (0.031) & (-) \\ -0.092 & 0.01 \\ (0.237) & (0.005) \\ -0.0024 & 0.00 \\ (0.006) & (0.0) \\ 0.008 & 0.0023 \\ (0.022) & (0.001) \\ -0.042 & -0.004 \\ (0.105) & (0.002) \end{bmatrix} \begin{bmatrix} \hat{\tau}_{y,t} \\ \hat{\tau}_{\pi,t} \end{bmatrix} + \Phi(L)v_t$$

4.4.5 The dynamics of the structural model

In this section we show how the VAR system reacts to various impulses. The impulse response function, along with asymptotic 95 percent confidence bands are given in Figures 4.1-4.10.

As have seen in Section 4.3, the impulse response analysis can be performed by rewriting the moving average model (i.e. equation (4.3.8)) as equation (4.3.10).

To identify F_r we have used the following specification

$$F_r = Q^{-1}\omega' \sum^{-1},$$

where $\underset{(n\times r)}{\omega} = \underset{(r\times r)}{[\alpha \quad \gamma]}$ so that the permanent and the transitory innovations are uncorrelated. Indeed, $E[vv']$ is given by

$$E[vv'] = F_r \sum F_r' = Q^{-1}\omega' \sum^{-1} \omega \left(Q^{-1}\right)' = I,$$

where the elements of Q are obtained from the Cholesky decomposition of $\omega' \sum^{-1} \omega$.

The second assumption we need is that only the first k structural shocks have long effects, whereas the last r shocks do not. This implies that

$$R(1) = C(1) F^{-1} = \left[\underset{(n\times k)}{A} \quad \underset{(n\times r)}{0} \right].$$

As far the results are concerned, from Figure 4.11 we can see that a wage shock in Italy reduces the profit rate and the user cost of capital by about 2%. Moreover, it produces a fall in labour productivity of about 0.25%, while unemployment is not affected. By contrast, in US a shock to wages does not affect the profit rate, but real wages, user cost

of capital, and unemployment increase. In particular, although as a response to a positive wage shock, real wages rapidly adjust to their long run level, after less then one year the unemployment rate starts increasing, and it takes approximately four years to reach the long run level. For Germany, we find that a shock to wages increases unemployment by about 1%. Adjustment to equilibrium unemployment shocks occurs within approximately 2 years.

Coming to the shock to the user cost of capital, we can see that a unit shock to the user cost significantly reduces the profit rate in UK (0.1%), Japan (1%), and Canada (0.4%). Moreover, we find significant response in unemployment for Germany (4%),UK (approximately 0.05%), US (3%), Canada (about 0.2%), Japan (0.06%)

From an inspection of Figures 4.12 and 4.15 it is evident that in general the technology trend and the markup trend do not affect unemployment, since we do not record any significant response for the first 10 years in these countries[36]. The only exception is the US, where a productivity shock reduces unemployment by about 0.2% in the long run. The fact that a technology shock does not affect unemployment in the long run is in line with the theoretical model by Layard *et al.*(1991), and some empirical results for the Scandinavian unemployment by Jacobson *et al.* (1997). The result from the markup shock are not consistent with the estimated parameters in Section 4.2. Indeed, the magnitude of the estimated coefficients falls in a range between 0 for Japan and 0.003 for UK, even though the 95% confidence interval are very wide.

---

[36] In Figure 4.10,4.13,4.16 we report the impulse response function to one standard deviation shock to technology and profit rate for France, Germany, US. To save space we omit to report the ones for the other countries whose the result are not significant. They are available on request.

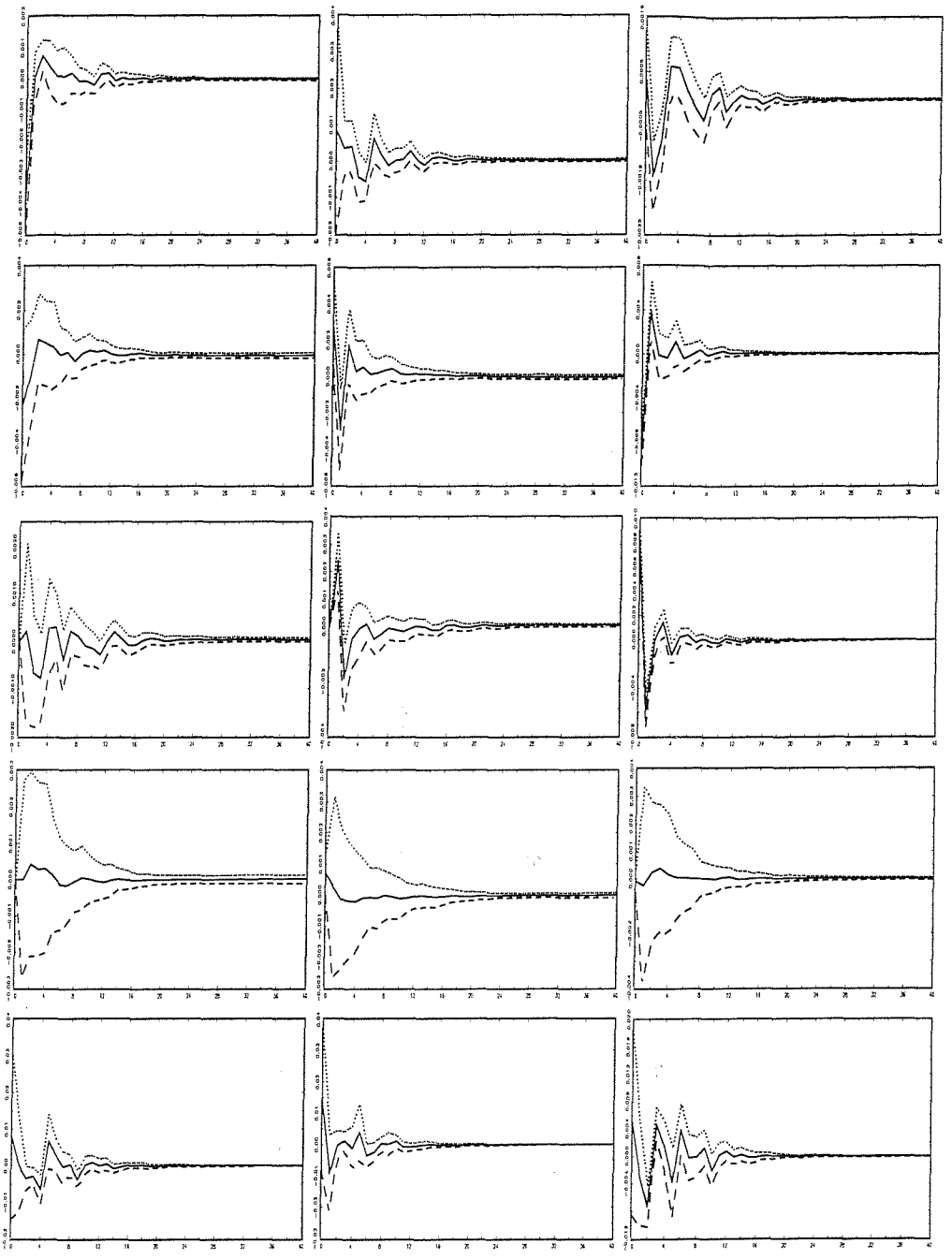

Figure 4.9. FRANCE: Impulse response function with 95% confidence interval from a one-standard deviation shock to the unemployment, wage innovation, and user cost of capital innovation. Response in $(y_t - n_t)$, $(\pi_t - k_t)$, u_t, w_t, uc_t respectively.

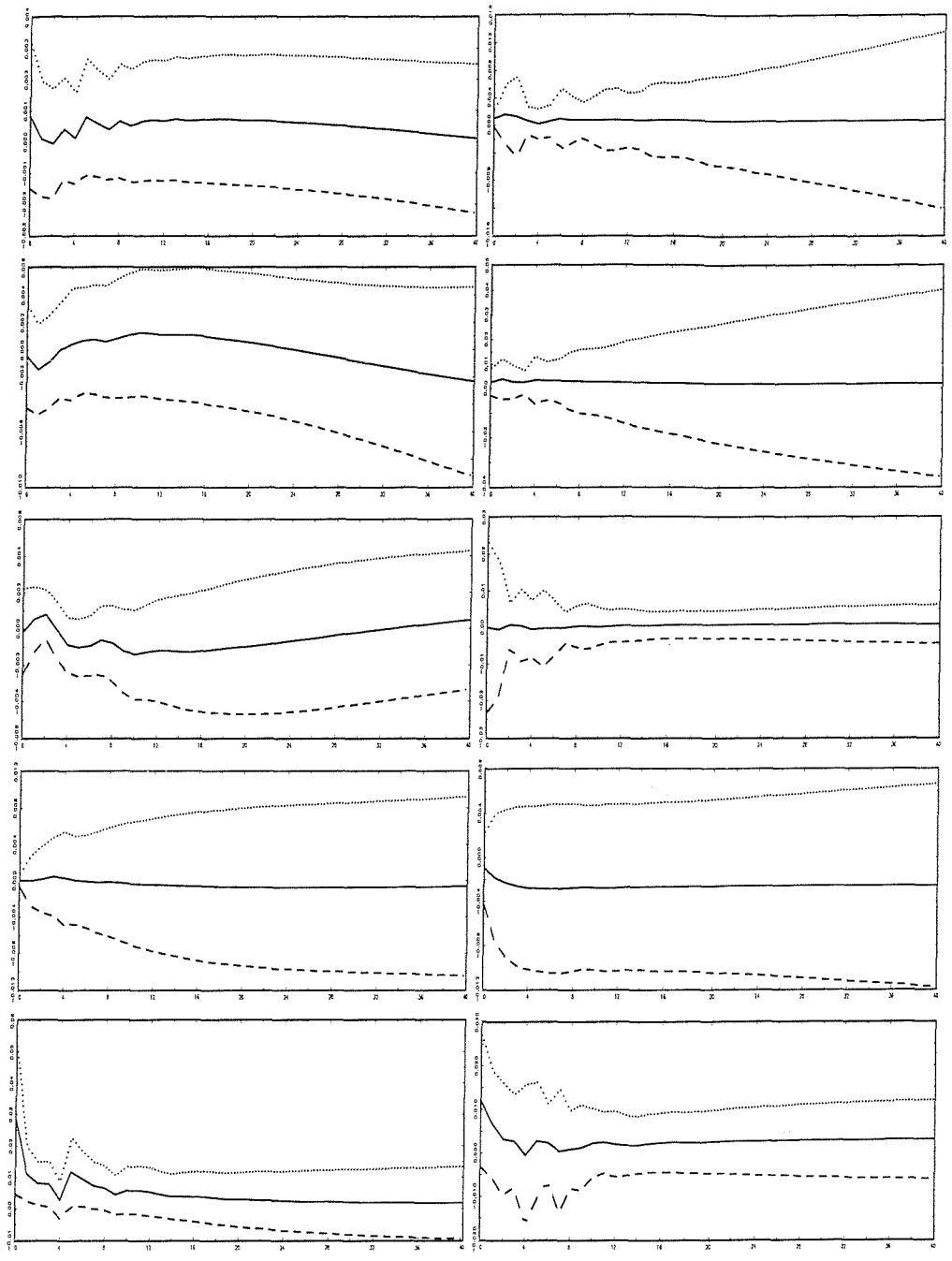

Figure 4.10. FRANCE: Impulse response function with 95% confidence interval from a one-standard deviation shock to the markup innovation and unemployment innovation. Response in $(y_t - n_t)$, $(\pi_t - k_t)$, w_t, u_t, uc_t respectively.

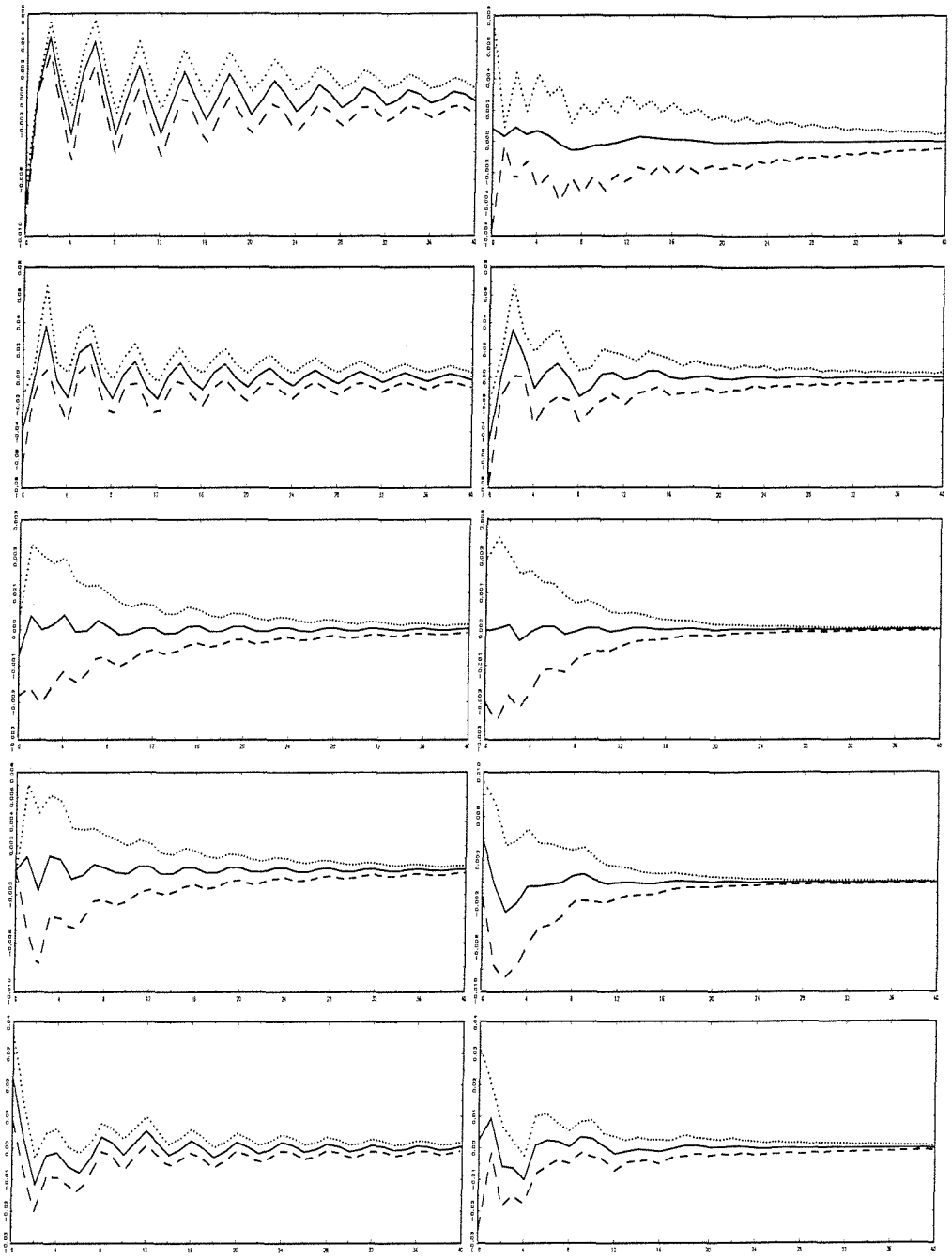

Figure 4.11. ITALY: Impulse response function with 95% confidence interval from a one-standard deviation shock to the wage innovation and user cost of capital innovation. Response in $(y_t - n_t)$, $(\pi_t - k_t)$, u_t, w_t, uc_t respectively.

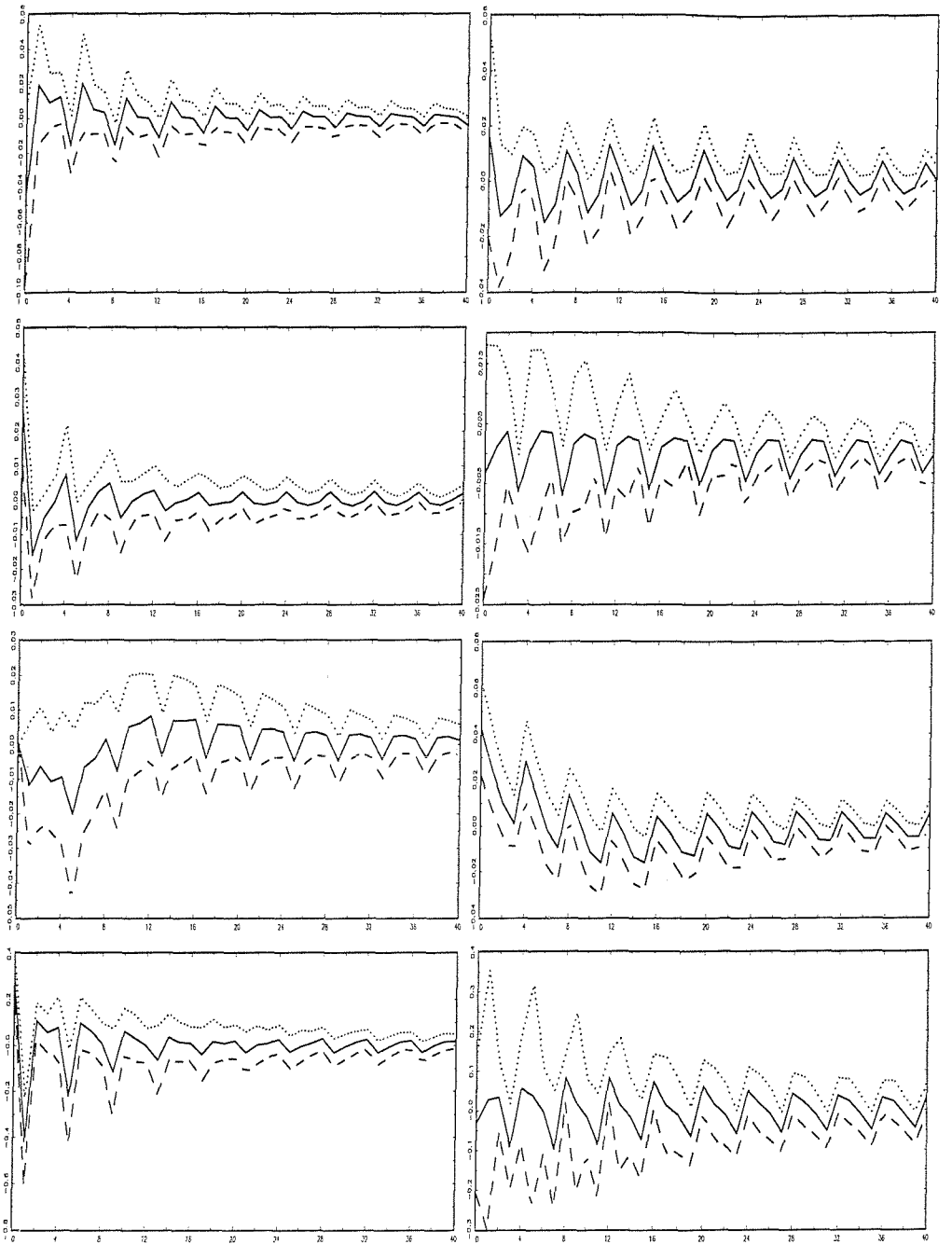

Figure 4.12. GERMANY: Impulse response function with 95% confidence interval from a one-standard deviation shock to the wage innovation, and user cost of capital innovation. Response in $(\pi_t \ - k_t)$, u_t, w_t, uc_t respectively.

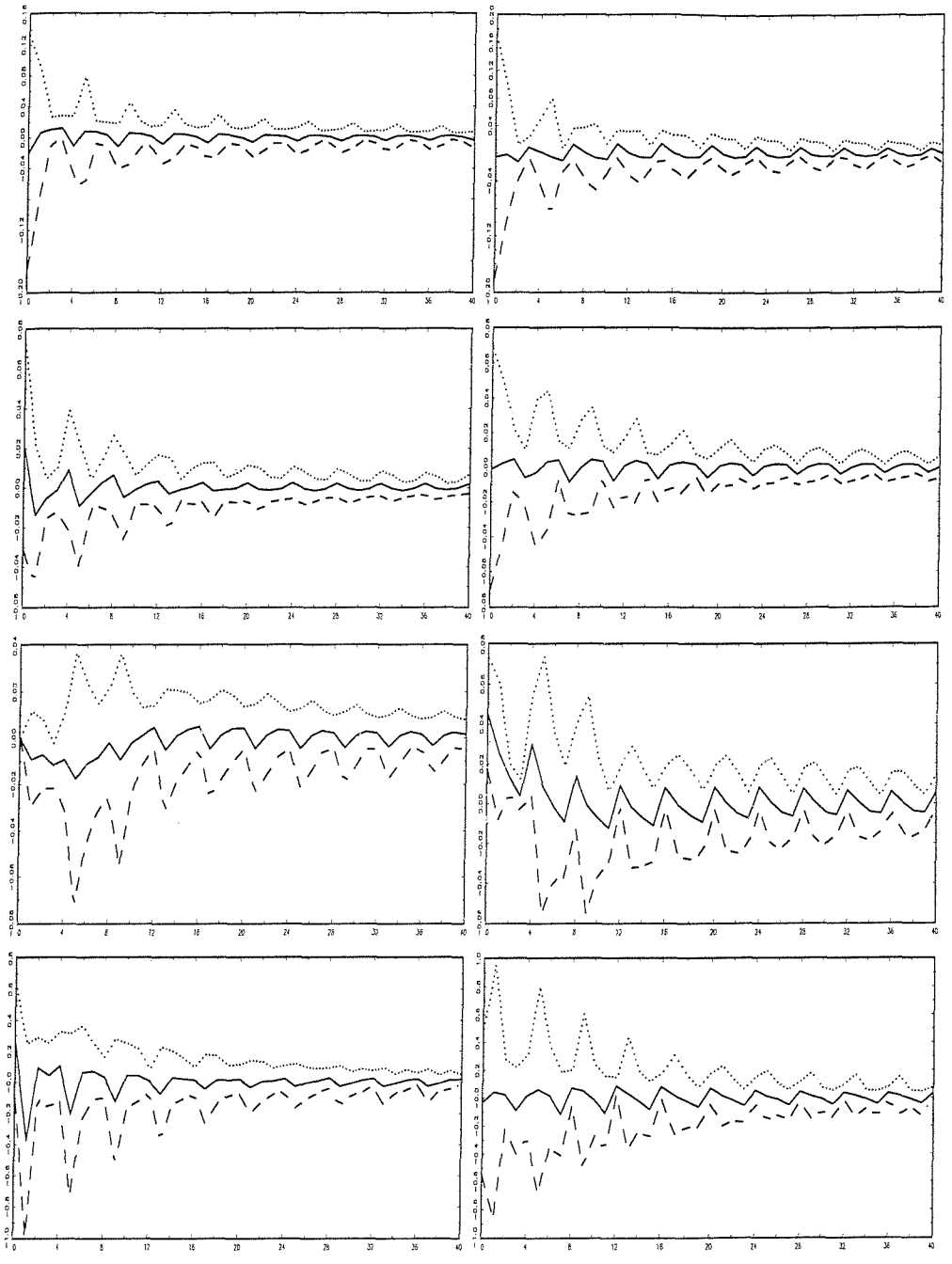

Figure 4.13. GERMANY: Impulse response function with 95% confidence interval from a one-standard deviation shock to the markup innovation and unemployment innovation. Response in $(\pi_t - k_t), w_t, u_t, uc_t$ respectively.

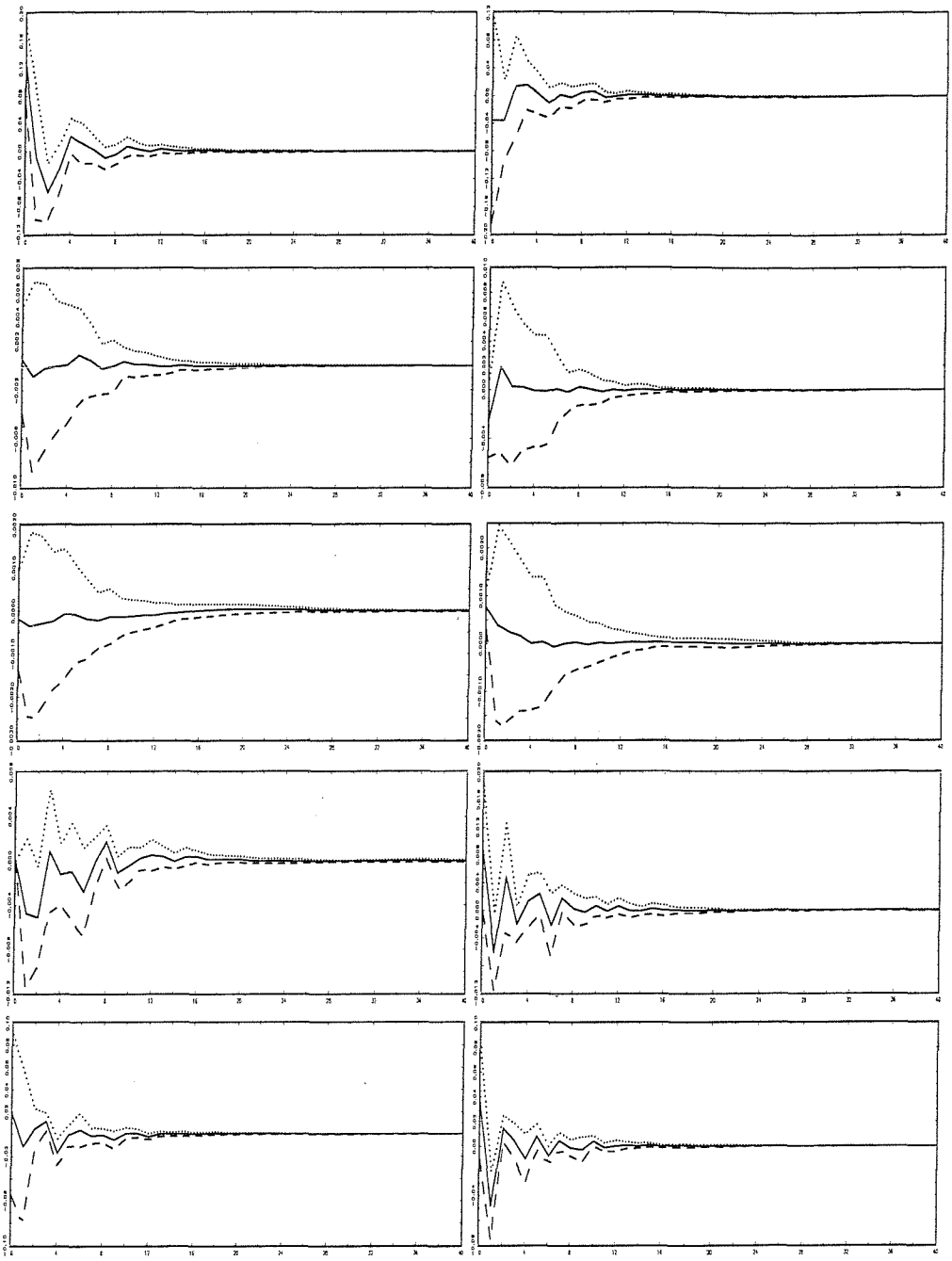

Figure 4.14. UK: Impulse response function with 95% confidence interval from a one-standard deviation shock to the wage innovation and user cost of capital innovation. Response in $(y_t - n_t)$, $(\pi_t - k_t)$, u_t, w_t, uc_t respectively.

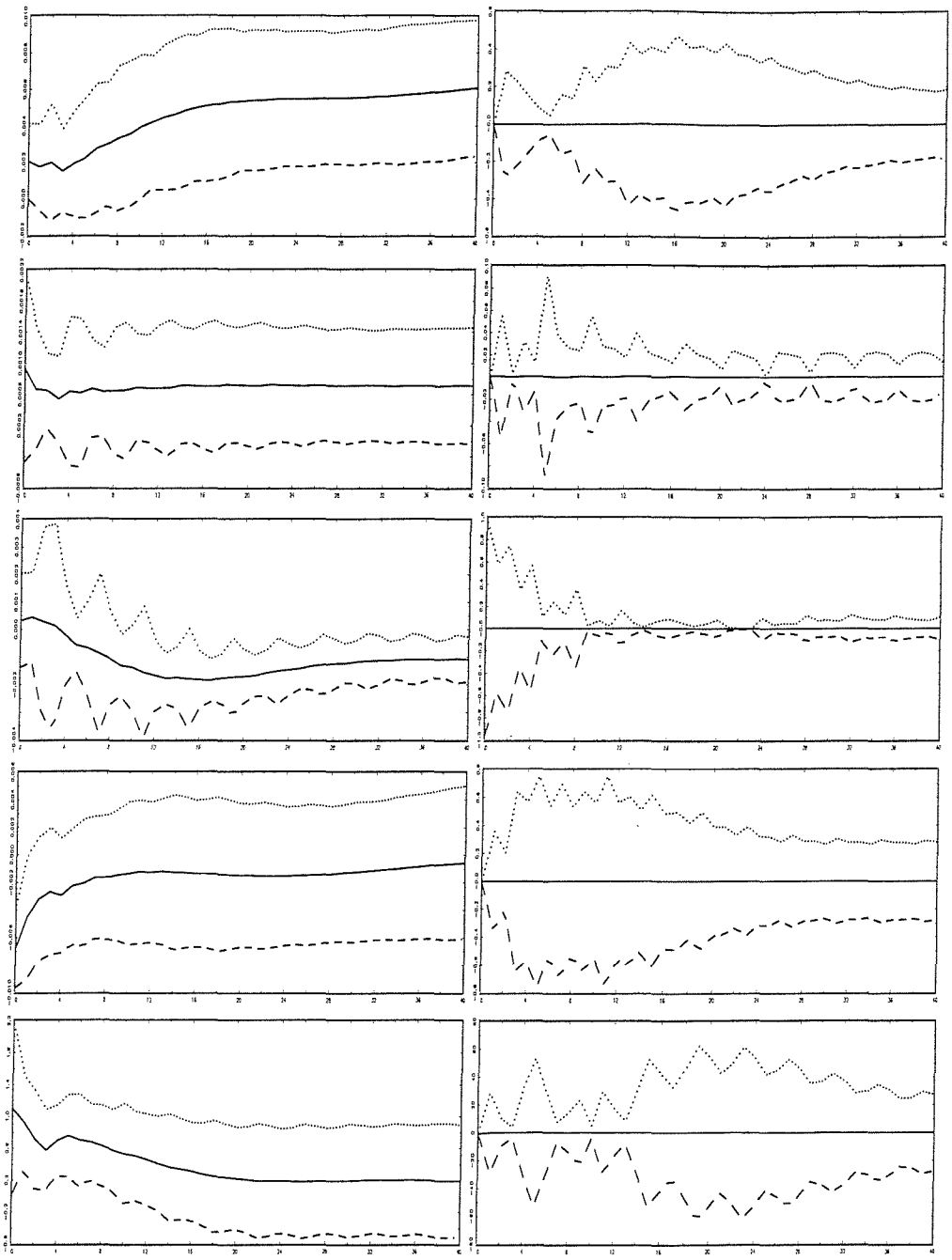

Figure 4.15. US: Impulse response function with 95% confidence interval from a one-standard deviation shock to technology innovation and markup innovation. Response in $(y_t - n_t)$, $(\pi_t - k_t)$, u_t, w_t, uc_t respectively.

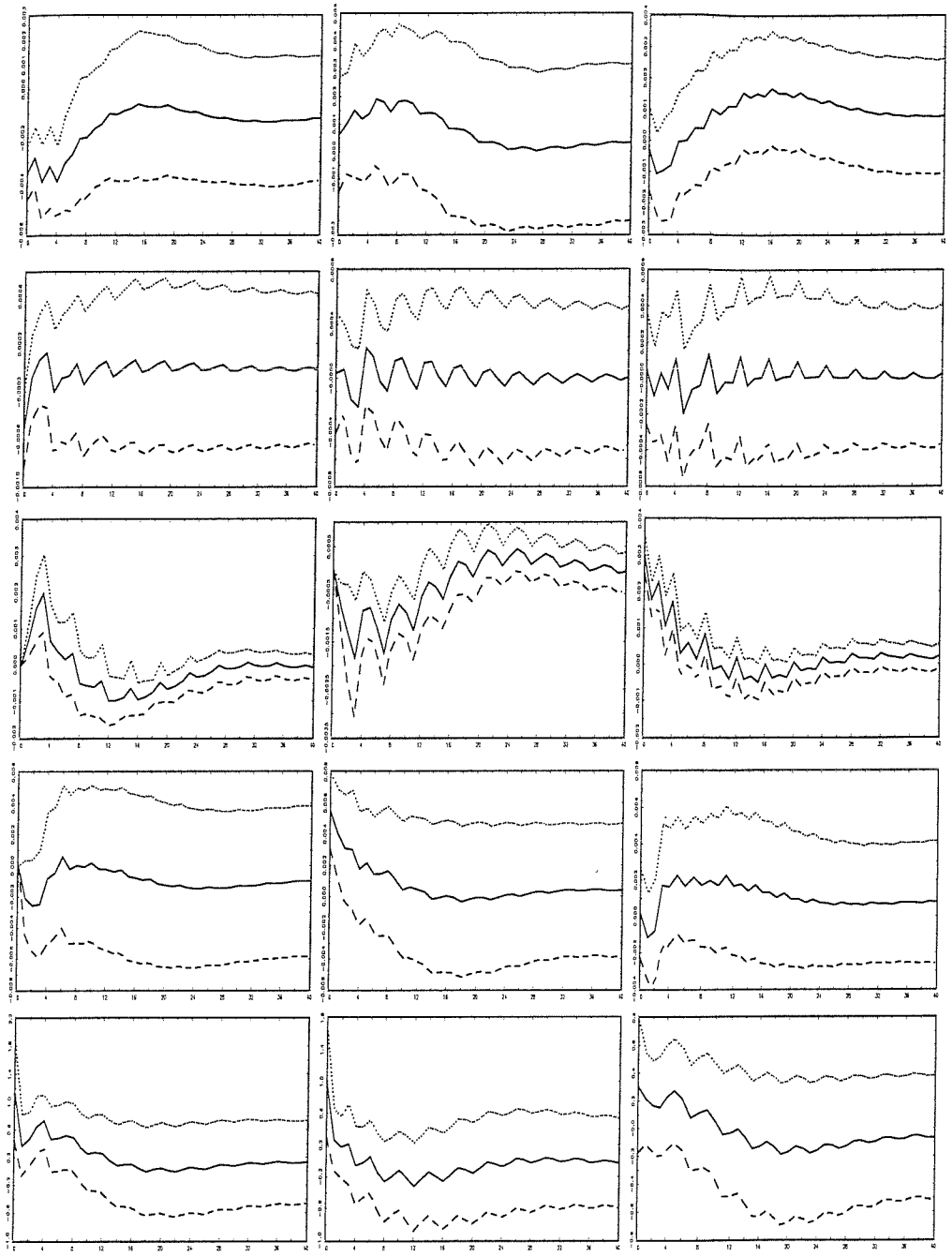

Figure 4.16. US: Impulse response function with 95% confidence interval from a one-standard deviation shock to unemployment innovation, wages innovation, and user cost of capital innovation. Response in $(y_t - n_t)$, $(\pi_t - k_t), u_t, w_t, uc_t$ respectively.

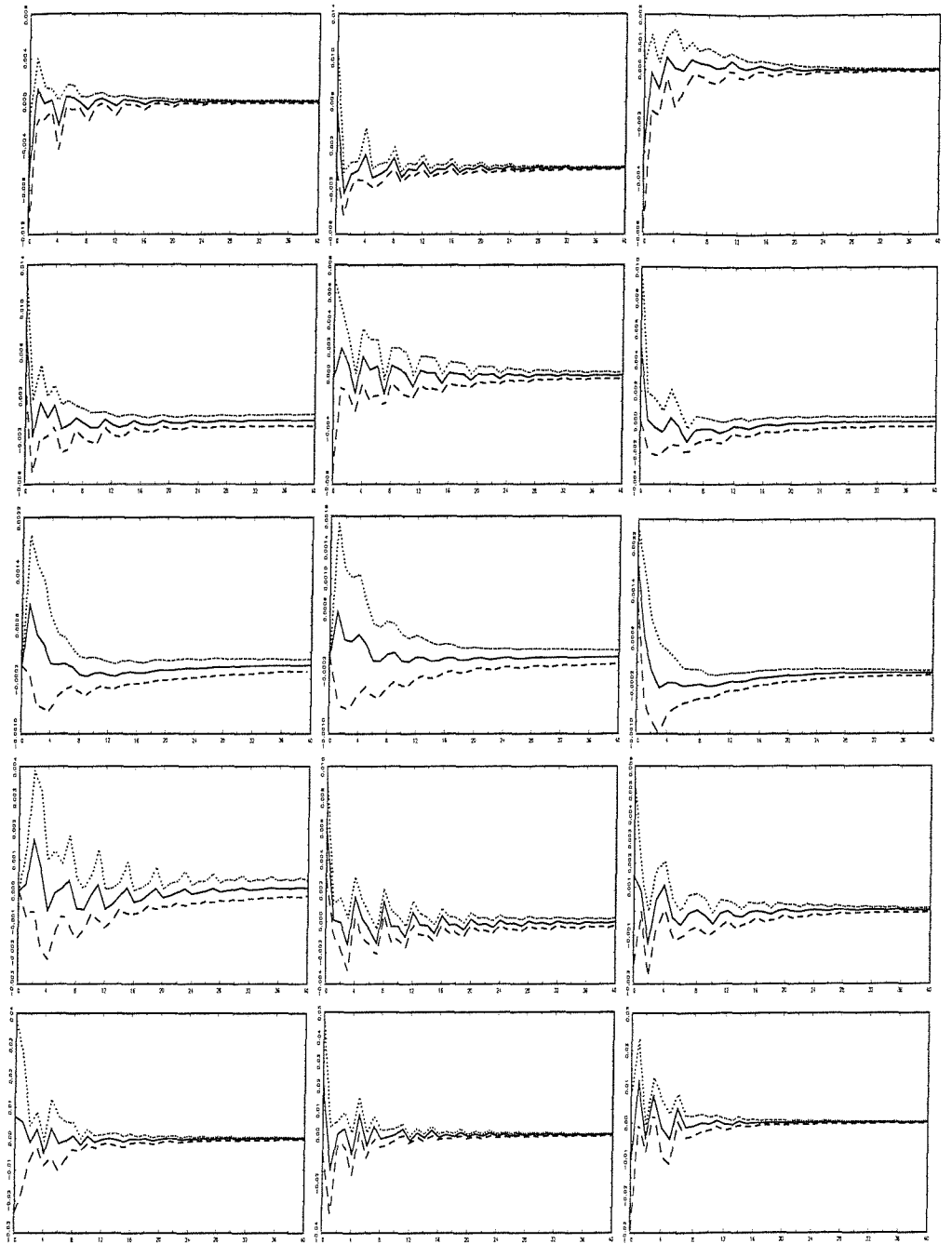

Figure 4.17. CANADA: Impulse response function with 95% confidence interval from a one-standard deviation shock to unemployment innovation, wage innovation, and user cost of capital innovation. Response in $(y_t - n_t)$, $(\pi_t - k_t), u_t, w_t, uc_t$ respectively.

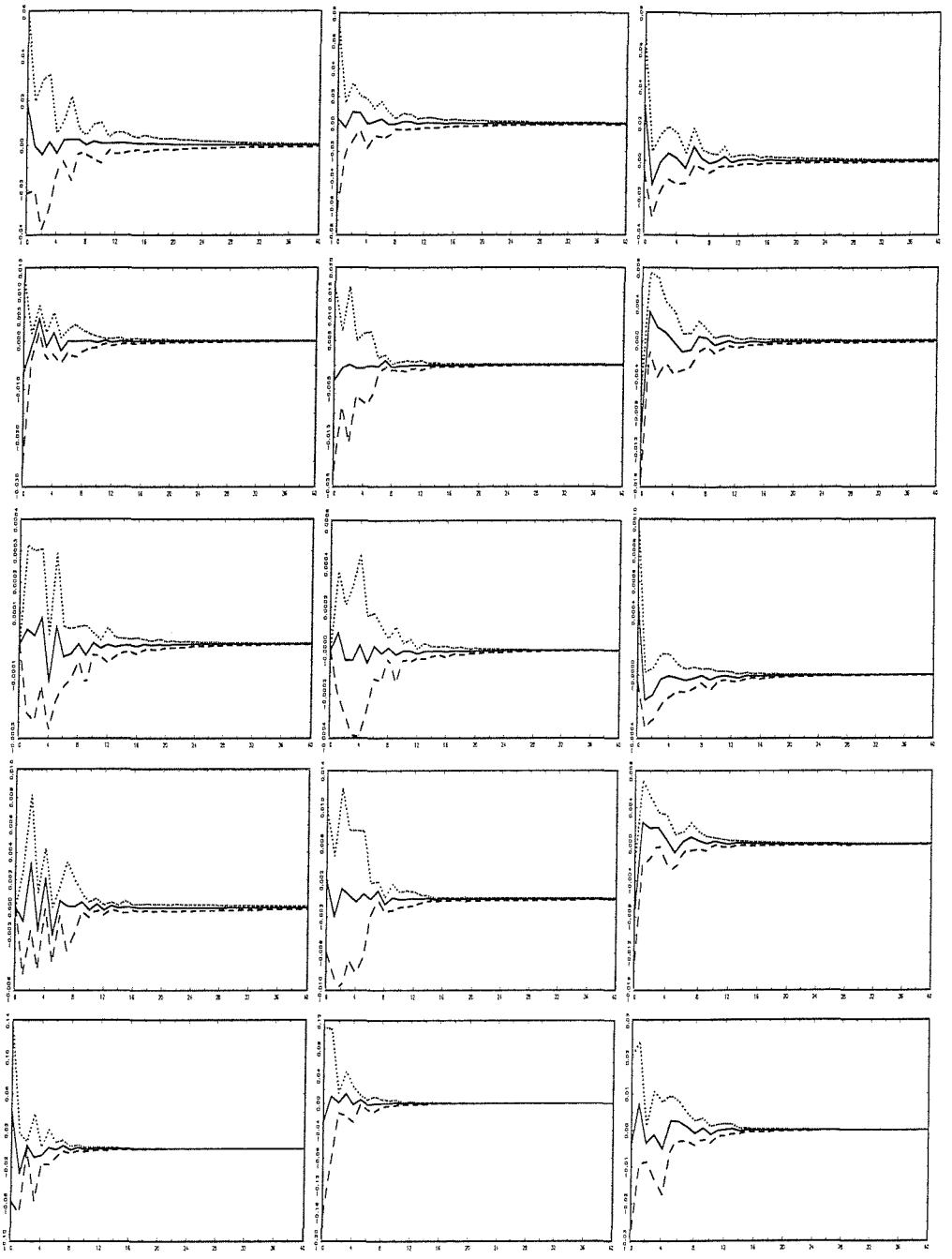

Figure 4.18. JAPAN: Impulse response function with 95% confidence interval from a one-standard deviation shock to unemployment, wage innovation and user cost of capital innovation. Response in $(y_t - n_t)$, $(\pi_t - k_t), u_t, w_t, uc_t$ respectively.

4.5 Concluding remarks

In this chapter we analyse the G7 labour market using a structural cointegrated VAR model. In the literature a lot of work has been done on this subject. However, much of the econometric research on the increase in unemployment has taken the form of estimation across countries and time of a reduced form equation for the unemployment rate as a function of a number of variables. In this chapter we suggests a different approach which involves identifying separately supply and demand shocks, and we try to explain the effect of each of these shocks separately. Moreover, the econometric procedure we use allows us to distinguish between the effects of short run shocks (i.e. deviation from the equilibrium relationships) and long run shocks (i.e. structural shocks) to unemployment. With respect to other related works, the novelty of this work is that we model directly some of the variables which affect firms' investment decisions, such as the profit rate and the factor prices, even thought we include in our model the variables that are traditionally regarded as the main "culprit" for high unemployment rate (i.e. wages rigidities and productivity shocks). In the light of the tight monetary policy which characterized the policy decisions in particular starting from the 80s, it seems to us that focusing only on the "labour market rigidities" is able to explain only part of the story. In particular, it fails to explain why the result of these policy decisions translates to high unemployment rate in the European countries, but did not increase unemployment rate in the UK and the US.

As far as the results are concerned we find that wages and real interest rate shocks are the most important source of uncertainty for unemployment. However, there are pronounced differences among the G7 countries.

4.6 Appendix A: Some methodological notes

Let us consider the following the $VAR(1)$ model

$$\Delta x_t = \mu + \Pi x_{t-1} + \varepsilon_t, \tag{4.6.1}$$

under the assumption of cointegration we can express the matrix Π as

$$\Pi = \alpha\beta'$$

so (4.6.1) can be written as

$$\Delta x_t = \mu + \alpha\beta' x_{t-1} + \varepsilon_t, \tag{4.6.2}$$

which can be rewritten as

$$x_t = \mu + (I_n + \alpha\beta')\, x_{t-1} + \varepsilon_t.$$

Premultiplying this system by β' give us a $VAR(1)$ model for the cointegration relations.

$$\begin{aligned} \beta' x_t &= \beta'\mu + \beta'\,(I_n + \alpha\beta')\, x_{t-1} + \beta'\varepsilon_t \\ &= \beta'\mu + (I_r + \beta'\alpha)\,\beta' x_{t-1} + \beta'\varepsilon_t, \end{aligned}$$

so that

$$(I_r - (I_r + \beta'\alpha)L)\,\beta' x_t = \beta'\mu + \beta'\varepsilon_t$$

where L stands for the lag operator. Since the matrix $(I_n + \alpha\beta')$ has all eigenvalues inside the unit circle, then the polynomial $(I_r - (I_r + \beta'\alpha)L)$ is invertible. It follows that

$$\beta' x_t = (I_r - (I_r + \beta'\alpha)L)^{-1}\,(\beta'\mu + \beta'\varepsilon_t)\,,$$

and using the formula for the sum of a geometric series we get

$$\begin{aligned}\beta' x_t &= \sum_{i=0}^{\infty} (I_r + \beta'\alpha)^i L^i (\beta'\mu + \beta'\varepsilon_t) \\ &= \sum_{i=0}^{\infty} (I_r + \beta'\alpha)^i \beta'\mu + \sum_{i=0}^{\infty} (I_r + \beta'\alpha)^i \beta'\varepsilon_{t-i} \\ &= -(\beta'\alpha)^{-1} \beta'\mu + \sum_{i=0}^{\infty} (I_r + \alpha\beta')^i \beta'\varepsilon_{t-i},\end{aligned} \tag{4.6.3}$$

so that we have found an $MA(\infty)$ representation for the r cointegrating relations.

Substituting equation (4.6.3) for $\beta' x_{t-1}$ in equation (4.6.1) we have found the MA representation for Δx_t, given by

$$\begin{aligned}\Delta x_t &= \mu + \alpha \left[-(\beta'\alpha)^{-1} \beta'\mu + \sum_{i=0}^{\infty} (I_r + \beta'\alpha)^i \beta'\varepsilon_{t-i-1} \right] + \varepsilon_t \\ &= \left(I_n - \alpha (\beta'\alpha)^{-1} \beta' \right) \mu + \varepsilon_t + \sum_{i=1}^{\infty} \alpha (I_r + \beta'\alpha)^{i-1} \beta'\varepsilon_{t-i} \\ &= \xi + \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}\end{aligned} \tag{4.6.4}$$

where $C_0 = I_n$.

To show that $C(z) = \sum_{i=0}^{\infty} C_i z^i$ has unit roots, note first that

$$\begin{aligned}C(1) &= I_n + \sum_{i=0}^{\infty} \alpha (I_r + \beta'\alpha)^i \beta' \\ &= I_n - \alpha (\beta'\alpha)^{-1} \beta'.\end{aligned}$$

Second, Johansen (1995) p.39 shows that for any $\alpha_\perp, \beta_\perp \in R^{n\times(n-r)}$ of rank $(n-r)$ such that $\alpha_\perp \alpha = 0$ and $\beta_\perp \beta = 0$ it holds that

$$I_n = \beta_\perp (\alpha_\perp' \beta_\perp)^{-1} \alpha_\perp' + \alpha (\beta'\alpha)^{-1} \beta'.$$

This can be verified in part through premultiplication of both sides by $\alpha_\perp$ or β' or through post-multiplication by α or $\beta_\perp$. The choice of the basis is irrelevant since $\alpha_\perp^* = \alpha_\perp \zeta$, $\beta_\perp^* = \beta_\perp \vartheta$ (where ζ and ϑ are nonsingular $(n-r) \times (n-r)$ matrices) satisfy

$$\beta_\perp^* \left(\alpha_\perp^{*\prime} \beta_\perp^*\right)^{-1} \alpha_\perp^{*\prime} = \beta_\perp \left(\alpha_\perp' \beta\right)^{-1} \alpha_\perp'.$$

Thus $C(1) = \beta_\perp \left(\alpha_\perp' \beta\right)^{-1} \alpha_\perp'$ and $\left(I_n - \alpha \left(\beta' \alpha\right)^{-1} \beta'\right) = C(1) = C$. In Johansen (1995), *Lemma 4.1* (p. 47), it is stated that the $C(z)$ matrix polynomial can be expressed as

$$C(z) = C + (1-z) C^*(z),$$

so that $C(1) = C$, and substituting for $C(z)$ in equation (4.6.4) we get

$$\begin{aligned}
x_t - x_{t-1} &= C\mu + C\varepsilon_t + \sum_{j=0}^{\infty} C_j^* \left(\varepsilon_{t-j} - \varepsilon_{t-j-1}\right) \Rightarrow \qquad (4.6.5)\\
\sum_{i=1}^{t} \Delta x_t &= \sum_{i=1}^{t} (C\mu + C\varepsilon_t) + \sum_{j=0}^{\infty} C_j^* \sum_{i=1}^{t} \left(\varepsilon_{t-j} - \varepsilon_{t-j-1}\right)\\
x_t - x_0 &= C\mu t + C \sum_{i=1}^{t} \varepsilon_t + \sum_{j=0}^{\infty} C_j^* (\varepsilon_{t-j} - \varepsilon_0)\\
x_t &= \bar{x}_0 + C\mu t + C \sum_{i=1}^{t} \varepsilon_t + \sum_{j=0}^{\infty} C_j^* \varepsilon_{t-j}.
\end{aligned}$$

where

$$\bar{x}_0 = x_0 + \sum_{j=0}^{\infty} C_j^* \varepsilon_0$$

So that we have found that the MA representation contains:

(i) an $I(1)$ component: $C\left(\mu t + \sum_{i=1}^{t} \varepsilon_t\right)$;

(ii) an $I(0)$ component: $\left(\sum_{j=0}^{\infty} C_j^* \varepsilon_{t-j}\right)$;

(iii) an initial values denoted by $\bar{x}_0$.

The $I(1)$ component of the MA representation can also be expressed as

$$C\left(\mu t+\sum_{i=1}^{t}\varepsilon_t\right)=\beta_{\perp}\left(\alpha_{\perp}'\beta\right)^{-1}\left(\alpha_{\perp}'\mu t+\sum_{i=1}^{t}\alpha_{\perp}'\varepsilon_t\right).$$

To illustrate this consider the following process

$$\begin{bmatrix}\Delta x_t\\ \Delta y_t\end{bmatrix}=\begin{bmatrix}\mu_y\\ \mu_y\end{bmatrix}+\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}u_t\\ w_t\end{bmatrix}+\begin{bmatrix}0 & 0\\ 1 & -1\end{bmatrix}\begin{bmatrix}u_{t-1}\\ w_{t-1}\end{bmatrix},$$

or

$$\Delta x_t=\delta+(I+CL)\,\varepsilon_t$$

Suppose that $C(z)$ has reduced rank, that is

$$C(1)=\begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}.$$

The $I(1)$ component of the MA representation can be expressed as

$$\begin{aligned}
x_t^p &= \begin{bmatrix}\mu_y\\ \mu_y\end{bmatrix}t+\begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}\sum_{i=1}^{t}\begin{bmatrix}u_t\\ w_t\end{bmatrix}\\
&= \begin{bmatrix}1\\ 1\end{bmatrix}\left(\mu_y t+\sum_{i=1}^{t}u_i\right)\\
&= \begin{bmatrix}1\\ 1\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}\mu_y\\ \mu_y\end{bmatrix}t+\begin{bmatrix}1\\ 1\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix}\sum_{i=1}^{t}\begin{bmatrix}u_t\\ w_t\end{bmatrix}\\
&= \beta_{\perp}\alpha_{\perp}'\mu t+\beta_{\perp}\alpha_{\perp}'\sum_{i=1}^{t}\varepsilon_t\\
&= \beta_{\perp}\alpha_{\perp}'\left(\mu t+\sum_{i=1}^{t}\varepsilon_t\right).
\end{aligned}$$

The term $\alpha_{\perp}'\left(\mu t+\sum_{i=1}^{t}\varepsilon_t\right)$ represent the common trend. So, in general without loss of generality we can rewrite (4.6.1) as

$$x_t=x_t^p+x_t^s,$$

where

$$x_t^p = C\left(\mu t + \sum_{i=1}^{t} \varepsilon_t\right),$$
$$x_t^s = \bar{x}_0 + \sum_{j=0}^{\infty} C_j^* \varepsilon_{t-j}.$$

It is important to note that if β is a cointegrating vector, then $\beta' x_t^p = 0$ for $\beta' x_t = \beta' x_t^s$ to be stationary.

4.7 Appendix B: The bootstrap experiment

In this appendix we describe the simulation experiment for the rank tests. For ease of notation we report the $VECM$ given in equation (4.3.3) in Section 4.3

$$\Gamma(L)\Delta x_t = \mu + \Pi x_{t-k} + \epsilon_t \tag{4.7.1}$$

where μ is an intercept; x_t, and ϵ_t are $(n \times 1)$ vectors, $\Gamma(L)$ is an $(n \times n)$ matrix polynomial in the lag operator L is the lag operator, $\Pi = \alpha\beta'$, $\Delta x_t = x_t - x_{t-1}$,and $\epsilon_t \backsim NID(0, \Sigma)$. The matrix Π determines whether or not, and to what extent, the system (5.7.1) is cointegrated. Suppose that Π has reduced rank r, the hypothesis of r cointegrating vectors β can be written as:

$$H_0 : \Pi = \alpha\beta',$$

where α and β are $(n \times r)$ matrices. As we have seen in Section 4.3, the rows of β' can be interpreted as the distinct cointegrating vectors of x_t (i.e. such that the linear combinations $\beta' x_t$ are $I(0)$) and the elements of α represent the weights of each of these r cointegrating relations in the n component equations (4.7.1).

A test for the number r of cointegrating vectors can be based on the n eigenvalues $\hat{\lambda}_1 > \ldots > \hat{\lambda}_n > 0$. As seen in Chapter 2 a likelihood ratio (LR) test of the hypothesis that there are at most r cointegration vectors by testing that the $(n-r)$ smallest eigenvalues $\lambda_{r+1}, \ldots, \lambda_n$ are zero against the assumption that $\lambda_i \geq 0$ for $i = 1, \ldots, n$. The LR test statistic for this is known as the trace tests, defined as

$$LR(trace)_r = -T \sum_{i=r+1}^{n} \ln\left(1 - \hat{\lambda}_i\right).$$

In addition, the maximum eigenvalue test statistic is given by

$$LR(\max)_r = -T \ln\left(1 - \hat{\lambda}_{r+1}\right)$$

and can be used to test the null $H_0(r) : rank(\Pi) = r$ against the alternative $H_1(r+1) : rank(\Pi) = r + 1$.

The bootstrap can be used to approximate the finite sample distribution of the $\lambda_{\max}$ and trace statistic under the null. The idea is to approximate the finite sample distribution of the $\widehat{LR}$ tests by drawing B bootstrap realizations $\left\{\widehat{LR_i}^*\right\}$ for $i = 1, 2, ..., B$ bootstrap samples $\left\{\left(\Delta x^*, x^*_{t-1}\right)_i\right\}$. The bootstrap algorithm we use can be summarised as follows:

1) Estimate the error correction model given by $(A2.1)$ and compute $\widehat{LR}$.

2) Resample the residual from $(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_t)$ independently with replacement to obtain a bootstrap sample $(\varepsilon_1^*, ..., \varepsilon_t^*)$. Generate the bootstrap sample $(x_1^*, ..., x_t^*)$ recursively from $x_0 = 0$ and $(\varepsilon_1^*, ..., \varepsilon_t^*)$ using the estimated restricted model

$$\Delta x_t^* = \hat{\mu} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i}^* + \tilde{\alpha}\tilde{\beta}' x_{t-1}^* + \varepsilon_t^*$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ denote the estimates under the null hypothesis.

4) Compute the bootstrap replication of $\left\{\widehat{LR}^{*}\right\}$,using $(x_1^*, ..., x_t^*)$

5) Get the distribution of the trace and the λ_{max} tests under the null by repeating steps 2-4 B times.

The p-value is given by

$$p^*\text{-}value = \frac{\#(\widehat{LR}^{*} > \widehat{LR})}{B+1}$$

where $\#\left(\widehat{LR_i}^{*} > \widehat{LR}\right)$ indicate the number of occurrences of eventuality $\left(\widehat{LR_i}^{*} > \widehat{LR}\right)$ over B bootstrap replications.

Chapter 5
Summary and Conclusions

There are five chapters in this thesis. Chapter 1 contains a brief survey of bootstrap inference procedures in econometric models. Chapters 2-4 contain the main body of the research. In this chapter the main results of this work are summarized.

Chapter 1 is an introductory chapter where the fundamental concepts of the bootstrap method are highlighted. The Chapter starts with an introduction of the bootstrap principle, and after discussing the first-order asymptotic properties of the bootstrap the higher order properties are considered in some detail. We then introduce a number of different bootstrap procedures used in the context of time series models. In particular, we distinguish between the residual based bootstrap which requires assuming a particular specification of the model in use, and techniques such as the block bootstrap and the stationary bootstrap which generate the bootstrap observation from the observed time series directly (i.e. without assuming a particular model specification). Throughout the chapter, empirical applications are provided to illustrate the methods and their applicability.

The purpose of Chapter 2 is twofold. Firstly, we use the bootstrap hypothesis testing as a way to reduce the size distortion of the tests for linear restrictions on the cointegrating space. Secondly, we consider the Johansen LR and Wald test statistics as well as the small sample corrected version of these tests, and we explore the robustness of the inference procedure in a situation where we allow for potential over-fitting and under-fitting of the number of cointegrating vectors included in the restricted model.

As far as the results are concerned we find that when the number of cointegrating relationships is correctly specified inference based on first-order asymptotic critical values for Wald and the LR statistics is markedly inaccurate. This is particularly true for the Wald statistic for which, when $T = 50$, the empirical size can be 3 times as large as the nominal level. The LR test performs better, but the overall impression is that the asymptotic theory is uniformly satisfactory only for $T \geq 200$, which is a sample size well above the sample size generally available to practitioners. Psaradakis's small sample corrected LR and Wald statistics have smaller size distortion than the uncorrected version of these tests. However, for all sample sizes, the empirical sizes that the F-type tests and the bootstrap test deliver are much closer to the 5% nominal size of the test. Turning to the power properties of the tests we find that for $T \geq 150$, both the small sample corrected and the bootstrap test have slightly lower 'power' (rejection frequencies) than the tests based on first-order asymptotic critical values. The picture changes when we come to the misspecified model. In this case we find that: (i) when the cointegrating rank is overfitted the size distortion of the tests is so large that it calls into question the use of the tests, since we find sizes over 30%, (ii) when the number of cointegrating vectors is underfitted the size distortion of the tests asymptotically vanishes, but the power-loss in this case is substantial, (iii) when the model is misspecified using the small sample corrected tests or the bootstrap test does not help, since the power of both these procedures mimic the behavior of the asymptotic tests.

Chapter 3 is closely related to Chapter 2. In the first part of Chapter 3 we propose approximating the finite sample expectation of the LR test statistic using the bootstrap and we compare the finite sample properties of the asymptotic, the bootstrap, and the bootstrap

Bartlett corrected likelihood ratio test. The Monte Carlo evaluation of the bootstrap and the bootstrap Bartlett corrected LR tests deliver remarkably accurate inference for the test statistics considered. Furthermore, the evaluation of the power reveals that the power of the bootstrap, and bootstrap Bartlett corrected likelihood, is almost as good as the asymptotic power, although in some situations the bootstrap Bartlett corrected LR test shows higher power than the bootstrap test. In the second part of the chapter we propose bootstrapping the Bartlett corrected likelihood ratio test, but in this case the Bartlett correction is calculated analytically using the correction factor proposed by Johansen (1999). According to theoretical arguments in Beran (1988) this procedure may produce an error of rejection probability of order $O(T^{-2})$, which is considerably smaller than the error of conventional first order approximation. The simulation results reveal that the bootstrap procedure works remarkably well, although the response surface analysis reveals that the size distortion of the test heavily depends on the parameter space values: there are regions of the parameter space were the usual asymptotic χ^2 approximation works reasonably well, whereas there are parameters points close to the boundary of the parameter space where the distribution of the LR test is very sensitive to the parameter values. In this case the first order approximation is quite inaccurate, as is the Bartlett corrected LR test.

Chapter 4 is a self contained chapter where an empirical application of the bootstrap test is undertaken using real data instead of the simulated ones. The idea is to analyse the effects of macroeconomic shocks on unemployment, and in particular the effects of shifts in labour supply and labour demand on the rise of European unemployment. The econometric model considered is a structural VAR with cointegrated constraints. This model allows us

to distinguish between the effects of transitory and permanent shocks to unemployment. Inference about the cointegrating rank is, once again, carried out using Johansen's (1988) procedure. To improve the robustness of our inference we also used the non-parametric bootstrap. Turning to the results, we find that the conclusions about the cointegrating rank are in agreement with Reimers's small-sample corrected tests, while the Johansen tests in some cases tend to over-estimate the number of cointegrating vectors.

In the various chapters, we come to the conclusion that the bootstrap can successfully eliminate the size distortion problem of the test statistics employed in cointegrated models without involving substantial loss in power. We were primarily interested in the problem of reducing the error in rejection probability of the asymptotic tests, so we only investigated the performance of the residual-based bootstrap. However, the residual-based bootstrap assumes that the dynamics of the VAR model is correctly specified. Under uncertainty with respect to the model specification, other bootstrap procedures such as the stationary bootstrap, seem to be more suitable than the residual-based bootstrap. With this in mind it would be interesting to extend the results in Chapter 2 and 3 to more complicated DGPs, using for example block bootstrap procedures and perhaps analysing the robustness of our conclusions to misspecification in the underspecification or overspecification of the dynamics of the VAR model. Moreover, throughout this thesis in our experiment design we only control the number of cointegrating vectors in the DGP and the sample size: it would be interesting to extend the analysis by controlling the dimension of the VAR model.

Nonetheless, the analysis in this thesis demonstrates that, provided that it is carefully executed, the bootstrap offers a promising option for conducting relatively accurate inference in cointegrated models.

Bibliography

Andrade, I.C., O'Brien, R.J. and Podivinsky, J.M. (1994). "Cointegration Tests and Mean Shifts". *University of Southampton Discussion Paper N. 9405.*

Akaike, H. (1969). "Fitting Autoregressive Model for Prediction". *Annals of the Institute for Statistical Mathematics*, 21, 243-247.

Alogoskoufis, G. S. and Manning, A. (1988). "Wage Setting and Unemployment Persistence in Europe, Japan, and USA". *European Economic Review*, 32, 698-706.

Babu, G.J and Singh, K.(1984). "On One-Term Edgeworth Correction by Efron's Bootstrap". *Sankhya, Ser. A.*, 46, 219-232.

Basawa, I. V., Mallik, A. K. , McCormick, W.P. and Taylor, R.L. (1989). "Bootstrapping explosive autoregressive processes". *Annals of Statistics* 17, 1479-1486.

Basawa, I. V., Mallik, A. K. , McCormick, W.P. and Taylor, R.L. (1991a). "Bootstrapping Unstable First Order Autoregressive Processes". *Annals of Statistics,* 19, 1098-1101.

Basawa, I. V., Mallik, A. K. , McCormick, W.P. and Taylor, R.L. (1991b). "Bootstrap Test of Significance and Sequential Bootstrap Estimation for Unstable First Order Autoregressive Processes". *Communication in Statistics: Theory and Methods*, 20, 1015-1026.

Bartlett, M.S. (1937). "Properties of Sufficiency and Statistical Tests". *Proceedings of the Royal Society*, 160, 263-268.

Bean, C. (1994). "European Unemployment: a Survey". *Journal of Economic Literature*, 32, 573-619.

Beran, R. (1982). "Estimated Sampling Distributions: The Bootstrap and Competitors". *The Annals of Statistics,* 10, 212-225.

Beran, R. (1988). "Prepivoting Test Statistic: A bootstrap View of Asymptotic Refinements". *Journal of the American Statistical Association*, 83, 687-697.

Berkowitz, J. and Kilian, L.K. (2000). "Recent Developments in Bootstrapping Time Series". *Econometric Reviews*, 19, 1-48.

Beveridge, S. and Nelson, C. (1981). "A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components with Particular Attention to Measurement of Business cycle". *Journal of Monetary Economics*, 7, 151-74.

Bickel, P. and Freedman, D.A. (1981). "Some Asymptotic Theory for the Bootstrap". *The Annals of Statistics*, 9, 1196-1217.

Blanchard, O. (1997). "The Medium Run". *Brookings Papers on Economic Activity, 2, 7-57.*

Blanchard, O. (1998)."Revising European Unemployment. Unemployment, Capital Accumulation, and Factor Prices". Geary Lecture, ESRI (June 1998).

Bose, A. (1988). "Edgeworth Correction by Bootstrap in Autoregressions". *Annals of Statistics*, 16, 1709-1722.

Bose, A. (1990). "Bootstrap in Moving Average Models". *The Annals of the Institute of Statistical Mathematics*, 42, 753-768.

Bühlmann, P. (1997). "Sieve Bootstrap for Time Series". *Bernulli*, 3, 123-148.

Caner, M. and Hansen, B. (2001). "Threshold Autoregression with a Unit Root". *Econometrica. 6,1555-1596.*

Carlstein, E. (1986). "The Use of Subseries Values for Estimating the Variance of a General Statistics from a Stationary Sequence". *The Annals of Statistics*, 14, 1171-1194.

Cheung, Y. W. and Lai, K.S. (1993). "Finite Sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration". *Oxford Bulletin of Economics and Statistics*, 55, 313-28.

Choi, E. and Hall, P. (2000). "Bootstrap Confidence Regions Computed from Autoregressions of Arbitrary Order". *Journal of Royal Statistics Society*, series B, 62, 461-477.

Cox, D.R. and Reid, N. (1987). "Approximations to Non-Central Distributions". *Canadian Journal of Statistics*, 15, 105-114.

Datta, S. (1996). "On Asymptotic Properties of Bootstrap for $AR(1)$ Processes". *Journal of Statistical planning and inference*, 53, 361-374.

Davidson, R. and MacKinnon, J.G. (1996a). "The Size and Power of Bootstrap Tests". *Discussion paper, Queen's University, Kingston*, Ontario.

Davidson, R. and MacKinnon, J.G. (1996b). "The Power of Bootstrap Tests". *Discussion paper Queen's University, Kingston*, Ontario.

Davidson, R. and MacKinnon, J.G. (1998). "The Size Distortion of Bootstrap Tests". *Econometric Theory*, 15, 361-376.

Doornik, J. and Hendry, D. (1994). *PcFiml 8.0 Interactive Econometric Modelling of Dynamics systems*. Institute of Economics and Statistics, University of Oxford.

Doornik, J. and O'Brien, R. J. (2002). "Numerically Stable Cointegration Analysis". Forthcoming in *Computational Statistics and Data.*

Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife" *The Annals of Statistics, 7, 1-26.*

Efron, B. (1981). "Censored Data and Bootstrap". *Journal of American Statistical Association,* 76, 312-319.

Fachin, S. (1997). "Bootstrap and Asymptotic Tests of Long-Run Relationships in Cointegrated Systems". *Discussion paper, University of Rome*

Ferretti, N. and Romo, J. (1996). "Unit Root Test for $AR(1)$ Models". *Biometrika*, 83, 849-860.

Fisher, N.I. and Hall, P. (1990). "On Bootstrapping Hypothesis Testing". *Australian Journal of Statistics*, 84, 996-1002.

Fitoussi, J. and Phelps E. (1988). "*The Slump in Europe*". Oxford & New York: Basil Blackwell.

Freedman, D.A. (1981). "Bootstrapping Regression Models". *The Annals of Statistics*, 9, 1218-1228.

Gredenhoff, M., Jacobson, T. (1998). "Bootstrap Approximate Finite Sample Distributions for Tests of Linear Restrictions on Cointegrating Vectors". *Discussion paper, Stockholm School of Economics.*

Gregory, A.W. and Veall, M.R. (1985). "Formulating Wald Tests of Non-Linear Restrictions". *Econometrica*, 53, 1465-8.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* New York, Springer-Verlag.

Hannan, H. and Quinn B.G. (1979). "The Determination of the Order of an Autoregression". *Journal of Statistics*,16, 927-953.

Hansen, B. E. (1996). "Inference when a Nuisance Parameter is not Identified Under the Null Hypothesis". *Econometrica*, 64, 413-500.

Harris, R.I.D. (1992). "Small Sample Testing for Unit Roots". *Oxford Bulletin of Economics and Statistics*, 54, 615-625.

Hillier, G.H. (1985). "On the Joint and Marginal Densities of Instrumental Variables Estimators in a General Structural Equation". *Econometric Theory*, 1, 53-72.

Horowitz, J.L. (1994). "Bootstrap-Based Critical Values for the Information Matrix Test", *Journal of Econometrics*, 61, 395-441.

Horowitz, J.L. (1997). "Bootstrap Methods in Econometrics: Theory and Numerical Performance", in D.M. Kreps and K.F. Wallis, Eds. *Advances in Economics and Econometrics: Theory and |Applications*, 7th World Congress of the Econometric Society Vol. III, Cambridge University Press.

Horowitz, J. L. (1999). "The Bootstrap" *Handbook of Econometrics, Vol 5*. North Holland.

Jacobson and Larsson (1996). "Bartlett Corrections in Cointegration Testing". *Discussion Paper University of Stockholm.*

Jacobson, T. Vredin, A. Warne, A. (1997). "Common Trends and Hysteresis in Scandinavian Unemployment". *European Economic Review,* 41, 1481-1816.

Jacobson, T. (1998). "Bootstrap Testing and Approximate Finite Sample Distributions for Tests of Linear Restrictions on Cointegrating Vectors". Paper Delivered at the European Meeting of the Econometric Society, Berlin.

Johansen, S. (1988). "Statistical Analysis of Cointegrated vectors". *Journal of Economic Dynamic and Control,* 12, 231-254.

Johansen, S. and Juselius, K. (1990). "Maximum Likelihood Estimation and Inference on Cointegration with Application to the Demand for Money". *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

Johansen, S. (1995). *Likelihood Inference in Cointegrated Vector Auto-Regressive Models.* Oxford University Press, Oxford.

Johansen, S., (1999). "A Bartlett Correction Factor for Tests on the Cointegrating Relations", European Institute - EUI Working Papers in Economics 99/10, European Institute - European Forum.

Johansen, S. (2000). "A Bartlett Correction Factor for Tests of on the Cointegrating Relations". *Econometric Theory*, 16, 740-778.

Kievet, M.L. (1984). "Bootstrapping Inference in Lagged Dependent Variable". *Working Paper University of Amsterdam.*

King, R., Plosser, C., Stock, J. (1987). "Stochastic Trends and Economics Fluctuations". *NBER Working Paper No. 2229.*

Kyndland, F. E. and Prescott, E. F. (1982). "Time to Build and Aggregate Fluctuations". *Econometrica*, 50, 936-56.

Kreiss, J.P. and Franke, J. (1992). "Bootstrapping Stationary Autoregressive Models". *Journal of Time Series Analysis*, 13, 297-317.

Kreiss, J.P. (1992). "Bootstrap Procedures for $AR(\infty)$ Processes" in K.H. Jöckel, G. Rothe, and W Sender (eds) *Bootstrapping and Related Techniques*, Lectures Notes in Economics and Mathematical Systems, 376 (Springer-Verlag, Heidelberg) pp. 107-113.

Künsch, H. R. (1991). "The Jackknife and the Bootstrap for General Stationary Observations". *The Annals of Statistics*, 17, 1217-1241.

Lafontaine, F. and White, K. J. (1986). "Obtaining any Wald Statistic you Want". *Economics Letters*, 21, 35-48.

Lahiri, S.N. (1999). "Theoretical Comparison of Block Bootstrap Methods". *Annals of Stastistics*, 27,386-404.

Layard, P.R.G. and Nickell, S.J. (1986). "Unemployment in Britain". *Economica, Supplement,* 53, 121-169.

Layard, P.R.G. and Nickell, S.J. Jackman R. (1991). *Unemployment Macroeconomic Performance and the Labour Market.* Oxford University Press, Oxford.

Lawley, A.N. (1989). "A General Method for Approximating to the Distribution of the Likelihood Ratio Criteria". *Biometrika*, 43, 555-563.

Li, H. and Xiao, Z. (2001). "Bootstrapping Time Series Regressions with Integrated Processes" *Journal of Time Series Analysis, 22,461-480.*

Li, H. and Maddala, G.S. (1997). "Bootstrapping Cointegrating Regressions", *Journal of Econometrics*, 80, 297-318.

Liu, R. Y. and Singh, K. (1992). "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence". In LePage, R. and Billard, L. (Eds), *Exploring the Limits of Bootstrap*, pp. 225-248. New York: John Wiley.

Mammen, E. (1992). *When Does the Bootstrap Work? Asymptotic Results and Simulations*. New York: Springer Verlag.

Marcellino, M. and Mizon, G. (1999). "Modelling Shifts in the Wage-Price and Unemployment Relationships in Italy, Poland, and the UK". *Discussion Papers in Economics and Econometrics No 9919. University of Southampton.*

Nielsen, B. (1998). "On the Distribution of Tests for Cointegrating Rank". *Mimeo*, Nuffield College

O'Brien, R.J. (1996). "The Curious Case of Spurious Cointegration". *Discussion Papers in Economics and Econometrics No 9405. University of Southampton.*

O'Brien, R.J. (1997). "Further Curiosa in Spurious Cointegration". *Discussion Papers in Economics and Econometrics No 9704.*

O'Brien, R.J. (1999). "The Curious Case of Spurious Cointegration". Paper Presented at the Far Eastern Meeting of Econometric Society, Singapore, 1999.

Osterwald-Lenum, M. (1992). "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistic: Four Cases". *Oxford Bulletin of Economics and Statistics*, 49,103-28.

Paparoditis, E. (1996). "Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes". *Journal of Multivariate Analysis*, 57,227-296.

Politis, D.N. and Romano, J.P. (1994). "The Stationary Bootstrap". *Journal of American Statistical Association*, 83, 111-115.

Podivinsky, J.M. (1992). "Small Sample Properties of Tests of Linear Restrictions on Cointegrating Vectors and Their Weights". *Economics Letters*, 39, 13-18.

Podivinsky, J.M. (1998). "Testing Misspecified Cointegrating Relationships". *Economics Letters* 60, 1-9.

Psaradakis, Z. (1994). "A Comparison of Tests of Linear Hypotheses in Cointegrated Vector Autoregressive Model". *Economics Letters,* 45, 137-144.

Psaradakis, Z. (2001). "On Bootstrap Inference in Cointegrating Regressions". *Economics Letters,* 72, 1-10.

Rayner, R.K. (1990). "Bartlett Correction and the Bootstrap in Normal Linear Regression Models". *Economic Letters.* 33, 255-258.

Reimers, H-E. (1992). "Comparison of Test for Cointegration". *Statistical Papers*, 33, 335-359.

Rocke, D.M. (1989). "Bootstrap Bartlett Adjustment in Seemingly Unrelated Regression". *Journal of the American Statistical Association*, 84, 598-601.

Singh, K. and Babu, G.J (1990). "On Asymptotic Optimality of the Bootstrap". *Journal of American Statistical Association*, 17, 1-9.

Stock J. and Watson, M. (1988). "Testing for Common Trends". *Journal of the American Statistical Association*, 83, 1097-107.

Van Giersbergen, N. (1996). "Bootstrapping the Trace Statistic in VAR Models: Monte Carlo Results and Applications". *Oxford Bulletin of Economics and Statistics*, 58, 391-408.

Veall, M.R. (1987). "Bootstrapping the Probability Distribution of Peak Electricity Demand". *International Economic Review,* 28, 203-212.

Veall, M.R. (1998). "Application of the Bootstrap in Econometrics and Economic Statistics", in *Handbook of Applied Economic Statistics*, Edited by Ullah A. and Giles, D.E.A.

Warne A. (1993). "A Common Trends Model: Identification, Estimation, and Inference". *Seminar paper no. 555, Institute of International Economics Studies, Stockholm University.*

White, J. S. (1961). "Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient". *Biometrika*, 48, 85-94.

Zaman, A. (1996). *Statistical Foundations for Econometric Techniques.* Academic Press, Inc.