Bootstrapping Nonstationary Time Series

Stephan Smeekes
Bootstrapping Nonstationary Time Series

PROEFSCHRIFT

der verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
  Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op donderdag 2 juli 2009 om 12.00 uur

door

Stephan Jacobus Maria Smeekes

Universiteit Maastricht
Promotores:
   Prof. dr. J.R.Y.J. Urbain
   Prof. dr. F.C. Palm

Beoordelingscommissie:
   Prof. dr. P.C. Schotman (voorzitter)
   Prof. dr. B. Candelon
   Prof. dr. A.R. Swensen (University of Oslo)

Dit onderzoek werd financieel mogelijk gemaakt door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) en de Maastricht Research School of Economics of Technology and Organizations (METEOR).
Acknowledgements

When I came to the end of my studies in 2004, I realized that I liked doing (theoretical) research very much, and therefore decided to pursue a Ph.D. degree. It has been a choice that I have never regretted since. Of course, sometimes you get stuck, or the methods you want to use just do not seem to work that well, and this can be very frustrating. However, the times of frustration have been greatly outnumbered by the pleasant times. Therefore I can say that in general it has been a very valuable and rewarding experience, but most of all it has been fun. I truly enjoyed the past five years and learned very much in this period. There are several people who contributed significantly to my enjoyment of the last five years, and I would like to thank some of them here.

First of all, I am very grateful to my supervisors Franz Palm and Jean-Pierre Urbain. It has truly been a privilege for me to work with such supervisors. Franz and Jean-Pierre always had time to discuss the latest progress and encouraged me to get the best out of myself. In our stimulating weekly meetings they managed to improve my work with ever constructive comments, challenging ideas and helpful contributions. I could not have wished for better supervision and I hope we can continue our collaboration in the future.

I would like to thank the members of the reading and assessment committee, Peter Schotman, Anders Swensen and Bertrand Candelon, for their careful reading of the manuscript. I thank all my colleagues from the Department of Quantitative Economics for the pleasant and stimulating working atmosphere. I want to thank Karin van den Boorn and Haydeé Hallmanns for all administrative assistance.

Special thanks go to Jeroen van den Berg, Rik Frehen, Christian Gengenbach, Kasper Leufkeus, Joyce van Loon and Jan Piplack; I thank them for their friendship, but also for making my life at the university for the last five years (nine for Jeroen, Rik and Joyce) a lot easier with their ever willingness to help and assist me. I really benefited from this tremendously and it has significantly smoothed my functioning at the university over these last five years.

I would also like to thank all colleagues, students and other people who have helped me now and then around the university, and who are simply too many to mention everyone by name. I would have had a considerably more difficult time...
without all the help.

Last but not least, I would like to thank my parents John and Connie and my sister Linda for their care and support over all these years. The fact that my parents accompanied me on all trips to conferences that I would not have been able to visit otherwise, and from which this thesis directly benefited, is but one of the many reasons to express my gratitude.

Stephan Smeekes
Maastricht, May 2009
# Contents

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>vii</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 The bootstrap</td>
<td>2</td>
</tr>
<tr>
<td>1.1.1 Asymptotic validity of the bootstrap</td>
<td>5</td>
</tr>
<tr>
<td>1.1.2 Asymptotic refinements of the bootstrap</td>
<td>7</td>
</tr>
<tr>
<td>1.2 The bootstrap in time series</td>
<td>8</td>
</tr>
<tr>
<td>1.2.1 The bootstrap for stationary time series</td>
<td>8</td>
</tr>
<tr>
<td>1.2.2 Bootstrapping nonstationary time series: contribution of the thesis</td>
<td>10</td>
</tr>
<tr>
<td><strong>2 Bootstrap Unit Root Tests: Comparison and Extensions</strong></td>
<td>13</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>2.2 The tests</td>
<td>15</td>
</tr>
<tr>
<td>2.2.1 DF sieve bootstrap test</td>
<td>15</td>
</tr>
<tr>
<td>2.2.2 ADF sieve bootstrap test</td>
<td>17</td>
</tr>
<tr>
<td>2.2.3 (A)DF block bootstrap test</td>
<td>19</td>
</tr>
<tr>
<td>2.2.4 DF stationary bootstrap test</td>
<td>22</td>
</tr>
<tr>
<td>2.2.5 Comparison of tests considered</td>
<td>23</td>
</tr>
<tr>
<td>2.3 Finite sample performance</td>
<td>26</td>
</tr>
<tr>
<td>2.3.1 Monte Carlo setup</td>
<td>26</td>
</tr>
<tr>
<td>2.3.2 Results</td>
<td>27</td>
</tr>
<tr>
<td>2.4 Conclusion</td>
<td>36</td>
</tr>
<tr>
<td>2.A Appendix: Proofs</td>
<td>36</td>
</tr>
<tr>
<td>2.A.1 Proof of Theorem 1</td>
<td>37</td>
</tr>
<tr>
<td>2.A.2 Proof of Theorem 2</td>
<td>41</td>
</tr>
</tbody>
</table>


## CONTENTS

### 3 Detrending Bootstrap Unit Root Tests

3.1 Introduction ........................................ 45
3.2 The model ........................................... 48
3.3 Test statistics ....................................... 48
  3.3.1 OLS / GLS detrending ............................ 49
  3.3.2 Recursive detrending ............................. 50
  3.3.3 Asymptotic properties ........................... 51
3.4 Bootstrap tests ....................................... 54
  3.4.1 Bootstrap algorithm ............................... 54
  3.4.2 Detrending within the bootstrap ................ 57
  3.4.3 Bootstrap validity ................................ 58
  3.4.4 Bootstrap tests under the alternative .......... 60
3.5 Simulations .......................................... 61
  3.5.1 Setup ........................................... 61
  3.5.2 Results .......................................... 62
3.6 Conclusion ........................................... 67
3.A Appendix: Proofs .................................... 70
  3.A.1 Proofs for Section 3.3 ........................... 70
  3.A.2 Proofs for Section 3.4 ........................... 84

### 4 A Sieve Bootstrap Test for Cointegration in a Conditional Error Correction Model

4.1 Introduction ........................................ 97
4.2 The model ........................................... 99
4.3 The bootstrap test and asymptotics ................ 101
  4.3.1 Test statistic .................................. 101
  4.3.2 Bootstrap method ................................. 102
  4.3.3 Asymptotic results ............................... 105
4.4 Simulations .......................................... 107
4.5 Deterministic components ............................ 112
4.6 Conclusion .......................................... 114
4.A Appendix: Proofs .................................... 116

### 5 Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests

5.1 Introduction ........................................ 137
5.2 Cross-sectionally dependent panels ................ 139
5.3 Bootstrap unit root tests in panels ............... 142
  5.3.1 Test statistics .................................. 142
  5.3.2 Bootstrap algorithm ............................. 143
5.4 Asymptotic properties ............................... 145
  5.4.1 Asymptotic properties under the main null hypothesis .... 145
  5.4.2 Asymptotic properties of the tests under cross-unit cointegration ........................................ 147
  5.4.3 Asymptotic properties under the alternative hypothesis .... 150
## CONTENTS

5.5 Small sample performance .................................. 153  
5.5.1 Monte Carlo design ........................................ 153  
5.5.2 Monte Carlo results ........................................ 155  
5.5.3 Block length selection .................................... 159  
5.6 Conclusion .................................................... 168  
5.A Appendix: Proofs ............................................. 168  

6 Conclusion 187  

Bibliography 191  

Nederlandse samenvatting 199  

Curriculum Vitae 203
Chapter 1

Introduction

The analysis of nonstationary time series is one of the major research topics in time series econometrics. Stationarity, which is the property that (certain aspects of) the joint distributions of subsets of the data do not change over time, is violated for many important (macro)economic time series.¹ Data on variables such as real GDP, inflation, exchange rates and stock markets give rise to one specific type of nonstationarity. They can be classified by the fact that their difference from one year to another, or their growth rate, is stationary. Loosely speaking, we call such data integrated time series, and their analysis will take up much of the remainder of this thesis. Alternatively, such data are said to contain an autoregressive unit root,² which is the terminology we will mainly employ in this thesis.³

If data are integrated, they contain a (stochastic) trend which is responsible for the behavior in the long run. It is therefore a very natural concept to describe macroeconomic variables. Integrated time series lead, statistically speaking, both to new problems and new opportunities for both theoretical and empirical research. The major problem is that most of the asymptotic theory developed for stationary time series no longer applies. The most interesting opportunity from an economic perspective, is the concept of cointegration (Engle and Granger, 1987). Two time series are said to be cointegrated if they share the same stochastic trend. The consequence is that the two series move together in the long-run, but not necessarily in the short-run. This interrelation provides an excellent description of many relations between economic variables, such as output and consumption.

¹Formally, we can distinguish between covariance stationarity and strict stationarity; see Davidson (2000, Section 4.4.1) for definitions.
²Let \(x_1, \ldots, x_n\) denote the time series of interest. The fact that the difference of year \(t-1\) to year is stationary can be expressed as

\[
x_t - x_{t-1} = u_t,
\]

where \(u_t\) is some stationary process. Defining \(L\) as the lag operator \(Lx_t = x_{t-1}\), we can write \((1-L)x_t = u_t\). It can now be seen that the lag polynomial \((1-L)\) has a root of one, which explains the name unit root. See Davidson (2000, Chapter 14) for a detailed exposition on unit roots.

³We usually leave out the part “autoregressive” whenever no confusion can arise.
Prize in Economics that Clive Granger (joint with Robert Engle) received in 2003 for his work on the analysis of nonstationary time series illustrates the importance of the topic for economic analysis.

This thesis considers methods to test for unit roots and cointegration. The contribution of this thesis lies in the development and improvement of tools for analyzing nonstationary time series using a different statistical technique than the one commonly used. The technique that will be studied throughout the dissertation to improve the tools for analyzing nonstationary time series, is the bootstrap. In a seminal paper, this statistical technique was formalized by Bradley Efron (1979). The name derives from the expression “to pull one self up by his own bootstraps” and ultimately from the famous tales of Baron von Münchhausen, who claimed that he pulled himself up out of a swamp by his own bootstraps.

One can derive two conclusions from this name. First, the name seems to indicate that, in one way or the other, bootstrapping works “on its own”, that nothing else is needed than the data themselves. This is indeed (partly) true, in a way that will be described in more detail below. Second, it gives the suggestion of cheating, of doing something that is not possible and should not be done. While this might appear to be true when one first looks at the bootstrap, it is definitely not the case, and hopefully after finishing this thesis, the sceptical reader will agree.

Let us now go into more detail. First the main principles of bootstrap will be explained. We explain these principles in the setting that the bootstrap was originally intended for, and we establish some general concepts that will return frequently in the following chapters. Next we will discuss how to apply the bootstrap to time series data. After discussing how the bootstrap can be applied to stationary time series, we will discuss the extension to nonstationary time series and link to the following chapters.

1.1 The bootstrap

The bootstrap is a method to perform statistical inference, and can be used as an alternative to classical asymptotic inference. Perhaps the most striking (although not defining) feature of the bootstrap is that it requires performing simulations in which one draws new samples based on the original data sample. The statistical properties of the original sample are then determined on the basis of these newly drawn samples, which are called the bootstrap samples. It is the ideas described above that give rise to the name bootstrap; one uses the sample and nothing else to perform inference. We will explain this in more detail below. Our exposition mainly follows Horowitz (2001), who provides an excellent survey on the use of bootstrap methods in econometrics.

---

4A different alternative to classical - or frequentist - statistics is Bayesian statistics. Bayesian statistics is based on a different philosophy than classical statistics. It will not be considered in this thesis, as all the analysis will be performed in a classical framework. As it relies on the same philosophy, the bootstrap can be classified as a frequentist device; although there also exists a Bayesian bootstrap (Rubin, 1981).
Suppose one has a sample of data $x_1,\ldots,x_n$. Let us assume that this sample is randomly drawn from a probability distribution with cumulative distribution function (CDF) $F(x) = P(x_i \leq x)$. This is obviously not true for economic time series data, and we will need to relax this assumption later. However, this is the setting for which the bootstrap was originally designed. Moreover, all of the concepts developed for the simple setting easily carry over to more general settings, but for expositional simplicity it is easier to discuss them in the simple framework.

Suppose we have some statistic of interest $T_n(x_1,\ldots,x_n)$ which is a function of the data. This statistic $T_n$ usually involves an estimator for some unknown parameter that depends on the distribution of the data $F$. The simplest example is the sample average as an estimator of the mean of the data, in which case $T_n = \frac{1}{n} \sum_{i=1}^{n} x_i$. As the statistic $T_n$ depends on random data, it is itself also random and therefore also has a distribution. This distribution of the statistic, $G_n(x) = P(T_n \leq x)$, usually depends on $F(x)$, the distribution of the data. To illustrate this dependence, we write $G_n(x,F)$. If $G_n(x,F)$ does not depend on $F$, hence if $G_n(x,F_1) = G_n(x,F_2)$ for any $F_1$ and $F_2$ within a class of CDFs $\mathcal{F}$, the statistic is called pivotal (for this class $\mathcal{F}$). Most statistics encountered in econometrics are not pivotal however. Usually the exact distribution of $T_n$ depends on the unknown $F$ and cannot be calculated. Therefore, $G_n$ must usually be approximated in one way or the other.

The most common way to approximate $G_n$ is to consider asymptotic theory, in which one lets $n \to \infty$. There are several results available for the setting $n \to \infty$ (usually involving some form of the central limit theorem) by which we can approximate the distribution $G_n$. Most asymptotic distributions of statistics ($G_\infty(x)$) do not depend on $F$. We call such statistics asymptotically pivotal. These distributions, which are free of nuisance parameters, are then used as approximation to the finite sample distributions. It is obvious that such approximations only make sense when the sample size $n$ is large. How large $n$ has to be however, depends on the situation. The sample sizes encountered in practice are often too small for the asymptotic distributions to be good approximations.

The alternative is to estimate $G_n$ using the bootstrap. Instead of replacing the unknown $G_n$ by the known $G_\infty$, bootstrapping replaces the unknown $F$ with an estimator $\hat{F}_n$. There are two main ways in which one can estimate $F$. The parametric bootstrap can be used if $F$ is supposed to be known up to certain parameters. For example, if $F$ is known to be normal, but the mean and standard deviation are unknown, $\hat{F}_n$ would be a normal distribution with estimated mean and standard deviation. The nonparametric bootstrap can be used without knowledge about $F$. In this case $F$ is estimated by the empirical distribution function (EDF): 

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

where $I(A)$ is equal to 1 if $A$ is true and zero otherwise. The nonparametric bootstrap will be the main focus of this paper, so we will restrict ourselves to it in the remainder.
Once one has obtained an estimate \( \hat{F}_n \), one can obtain \( G_n(x, \hat{F}_n) \). The distribution \( G_n(x, \hat{F}_n) \) is called the \textit{bootstrap distribution} and often denoted by \( G^*(x) \).

In general, bootstrap quantities are denoted using a superscript ‘*’ and we will follow this tradition in this thesis. Unfortunately, the distribution of the statistic is often too complex to obtain an analytical expression for \( G_n(x, \hat{F}_n) \). This is where the simulations come into play. New (pseudo) data can be simulated from the distribution \( \hat{F}_n \). Call such a sample of data the \textit{bootstrap sample} and denote it by \( x^*_1, \ldots, x^*_n \). Then, based on the frequentist notion of repeated sampling, \( G_n(x, \hat{F}_n) \) can be approximated by drawing many bootstrap samples, and for each calculating the bootstrap statistic \( T^*_n = T_n(x^*_1, \ldots, x^*_n) \).

If the EDF is chosen as \( \hat{F}_n \), drawing from \( \hat{F}_n \) means drawing from the original sample \( x_1, \ldots, x_n \) with replacement, as the EDF puts a mass of \( \frac{1}{n} \) on each observation. Let us call this method the i.i.d. bootstrap, as it can be applied to \textit{independent and identically distributed} (i.i.d.) data. Formally, the following steps are required.

\textbf{Bootstrap Algorithm 1.1 (i.i.d. bootstrap).}

1. Generate a bootstrap sample \( x^*_1, \ldots, x^*_n \) by randomly drawing with replacement from \( x_1, \ldots, x_n \).
2. Calculate \( T^*_n = T_n(x^*_1, \ldots, x^*_n) \).
3. Repeat steps 1 and 2 \( B \) times. Let \( T^*_{n,b} \) denote the bootstrap statistic obtained in the \( b \)-th replication. Estimate \( G_n(x, \hat{F}_n) = P^*(T^*_n \leq x) \) by

\[
\frac{1}{B} \sum_{i=1}^{B} I(T^*_{n,b} \leq x).
\]

The number of replications \( B \) is a trade-off between precision of the estimate and time consumption. The higher \( B \) is, the more precise the estimate of \( G^* \) will be, but also the more time consuming the calculations will be. At some point, increasing \( B \) will only lead to a minor increase in precision, not worth the amount of additional time needed anymore. Where exactly this point is depends on the application (how much precision is needed) and on the computer used to perform the calculations (how much time is needed). There is some theoretical work on the optimal number of bootstrap replications (Davidson and MacKinnon, 2000; Andrews and Buchinsky, 2000), but given the recent increase in computer power this is often not very relevant in practice anymore. Any relatively modern computer can easily perform a large enough number of replications without much trouble, unless the application to which the bootstrap is applied is very time-consuming.

\footnote{We will generally refer to simulations in a bootstrap setting as (bootstrap) replications, to distinguish from Monte Carlo simulations that will be discussed in later chapters.}
1.1 The bootstrap

What element of $G_n$ is needed depends on the type of application; one can think of bias reduction (point estimation), variance estimation, confidence intervals and hypothesis testing. The bootstrap can be used for all. In this thesis the focus will be exclusively on hypothesis testing. That is, one has a null hypothesis $H_0$ that one wants to test against an alternative hypothesis $H_1$. $H_0$ is rejected in favor of $H_1$ if the value of the test statistic is in a critical region specified by the person performing the hypothesis test.\textsuperscript{6} Bootstrap critical values are simply selected as the relevant quantiles of the bootstrap distribution (see the following chapters for more details). There is however one additional difficulty in applying the bootstrap to hypothesis testing, and that is that the bootstrap data must satisfy the null hypothesis, whether or not the actual data do. We will return to this point on several occasions in the following chapters.

We have described above how the bootstrap is performed. It is now time to turn to the question when and why we should use it.

1.1.1 Asymptotic validity of the bootstrap

The minimal condition that the bootstrap has to satisfy is called consistency or asymptotic validity. If this condition is not satisfied, using the bootstrap leads to incorrect results. There are several ways to define consistency, although all are essentially the same. Loosely speaking, we have bootstrap consistency if $G_n(x, \hat{F}_n) \overset{d*}{\longrightarrow} G_\infty(x, F)$ as $n \to \infty$. We need to discuss the method of convergence in somewhat more detail. This convergence consists of two parts; the first part is the convergence of $G_n$ to $G_\infty$, i.e. $G_n \overset{d}{\to} G_\infty$ as $n \to \infty$. This is the “standard” convergence in distribution, or weak convergence, which is denoted by $d$. We write $d^*$ instead of $d$ to indicate that the convergence takes place in the “bootstrap world”, i.e. conditional on the original sample. After this first part, we have that $G_n(x, \hat{F}_n) \overset{d^*}{\longrightarrow} G_\infty(x, \hat{F}_n)$. To get to the condition for consistency, we now need a second part. We require that the estimator of the distribution of the data, $\hat{F}_n$, converges to the true distribution, $F$, i.e. $\hat{F}_n \to F$ as $n \to \infty$. This convergence can take place in probability or almost surely, depending of the estimator and on the setting. Putting these two elements together, we can state that

$$G_n(x, F_n) \overset{d^*}{\longrightarrow} G_\infty(x, F) \quad \text{in probability / almost surely.} \quad (1.1)$$

Hence, the quantifier ‘in probability’ (or ‘almost surely’) is needed when we discuss convergence of bootstrap distributions, in order to take the convergence of $\hat{F}_n$ to $F$ into account. In this thesis we will focus exclusively on the weak form, so convergence in probability.

We can now give a definition of consistency.\textsuperscript{7}

---

\textsuperscript{6}The critical region is defined by a critical value that is chosen such that the probability that a value of the test statistic $T_n$ in the critical region occurs is very small if the data actually satisfy the null hypothesis. How small this probability should be is up to the tester who has to set the significance level of the test.

\textsuperscript{7}Our definition is very similar to the one given in Horowitz (2001). The difference is that he
Definition 1.1 (Consistency of the bootstrap). Let \( \hat{F}_n \) be an estimator of \( F \). The bootstrap estimator \( G_n(x, \hat{F}_n) \) of \( G_n(x, F) \) is consistent if

\[
\sup_x \left| G_n(x, \hat{F}_n) - G_\infty(x, F) \right| \xrightarrow{p} 0.
\]

Horowitz (2001, Section 2.1) discusses conditions under which the bootstrap is consistent. For random samples, one can establish very general theorems; for example, asymptotic normality of the statistic is in many situations sufficient (see Horowitz (2001) for more details).\(^8\)

When we consider hypothesis tests, there are two complications. First, the terminology consistency is confusing in a testing setup as a consistent hypothesis test is a test that has the property that it will reject the null hypothesis with probability 1 if the data are generated under \( H_1 \) if \( n \to \infty \). Second, the bootstrap should not just replicate the asymptotic distribution of the test statistic, but it should do so under the null hypothesis. Therefore we will, in the remainder of this thesis, talk about asymptotic validity instead of consistency when we consider the convergence of the bootstrap distribution as \( n \to \infty \). When we discuss consistency in the remainder of this thesis, it refers to the concept mentioned for hypothesis tests (including bootstrap tests).

We will now give our definition of asymptotic validity. Let \( F_0 \) denote a distribution that satisfies the null hypothesis \( H_0 \). Let \( x_1, \ldots, x_n \) be generated by \( F_0 \), and let the bootstrap sample \( x^*_1, \ldots, x^*_n \) and \( T^*_n = T_n(x^*_1, \ldots, x^*_n) \) be conditional on these data that satisfy the null hypothesis.

Definition 1.2 (Asymptotic validity of the bootstrap). Let \( H_0 \) hold. Then the bootstrap is asymptotically valid if

\[
T^*_n \xrightarrow{d^*} G_\infty(x, F_0) \text{ in probability.}
\]

Two remarks are in order. First, the definition above only makes a statement about the situation where the data actually satisfy the null hypothesis. Ideally, we want the bootstrap distribution to converge to \( G_\infty(x, F_0) \) as well if the null hypothesis does not hold. However, this is not required for asymptotic validity (nor for consistency of the hypothesis test).\(^9\) Second, we did not define asymptotic validity in terms of the sup-norm (or some other norm) as we did above in the definition of consistency in (1.2). However, condition (1.3), together with continuity of \( G_\infty \), implies condition (1.2) (cf. Politis and Romano, 1994a, Theorem 1). Moreover, showing that condition (1.3) holds, is sufficient to show that bootstrap tests are asymptotically correctly sized; see Chang and Park (2003, p. 389) for a discussion in a unit root setting.

\(^8\)Bickel and Freedman (1981) were the first who established and utilized a framework of consistency for the bootstrap.

\(^9\)Chapter 3 and 5 in particular will go into more detail on the difference between bootstrap convergence under the null and the alternative.
Whenever we discuss asymptotic validity in the upcoming chapters, it refers to Definition 1.2. Note that asymptotic validity or consistency of a bootstrap procedure is only a necessary condition to be able to use it properly. It does not say anything about how well the bootstrap performs. This we will discuss in the next section.

1.1.2 Asymptotic refinements of the bootstrap

The reason to prefer the bootstrap over asymptotic inference would normally be that $G_n^*$ is a better approximation of $G_n$ than $G_\infty$. Unfortunately, we cannot directly analyze this theoretically, and we again have to resort to asymptotic theory.

What we can analyze using asymptotic theory, under certain (fairly restrictive) conditions, is how fast the error made by the bootstrap approximation and the error made by the asymptotic approximation converge to zero as $n \to \infty$. If the error made by the bootstrap approximation converges to zero faster than the asymptotic approximation, the bootstrap is said to offer asymptotic refinements. The analysis involves complex asymptotic expansions, which are outside the scope of this thesis.\(^\text{10}\) It suffices to say here that the bootstrap generally only offers refinements if the statistic of interest is asymptotically pivotal.

The concept of asymptotic refinements has traditionally been the benchmark with which to evaluate the performance of bootstrap methods (including versus each other). Methods with higher order refinements have generally been advocated as the preferred method. However, asymptotic refinements are still an asymptotic concept and finite sample performance may differ from what is expected from the refinements. So the bootstrap may be useful even when there are no (or small refinements), and the other way around. Monte Carlo simulations assessing the bootstrap’s performance in finite samples do not always agree with the asymptotic refinements.

Notwithstanding the above qualification, a general good advice is to avoid bootstrapping statistics that are not asymptotically pivotal statistics if asymptotically pivotal statistics are available; this is confirmed for unit root testing in Chapter 2. However, if no asymptotically pivotal statistics are available in the setting one is analyzing, the bootstrap has other advantages. The fact that no asymptotically pivotal statistic is available means that for each relevant statistic, the limit distribution of the statistic depends on nuisance parameters. Unless there is some way to consistently estimate these nuisance parameters, asymptotic inference is not feasible; for every new set of data one would need a different set of unknown critical values for example. The bootstrap by construction allows for inference, as it automatically takes the nuisance parameters into account by replicating them in the bootstrap distribution. Thus it can be used in settings where asymptotic inference is infeasible.

\(^{10}\) Hall (1992) discusses asymptotic refinements of the bootstrap in detail.
1 Introduction

1.2 The bootstrap in time series

The bootstrap as discussed above, in particular the bootstrap algorithm for the
i.i.d. bootstrap, is clearly not valid for time series data. In contrast to random
samples, there is a logical ordering in time series data, and usually also a depen-
dence between different time points. The i.i.d. bootstrap clearly violates this, and
by applying it to time series data all dependence structure will be lost.

1.2.1 The bootstrap for stationary time series

Several techniques have been developed to deal with stationary time series data.\textsuperscript{11}
We will discuss two techniques that will be used in the later chapters of the thesis.
The first is a nonparametric technique called the block bootstrap, first developed
by Carlstein (1986) and Künsch (1989). The basic idea behind it is to divide the
data into blocks of consecutive observations and resample these blocks instead
of individual observations. Within each block, the structure of the series will be
preserved. As such, the block bootstrap works for fairly general processes; loosely
speaking, it works if enough of the dependence structure can be preserved within
one block and if the data are stationary.\textsuperscript{12}

We will here give a short algorithm of the most popular form, the moving-blocks

Bootstrap Algorithm 1.2 (moving-blocks bootstrap).

1. Divide the data $x_1, \ldots, x_n$ into overlapping blocks of length $b$; so the first
   block is $x_1, \ldots, x_b$, the second block is $x_2, \ldots, x_{b+1}$, and so on.
2. Randomly draw blocks with replacement and lay them end-to-end to obtain
   the bootstrap sample $x_1^*, \ldots, x_n^*$.
3. Calculate $T_n^* = T_n(x_1^*, \ldots, x_n^*)$.
4. Repeat steps 2 and 3 $B$ times. Let $T_{n,b}^*$ denote the bootstrap statistic ob-
   tained in the $b$-th replication. Estimate $G_n(x, \hat{F}_n) = P^*(T_n^* \leq x)$ by

   \[ \frac{1}{B} \sum_{i=1}^{B} I(T_{n,b}^* \leq x). \]

The block length $b$ has to be selected by the user; more on this will be said
in Chapter 2 and especially Chapter 5. There are many variants on this method
but all rely the same underlying principle. Here we just mention the stationary
bootstrap of Politis and Romano (1994b) which uses random block lengths, as it
returns in Chapter 2.

\textsuperscript{11}See the surveys of Bühmann (2002), Härdle, Horowitz, and Kreiss (2003) and Politis (2003)
\textsuperscript{12}Formal conditions will be given in Chapter 2 and 5.
The second technique is a semi-parametric technique developed by Kreiss (1992) and Bühlmann (1997). While it is also often labeled the recursive bootstrap, we will use the name sieve bootstrap as introduced by Bühlmann (1997). The underlying idea is that by modeling and estimating the dependence structure, one can filter it out and treat the residuals of the model as if they were i.i.d. After applying the i.i.d. bootstrap, the estimated model can then be used to build a bootstrap sample that has the same dependence structure as the original sample. The filter that is being used by Kreiss (1992) and Bühlmann (1997) is an autoregressive model. In order to be able to use this autoregressive model, one must assume that the process generating the data can be described as an invertible linear process (Phillips and Solo, 1992). The corresponding algorithm looks as follows.

**Bootstrap Algorithm 1.3** (sieve bootstrap).

1. Estimate an autoregressive model of order $p$ for the sample $x_1, \ldots, x_n$:
   \[
   x_t = \sum_{j=1}^{p} \hat{\phi}_j x_{t-j} + \hat{\epsilon}_t. \tag{1.4}
   \]

2. Draw with replacement from the residuals $\hat{\epsilon}_{p+1}, \ldots, \hat{\epsilon}_n$ (possibly after subtracting their mean first) to create the bootstrap errors $\epsilon^*_t$.

3. Construct $x^*_t$ as
   \[
   x^*_t = \sum_{j=1}^{p} \hat{\phi}_j x^*_{t-j} + \epsilon^*_t. \tag{1.5}
   \]

4. Calculate $T^*_n = T_n(x^*_1, \ldots, x^*_n)$.

5. Repeat steps 2 to 4 $B$ times. Let $T^*_n,b$ denote the bootstrap statistic obtained in the $b$-th replication. Estimate $G_n(x, \hat{F}_n) = P^*(T^*_n \leq x)$ by
   \[
   \frac{1}{B} \sum_{i=1}^{B} I(T^*_n,b \leq x). \]

The order of the autoregressive model $q$ must be selected by the user (see Chapter 2, 3 and 4). Note that no specific model is assumed (such as an autoregressive model); instead it is assumed that the data generating process belongs to a class of processes. Therefore it would be incorrect to call this method a parametric method. One might even call it a nonparametric method (in fact Bühlmann

---

13 The name derives from the fact that the method is based on approximating an infinite-dimensional model by a sequence of finite-dimensional models, which is known as the method of sieves (cf. Geman and Hwang, 1982).

14 Formal conditions will be given in Chapter 2, 3 and 4.

15 For a finite-order autoregressive model, this bootstrap method is equivalent to the model-based bootstrap considered by Freedman (1981, 1984) and could therefore be called parametric in that particular setting.
1 Introduction

(1997) does), but as it allows for considerably less general processes than the block bootstrap we settle for semi-parametric.

As the methods are compared extensively in the unit root setting in Chapter 2, we will not discuss them in detail here. We only mention here that, under the appropriate assumptions, both methods have been shown to be consistent for a wide array of statistics of interest. Results that are available on asymptotic refinements indicate that the sieve bootstrap (when it is consistent) is more precise than the block bootstrap for stationary time series.\(^{16}\)

1.2.2 Bootstrapping nonstationary time series: contribution of the thesis

The methods described above cannot be applied to nonstationary time series directly. For integrated time series not enough of the dependence can be captured in blocks (the stochastic trend is “broken” between blocks), leading to invalidity of the block bootstrap. Also, an integrated process is not invertible, thereby invalidating the sieve bootstrap.

Moreover, several additional difficulties arise with applying the bootstrap to integrated processes. The most notable is that it is no longer possible, at least at this point in time, to give general theorems for asymptotic validity for a range of underlying processes and test statistics. Instead, for every specific application one has to show the validity of the bootstrap. This is one of the major tasks we will take up in the upcoming chapters, and is the first theme of the thesis. Another issue arising is that virtually nothing known about asymptotic refinements (with the exception of Park, 2003) in this setting. The derivation of asymptotic refinements in this setting is extremely difficult and will therefore not be considered. Instead we will analyze the performance of the bootstrap in finite samples through Monte Carlo simulations, which is the second theme of this thesis.

We can state three main contributions of the thesis. First, for several specific cases theoretical proofs of validity are provided. These specific cases will be described below. Second, the methodology employed in this thesis establishes a general framework to analyze the bootstrap in nonstationary time series and can therefore be used beyond the cases considered here. Third, simulation studies provide an analysis of the performance of the methods in finite samples. These contributions are not only useful for theoretical research but will be of value for practical applications as well. The first contribution gives practitioners the confirmation that those bootstrap methods are valid and may therefore be used; the second allows researchers to look beyond the specific cases considered here and evaluate new methods using the framework established here; the third gives practical recommendations about performance in finite samples. Hence this thesis also provides guidelines and recommendations for practitioners on which tests to apply for empirical research.

In Chapter 2 we will analyze several bootstrap methods that have been proposed for unit root testing. The considered tests differ in several elements, and it

\(^{16}\)Details can also be found in Härdle et al. (2003).
is analyzed which options for these elements work best. One of the conclusions is that the sieve bootstrap is generally preferable to the block bootstrap, which is why we will focus on it in Chapter 3 and 4, although not in Chapter 5 as we work in a different setting there.

Chapter 3 builds on the results of Chapter 2. Using one of the tests based on the sieve bootstrap that was found to perform well in Chapter 2, we consider an aspect that was largely ignored in Chapter 2, namely the treatment of deterministic trends in unit root testing, which is very relevant for empirical applications. It is analyzed both theoretically and through simulations how this affects the bootstrap test.

We consider testing for cointegration in Chapter 4. Again using the sieve bootstrap, we propose a bootstrap test for cointegration based on a single equation error correction model. The asymptotical validity of the method is established and its performance in finite samples is analyzed in relation to other bootstrap tests and its asymptotic counterpart.

Chapter 5 returns to unit root testing but leaves the sieve bootstrap behind. Instead of staying in the pure time series setup, we consider unit root testing using panel data. Adding the cross-sectional dimension to the data allows for more opportunities, but also leads to additional issues to be solved, most notably the issue of cross-sectional dependence. In Chapter 5 we claim that the block bootstrap instead of the sieve bootstrap is the preferred method to deal with these issues. The asymptotic validity of our block bootstrap panel unit root tests is established for a wide range of underlying processes, and the performance of the methods in finite samples is analyzed using simulations.

Chapter 6 concludes the thesis. Finally, some further remarks on the structure of the thesis. Proofs of the theoretical results are contained in appendices at the end of the chapters for which the proofs are relevant. The notation used, although fairly similar, may differ for each chapter, but within each chapter the notation is consistent. Relevant notation is therefore defined within the chapters.
Chapter 2

Bootstrap Unit Root Tests: Comparison and Extensions

In this chapter we study and compare the properties of several bootstrap unit root tests recently proposed in the literature. The tests are Dickey-Fuller or Augmented DF-tests, either based on residuals from an autoregression and the use of the block bootstrap or on first differenced data and the use of the stationary bootstrap or sieve bootstrap. We extend the analysis by interchanging the data transformations (differences versus residuals), the types of bootstrap and the presence or absence of a correction for autocorrelation in the tests.

We show that two sieve bootstrap tests based on residuals remain asymptotically valid. In contrast to the literature which focuses on a comparison of the bootstrap tests with an asymptotic test, we compare the bootstrap tests among them using response surfaces for their size and power in a simulation study.

This study leads to the following conclusions: (i) augmented DF-tests are always preferred to standard DF-tests; (ii) the sieve bootstrap performs better than the block bootstrap; (iii) difference-based tests appear to have slightly better size properties but residual-based tests appear more powerful.\footnote{This chapter is based on the paper Palm, Smeekes, and Urbain (2008a) published in Journal of Time Series Analysis.}

2.1 Introduction

Due to the good performance of the bootstrap in finite samples for stationary processes, its application to nonstationary series has recently become increasingly popular. In this chapter we study and compare the properties of some bootstrap unit root tests that have recently been proposed in the literature. We also introduce some new tests, show their first order asymptotic validity and compare them to existing tests. The tests considered are Dickey-Fuller (DF) or Augmented Dickey-Fuller (ADF) tests, either based on residuals from an autoregression and...
the use of the block bootstrap (Paparoditis and Politis, 2003) or on first differenced data and the use of the stationary bootstrap (Swensen, 2003a) or sieve bootstrap (Psaradakis, 2001; Chang and Park, 2003). As mentioned, these papers differ in the way the bootstrap unit root tests are constructed. Besides showing the asymptotic validity, all these papers compare the finite sample performance of their test(s) to the asymptotic counterpart(s), and the results are overall encouraging. It is however less clear how these tests perform compared to each other. The goal of this chapter is to find out which tests perform best under circumstances to be given, and which aspects of the tests determine their finite sample performance. We will analyze and compare the asymptotic properties of these tests, and we will also consider Monte Carlo simulations.

We distinguish three main features of the tests. The first feature is the actual test statistic. Some tests use the DF test, others the ADF. As the ADF statistic is asymptotically pivotal, whereas the DF is not, we might expect a bootstrap ADF test to offer asymptotic refinements over the bootstrap DF test and asymptotic tests (Horowitz, 2001). The second feature is which series exactly should be resampled. Bootstrapping a nonstationary series directly is not valid (Basawa, Mallik, McCormick, Reeves, and Taylor, 1991b). Therefore a stationary series has to be constructed first. Some tests use residuals from a first-order autoregression of the series, others use first-differences of the series. Swensen (2003b) shows that power functions are the same for both cases if the innovations are i.i.d. However as shown by Paparoditis and Politis (2003, 2005), the use of differences leads to poor behavior of the bootstrap tests under the alternative. The third feature is the time series bootstrap method that is employed. Some tests that we consider use some form of the block bootstrap, in which blocks of (restricted) residuals are resampled. Other tests use the sieve bootstrap, that fits an AR model to the (restricted) residuals and resamples the residuals of this AR model. The sieve bootstrap is somewhat easier to use and performs better when valid, but the block bootstrap is valid for more general processes.

Currently, to our knowledge no tests that use the sieve bootstrap based on residuals have been shown to be asymptotically valid for the Data Generating Processes (DGPs) considered in this chapter. We adapt the sieve bootstrap tests by Psaradakis (2001) and Chang and Park (2003) by constructing them using residuals instead of differences and show that these new tests are asymptotically valid. As residual-based tests may have better properties under the alternative than difference-based tests, this is an important extension. With these results, all the tests considered in this chapter have been shown to be asymptotically valid.

A word on notation. We denote weak convergence by \( \overset{d}{\rightarrow} \), convergence in probability by \( \overset{p}{\rightarrow} \), and almost sure convergence by \( \overset{a.s.}{\rightarrow} \). \( W(r) \) indicates a standard Brownian motion. As usual, we use the superscript ‘\( * \)’ to denote bootstrap quantities, both for bootstrap samples and statistics calculated for bootstrap samples.

---

\(^2\)We call a test **asymptotically valid** if the bootstrap distribution under the null converges to the asymptotic null distribution.

\(^3\)Park (2003) shows that bootstrap ADF tests offer asymptotic refinements under the assumption the errors are a finite AR process with known order.
2.2 The tests

Similarly, ‘$d^*\to$’ indicates weak convergence of a bootstrap statistic conditional on the original series.

The structure of the chapter is as follows. In Section 2.2, we discuss the bootstrap unit root tests, highlight several features of these tests and prove the asymptotic validity of the new tests proposed. Section 2.3 contains an extensive Monte Carlo simulation analysis of the various bootstrap unit root tests. The results are summarized using response surfaces. Section 2.4 concludes. All proofs are contained in the appendix.

2.2 The tests

In this section we discuss several bootstrap unit root tests from a theoretical point of view.

2.2.1 DF sieve bootstrap test

Difference-based DF sieve bootstrap test: Psaradakis (2001)

Psaradakis (2001) considers the following DGP for the time series $y_t, t = 1, \ldots, n$:

$$y_t = d_t + v_t, \quad v_t = \rho v_{t-1} + u_t,$$

(2.1)

where $d_t$ consists of deterministic components. Three cases for the deterministic components are considered: the first case is without deterministics, $d_t = 0$, the second case is with only a constant term, $d_t^* = \delta_0$, and the third case is with constant term and linear time trend, $d_t^* = \delta_0 + \delta_1 t$. The process $u_t$ is assumed to satisfy the following condition with $r = 4$ and $s = 1$:

Assumption 2.1.

(i) the process $u_t$ is generated by $u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, with $\varepsilon_t$ a sequence of i.i.d. random variables with $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma_\varepsilon^2 > 0$ and $E[\varepsilon_t^r] < \infty$.

(ii) (A) Let $\psi_0 = 1, \sum_{j=1}^{\infty} j^s |\psi_j| < \infty$ and $\sum_{j=0}^{\infty} |\psi_j| \neq 0.$

(B) $\sum_{j=0}^{\infty} \psi_j z^j$ is bounded, and bounded away from zero for $\{z \in \mathbb{C}: |z| \leq 1\}.$

Note that this assumption implies that $u_t$ is an invertible linear process; see Phillips and Solo (1992) for more details. We can rewrite the model (2.1) into the following form

$$y_t = \rho y_{t-1} + d_t^* + u_t,$$

(2.2)

where $d_t^* = \gamma_0 + \gamma_1 t := (1 - \rho)\delta_0 + \rho \delta_1 + (1 - \rho)\delta_1 t$ (in the first case $\delta_0 = \delta_1 = 0$, in the second case $\delta_1 = 0$). Psaradakis considers the DF coefficient test $n(\hat{\rho} - 1)$ and t-test in equation (2.2) for testing $\rho = 1$. As stated above, the assumptions on the innovations allow for a sieve bootstrap.

Psaradakis (2001) furthermore needs the following assumption on the order of the autoregression:
Assumption 2.2. The order \( p \) of the autoregressive approximation is such that \( p = p(n) \to \infty \) as \( n \to \infty \) with \( p(n) = o((n/\ln n)^{1/4}) \).

The exact bootstrap procedure can be described as follows.

**Bootstrap Algorithm 2.1** (Psaradakis, 2001).

1. Fit an AR(\( p \)) model to \( \hat{u}_t \), where \( \hat{u}_t = \Delta y_t \) if the deterministic part consists of at most a constant term, and \( \hat{u}_t = \Delta y_t - n^{-1} \sum_{t=1}^{n} \Delta y_t \) if the deterministic part contains both a constant term and a linear time trend, to obtain estimates \( \hat{\phi}_{j,n} \) and

\[
\hat{\varepsilon}_{t,n} = \hat{u}_t - \sum_{j=1}^{p} \hat{\phi}_{j,n} \hat{u}_{t-j}, \quad t = 1 + p, \ldots, n.
\]

2. Generate an i.i.d. sample \( \tilde{\varepsilon}_{t,n} \) by drawing randomly with replacement from \( \hat{\varepsilon}_{t,n} - (n - p)^{-1} \sum_{t=1+p}^{n} \hat{\varepsilon}_{t,n} \).

3. Construct bootstrap errors by the recursion

\[
\tilde{u}_{t,n} = \sum_{j=1}^{p} \hat{\phi}_{j,n} \tilde{u}_{t-j,n} + \tilde{\varepsilon}_{t,n}. \tag{2.3}
\]

4. The bootstrap sample \( \tilde{y}_{t,n} \) is generated recursively by

\[
\tilde{y}_{t,n} = \tilde{y}_{t-1,n} + \tilde{u}_{t,n}
\]

in case of no deterministic components or an intercept only, and by

\[
\tilde{y}_{t,n} = n^{-1} \sum_{t=1}^{n} \Delta y_t + \tilde{y}_{t-1,n} + \tilde{u}_{t,n}
\]

in case of a constant term and a linear trend.

5. Calculate the DF coefficient test and t-test using the bootstrap sample for the previously specified deterministic specification.

6. Repeat steps 2 to 5 \( B \) times to find the bootstrap distributions where \( B \) denotes the number of bootstrap replications.

Psaradakis (2001) suggests to estimate the AR(\( p \)) model in step 1 using the Yule-Walker equations to ensure that the generated innovations \( \tilde{u}_{t,n} \) admit a one-sided MA(\( \infty \)) representation. The asymptotic distribution of the bootstrap statistics under the null is shown to be the same as the asymptotic distribution of the original DF statistics. Note that although the limiting distributions contain nuisance parameters, this does not matter for the bootstrap approach as the critical values for testing are based on the (empirical) distributions of the bootstrap tests that can be approximated by simulation with any accuracy desired.
Residual-based DF sieve bootstrap test: Psaradakis modified

The test we propose here is very similar to the test by Psaradakis (2001), except that it is based on residuals. Paparoditis and Politis (2005) have proposed an ADF coefficient test, and we construct our test in the same way as they do. We will show that our test is asymptotically valid when considering the assumptions made by Psaradakis (2001).

The new algorithm differs from that for the tests by Psaradakis (2001) only in step 1:

**Bootstrap Algorithm 2.2** (Residual-based DF sieve bootstrap procedure).
Replace step 1 from Bootstrap Test 2.1 by calculating the residuals from the regression

$$
\hat{\epsilon}_{t,n} = \tilde{y}_t - \hat{\rho}_n \tilde{y}_{t-1} - \sum_{j=1}^{p} \hat{\phi}_{j,n} \Delta \tilde{y}_{t-j}, \quad t = 1 + p, \ldots, n,
$$

where \( \tilde{y}_t = y_t \) in the case of a (possibly zero) intercept and \( \tilde{y}_t = y_t - \hat{\gamma}_0 - \hat{\gamma}_1 t \) in the case of a linear trend, and \( \hat{\gamma}_0 \) and \( \hat{\gamma}_1 \) are the corresponding OLS estimates.

The next theorem shows that, under the assumptions given above, the bootstrap distributions converge to the same limit distribution as the standard test statistics:

**Theorem 2.1.** Let \( \tau^*_{\alpha} = n(\hat{\rho}^*_n - 1) \) and \( t^*_{\alpha} \) be the coefficient and t-statistic, respectively, that follow from Bootstrap Procedure 2.2. Let \( \sigma_n^2 = \text{E}[u_t^2] \) and \( \sigma^2 = \lim_{n \to \infty} n^{-1} \text{E}[\sum_{t=1}^{n} u_t^2] \). Under Assumptions 2.1 with \( r = 4 \) and \( s = 1 \) and 2.2, we have that

$$
\tau^*_{\alpha} \xrightarrow{d^*} \int_{0}^{1} W(r)dW(r) + \left(\sigma^2 - \sigma_n^2\right)/2\sigma^2
$$

in probability

$$
t^*_{\alpha} \xrightarrow{d^*} \int_{0}^{1} W(r)dW(r) + \left(\sigma^2 - \sigma_n^2\right)/2\sigma^2

\left(\sigma_n^2/\sigma^2\right) \int_{0}^{1} W(r)^2dr \right)^{1/2}
$$

in probability.

where \( W(r) \) is a standard Brownian motion on \([0,1]\).

We have shown that the DF sieve test as constructed by Psaradakis (2001) remains asymptotically valid if it is based on residuals instead of differences.

2.2.2 ADF sieve bootstrap test

**Difference-based ADF sieve bootstrap test: Chang and Park (2003)**

Chang and Park (2003) consider the DGP

$$
y_t = \rho y_{t-1} + u_t,
$$

(2.5)
where \( u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \). They employ Assumption 2.1 with \( r \geq 4 \) and \( s \geq 1 \). For the order of the autoregressive approximation, Chang and Park (2003) consider two different assumptions:

**Assumption 2.3.** Let \( p(n) \to \infty \) and \( p(n) = o(n^\kappa) \) with \( \kappa < \frac{1}{2} \) as \( n \to \infty \).

The following assumption is stronger.

**Assumption 2.4.** Let \( p(n) = cn^\kappa \) for some constant \( c \) and \( 1/rs < \kappa < \frac{1}{2} \).

The bootstrap procedures of Chang and Park (2003) and Psaradakis (2001) are very similar:

**Bootstrap Algorithm 2.3** (Chang and Park, 2003).

Follow the same steps as in Bootstrap Test 2.1, but only for the deterministic specification \( d_t = 0 \). Replace step 5 by

5. Calculate the ADF coefficient statistic \( (1 - \sum_{j=1}^{p} \hat{\phi}_j,n)\Delta y_{t-j,n} + \varepsilon^*_t \) and the corresponding t-statistic\(^4\) from the ADF regression

\[
y_{t,n} = \rho^* y_{t-1,n} + \sum_{j=1}^{p} \hat{\phi}_j^* \Delta y_{t-j,n} + \varepsilon^*_t.
\]

Chang and Park (2003) show that their bootstrap tests converge to the same asymptotic distributions under the null as the asymptotic tests. The convergence is shown to hold almost surely under the strong assumptions, and in probability under the weaker assumptions. They claim that their tests are also valid when applied to demeaned or detrended data, but they do not provide any further analysis.

**Residual-based ADF sieve bootstrap test: Chang and Park modified**

Similar to the previous section, we construct a residual-based test that is based on the test by Chang and Park (2003) and resembles the residual-based ADF test of Paparoditis and Politis (2005) strongly.

**Bootstrap Algorithm 2.4** (Residual-based ADF sieve bootstrap test).

Replace step 1 from Bootstrap Test 2.3 by calculating the residuals from an ADF regression as in the following equation

\[
\hat{\varepsilon}_{t,n} = y_t - \hat{\rho}_n y_{t-1} - \sum_{j=1}^{p} \hat{\phi}_{j,n} \Delta y_{t-j}, \quad t = 1 + p, \ldots, n.
\]

\(^4\)Chang and Park suggest using \( \hat{\sigma}^2_{\varepsilon,n} \) (calculated from the original sample) for the t-test instead of \( \hat{\sigma}^2_{\varepsilon,n}^* \) (calculated from the bootstrap sample), although both are appropriate. Similarly, it is possible to use \( 1 - \sum_{j=1}^{p} \hat{\phi}^*_{j,n} \) for the coefficient test.
The next theorem shows that, under the assumptions given above, the bootstrap distributions converge to the same limit distributions as the asymptotic test statistics.

**Theorem 2.2.** Let $\tau_n^*$ and $t_n^*$ be the bootstrap coefficient statistic and t-statistic, respectively, that follow from Bootstrap Test 2.4. Let Assumptions 2.1 with $r \geq 4$ and $s \geq 1$ and 2.3 hold. Then

\[
\tau_n^* \xrightarrow{d} \int_0^1 W(r) dW(r) \quad \text{in probability,}
\]

\[
t_n^* \xrightarrow{d} \left( \int_0^1 W(r) dW(r) \right)^{1/2} \quad \text{in probability.}
\]

We have shown that the ADF sieve test as constructed by Chang and Park (2003) is also asymptotically valid if it is based on residuals. In Theorem 2 we have obtained convergence in probability whereas Chang and Park (2003) proved a.s. convergence for their strong assumptions. By imposing the unit root restriction difference-based tests rely on stationary series for which a.s. convergence holds. Although not imposing the unit root when applying the sieve bootstrap is certainly a drawback, our result is worthwhile as it does provide justification for using a residual-based sieve bootstrap, even if it is not the same justification as Chang and Park (2003) provide for their tests.

For finite order AR($p$) processes, Paparoditis and Politis (2005) show that under fixed alternatives the bootstrap distribution of the residual-based sieve bootstrap coefficient test is the same as that under the null. For the difference-based sieve bootstrap the distribution under the null differs from that under the alternative for the coefficient test, but not for the t-test. This results in a loss of power for the difference-based sieve bootstrap coefficient test. For the t-tests, both methods are asymptotically equivalent. For AR(\infty) processes, Paparoditis and Politis (2005) do not discuss the residual-based sieve test, but they state that the difference-based sieve bootstrap is inappropriate as the differenced process is not invertible if the alternative is true.

### 2.2.3 (A)DF block bootstrap test

**Residual-based (A)DF block bootstrap test: Paparoditis and Politis (2003)**

Paparoditis and Politis (2003) propose a block bootstrap method to test for unit roots. Their method, the residual-based block bootstrap (RBB), is a block bootstrap method applied to residuals of a regression of the series $y_t$ on its first lag. We first state the assumptions under which the RBB is appropriate. Two sets of assumptions are considered, such that one of these should be satisfied by the process $y_t$ to validate the use of the RBB. Paparoditis and Politis (2003) consider
the process $y_t = \alpha + \rho y_{t-1} + u_t$ where if $\alpha \neq 0$ there is a drift under the null of $\rho = 1$.

Paparoditis and Politis (2003) consider two sets for $u_t$. The first is that Assumption 2.1 (i) and (ii)(A) hold with $r = 4$ and $s = 1$ under the null. Under the alternative these assumptions should hold for $y_t$ as well. Under the additional assumption (ii)(B) the process is invertible as well. This assumption is similar to those Psaradakis and Chang and Park employ.

The second assumption that Paparoditis and Politis (2003) use, is that $u_t$ is strong mixing:

**Assumption 2.5.** For each value of $\rho$, the series $u_t$ is strong mixing and satisfies the following conditions: $E[u_t] = 0$, $E[|u_t|^r] < \infty$ for some $r > 2$, $f_u(0) > 0$, where $f_u$ denotes the spectral density of $u_t$, i.e., $f_u(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_u(h) \exp(i\lambda h)$ and $\gamma_u(h) = E[u_t u_{t+h}]$. Furthermore, $\sum_{k=0}^{\infty} \alpha(k)^{1-2/r} < \infty$, where $\alpha(\cdot)$ denotes the strong mixing coefficient of $u_t$.

In contrast to the condition needed for the sieve bootstrap, one should note that the generating process of $u_t$ does not have to belong to the class of linear processes to satisfy this condition. Hence, we see here a class of processes (possibly non-linear) for which the block bootstrap is valid but the sieve bootstrap is not.

The procedure proposed by Paparoditis and Politis (2003) can be described as follows.

**Bootstrap Algorithm 2.5** (Paparoditis and Politis, 2003).

1. Calculate the centered residuals

$$\hat{u}_{t,n} = \hat{u}_{t,n} - \frac{1}{n-1} \sum_{j=2}^{n} \hat{u}_{t,n} = (y_t - \hat{\rho}_n y_{t-1}) - \frac{1}{n-1} \sum_{j=2}^{n} (y_t - \hat{\rho}_n y_{t-1}),$$

where $\hat{\rho}_n$ is a consistent estimator of $\rho$.

2. Choose the block length $b$, and draw points $i_0, i_1, \ldots, i_{k-1}$, where $k = \lfloor (n - 1)/b \rfloor$, from the uniform distribution on the set $\{1, 2, \ldots, n - b\}$. These points will serve as the beginning points of the blocks of centered residuals:

$$y^*_{t,n} = \begin{cases} y_t & \text{for } t = 1 \\ \alpha^* + y^*_{t-1,n} + \hat{u}_{i_{m+s,n}} & \text{for } t = 2, 3, \ldots, n \end{cases},$$

(2.7)

where $m = \lfloor (t - 2)/b \rfloor$, $s = t - mb - 1$, and $\alpha^*$ is a drift parameter that is either set equal to zero or it is a consistent estimator of $\alpha$.

3. From the bootstrap series $y^*_{t,n}$ compute the desired statistics.

4. Repeat steps 2 to 3 $B$ times to find the bootstrap distribution.

---

The bootstrap sample $y^*_t$ will have total length $l = kb + 1$. 

---

5The bootstrap sample $y^*_t$ will have total length $l = kb + 1$. 

---

20
2.2 The tests

Although most bootstrap unit root tests are based on differences, Paparoditis and Politis (2003) formally show that the residual-based block bootstrap coefficient test performs well asymptotically both under the null and under contiguous alternatives whereas the asymptotic distribution of the difference-based block bootstrap (DBB) statistic differs from that of the RBB statistic under the alternative, leading to a loss of power of the DBB test. Moreover, the convergence rate for this DBB bootstrap test is slower under the alternative than for the RBB test. For fixed alternatives, the slower rate of convergence leads to a loss of power of the DBB test compared to the RBB test. For sequences of \( n^{-1} \) local alternatives, the two tests have the same power.

In step 1 of the bootstrap procedure, \( \hat{\rho}_n \) should be a consistent estimator of \( \rho \). Furthermore, it is required that \( \hat{\rho}_n \) is \( o_p(1) \) if \( \rho \neq 1 \), \( O_p(n^{-1}) \) if \( \rho = 1 \) and \( \alpha = 0 \), and \( O_p(n^{-3/2}) \) if \( \rho = 1 \) and \( \alpha \neq 0 \). Many estimators satisfy these conditions. Paparoditis and Politis (2003) focus on what they call the least squares (LS) estimator, which is just the DF estimator, and the ADF estimator (they call this the DF estimator). For the validity of the ADF estimator the additional condition (ii)(B) is needed to ensure invertibility.

They prove the consistency of the RBB for the DF coefficient test and the ADF coefficient test. For models where \( \alpha = 0 \), they recommend to use the OLS estimator of \( \rho \) in

\[
y_t = \alpha + \rho y_{t-1} + u_t \tag{2.8}
\]

or the ADF equivalent as \( \hat{\rho}_n \), which is used to construct the residuals. In the second step \( \alpha^* \) is set to zero, as there should be no drift. They also recommend for the RBB ADF test to use the block bootstrap of \( y_t - y_{t-1} \) directly as lagged differences instead of \( y^*_{t,n} - y^*_{t-1,n} \). For both the tests without deterministic components and the tests with a constant included, the consistency of the DF and ADF RBB tests is proved.

For the case of \( \alpha \neq 0 \), Paparoditis and Politis (2003) recommend using the same estimator for \( \hat{\rho}_n \) as before but setting \( \alpha^* = \hat{\alpha}_n \) where \( \hat{\alpha}_n \) is the estimator of \( \alpha \) in (2.8). They prove the consistency of the DF coefficient test with constant and trend and claim the consistency of the corresponding ADF test can be established similarly.

**Difference-based (A)DF block bootstrap test: Paparoditis and Politis (2003)**

Again we consider an alternative version of the tests by Paparoditis and Politis (2003). As the original tests are based on residuals, the modified tests will be based on differences. The new procedure simply replaces \( \hat{\rho}_n \) by 1. Paparoditis and Politis (2003) already showed the asymptotic validity of these alternative tests (also see the discussion of power above).

---

6 Depending on which unit root test is performed.

7 Using the same blocks as for the residuals.
2 Bootstrap Unit Root Tests: Comparison and Extensions

2.2.4 DF stationary bootstrap test

Difference-based DF stationary bootstrap test: Swensen (2003a)

Swensen (2003a) considers a unit root test without deterministic components based on the stationary bootstrap of Politis and Romano (1994b). He assumes the DGP $y_t = \rho y_{t-1} + u_t$ with the following assumptions on $u_t$.

Assumption 2.6.

(i) The process $u_t$ is strictly stationary with $E[u_t] = 0$ for all $t$.

(ii) If $\gamma(k) = E[u_t u_{t+k}]$, then $\gamma_0 + \sum_{r=0}^{\infty} |r \gamma(r)| < \infty$

(iii) $\sum_{r,s,t} \kappa_4(r,s,t) = K < \infty$ where $\kappa_4(r,s,t)$ is the fourth cumulant of the distribution of $(u_j, u_{j+r}, u_{j+r+s}, u_{j+r+s+t})$.

Assumption (iii) is used to ensure that the variance of $\frac{1}{n} \sum_t u_t^2$ tends to zero and implies that $\sigma^2$ can be consistently estimated. Under these conditions Swensen (2003a) proves the consistency of the DF tests without deterministic components based on the stationary bootstrap. The conditions needed are significantly weaker than those needed for the sieve bootstrap.

The algorithm can be described as follows:

Bootstrap Algorithm 2.6 (Swensen, 2003).

1. Compute centered differences

$$\tilde{u}_t = \Delta y_t - (n-1)^{-1} \sum_{j=2}^{n} \Delta y_j.$$

2. Apply the stationary bootstrap of Politis and Romano (1994b) to the centered residuals to obtain bootstrap errors $u_{t,n}^*$:

(a) Draw the index of the starting points of the blocks, $i_1, i_2, \ldots$, from the uniform distribution $P(t_1 = t) = \frac{1}{n}, t = 1, \ldots, n$. Let $p_L$ be a fixed number between 0 and 1. Draw the length of the blocks $b_1, b_2, \ldots$ from the geometric distribution $P(b_1 = l) = (1 - p_L)^{l-1} p_L$. The expected block length is $1/p_L$.

(b) Form blocks using the drawn starting points and block lengths. For block $m+1$ we have

$$u_{t,n}^* = \tilde{u}_{i_{m+1}+t-b(m)-1}$$

where $t = b(m) + 1, \ldots, b(m) + b_{m+1}$ and $b(m) = \sum_{j=1}^{m} b_j$.

(c) Stop after generating $B$ blocks if $l_B = \sum_{j=1}^{B} b_j \geq n$. Lay the blocks end-to-end in the order sampled; cut off the resulting series $u_{t,n}^*, \ldots, u_{B,n}^*$ at $u_{n,n}^*$ if $l_B < n$. 

22
2.2 The tests

3. Construct the bootstrap sample $y_{t,n}^*$ with the recursion $y_{t,n}^* = y_{t-1,n}^* + u_{t,n}^*$.

4. Compute the bootstrap DF coefficient and t-statistic.

5. Repeat steps 2 to 4 $B$ times to find the bootstrap distribution.

Residual-based DF stationary bootstrap test: Parker, Paparoditis, and Politis (2006)

Again, we also consider a modified version of these tests. Here we base the modified version on residuals instead of differences. Instead of the centered differences we calculate in step 1 centered residuals as in the bootstrap procedure of Paparoditis and Politis (2003). This test has been proposed by Parker et al. (2006) who also show its asymptotic validity.

2.2.5 Comparison of tests considered

When comparing these tests, we will mainly focus on three aspects: whether differences or residuals are used, the bootstrap method and the test statistic.

Table 2.1 summarizes all the test statistics and their main features. A note on the notation: we use $\tau$ for a coefficient test and $t$ for a t-test. The first subscript indicates the bootstrap method: $S$ stands for sieve bootstrap, $B$ for block bootstrap, and $St$ for stationary bootstrap; the second subscript indicates whether a test is based on differences ($d$) or residuals ($r$). A superscript $a$ states that the test is an augmented DF test.

The test statistic of interest - Asymptotic Refinements

Asymptotic refinements only occur if the statistic of interest is asymptotically pivotal, i.e. the limiting distribution does not depend on nuisance parameters (see for example Horowitz (2001)). Of the tests we consider, only the ADF tests proposed by Chang and Park (2003) and the modification based on residuals provide asymptotically pivotal statistics. Psaradakis (2001) and Swensen (2003a) consider DF tests, whereas the ADF coefficient test of Paparoditis and Politis (2003) does not lead to an asymptotically pivotal statistic for infinite AR models.\(^8\) Therefore, based on possible asymptotic refinements, we might expect that the tests by Chang and Park (2003) and the modification based on residuals will have a better finite sample performance than the other tests.

This is however only a conjecture, as there are little theoretical results about asymptotic refinements for bootstrap unit root tests. Park (2003) shows that for finite AR innovations, where the order is known in advance, bootstrap tests offer asymptotic refinements. This assumption is however rather restrictive, as it means that by fitting an AR model of the correct order all dependency can be removed,\(^8\)

\(^8\)Still, the ADF test of Paparoditis and Politis (2003) takes the dependence in the errors into account, so we might still expect a better performance of this test than that of the DF tests. Besides, for finite AR models, which we also consider, the statistic is asymptotically pivotal.
2 Bootstrap Unit Root Tests: Comparison and Extensions

Table 2.1: Main features of the tests.

<table>
<thead>
<tr>
<th>Test(^a)</th>
<th>Bootstrap Method</th>
<th>Based on</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\tau_{S,d})</td>
<td>Sieve</td>
<td>Differences</td>
<td>DF (\tau)</td>
</tr>
<tr>
<td>(t_{S,d})</td>
<td>Sieve</td>
<td>Differences</td>
<td>DF (t)</td>
</tr>
<tr>
<td>(\tau_{S,r})</td>
<td>Sieve</td>
<td>Residuals</td>
<td>DF (\tau)</td>
</tr>
<tr>
<td>(t_{S,r})</td>
<td>Sieve</td>
<td>Residuals</td>
<td>DF (t)</td>
</tr>
<tr>
<td>(\tau_{S,d}^a)</td>
<td>Sieve</td>
<td>Differences</td>
<td>ADF (\tau)</td>
</tr>
<tr>
<td>(t_{S,d}^a)</td>
<td>Sieve</td>
<td>Differences</td>
<td>ADF (t)</td>
</tr>
<tr>
<td>(\tau_{S,r}^a)</td>
<td>Sieve</td>
<td>Residuals</td>
<td>ADF (\tau)</td>
</tr>
<tr>
<td>(t_{S,r}^a)</td>
<td>Sieve</td>
<td>Residuals</td>
<td>ADF (t)</td>
</tr>
<tr>
<td>(\tau_{B,r})</td>
<td>Block</td>
<td>Residuals</td>
<td>DF (\tau)</td>
</tr>
<tr>
<td>(\tau_{B,d})</td>
<td>Block</td>
<td>Differences</td>
<td>DF (\tau)</td>
</tr>
<tr>
<td>(\tau_{B,r}^a)</td>
<td>Block</td>
<td>Residuals</td>
<td>ADF (\tau)</td>
</tr>
<tr>
<td>(\tau_{B,d}^a)</td>
<td>Block</td>
<td>Differences</td>
<td>ADF (\tau)</td>
</tr>
<tr>
<td>(\tau_{St,d})</td>
<td>Stationary</td>
<td>Differences</td>
<td>DF (\tau)</td>
</tr>
<tr>
<td>(t_{St,d})</td>
<td>Stationary</td>
<td>Differences</td>
<td>DF (t)</td>
</tr>
<tr>
<td>(\tau_{St,r})</td>
<td>Stationary</td>
<td>Residuals</td>
<td>DF (\tau)</td>
</tr>
<tr>
<td>(t_{St,r})</td>
<td>Stationary</td>
<td>Residuals</td>
<td>DF (t)</td>
</tr>
</tbody>
</table>

\(^a\)We use \(\tau\) for a coefficient test and \(t\) for a t-test. The first subscript indicates the bootstrap method: \(S\) stands for sieve bootstrap, \(B\) for block bootstrap, and \(St\) for stationary bootstrap; the second subscript indicates whether a test is based on differences (\(d\)) or residuals (\(r\)). A superscript \(a\) states that the test is an augmented DF test.
and time series bootstrap methods are not needed anymore. Results for more realistic cases are not yet available. Hence we have little theoretical guidance here when interpreting the results from our simulations.

The time series bootstrap method

We can broadly divide our tests into sieve bootstrap and block bootstrap tests. The major advantage of the block bootstrap is that it is valid under very general assumptions. One disadvantage is that the resulting bootstrap sample is not stationary due to the cut-off points at the end of the blocks. The stationary bootstrap, introduced by Politis and Romano (1994b), and employed by Swensen (2003a), solves this issue by using random block lengths. The disadvantage is that the additional randomness of the block lengths causes the stationary bootstrap to work less efficiently than the overlapping blocks bootstrap (see, for example, Lahiri (1999)).

A general disadvantage of the block bootstrap is the difficult choice of the (expected) block length. This choice is not intuitive at all. There are some methods available for selecting block lengths, but most of these apply to estimating variance. An easy to use method applicable to (unit root) testing has yet to be developed.9

The main disadvantage of the sieve bootstrap developed by Kreiss (1992) and Bühlmann (1997) is that it is valid only for the class of stationary linear invertible processes, as we have seen in the assumptions given above. The advantages of the sieve bootstrap include higher order asymptotic refinements than the block bootstrap, a stationary bootstrap series, and an easy to select lag order \( p \) (using for example the well-known information criteria like \( AIC \) or \( BIC \)).

Power considerations

Intuitively, one might argue that taking first differences in the bootstrap should perform better when the null hypothesis is true (if there is indeed a unit root, imposing it leads to the best result), while residuals might perform better under the alternative hypothesis, as imposing a unit root would be false in that situation.

Swensen (2003b) shows that power functions are the same for both cases if the innovations are i.i.d. However, Paparoditis and Politis (2005) show that first differences lead to poor behavior of sieve bootstrap tests under the alternative. If the true model is a finite AR(\( p \)) model, using differences leads to a lower power. If the true model is an infinite AR model, a difference-based test even becomes inappropriate to use as the incorrectly first-differenced process contains a unit moving average root, and is therefore not invertible and hence has no autoregressive approximation. This confirms our intuition that the difference-based tests might suffer from power problems. We will also analyze the choice of differences versus residuals in our simulations.

---

9We discuss some methods in Chapter 5 in relation to unit root testing in panel data.
2 Bootstrap Unit Root Tests: Comparison and Extensions

2.3 Finite sample performance: Monte Carlo results

We analyze and compare the finite sample behavior of the tests by Monte Carlo simulations.

2.3.1 Monte Carlo setup

We generate a series \( y_t, t = 1, \ldots, n \), according to the recursion

\[
y_t = \rho y_{t-1} + u_t, \quad y_0 = 0,
\]

(2.10)

where different values for \( \rho \) are used: 1, 0.99, 0.95, 0.9 and 0.8. We let \( u_t \) be generated by an ARMA(1,1) process:

\[
u_t = \phi u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1},
\]

(2.11)

where \( \varepsilon_t \sim \mathcal{N}(0,1) \). The values used for \( \phi \) and \( \theta \) vary from -0.8 to 0.8. \(^{10}\)

As sample sizes we consider \( n = 50, 100 \) and 250. We use three different significance levels: 0.01, 0.05 and 0.10. All experiments will be based on 5000 simulations and 999 bootstrap replications. All simulations are performed using GAUSS 6.0.

AIC is used to select the lag length for the sieve bootstrap. We estimate the AR(\( p \)) models by OLS. \(^{11}\) For the lag length in the ADF tests we use the modified AIC by Ng and Perron (2001), both outside and inside the bootstrap procedures. For the block length we choose fixed numbers: 5 for \( n = 50 \), 8 for \( n = 100 \) and 15 for \( n = 250 \). The fact that there is no easy way to estimate block lengths remains a problem.

We perform two sets of simulations with these models. The first set considers the tests based on models without deterministic components. In the second set of simulations the DGPs remain unchanged but the tests are based on models with a constant and a trend. These extensions are not discussed in all papers, so that not all tests we consider have been shown to be theoretically valid. For the ADF test of Paparoditis and Politis (2003), we follow their instructions on how to handle the test allowing for a trend. Chang and Park (2003) indicate that their tests can be applied for the model with trend by applying the bootstrap test to the detrended data. We detrend both the original series (by OLS) and the bootstrap series. \(^{12}\)

\(^{10}\) Specific values used for \( (\phi, \theta) \) are: (0.0), (0.0, 0.8), (0, 0.4, 0), (0.4, 0.4), (0.8, 0), (0, 0.8), (0, 0.4), (0, 0.8), (0.8, 0), (0, -0.4), (-0.4, 0), (-0.4, 0.4), (-0.4, -0.4).

\(^{11}\) The estimated AR(\( p \)) model may not be invertible. A solution could be to impose a root bound as in Burridge and Taylor (2004). This is however mainly important for empirical work, as in a large simulation study as ours the number of cases in which the estimated process is not invertible, is very small.

\(^{12}\) It is crucial to detrend the bootstrap series as well, otherwise the bootstrap distribution will not converge to the correct asymptotic distribution.
For the test proposed by Swensen (2003a), deterministic components are added in the same way as in Psaradakis (2001).\textsuperscript{13}

Note however that this is in fact not necessary, as the tests applied are actually invariant to the deterministic components present in the DGP, provided sufficient deterministics are included in the test regression. Therefore, the bootstrap test statistics are also invariant to the deterministics in the bootstrap DGP as long they are correctly specified in the bootstrap test regression.

The large number of DGP’s and tests statistics in our simulations leads to a huge number of results that is rather hard to analyze in standard tables. We circumvent this problem by estimating different response surfaces for the rejection frequencies observed in our simulations for each of the test statistics. Because the empirical rejection frequency $\hat{P}$ lies between 0 and 1, we use the following transformation:

$$L(\hat{P}) = \ln\left(\frac{\hat{P}}{1 - \hat{P}}\right).$$

The dependent variable is $L(\hat{P})$. As explanatory variables we consider several functions of the nominal level and the parameters in the underlying DGP. We will provide more details below. The specific form of the response surfaces is test specific. To avoid lengthy specification searches, we rely on PcGets (Hendry and Krolzig, 2001) to select the most appropriate specification from a large set of possible variables. The reported standard errors are White’s heteroscedasticity consistent standard errors. Apart from the coefficient estimates and their standard errors, the adjusted $R^2$ of the regression is also reported.

### 2.3.2 Results

In this section we will give the main findings of our simulation study. We focus here on the results for the tests without deterministic trends. The results for the tests allowing for deterministic trends will be briefly discussed below.\textsuperscript{14}

**Size**  Tables 2.2 give a summary of the response surfaces for the size. We consider the following response surface for the size:

$$L(\hat{P}_i) = \beta_1 L(P_{a,i}) + \beta_2^f f(L(P_{a,i}), \phi_i, \theta_i, n_i) + \nu_i, \quad i = 1, \ldots, M,$$

where $P_a$ is the nominal size of the test, $f(L(P_{a,i}), \phi_i, \theta_i, n_i)$ is a vector of functions (all of order $O(n^{-1/2})$ and $O(n^{-1/2})$) of $L(P_a)$, the ARMA parameters $\phi_i$ and $\theta_i$, and the sample size $n_i$ and $\nu_i$ denotes a disturbance. The number of simulation experiments $M$ is 99.

\textsuperscript{13}More efficient detrending methods are not considered in this chapter; they are discussed in the next chapter.

\textsuperscript{14}The collection of all simulation results, response surfaces and graphical analyses is available on the Internet: \texttt{www.personeel.unimaas.nl/s.smeekes/outputreport.pdf}
The term $\beta_1 f(\cdot)$ captures the deviations of the actual size from the nominal size as a function of the parameters of the DGP and sample size. $\beta_1 L(P_{n,i})$ gives an indication of the asymptotic size of the tests. When $\beta_1$ is equal to 1, the empirical size of the test is equal to the nominal size for large $n$. The table gives the estimate of $\beta_1$ and its standard error, as well a measure of the fit.

Several things can be seen from the tables. We see that for some tests $\beta_1$ is significantly different from 1, although for most it is close to it. The estimates for the residual-based sieve tests are all not significantly different from one. Most of the other estimates are different from one, where especially the estimates for the DF test are far away from one. Note that these are all DF tests, as opposed to the ADF tests for which $\beta_1$ is much closer to 1. The coefficient and t-tests appear to have similar size in most cases. For the block tests, the value for $\beta_1$ is higher.

\begin{table}[h]
<table>
<thead>
<tr>
<th></th>
<th>$\tau_{S,d}$</th>
<th>$\tau_{S,r}$</th>
<th>$\tau_{a,S,r}$</th>
<th>$\tau_{S,d}^a$</th>
<th>$\tau_{S,r}^a$</th>
<th>$\tau_{S,r}^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>1.15</td>
<td>1.15</td>
<td>0.95</td>
<td>1.01</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td>$P_{n}$</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.76</td>
<td>0.77</td>
<td>0.66</td>
<td>0.66</td>
<td>0.82</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.86</td>
<td>0.87</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>$P_{n}$</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.88</td>
<td>0.87</td>
<td>0.98</td>
<td>1.00</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>$P_{n}$</td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\tau_{B,r}$</th>
<th>$\tau_{B,d}$</th>
<th>$\tau_{B,d}^a$</th>
<th>$\tau_{S,t}$</th>
<th>$\tau_{S,t}^a$</th>
<th>$\tau_{S,t}^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.68</td>
<td>1.05</td>
<td>1.04</td>
<td>1.34</td>
<td>1.20</td>
<td>1.23</td>
</tr>
<tr>
<td>$P_{n}$</td>
<td>(0.08)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.12</td>
<td>0.40</td>
<td>0.48</td>
<td>0.45</td>
<td>0.53</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>1.03</td>
<td>1.17</td>
<td>1.72</td>
<td>1.15</td>
<td>1.12</td>
</tr>
<tr>
<td>$P_{n}$</td>
<td>(0.12)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.12)</td>
<td>(0.06)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.88</td>
<td>0.93</td>
<td>0.85</td>
<td>0.71</td>
<td>0.93</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>1.07</td>
<td>0.92</td>
<td>1.30</td>
<td>1.23</td>
<td>1.19</td>
</tr>
<tr>
<td>$P_{n}$</td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.95</td>
<td>0.94</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Table 2.2: Response surfaces of size
for the difference-based version than the residual-based version, which indicates that in general the residual-based block tests give higher rejection frequencies than difference-based block tests.

We also see for all tests that the fit increases when we include variables of higher order. Especially the increase in the fit from the first setting to the second is noticeable. This shows that all tests suffer from finite-sample distortions, although some more than others. As can be clearly seen, the adjusted $R^2$ for the regression on only the nominal size is much higher for the sieve tests than for the block tests. This shows that the (especially ADF) sieve tests suffer less from finite sample distortions than the other tests.

Figure 2.1 and 2.2 show graphs of the fitted size plotted against the autoregressive and moving-average parameters $\phi$ and $\theta$. As nominal level we take 0.05 and as sample size we take 100. The light grey area indicates a size between 0.03 and 0.07, the black area indicates a size below that range and the dark grey area above that range.

The fitted transformed sizes are calculated from the response surfaces (2.13) for specific values $\phi_0$, $\theta_0$, $n_0$ and $P_{a,0}$, substituting estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ for $\beta_1$ and $\beta_2$ respectively and dropping the disturbance $\nu$. Next we apply the inverse of the $L(\cdot)$ transformation to the fitted values to obtain the fitted size.

As well as confirming what the tables tell us, the graphs show how the AR and MA parameter influence the empirical size. For all the tests, we see the well-known size distortions for a large negative MA parameter. The extent of this size distortion differs however. The stationary bootstrap tests and the DF block tests have massive size distortions, that also increase when the AR parameter becomes large and negative. We see that these tests are much more sensitive to the values of $\phi$ and $\theta$, as they also exhibit a large undersize for large positive values. The ADF block tests mainly exhibit large undersize, especially for large absolute values of $\phi$ and $\theta$. The sieve tests can be seen to perform quite well; especially the ADF sieve tests, for which the graphs are quite flat and in the correct range. We can also see that in general residual-based tests have higher rejection frequencies than the difference-based tests, except for the ADF sieve tests where both perform equally well.

Power Tables 2.3 and 2.4 give summaries of the response surfaces for the power. We choose to report only unadjusted power as we feel this is the most relevant, because this is what matters in practice. We now estimate the following response surface:

$$L(\hat{P}_i) = \beta_0 + \beta_1L(P_{a,i}) + \sum_{k=1}^{3} \beta_{2,k}(\rho_i - 1)^k + \beta_{3}f(\rho_i, L(P_{a,i}), \phi_i, \theta_i, n_i) + \nu_i, \quad i = 1, \ldots, M. \quad (2.14)$$

The number of simulation experiments $M$ is 396. As for size, all variables in $f(\rho, L(P_{a,i}), \phi_i, \theta_i, n_i)$ are either of order $O(n^{-1/2})$ or $O(n^{-1})$. So in this case the asymptotic behavior can be deduced from $\beta^a = (\beta_0, \beta_1, \beta_{2}')'$. The tables give
Figure 2.1: Size as a function of $\phi$ and $\theta$ for sieve tests
2.3 Finite sample performance

(a) Block tests

(b) Stationary tests

Figure 2.2: Size as a function of $\phi$ and $\theta$ for block-type tests
the estimates plus standard error for the $O(1)$ variables and a measure of the fit. Again we see that the fit increases when we add higher order terms.

In Figures 2.3 to 2.4 we give power curves for varying sample sizes. These plots are derived from the response surfaces in the same way as the surface graphs for the size. For all cases, we have taken $\phi = \theta = 0$.\footnote{Unreported results show the dependence of the empirical power on $\phi$ and $\theta$ is similar as in the case of the size.} Most of the graphs show that the residual-based tests have higher power than the difference-based tests. However, as we also found that the residual-based tests have larger size distortions than the difference-based tests in general, the higher power will partly be caused by the size distortions. In that respect, we see that the power difference between residual-based and difference-based ADF sieve tests is quite small, while for these tests the behavior under the null of residual-based tests and difference-based tests was also comparable. Hence, if there is a power advantage for residual-based tests, it is only small.

Deterministic trends  The tests allowing for deterministic trends give qualitatively similar results as the ones described above. All tests perform worse, however the effect of including the deterministic trends where in fact none are needed is
2.3 Finite sample performance

<table>
<thead>
<tr>
<th>( L(P_n) )</th>
<th>( t_s,d )</th>
<th>( t_s,r )</th>
<th>( t_r,d )</th>
<th>( t_r^2,d )</th>
<th>( t_r^3,d )</th>
<th>( t_r^2,r )</th>
<th>( t_r^3,r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.50</td>
<td>1.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((\rho - 1))</td>
<td>-50.58</td>
<td>-50.06</td>
<td>-58.09</td>
<td>-57.19</td>
<td>-54.45</td>
<td>-50.47</td>
<td>-51.80</td>
</tr>
<tr>
<td>((\rho - 1)^2)</td>
<td>-116.14</td>
<td>-116.80</td>
<td>-120.31</td>
<td>-117.36</td>
<td>-125.90</td>
<td>-123.84</td>
<td>-111.34</td>
</tr>
<tr>
<td>((\rho - 1)^3)</td>
<td>-131.44</td>
<td>-128.45</td>
<td>-162.95</td>
<td>-159.51</td>
<td>-140.67</td>
<td>-127.78</td>
<td>-139.64</td>
</tr>
<tr>
<td>Adj. ( R^2 )</td>
<td>0.55</td>
<td>0.55</td>
<td>0.51</td>
<td>0.50</td>
<td>0.49</td>
<td>0.51</td>
<td>0.49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( L(P_n) )</th>
<th>( t_s,d )</th>
<th>( t_s,r )</th>
<th>( t_r,d )</th>
<th>( t_r^2,d )</th>
<th>( t_r^3,d )</th>
<th>( t_r^2,r )</th>
<th>( t_r^3,r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.75</td>
<td>2.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((\rho - 1))</td>
<td>-131.44</td>
<td>-128.45</td>
<td>-162.95</td>
<td>-159.51</td>
<td>-140.67</td>
<td>-127.78</td>
<td>-139.64</td>
</tr>
<tr>
<td>((\rho - 1)^2)</td>
<td>-335.05</td>
<td>-328.89</td>
<td>-405.38</td>
<td>-395.02</td>
<td>-362.25</td>
<td>-336.56</td>
<td>-340.95</td>
</tr>
<tr>
<td>((\rho - 1)^3)</td>
<td>-180.89</td>
<td>-183.29</td>
<td>-272.02</td>
<td>-276.88</td>
<td>-242.31</td>
<td>-183.29</td>
<td>-238.07</td>
</tr>
<tr>
<td>Adj. ( R^2 )</td>
<td>0.95</td>
<td>0.94</td>
<td>0.93</td>
<td>0.93</td>
<td>0.89</td>
<td>0.90</td>
<td>0.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( L(P_n) )</th>
<th>( t_s,d )</th>
<th>( t_s,r )</th>
<th>( t_r,d )</th>
<th>( t_r^2,d )</th>
<th>( t_r^3,d )</th>
<th>( t_r^2,r )</th>
<th>( t_r^3,r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>10.49</td>
<td>9.12</td>
<td>1.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((\rho - 1))</td>
<td>-180.89</td>
<td>-235.58</td>
<td>-272.02</td>
<td>-276.88</td>
<td>-242.31</td>
<td>-183.29</td>
<td>-238.07</td>
</tr>
<tr>
<td>((\rho - 1)^2)</td>
<td>-177.17</td>
<td>-888.64</td>
<td>-742.83</td>
<td>-704.57</td>
<td>-362.25</td>
<td>-991.26</td>
<td>-1048.42</td>
</tr>
<tr>
<td>((\rho - 1)^3)</td>
<td>-473.46</td>
<td>-1240.03</td>
<td>-1484.91</td>
<td>-1616.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. ( R^2 )</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.92</td>
<td>0.93</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Table 2.3: Response surfaces of power - part I
### Table 2.4: Response surfaces of power - part II

<table>
<thead>
<tr>
<th>$L(P_n)$</th>
<th>$\tau_B,r$</th>
<th>$\tau_B,d$</th>
<th>$\tau_{B,r}$</th>
<th>$\tau_{B,d}$</th>
<th>$\tau_{St,d}$</th>
<th>$\tau_{St,r}$</th>
<th>$\tau_{St,r}$</th>
<th>$\tau_{St,r}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.79</td>
<td>1.48</td>
<td>1.83</td>
<td>2.27</td>
<td>2.02</td>
<td>1.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.78)</td>
<td>(0.55)</td>
<td>(0.55)</td>
<td>(0.53)</td>
<td>(0.82)</td>
<td>(0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\rho - 1)$</td>
<td>-71.93</td>
<td>-62.79</td>
<td>-52.52</td>
<td>-46.45</td>
<td>-59.59</td>
<td>-57.91</td>
<td>-70.66</td>
<td>-69.22</td>
</tr>
<tr>
<td></td>
<td>(11.42)</td>
<td>(8.34)</td>
<td>(7.03)</td>
<td>(6.90)</td>
<td>(8.41)</td>
<td>(8.11)</td>
<td>(11.98)</td>
<td>(11.76)</td>
</tr>
<tr>
<td>$(\rho - 1)^2$</td>
<td>-174.04</td>
<td>-123.95</td>
<td>-97.71</td>
<td>-81.61</td>
<td>-112.72</td>
<td>-112.41</td>
<td>-179.02</td>
<td>-169.64</td>
</tr>
<tr>
<td></td>
<td>(52.80)</td>
<td>(40.98)</td>
<td>(33.92)</td>
<td>(32.77)</td>
<td>(41.34)</td>
<td>(40.01)</td>
<td>(55.16)</td>
<td>(53.93)</td>
</tr>
<tr>
<td>$(\rho - 1)^3$</td>
<td>-178.26</td>
<td>-171.88</td>
<td>-133.75</td>
<td>-164.84</td>
<td>-161.46</td>
<td>-170.06</td>
<td>-265.95</td>
<td>-265.95</td>
</tr>
<tr>
<td></td>
<td>-562.84</td>
<td>-457.14</td>
<td>-334.33</td>
<td>-304.25</td>
<td>-416.20</td>
<td>-533.99</td>
<td>-2035.21</td>
<td>-2035.21</td>
</tr>
<tr>
<td></td>
<td>(65.82)</td>
<td>(52.36)</td>
<td>(48.79)</td>
<td>(74.40)</td>
<td>(55.18)</td>
<td>(56.45)</td>
<td>(67.15)</td>
<td>(71.28)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.31</td>
<td>0.47</td>
<td>0.50</td>
<td>0.49</td>
<td>0.49</td>
<td>0.29</td>
<td>0.28</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2.5: Response surfaces of power - part III

<table>
<thead>
<tr>
<th>$L(P_n)$</th>
<th>$\tau_B,r$</th>
<th>$\tau_B,d$</th>
<th>$\tau_{B,r}$</th>
<th>$\tau_{B,d}$</th>
<th>$\tau_{St,d}$</th>
<th>$\tau_{St,r}$</th>
<th>$\tau_{St,r}$</th>
<th>$\tau_{St,r}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.53</td>
<td>1.48</td>
<td>1.93</td>
<td>2.40</td>
<td>1.72</td>
<td>1.52</td>
<td>1.46</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.05)</td>
<td>(0.14)</td>
<td>(0.17)</td>
<td>(0.05)</td>
<td>(0.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\rho - 1)$</td>
<td>-351.34</td>
<td>-324.64</td>
<td>-204.24</td>
<td>-172.79</td>
<td>-313.36</td>
<td>-287.02</td>
<td>-276.15</td>
<td>-275.72</td>
</tr>
<tr>
<td></td>
<td>(39.12)</td>
<td>(21.19)</td>
<td>(8.50)</td>
<td>(8.93)</td>
<td>(21.05)</td>
<td>(23.15)</td>
<td>(35.27)</td>
<td>(34.28)</td>
</tr>
<tr>
<td>$(\rho - 1)^2$</td>
<td>-1979.17</td>
<td>-1068.85</td>
<td>-334.33</td>
<td>-992.24</td>
<td>-889.42</td>
<td>-2067.15</td>
<td>-2150.97</td>
<td>-488.18</td>
</tr>
<tr>
<td></td>
<td>(350.62)</td>
<td>(130.46)</td>
<td>(36.81)</td>
<td>(42.35)</td>
<td>(129.22)</td>
<td>(133.88)</td>
<td>(505.94)</td>
<td>(488.18)</td>
</tr>
<tr>
<td>$(\rho - 1)^3$</td>
<td>-3170.25</td>
<td>(782.36)</td>
<td>(36.81)</td>
<td>(42.35)</td>
<td>(129.22)</td>
<td>(133.88)</td>
<td>(505.94)</td>
<td>(488.18)</td>
</tr>
<tr>
<td></td>
<td>-5177.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.89</td>
<td>0.93</td>
<td>0.90</td>
<td>0.80</td>
<td>0.93</td>
<td>0.92</td>
<td>0.89</td>
<td>0.80</td>
</tr>
</tbody>
</table>
2.3 Finite sample performance

Figure 2.4: Power curves for block-type tests
similar for all tests. Power becomes lower in general, and size seems to fluctuate more for different AR and MA parameters.

2.4 Conclusion

We have analyzed the behavior of a set of bootstrap unit root tests in finite samples. Moreover, we have shown the validity of two procedures that turn out to work well in finite samples.

From our simulation study we can draw several conclusions. First, ADF tests clearly perform better than DF tests, which is what we expected from our discussion about asymptotically pivotal statistics. We do not observe a clear difference between the coefficient tests and t-tests.

Second, it seems that sieve tests perform better in terms of size than block tests for ARMA models, which is in line with the results for stationary series. We also see that the stationary bootstrap test performs worse in terms of size than the block bootstrap. Added to this, there is also a practical reason to use the sieve bootstrap. The selection of the lag length can be done quite easily, and appears to work if based on an information criterion like AIC or modified AIC. On the other hand, choosing the block length on the basis of intuition is difficult, and there exist no satisfactory methods for it. Taking all this into account, for our set of models the sieve bootstrap is preferable over the block bootstrap.

Third, the choice between difference-based tests and residual-based tests is less obvious. While the residual-based tests have higher power than the difference-based tests, these tests also have higher size distortions. However, when we consider ADF sieve bootstrap test, the residual-based tests perform similarly as the difference-based tests both in terms of size and in terms of power.

These findings are in line with the simulation results reported in the previous studies mentioned in the introduction, in the way the tests perform for different ARMA parameters. Our findings however allowed us to systematically compare existing and newly proposed tests. On the basis of previous studies only, it was not clear how the various tests compared.

Summarizing, for the type of processes considered, the ADF sieve tests perform best in our simulation study. Therefore, for settings comparable to ours, we can recommend to use either the tests by Chang and Park (2003) or the ADF sieve tests based on residuals that we proposed. For other types of processes, allowing for broken trends, heteroscedasticity, etc., further research is needed.

2.A Appendix: Proofs

Our proofs are adaptations of the proofs of Psaradakis (2001) and Chang and Park (2003) (which in turn depends on Park (2002)). We only elaborate where our proofs differ from theirs due to the use of residuals instead of differences. To be specific, the residuals to
be resampled in our tests are constructed as
\[ \hat{\varepsilon}_{t,n} = y_t - \hat{\rho}_n y_{t-1} - \sum_{j=1}^{p} \hat{\phi}_{j,n} \Delta y_{t-j}, \]
(2.15)

where \( \hat{\rho}_n, \hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n} \) are the OLS estimates from the (augmented Dickey-Fuller) regression of \( y_t \) on \( y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p} \).

The residuals to be resampled in the tests of Psaradakis (2001) and Chang and Park (2003) are constructed as
\[ \tilde{\varepsilon}_{t,n} = \Delta y_t - \sum_{j=1}^{p} \tilde{\phi}_{j,n} \Delta y_{t-j}, \]
(2.16)

where \( \tilde{\phi}_{1,n}, \ldots, \tilde{\phi}_{p,n} \) are the OLS (or Yule-Walker) estimates from the regression of \( \Delta y_t \) on \( \Delta y_{t-1}, \ldots, \Delta y_{t-p} \).

Let \( \tilde{\phi}_n = (\tilde{\phi}_{1,n}, \ldots, \tilde{\phi}_{p,n})' \), \( \hat{\phi}_n = (\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n})' \) and \( x_{p,t} = (\Delta y_{t-1}, \ldots, \Delta y_{t-p})' \). Then \( \tilde{\phi}_n \) and \( \hat{\phi}_n \) are related by
\[ \hat{\phi}_n = \tilde{\phi}_n + (\hat{\rho}_n - 1) \left( \frac{1}{n} \sum_{t=p+1}^{n} x_{p,t} x_{p,t}' \right)^{-1} \left( \frac{1}{n} \sum_{t=1}^{n} x_{p,t} y_{t-1} \right) \]
(2.17)
as in Chang and Park (2002, Proof of Lemma 3.5). From this we can deduce that
\[ \hat{\phi}_{j,n} = \tilde{\phi}_{j,n} + O_p(n^{-1}) \]
(2.18)
under the null hypothesis of a unit root.

One consequence of not imposing the unit root restriction is that \( \rho \) has to be estimated so that we are only able to show some results in terms of convergence in probability instead of almost sure convergence.

Note that we only focus on the bootstrap distributions under the null, in line with most of the literature and specifically the papers that we base the new tests on. To analyze power properties, one needs to look at the bootstrap distribution under alternatives as well. In the main text we discuss the findings of Paparoditis and Politis (2005), who consider the power of these type of tests.

### 2.A.1 Proof of Theorem 1

In order to prove this theorem we need the following lemmas:

**Lemma 2.A.1.** Suppose Assumptions 2.1 (with \( r = 4 \) and \( s = 1 \)) and 2.2 hold. Then
\[ E^*[\varepsilon_{t,n}^2w] = E[(\varepsilon_t)^{2w}] + o_p(1) \]
for \( w = 1, 2 \).
(2.19)

**Proof of Lemma 2.A.1.** We adapt the proof of Bühlmann (1997, Proof of Lemma 5.3). First note that
\[ E^*[\varepsilon_{t,n}^2w] = (n-p)^{-1} \sum_{t=p+1}^{n} (\hat{\varepsilon}_{t,n} - \hat{\varepsilon}_n^{(i)})^{2w}, \]
(2.20)
where \( \hat{\varepsilon}_n^{(i)} = (n-p)^{-1} \sum_{t=p+1}^{n} \hat{\varepsilon}_{t,n} \).
We first show that
\[ \hat{e}_n^{(i)} = o_p(1). \]  
(2.21)

Note that under the null
\[ \varepsilon_t = \Delta y_t - \sum_{j=1}^{\infty} \phi_j \Delta y_{t-j} \]  
(2.22a)
\[ \hat{e}_{t,n} = y_t - \hat{\rho}_n y_{t-1} - \sum_{j=1}^{p} \hat{\phi}_{j,n} \Delta y_{t-j}. \]  
(2.22b)

Then we write
\[ \hat{e}_n^{(i)} = (n-p)^{-1} \sum_{t=p+1}^{n} (\varepsilon_t - \varepsilon_t + \hat{\varepsilon}_{t,n}) \]
\[ = (n-p)^{-1} \sum_{t=p+1}^{n} \left( \varepsilon_t - (\Delta y_t - \sum_{j=1}^{\infty} \phi_j \Delta y_{t-j}) + (y_t - \hat{\rho}_n y_{t-1} - \sum_{j=1}^{p} \hat{\phi}_{j,n} \Delta y_{t-j}) \right) \]
\[ = (n-p)^{-1} \sum_{t=p+1}^{n} \left( \varepsilon_t - (\Delta y_t - (y_t - \hat{\rho}_n y_{t-1})) \right) \]
\[ - \sum_{j=1}^{p} (\hat{\phi}_{j,n} - \phi_j) \Delta y_{t-j} + \sum_{j=p+1}^{\infty} \phi_j \Delta y_{t-j} \]
\[ = (n-p)^{-1} \sum_{t=p+1}^{n} (A_{t,n} + B_{t,n} + C_{t,n} + D_{t,n}) \]  
(2.23)

Hence it has to be shown that the four right hand side components in (2.23) are \( o_p(1) \).

It is trivial that \( (n-p)^{-1} \sum_{t=p+1}^{n} A_{t,n} \) and \( (n-p)^{-1} \sum_{t=p+1}^{n} D_{t,n} \) are \( o_p(1) \). Next we turn to \( B_{t,n} \):
\[ B_{t,n} = -\Delta y_t + y_t - \hat{\rho}_n y_{t-1} \]
\[ = (1 - \hat{\rho}_n) y_{t-1} \]  
(2.24)

Under the null \( 1 - \hat{\rho}_n = O_p(n^{-1}) \) (Chang and Park, 2002), so that
\[ (n-p)^{-1} \sum_{t=p+1}^{n} (1 - \hat{\rho}_n) y_{t-1} = (1 - \hat{\rho}_n)(n-p)^{-1} \sum_{t=p+1}^{n} y_{t-1} = o_p(1). \]  
(2.25)

Finally, we consider \( C_{t,n} \). By the Cauchy-Schwartz inequality,
\[ \left| (n-p)^{-1} \sum_{t=p+1}^{n} \sum_{j=1}^{p} (\hat{\phi}_{j,n} - \phi_j) \Delta y_{t-j} \right| \]
\[ \leq \left[ \sum_{j=1}^{p} (\hat{\phi}_{j,n} - \phi_j)^2 \right]^{1/2} \left[ (n-p)^{-1} \sum_{t=p+1}^{n} \sum_{j=1}^{p} (\Delta y_{t-j})^2 \right]^{1/2}. \]  
(2.26)
As \( \hat{\phi}_{j,n} - \phi_j = O_p\left(\left(\ln n/n\right)^{1/2}\right) + o\left(p^{-1}\right) \) (Chang and Park, 2002, Lemma 3.5) and \( p(n) = o\left(n / \ln n\right)^{1/4} \), we have

\[
\left[ \sum_{j=1}^{p} (\hat{\phi}_{j,n} - \phi_j)^2 \right]^{1/2} = \left[ \sum_{j=1}^{p} \left\{ O_p\left(\left(\ln n/n\right)^{1/2}\right) + o\left(p^{-1}\right) \right\} \right]^{1/2}
\]

\[
= \left[ \sum_{j=1}^{p} \left\{ O_p\left(\left(p(n)\right)^{1/4}\right) + o_p\left(p(n)\right)^{1/4} \right\} \right]^{1/2}
\]

\[
= O_p\left(p(n)\right)^{1/4} + o_p\left(p(n)\right)^{1/4}
\]

\[
= O_p\left(\left(\ln n/n\right)^{1/2}\right) + o\left(p^{-1}\right) \]^{1/2}.


(2.27)

Therefore,

\[
(n - p)^{-1} \sum_{t=p+1}^{n} C_t = \left[ O_p\left(\left(\ln n/n\right)^{1/2}\right) + o(p^{-1}) \right]^{1/2} O_p(p^{1/2}) = o_p(1). \tag{2.28}
\]

Having shown (2.21), we now need to show that

\[
(n - p)^{-1} \sum_{t=p+1}^{n} (\hat{\epsilon}_{t,n})^{2w} = E[(\epsilon_t)^{2w}] + o_p(1). \tag{2.29}
\]

As in (2.23), write

\[
\hat{\epsilon}_{t,n} = \epsilon_t - (\Delta y_t - (y_t - \hat{\rho}_n y_{t-1})) - \sum_{j=1}^{p} (\hat{\phi}_{j,n} - \phi_j) \Delta y_{t-j} + \sum_{j=p+1}^{\infty} \phi_j \Delta y_{t-j} \tag{2.30}
\]

Using the arguments in (2.24) to (2.28), we have

\[
(n - p)^{-1} \sum_{t=p+1}^{n} |B_{t,n}|^{2w} = o_p(1), \tag{2.31a}
\]

\[
(n - p)^{-1} \sum_{t=p+1}^{n} |C_{t,n}|^{2w} = o_p(1), \tag{2.31b}
\]

\[
(n - p)^{-1} \sum_{t=p+1}^{n} |D_{t,n}|^{2w} = o_p(1). \tag{2.31c}
\]

Then

\[
(n - p)^{-1} \sum_{t=p+1}^{n} (\hat{\epsilon}_{t,n})^{2w} = (n - p)^{-1} \sum_{t=p+1}^{n} (A_{t,n} + B_{t,n} + C_{t,n} + D_{t,n})^{2w}.
\]

If we expand the right-hand side of this last equation, we will get a sum of terms of the form

\[
(n - p)^{-1} \sum_{t=p+1}^{n} A_{t,n}^p B_{t,n}^p C_{t,n}^p D_{t,n}^p
\]

39
where \( a, b, c, d \geq 0 \) and \( a + b + c + d = 2w \). Next we apply Hölder’s inequality to each of these terms. Then, because of (2.31), all these terms are \( o_p(1) \) apart from the term \( (n - p)^{-1} \sum_{t=p+1}^{n} A_{t,n}^{2w} \). Hence,

\[
(n - p)^{-1} \sum_{t=p+1}^{n} (\hat{\varepsilon}_{t,n})^{2w} = (n - p)^{-1} \sum_{t=p+1}^{n} (\varepsilon_{t,n})^{2w} + o_p(1).
\]

We can then establish (2.29) by applying the weak law of large numbers.

Finally, we expand the right-hand side of (2.20) and again apply Hölder’s inequality to the cross-terms. The proof is then completed using (2.29) and (2.21).

Lemma 2.A.2. Suppose Assumptions 2.1 (with \( r = 4 \) and \( s = 1 \)) and 2.2 hold. Then

(a) there exists a random variable \( n_0 \) such that \( \sup_{n \geq n_0} \sum_{j=0}^{\infty} j|\hat{\psi}_{j,n} - \psi_{j}| < \infty \) in probability;

(b) \( \sup_{0 \leq j \leq \infty} |\hat{\psi}_{j,n} - \psi_{j}| = o_p(1) \);

(c) \( \text{Var}^*[\hat{\epsilon}_{t,n}] - \sigma^2 = o_p(1) \);

(d) \( \text{Var}^*[n^{-1/2} \sum_{t=1}^{n} \hat{\epsilon}_{t,n}] - \sigma^2 = o_p(1) \).

Proof of Lemma 2.A.2. We start with (a). As shown in Bühlmann (1995, Lemma 2.2) it is sufficient to prove that

\[
\sum_{j=0}^{\infty} j|\hat{\psi}_{j,n} - \psi_{j}| = o_p(1).
\]

From the triangular inequality, we have

\[
\sum_{j=0}^{\infty} j|\hat{\psi}_{j,n} - \phi_{j}| \leq \sum_{j=0}^{\infty} j|\hat{\psi}_{j,n} - \hat{\phi}_{j,n}| + \sum_{j=0}^{\infty} j|\hat{\phi}_{j,n} - \phi_{j}|.
\]

Bühlmann (1995, Proof of Lemma 3.1) shows that \( \sum_{j=0}^{\infty} j|\hat{\psi}_{j,n} - \hat{\phi}_{j,n}| = o(1) \) a.s. and furthermore we have that

\[
\sum_{j=0}^{\infty} j|\hat{\phi}_{j,n} - \phi_{j}| \leq p^2 \max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j}| = o_p(1).
\]

Hence, \( \sum_{j=0}^{\infty} j|\hat{\psi}_{j,n} - \phi_{j}| = o_p(1) \) and the proof of part (a) is completed.

For (b), see Bühlmann (1995, Proof of Theorem 3.2) for the proof.

Using Lemma 2.A.1 and parts (a) and (b), the results in (c) and (d) follow as in Psaradakis (2001, Proof of Lemma 2).

Lemma 2.A.3. Let \( S_n(r) = n^{-1/2} \sum_{t=1}^{[nr]} u_{t,n}^* \) and suppose Assumptions 2.1 (with \( r = 4 \) and \( s = 1 \)) and 2.2 hold. Then

\[
S_n(r) \Rightarrow \sigma W(r).
\]

2.A Appendix: Proofs

2.A.2 Proof of Theorem 2

Again we first need several lemmas.

Lemma 2.A.5. Let Assumption 2.1 (with \( r \geq 4 \) and \( s \geq 1 \)) hold and let \( p(n) = o((n/\ln n)^{1/2}) \). Then it follows that

(a) \( \max_{1 \leq j \leq p} |\hat{\phi}_j - \phi_j| = O_p((\ln n/n)^{1/2}) + o(p^{-s}) \)

(b) \( \hat{\sigma}_n^2 = \sigma^2 + O_p((\ln n/n)^{1/2}) + o(p^{-s}) \)

(c) \( \sum_{j=1}^p \hat{\phi}_j = \sum_{j=1}^p \phi_j + O_p((\ln n/n)^{1/2}) + o(p^{-s}) \).

Proof of Lemma 2.A.5. Part (a) and (c) follow from Chang and Park (2002, Lemma 3.5); part (b) follows from Bühlmann (1995, Proof of Theorem 3.2).

Lemma 2.A.6. Let Assumption 2.1 (with \( r \geq 4 \) and \( s \geq 1 \)) hold and let \( p(n) = o((n/\ln n)^{1/2}) \). Then \( n^{1-r/2} \mathbb{E} |\hat{\varepsilon}_{t,n}|^r \xrightarrow{p} 0 \) and

\[
W^*_n(i) = \frac{1}{\sigma_n \sqrt{n}} \sum_{k=1}^{[n]} \hat{\varepsilon}_{k,n} \xrightarrow{d} W(i) \quad \text{as} \quad n \to \infty.
\]

Proof of Lemma 2.A.6. As shown in Park (2002, Theorem 2.2) we only need to show

\[
n^{1-r/2} \mathbb{E} |\hat{\varepsilon}_{t,n}|^r = n^{1-r/2} \left( \frac{1}{n} \sum_{t=1}^{n} \hat{\xi}_{t,n} - \frac{1}{n} \sum_{t=1}^{n} \hat{\xi}_{t,n} \right)^r \xrightarrow{p} 0.
\]

Our proof will follow the lines of Park (2002, Proof of Lemma 3.2). We have that

\[
\frac{1}{n} \sum_{t=1}^{n} |\hat{\xi}_{t,n} - \frac{1}{n} \sum_{t=1}^{n} \hat{\xi}_{t,n}|^r \leq c(A_n + B_n + C_n + D_n)
\]

where

\[
A_n = \frac{1}{n} \sum_{t=1}^{n} |\varepsilon_t|^r, \quad B_n = \frac{1}{n} \sum_{t=1}^{n} |\varepsilon_{t,n} - \varepsilon_t|^r,
\]

\[
C_n = \frac{1}{n} \sum_{t=1}^{n} |\hat{\xi}_{t,n} - \varepsilon_{t,n}|^r, \quad D_n = \frac{1}{n} \sum_{t=1}^{n} |\hat{\xi}_{t,n}|^r.
\]
and

$$\varepsilon_{t,n} = \Delta y_t - \sum_{j=1}^{p} \phi_j \Delta y_{t-j}. \tag{2.38}$$

$\hat{\varepsilon}_{t,n}$ is defined in (2.15). Furthermore, let $\phi_{j,n}$ be defined such that in

$$\Delta y_t = \sum_{j=1}^{p} \phi_{j,n} \Delta y_{t-j} + \varepsilon_{t,n}, \tag{2.39}$$

$\varepsilon_{t,n}$ is uncorrelated with $\Delta y_{t-1}, \ldots, \Delta y_{t-p}$.

Hence we have to show that

$$n^{1-r/2} A_n, n^{1-r/2} B_n, n^{1-r/2} C_n \text{ and } n^{1-r/2} D_n \overset{p}{\longrightarrow} 0.$$ The results for $A_n$ and $B_n$ are shown in Park (2002, Proof of Lemma 3.2).

Next we turn to $C_n$. We write

$$\hat{\varepsilon}_{t,n} = y_{t} - \hat{\rho}_n y_{t-1} - \sum_{j=1}^{p} \hat{\phi}_{j,n} \Delta y_{t-j},$$

$$= (y_{t} - \hat{\rho}_n y_{t-1} - \Delta y_{t}) + (\Delta y_{t} - \sum_{j=1}^{p} \hat{\phi}_{j,n} \Delta y_{t-j}), \tag{2.40}$$

$$= -(\hat{\rho}_n - 1) y_{t-1} + (\varepsilon_{t,n} - \sum_{j=1}^{p} (\hat{\phi}_{j,n} - \phi_j) \Delta y_{t-j} - \sum_{j=1}^{p} (\phi_{j,n} - \phi_j) \Delta y_{t-j})$$

It then follows that

$$|\hat{\varepsilon}_{t,n} - \varepsilon_{t,n}|^r \leq c \left( |(\hat{\rho}_n - 1) y_{t-1}|^r + \sum_{j=1}^{p} (|\hat{\rho}_{j,n} - \phi_j)| \Delta y_{t-j}|^r \right) + \sum_{j=1}^{p} (|\hat{\phi}_{j,n} - \phi_j)| \Delta y_{t-j}|^r, \tag{2.41}$$

where $c = 3^{r-1}$. We define

$$C_{0n} = \frac{1}{n} \sum_{t=1}^{n} |(\hat{\rho}_n - 1) y_{t-1}|^r \tag{2.42a}$$

$$C_{1n} = \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{p} (|\hat{\rho}_{j,n} - \phi_j)| \Delta y_{t-j}|^r \tag{2.42b}$$

$$C_{2n} = \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{p} (|\hat{\phi}_{j,n} - \phi_j)| \Delta y_{t-j}|^r \tag{2.42c}$$

so that it needs to be shown that $n^{1-r/2} C_{in} \overset{a.s.}{\longrightarrow} 0$ for $i = 0, 1, 2$. The result for $C_{2n}$ follows from Park (2002, Proof of Lemma 3.2).
Again following Park (2002, Proof of Lemma 3.2), \( C_{1n} \) is majorized by

\[
\left( \max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}|^{r} \right) \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{p} |\Delta y_{t-j}|^{r} \leq \left( \max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}|^{r} \right) \frac{p}{n} \left( \sum_{t=0}^{n-1} |\Delta y_{t}|^{r} + \sum_{t=1}^{1-p} |\Delta y_{t}|^{r} \right) \leq c \left( \max_{1 \leq j \leq p} (|\hat{\phi}_{j,n} - \phi_{j,n}|^{r} + |\hat{\phi}_{j,n} - \tilde{\phi}_{j,n}|^{r}) \right) \frac{p}{n} \left( \sum_{t=0}^{n-1} |\Delta y_{t}|^{r} + \sum_{t=1}^{1-p} |\Delta y_{t}|^{r} \right) = O \left( (\ln n/n)^{r} + O_{p}(n^{-r}) \right) (p/n) = O \left( p \ln(n/n)^{r} \right) = o_{p} \left( (\ln n/n)^{r-1/2} \right).
\]

As \( r \geq 4 \), \( C_{1n} \rightarrow 0 \).

Next we consider \( C_{0n} \). Rewrite the expression in (2.42a) for \( C_{0n} \) as

\[
\frac{1}{n} \sum_{t=1}^{n} \left( |(\hat{\rho}_{n} - 1)y_{t-1}|^{r} - \hat{\rho}_{n} - 1 \right) \frac{1}{n} \sum_{t=1}^{n} |y_{t-1}|^{r} = O_{p}(1). \tag{2.43}
\]

This proves that \( n^{1-r/2}C_{n} \rightarrow 0 \).

For \( D_{n} \), we need to prove that

\[
\frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{t,n} = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_{t,n} + o_{p}(1) = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_{t} + o_{p}(1), \tag{2.44}
\]

which, by (2.40) and the result that \( \varepsilon_{t,n} = \varepsilon_{t} + \sum_{j=p+1}^{\infty} \phi_{j} \Delta y_{t-j} \), holds if

\[
\frac{1}{n} \sum_{t=1}^{n} \sum_{j=p+1}^{\infty} \phi_{j} \Delta y_{t-j} \overset{p}{\rightarrow} 0, \tag{2.45}
\]

\[
\frac{1}{n} \sum_{t=1}^{n} \sum_{j=p+1}^{\infty} (\hat{\phi}_{j,n} - \phi_{j,n}) \Delta y_{t-j} \overset{p}{\rightarrow} 0, \tag{2.46}
\]

\[
\frac{1}{n} \sum_{t=1}^{n} \sum_{j=p+1}^{\infty} (\hat{\phi}_{j,n} - \tilde{\phi}_{j,n}) \Delta y_{t-j} \overset{p}{\rightarrow} 0, \tag{2.47}
\]

\[
\frac{1}{n} \sum_{t=1}^{n} (1 - \hat{\rho}_{n}) y_{t-1} \overset{p}{\rightarrow} 0, \tag{2.48}
\]

where (2.45) and (2.46) follow from Park (2002, Proof of Lemma 3.2). For (2.47) we define

\[
N_{n} = \sum_{j=1}^{p} (\hat{\phi}_{j,n} - \phi_{j,n}) \sum_{t=1}^{n} \Delta y_{t-j} \tag{2.49}
\]

and

\[
Q_{n} = \sum_{j=1}^{p} \sum_{t=1}^{n} \Delta y_{t-j} \tag{2.50}
\]
Then \( N_n \) is dominated by

\[
Q_n \max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}|.
\] (2.51)

Park (2002, Proof of Lemma 3.2) shows that \( Q_n = o \left( pn^{1/2} (\ln n)^{1/r} (\ln \ln n)^{1+4/r} \right) \) a.s. for any \( \delta > 0 \). Furthermore,

\[
\max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}| \leq \max_{1 \leq j \leq p} (|\hat{\phi}_{j,n} - \phi_{j,n}| + |\hat{\phi}_{j,n} - \tilde{\phi}_{j,n}|)
\]

\[
= O \left( (\ln n/n)^{1/2} \right) \text{ a.s.} + O_p \left( n^{-1} \right),
\] (2.52)

from which we can conclude that \( N_n = o_p(n) \), which proves the result.

Finally, from (2.25) it is easy to see (2.48) holds as well. This completes the proof of \( D_n \) and hence of Lemma 2.A.6.

**Lemma 2.A.7.** Let Assumption 2.1 (with \( r \geq 4 \) and \( s \geq 1 \)) hold and let \( p(n) = O(\left( n/\ln n \right)^{1/2}) \). Then

\[
V_n^*(i) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor ni \rfloor} u_{k,n}^* \overset{d}{\rightarrow} \sigma \left( \sum_{j=0}^{\infty} \psi_j \right) W(i) \quad \text{as } n \to \infty.
\]

**Proof of Lemma 2.A.7.** Given Lemma 2.A.2, see Park (2002, Proof of Theorem 3.3).

**Lemma 2.A.8.** Let \( \omega^2 = (1/n) \sum_{t=1}^{n} \Delta y_t \) and \( \omega^* = (1/n) \sum_{t=1}^{n} \Delta y_t^* \). Furthermore assume that Assumption 2.1 holds with \( r \geq 4 \) and \( s \geq 1 \). and \( p(n) = o \left( (n/\ln n)^{1/2} \right) \).

Then we have for any \( \delta > 0 \), \( P[|\omega^* - \omega^2| \geq \delta] \overset{\Delta}{\rightarrow} 0 \).


**Proof of Theorem 2.2.** Given Lemmas 2.A.5 to 2.A.8, all the relevant lemmas found in Chang and Park (2003) are valid for the test with residuals. The proof then concludes by Chang and Park (2003, Proof of Theorem 2).
Chapter 3

Detrending Bootstrap Unit Root Tests

The role of detrending in bootstrap unit root tests is investigated. When bootstrapping, detrending must not only be done for the construction of the test statistic, but also in the first step of the bootstrap algorithm. It is argued that the two points should be treated separately. Asymptotic validity of sieve bootstrap ADF unit root tests is shown for test statistics based on OLS, GLS and recursive detrending. It is also shown that these tests are valid for a wide range of detrending methods in the first step of the bootstrap algorithm, and that this detrending method may differ from the one used in the construction of the test statistic. A simulation study is conducted to analyze the effects of detrending on finite sample performance of the bootstrap test. The results are that the detrending in the bootstrap algorithm only has an impact on the size properties of the test, while the power properties of the bootstrap test are completely determined by the power properties of their asymptotic counterparts.

3.1 Introduction

In recent years we have seen a large number of papers on the application of the bootstrap to nonstationary time series. The good performance of bootstrap methods in stationary time series has led people to adapt the methods to a nonstationary setting. Especially in the field of unit root testing, where finite sample size distortions are known to occur frequently, a large literature has arisen. The literature has focused mainly on how to deal with serial correlation, but it stays relatively silent on an important aspect of unit root testing, that is how to deal with deterministic trends. Our aim in this chapter is to investigate how the method of detrending impacts the performance of bootstrap unit root tests.

Allowing for deterministic trends is very relevant in practice. Many economic series such as real GDP can be thought of as containing a linear trend, while the
inclusion of an intercept is relevant for virtually every economic time series. It is therefore crucial to have tests that can take such trends into account. One way to take a trend into account is to include it in the unit root equation and make it part of the testable hypothesis, such as the Φ-tests of Dickey and Fuller (1981). The alternative way, that has become the most popular in the recent years, is to perform an initial step of detrending, with the goal of eliminating the deterministic components, and then performing the unit root test on the detrended series.

It is well known in the unit root literature that the method of detrending can have a major impact on the power of the tests. The seminal work on this topic is Elliott, Rothenberg, and Stock (1996), who show that GLS (or quasi-difference) detrending is optimal under certain settings in terms of local asymptotic power. It is also confirmed using simulations that the finite sample power of GLS detrended tests, in particular the DF-GLS test, is higher than that of their OLS detrended counterparts. Another method that has been proposed is recursive demeaning by Shin and So (2001), extended to recursive detrending by Sul (2008). Shin and So (2001) show that with recursive detrending the bias of the estimate of the autoregressive parameter decreases and correspondingly the power of the test increases.

While one might expect the power properties of the asymptotic tests to carry over to the bootstrap setting, it might also be that the method of detrending in the actual bootstrap procedure has an effect on the size of the bootstrap tests as well. The argument of Shin and So (2001) that the autoregressive parameter is estimated more precisely, is for example something which one could easily imagine to lead to an improvement in size properties of the bootstrap tests as well.

As mentioned before, the work on bootstrap unit root testing has become quite extensive. It began with Basawa, Mallik, McCormick, Reeves, and Taylor (1991a); Basawa et al. (1991b) and Ferretti and Romo (1996), who considered settings with simple correlation structures. Their work was later extended to fairly general settings by Park (2002), Chang and Park (2003), Paparoditis and Politis (2003), Swensen (2003a) and Parker et al. (2006). Some tests use the sieve bootstrap, others the moving-blocks or stationary bootstrap. All these tests are based on Dickey-Fuller type of test statistics, although some methods use the augmented DF test while others use the non-augmented. Finally, the methods differ in whether estimation is done under the null or under the alternative.

Given the large array of options, the questions becomes how to deal with that in this chapter. We choose to use one single bootstrap test, based on the following. In Palm et al. (2008a) these tests are compared and it is found that using ADF tests is clearly preferable to DF tests. Furthermore, it is found that the sieve bootstrap usually outperforms the block bootstrap, especially for linear models. Regarding the use of differences and residuals, it is strongly argued in Paparoditis and Politis (2005) to use residuals as using differences leads to a misspecified model if the alternative is true. For these reasons, we focus here on the residual-based ADF sieve bootstrap t-test, a test that performed well in the simulation study of

---

1 This conclusion is not surprising given that ADF tests are asymptotically pivotal and therefore may provide asymptotic refinements (Park, 2003) whereas DF tests are not.
Palm et al. (2008a) and was advocated by Paparoditis and Politis (2005).

The framework covered by Palm et al. (2008a) is obviously not complete by any means. Cavaliere and Taylor (2009) propose bootstrap versions of the $M$ unit root tests of Ng and Perron (2001) based on GLS detrending. Richard (2007) proposes an ARMA sieve bootstrap unit root test, instead of the regular AR sieve method. Simulations indicate that the method has quite some potential. Another interesting extension is to allow for nonstationary volatility, and apply the tests of Cavaliere and Taylor (2008). However, we would like to restrict ourselves to one specific bootstrap unit root test, in order to analyze the effects of detrending only without having to consider differences in bootstrap tests.

We extend the proof of asymptotic validity given in Palm et al. (2008a) to a setting with deterministic components, allowing for OLS, GLS and recursive detrending (RD). Most of the bootstrap unit root tests considered in the literature are based on OLS detrending (if deterministic components are taken into account at all). The exceptions are Swensen (2003b) who also considers a DF test based on GLS detrending, although in a setting without serial correlation, and Cavaliere and Taylor (2009). To our knowledge no bootstrap version of a test based on recursive detrending has yet been proposed. As a side-product we obtain a rigorous derivation of the asymptotic distribution of the ADF-GLS t-statistic and the ADF-RD t-statistic. While the limiting distributions are well known and accepted, to our knowledge no such rigorous derivations as Chang and Park (2002) can be found in the literature for ADF t-tests with deterministic components.

A simulation study investigates the impact of the method of detrending on the performance of the bootstrap unit root test. By allowing for a different method of detrending in the first step of the bootstrap procedure than in the calculation of the test statistic, we can analyze the two points separately.

An interesting question is when to apply the tests with just an intercept, and when with both an intercept and a trend. As analyzed by, among others, Harvey, Leybourne, and Taylor (2009), unnecessarily estimating the model with trend leads to a significant loss of power compared to the model with only an intercept. On the other hand the tests with intercept only are not invariant to the presence of a trend in the DGP and should therefore not be applied in this setting. This is therefore a very interesting and empirically relevant issue. However, we will not analyze this issue explicitly in combination with the bootstrap as the problem is essentially the same whether one uses the bootstrap or not. As such, the conclusions of Harvey et al. (2009) remain relevant with the application of the bootstrap as well.

The outline of the chapter is as follows. Section 3.2 will describe the model used for the theoretical analysis. The tests will be explained and their limit distribution derived in Section 3.3. The bootstrap tests are the topic of Section 3.4. In Section 3.5 a simulation study will be undertaken. Section 3.6 concludes. All proofs are contained in the appendix.

A word on notation. $\lfloor x \rfloor$ is the largest integer smaller than or equal to $x$. Convergence in distribution (probability) is denoted by $\xrightarrow{d}$ ($\xrightarrow{p}$). Bootstrap quantities (conditional on the original sample) are indicated by appending a superscript * to the standard notation. $W(r)$ denotes a univariate standard Brownian motion.
3.2 The model

We consider the following Data Generating Process (DGP), where $y_t$ is a scalar variable.

\[ y_t = x_t + \beta' z_t \]
\[ x_t = \rho x_{t-1} + u_t \]
\[ u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \psi(L) \varepsilon_t. \]  

(3.1)

The process $z_t$ is a deterministic process. In particular, we consider $z_t = 1$ and $z_t = (1, t)'$. In the remainder of the chapter we will focus on the case with linear trend, but it is clear that all results will also hold for the intercept only case.

We need the following assumption on the linear process $\psi(z)$.

Assumption 3.1.

(i) Let $\varepsilon_t$ be i.i.d. with $E\varepsilon_t = 0$, $E\varepsilon_t^2 = \sigma$ and $E\varepsilon_t^4 < \infty$.

(ii) $\psi(z) \neq 0$ for all $|z| \leq 1$, and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$.

These assumptions, which are comparable to those found in the literature (cf. Phillips and Solo, 1992; Chang and Park, 2002, 2003), are sufficient for the derivation of the asymptotic distribution of the test statistic and its bootstrap counterpart.

The null hypothesis $H_0 : \rho = 1$ corresponds to a unit root, possibly in the presence of a deterministic trend. Under the alternative $H_1 : |\rho| < 1$, with the conditions on $\psi(z)$, the process is integrated of order zero.

The treatment of the deterministic components is comparable to Elliott et al. (1996). Moreover, as in Elliott et al. (1996), we assume that the initial condition is zero, i.e. $x_0 = 0$. While this is an innocuous assumption under the null hypothesis as $x_0$ cannot be identified if a constant is included in the model, this is a crucial assumption under the alternative for the optimality of the approach of Elliott et al. (1996), as discussed by Elliott (1999), Müller and Elliott (2003), Elliott and Müller (2006) and Harvey et al. (2009) among others. A theoretical discussion on the role of the initial condition for the optimality of the tests is beyond the scope of this chapter, but we will return to the point in the simulation study in Section 3.5.

3.3 Test statistics

We consider ADF statistics with different methods of detrending. We consider two main approaches. The first approach constructs $\hat{\beta}$ as an estimator of $\beta$, and uses this to detrend the series by subtracting $\beta' z_t$ from each observation. This approach includes OLS and GLS detrending as special cases.

The second approach is based on a recursive trend adjustment. As a different term is subtracted from each observation in the series, this approach does not fit
3.3 Test statistics

into the first approach, although there are many similarities in the way we treat them.

We will first describe the OLS / GLS detrending approach and then the recursive detrending approach. Afterwards we will derive the asymptotic distributions for all methods of detrending. In the following we will focus on the t-statistic, as this is the most popular in practice. It should be clear however that all results apply to the ADF coefficient test, for example as discussed by Xiao and Phillips (1998) with GLS detrending, as well. Note that while studentizing can be problematic if the block bootstrap is employed, as discussed in Section 3.1.2 of Härdle et al. (2003), this is not the case with the sieve bootstrap.

3.3.1 OLS / GLS detrending

The OLS estimator, which is the one usually applied to detrend \( y_t \), is obtained as

\[
\hat{\beta} = \left( \sum_{t=1}^{T} z_t z_t' \right)^{-1} \left( \sum_{t=1}^{T} z_t y_t \right).
\]

(3.2)

Elliott et al. (1996) consider the construction of unit root tests that are point optimal against a local alternative \( \rho = 1 + \bar{c}T^{-1} \). Local alternatives are the relevant framework if one is interested in alternatives that are close to the null hypothesis.

Under this local alternative, we let \( z_{c,1} = z_1 \) and \( z_{c,t} = z_t - (1 + c T^{-1}) z_{t-1} = \Delta z_t - c T^{-1} z_{t-1} \) for \( t = 2, \ldots, T \). Similarly, define \( y_{c,1} = y_1 \) and \( y_{c,t} = \Delta y_t - c T^{-1} y_{t-1} \) for \( t = 2, \ldots, T \). The GLS estimator, as considered by Elliott et al. (1996), is defined as

\[
\hat{\beta} = \left( \sum_{t=1}^{T} z_{c,t} z_{c,t}' \right)^{-1} \left( \sum_{t=1}^{T} z_{c,t} y_{c,t} \right).
\]

(3.3)

The parameter \( \bar{c} \) has to be selected by the user. Elliott et al. (1996) recommend using \( \bar{c} = -7 \) for the intercept only case and \( \bar{c} = -13.5 \) for the linear trend case, as the power functions of the DF-GLS test are very close to the power envelope for these values. As these values are commonly accepted we will use them as well.

Given the choice of \( \hat{\beta} \), we can define the detrended series \( y_d^t \) as

\[
y_d^t = y_t - \hat{\beta}' z_t.
\]

(3.4)

Let \( \alpha = \rho - 1 \). The estimate for \( \alpha \) is then obtained from the OLS regression

\[
\Delta y_d^t = \alpha y_d^{t-1} + \sum_{j=1}^{p} \phi_j \Delta y_d^{t-j} + \epsilon_d^{p,t}.
\]

(3.5)

Letting \( w_{p,t}^d = (\Delta y_{d,t-1}^d, \ldots, \Delta y_{d,t-p}^d)' \), we can define

\[
\hat{\alpha} = A_T B_T^{-1},
\]

(3.6)

49
where

\[ A_T = \sum_{t=1}^{T} y_{t-1}^d \Delta y_t^d - \left( \sum_{t=1}^{T} y_{t-1}^d w_{p,t}^d \right) \left( \sum_{t=1}^{T} w_{p,t}^d w_{p,t}^d \right)^{-1} \left( \sum_{t=1}^{T} w_{p,t}^d \Delta y_t^d \right) \]

\[ B_T = \sum_{t=1}^{T} y_{t-1}^{d^2} - \left( \sum_{t=1}^{T} y_{t-1}^d w_{p,t}^d \right) \left( \sum_{t=1}^{T} w_{p,t}^d w_{p,t}^d \right)^{-1} \left( \sum_{t=1}^{T} w_{p,t}^d y_{t-1}^d \right). \]

(3.7)

We can then define the ADF t-statistic as

\[ t_{ADF} = \alpha \left[ \hat{\sigma}^2 \text{Var}(\hat{\alpha}) \right]^{-1/2} = A_T \hat{\sigma}^{-1} B_T^{-1/2}, \]

(3.8)

where \( \hat{\sigma}^2 \) is the OLS residual variance estimator in (3.5).

### 3.3.2 Recursive detrending

The second method we consider is recursive detrending, originally introduced as recursive demeaning by Shin and So (2001). Their main argument for recursive demeaning is to avoid that the explanatory variable (the first lag) is correlated to the error term, which is the case for OLS demeaning through the subtraction of the overall mean estimate. They showed using simulations that the first order autoregressive estimator under recursive demeaning is less biased than under OLS demeaning, and as a consequence, unit root tests based on recursive demeaning are more powerful. Sul (2008) extended recursive demeaning to recursive detrending in the way we will describe below.

Shin and So (2001) and Sul (2008) argued that recursive detrending should be applied slightly differently to the dependent variable and the explanatory variables; both should be detrended using the recursive mean at time \( t - 1 \). This is to achieve that, in the case of recursive demeaning, the error term in the regression of the demeaned variables is independent of the regressor. So for the dependent variable, we construct

\[ \tilde{y}_t^{rd} = y_t - 2(t - 1)^{-1} \sum_{s=1}^{t-1} y_s - T^{-1} \sum_{t=1}^{T} \left( y_t - 2(t - 1)^{-1} \sum_{s=1}^{t-1} y_s \right). \]

(3.9)

Subtracting twice the recursive mean will eliminate the linear trend. An overall demeaning is then still necessary to eliminate the intercept. For the explanatory variables we use

\[ y_t^{rd} = y_t - 2t^{-1} \sum_{s=1}^{t} y_s - T^{-1} \sum_{t=1}^{T} (y_t - 2t^{-1} \sum_{s=1}^{t} y_s). \]

(3.10)

Now define \( \tilde{\Delta} y_t^{rd} = \tilde{y}_t^{rd} - \tilde{y}_{t-1}^{rd} \). Also let \( w_{p,t}^{rd} = (\Delta y_{t-1}^{rd}, \ldots, \Delta y_{t-p}^{rd})' \). Then we
3.3 Test statistics

can give the ADF t-statistic as in (3.8), but now with

\[ A_T = \sum_{t=1}^{T} \tilde{y}_t \Delta y_t - \left( \sum_{t=1}^{T} \tilde{y}_{t-1} w_{p,t} \right) \left( \sum_{t=1}^{T} w_{p,t} \tilde{y}_{t-1} \right)^{-1} \left( \sum_{t=1}^{T} w_{p,t} \Delta y_t \right) \]

\[ B_T = \sum_{t=1}^{T} \tilde{y}_{t-1} - \left( \sum_{t=1}^{T} \tilde{y}_{t-1} w_{p,t} \right) \left( \sum_{t=1}^{T} w_{p,t} \tilde{y}_{t-1} \right)^{-1} \left( \sum_{t=1}^{T} w_{p,t} \Delta y_{t-1} \right) \]  

(3.11)

3.3.3 Asymptotic properties

In this section we will derive the limiting distributions of the test statistics. Our first goal is to derive an autoregressive approximation for the detrended series, on which the ADF test is based.

As Assumption 3.1 implies that \( \psi(z) \) is invertible, we can define \( \phi(z) = \psi(z)^{-1} = 1 - \sum_{j=1}^{\infty} \phi_j z^j \) and write

\[ u_t = \sum_{j=1}^{\infty} \phi_j u_{t-j} + \varepsilon_t. \]  

(3.12)

Now define \( \varepsilon_{p,t} \) such that

\[ u_t = \sum_{j=1}^{p} \phi_j u_{t-j} + \varepsilon_{p,t}. \]  

(3.13)

Combining (3.12) and (3.13) we obtain

\[ \varepsilon_{p,t} = \varepsilon_t + \sum_{j=p+1}^{\infty} \phi_j u_{t-j}. \]  

(3.14)

Let us first focus on OLS / GLS detrending. As \( y_t^d = y_t - \hat{\beta}' z_t \) and \( y_t = x_t + \beta' z_t \), we have that

\[ y_t^d = x_t + \beta' z_t - \hat{\beta}' z_t = x_t - (\hat{\beta} - \beta)' z_t. \]  

(3.15)

Then \( u_t = \Delta x_t = \Delta y_t^d + (\hat{\beta} - \beta)' \Delta z_t \). Now we can write

\[ \varepsilon_{p,t} = u_t - \sum_{j=1}^{p} \phi_j u_{t-j} = (\Delta y_t^d + (\hat{\beta} - \beta)' \Delta z_t) - \sum_{j=1}^{p} \phi_j (\Delta y_{t-j}^d + (\hat{\beta} - \beta)' \Delta z_{t-j}) \]

\[ = \Delta y_t^d - \sum_{j=1}^{p} \phi_j \Delta y_{t-j}^d + (\hat{\beta} - \beta)' \Delta z_t - \sum_{j=1}^{p} \phi_j (\hat{\beta} - \beta)' \Delta z_{t-j}. \]

Then, letting \( \phi_p(z) = 1 - \sum_{j=1}^{p} \phi_j z^j \), we can write \( \varepsilon_{p,t} = \varepsilon_{p,t} - (\hat{\beta} - \beta)' \phi_p(L) \Delta z_t \) such that

\[ \Delta y_t^d = \sum_{j=1}^{p} \phi_j \Delta y_{t-j}^d + \varepsilon_{p,t}. \]
Similarly we can define $\varepsilon^d_t$ such that

$$\varepsilon^d_t = \Delta y^d_t - \sum_{j=1}^{\infty} \phi_j \Delta y^d_{t-j} = \varepsilon_t - (\beta - \beta) L \Delta z_t.$$  

For the tests based on recursive detrending we can perform a similar analysis. First, note that

$$y^d_{t-1} = x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s - T^{-1} \sum_{t=1}^{T} (x_t - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s)$$

and

$$y^d_{t-i} = x_{t-i} - 2(t-i)^{-1} \sum_{s=1}^{t-i} x_s - T^{-1} \sum_{t=1}^{T} (x_t - 2(t-i)^{-1} \sum_{s=1}^{t-i} x_s).$$

Furthermore, we have that

$$\tilde{\Delta} y^d_t = u_t - T^{-1} \sum_{t=1}^{T} u_t = u_t - \bar{u} \quad \text{and} \quad \Delta y^d_{t-j} = u_{t-j} - \bar{u}_{t-j} - (f_{t-j} - \bar{f}_{t-j}),$$

where

$$f_{t-j} = 2[(t-j)(t-j-1)]^{-1} \sum_{s=1}^{t-j-1} x_s + (t-j)^{-1} x_{t-j}$$

and $\bar{f}_{t-j} = T^{-1} \sum_{t=1}^{T} f_{t-j}.$

Then, by plugging the above results into (3.13), we obtain

$$\tilde{\Delta} y^d_t = \sum_{j=1}^{p} \phi_j \Delta y^d_{t-j} = \left( \bar{u} - \sum_{j=1}^{p} \phi_j (u_{t-j} + (f_{t-j} - \bar{f}_{t-j})) \right) + \varepsilon_{p,t}$$

and

$$\Delta y^d_{t-j} = u_{t-j} - \bar{u}_{t-j} - (f_{t-j} - \bar{f}_{t-j}).$$

Similarly, we can define $\varepsilon^d_t = \varepsilon_t - \phi(L) g_t + (f_t - \bar{f})$ such that

$$\tilde{\Delta} y^d_t = \sum_{j=1}^{\infty} \phi_j \Delta y^d_{t-j} + \varepsilon^d_{p,t}.$$  

Let $y_t^{(r)d}$ denote a detrended series which is either $y_t^d$ or $y_t^{rd}$ as defined above, and let $\tilde{\Delta} y_t^{(r)d}$ be equal to either $\Delta y_t^d$ or $\Delta y_t^{rd}$. Then, letting $\tilde{\Delta} Y^{(r)d} = 52$.
Lemma 3.1. Let Assumption 3.1 and 3.2 hold. Then

(a) \( T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d} d_{p,t} \xrightarrow{d} \psi(1)^2 \sigma^4 \int_0^1 W_\gamma(r)^2 dr. \)

(b) \( T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d} \varepsilon_{p,t} \xrightarrow{d} \frac{1}{8} \psi(1) \sigma^2 (W_\gamma(1)^2 - 1). \)

(c) \( \left\| T^{-1} \sum_{t=1}^{T} w_{p,t}^{(r)d} w_{p,t}^{(r)d} \right\| = O_p(1), \)

(d) \( T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d} \varepsilon_{p,t}^{(r)d} = O_p(p^{1/2}), \)

(e) \( T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d} \varepsilon_{p,t}^{(r)d} = o_p(p^{-1/2}), \)

Assumption 3.2. Let \( p \to \infty \) and \( p = o(n^{1/2}) \) as \( n \to \infty. \)

Using the expressions developed above, one can derive the asymptotic distribution of the test statistics. In the following we will derive the limiting distributions of the ADF-OLS, ADF-GLS and ADF-RD t-statistics.

The first step in deriving the limiting distribution of the ADF t-statistics is to consider the limiting behavior of the elements of \( A_T \) and \( B_T, \) as considered in the following lemma.

Lemma 3.1. Let Assumption 3.1 and 3.2 hold. Then
3 Detrending Bootstrap Unit Root Tests

where \( \gamma = \text{ols}, \text{gls}, \text{rd} \) and

\[
\begin{align*}
W_{\text{ols}}(r) &= W(r) - (4 - 6r)\psi(1)\sigma \int_0^1 W(s)ds - (12r - 6)\psi(1)\int_0^1 sW(s)ds,
W_{\text{gls}}(r) &= W(r) - r(1 - \bar{c} + \frac{1}{3}r^2)^{-1} \left[(1 - \bar{c})W(1) + \bar{c}^2 \int_0^1 sW(s)ds\right],
W_{\text{rd}}(r) &= W(r) - 2r^{-1} \int_0^r W(s)ds - \int_0^1 \left[W(r) - 2r^{-1} \int_0^r W(s)ds\right] dr.
\end{align*}
\]

The next step is to show the consistency of the residual variance estimator, as done in the following lemma.

**Lemma 3.2.** Let Assumption 3.1 and 3.2 hold. Let \( \hat{\sigma}^2 \) be defined as

\[
\hat{\sigma}^2 = T^{-1}(\hat{\Delta}Y^{(r)d} - Y_{-1}^{(r)d})' (I - M_p^{(r)d}(M_p^{(r)d})^{-1} M_p^{(r)d}) (\hat{\Delta}Y^{(r)d} - Y_{-1}^{(r)d})
\]

Then \( \hat{\sigma}^2 \xrightarrow{} \sigma^2 \).

We can then straightforwardly derive the limiting distribution of the ADF-GLS t-statistic as given below.

**Theorem 3.1.** Let \( t_{\text{ADF,} \gamma} \) be defined as in (3.8) with detrending performed using \( \gamma = \text{ols}, \text{gls}, \text{rd} \). Let Assumption 3.1 and 3.2 hold. Then, as \( T \to \infty \), we have that

\[
t_{\text{ADF,} \gamma} \xrightarrow{d} \frac{W_{\gamma}(1)^2 - 1}{2 \left(\int_0^1 W_{\gamma}(r)^2 dr\right)^{1/2}}.
\]

### 3.4 Bootstrap tests

#### 3.4.1 Bootstrap algorithm

The bootstrap algorithm we consider is an extension of Bootstrap Test 4 given in Palm et al. (2008a). The extension is Step 1, on the treatment of deterministic components. We first give the algorithm for the class of detrending that includes OLS and GLS, and later we present the modification needed to deal with recursive detrending.

**Bootstrap Algorithm 3.1** (OLS / GLS detrending).

1. Calculate

\[
y_t^d = y_t - \tilde{\beta}' z_t.
\]

where \( \tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2)' \) is any estimator of \( \beta \) that satisfies the conditions

\[
\tilde{\beta}_1 - \beta_1 = O_p(T^{1/2}) \quad \text{and} \quad \tilde{\beta}_2 - \beta_2 = O_p(T^{-1/2}).
\]

54
2. Estimate an ADF regression of order $q$ for $y^d_t$ by OLS and save the residuals
\[ \hat{\varepsilon}_{q,t} = \Delta y^d_t - \hat{\alpha} y^d_{t-1} - \sum_{j=1}^{q} \hat{\phi}_j \Delta y^d_{t-j}, \] (3.20)

Recenter the residuals $\hat{\varepsilon}_{q,t}$ and save the centered residuals $\tilde{\varepsilon}_{q,t} = \hat{\varepsilon}_{q,t} - (n - q - 1)^{-1} \sum_t \hat{\varepsilon}_{q,t}.$

3. Resample with replacement from $\tilde{\varepsilon}_{q,t}$ to obtain bootstrap errors $\varepsilon^*_t$.

4. Build $u^*_t$ recursively as
\[ u^*_t = \sum_{j=1}^{q} \hat{\phi}_j u^*_{t-j} + \varepsilon^*_t, \] (3.21)

using the estimated parameters $\hat{\phi}_j$ from Step 2, and build $y^*_t$ as
\[ x^*_t = x^*_{t-1} + u^*_t. \] (3.22)

Finally let
\[ y^*_t = x^*_t + \beta^* z_t, \] (3.23)

See Remark 3.1 for the choice of $\beta^*$.

5. Using the bootstrap sample $y^*_t$, apply the desired method of detrending to obtain the detrended bootstrap series $y^{(r)d}_t$.

Estimate by OLS the ADF regression of order $p^*$
\[ \Delta y^{(r)d}_t = \alpha^{(r)} y^{(r)d}_{t-1} + p^* \sum_{j=1}^{p^*} \phi^*_{j} \Delta y^{(r)d}_{t-j} + \varepsilon^{(r)d}_{p^*,t}. \] (3.24)

and calculate the ADF test statistic as
\[ t^*_{ADF} = \hat{\alpha}^* \left[ \hat{\sigma}^{2*} \text{Var}(\hat{\alpha}^*) \right]^{-1/2} = A^*_T \hat{\sigma}^{-1} B^*_{T}^{-1/2}, \] (3.25)

where
\[ A^*_T = \sum_{t=1}^{T} y^{(r)d}_{t-1} \Delta y^{(r)d}_t - \left( \sum_{t=1}^{T} y^{(r)d}_{t-1} w^{(r)d}_{p^*,t} \right) \] \[ \times \left( \sum_{t=1}^{T} w^{(r)d}_{p^*,t} u^{(r)d}_{p^*,t} \right)^{-1} \left( \sum_{t=1}^{T} w^{(r)d}_{p^*,t} \Delta y^{(r)d}_t \right) \] (3.26)
\[ B^*_T = \sum_{t=1}^{T} y^{(r)d}_{t-1} \Delta y^{(r)d}_t - \left( \sum_{t=1}^{T} y^{(r)d}_{t-1} u^{(r)d}_{p^*,t} \right) \] \[ \times \left( \sum_{t=1}^{T} w^{(r)d}_{p^*,t} u^{(r)d}_{p^*,t} \right)^{-1} \left( \sum_{t=1}^{T} w^{(r)d}_{p^*,t} y^{(r)d}_{t-1} \right). \]

where $w^{(r)d}_{p^*,t} = (\Delta y^{(r)d}_{t-1}, \ldots, \Delta y^{(r)d}_{T-p^*})'$. 

55
6. Repeat Steps 3 to 5 $B$ times, obtaining bootstrap test statistics $t_{ADF}^{b}$ for $b = 1, \ldots, B$, and select the bootstrap critical value $c_{\alpha}^{*}$ as $c_{\alpha}^{*} = \max \{ c : \sum_{b=1}^{B} I(t_{ADF}^{b} < c) \leq \alpha \}$, or equivalently as the $\alpha$-quantile of the ordered $t_{ADF}^{b}$ statistics. Reject the null of a unit root if $t_{ADF}$ is smaller than $c_{\alpha}^{*}$, where $\alpha$ is the nominal level of the test.

The modification needed for recursive detrending is as follows.

**Bootstrap Algorithm 3.2 (Recursive detrending).** Replace Step 1 and 2 of the previous bootstrap algorithm by the following.

Step 1: Calculate $\tilde{y}_{t}^{rd}$ and $y_{t}^{rd}$ as in (3.9) and (3.10).

Step 2: Estimate an ADF regression of order $q$ with recursive trend adjustment by OLS and save the residuals

$$\hat{\varepsilon}_{q,t} = \tilde{\Delta}y_{t}^{rd} - \hat{\alpha}y_{t-1}^{rd} - \sum_{j=1}^{q} \hat{\phi}_{j} \Delta y_{t-j}^{rd}, \quad (3.27)$$

Recenter the residuals $\hat{\varepsilon}_{q,t}$ and save the recentered residuals $\tilde{\varepsilon}_{q,t} = \hat{\varepsilon}_{q,t} - (n - q - 1)^{-1} \sum_{t} \hat{\varepsilon}_{q,t}$.

As can be seen from the algorithms above, we allow for a different lag length in the sieve bootstrap ($q$) than in the calculation of the test statistic ($p$). Moreover, we allow for a different lag length in the calculation of the bootstrap test statistic ($p^{*}$). In general it will be a logical choice to set $q = p$, as both are based on an ADF regression. However we do not wish to impose this a priori in order to be as general as possible. For example, if the methods of detrending differ, in finite samples one might a different $p$ and $q$.

What is more important however is to allow for lag length selection of $p^{*}$ within the bootstrap, as this will improve the finite sample properties of the test. In the following we will simply denote $p^{*}$ by $p$ to lighten the notational load. This is a harmless simplification as we require $p^{*}$ to satisfy Assumption 3.2 as well, and moreover $p$ and $p^{*}$ will never be in the same part of the proof anyway. The finite sample performance of the tests might improve by imposing certain restrictions on the relation between $p$ and $p^{*}$; see Richard (2008) for more details. We will not explore this here any further.

We need the following assumption on the lag length $q$.

**Assumption 3.3.** Let $q \rightarrow \infty$ and $q = o((n/\ln n)^{1/3})$ as $n \rightarrow \infty$.

We also need the following assumption to relate $q$ to $p$ ($p^{*}$).

**Assumption 3.4.** Let $p/q \rightarrow \kappa > 1$ as $T \rightarrow \infty$, where $\kappa$ may be infinite.

This assumption essentially states that, for large $T$, $p$ should be at least as large as $q$. 

56
3.4 Bootstrap tests

**Remark 3.1.** It is unnecessary to include deterministic components in Step 4 of the bootstrap algorithm, as the tests we consider are invariant regarding the true deterministic components in the (bootstrap) DGP. Therefore we recommend setting $\beta^* = 0$ for simplicity. Note however that the arguments still hold for different values of $\beta^*$, such as $\beta^* = \tilde{\beta}$.

### 3.4.2 Detrending within the bootstrap

It is important to note that the detrending performed in the first step of the bootstrap test does not have to be the same method of detrending as the one performed in the test. As mentioned in the algorithm, the crucial aspect is that the estimator of $\beta$ satisfies the conditions $\tilde{\beta}_1 - \beta_1 = O_p(T^{1/2})$ and $\tilde{\beta}_2 - \beta_2 = O_p(T^{-1/2})$. This is the case for both OLS and GLS detrending. This statement is formalized in the following lemma.

**Lemma 3.1.** Define $\hat{\phi}_j$, $j = 1, \ldots, q$ as the OLS estimators in a regression of $u_t$ on $u_{t-1}, \ldots, u_{t-q}$ and $\tilde{\varepsilon}_{q,t}$ as the corresponding residuals. Let $\hat{\phi}_j$ be defined as in (3.20). Let $\hat{\beta}$ satisfy 3.19 and let Assumption 3.1 and 3.3 hold. Then

$$\hat{\phi}_j = \tilde{\phi}_j + O_p(T^{-1}q^{1/2}),$$

uniformly in $j = 1, \ldots, q$. Moreover,

$$\max_{1 \leq t \leq T} |\hat{\varepsilon}_{d,t} - \tilde{\varepsilon}_{q,t}| = O_p(T^{-1/2}).$$

Using the above lemma we can use the results on autoregressive approximation and the sieve bootstrap as established by Hannan and Kavalieris (1986) and Bühlmann (1995, 1997), used in a unit root setting by Park (2002) and Chang and Park (2003) (also see Remark 3.2). Given Lemma 3.1 and the results mentioned above, we can establish the limit distribution of OLS and GLS detrended ADF bootstrap tests.

We can establish a similar result for the use of recursive detrending in the first step of the bootstrap algorithm.

**Lemma 3.2.** Let $\hat{\phi}_j$ be defined as in (3.27). Let Assumption 3.1 and 3.3 hold. Then

$$\hat{\phi}_j = \tilde{\phi}_j + O_p(T^{-1}q^{1/2}),$$

uniformly in $j = 1, \ldots, q$. Moreover,

$$|\hat{\varepsilon}_{d,t} - \tilde{\varepsilon}_{q,t}| = \max_{1 \leq j \leq q} |g_{t-j}| \sum_{j=1}^q |\hat{\phi}_j| + O_p(T^{-1/2}).$$

**Remark 3.2.** One might consider using Yule-Walker instead of OLS in the sieve bootstrap to ensure that the estimated autoregression is invertible. In fact, the disadvantage of Yule-Walker is that it may have substantial finite sample bias (Poskitt, 1994). Another option if one is worried about the noninvertibility of the OLS estimates is to impose a root bound as in Burridge and Taylor (2004).
results of Hannan and Kavalieris (1986) and Bühlmann (1995, 1997) are derived for Yule-Walker estimators. However, Theorem 1 of Poskitt (1994) implies these results are valid for OLS estimation as well.

3.4.3 Bootstrap validity

In this section we want to show that the bootstrap tests are asymptotically valid. In order to establish asymptotic validity we need to show that the bootstrap t-statistic converges to the same distribution as its asymptotic counterpart if the null hypothesis is true.

The first step in the derivation of the bootstrap limit distribution is the construction of an invariance principle for $y_{t}^{rd}$ and $y_{t}^{rd}$. The several steps that are needed for the construction are detailed in the Appendix. Here we give the final invariance principle.

**Lemma 3.3.** Let Assumption 3.1 and 3.3 hold. Then

$$T^{-1/2}y_{[Tr]}^{*d} \xrightarrow{d} \sigma \psi(1)W_{\gamma}(r) \text{ in probability},$$

where $\gamma = \text{ols, gls, rd}$.

Next we must derive the autoregressive approximation on which the ADF regression is based, as for the asymptotic test. For the bootstrap error process $u_{t}^{*}$, we can write

$$u_{t}^{*} = \sum_{j=1}^{q} \hat{\phi}_{j} u_{t-j}^{*} + \epsilon_{*}^{*}. \quad (3.28)$$

In analogy with the original sample, define $\epsilon_{p,t}^{*}$ such that

$$u_{t}^{*} = \sum_{j=1}^{p} \hat{\phi}_{j} u_{t-j}^{*} + \epsilon_{p,t}^{*}. \quad (3.29)$$

Combining (3.28) and (3.29) we obtain

$$\epsilon_{p,t}^{*} = \epsilon_{t}^{*} + \sum_{j=p+1}^{q} \hat{\phi}_{j} u_{t-j}^{*}. \quad (3.30)$$

However, it is clear from our Assumption 3.4 that for large $T$ one obtains $\epsilon_{p,t}^{*} = \epsilon_{t}^{*}$. Therefore our proofs can proceed as if we set $p = q$.

In analogy with the original sample we can derive that

$$\tilde{\Delta} y_{t}^{*d} = \sum_{j=1}^{p} \hat{\phi}_{j} \Delta y_{t-j}^{*d} + \epsilon_{p,t}^{*d}$$

58
where, letting \( \hat{\phi}_p(z) = 1 - \sum_{j=1}^p \hat{\phi}_j z^j \),
\[
\varepsilon_{p,t}^{\varepsilon d} = \varepsilon_{p,t}^\ast - (\hat{\beta}^\ast - \beta^\ast)' \hat{\phi}_p(L) \Delta z_t
\]
and
\[
\varepsilon_{p,t}^{\varepsilon rd} = \varepsilon_{p,t}^\ast - \hat{\phi}_p(L) \hat{g}_t^* + f_t^* - \bar{f}_{t-j}^*.
\]
with \( \hat{g}_{t-j}^* = \bar{g}_{t-j}^* - (f_{t-j}^* - \bar{f}_{t-j}^*) \) and
\[
f_{t-j}^* = 2[(t-j)(t-j-1)]^{-1} \sum_{s=1}^{t-j-1} x_s^* + (t-j)^{-1} x_{t-j}^*.
\]
Similarly we can define
\[
\varepsilon_t^{\varepsilon d} = \varepsilon_t^\ast - \sum_{j=1}^q \hat{\phi}_j (\hat{\beta}^\ast - \beta^\ast)' \Delta z_{t-j}
\]
and
\[
\varepsilon_t^{\varepsilon rd} = \varepsilon_t^\ast - \sum_{j=1}^q \hat{\phi}_j g_{t-j}^*.
\]
(3.31)

It will then also be clear that for large \( T \) we have \( \varepsilon_{p,t}^{(r)d} = \varepsilon_t^{(r)d} \).

Now let \( \tilde{\Delta} Y^{\varepsilon d}, Y_{-1}^{\varepsilon d}, M_p^{\varepsilon d}, \varepsilon_p^{(r)d} \) and \( \hat{\Phi}_p \) be defined analogously as their original sample counterparts. Then
\[
\tilde{\Delta} Y^{\varepsilon (r)d} = M_p^{\varepsilon (r)d} \hat{\Phi}_p + \varepsilon_p^{(r)d}
\]
and
\[
A_T^\ast = Y_{-1}^{(r)d} \varepsilon_p^{(r)d} - Y_{-1}^{(r)d} M_p^{\varepsilon (r)d} \left(M_p^{(r)d} M_p^{\varepsilon (r)d}\right)^{-1} M_p^{(r)d} \varepsilon_p^{(r)d}
\]
\[
B_T^\ast = Y_{-1}^{(r)d} Y_{-1}^{\varepsilon d} - Y_{-1}^{(r)d} M_p^{\varepsilon (r)d} \left(M_p^{(r)d} M_p^{\varepsilon (r)d}\right)^{-1} M_p^{(r)d} Y_{-1}^{\varepsilon d}
\]
(3.32)

\[
\hat{\sigma}^2 = T^{-1}(\tilde{\Delta} Y^{\varepsilon (r)d} - Y_{-1}^{\varepsilon (r)d} \hat{\sigma}^\ast)' (I - M_p^{(r)d} M_p^{\varepsilon (r)d} M_p^{(r)d} M_p^{\varepsilon (r)d})^{-1} M_p^{(r)d} Y_{-1}^{\varepsilon d} \hat{\sigma}^\ast,
\]
as
\[
t_{ADF}^* = A_T^\ast \hat{\sigma}^\ast^{-1} B_T^\ast^{-1/2}.
\]

Next we can establish the bootstrap counterparts of Lemma 3.1 and 3.2.

**Lemma 3.4.** Let Assumption 3.1, 3.3, 3.2 and 3.4 hold. Then
(a) \( T^{-1} \sum_{t=1}^T \tilde{Y}_{t-1}^{(r)d} d \xrightarrow{d} \psi(1)^2 \sigma_1^4 \int_0^1 W_\gamma(r)^2 dr \) in probability.
3 Detrending Bootstrap Unit Root Tests

(b) \( T^{-1} \sum_{t=1}^{T} y_{t-1} = \frac{1}{2} \psi(1) \sigma^2 (W_{\gamma}(1)^2 - 1) \) in probability.

(c) \( \left\| \left( T^{-1} \sum_{t=1}^{T} w^r_{p,t} d^p_{t} \right)^{-1} \right\| = O_p^*(1), \)

(d) \( \left| T^{-1} \sum_{t=1}^{T} y_{t-1} w^r_{p,t} d^p_{t} \right| = O_p^*(p^{1/2}), \)

(e) \( \left| T^{-1} \sum_{t=1}^{T} w^r_{p,t} \varepsilon^r_{p,t} \right| = O_p^*(T^{-1/2} p^{1/2}). \)

Lemma 3.5. Let Assumption 3.1, 3.3, 3.2 and 3.4 hold. Let \( \hat{\sigma}^2 \) be defined as in (3.32). Then \( \hat{\sigma}^2 \overset{p^*}{\rightarrow} \sigma^2. \)

This leads to the following theorem on the asymptotic distribution of the bootstrap ADF t-statistics. Note that, as the limit distributions of the bootstrap statistic are the same as those of their asymptotic counterparts, this theorem establishes the asymptotic validity of the bootstrap ADF test.

Theorem 3.2. Let \( t^*_{ADF,\gamma} \) be defined as in (3.25) with detrending performed using \( \gamma = \text{ols, gls, rd} \). Let Assumption 3.1, 3.3, 3.2 and 3.4 hold. Then, as \( T \to \infty \), we have that

\[ t^*_{ADF,\gamma} \overset{d^*}{\rightarrow} \frac{W_{\gamma}(1)^2 - 1}{2 \left( \int_{0}^{1} W_{\gamma}(r)^2 dr \right)^{1/2}} \] in probability.

### 3.4.4 Bootstrap tests under the alternative

The asymptotic validity of the bootstrap tests that we established in the previous section is purely a property of the bootstrap tests under the null hypothesis; it does not say anything about how the bootstrap performs under the alternative hypothesis. This is what we will investigate in this section. We can discern two different alternative hypotheses, local and fixed alternatives.

Under local alternatives we want the bootstrap tests to have the same asymptotic distribution as under the null hypothesis. It is only then that the bootstrap tests will have the same asymptotic local power function as the asymptotic tests. Swensen (2003b) shows that this is the case for the OLS and GLS tests we considered here when there is no correlation in the residuals.

Under fixed alternatives we need that the bootstrap tests converge to some limiting distribution (i.e. it does not diverge) in order to achieve consistency. However, to have the highest power possible one wants again that the limit distribution is the same as under the null.

We will not go into the technical details in this chapter but we will try to show intuitively why the bootstrap tests considered here satisfy these requirements. Let us start with local alternatives. It is not difficult to see that the bootstrap tests will have the same asymptotic distribution as under the null hypothesis. Under local alternatives all rates of convergence remain the same as under the null hypothesis,
3.5 Simulations

including those of the trend estimators, which will ensure that all results, including Lemma 3.1 and Lemma 3.2, remain valid. It then follows from these Lemmas that the bootstrap tests will have the same distributions as under the null.

For fixed alternatives we may write

\[ x_t = \psi^*(L)\varepsilon_t = (1 - \rho L)^{-1}\psi(L)\varepsilon_t, \]

where \( \psi^*(L) \) is an invertible polynomial. Therefore one may approximate \( x_t \) with a finite order autoregressive model, or in other words, directly apply the sieve bootstrap of Bühlmann (1997) to it. Our ADF regression is equivalent to the direct autoregressive approximation and therefore valid as well. As such, the estimates \( \hat{\phi}_j \) will converge to their population counterparts with rates as in Hannan and Kavalieris (1986). The only complication arising is the detrending, as the trend estimators have different properties in the stationary setting. However, the trend estimators will converge at higher rates,\(^3\) which means that this will not cause any problems. For these reasons the bootstrap tests will have the same distributions under fixed alternatives as under the null hypothesis.

3.5 Simulations

3.5.1 Setup

In this section we will perform a Monte Carlo study to investigate the performance of the methods in finite samples. Our goal is twofold. First, we wish to investigate whether the power properties of the asymptotic tests carry over to the bootstrap setting. For example, it is well known that the GLS detrended test is more powerful than the OLS detrended test if the initial condition, the deviation of the initial observation from the deterministic components, is small, while it is the other way around if the initial condition is large (cf. Müller and Elliott, 2003). Therefore we will perform two sets of simulations, the first with a small (zero) initial condition, the second with a large initial condition. Our goal is certainly not to give a complete analysis of the power properties of the tests, but simply to get an idea of whether power properties carry over to the bootstrap.

The second goal is to investigate whether the method of detrending in the first step of the bootstrap procedure has an impact on the performance of the test (both size and power). As discussed in the previous section, the method of detrending in the bootstrap does not have to be the same as the method performed for the construction of the test statistic.

In order to investigate this we will consider all combinations of OLS, GLS and recursive detrending for use in the bootstrap and the construction of the test statistic, including their asymptotic variants. In particular, the tests in the tables are denoted as \( t_{a,b} \) where \( a \) is the method of detrending in the first step of the bootstrap with options

\(^3\)See for example Hamilton (1994, Chapter 16) for the OLS estimator in a model with intercept and trend.
3 Detrending Bootstrap Unit Root Tests

- $o$: OLS detrending,
- $g$: GLS detrending,
- $r$: recursive detrending,
- $a$: asymptotic test (bootstrap detrending n.a.),

while $b$ gives the method of detrending used to construct the test statistic (with options $o$, $g$ and $r$). For GLS detrending we use $\bar{c} = -13.5$ as Elliott et al. (1996) suggest.

The DGP we use in our simulations is almost identical as the one given in (3.1), except that we restrict $u_t$ to be a (stationary and invertible) ARMA(1,1) process and we generalize the initial condition. The DGP is given below.

\[
\begin{align*}
y_t &= x_t + \beta' z_t \\
x_t &= \rho x_{t-1} + u_t \\
u_t &= \phi u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}
\end{align*}
\]

where $\varepsilon_t \sim N(0,1)$ and $\rho = 1 + cT^{-1}$. We set the true deterministic components equal to zero (take $\beta = (0, 0)'$); as we perform all tests under the assumption that $z_t = (1, t)$, all tests are invariant to the true value of $\beta$.

For the first set of simulations we set the initial condition equal to zero, i.e. $x_0 = 0$. For the second set of simulations, we follow Harvey et al. (2009) and set

\[
x_0 = a \sqrt{\omega_u/(1 - \rho^2)},
\]

where $\omega_u = \lim_{T \to \infty} T^{-1} E(\sum_{t=1}^{T} u_t)^2$. We set $a = 2.5$, a value that gives a clear power advantage to the OLS test in Harvey et al. (2009).

Lag lengths are selected using the MAIC proposed by Ng and Perron (2001); we allow for separate selection of lag lengths outside and within the bootstrap. All results are obtained using 5000 simulations and the Warp-speed bootstrap method of Giacomini, Politis, and White (2007).

### 3.5.2 Results

Table 3.1 presents results for size ($c = 0$). It can be seen that all bootstrap tests perform better than the asymptotic tests. The size of the asymptotic tests is quite sensitive to the values in the simulation DGP of both the AR and MA parameter. There is undersize for most parameter combinations, although generally not too severe, while there is the familiar oversize for negative MA parameters. From the asymptotic tests, the test based on recursive detrending seems to be the most sensitive to the parameters of the dynamics, while the GLS detrended test seems to be the least sensitive.

The bootstrap tests are far less sensitive to the values of the AR and MA parameters, and have size close to the nominal level in general. The exception is the DGP with the large negative MA parameter, where there is still oversize,
3.5 Simulations

although considerably less than for the asymptotic tests. What is also noticeable
is that the bootstrap not only corrects oversize of the asymptotic tests, but also
undersize.

There are also differences between the detrending methods in terms of size. First we consider the method of detrending in the first step of the bootstrap.
Focusing on the DGP with negative MA parameters, we can see that there is clearly less oversize with GLS detrending and recursive detrending than with OLS
detrending. For the other parameter combinations the differences between the
detrending methods within the bootstrap are small.

The detrending performed in the construction of the test statistic also matters
for the size properties, and the effect is again the most noticeable for the models
with negative MA parameters. Using OLS detrending leads to the most size
distortions in DGP with negative MA parameters, while GLS detrending gives
the lowest size distortions. For the other parameter combinations the differences
between detrending methods are very small.

Tables 3.2 and 3.3 give size-adjusted powers for $c = -5$ and $c = -10$ for a
model with zero initial condition. The size-adjusted powers of the GLS and recur-
sive detrended asymptotic tests are higher in general than the power of the OLS
detrended test. This is in line with results from the literature on unit root testing
(Elliott et al., 1996; Shin and So, 2001). Which of GLS and recursive detrend-
ing has higher power depends on the specific DGP; there is not one test clearly
preferable although there seems to be a slight advantage for recursive detrending.

It can also be seen that the size-adjusted powers of the bootstrap tests are quite
close to those of the asymptotic tests. Again tests based on OLS detrending are
less powerful in general than tests based on GLS and recursive detrending. There is
not a clear advantage for either GLS or recursive detrending. Interestingly, power
properties of the bootstrap tests do not systematically depend on the method of
detrending in the bootstrap, unlike the size properties.

Based on these results, it seems that the power properties of the bootstrap
tests are completely determined by the power properties of their asymptotic coun-
terparts. We will try to confirm this conclusion by next looking at models with
large initial conditions.

Tables 3.4 and 3.5 give the size-adjusted powers for $c = -5$ and $c = -10$ for
the model with a large initial condition. The initial condition used is based on
Harvey et al. (2009), where this value led to a clear power advantage of OLS over
GLS detrending. It is not clear yet from the literature how recursive detrending
performs for such a large initial condition.

Considering the asymptotic tests first, we see that the test based on OLS de-
trending is clearly the most powerful now. The power advantage of OLS detrending
over GLS detrending is in line with the results in Harvey et al. (2009). Remark-
ably, recursive detrending performs very similarly to GLS detrending. It therefore
seems that the power of recursive detrending also decreases with an increasing
initial condition, although our simulation study here is of course too small to draw
any conclusions about that.

It can again be seen that the bootstrap tests follow their asymptotic counter-
### Table 3.1: Size ($c = 0$); lag lengths selected by MAIC

|        | 100.0 | 200.0 | 300.0 | 400.0 | 500.0 | 600.0 | 700.0 | 800.0 | 900.0 | 1000 | 1100 | 1200 | 1300 | 1400 | 1500 | 1600 | 1700 | 1800 | 1900 | 2000 | 2100 | 2200 | 2300 | 2400 | 2500 | 2600 | 2700 | 2800 | 2900 | 3000 |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
|        | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
|        | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
|        | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
|        | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |

---
### 3.5 Simulations

<table>
<thead>
<tr>
<th>n</th>
<th>$\theta$</th>
<th>$\phi$</th>
<th>$\theta$</th>
<th>$\phi$</th>
<th>$\theta$</th>
<th>$\phi$</th>
<th>$\theta$</th>
<th>$\phi$</th>
<th>$\theta$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0</td>
<td>0.077</td>
<td>0.105</td>
<td>0.093</td>
<td>0.089</td>
<td>0.106</td>
<td>0.086</td>
<td>0.106</td>
<td>0.086</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td>0</td>
<td>0.074</td>
<td>0.099</td>
<td>0.082</td>
<td>0.079</td>
<td>0.087</td>
<td>0.083</td>
<td>0.077</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>-0.4</td>
<td>0</td>
<td>0.077</td>
<td>0.086</td>
<td>0.075</td>
<td>0.076</td>
<td>0.080</td>
<td>0.074</td>
<td>0.079</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.088</td>
<td>0.086</td>
<td>0.085</td>
<td>0.086</td>
<td>0.079</td>
<td>0.082</td>
<td>0.086</td>
<td>0.082</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0</td>
<td>0.083</td>
<td>0.074</td>
<td>0.075</td>
<td>0.077</td>
<td>0.071</td>
<td>0.074</td>
<td>0.076</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0</td>
<td>0.087</td>
<td>0.085</td>
<td>0.088</td>
<td>0.087</td>
<td>0.088</td>
<td>0.087</td>
<td>0.086</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>0</td>
<td>0.091</td>
<td>0.087</td>
<td>0.090</td>
<td>0.088</td>
<td>0.088</td>
<td>0.087</td>
<td>0.088</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>0</td>
<td>0.092</td>
<td>0.089</td>
<td>0.091</td>
<td>0.091</td>
<td>0.092</td>
<td>0.091</td>
<td>0.091</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0</td>
<td>0.097</td>
<td>0.093</td>
<td>0.095</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
</tr>
</tbody>
</table>

| 100 | 0      | 0.064 | 0.105  | 0.093  | 0.089  | 0.106  | 0.086  | 0.106  | 0.086  | 0.106  |
|     | -0.8   | 0     | 0.061 | 0.099  | 0.082  | 0.079  | 0.087  | 0.083  | 0.077  | 0.084  |
|     | -0.4   | 0     | 0.074 | 0.093  | 0.091  | 0.093  | 0.091  | 0.091  | 0.091  | 0.091  |
|     | 0      | 0.086 | 0.090  | 0.086  | 0.086  | 0.084  | 0.084  | 0.084  | 0.084  | 0.084  |
|     | 0.4    | 0     | 0.073 | 0.087 | 0.083  | 0.067 | 0.091  | 0.091  | 0.091  | 0.091  |
|     | 0.8    | 0     | 0.075 | 0.089  | 0.087  | 0.082  | 0.088  | 0.086  | 0.086  | 0.086  |
|     | 1.2    | 0     | 0.090 | 0.093  | 0.091  | 0.091  | 0.091  | 0.091  | 0.091  | 0.091  |
|     | 1.6    | 0     | 0.096 | 0.094  | 0.095  | 0.096  | 0.096  | 0.096  | 0.096  | 0.096  |
|     | 2.0    | 0     | 0.097 | 0.093  | 0.095  | 0.096  | 0.096  | 0.096  | 0.096  | 0.096  |

### Table 3.2: Size-adjusted power ($c = -5$); lag lengths selected by MAIC

- $n$ represents the number of observations.
- $\theta$ and $\phi$ are the autoregressive and moving average parameters, respectively.
- The values are estimated using simulations and adjusted for size.

Note: The table provides values for various combinations of $n$, $\theta$, and $\phi$, illustrating how the size-adjusted power changes with different parameter settings.

---

For a comprehensive analysis, consult the full paper for detailed methodology and results.
### Table 3.3: Size-adjusted power ($c = -10$); lag lengths selected by MAIC

<table>
<thead>
<tr>
<th>T</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.192</td>
<td>0.275</td>
<td>0.238</td>
<td>0.200</td>
<td>0.185</td>
<td>0.154</td>
<td>0.113</td>
<td>0.080</td>
<td>0.050</td>
<td>0.030</td>
<td>0.010</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.212</td>
<td>0.238</td>
<td>0.209</td>
<td>0.189</td>
<td>0.167</td>
<td>0.134</td>
<td>0.101</td>
<td>0.069</td>
<td>0.038</td>
<td>0.017</td>
<td>0.005</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.228</td>
<td>0.252</td>
<td>0.222</td>
<td>0.199</td>
<td>0.175</td>
<td>0.141</td>
<td>0.107</td>
<td>0.074</td>
<td>0.043</td>
<td>0.018</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.240</td>
<td>0.265</td>
<td>0.235</td>
<td>0.210</td>
<td>0.185</td>
<td>0.150</td>
<td>0.116</td>
<td>0.084</td>
<td>0.053</td>
<td>0.025</td>
<td>0.006</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.250</td>
<td>0.276</td>
<td>0.246</td>
<td>0.221</td>
<td>0.195</td>
<td>0.161</td>
<td>0.126</td>
<td>0.095</td>
<td>0.065</td>
<td>0.034</td>
<td>0.010</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.260</td>
<td>0.286</td>
<td>0.256</td>
<td>0.231</td>
<td>0.205</td>
<td>0.171</td>
<td>0.136</td>
<td>0.106</td>
<td>0.076</td>
<td>0.045</td>
<td>0.012</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.270</td>
<td>0.298</td>
<td>0.266</td>
<td>0.241</td>
<td>0.215</td>
<td>0.181</td>
<td>0.146</td>
<td>0.116</td>
<td>0.087</td>
<td>0.057</td>
<td>0.025</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.280</td>
<td>0.311</td>
<td>0.276</td>
<td>0.251</td>
<td>0.225</td>
<td>0.191</td>
<td>0.156</td>
<td>0.126</td>
<td>0.098</td>
<td>0.068</td>
<td>0.037</td>
<td>0.013</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.290</td>
<td>0.325</td>
<td>0.287</td>
<td>0.262</td>
<td>0.237</td>
<td>0.202</td>
<td>0.167</td>
<td>0.137</td>
<td>0.109</td>
<td>0.081</td>
<td>0.051</td>
<td>0.021</td>
<td>0.004</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.300</td>
<td>0.339</td>
<td>0.298</td>
<td>0.276</td>
<td>0.251</td>
<td>0.217</td>
<td>0.182</td>
<td>0.152</td>
<td>0.123</td>
<td>0.096</td>
<td>0.067</td>
<td>0.036</td>
<td>0.014</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.310</td>
<td>0.354</td>
<td>0.310</td>
<td>0.292</td>
<td>0.267</td>
<td>0.233</td>
<td>0.198</td>
<td>0.168</td>
<td>0.139</td>
<td>0.109</td>
<td>0.085</td>
<td>0.060</td>
<td>0.028</td>
<td>0.004</td>
<td></td>
</tr>
</tbody>
</table>

The table above shows the size-adjusted power for different values of $T$ and $c$ with $c = -10$. The lag lengths selected by MAIC are also indicated.
3.6 Conclusion

We have investigated the role of detrending in bootstrap unit root tests. We have shown that the method of detrending used for the construction of the test statistic does not have to be the same as the method of detrending performed in the first step of the bootstrap algorithm. The bootstrap has been shown to be valid for a wide range of possible detrending methods, irrespective of the method used in the construction of the test statistic.

A simulation study has been conducted to investigate the impact of detrending on the size and power properties of the bootstrap unit root tests. The first important conclusion is that the method of detrending in the first step of the bootstrap algorithm has an impact on the size of the test but not on the power of the test. The second important conclusion is that the method of detrending used for the construction of the test statistic has a major impact on the power of the test, while having a minor impact on the size. Moreover, the power properties of the bootstrap tests are determined by the power properties of their asymptotic counterparts.

These two conclusions have the following implications. First, the choice of detrending used in the bootstrap algorithm should only be based on size considerations. In our analysis there were only small differences between the detrending methods considered. Based on the models with negative MA parameters one would recommend the use of GLS or recursive detrending, as OLS detrending led to the largest size distortions for those models. Second, the choice of the detrending used in the construction of the test statistic should be based on power considerations. As the power properties of the asymptotic tests carry over to the bootstrap setting, the choice of the detrending method for the bootstrap tests should be based on the same considerations as for the asymptotic tests. For example, one could simply adapt the arguments used in Harvey et al. (2009) when there is uncertainty over the initial condition to the bootstrap setting. Also uncertainty over the deterministic specification can be taken into account as in Harvey et al. (2009).

There are several interesting extensions possible to this chapter. First, one could consider alternative methods of detrending. We have limited our analysis to OLS, GLS and recursive detrending, but one can easily imagine other methods. Of course, our Lemma 3.1, which is the key to the proof of asymptotic validity, already allows for more general detrending methods. Second, one could extend the analysis to other types of unit root tests. Third, we could explicitly use the bootstrap to tackle the problem of uncertainty about deterministic trends and the
<table>
<thead>
<tr>
<th>$T$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
<th>$e$</th>
<th>$f$</th>
<th>$g$</th>
<th>$h$</th>
<th>$i$</th>
<th>$j$</th>
<th>$k$</th>
<th>$l$</th>
<th>$m$</th>
<th>$n$</th>
<th>$o$</th>
<th>$p$</th>
<th>$q$</th>
<th>$r$</th>
<th>$s$</th>
<th>$t$</th>
<th>$u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>160</td>
<td>120</td>
<td>210</td>
<td>110</td>
<td>50</td>
<td>350</td>
<td>90</td>
<td>30</td>
<td>60</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>25</td>
<td>50</td>
<td>30</td>
<td>40</td>
<td>100</td>
<td>25</td>
<td>50</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>10</td>
<td>150</td>
<td>110</td>
<td>200</td>
<td>100</td>
<td>40</td>
<td>360</td>
<td>95</td>
<td>35</td>
<td>65</td>
<td>25</td>
<td>45</td>
<td>105</td>
<td>20</td>
<td>45</td>
<td>35</td>
<td>55</td>
<td>105</td>
<td>20</td>
<td>45</td>
<td>35</td>
<td>55</td>
</tr>
<tr>
<td>20</td>
<td>140</td>
<td>100</td>
<td>190</td>
<td>90</td>
<td>30</td>
<td>370</td>
<td>90</td>
<td>30</td>
<td>60</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>20</td>
<td>40</td>
<td>30</td>
<td>50</td>
<td>100</td>
<td>20</td>
<td>40</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>30</td>
<td>130</td>
<td>90</td>
<td>180</td>
<td>80</td>
<td>20</td>
<td>380</td>
<td>85</td>
<td>25</td>
<td>55</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>15</td>
<td>35</td>
<td>25</td>
<td>45</td>
<td>100</td>
<td>15</td>
<td>35</td>
<td>25</td>
<td>45</td>
</tr>
<tr>
<td>40</td>
<td>120</td>
<td>80</td>
<td>170</td>
<td>70</td>
<td>10</td>
<td>390</td>
<td>80</td>
<td>20</td>
<td>50</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>10</td>
<td>30</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>10</td>
<td>30</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>50</td>
<td>110</td>
<td>70</td>
<td>160</td>
<td>60</td>
<td>0</td>
<td>380</td>
<td>75</td>
<td>15</td>
<td>45</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>0</td>
<td>25</td>
<td>15</td>
<td>35</td>
<td>100</td>
<td>0</td>
<td>25</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>60</td>
<td>100</td>
<td>60</td>
<td>150</td>
<td>50</td>
<td>10</td>
<td>390</td>
<td>70</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>10</td>
<td>30</td>
<td>10</td>
<td>40</td>
<td>100</td>
<td>10</td>
<td>30</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>70</td>
<td>90</td>
<td>50</td>
<td>140</td>
<td>40</td>
<td>20</td>
<td>380</td>
<td>65</td>
<td>5</td>
<td>35</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>20</td>
<td>50</td>
<td>5</td>
<td>40</td>
<td>100</td>
<td>20</td>
<td>50</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
<td>40</td>
<td>130</td>
<td>30</td>
<td>30</td>
<td>390</td>
<td>60</td>
<td>10</td>
<td>30</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>30</td>
<td>60</td>
<td>10</td>
<td>40</td>
<td>100</td>
<td>30</td>
<td>60</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>90</td>
<td>70</td>
<td>30</td>
<td>120</td>
<td>20</td>
<td>40</td>
<td>380</td>
<td>55</td>
<td>15</td>
<td>25</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>40</td>
<td>70</td>
<td>15</td>
<td>35</td>
<td>100</td>
<td>40</td>
<td>70</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>20</td>
<td>110</td>
<td>10</td>
<td>50</td>
<td>390</td>
<td>50</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>40</td>
<td>100</td>
<td>50</td>
<td>80</td>
<td>20</td>
<td>30</td>
<td>100</td>
<td>50</td>
<td>80</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 3.4: Size-adjusted power ($c = -1$); lag lengths selected by MAIC; large initial condition.
Table 3.5: Size-adjusted power ($c = -10$); lag lengths selected by MAIC; large initial condition
initial condition. Instead of simply adapting the ideas of Harvey et al. (2009) to the bootstrap test, we could explicitly use the bootstrap to control size exactly when the rejection strategy is based on the union of rejections of individual tests as in Harvey et al. (2009). To do so however would not be trivial. Finally, one could view detrending in a broader perspective and analyze more general trends, such as polynomial trends of higher order or broken trends.

3.A Appendix: Proofs

3.A.1 Proofs for Section 3.3

For completeness, we start with two results that are well known in the literature (Phillips and Solo, 1992). We let $W(r)$ denote a standard Brownian motion.

**Lemma 3.A.1.** Let Assumption 3.1 hold. Then

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t d \rightarrow \sigma W(r).$$

**Lemma 3.A.2.** Let Assumption 3.1 hold. Then

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \nu_t d \rightarrow \sigma \psi(1) W(r).$$

The first step is to derive the distribution of the estimator of $\beta$. Lemma 3.A.3 gives the limiting distribution for the OLS estimator of $\beta$, which is a standard result (cf. Stock, 1994). Lemma 3.A.4, in which we derive the limit distribution of the GLS estimator of $\beta$, is essentially the same as the first part of Lemma A.4 of Elliott et al. (1996). Both are included for completeness.

**Lemma 3.A.3.** Let $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$ be defined as in (3.2). Let Assumption 3.1 hold. Then

$$\left( T^{-1/2}(\hat{\beta}_1 - \beta_1) \right) \rightarrow 4\psi(1)\sigma \int_0^1 W(r) dr - 6\psi(1)\sigma \int_0^1 rW(r) dr.$$  

Proof of Lemma 3.A.3. Note that

$$\hat{\beta} - \beta = \left( \sum_{t=1}^T z_t z'_t \right)^{-1} \left( \sum_{t=1}^T z_t x_t \right).$$

Now

$$\sum_{t=1}^T z_t z'_t = \left( \sum_{t=1}^T z_t^2 \right)^{1/2} = \left( \sum_{t=1}^T z_t^2 \right) = \left( T \sum_{t=1}^T z_t^2 \right) = \left( \frac{T}{2}(T+1) \right),$$

and

$$\left( \sum_{t=1}^T z_t z'_t \right)^{-1} = 12(T+1)^{-1}(T-1)^{-1} \left( \frac{T}{2}(T+1)(2T+1) \right) = \left( \begin{array}{cc} \frac{1}{4}T(T+1) & \frac{1}{4}T(T+1) \\ -\frac{1}{4}T(T+1) & T \end{array} \right).$$

70
As
\[
\left( \sum_{t=1}^{T} z_{1t} x_t \right) = \left( \sum_{t=1}^{T} x_t \right), \tag{3.35}
\]
we have that
\[
\left( T^{-1/2} (\tilde{\beta}_1 - \beta_1) \right) /
\left( T^{1/2} (\tilde{\beta}_2 - \beta_2) \right) = \left( \begin{array}{c} 4T^{-3/2} \sum_{t=1}^{T} x_t - 6T^{-5/2} \sum_{t=1}^{T} t x_t + 12T^{-3/2} \sum_{t=1}^{T} t^3 x_t \end{array} \right) + o_p(1)
\]
\[
\left( 4\psi(1) \sigma \int_0^1 W(r) dr - 6\psi(1) \sigma \int_0^1 t W(r) dr \right) \left( -6\psi(1) \sigma \int_0^1 W(r) dr + 12\psi(1) \sigma \int_0^1 rW(r) dr \right), \tag{3.36}
\]
which follows from Lemma 3.A.2 and the continuous mapping theorem. □

**Lemma 3.A.4.** Let \( \hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)' \) be defined as in (3.3). Let Assumption 3.1 hold. Then
\[
\left( T^{-1/2} (\hat{\beta}_1 - \beta_1) \right) /
\left( T^{1/2} (\hat{\beta}_2 - \beta_2) \right) = \left( \begin{array}{c} \sigma \psi(1)(1 - \bar{c} + \frac{\bar{c}^2}{4})^{-1} \left( (1 - \bar{c})W(1) + \bar{c}^2 \int_0^1 sW(s) ds \right) \end{array} \right). \tag{3.37}
\]

**Proof of Lemma 3.A.4.** Note that
\[
\hat{\beta} - \beta = \left( \sum_{t=1}^{T} z_{t} \right)^{-1} \left( \sum_{t=1}^{T} z_{t} x_t \right), \tag{3.38}
\]
where \( x_{t,1} = x_1 \) and \( x_{t,t} = \Delta x_t - \bar{c}T^{-1} x_{t-1} \) for \( t = 2, \ldots, T \). Now define \( \bar{Z} = (z_{1,1}', \ldots, z_{T,1}')' \) and \( \bar{X} \) accordingly, such that \( \hat{\beta} - \beta = (Z' \bar{Z})^{-1} Z' \bar{X} \). Also let \( N_T = \text{diag}(T^{1/2}, 1) \) and let \( \hat{Z} = N_T \bar{Z} \). Then,
\[
\hat{\beta} - \beta = N_T (T^{-1} Z' Z)^{-1} T^{-1} Z \bar{X},
\]
or equivalently
\[
\left( T^{-1/2} (\hat{\beta}_1 - \beta_1) \right) /
\left( T^{1/2} (\hat{\beta}_2 - \beta_2) \right) = \left( \begin{array}{cc} T^{-1/2} \sum_{t=1}^{T} z_{t,1}^2 & T^{-1/2} \sum_{t=1}^{T} z_{t,1} z_{t,2t} \\ T^{-1/2} \sum_{t=1}^{T} z_{t,2t} z_{t,1} & T^{-1/2} \sum_{t=1}^{T} z_{t,2t} \end{array} \right)^{-1}
\times \left( T^{1/2} \sum_{t=1}^{T} z_{t,1} x_t \right) \left( T^{1/2} \sum_{t=1}^{T} z_{t,2t} x_t \right). \]

We first consider \( \sum_{t=1}^{T} z_{t,1} z_{t,2t} \). Note that
\[
\sum_{t=1}^{T} z_{t,1} z_{t,2t} = \left( \sum_{t=1}^{T} z_{t,1}^2 \right) \left( \sum_{t=1}^{T} z_{t,2t} \right) \left( \sum_{t=1}^{T} z_{t,2t} \right). \]

Before looking at the different parts, note that \( z_{1,1} = (1, 1)' \). Also we have that
\[
z_{1,1} = \Delta z_{1t} - \bar{c}T^{-1} z_{1,t-1} = -\bar{c}T^{-1} \text{ and } z_{2,2t} = \Delta z_{2t} - \bar{c}T^{-1} z_{2,t-1} = 1 - \bar{c}T^{-1} (t - 1). \]
Then
\[
\sum_{t=1}^{T} z_{t,1}^2 = 1 - \sum_{t=2}^{T} \bar{c}^2 T^{-2} = 1 - \bar{c}^2 T^{-2} (T - 1). \]

71
3 Detrending Bootstrap Unit Root Tests

Furthermore

\[
\sum_{t=1}^{T} z_{\epsilon,1t} z_{\epsilon,2t} = 1 - \bar{c}T^{-1} \sum_{t=2}^{T} (1 - \bar{c}T^{-1}t) = 1 - \bar{c} + \frac{1}{2} \bar{c}^2 T^{-1}(T - 1).
\]

Finally,

\[
\sum_{t=1}^{T} z_{\epsilon,2t}^2 = 1 + \sum_{t=2}^{T} (1 - 2\bar{c}T^{-1}t + \bar{c}^2 T^{-2}t^2) = \bar{c} + (1 - \bar{c})T + \frac{1}{6} \bar{c}^2 T^{-1}(T - 1)(2T - 1).
\]

Therefore we have that

\[
\left( T^{-1/2} \sum_{t=1}^{T} z_{\epsilon,1t}^2 \quad T^{-1/2} \sum_{t=1}^{T} z_{\epsilon,1t} z_{\epsilon,2t} \right) \rightarrow \left( \begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right) \quad \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}}. \quad (3.39)
\]

Next we consider \( \sum_{t=1}^{T} z_{\epsilon,1x_{\epsilon,t}} \). Note that

\[
\sum_{t=1}^{T} z_{\epsilon,1t} x_{\epsilon,t} = x_1 - \bar{c}T^{-1} \sum_{t=2}^{T} (\Delta x_t - \bar{c} T^{-1} x_{t-1}) = x_1 - \bar{c} T^{-1} (x_T - x_1) + \bar{c} T^{-2} \sum_{t=2}^{T} x_{t-1}
\]

and

\[
\sum_{t=1}^{T} z_{\epsilon,2t} x_{\epsilon,t} = x_1 + \sum_{t=2}^{T} (\Delta x_t - \bar{c} T^{-1} x_{t-1}) - \bar{c} T^{-1} \sum_{t=2}^{T} (t - 1)(\Delta x_t - \bar{c} T^{-1} x_{t-1})
\]

\[
= x_T - \bar{c} T^{-1} \sum_{t=2}^{T} (x_{t-1} + (t - 1) \Delta x_t) + \bar{c}^2 T^{-2} \sum_{t=2}^{T} (t - 1) x_{t-1}
\]

\[
= (1 - \bar{c})x_T + \bar{c} T^{-1} x_T + \bar{c}^2 T^{-1} \sum_{t=2}^{T} \frac{t - 1}{T} x_{t-1},
\]

as

\[
\sum_{t=2}^{T} (x_{t-1} + (t - 1) \Delta x_t) = \sum_{t=2}^{T} x_{t-1} + \sum_{t=2}^{T} (t - 1) x_t - \sum_{t=2}^{T} (t - 1) x_{t-1}
\]

\[
= \sum_{t=2}^{T} x_{t-1} + \sum_{t=1}^{T} (t - 1) x_t - \sum_{t=1}^{T} t x_t + T x_T = \sum_{t=2}^{T} x_{t-1} - \sum_{t=1}^{T} x_t + T x_T
\]

\[
= \sum_{t=2}^{T} x_{t-1} - \sum_{t=2}^{T} x_{t-1} - T x_T + T x_T = (T - 1) x_T.
\]

Hence,

\[
\left( T^{-1/2} \sum_{t=1}^{T} z_{\epsilon,1x_{\epsilon,t}} \right) \overset{d}{\rightarrow} \left( \begin{array}{c} 0 \\ \sigma \psi(1) (1 - \bar{c})W(1) + \sigma \psi(1) \bar{c}^2 \int_{0}^{1} s W(s) ds \end{array} \right), \quad (3.40)
\]

where the second line follows from Lemma 3.A.2 and the continuous mapping theorem. The result now follows by combining (3.39) and (3.40).

Lemma 3.A.5 provides the invariance principle for \( y_t^{(r)d} \), dependent on the method of detrending.

72
Lemma 3.A.5. Let Assumption 3.1 hold. We have that

\[ T^{-1/2} y^{(r)d}_{[Tr]} \overset{d}{\to} \sigma \psi(1) W_{\gamma}(r), \]

where \( \gamma = \text{ols}, \text{gls}, \text{rd} \) and

\[
W_{\text{ols}}(r) = W(r) - (4 - 6r)\psi(1)\sigma \int_0^1 W(s)ds - (12r - 6)\psi(1)\sigma \int_0^1 sW(s)ds,
\]

\[
W_{\text{gls}}(r) = W(r) - r(1 - \bar{c} + \frac{1}{2}\bar{c}^2)^{-1} \left[ (1 - \bar{c})W(1) + \bar{c}^2 \int_0^1 sW(s)ds \right],
\]

\[
W_{\text{rd}}(r) = W(r) - 2r^{-1} \int_0^r W(s)ds - \int_0^1 \left[ W(r) - 2r^{-1} \int_0^r W(s)ds \right] dr.
\]

Proof of Lemma 3.A.5. Let us start with OLS and GLS detrending. Note that

\[
T^{-1/2} y_{[Tr]} = T^{-1/2} y_{[Tr]} - T^{-1/2} \beta^T \varepsilon_{[Tr]}
\]

for recursive detrending, note that

\[
T^{-1/2} x_{[Tr]} = T^{-1/2} x_{[Tr]} - T^{-1/2} (\hat{\beta}_1 - \beta_1) - \frac{[Tr]}{T} T^{1/2} (\hat{\beta}_2 - \beta_2)
\]

By Lemma 3.A.2 we have that

\[
T^{-1/2} x_{[Tr]} \overset{d}{\to} \sigma \psi(1) W(r).
\]

The result for OLS and GLS detrending then follows straightforwardly from Lemma 3.A.3 and 3.A.4 respectively.

For recursive detrending, note that

\[
T^{-1/2} x_{[Tr]} = T^{-1/2} x_{[Tr]} - 2T^{-1/2} [Tr]^{-1} \sum_{s=1}^{[Tr]} x_s - T^{-3/2} \sum_{t=1}^{T} (x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s).
\]

Now

\[
T^{-1/2} x_{[Tr]} \overset{d}{\to} \sigma \psi(1) W(r)
\]

\[
T^{-1/2} [Tr]^{-1} \sum_{s=1}^{[Tr]} x_s \overset{d}{\to} \sigma \psi(1) r^{-1} \int_0^r W(s)ds
\]

\[
T^{-3/2} \sum_{t=1}^{T} (x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s) \overset{d}{\to} \sigma \psi(1) \left[ \int_0^1 \left( W(r) - 2r^{-1} \int_0^r W(s)ds \right) dr \right],
\]

which follows directly from Lemma 3.A.2 and the continuous mapping theorem.

Proof of Lemma 3.1. Part (a) follows directly from Lemma 3.A.5 using the continuous mapping theorem.

For part (b), we write

\[
T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d} \varepsilon_{p,t} = T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d} \varepsilon_{t} + T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d} (\hat{\varepsilon}_{p,t} - \varepsilon_{p,t}).
\]

We want to show that \( T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d} (\hat{\varepsilon}_{p,t} - \varepsilon_{p,t}) = o_p(1) \). Let us start with OLS and GLS detrending. Note that

\[
\varepsilon_{p,t} - \varepsilon_t = \sum_{j=p+1}^{\infty} \phi_j \varepsilon_{t-j} + \sum_{j=p+1}^{\infty} \phi_j (\hat{\beta} - \beta)' \Delta z_{t-j}.
\]
As $y_t^d = x_t - (\hat{\beta} - \beta)' z_t$, we have

$$T^{-1} \sum_{t=1}^{T} y_{t-1}(\hat{\epsilon}_{p,t}^d - \hat{d}_t) = T^{-1} \sum_{t=1}^{T} x_{t-1} \left( \sum_{j=p+1}^{\infty} \phi_j u_{t-j} + x_{t-1} \sum_{j=p+1}^{\infty} \phi_j (\hat{\beta} - \beta)' \Delta z_{t-j} \right)$$

$$- (\hat{\beta} - \beta)' z_{t-1} \sum_{j=p+1}^{\infty} \phi_j u_{t-j}$$

$$- (\hat{\beta} - \beta)' z_{t-1} \sum_{j=p+1}^{\infty} \phi_j (\hat{\beta} - \beta)' \Delta z_{t-j}$$

$$= A_T^b + B_T^b - C_T^b - D_T^b.$$ 

It follows from Chang and Park (2002, Proof of Lemma 3.1a) that $A_T^b = o_p(1)$. Also note that $\Delta z_t = (0, 1)'$. Then,

$$B_T^b = T^{-1} \sum_{t=1}^{T} x_{t-1}(\hat{\beta}_2 - \beta_2) \sum_{j=p+1}^{\infty} \phi_j = (\hat{\beta}_2 - \beta_2) \sum_{j=p+1}^{\infty} \phi_j T^{-1} \sum_{t=1}^{T} x_{t-1}$$

$$= O_p(T^{-1/2})o(p^{-1})O_p(T^{1/2}) = o_p(p^{-1}).$$

Furthermore,

$$C_T^b = T^{-1} \sum_{t=1}^{T} (\hat{\beta}_1 - \beta_1) \left( \sum_{j=p+1}^{\infty} \phi_j u_{t-j} \right) + T^{-1} \sum_{t=1}^{T} (\hat{\beta}_2 - \beta_2) t \left( \sum_{j=p+1}^{\infty} \phi_j u_{t-j} \right)$$

$$= C_T^{1b} + C_T^{2b}.$$ 

Define $\psi_{p,j}$ such that

$$\sum_{j=p+1}^{\infty} \phi_j u_{t-j} = \sum_{j=p+1}^{\infty} \psi_{p,j} \hat{\epsilon}_{t-j}$$

and note that $\sum_{j=p+1}^{\infty} \psi_{p,j} = o(p^{-1})$ (Chang and Park, 2002, Proof of Lemma 3.1a). Then

$$C_T^{1b} = T^{-1} (\hat{\beta}_1 - \beta_1) \sum_{j=p+1}^{\infty} \psi_{p,j} \sum_{t=1}^{T} \hat{\epsilon}_{t-j} = T^{-1} O_p(T^{1/2})o(p^{-1})O_p(T^{1/2}) = o_p(p^{-1})$$

and

$$C_T^{2b} = T^{-1} (\hat{\beta}_2 - \beta_2) \sum_{j=p+1}^{\infty} \phi_j \sum_{t=1}^{T} t \hat{\epsilon}_{t-j} = T^{-1} O_p(T^{-1/2})o(p^{-1})O_p(T^{3/2}) = o_p(p^{-1}).$$

Finally,

$$D_T^b = T^{-1} \sum_{t=1}^{T} (\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) \sum_{j=p+1}^{\infty} \phi_j + T^{-1} \sum_{t=1}^{T} (\hat{\beta}_2 - \beta_2)^2 t \sum_{j=p+1}^{\infty} \phi_j$$

$$= D_T^{1b} + D_T^{2b}.$$
3.A Appendix: Proofs

Then

\[ D_{1T}^b = (\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) \sum_{j=p+1}^{\infty} \phi_j = O_p(T^{1/2}O_p(T^{-1/2})o(p^{-1}) = o_p(p^{-1}), \]

\[ D_{2T}^b = (\hat{\beta}_2 - \beta_2)^2 T^{-1} \sum_{t=1}^{T} \sum_{j=p+1}^{\infty} \phi_j = O_p(T^{-1})O_T o(p^{-1}) = o_p(p^{-1}). \]

For recursive detrending, note that

\[ \varepsilon_{p,t} - \varepsilon_{t \tau}^d = \sum_{j=p+1}^{\infty} \phi_j u_{t-j} + \sum_{j=p+1}^{\infty} \phi_j g_{t-j}. \]

As \( y_{t-1}^d = x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s - T^{-1} \sum_{s=1}^{T} (x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s), \) we have

\[ T^{-1} \sum_{t=1}^{T} y_{t-1}^d (\varepsilon_{p,t} - \varepsilon_{t \tau}^d) = T^{-1} \sum_{t=1}^{T} \left[ x_{t-1} \sum_{j=p+1}^{\infty} \phi_j u_{t-j} + x_{t-1} \sum_{j=p+1}^{\infty} \phi_j g_{t-j} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s \sum_{j=p+1}^{\infty} \phi_j u_{t-j} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s \sum_{j=p+1}^{\infty} \phi_j g_{t-j} \right] \]

\[ - T^{-1} \sum_{t=1}^{T} (x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s) \sum_{j=p+1}^{\infty} \phi_j u_{t-j} \]

\[ - T^{-1} \sum_{t=1}^{T} (x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s) \sum_{j=p+1}^{\infty} \phi_j g_{t-j} \]

\[ = A_T^b + B_T^b - C_T^b - D_T^b - E_T^b - F_T^b. \]

It again follows from Chang and Park (2002, Proof of Lemma 3.1a) that \( A_T^b = o_p(1). \)

Then,

\[ B_T^b = T^{-1} \sum_{t=1}^{T} x_{t-1} \sum_{j=p+1}^{\infty} \phi_j g_{t-j} = \sum_{j=p+1}^{\infty} \phi_j T^{-1} \sum_{t=1}^{T} x_{t-1} g_{t-j} \]

\[ = \sum_{j=p+1}^{\infty} \phi_j T^{-1} \sum_{t=1}^{T} x_{t-1} (\bar{u}_t - f_{t-j} + \bar{f}_{t-j}) = \sum_{j=p+1}^{\infty} \phi_j T^{-2} \sum_{t=1}^{T} x_{t-1} \sum_{t=1}^{T} u_{t-j} \]

\[ + \sum_{j=p+1}^{\infty} \phi_j T^{-1} \sum_{t=1}^{T} x_{t-1} \left( 2[(t-j)(t-j-1)]^{-1} \sum_{s=1}^{t-j-1} x_s + (t-j)^{-1} x_{t-j} \right) \]

\[ + \sum_{j=p+1}^{\infty} \phi_j T^{-2} \sum_{t=1}^{T} x_{t-1} \sum_{t=1}^{T} \left( 2[(\tau-j)(\tau-j-1)]^{-1} \sum_{s=1}^{\tau-j-1} x_s + (\tau-j)^{-1} x_{t-j} \right) \]

\[ = o(p^{-1})O_p(T^{-1/2})O_p(T^{1/2}) + o(p^{-1})O_T o(p^{-1})O_p(T^{-1/2})O_p(T^{1/2}) \]

\[ = o_p(p^{-1}), \]

75
which follows straightforwardly from Lemma 3.A.2 and the continuous mapping theorem. Similarly,

\[ C_{rb}^{xh} = T^{-1} \sum_{t=1}^{T} 2(t-1)^{-1} \sum_{s=1}^{t-1} \sum_{j=p+1}^{\infty} \phi_j u_{t-j} \]

\[ = T^{-1} \sum_{j=p+1}^{\infty} \psi_{p,j} \sum_{t=1}^{T} 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s \varepsilon_{t-j} = T^{-1} o(p^{-1})O_p(T) = o_p(p^{-1}). \]

Also,

\[ D_{rb}^{xh} = T^{-1} \sum_{t=1}^{T} (2(t-1)^{-1} \sum_{s=1}^{t-1} \sum_{j=p+1}^{\infty} \phi_j g_{t-j} \]

\[ = T^{-1} \sum_{j=p+1}^{\infty} \phi_j \sum_{t=1}^{T} 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s g_{t-j} \]

\[ = T^{-1} o(p^{-1})O_p(T) = o_p(p^{-1}). \]

Next consider

\[ E_{rb}^{xh} = T^{-2} \sum_{\tau=1}^{T} (x_{\tau-1} - 2(\tau-1)^{-1} \sum_{s=1}^{\tau-1} \sum_{j=p+1}^{\infty} \phi_j u_{t-j} \]

\[ = T^{-2} \sum_{\tau=1}^{T} (x_{\tau-1} - 2(\tau-1)^{-1} \sum_{s=1}^{\tau-1} \sum_{j=p+1}^{\infty} \psi_{p,j} \sum_{t=1}^{T} \varepsilon_{t-j} \]

\[ = T^{-2} O_p(T^{3/2})o(p^{-1})O_p(T^{1/2}) = o_p(p^{-1}). \]

Finally,

\[ F_{rb}^{xh} = T^{-2} \sum_{\tau=1}^{T} (x_{\tau-1} - 2(\tau-1)^{-1} \sum_{s=1}^{\tau-1} \sum_{j=p+1}^{\infty} \phi_j g_{t-j} \]

\[ = T^{-2} O_p(T^{3/2})o(p^{-1})O_p(T^{1/2}) = o_p(p^{-1}). \]

Hence,

\[ T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d,(r)d} = T^{-1} \sum_{t=1}^{T} y_{t-1}^{(r)d,(r)d} + o_p(1). \]

The next part we will only show in detail for GLS detrending, but the other two methods follow in exactly the same way. To lighten the notational load, let

\[ G(\bar{c}) = (1 - \bar{c} + \frac{1}{3} \bar{c}^2)^{-1} \left[ (1 - \bar{c})W(1) + \bar{c}^2 \int_0^1 sW(s)ds \right], \]

such that \( W_{gls}(r) = W(r) - rG(\bar{c}). \)

76
Finally, Lemma 3.2), i.e. $\Omega_{pp}$

$$T^{-1} \sum_{t=1}^{T} y_{t-1} \varepsilon_t = T^{-1} \sum_{t=1}^{T} x_{t-1} \varepsilon_t - T^{-1} \sum_{t=1}^{T} x_{t-1} \phi(L)(\hat{\beta} - \beta)' \Delta z_t$$

$$- T^{-1} \sum_{t=1}^{T} (\hat{\beta} - \beta)' z_t \varepsilon_t + T^{-1} \sum_{t=1}^{T} (\hat{\beta} - \beta)' z_t \phi(L)(\hat{\beta} - \beta)' \Delta z_t$$

$$= E_T^b - F_T^b - G_T^b + H_T^b.$$

Then

$$E_T^b = T^{-1} \sum_{t=1}^{T} x_{t-1} \varepsilon_t \xrightarrow{d} \frac{1}{2} \psi(1) \sigma^2 (W(1)^2 - 1)$$

by Chang and Park (2002, Lemma 3.1a). Furthermore,

$$F_T^b = T^{-1} (\hat{\beta} - \beta) \phi(1) \sum_{t=1}^{T} x_{t-1} = T^{1/2} (\hat{\beta} - \beta) \psi(1) \sum_{t=1}^{T} x_t$$

$$\xrightarrow{d} \psi(1) \sigma^2 G(\hat{\varepsilon}) \int_0^1 W(r) dr,$$

by Lemma 3.A.4, Lemma 3.A.1 and the continuous mapping theorem. Furthermore,

$$G_T^b = T^{-1/2} (\hat{\beta} - \beta) T^{-1/2} \sum_{t=1}^{T} \varepsilon_t + T^{1/2} (\hat{\beta} - \beta) T^{-3/2} \sum_{t=1}^{T} \varepsilon_t$$

$$\xrightarrow{d} 0 + \psi_1 \sigma^2 G(\hat{\varepsilon}) \int_0^1 r dW(r).$$

Finally,

$$H_T^b = T^{-1} \sum_{t=1}^{T} (\hat{\beta} - \beta) \phi(1)(\hat{\beta} - \beta) + T^{-1} \sum_{t=1}^{T} \phi(1)(\hat{\beta} - \beta)^2 t$$

$$= \psi(1)^{-1} T^{1/2} (\hat{\beta} - \beta) T^{1/2} (\hat{\beta} - \beta) + \psi(1)^{-1} T (\hat{\beta} - \beta)^2 \frac{1}{2} (1 + T^{-1})$$

$$\xrightarrow{d} 0 + \frac{1}{2} \psi(1) \sigma^2 G(\hat{\varepsilon})^2.$$

Putting the above results together and using Itô’s Lemma we get that

$$T^{-1} \sum_{t=1}^{T} y_{t-1} \varepsilon_t \xrightarrow{d} \frac{1}{2} \psi(1) \sigma^2 \left\{ W(1)^2 - 1 - 2 G(\hat{\varepsilon}) \left[ \int_0^1 W(r) dr + \int_0^1 r dW(r) \right] + G(\hat{\varepsilon})^2 \right\}$$

$$- \frac{1}{2} \psi(1) \sigma^2 \left\{ W(1)^2 - 2 G(\hat{\varepsilon}) W(1) + G(\hat{\varepsilon})^2 - 1 \right\}$$

$$= \frac{1}{2} \psi(1) \sigma^2 (W_{st}(1)^2 - 1).$$

This concludes the proof of part (b).

We continue with (c). Let $\Omega_{pp}$ be defined as in Chang and Park (2002, Proof of Lemma 3.2), i.e. $\Omega_{pp} = (\Gamma_{i-j})_{i,j=1}^{p}$ where $\Gamma_k = E(u_t u_{t-k}).$
Again we start with OLS / GLS detrending. Let \( \Delta z_{p,t} = (\Delta z_{t-1}, \ldots, \Delta z_{t-p})' \). Then \( w_{p,t}^d = w_{p,t} - \Delta z_{p,t}(\hat{\beta} - \beta) \) and

\[
\left| T^{-1} \sum_{t=1}^{T} w_{p,t}^d w_{p,t}^d - \Omega_{pp} \right| \leq \left| T^{-1} \sum_{t=1}^{T} w_{p,t} w_{p,t} - \Omega_{pp} \right| + 2 \left| T^{-1} \sum_{t=1}^{T} w_{p,t}(\hat{\beta} - \beta)' \Delta z_{p,t}' \right| \\
+ \left| T^{-1} \sum_{t=1}^{T} \Delta z_{p,t}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \Delta z_{p,t}' \right|
\]

= \( A_T^r + 2B_T^r + C_T^r \).

By Berk (1974, Proof of Lemma 3) and Chang and Park (2002, Proof of Lemma 3.2a) \( A_T^r = O_p(T^{-1/2}p) \). Next consider

\[
T^{-1} \sum_{t=1}^{T} u_{t-1}(\hat{\beta} - \beta)' \Delta z_{t-1-j} = (\hat{\beta}_2 - \beta_2)T^{-1} \sum_{t=1}^{T} u_{t-1} = O_p(T^{-1}).
\]

As this holds uniformly in \( i, j = 1, \ldots, p \), we can conclude that \( B_T^r = O_p(T^{-1}p) \). Finally consider

\[
T^{-1} \sum_{t=1}^{T} (\hat{\beta} - \beta)' \Delta z_{t-1} (\hat{\beta} - \beta)' \Delta z_{t-1-j} = (\hat{\beta}_2 - \beta_2)^2 = O_p(T^{-1}).
\]

As this holds uniformly in \( i, j = 1, \ldots, p \), we can conclude that \( C_T^r = O_p(T^{-1}p) \).

For recursive detrending, we can write

\[
\left| T^{-1} \sum_{t=1}^{T} w_{p,t}^d w_{p,t}^d - \Omega_{pp} \right| \leq \left| T^{-1} \sum_{t=1}^{T} w_{p,t} w_{p,t} - \Omega_{pp} \right| + 2 \left| T^{-1} \sum_{t=1}^{T} w_{p,t} g_{p,t} \right| \\
+ \left| T^{-1} \sum_{t=1}^{T} g_{p,t} g_{p,t} \right|
\]

= \( A_T^{rc} + 2B_T^{rc} + C_T^{rc} \),

where \( g_{p,t} = (g_{t-1}, \ldots, g_{t-p})' \) is the recursive trend adjustment.

Of course again \( A_T^r = O_p(T^{-1/2}p) \). Next consider

\[
T^{-1} \sum_{t=1}^{T} u_{t-1} g_{t-1-j} = O_p(T^{-1}).
\]

As this holds uniformly in \( i, j = 1, \ldots, p \), we can conclude that \( B_T^{rc} = O_p(T^{-1}p) \). Finally consider

\[
T^{-1} \sum_{t=1}^{T} g_{t-1} g_{t-1-j} = O_p(T^{-1}).
\]

As this holds uniformly in \( i, j = 1, \ldots, p \), we can conclude that \( C_T^{rc} = O_p(T^{-1}p) \).

The proof (for all detrending methods) now follows as in Chang and Park (2002, Proof of Lemma 3.2a).
Next we look at (d). For OLS / GLS detrending, we have that

\[
T^{-1} \sum_{t=1}^{T} y_{t-1} u_{p,t}^d \leq T^{-1} \sum_{t=1}^{T} x_{t-1} w_{p,t} + T^{-1} \sum_{t=1}^{T} x_{t-1} \Delta z_{p,t} (\hat{\beta} - \beta) \\
+ T^{-1} \sum_{t=1}^{T} (\beta' - \beta') z_{t} w_{p,t} + T^{-1} \sum_{t=1}^{T} (\beta' - \beta') z_{t} \Delta z_{p,t} (\hat{\beta} - \beta) \\
= A_T^d + B_T^d + C_T^d + D_T^d.
\]

Now \(A_T^d = O_p(p^{1/2})\) by Chang and Park (2002, Proof of Lemma 3.2b). Next look at

\[
T^{-1} \sum_{t=1}^{T} x_{t-1} (\hat{\beta} - \beta') \Delta z_{t-j} = T^{-1} (\hat{\beta}_2 - \beta_2) \sum_{t=1}^{T} x_{t-1} \\
= T^{1/2} (\hat{\beta}_2 - \beta_2) T^{-3/2} \sum_{t=1}^{T} x_{t-1} = O_p(1)
\]

from which we can conclude that \(B_T^d = O_p(p^{1/2})\). Next consider

\[
T^{-1} \sum_{t=1}^{T} (\beta' - \beta') z_{t} u_{t-j} = T^{-1/2} (\hat{\beta}_1 - \beta_1) T^{-1/2} \sum_{t=1}^{T} u_{t-j} \\
+ T^{1/2} (\hat{\beta}_2 - \beta_2) T^{-3/2} \sum_{t=1}^{T} t u_{t-j} = O_p(1).
\]

This shows that \(C_T^d = O_p(p^{1/2})\). Finally, we have

\[
T^{-1} \sum_{t=1}^{T} (\hat{\beta} - \beta') z_{t} (\hat{\beta} - \beta') \Delta z_{t-j} = (\hat{\beta}_1 - \beta_1) (\hat{\beta}_2 - \beta_2) + T (\hat{\beta}_2 - \beta_2)^2 T^{-1} \sum_{t=1}^{T} t \\
= O_p(1)
\]

by which \(D_T^d = O_p(p^{1/2})\). This concludes the proof for part (d).

For recursive detrending we have

\[
T^{-1} \sum_{t=1}^{T} y_{t-1}^d w_{p,t}^d \leq T^{-1} \sum_{t=1}^{T} x_{t-1} w_{p,t} + T^{-1} \sum_{t=1}^{T} x_{t-1} g_{p,t} \\
+ 2 \left| T^{-1} \sum_{t=1}^{T} (t-1) x_{t-1} w_{p,t} \right| + 2 \left| T^{-1} \sum_{t=1}^{T} (t-1) x_{t-1} g_{p,t} \right| \\
+ 2 \left| \frac{T}{T-1} \sum_{t=1}^{T} \left( x_{t-1} - 2(t-1) \frac{x_{t-1}}{2} \right) w_{p,t} \right| \\
+ \left| \frac{2}{T-1} \sum_{t=1}^{T} \left( x_{t-1} - 2(t-1) \frac{x_{t-1}}{2} \right) g_{p,t} \right| \\
= A_T^{rd} + B_T^{rd} + C_T^{rd} + 2D_T^{rd} + F_T^{rd}.
\]

79
Again we have that $A_T^d = O_p(p^{1/2})$ by Chang and Park (2002, Proof of Lemma 3.2b). Next consider

$$T^{-1} \sum_{t=1}^T x_{t-1} g_{t-j} = T^{-1} \sum_{t=1}^T x_{t-1} = O_p(1),$$

from which we can conclude that $B_T^d = O_p(p^{1/2})$. As

$$T^{-1} \sum_{t=1}^T (t - 1)^{-1} \sum_{s=1}^t x_s u_{t-j} = O_p(1)$$

and

$$T^{-1} \sum_{t=1}^T (t - 1)^{-1} \sum_{s=1}^t x_s g_{t-j} = O_p(1),$$

we have that $C_T^d = O_p(p^{1/2})$ and $D_T^d = O_p(p^{1/2})$. Moreover, as

$$T^{-1} \sum_{t=1}^T \left( x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_{s+1} \right) T^{-1} \sum_{t=1}^T u_{t-j} = O_p(T^{1/2})O_p(T^{-1/2}) = O_p(1)$$

and

$$T^{-1} \sum_{t=1}^T \left( x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_{s+1} \right) T^{-1} \sum_{t=1}^T g_{t-j} = O_p(T^{1/2})O_p(T^{-1/2}) = O_p(1),$$

we have that $E_T^d = O_p(p^{1/2})$ and $F_T^d = O_p(p^{1/2})$. This concludes the proof for part (d).

For part (e) we can write for OLS / GLS detrending

$$\left| T^{-1} \sum_{t=1}^T w_{p,t} \varepsilon_{p,t} \right| \leq \left| T^{-1} \sum_{t=1}^T w_{p,t} \varepsilon_{p,t} \right| + \left| T^{-1} \sum_{t=1}^T \Delta z_{p,t}(\hat{\beta} - \beta) \varepsilon_{p,t} \right|$$

$$+ \left| T^{-1} \sum_{t=1}^T w_{p,t} \phi_p(L)(\hat{\beta} - \beta)' \Delta z_t \right|$$

$$+ \left| T^{-1} \sum_{t=1}^T \Delta z_{p,t}(\hat{\beta} - \beta) \phi_p(L)(\hat{\beta} - \beta)' \Delta z_t \right|$$

$$= A_T^d + B_T^d + C_T^d + D_T^d.$$

By Chang and Park (2002, Proof of Lemma 3.2c) we have that $A_T^d = o_p(p^{-1/2})$. Now consider

$$T^{-1} \sum_{t=1}^T (\hat{\beta} - \beta)' \Delta z_{t-j} \varepsilon_{p,t} = (\hat{\beta}_2 - \beta_2) T^{-1} \sum_{t=p+1}^\infty \left( \sum_{j=p+1}^\infty \psi_{p,j} \varepsilon_{t-j} + \varepsilon_t \right)$$

$$= (\hat{\beta}_2 - \beta_2) \sum_{j=p+1}^\infty \psi_{p,j} T^{-1} \sum_{t=1}^T \varepsilon_{t-j} + (\hat{\beta}_2 - \beta_2) T^{-1} \sum_{t=1}^T \varepsilon_t$$

$$= O_p(T^{-1/2})O_p(T^{-1/2}) + O_p(T^{-1/2})O_p(T^{-1/2}) = O_p(T^{-1}).$$
Hence, $B_T \equiv O_p(T^{-1} p^{1/2})$. Next, consider

$$T^{-1} \sum_{l=1}^T u_{t-j} \phi_p(L)(\beta - \beta)' \Delta z_t = (\beta_2 - \beta_2) T^{-1} \sum_{l=1}^T u_{t-j} \left( 1 - \frac{p}{i=1} \phi_i \right) = O_p(T^{-1}),$$

which shows that $C_T^\epsilon = O_p(T^{-1} p^{1/2})$. Finally,

$$T^{-1} \sum_{l=1}^T (\beta - \beta)' \Delta z_t \phi_p(L)(\beta - \beta)' \Delta z_t = (\beta_2 - \beta_2)^2 \left( 1 - \frac{p}{i=1} \phi_i \right) = O_p(T^{-1}),$$

by which $D_T^\epsilon = O_p(T^{-1} p^{1/2})$.

For recursive detrending we can write

$$\left| T^{-1} \sum_{l=1}^T w_{p,t} \xi_p,t \right| \leq \left| T^{-1} \sum_{l=1}^T w_{p,t} \xi_p,t \right| + \left| T^{-1} \sum_{l=1}^T g_{p,t} \xi_p,t \right|
+ \left| T^{-1} \sum_{l=1}^T w_{p,t} \left( \phi_p(L)g_t - (f_t - \bar{f}) \right) \right|
+ \left| T^{-1} \sum_{l=1}^T g_{p,t} \left( \phi_p(L)g_t - (f_t - \bar{f}) \right) \right|
= A_T^{\epsilon r} + B_T^{\epsilon r} + C_T^{\epsilon r} + D_T^{\epsilon r}.$$

We have that $A_T^{\epsilon r} = o_p(p^{-1/2})$. For $B_T^{\epsilon r}$, consider

$$T^{-1} \sum_{l=1}^T g_{t-j} \xi_p,t = T^{-1} \sum_{l=1}^T g_{t-j} \left( \sum_{j=p+1}^{\infty} \psi_{p,j} \xi_{t-j} + \xi_t \right)
= \sum_{j=p+1}^{\infty} \psi_{p,j} T^{-1} \sum_{l=1}^T g_{t-j} \xi_{t-j} + T^{-1} \sum_{l=1}^T g_{t-j} \xi_t
= o(p^{-1})O_p(T^{-1}),$$

by which we may conclude that $B_T^{\epsilon r} = o_p(T^{-1} p^{-1/2})$. Next consider

$$T^{-1} \sum_{l=1}^T u_{t-j} \left( \psi_p(L)g_t - (f_t - \bar{f}) \right) = T^{-1} \sum_{l=1}^T u_{t-j} (\bar{u} - \sum_{i=1}^p \phi \phi_{t-i}) = O_p(T^{-1}),$$

which shows that $C_T^{\epsilon r} = O_p(T^{-1} p^{1/2})$. Finally,

$$T^{-1} \sum_{l=1}^T g_{t-j} \left( \psi_p(L)g_t - (f_t - \bar{f}) \right) = g_{t-j} (\bar{u} - \sum_{i=1}^p \phi \phi_{t-i}) = O_p(T^{-1}),$$

by which $D_T^{\epsilon r} = O_p(T^{-1} p^{1/2})$. This concludes the proof.

\[ \square \]

Corollary 3.3.1. Let Assumptions 3.1 and 3.2 hold. Let $A_T$ and $B_T$ be defined as in (3.17). Then

1. $T^{-1} \Delta T \rightarrow \frac{1}{2} \psi(1) \sigma^2 [W(1^2 - 1)]$,
2. \( T^{-2}B_T \overset{d}{\rightarrow} \psi(1)^2 \sigma^2 \int_0^1 W_\gamma(r)^2 dr \),
where \( \gamma = \text{o}ls, \text{gls}, \text{rd} \).

**Proof of Corollary 3.3.1.** Given the expressions for \( A_T \) and \( B_T \) in (3.17), it follows immediately from Lemma 3.1 that

\[
T^{-1}A_T = T^{-1}Y_{t-1}^{(r)d} \hat{e}_p^{(r)d} - T^{-1}Y_{t-1}^{(r)d} M_p^{(r)d} \left( T^{-1}M_p^{(r)dr} M_p^{(r)d} \right)^{-1} T^{-1}M_p^{(r)dr} \hat{e}_p^{(r)d} \\
= T^{-1}Y_{t-1}^{(r)d} \hat{e}_p^{(r)d} + \hat{O}_p(p^{1/2})O_p(p^{-1/2}) = T^{-1}Y_{t-1}^{(r)d} \hat{e}_p^{(r)d} + o_p(1)
\]

\[
d \overset{d}{\rightarrow} \psi(1)^2 \sigma^2 \int_0^1 W_\gamma(r)^2 dr.
\]

This completes the proof. \( \square \)

**Proof of Lemma 3.2.** We look at OLS / GLS detrending first. Note that

\[
T \hat{\sigma}^2 = (\Delta Y^d - Y_{t-1}^d \hat{\alpha})(I - M_p^{(r)dr} M_p^{(r)d} - M_p^{(r)d} \Delta Y^d - M_p^{(r)d} Y_{t-1}^d \hat{\alpha}) \\
= \Delta Y^d (I - M_p^{(r)dr} M_p^{(r)d} - M_p^{(r)d} \Delta Y^d - M_p^{(r)d} Y_{t-1}^d \hat{\alpha}) \\
- \hat{\alpha} Y_{t-1}^d (I - M_p^{(r)dr} M_p^{(r)d} - M_p^{(r)d} \Delta Y^d - M_p^{(r)d} Y_{t-1}^d \hat{\alpha}) \\
= \hat{\epsilon}_p^{(r)} (I - M_p^{(r)dr} M_p^{(r)d} - M_p^{(r)d} \Delta Y^d - M_p^{(r)d} Y_{t-1}^d \hat{\alpha}) \\
- \hat{\alpha} Y_{t-1}^d (I - M_p^{(r)dr} M_p^{(r)d} - M_p^{(r)d} \Delta Y^d - M_p^{(r)d} Y_{t-1}^d \hat{\alpha}),
\]

which we can write as

\[
\hat{\sigma}^2 = C_T - 2D_T + E_T
\]

We first consider \( C_T \). Write

\[
C_T = T^{-1} \hat{\epsilon}_p^{(r)} \hat{\epsilon}_p^{(r)} - T^{-1} \hat{\epsilon}_p^{(r)} M_p^{(r)dr} M_p^{(r)d} \hat{\epsilon}_p^{(r)}.
\]

It follows from Corollary 3.3.1 that \( \hat{\alpha} = T^{-1}(T^{-1}A_T)(T^{-2}B_T)^{-1} = O_p(T^{-1}) \). Given the results from Lemma 3.1, we have that

\[
T^{-1} \left| \hat{\epsilon}_p^{(r)dr} M_p^{(r)dr} M_p^{(r)d} \hat{\epsilon}_p^{(r)} \right|^2 \leq T^{-1} \hat{\epsilon}_p^{(r)dr} M_p^{(r)dr} \left\| T^{-1} \hat{\epsilon}_p^{(r)dr} M_p^{(r)dr} \right\| \left\| T^{-1} M_p^{(r)d} \hat{\epsilon}_p^{(r)} \right\| \\
= o_p(p^{-1/2})O_p(1)o_p(p^{-1/2}) = o_p(p^{-1}).
\]

Hence,

\[
C_T = T^{-1} \hat{\epsilon}_p^{(r)dr} \hat{\epsilon}_p^{(r)} + o_p(1).
\]

Next we turn to \( D_T \). We can write \( D_T \) as

\[
D_T = T^{-1} \hat{\epsilon}_p^{(r)dr} \hat{\alpha} - T^{-1} \hat{\epsilon}_p^{(r)dr} M_p^{(r)d} \hat{\alpha} + \hat{\alpha} Y_{t-1}^d (I - M_p^{(r)dr} M_p^{(r)d} - M_p^{(r)d} \Delta Y^d - M_p^{(r)d} Y_{t-1}^d \hat{\alpha}).
\]
Again using Lemma 3.1 and \( \hat{\alpha} = O_p(T^{-1}) \), we have
\[
|D_T| \leq \left| T^{-1} \frac{d}{dT} y_{t-1} \right| |\hat{\alpha}| + \left| T^{-1} \frac{d}{dT} M_p \right| \left| T^{-1} (M_p^d M_p^d)^{-1} \right| \left| T^{-1} M_p^d y_{t-1} \right| |\hat{\alpha}|
\]
\[
= O_p(1) O_p(T^{-1}) + \sigma^2(p^{-1/2}) O_p(1) O_p(1) O_p(T^{-1}) = O_p(T^{-1}).
\]
Finally we look at \( E_T \):
\[
E_T = \hat{\alpha} y_{t-1} y_{t-1} - \hat{\alpha} y_{t-1} M_p^d (M_p^d M_p^d)^{-1} M_p^d y_{t-1} \hat{\alpha}.
\]
As before, we use the results from Lemma 3.1 and \( \hat{\alpha} = O_p(T^{-1}) \) to obtain
\[
|E_T| \leq T |\hat{\alpha}| \left| T^{-2} y_{t-1} y_{t-1} \right| |\hat{\alpha}| + |\hat{\alpha}| \left| T^{-1} y_{t-1} M_p \right| \left| (T^{-1} M_p^d M_p^d)^{-1} \right| \left| T^{-1} M_p^d y_{t-1} \right| |\hat{\alpha}|
\]
\[
= T O_p(T^{-1}) O_p(1) O_p(T^{-1}) + O_p(T^{-1}) O_p(p^{-1/2}) O_p(1) O_p(p^{-1/2}) O_p(T^{-1}) = O_p(T^{-1}).
\]
Therefore, we have that
\[
\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{p,t}^2 + o_p(1).
\]
For recursive detrending we can obtain in the same way that
\[
\hat{T}\hat{\sigma}^2 = \frac{\varepsilon_{p}^{rd}(1 - M_p^d (M_p^d M_p^d)^{-1} M_p^d)\varepsilon_{p}^{rd}}{T} - \varepsilon_{p}^{rd}(1 - M_p^d (M_p^d M_p^d)^{-1} M_p^d)\varepsilon_{p}^{rd} \hat{\alpha} - \hat{\alpha} \varepsilon_{p}^{rd}(1 - M_p^d (M_p^d M_p^d)^{-1} M_p^d)\varepsilon_{p}^{rd} + \hat{\alpha} \varepsilon_{p}^{rd}(1 - M_p^d (M_p^d M_p^d)^{-1} M_p^d)\varepsilon_{p}^{rd} \Delta \varepsilon_{t-1} \hat{\alpha}.
\]
and correspondingly
\[
\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{p,t}^{rd2} + o_p(1).
\]
Now
\[
\left| \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{p,t}^{rd2} \right|^{1/2} - \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{p,t}^{rd2} \right)^{1/2} \leq \left[ \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{p,t}^{rd2} - \varepsilon_{p,t}^{rd2} \right]^{1/2}.
\]
Then, for OLS / GLS detrending we have that
\[
\frac{1}{T} \sum_{t=1}^{T} (\varepsilon_{p,t} - \varepsilon_{p,t})^2 = \frac{1}{T} \sum_{t=1}^{T} \left[ (\hat{\beta} - \beta)' \Delta z_t - \sum_{j=1}^{p} \phi_j (\hat{\beta} - \beta)' \Delta z_{t-j} \right]^2
\]
\[
= (\hat{\beta} - \beta)^2 \left( 1 - \sum_{j=1}^{p} \phi_j \right)^2 = O_p(T^{-1}),
\]
while for recursive detrending we have that
\[
\frac{1}{T} \sum_{t=1}^{T} (\varepsilon_{p,t} - \varepsilon_{p,t})^2 = \frac{1}{T} \sum_{t=1}^{T} \left[ \bar{u} - \sum_{j=1}^{p} \phi_j g_{t-j} \right]^2
\]
\[
= O_p(T^{-1}).
\]
Then, by Chang and Park (2002, Proof of Lemma 3.1c), we have that
\[
\frac{1}{T} \sum_{t=1}^{T} \varepsilon^2_{p,t} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon^2_{t} + o_p(1),
\]
and by the law of large numbers \( \frac{1}{T} \sum_{t=1}^{T} \varepsilon^2_{t} \xrightarrow{P} \sigma^2 \).

**Proof of Theorem 3.1.** We have that
\[
\frac{T^{-1} A_T}{(T^{-2} B_T \sigma^2)^{1/2}} \xrightarrow{d} \sqrt{T} \psi(1) \sigma^2 (W_\gamma(1)^2 - 1) / \left( \psi(1)^2 \sigma^2 \int_0^1 W_\gamma(r)^2 dr \right)^{1/2} \xrightarrow{d} \frac{W_\gamma(1)^2 - 1}{2 \left( \int_0^1 W_\gamma(r)^2 dr \right)^{1/2}},
\]
which follows straightforwardly from Corollary 3.A.1 and Lemma 3.2. □

### 3.A.2 Proofs for Section 3.4

We start with the proofs of the two lemmas that demonstrate the equivalence of the different detrending techniques in the first step in the bootstrap.

**Proof of Lemma 3.1.** For the first part, we first make the step from ADF estimation to estimation under the null of a unit root (denote the vector of autoregressive estimators as \( \Phi_\gamma \)). Then, as shown by Chang and Park (2002) for the case without deterministic components,
\[
\Phi_\gamma = \Phi_\gamma + (M_{q}' M_{q})^{-1} M_{q}' \Delta Y = \Phi_\gamma + O_p(T^{-1}).
\]

The next step is to show that \( \Phi_\gamma = \Phi_\gamma + O_p(T^{-1} q) \). Note that
\[
\Phi_\gamma - \Phi_\gamma = (M_{q}' M_{q})^{-1} M_{q}' \Delta Y - (M_{q}' M_{q})^{-1} M_{q}' u
\]
\[
= (M_{q}' M_{q})^{-1} M_{q}' \varepsilon_{q} - (M_{q}' M_{q})^{-1} M_{q}' \varepsilon_{p}
\]
\[
= [((M_{q}' M_{q})^{-1} - (M_{q}' M_{q})^{-1}) \varepsilon_{q} - (M_{q}' M_{q})^{-1} \left[ M_{q}' \phi_{q}(L) \Delta z_q (\beta - \beta) \right]
\]
\[
+ (\beta - \beta) Y \Delta z_q \varepsilon_{q} - (\beta - \beta) Y \Delta z_q \phi_{p}(L) \Delta z_q (\beta - \beta)]
\]
\[
= A_T + B_T + C_T + D_T.
\]

Then
\[
|\tilde{\phi} - \phi| = |e_{\gamma}(\tilde{\phi}_\gamma - \phi_\gamma)| \leq |e_{\gamma}|(||A_T|| + ||B_T|| + ||C_T|| + ||D_T||).
\]

First we look at \( A_T \). Note that
\[
(T^{-1} M_{q}' M_{q})^{-1} - (T^{-1} M_{q}' M_{q})^{-1} = (T^{-1} M_{q}' M_{q})^{-1} 
\]
\[
\times (T^{-1} M_{q}' M_{q} - T^{-1} M_{q}' M_{q}(T^{-1} M_{q}' M_{q})^{-1}).
\]

Hence,
\[
\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{p,t}^2 = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{p,t}^2 + o_p(1).
\]
Furthermore,
\[ T^{-1} M'_q M_q - T^{-1} M'_q M'_q M_q = -T^{-1} M'_q \Delta z_q (\bar{\beta} - \beta) \]
\[ - T^{-1}(\bar{\beta} - \beta)' \Delta z'_q M_q + (\bar{\beta} - \beta)' \Delta z'_q \Delta z_q (\bar{\beta} - \beta). \]

Therefore
\[ ||A_T|| \leq ||T^{-1} M'_q \varepsilon_q|| ||(T^{-1} M'_q M'_q)^{-1}|| ||(T^{-1} M'_q M_q)^{-1}|| \]
\[ \times 2 \left( ||T^{-1} M'_q \Delta z_q (\bar{\beta} - \beta)|| + ||(\bar{\beta} - \beta)' \Delta z'_q \Delta z_q (\bar{\beta} - \beta)|| \right) \]
\[ = o_p(q^{-1/2}) O_p(1) O_p(T^{-1} q) + O_p(T^{-1} q) = o_p(T^{-1} q^{1/2}) \]
which follows directly from the proof of Lemma 3.1(c) and (e). It also follows directly from the proof of Lemma 3.1(c) and (e) that
\[ ||B_T|| \leq ||(T^{-1} M'_q M'_q)^{-1}|| ||T^{-1} M'_q \phi_q(L) \Delta z_q (\bar{\beta} - \beta)|| = O_p(T^{-1} q^{1/2}), \]
\[ ||C_T|| \leq ||(T^{-1} M'_q M'_q)^{-1}|| ||T^{-1} (\bar{\beta} - \beta)' \Delta z'_q \varepsilon_q|| = O_p(T^{-1} q^{1/2}), \]
\[ ||D_T|| \leq ||(T^{-1} M'_q M'_q)^{-1}|| ||T^{-1}(\bar{\beta} - \beta)' \Delta z'_q \phi_q(L) \Delta z_q (\bar{\beta} - \beta)|| = O_p(T^{-1} q^{1/2}). \]

Therefore we may conclude that \( \tilde{\varphi}_j = \tilde{\varphi}_{0,j} + O_p(T^{-1} q^{1/2}) \) and consequently that \( \varphi_j = \tilde{\varphi}_j + O_p(T^{-1} q^{1/2}) \) uniformly in \( j, 1 \leq j \leq q \).

For the second part we have that
\[ \tilde{e}_{q,t} - \bar{e}_{q,t} = -\hat{\alpha} y_{t-1} + \sum_{j=1}^{q} (\hat{\varphi}_j - \tilde{\varphi}_j) u_{t-j} - \left( (\bar{\beta} - \beta)' \Delta z_t \right) \sum_{j=1}^{q} \tilde{\varphi}_j (\bar{\beta} - \beta)' \Delta z_{t-j}, \]
from which we can conclude that
\[
\max_{1 \leq t \leq T} |\tilde{e}_{q,t} - \bar{e}_{q,t}| \leq |\hat{\alpha}| \max_{1 \leq t \leq T} |y_{t-1}| + \max_{1 \leq t \leq T} \left| u_{t} \right| \sum_{j=1}^{q} |\hat{\varphi}_j - \tilde{\varphi}_j| + |\bar{\beta}_2 - \beta_2| (1 - \sum_{j=1}^{q} |\tilde{\varphi}_j|) \\
= O_p(T^{-1/2}) + O_p(T^{-1} q^{3/2}) + O_p(T^{-1} q^{1/2}).
\]

This completes the proof. \( \square \)

**Proof of Lemma 3.2.** As in the proof of Lemma 3.1 we have that
\[ \Phi_q = \Phi_q + (M'_q M'_q)^{-1} W_{r,d} \Delta Y^{-d} \alpha = \Phi_q + O_p(T^{-1}), \]
and furthermore that
\[ \tilde{\varphi}_q - \bar{\varphi}_q = [(M'_q M'_q)^{-1} - (M'_q M_q)^{-1}] M'_q \varepsilon_q \]
\[ - (M'_q M'_q)^{-1} \left[ M'_q \phi_q(L) g_q + g'_q \varepsilon_q - g'_q \phi_q(L) g_q \right] \\
= A_T + B_T + C_T + D_T. \]

For \( A_T \) we have that
\[ ||A_T|| \leq ||T^{-1} M'_q \varepsilon_q|| ||(T^{-1} M'_q M'_q)^{-1}|| \]
\[ \times 2 \left( ||T^{-1} M'_q \varepsilon_q|| + ||g'_q \varepsilon_q|| \right) \]
\[ = o_p(q^{-1/2}) O_p(1) O_p(T^{-1} q) + O_p(T^{-1} q) = o_p(T^{-1} q^{1/2}) \]
which again follows directly from the proof of Lemma 3.1(c) and (e). Similarly we have that
\[ \|B_T\| \leq \| (T^{-1} M_q^{rd} M_q^{rd})^{-1} L_0 \| = O_p(T^{-1} q^{1/2}), \]
\[ \|C_T\| \leq \| (T^{-1} M_q^{rd} W_q^{rd})^{-1} L_0 \| = O_p(T^{-1} q^{1/2}), \]
\[ \|D_T\| \leq \| (T^{-1} M_q^{rd} M_q^{rd})^{-1} \| = O_p(T^{-1} q^{1/2}). \]

Therefore we may conclude that \( \tilde{\phi}_j = \hat{\phi}_j + O_p(T^{-1} q^{1/2}) \) and consequently that \( \hat{\phi}_j = \hat{\phi}_j + O_p(T^{-1} q^{1/2}) \) uniformly in \( j, 1 \leq j \leq q. \)

For the second part we have that
\[ \hat{\varepsilon}_{q,t}^d - \tilde{\varepsilon}_{q,t} = -\hat{\alpha}_y t - \sum_{j=1}^q (\hat{\phi}_j - \tilde{\phi}_j) u_{t-j} - \left( \bar{u} - \sum_{j=1}^q \tilde{\phi}_j g_{t-j} \right), \]

from which we can conclude that
\[ |\varepsilon_{q,t} - \tilde{\varepsilon}_{q,t}| \leq |\hat{\alpha}| \max_{1 \leq t \leq T} |y_{t-1}| + \max_{1 \leq t \leq T} |u_t| \sum_{j=1}^q |\hat{\phi}_j - \tilde{\phi}_j| + |\bar{u}| + \max_{1 \leq j \leq q} |g_{t-1}| \sum_{j=1}^q |\tilde{\phi}_j| \]
\[ = O_p(T^{-1/2}) + O_p(T^{-1} q^{1/2}) + O_p(T^{-1/2}) + \max_{1 \leq j \leq q} |g_{t-1}| \sum_{j=1}^q |\tilde{\phi}_j|. \]

This completes the proof. \( \square \)

The first step towards an invariance principle for \( u_t^* \) is to show that higher than second order moments exist for \( \varepsilon_t^* \).

**Lemma 3.A.6.** Under Assumptions 3.1 and 3.3 we have for any \( 2 < a \leq 4 \)
\[ \mathbb{E}^* |\varepsilon_t|^a = O_p(1). \]

**Proof of Lemma 3.A.6.** We have that
\[ \mathbb{E}^* |\varepsilon_t|^a = T^{-1} \sum_{t=1}^T \|\tilde{\varepsilon}_{q,t} - \tilde{\varepsilon}_{q,t}^d \|_{\mathbb{L}_0} \leq 2^{a-1} T^{-1} \sum_{t=1}^T \|\tilde{\varepsilon}_{q,t} - \tilde{\varepsilon}_{q,t}^d \|_{\mathbb{L}_0} \]
\[ - T^{-1} \sum_{t=1}^T \|\tilde{\varepsilon}_{q,t} - \tilde{\varepsilon}_{q,t}|^a \|_{\mathbb{L}_0} + 2^{a-1} T^{-1} \sum_{t=1}^T \|\tilde{\varepsilon}_{q,t} - \tilde{\varepsilon}_{q,t}^d \|^a, \]

where the first part is \( o_p(1) \) by Lemma 3.1 and 3.2. To see the latter result, note that
\[ T^{-1} \sum_{t=1}^T \|\tilde{\varepsilon}_{q,t} - \tilde{\varepsilon}_{q,t}^d \|^a \leq \max_{1 \leq j \leq q} T^{-1} \sum_{t=1}^T |g_{t-1}|^a + o_p(1) = o_p(1). \]

The second part is \( O_p(1) \) by Park (2002, Lemma 3.2).

**Lemma 3.A.7.** Let Assumption 3.1 and 3.3 hold. Then \( \sigma^* \overset{p}{\rightarrow} \sigma. \)

**Proof of Lemma 3.A.7.** Follows directly from Lemma 3.2. \( \square \)
Lemma 3.A.8. Let Assumption 3.1 and 3.3 hold. Then
\[ T^{-1/2} \sum_{t=1}^{[Tr]} \varepsilon_t^* \overset{d}{\to} \sigma W(r) \text{ in probability.} \]


Lemma 3.A.9. Let Assumption 3.1 and 3.3 hold. Then
\[ T^{-1/2} \sum_{t=1}^{[Tr]} u_t^* \overset{d}{\to} \sigma \psi(1) W(r) \text{ in probability.} \]

Proof of Lemma 3.A.9. As in Park (2002, p. 478) we need to show that
\[ \hat{\phi}(1) \overset{p}{\to} \phi(1) \quad (3.42) \]
\[ P^* \left\{ \max_{1 \leq t \leq T} |T^{-1/2} \bar{u}_t^*| > \epsilon \right\} = o_p(1) \quad (3.43) \]
where \( \bar{u}_t^* = \hat{\phi}(1)^{-1} \sum_{i=1}^q (\sum_{j=1}^q \hat{\phi}_j) u_{t-i+1}^* \).

We first show (3.42). Park (2002, Lemma 3.1) shows that \( |\tilde{\phi}(1) - \phi(1)| = o_p(1) \); therefore we have that
\[ |\hat{\phi}(1) - \phi(1)| \leq |\hat{\phi}(1) - \tilde{\phi}(1)| + |\tilde{\phi}(1) - \phi(1)| = O_p(T^{-1/2}) + o_p(1), \]
where the first part follows from Lemma 3.1 and 3.2. Hence \( \tilde{\phi}(1) = \phi(1) + o_p(1) \). This proves (3.42).

To prove (3.43) we have, as in Park (2002, Proof of Theorem 3.3),
\[ P^* \left\{ \max_{1 \leq t \leq T} |T^{-1/2} \bar{u}_t^*| > \epsilon \right\} \leq T P^* \left\{ |T^{-1/2} \bar{u}_t^*| > \epsilon \right\} \leq (1/\epsilon^a) T^{1-a/2} E^* |\bar{u}_t^*|^a. \]

Hence, we have to show that
\[ T^{1-a/2} E^* |\bar{u}_t^*|^a = o_p(1). \]

As in Palm, Smeekes, and Urbain (2008c) this amounts to showing that
\[ \sum_{j=1}^q j^{1/2} |\hat{\phi}_j| = O_p(1). \]

We can write
\[ \sum_{j=1}^q j^{1/2} |\hat{\phi}_j| \leq \sum_{j=1}^q j^{1/2} |\phi_j - \hat{\phi}_j| + \sum_{j=1}^q j^{1/2} |\hat{\phi}_j| = O_p(T^{-1} q^2) + O_p(1) = O_p(1), \]
where the first part follows from Lemma 3.1 and 3.2 and the second part follows from Palm et al. (2008c). This concludes the proof of this theorem. \( \square \)
Lemma 3.A.10. Let \( \hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)' \) be defined as

\[
\hat{\beta}^* = \left( \sum_{t=1}^{T} z_t z_t' \right)^{-1} \left( \sum_{t=1}^{T} z_t y_t' \right).
\]

Let Assumption 3.1 hold. Then

\[
\left( T^{-1/2}(\hat{\beta}_1 - \beta_1), T^{-1/2}(\hat{\beta}_2 - \beta_2) \right) \xrightarrow{d^*} \left( 4\psi(1)\sigma \int_0^1 W(r)dr - 6\psi(1)\sigma \int_0^1 rW(r)dr, -6\psi(1)\sigma \int_0^1 W(r)dr + 12\psi(1)\sigma \int_0^1 rW(r)dr \right)
\]
in probability.

Proof of Lemma 3.A.3. As in the proof of Lemma 3.A.3, note that

\[
\text{Let Assumption 3.1 hold. Then}
\]

\[
\begin{align*}
\hat{\beta}^* - \beta^* &= \left( \sum_{t=1}^{T} z_t z_t' \right)^{-1} \left( \sum_{t=1}^{T} z_t x_t' \right).
\end{align*}
\]

From the proof of Lemma 3.A.3 we have that

\[
\left( \sum_{t=1}^{T} z_t z_t' \right)^{-1} = \begin{pmatrix}
2(2T+1)T^{-1}(T-1)^{-1} & -6T^{-1}(T-1)^{-1} \\
-6T^{-1}(T-1)^{-1} & 12T^{-1}(T+1)^{-1}(T-1)^{-1}
\end{pmatrix}.
\]

Applying (3.35) to the bootstrap sample we obtain

\[
\left( T^{-1/2}(\hat{\beta}_1^* - \beta_1^*), T^{-1/2}(\hat{\beta}_2^* - \beta_2^*) \right) \xrightarrow{d^*} \left( 4\psi(1)\sigma \int_0^1 W(r)dr - 6\psi(1)\sigma \int_0^1 rW(r)dr, -6\psi(1)\sigma \int_0^1 W(r)dr + 12\psi(1)\sigma \int_0^1 rW(r)dr \right)
\]
in probability, which follows from Lemma 3.A.9 and the continuous mapping theorem. 

Lemma 3.A.11. Let \( \hat{\beta} = (\hat{\beta}_1^*, \hat{\beta}_2^*)' \) be defined as

\[
\hat{\beta}^* = \left( \sum_{t=1}^{T} z_{t,t} z_{t,t}' \right)^{-1} \left( \sum_{t=1}^{T} z_{t,t} y_{t,t}' \right).
\]

where \( y_{t,1} = y_t' \) and \( y_{t,t} = \Delta y_t - \bar{c} T^{-1} y_{t-1}^* \) for \( t = 2, \ldots, T \). Let Assumption 3.1 and 3.3 hold. Then

\[
\left( T^{-1/2}(\hat{\beta}_1^* - \beta_1^*), T^{-1/2}(\hat{\beta}_2^* - \beta_2^*) \right) \xrightarrow{d^*} \left( 0, \sigma \psi(1) G(\bar{c}) \right)
\]
in probability,

(3.45)

where

\[
G(\bar{c}) = (1 - \bar{c} + \frac{1}{3} \bar{c}^3)^{-1} \left[ (1 - \bar{c}) W(1) + \bar{c}^2 \int_0^1 sW(s)ds \right].
\]

Proof of Lemma 3.A.11. The proof follows along the same lines as in the proof of Lemma 3.A.4 and is essentially the same as the first part of the proof of Lemma 4 of Swensen (2003b). We have

\[
\hat{\beta}^* - \beta^* = \left( \sum_{t=1}^{T} z_{t,t} z_{t,t}' \right)^{-1} \left( \sum_{t=1}^{T} z_{t,t} x_{t,t}' \right),
\]

(3.46)
where \( x_{*,1}^t = x_1^t \) and \( x_{*,t}^* = \Delta x_1^t = \bar{c}T^{-1}x_{t-1}^* \) for \( t = 2, \ldots, T \). Then

\[
\begin{pmatrix}
T^{-1/2} \beta_*^t \\
 T_{1/2}^2 \rho_*^t
\end{pmatrix}
= \begin{pmatrix}
 T^{-1/2} \sum_{t=1}^T \bar{z}_{*,1t}^2 & T^{-1/2} \sum_{t=1}^T \bar{z}_{*,1t} \bar{z}_{*,2t} \\
 T^{-1/2} \sum_{t=1}^T \bar{z}_{*,2t} \bar{z}_{*,1t} & T^{-1/2} \sum_{t=1}^T \bar{z}_{*,2t}^2
\end{pmatrix}^{-1}
\times \begin{pmatrix}
 T^{-1/2} \sum_{t=1}^T \bar{z}_{*,1t} x_{*,t}^* \\
 T^{-1/2} \sum_{t=1}^T \bar{z}_{*,2t} x_{*,t}^*
\end{pmatrix}.
\]

As in the proof of Lemma 3.A.4 we have that

\[
\begin{pmatrix}
 T^{-1/2} \sum_{t=1}^T \bar{z}_{*,1t}^2 & T^{-1/2} \sum_{t=1}^T \bar{z}_{*,1t} \bar{z}_{*,2t} \\
 T^{-1/2} \sum_{t=1}^T \bar{z}_{*,2t} \bar{z}_{*,1t} & T^{-1/2} \sum_{t=1}^T \bar{z}_{*,2t}^2
\end{pmatrix}
= \begin{pmatrix}
 T^{-1/2}(1-\bar{c}) + \frac{1}{\bar{c}} T^{-3/2} (T-1) & T^{-1/2}(1-\bar{c}) + \frac{1}{\bar{c}} T^{-3/2} (T-1) \\
 T^{-1/2}(1-\bar{c}) + \frac{1}{\bar{c}} T^{-3/2} (T-1) & T^{-1/2}(1-\bar{c}) + \frac{1}{\bar{c}} T^{-3/2} (T-1) (2T-1)
\end{pmatrix}
\]

\[(3.47)\]

and

\[
\begin{pmatrix}
 T^{-1/2} \sum_{t=1}^T \bar{z}_{*,1t} x_{*,t}^* \\
 T^{-1/2} \sum_{t=1}^T \bar{z}_{*,2t} x_{*,t}^*
\end{pmatrix}
= \begin{pmatrix}
 T^{-1/2} x_1^* - \bar{c} T^{-3/2} (x_T^* - x_1^*) + \frac{1}{\bar{c}} T^{-3/2} \sum_{t=2}^T x_t^* \\
 T^{-1/2} (1-\bar{c}) x_T^* + \frac{1}{\bar{c}} T^{-3/2} x_T^* + \frac{1}{\bar{c}} T^{-3/2} \sum_{t=2}^T x_{t-1}^*
\end{pmatrix}
\]

\[(3.48)\]

where the second line follows from Lemma 3.A.9 and the continuous mapping theorem. The result now follows by combining (3.47) and (3.48).

**Proof of Lemma 3.3.** For OLS / GLS detrending the result follows trivially from Lemma 3.A.11 and 3.A.9 using the proof of Lemma 3.A.5.

For recursive detrending, note that

\[
T^{-1/2} y_{[T]}^{*,r} = T^{-1/2} x_T^* \rightarrow -2T^{-1/2} [\{T_{r}\}]^{-1} \sum_{s=1}^{[T_{r}]} x_s^* - T^{-3/2} \sum_{t=1}^{T} (x_t^* - 2t^{-1} \sum_{s=1}^{t} x_s^*).
\]

The result now follows straightforwardly from Lemma 3.A.9 and the continuous mapping theorem as in the proof of Lemma 3.A.5.

**Proof of Lemma 3.4.** Part (a) follows from Lemma 3.3 using the continuous mapping theorem.

For part (b), we write

\[
T^{-1} \sum_{t=1}^{T} y_{t-1}^{*,r} z_{*,s}^{*,r} = T^{-1} \sum_{t=1}^{T} y_{t-1}^{*,r} \varepsilon_t^{*,r} \varepsilon_t^{*,r} + T^{-1} \sum_{t=1}^{T} y_{t-1}^{*,r} (\varepsilon_{p,t}^{*,r} \varepsilon_t^{*,r} - \varepsilon_t^{*,r} \varepsilon_t^{*,r}).
\]

Note that

\[
\varepsilon_{p,t}^{*,r} \varepsilon_t^{*,r} - \varepsilon_t^{*,r} \varepsilon_t^{*,r} = \sum_{j=p+1}^{q} \phi_j y_{t-j}^* + \sum_{j=p+1}^{q} \phi_j (\beta^* - \beta^*)^\prime \Delta z_{t-j}.
\]
3 Detrending Bootstrap Unit Root Tests

By Assumption 3.4 there is some $\tilde{T}$ such that for all $T > \tilde{T}$ we have that $e_{p,t}^{r}(d) - e_{t}^{r}(d) = 0$.

We can now proceed as in the proof of Lemma 3.1(b). For the specific case of GLS detrending, this amounts to showing that

$$T^{-1} \sum_{t=1}^{T} u_{t-1} e_{t}^{*} = T^{-1} \sum_{t=1}^{T} x_{t-1} \epsilon_{t}^{*} - T^{-1} \sum_{t=1}^{T} x_{t-1} \phi(L)(\hat{\beta}^{*} - \beta)^{'} \Delta z_{t}$$

$$- T^{-1} \sum_{t=1}^{T} (\hat{\beta}^{*} - \beta)^{'} z_{t} \epsilon_{t}^{*} + T^{-1} \sum_{t=1}^{T} (\hat{\beta}^{*} - \beta)^{'} \phi(L)(\hat{\beta}^{*} - \beta)^{'} \Delta z_{t}$$

$$= E_{T}^{b} - F_{T}^{b} - G_{T}^{b} + H_{T}^{b},$$

and consequently

$$E_{T}^{b} = T^{-1} \sum_{t=1}^{T} x_{t-1} \epsilon_{t}^{*} - \frac{1}{2} \psi(1) \sigma^{2} (W(1)^{2} - 1),$$

$$F_{T}^{b} = T^{-1} (\hat{\beta}^{*} - \beta) \phi(1) \sum_{t=1}^{T} x_{t-1} \epsilon_{t}^{*} + \frac{1}{2} \psi(1) \sigma^{2} G(\bar{e}) \int_{0}^{1} W(r) dr,$n

$$G_{T}^{b} = T^{-1} (\hat{\beta}^{*} - \beta) \sum_{t=1}^{T} \epsilon_{t}^{*} + T^{-1} (\hat{\beta}^{*} - \beta) \sum_{t=1}^{T} t \epsilon_{t}^{*} + \psi_{1} \sigma^{2} G(\bar{e}) \int_{0}^{1} r dW(r),$$

$$H_{T}^{b} = T^{-1} \sum_{t=1}^{T} (\hat{\beta}^{*} - \beta) \phi(1) (\hat{\beta}^{*} - \beta) + T^{-1} \sum_{t=1}^{T} \phi(1) (\hat{\beta}^{*} - \beta)^{2} t \epsilon_{t}^{*} - \frac{1}{2} \psi(1) \sigma^{2} G(\bar{e})^{2},$$

which follow by Chang and Park (2003, Lemma 2a), Lemma 3.1a, equation (3.42) and the continuous mapping theorem. The other methods of detrending follow similarly.

We continue with (c). Let $\Omega_{pp}^{*} = (\Gamma_{t}^{*})_{t=1}^{T}$ where $\Gamma_{t}^{*} = E^{*}(u_{t}^{*} u_{t-1}^{*})$. Then, as in the proof of Lemma 3.1(c), we can write for OLS / GLS detrending

$$\left| T^{-1} \sum_{t=1}^{T} u_{t-1} w_{p,t}^{*} - \Omega_{pp}^{*} \right| \leq \left| T^{-1} \sum_{t=1}^{T} u_{t-1} w_{p,t}^{*} - \Omega_{pp}^{*} \right|$$

$$+ 2 \left| T^{-1} \sum_{t=1}^{T} u_{t-1} (\hat{\beta}^{*} - \beta)^{'} \Delta z_{t} \right|$$

$$+ \left| T^{-1} \sum_{t=1}^{T} \Delta z_{t} (\hat{\beta}^{*} - \beta)^{'} \Delta z_{t} \right|$$

$$= A_{T}^{*} + 2B_{T}^{*} + C_{T}^{*}.$$

By Chang and Park (2003, Proof of Lemma 3a) and Lemma 3.1 $A_{T}^{*} = O_{p}^{*}(T^{-1/2})$. Next consider

$$T^{-1} \sum_{t=1}^{T} u_{t-1}(\hat{\beta}^{*} - \beta)^{'} \Delta z_{t-1} = (\hat{\beta}^{*} - \beta) T^{-1} \sum_{t=1}^{T} u_{t-1} = O_{p}^{*}(T^{-1}),$$

as $\sum_{t=1}^{T} u_{t-1} = O_{p}(T^{1/2})$ by Lemma 3.1a. Hence $B_{T}^{*} = O_{p}^{*}(T^{-1})$. Finally consider

$$T^{-1} \sum_{t=1}^{T} (\hat{\beta}^{*} - \beta)^{'} \Delta z_{t-1} (\hat{\beta}^{*} - \beta)^{'} \Delta z_{t-1} = (\hat{\beta}^{*} - \beta)^{2}.$$
3.A Appendix: Proofs

and therefore $C_T^{**} = O_p(T^{-1/2})$.

For recursive detrending we write

$$
\left| T^{-1} \sum_{t=1}^{T} u^r_{p,t} w^r_{p,t} - \Omega_{pp}^* \right| \leq \left| T^{-1} \sum_{t=1}^{T} u^r_{p,t} w^r_{p,t} - \Omega_{pp}^* \right| + 2 \left| T^{-1} \sum_{t=1}^{T} u^r_{p,t} g^r_{p,t} \right|
$$

$$= A_T^{**} + 2B_T^{**} + C_T^{**},
$$

where $g^r_{p,t} = (g^r_{p,t-1}, \ldots, g^r_{p,1})^\prime$.

Again by Chang and Park (2003, Proof of Lemma 3a) and Lemma 3.2 $A_T^{**} = O_p(T^{-1/2})$. Next consider

$$T^{-1} \sum_{t=1}^{T} u^r_{t-j} g^r_{t-j} = O_p(T^{-1})$$

$$T^{-1} \sum_{t=1}^{T} g^r_{t-j} g^r_{t-j} = O_p(T^{-1}).$$

As this holds uniformly in $i, j = 1, \ldots, p$, we can conclude that $B_T^{**} = O_p(T^{-1/2})$ and $C_T^{**} = O_p(T^{-1/2})$.

Hence $\left| T^{-1} \sum_{t=1}^{T} u^r_{p,t} w^r_{p,t} - \Omega_{pp}^* \right| = O_p(T^{-1/2})$ and we can conclude the proof as in Chang and Park (2003, Proof of Lemma 3a).

Next we look at (d). We start again with OLS / GLS detrending. We have

$$\left| T^{-1} \sum_{t=1}^{T} y^*_{t-1} w^*_{p,t} \right| \leq \left| T^{-1} \sum_{t=1}^{T} x^*_{t-1} w^*_{p,t} \right| + \left| T^{-1} \sum_{t=1}^{T} x^*_{t-1} \Delta z_{p,t}(\hat{\beta}^* - \beta^*) \right|
$$

$$+ \left| T^{-1} \sum_{t=1}^{T} (\hat{\beta}^* - \beta^*)' z_{t} w^*_{p,t} \right|
$$

$$+ \left| T^{-1} \sum_{t=1}^{T} (\hat{\beta}^* - \beta^*)' z_{t} \Delta z_{p,t}(\hat{\beta}^* - \beta^*) \right|
$$

$$= A_T^{**} + B_T^{**} + C_T^{**} + D_T^{**}.$$

We can easily show that $A_T^{**} = O_p(p^{1/2})$ along the same lines as Chang and Park (2003, Proof of Lemma 3b) using Lemma 3.1. Furthermore,

$$T^{-1} (\hat{\beta}_1^* - \beta_1^*) \sum_{t=1}^{T} x^*_{t-1} = O_p(1)$$

$$T^{-1} \sum_{t=1}^{T} (\hat{\beta}^* - \beta^*)' z_{t} (\hat{\beta}^* - \beta^*)' \Delta z_{t-j} = (\hat{\beta}_1^* - \beta_1^*) (\hat{\beta}_2^* - \beta_2^*) + (\hat{\beta}_2^* - \beta_2^*)^2 \frac{1}{2}(T + 1)
$$

$$= O_p(1)$$

$$T^{-1} \sum_{t=1}^{T} (\hat{\beta}^* - \beta^*)' z_{t} u^*_{t-j} = T^{-1} (\hat{\beta}_1^* - \beta_1^*) \sum_{t=1}^{T} u^*_{t-j} + T^{-1} (\hat{\beta}_2^* - \beta_2^*) \sum_{t=1}^{T} \sum_{t=1}^{T} t u^*_{t-j} = O_p(1).$$
from which we can conclude that $B_T^{dr} = O_p(p^{1/2})$, $C_T^{dr} = O_p(p^{1/2})$ and $D_T^{dr} = O_p(p^{1/2})$.

For recursive detrending we have

\[
T^{-1} \sum_{t=1}^T \Delta z_{p,t} \leq T^{-1} \sum_{t=1}^T \Delta z_{p,t}^{\ast} + T^{-1} \sum_{t=1}^T \Delta z_{p,t} (\hat{\beta}^{\ast} - \beta^{\ast})^T \varepsilon_{p,t}
\]

Finally, as

\[
T^{-1} \sum_{t=1}^T \left( x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s \right) \leq T^{-1} \sum_{t=1}^T u_{x,t} + T^{-1} \sum_{t=1}^T g_{x,t} - O_p(1)
\]

we have that $E_T^{dr} = O_p(p^{1/2})$ and $F_T^{dr} = O_p(p^{1/2})$. This concludes the proof for part (d).

Finally we consider part (e). For OLS / GLS detrending we can write

\[
T^{-1} \sum_{t=1}^T w_{p,t}^{\ast} \leq T^{-1} \sum_{t=1}^T w_{p,t}^{\ast} + T^{-1} \sum_{t=1}^T \Delta \hat{\rho}_p(L) (\hat{\beta}^{\ast} - \beta^{\ast})^T \varepsilon_{p,t}
\]

Finally, as

\[
T^{-1} \sum_{t=1}^T \left( x_{t-1} - 2(t-1)^{-1} \sum_{s=1}^{t-1} x_s \right) \leq T^{-1} \sum_{t=1}^T u_{x,t} + T^{-1} \sum_{t=1}^T g_{x,t} - O_p(1)
\]

we have that $E_T^{dr} = O_p(p^{1/2})$ and $F_T^{dr} = O_p(p^{1/2})$. This concludes the proof for part (d).

Finally we consider part (e). For OLS / GLS detrending we can write
Now by Assumption 3.4 we have that for large \( T \) we may write

\[
A_T^{**} = \left| T^{-1} \sum_{t=1}^{T} u_{p,t}^* \varepsilon_t^* \right|.
\]

Then for any \( 1 \leq j \leq p \)

\[
E^*( \sum_{t=1}^{T} u_{t-j}^* \varepsilon_t^* )^2 = T \sigma^* T_0^* t^2,
\]

by which it follows from Chang and Park (2003, Proof of Lemma 3c) that \( A_T^{**} = O_p(T^{-1/2} p^{1/2}) \). Now

\[
T^{-1} \sum_{t=1}^{T} (\hat{\beta}^* - \beta^*)' \Delta z_{t-j} \varepsilon_{t-j}^* = (\hat{\beta}_2^* - \beta_2^*) T^{-1} \sum_{t=1}^{T} \varepsilon_t^* = O_p(T^{-1}),
\]

\[
T^{-1} \sum_{t=1}^{T} u_{t-j}^* \hat{\phi}(L)(\hat{\beta}^* - \beta^*)' \Delta z_t = (\hat{\beta}_2^* - \beta_2^*) T^{-1} \sum_{t=1}^{T} u_t^* \sum_{j=1}^{p} \hat{\phi}_j = O_p(T^{-1}),
\]

\[
T^{-1} \sum_{t=1}^{T} (\hat{\beta}^* - \beta^*)' \Delta z_{t-j} \hat{\phi}(L)(\hat{\beta}^* - \beta^*)' \Delta z_t = (\hat{\beta}_2^* - \beta_2^*) T^{-1} \sum_{j=1}^{p} \hat{\phi}_j = O_p(T^{-1}),
\]

by which \( B_T^{**} = O_p(T^{-1/2} p^{1/2}) \), \( C_T^{**} = O_p(T^{-1} p^{1/2}) \) and \( D_T^{**} = O_p(T^{-1} p^{1/2}) \).

For recursive detrending we can write

\[
\left| \sum_{t=1}^{T} \sum_{l=1}^{T} u_{p,t}^* \varepsilon_{p,t}^* \right| \leq \left| \sum_{t=1}^{T} \sum_{l=1}^{T} g_{p,t}^* \varepsilon_{p,t}^* \right| + \left| \sum_{l=1}^{T} \sum_{t=1}^{T} g_{p,t}^* \left( \hat{\phi}_p(L) g_{t}^* - (f_t^* - f^*) \right) \right| + \left| \sum_{l=1}^{T} \sum_{t=1}^{T} g_{p,t}^* \left( \hat{\phi}_p(L) g_{t}^* - (f_t^* - f^*) \right) \right| = A_T^{**} + B_T^{**} + C_T^{**} + D_T^{**}.
\]

As for OLS / GLS detrending, we have that \( A_T^{**} = O_p(T^{-1/2} p^{1/2}) \) by Chang and Park (2003, Proof of Lemma 3c). Now consider

\[
T^{-1} \sum_{t=1}^{T} g_{t-j}^* \varepsilon_t^* = O_p(T^{-1}),
\]

\[
T^{-1} \sum_{t=1}^{T} u_{t-j}^* \left( \bar{u}^* - \sum_{j=1}^{q} \hat{\phi}_j \bar{y}_{t-j} \right) = O_p(T^{-1}),
\]

\[
T^{-1} \sum_{t=1}^{T} g_{t-j}^* \left( \bar{u}^* - \sum_{j=1}^{q} \hat{\phi}_j \bar{y}_{t-j} \right) = O_p(T^{-1}),
\]

by which we may conclude that \( B_T^{**} = O_p(T^{-1} p^{1/2}) \), \( C_T^{**} = O_p(T^{-1} p^{1/2}) \) and \( D_T^{**} = O_p(T^{-1} p^{1/2}) \). This concludes the proof. \( \square \)
Corollary 3.A.2. Let Assumptions 3.1, 3.3, 3.2 and 3.4 hold. Let $A_T^*$ and $B_T^*$ be defined as in (3.32). Then

1. $T^{-1} A_T^* \xrightarrow{d^*} \psi(1) \sigma^2 [W_\gamma(1)^2 - 1]$ in probability,
2. $T^{-2} B_T^* \xrightarrow{d^*} \psi(1)^2 \sigma^4 \int_0^1 W_\gamma(r)^2 dr$ in probability,

where $\gamma = \text{ols, gls, rd}$.

Proof of Corollary 3.A.2. The results follow immediately from Lemma 3.4, given the expressions for $A_T^*$ and $B_T^*$.

Proof of Lemma 3.5. As in the proof of Lemma 3.2, for OLS / GLS detrending we have that

$$T \hat{\sigma}^2 = \varepsilon_p^{*dr}(I - M_p^{*dr}(M_p^{*dr} M_p^{*dr})^{-1} M_p^{*dr}) \varepsilon_p^{*dr}$$

$$- \varepsilon_p^{*dr}(I - M_p^{*dr}(M_p^{*dr} M_p^{*dr})^{-1} M_p^{*dr}) Y_{-1} \hat{\alpha}^*$$

$$- \hat{\alpha}^* Y_{-1} (I - M_p^{*dr}(M_p^{*dr} M_p^{*dr})^{-1} M_p^{*dr}) \varepsilon_p^{*dr}$$

$$+ \hat{\alpha}^* Y_{-1}^* (I - M_p^{*dr}(M_p^{*dr} M_p^{*dr})^{-1} M_p^{*dr}) Y_{-1} \hat{\alpha}^*,$$

which we can write as

$$\hat{\sigma}^2 = C_T^* - 2D_T^* + E_T^*.$$

Using Lemma 3.4 and Corollary 3.A.2 we can show in the same way as in the proof of Lemma 3.2 that

$$C_T^* = T^{-1} \varepsilon_p^{*dr} \varepsilon_p^{*dr} + o_p^*(1),$$

$$D_T^* = o_p^*(1),$$

$$E_T^* = o_p^*(1).$$

Therefore, we have that

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{p,t}^{*dr} + o_p^*(1),$$

which, by Assumption 3.4, we can write as

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{*rd} + o_p^*(1).$$

For recursive detrending we obtain the same expression, i.e.

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{*rd} + o_p^*(1).$$

Now

$$\left| \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{(r)d} \right)^{1/2} - \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{*d} \right)^{1/2} \right| \leq \left( \frac{1}{T} \sum_{t=1}^{T} \left( \varepsilon_t^{(r)d} - \varepsilon_t^{*d} \right)^2 \right)^{1/2}.$$
Then for OLS / GLS detrending

\[
\frac{1}{T} \sum_{t=1}^{T} (\varepsilon_t^d - \hat{\varepsilon}_t^d)^2 = \frac{1}{T} \sum_{t=1}^{T} \left[ (\hat{\beta}_t^* - \beta_0^*)' \Delta \tilde{z}_t - \sum_{j=1}^{q} \hat{\phi}_j (\hat{\beta}_t^* - \beta_0^*)' \Delta \tilde{z}_{t-j} \right]^2 \\
= (\hat{\beta}_2^* - \beta_0^*)^2 \left( 1 - \sum_{j=1}^{p} \hat{\phi}_j \right) = O_p(T^{-1}),
\]

while for recursive detrending

\[
\frac{1}{T} \sum_{t=1}^{T} (\varepsilon_t^r - \hat{\varepsilon}_t^r)^2 = \frac{1}{T} \sum_{t=1}^{T} \left[ \bar{u}_t^* - \sum_{j=1}^{p} \hat{\phi}_j g_{t-j}^* \right]^2 = O_p(T^{-1}).
\]

Hence,

\[
\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{(r)d2} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 + o_p(1).
\]

Next we show that

\[
\left| T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 - \sigma^2 \right| \leq \left| T^{-1} \sum_{t=1}^{T} \varepsilon_t^{*2} - \sigma^{*2} \right| + |\sigma^{*2} - \sigma^2| = o_p(1).
\]

To show that \( T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 - \sigma^2 = o_p(1) \), note that we have

\[
P^* \left( T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 - \sigma^2 > \epsilon \right) \leq \epsilon^{-2} \left( T^{-1} \sum_{t=1}^{T} \varepsilon_t^{*2} - \sigma^{*2} \right) \leq \epsilon^{-2} T^{-2} \sum_{t=1}^{T} \left( \hat{\varepsilon}_t^4 - \left( \hat{\varepsilon}_t^* \right)^2 \right) = O_p(T^{-1}).
\]

It follows from Lemma 3.A.7 that \( \sigma^{*2} \xrightarrow{p} \sigma^2 \). This completes the proof.

**Proof of Theorem 3.2.** The result follows straightforwardly from Corollary 3.A.2 and Lemma 3.5.
Chapter 4

A Sieve Bootstrap Test for Cointegration in a Conditional Error Correction Model

In this chapter we propose a bootstrap version of the Wald test for cointegration in a single-equation conditional error correction model. The multivariate sieve bootstrap is used to deal with dependence in the series. We show that the introduced bootstrap test is asymptotically valid.

We also analyze the small sample properties of our test by simulation and compare it with the asymptotic test and several alternative bootstrap tests. The bootstrap test offers significant improvements in terms of size properties over the asymptotic test, while having similar power properties.

The sensitivity of the bootstrap test to the allowance for deterministic components is also investigated. Simulation results show that the tests with sufficient deterministic components included are insensitive to the true value of the trends in the model, and retain correct size.1

4.1 Introduction

In this chapter we present a bootstrap version of the single-equation error correction model (ECM) Wald test for cointegration originally proposed by Boswijk (1994).

Broadly speaking, tests proposed in the literature to test for the absence of cointegration can be classified in two groups. Tests that allow for more than one cointegrating vector under the alternative using for example a VAR framework,

1This chapter is based on the paper Palm et al. (2008c), forthcoming in Econometric Theory.
see e.g. Johansen (1995), and tests that consider single-equation models and assuming at most a single cointegrating vector under the alternative. Among the latter ones, we can further distinguish between approaches based on the triangular representation of a cointegration system that naturally leads to residual-based tests for cointegration (e.g. Phillips and Ouliaris, 1990) that make use of semi-parametric correction for endogeneity and serial correlation; and those based on fully specified parametric data generating processes that naturally lead to single equation dynamic models. The ECM test considered in this chapter falls in this category. As already discussed in the literature, ECM tests are an attractive option for cointegration testing, as, contrary to the more popular residual-based tests, ECM tests do not suffer from imposing potentially invalid common factor restrictions (Kremers, Ericsson, and Dolado, 1992; Banerjee, Dolado, and Mestre, 1998; Zivot, 2000). Moreover, Pesavento (2004) analyzes several tests which have as null hypothesis no cointegration, including the residual ADF test by Engle and Granger (1987) and the maximum eigenvalue test by Johansen and Juselius (1990), and finds that among these the ECM tests perform best in terms of power both in small and large samples, while performing similarly as the other tests in terms of size. ECM tests thus appear to be an appealing tool of testing for cointegration.

The ECM Wald test has as main advantage over the ECM t-test (Banerjee et al., 1998) that it is more intuitive and one does not have to add a redundant regressor if no particular cointegrating vector is specified. Although the Wald ECM test performs well in general, especially in terms of power, it still suffers from size distortions in finite samples (see for example Boswijk and Franses, 1992). It is well known that the bootstrap’s ability to provide asymptotic refinements often leads to a reduction of size distortions for hypothesis tests. Even under “non-favorable” conditions for the bootstrap, under which it is unclear whether it provides asymptotic refinements, such as when dealing with nonstationary time series, the bootstrap has been shown to reduce size distortions in finite samples (see for example the tests for unit roots considered in Chang and Park, 2003, Palm et al., 2008a or Paparoditis and Politis, 2003).2

Little is known so far about the application of the bootstrap to cointegration testing in error correction models. Swensen (2006) and Trenkler (2009) provide theoretical and simulation results on bootstrap versions of the trace test for cointegration rank by Johansen (1995). Their setting differs from ours in that we a priori assume that the cointegrating rank is at most one. Seo (2006) provides analytical and simulation results for a residual-based bootstrap test in a threshold vector error correction model. Closer to our setting, Mantalos and Shukur (1998) and Ahlgren (2000) consider a bootstrap version of the test with known cointegrating vector by Kremers et al. (1992), however they only provide simulation results for a simple model. In this chapter we will allow for more general dependence over time in our model, and we provide analytical as well as simulation results.

Our chapter relies on the sieve bootstrap introduced by Bühlmann (1997), a

\[\text{Eq. 1}\]

The notable exception to the lack of theoretical results is Park (2003), who shows that bootstrap ADF tests offer asymptotic refinements under the assumption that the errors are a finite AR process with known order.
method that can handle time series dependence in the form of a general linear process that is approximated by an autoregressive process. The sieve bootstrap method is easy to use and performs well relative to other time series bootstrap methods, especially the block bootstrap (for a comparison between methods in the unit root setting, see Palm et al., 2008a). The condition of linearity is fulfilled by a large class of processes, and is needed to validate the use of the Wald test without the bootstrap as well.

The contribution of the chapter is threefold. First, we prove that the sieve bootstrap version of the single-equation Wald test of no cointegration is asymptotically valid. The proofs are given in detail for the multivariate setting, such that proofs of other types of tests could be done along the same lines as presented here. Second, we provide simulation results showing that the bootstrap version of the Wald test has better properties in finite samples than the asymptotic test. Third, we investigate the sensitivity of the bootstrap to various specifications of deterministic components and alternative distributional assumptions.

The structure of the chapter is as follows. Section 4.2 explains the model and assumptions. The construction of the bootstrap test and the establishment of its asymptotic validity are discussed in Section 4.3. Our simulation study is presented in Section 4.4. The inclusion of deterministic components is discussed in Section 4.5. Section 4.6 concludes. All proofs are contained in Appendix 4.A.

Finally, a word on notation. We use $| \cdot |$ to denote the Euclidean norm for vectors and matrices, i.e. $|v| = (v'v)^{1/2}$ for a vector $v$ and $|M| = (\text{tr } M'M)^{1/2}$ for a matrix $M$. For matrices we also use the operator norm $||M|| = \max_v |Mv|/|v|$. $W(r) = (W_1(r), W_2(r)')'$ denotes a multivariate standard Brownian motion of dimension $(1 + l)$. $[x]$ is the largest integer smaller than or equal to $x$. Convergence in distribution (probability) is denoted by $\xrightarrow{d} (\xrightarrow{p})$. Bootstrap quantities (conditional on the original sample) are indicated by appending a superscript $*$ to the standard notation. Subscripts $p$ (or $q$) are used to indicate quantities depending on approximations of infinite order models by models of order $p$ (or $q$). For simplicity we suppress these subscripts whenever clarity allows it.

## 4.2 The model

Our Data Generating Process (DGP) is closely related to that of Pesavento (2004). We let our $(1 + l)$-dimensional time series $z_t = (y_t, x_t')'$ be described by the process

$$z_t = \mu + \tau t + \xi_t, \quad (4.1)$$

The stochastic component $\xi_t$ is given by

$$\Delta \xi_t = (\rho - 1)\alpha \beta' \xi_{t-1} + u_t, \quad (4.2)$$

where

$$u_t = \Psi(L)\varepsilon_t \quad (4.3)$$
with \( \Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j \). Furthermore we assume that \( \zeta_0 = 0 \). The null hypothesis is \( H_0 : \rho = 1 \), there is no cointegration. Under the alternative \( H_1 : \rho < 1 \) there is cointegration with a single cointegrating vector \( \beta \) and the error correction term must be present in the equation for \( y_t \). Also, we impose \( \alpha_1 = 1 \) and \( \alpha_2 = 0 \), which follows from the triangular representation of the model as in Pesavento (2004) and is needed for identification purposes. These points are formalized in Assumption 4.1.

**Assumption 4.1.** We assume

(i) \( \alpha \beta' \) is of rank 1, i.e. there is a single \((1+l)\)-dimensional cointegrating vector \( \beta \),

(ii) \( \beta \) is normalized on the coefficient of \( y_t \), i.e. \( \beta = (1, -\gamma')' \),

(iii) \( \alpha = (1, 0)' \).

It is important to remark that Assumption 1 is of no importance for the derivation of the null distribution of the tests as it only concerns the situation where cointegration is present in the system. It is however important to derive the equivalence between the triangular representation and the ECM form. Assumption 1 is also important to enable us to focus on a single equation ECM and to rule out cases where the ECM tests would trivially have low power. This would for example occur under the alternative if the cointegration vector only appears in the equation for the conditioning variables \( x_t \).

Equation (4.3) shows that we take \( u_t \) to be a linear process (Phillips and Solo, 1992). Assumption 4.2 ensures the invertibility of \( u_t \) and the existence of moments of \( \epsilon_t \). These assumptions are not very stringent and encompass many assumptions (including all finite VARMA models) that are often used in cointegration analysis.

**Assumption 4.2.** We assume

(i) \( \epsilon_t \) are i.i.d. with \( E(\epsilon_t) = 0, E(\epsilon_t \epsilon_t') = \Sigma \) and \( E|\epsilon_t|^4 < \infty \).

(ii) \( \det(\Psi(z)) \neq 0 \) for all \( |z| \leq 1 \), and \( \sum_{j=0}^{\infty} j |\Psi_j| < \infty \).

By Assumption 4.2 we may write \( \Phi(L) = \sum_{j=0}^{\infty} \Phi_j L^j = \Psi(L)^{-1} \). We may then substitute equation (4.1) into (4.2) and apply the Beveridge-Nelson decomposition to show as in Pesavento (2004) that this model can be rewritten in VECM form

\[
\Delta z_t = (\rho - 1) \Phi(1) \alpha \beta' (z_{t-1} - \mu - \tau (t - 1)) + \tilde{\tau} + \Phi^*(L) \Delta z_{t-1} + \epsilon_t \tag{4.4}
\]

where,

\[
\Phi^*(L) = \sum_{j=0}^{\infty} \left( 1 - \rho \right) \left( \sum_{i=j+1}^{\infty} \Phi_i \right) \alpha \beta' - \Phi_{j+1} L^j
\]

\footnote{This assumption is made for expositional simplicity only and can be extended to \( \zeta_0 = O_p(1) \).}

\footnote{Pesavento (2004) shows that this restriction corresponds to the assumption that \( x_t \) are not mutually cointegrated, as required under Assumption 4.1(i), and are known a priori to be \( I(1) \).}
4.3 The bootstrap test and asymptotics

and

\[ \hat{\tau} = \left( \sum_{j=0}^{\infty} \Phi_j + (\rho - 1) \left( \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} \Phi_i \right) \right) \alpha \beta' \tau. \]

It can be seen from the above representation that \( z_t \) has a drift if \( \tau \neq 0 \), and this drift leads to a linear trend in the cointegrating relation if \( \beta' \tau \neq 0 \). The constant \( \mu \) only appears in the cointegrating relation; note that the cointegrating relation has mean zero if \( \beta' \mu = 0 \).

Pesavento shows that the model can be written in triangular form as well, which makes it a very flexible model. As we do not need that representation here, we continue with the VECM representation (4.4) and condition on \( x_t \) to obtain

\[ \Delta y_t = (\rho - 1) \theta \beta'(z_{t-1} - \mu - \tau(t - 1)) + \hat{\tau} + \pi_0' \Delta x_t + \sum_{j=1}^{\infty} \pi_j' \Delta z_{t-j} + \xi_t, \quad (4.5) \]

where \( \xi_t = \varepsilon_{1,t} - \Sigma_{12} \Sigma_{22}^{-1} \varepsilon_{2,t} \sim \text{i.i.d. } (0, \omega^2) \) and \( \theta = \Phi_1(1) \alpha - \Sigma_{12} \Sigma_{22}^{-1} \Phi_2(1) \alpha \) with \( \Phi(1) = \Phi_1(1)', \Phi_2(1)' \).

The advantage of this framework is that its assumptions are weaker than what is usually assumed for tests based on a conditional ECM, as it does not impose that \( x_t \) are weakly exogenous for \( \beta \) under the alternative of cointegration. Under the null however, the error correction term does not appear in the marginal equations, which makes a test on the error correction term in the conditional model a valid test for cointegration (Boswijk, 1994).

4.3 The bootstrap test and asymptotics

4.3.1 Test statistic

The Wald test proposed by Boswijk (1994) is based on the conditional model (4.5). Consider the regression

\[ \Delta y_t = \delta' \tilde{z}_{t-1} + \lambda' D_t + \pi_0' \Delta x_t + \sum_{j=1}^{p} \pi_j' \Delta z_{t-j} + \xi_{p,t}, \quad (4.6) \]

where \( D_t \) are the (unrestricted) deterministic components included in the regression, \( \tilde{z}_{t-1} = (z'_{t-1}, D'_{t-1})' \) where \( D'_t \) are the deterministic components that are restricted to be equal to zero under the null (see Section 4.5) and \( \xi_{p,t} = \sum_{j=p+1}^{\infty} \pi_j' \Delta z_{t-j} + \xi_t \). If \( \rho = 1 \), \( \delta' = (\rho - 1) \theta \beta' = 0 \), which leads to the test statistic

\[ T_{\text{wald}} = \delta' \text{Var}(\hat{\delta})^{-1} \hat{\delta}, \quad (4.7) \]

Note that \( \omega^2 = \sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \). \( \Sigma \) and \( \varepsilon_t \) have been partitioned conformably with \( y_t \) and \( x_t \), i.e. \( \varepsilon_{1,t} \) is a scalar and \( \varepsilon_{2,t} \) is an \( l \)-dimensional vector.
where $\hat{\delta}$ is the OLS estimator of $\delta$ in (4.6) and $\text{Var}(\hat{\delta})$ is its estimated covariance matrix. The null hypothesis of no cointegration is then rejected for large values of $T_{\text{wald}}$.

We let the lag length $p$ in regression (4.6) grow to infinity at a controlled rate.

**Assumption 4.3.** Let $p \to \infty$ and $p = o(n^{1/2})$ as $n \to \infty$.

The limiting distribution of $T_{\text{wald}}$ can be found in Boswijk (1994) for the ECM with finite autoregressive dependence and in Pesavento (2004) for the infinite-order model. The asymptotic distribution of the test without the inclusion of any deterministic components (and with $\mu = \tau = 0$) is given for completeness in Lemma 4.1 without proof.

**Lemma 4.1.** Under Assumptions 4.2 and 4.3 we have that

$$T_{\text{wald}} \xrightarrow{d} \int_0^1 dW_1(r)W(r)^\prime \left[ \int_0^1 W(r)W(r)^\prime \,dr \right]^{-1} \int_0^1 W(r)dW_1(r)$$

where $T_{\text{wald}}$ is defined in equation (4.7).

### 4.3.2 Bootstrap method

The multivariate sieve bootstrap method we employ here is similar to the one employed by Chang, Park, and Song (2006). It is important to note that they study bootstrap inference on the cointegrating regressions and they do not consider bootstrap tests for no cointegration. The full algorithm is given below.

**Bootstrap Algorithm 4.1.**

Step 1: Fit a $\text{VAR}(q)$ process to $\Delta z_t$ by OLS and save the residuals

$$\tilde{\varepsilon}_{q,t} = \Delta z_t - \hat{\lambda}_sD^s_t - \sum_{j=1}^{q} \hat{\Phi}_j \Delta z_{t-j},$$

(4.8)

where $D^s_t$ are the deterministic components included in this sieve estimation (see Section 4.5 for details). Recenter the residuals $\tilde{\varepsilon}_{q,t}$ in the case where no constant is included to eliminate any drifts in the resampled series and save the recentered residuals $\tilde{\varepsilon}_{q,t} = \tilde{\varepsilon}_{q,t} - (n-q-1)^{-1} \sum_t \tilde{\varepsilon}_{q,t}$.\(^6\)

Step 2: Resample with replacement from $\tilde{\varepsilon}_{q,t}$ to obtain bootstrap errors $\varepsilon^*_t$.

Step 3: Build $u^*_t$ recursively as

$$u^*_t = \sum_{j=1}^{q} \hat{\Phi}_j u^*_{t-j} + \varepsilon^*_t,$$

(4.9)

\(^6\)In the cases where we do not include a constant in this regression the residuals may have a sample mean unequal to zero, even though their theoretical mean is zero. As the sample mean of the residuals becomes the population mean of the bootstrap errors, this may lead to (unwanted) drifts in the bootstrap sample.
4.3 The bootstrap test and asymptotics

using the estimated parameters $\hat{\Phi}_j$ from Step 1, and build $z^*_t$ as

$$z^*_t = z^*_{t-1} + u^*_t.$$  

(4.10)

Note that it is unnecessary to include deterministic components in this step, as the tests we consider are asymptotically similar (see Remark 4.8 in Section 4.5).

Step 4: Using the bootstrap sample $z^*_t$, obtain $\hat{\delta}^*$ from the regression

$$\Delta y^*_t = \delta^* z^*_{t-1} + \lambda^* D_t^* + \pi^*_0 \Delta x^*_t + \pi^*_j \Delta z^*_t - j + \zeta^*_{r, t},$$  

(4.11)

where $p^*$ is the lag length selected in the bootstrap regression (see Remark 4.6) and $z^*_{t-1} = (z^*_{t-1}, D^*_{r, t-1})'$. In order to get the correct asymptotic bootstrap distribution, one should always take $D_t^* = D_t$ and $D^*_r = D^*_r$.

Step 5: Repeat Steps 2 to 4 $B$ times, obtaining bootstrap test statistics $T^*_{\text{wald}, b}$, $b = 1, \ldots, B$, and select the bootstrap critical value $c^*_\alpha$ as $c^*_\alpha = \text{min}\{ c : \sum_{b=1}^B I(T^*_{\text{wald}, b} > c) \leq \alpha \}$, or equivalently as the $(1 - \alpha)$-quantile of the ordered $T^*_{\text{wald}, b}$ statistics. Reject the null of no cointegration if $T^*_{\text{wald}}$, calculated from equations (4.6) and (4.7), is larger than $c^*_\alpha$, where $\alpha$ is the nominal level of the test.

We need to allow the lag length $q$ in the sieve bootstrap to go to infinity at a controlled rate. We will use two assumptions.

**Assumption 4.4.** Let $q \to \infty$ and $q = o((n / \ln n)^{1/2})$ as $n \to \infty$.

**Assumption 4.4’.** Let $q \to \infty$ and $q = o((n / \ln n)^{1/3})$ as $n \to \infty$.

We also need an assumption on the relative speed of the lag lengths $p$ and $q$.

**Assumption 4.5.** Let $p/q \to \kappa > 1$ as $n \to \infty$, where $\kappa$ may be infinite.

Note that by allowing $\kappa$ to be infinite, we do not impose the same rate on $p$ and $q$. Assumption 4.5 imposes a lower bound but not an upper bound on the rate of $p$ (or equivalently an upper bound but not a lower bound on the rate of $q$).

**Remark 4.1.** In Step 3 we need to initialize $u^*_t$ in (4.9) and $z^*_t$ in (4.10). We propose to generate a large number of values of $u^*_t$ and delete the first generated values. This will ensure that $u^*_t$ is a stationary process. The initial values in (4.9) will then become unimportant as the realization of $u^*_t$ will not depend on them; hence they may be set equal to zero. An alternative is to take the first $q$ values
of $u^{\ast}_{t}$ equal to the first $q$ values of $u_{t}$; this however does not ensure stationarity of $u^{\ast}_{t}$.

As asymptotically the effect of $z^{\ast}_{0}$ disappears, we simply set $z^{\ast}_{0} = 0$. The logical alternative here would be to set $z^{\ast}_{0} = z_{0}$, especially in applications.

**Remark 4.2.** Instead of estimating the sieve under the null of no cointegration (which we impose by fitting the VAR model to $\Delta z_{t}$ in Step 1), we may also estimate it under the alternative of cointegration. In this case we would estimate the residuals as

$$
\hat{\varepsilon}_{q,t} = \Delta z_{t} - \hat{\lambda}_{q} D_{t}^{\ast} - \hat{\Phi}_{0} z_{t-1} - \sum_{j=1}^{q} \hat{\Phi}_{j} \Delta z_{t-j},
$$

where $\hat{\Phi}_{0}$ denotes the unrestricted OLS estimator and $D_{t}^{\ast}$ are the deterministic components included in this alternative-based sieve estimation. Note that even for the same deterministic setting, $D_{t}^{\ast}$ is not necessarily the same as $D_{t}^{\ast}$ in (4.8), as is explained in Section 4.5 (Remark 4.10).

In the context of unit root testing, Paparoditis and Politis (2005) advocate the use of such a “residual-based” estimation as opposed to the “difference-based” estimation in (4.8), claiming that the residual-based tests have better power properties. We will return to this point in our simulations in Section 4.4.

**Remark 4.3.** A second alternative bootstrap strategy would be to base the sieve bootstrap on the conditional/marginal ECM model instead of the VECM/VAR model. In this case we would need two separate equations to estimate residuals in Step 1. We would estimate the residuals from the conditional model as

$$
\hat{\varepsilon}_{1,q,t} = \Delta y_{t} - \hat{\lambda}_{s,1} D_{t}^{\ast,1} - \hat{\pi}_{0} \Delta x_{t} - \sum_{j=1}^{q} \hat{\pi}_{j} \Delta z_{t-j},
$$

and the residuals from the marginal model as

$$
\hat{\varepsilon}_{2,q,t} = \Delta x_{t} - \hat{\lambda}_{s,2} D_{t}^{\ast,2} - \sum_{j=1}^{q} \hat{\Phi}_{2,j} \Delta z_{t-j},
$$

for the difference-based alternative. We can of course also construct a residual-based version of this test. In the simulations in Section 4.4 we will look at these alternatives as well.

Although such an approach is closer in spirit to the single-equation Wald test statistic, it is basically just a reparametrization of the VECM approach, as the model on which the bootstrap is based is still completely specified. An alternative approach, which would be “truly conditional” on $x_{t}$, is to take $x_{t}$ as fixed and only resample $y_{t}$. To justify such an approach we would have to assume strong exogeneity, see Van Giersbergen and Kiviet (1996) for a discussion. This last approach will not be investigated in this chapter.
Remark 4.4. Although estimation under the alternative is an option in Step 1, it is not possible to build the bootstrap sample $z_t^*$ in Step 3 based on the alternative hypothesis, i.e. using

$$z_t^* = (I + \hat{\Phi}_0)z_{t-1}^* + u_t^*.$$  

(4.14)

Basawa et al. (1991b) show that if such an alternative-based recursion is used in the unit root setting, the limiting distribution of the bootstrap test statistic is random due to the discontinuity of the limiting distribution at the unit root. The same logic applies here, therefore the null hypothesis of no cointegration must be imposed in Step 3.

Remark 4.5. To obtain the theoretical results in the next subsection, we set all deterministic components equal to zero, both in the model ($\mu$ and $\tau$) and in the test (all variants of $D_t$). In Section 4.5 we will go into more detail about the inclusion of deterministic components, and present some simulation results. We conjecture that asymptotic validity still holds in the presence of deterministic components.

Remark 4.6. In Step 4 we specify the lag length in the bootstrap test regression (4.11) as $p^*$, in order to emphasize that this lag length does not have to be the same as the lag length in the original test regression (4.6). In finite samples the performance of the bootstrap test will be better if the lag length is allowed to be different. Just as for the original test regression (and the sieve bootstrap), the lag length can be chosen in practice using information criteria like AIC and BIC.

Obviously, $p^*$ has to fulfill the same conditions as $p$. Therefore, we can write $p^*$ as $p$ in the theoretical results, which is done for notational simplicity.

Remark 4.7. As we will see in the next subsection, Assumption 4.4 is sufficient to prove Theorem 4.1. However, to prove the second result needed for Theorem 4.2, we require the stronger assumption 4.4'. The result in the proof of Theorem 3.3 of Park (2002, p. 487, line 12), where it is stated (in Park’s notation) that

$$\sum_{k=1}^{p} k|\hat{\alpha}_{p,k}| = \sum_{k=1}^{p} k|\alpha_{p,k}| + o(1) \text{ a.s.,}$$

with $\hat{\alpha}_{p,k}$ being the OLS estimators of the $p$-th order autoregressive approximation of the univariate general linear process considered by Park (2002) with coefficients $\alpha_{p,k}$, does not go through with $p = o((n/\ln n)^{1/2})$. One needs a stronger restriction on $p$ to make the second part $o(1)$.

With our stronger Assumption 4.4' one can show that Theorem 3.3 of Park (2002) (and consequently our Theorem 4.2) holds.

4.3.3 Asymptotic results

In this section we will give the main theoretical results needed to show the asymptotic validity of the bootstrap test. As stated in Remark 4.5, we derive these results for the tests (and DGP) without deterministic components. The proofs of

7 We thank Anders Swensen for bringing this point to our attention in a personal communication.
all the results here plus additional lemmas can be found in Appendix A. Most of the proofs are based on the proofs in Chang et al. (2006), and the papers they refer to.

As we present all our proofs for vector processes, the theory employed in the chapter can be used to prove validity of other multivariate bootstrap procedures as well. Note that all our bootstrap weak convergence results hold in probability as we derive all underlying results in probability.

The first step in proving the asymptotic validity is the development of an invariance principle for the bootstrap errors $\varepsilon^*_t$.

**Theorem 4.1.** Under Assumptions 4.2 and 4.4, we have that

$$W_n^*(r) = n^{-1/2} \left[ \sum_{t=1}^{nr} \varepsilon^*_t d^r \right] L W(r) \quad \text{in probability}$$

where $L$ is a $(1 + l) \times (1 + l)$-dimensional lower triangular matrix such that the Cholesky decomposition of $\Sigma$ is equal to $LL'$.

We can show this result by first showing that $E^* |\varepsilon^*_t|^a = O_p(1)$ for some $a > 2$, and then referring to Einmahl (1987), who shows that an invariance principle holds if this condition is met.

From this result, with the help of the Beveridge-Nelson decomposition, we can construct an invariance principle for $u^*_t$.

**Theorem 4.2.** Under Assumptions 4.2 and 4.4' we have that

$$B_n^*(r) = n^{-1/2} \left[ \sum_{t=1}^{nr} u^*_t d^r \right] B(r) \quad \text{in probability},$$

where $B(r) = \Psi(1)LW(r)$ is an $(1 + l)$-dimensional Brownian motion.

Then, using Theorem 4.2, we can derive the limiting distributions of the elements of the test statistic, and finally show the consistency of the bootstrap variance estimator. With these results, we can then present Theorem 4.3 which establishes the asymptotic distribution of the bootstrap test statistic.

**Theorem 4.3.** Under Assumptions 4.2, 4.3, 4.4' and 4.5 we have that

$$T^n_{\text{wald}} \overset{d}{\to} \int_0^1 dW_1(r) W(r) W(r)' \left[ \int_0^1 W(r) W(r)' \, dr \right]^{-1} \int_0^1 W(r) dW_1(r) \quad \text{in probability}$$

where $T^n_{\text{wald}}$ is defined in equation (4.12).

Note that Theorem 4.3 shows that the bootstrap test statistic has the same asymptotic distribution as the original test statistic, which shows that the bootstrap test is asymptotically valid. Also note that the test statistic is asymptotically pivotal, which means that the bootstrap may offer asymptotic refinements, although this does not have to be so.
4.4 Simulations

We wish to study the small sample properties of our test by simulation. We compare our test with the test based on asymptotic critical values (provided by Boswijk, 1994) and with the three alternative bootstrap tests mentioned in Remarks 4.2 and 4.3. Our bootstrap test is denoted by $T_{v,n}^*$, where the subscript $v$ stands for estimation based on the VAR/VECM model, and the $n$ for estimation of the sieve bootstrap under the null. The alternative test discussed in Remark 4.2 is denoted by $T_{v,a}^*$, with the subscript $a$ indicating estimation under the alternative. Similarly, the two alternatives discussed in Remark 4.3 are given as $T_{c,n}^*$ and $T_{c,a}^*$, where the subscript $c$ indicates that these are based on the conditional/marginal model. Finally, the asymptotic test is denoted as $T_{as}^*$.

For the simulation study we use the same setup as Pesavento (2004). We let the bivariate series $(y_t, x_t)'$ be generated by the triangular system

\[ y_t = \gamma x_t + w_t, \]
\[ w_t = \rho w_{t-1} + v_{1t}, \]
\[ \Delta x_t = v_{2t}. \]

(4.15)

We take $\rho = 1$ to analyze the size of tests, and $\rho < 1$ for the power. For the local power analysis, $\rho = 1 + c/n$, where $n$, the sample size, is either 50 or 100. The tests are invariant to the true value of $\gamma$ as long as it is non-zero, therefore we set $\gamma = 1$. Furthermore we set $w_0 = x_0 = 0$.

The errors $v_t = (v_{1t}, v_{2t})'$ are generated as

\[ (1 - \Phi L)v_t = (1 + \Theta L)\varepsilon_t, \]

where $\varepsilon_t$ is generated as an i.i.d. sequence from a bivariate normal distribution with covariance matrix

\[ \Sigma = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}. \]

The exact parameter combinations considered are summarized in Table 4.1.

We can rewrite the above DGP in terms of the model in (4.2) by setting $\alpha = (1, 0)'$, $\beta = (1, -\gamma)'$ and $u_t = \begin{bmatrix} 1 & \gamma \end{bmatrix}' v_t$ in equation (4.2).

The lag lengths in (4.6), (4.8) and (4.11) are selected by BIC, with maximum lag lengths of 8 for $n = 50$ and 11 for $n = 100$. Each generated sample is used to perform all the tests, such that the lag length $p$ in (4.6) is always the same for all tests. Our results are based on 2000 simulations, with 999 bootstrap replications per simulation.

The results for the DGPs with white noise errors ($\Phi = \Theta = 0$) are given in Table 4.2. For this case, the asymptotic test has a reasonably good size, but the bootstrap tests clearly have sizes even closer to the nominal size, especially for $n = 50$. The rejection frequencies of the bootstrap tests are somewhat smaller than those of the asymptotic test under the alternatives considered, but it is difficult to compare powers as sizes are not equal. We therefore also report size-corrected
Table 4.1: Parameter combinations used in the simulation DGP

<table>
<thead>
<tr>
<th>$\Phi$</th>
<th>$\Theta$</th>
<th>$r$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>0 0</td>
<td>0, $\sqrt{0.3}$, $\sqrt{0.7}$, 0, -5, -10, -20</td>
<td></td>
</tr>
<tr>
<td>0 0</td>
<td>0 0</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0.2 0</td>
<td>0 0</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0 0.2</td>
<td>0 0</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0.8 0</td>
<td>0 0</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0 0.8</td>
<td>0 0</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0.2 0.5</td>
<td>0 0</td>
<td>$\sqrt{0.3}$</td>
<td>0, -5, -10, -20</td>
</tr>
<tr>
<td>0.5 0.2</td>
<td>0 0</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0 0</td>
<td>0.2 0</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0 0</td>
<td>0 0.2</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0 0</td>
<td>0.8 0</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0 0</td>
<td>0 0.8</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0 0</td>
<td>0.2 0.5</td>
<td>$\sqrt{0.3}$</td>
<td>0, -5, -10, -20</td>
</tr>
<tr>
<td>0 0</td>
<td>0.5 0.2</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0 0</td>
<td>-0.8 0</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
<tr>
<td>0 0</td>
<td>0 -0.8</td>
<td>$\sqrt{0.3}$</td>
<td>0</td>
</tr>
</tbody>
</table>

powers for the asymptotic test (in the Table as $T_{sc}$). The size-corrected power of the asymptotic test is close to the power of the bootstrap tests, which shows that the higher raw power of the asymptotic test is mainly due to the higher size distortions. All bootstrap tests perform similarly both in terms of size and power, indicating that there is no evidence of reduced power for the difference-based tests in this setting.

Table 4.3 gives the results for the size of the tests for DGPs with autoregressive and moving-average errors. For all DGPs considered here, there is a clear advantage of using the bootstrap, which virtually eliminates all size distortions except for the negative moving-average coefficients. Again note that the difference between the bootstrap and asymptotic test is the largest for $n = 50$. The bootstrap tests perform fairly similarly, with a minor advantage for the difference-based tests. This is especially noticeable for the DGP with negative moving-average coefficients.

To illustrate the power properties for DGPs allowing for some dependence in the errors, we selected one DGP with autoregressive and one with moving-average coefficients from the set considered above. The results are given in Table 4.4. We again have to be cautious when comparing raw powers as the sizes vary across the tests. We see that the asymptotic test has somewhat higher rejection frequencies than the bootstrap tests, but as in Table 4.2 the differences are due to high size distortions of the asymptotic test. This is confirmed by the size-corrected power of the asymptotic test, which is not better, and in some cases considerably worse, than the power of the bootstrap tests. The difference-based tests appear to have higher

---

8There is no need to correct the power of the bootstrap tests, as they have virtually no size distortions; their size-corrected powers would be almost the same as their raw powers.
Table 4.2: Size and power for white noise errors

<table>
<thead>
<tr>
<th>r</th>
<th>c</th>
<th>$T_{c,n}^*$</th>
<th>$T_{c,a}^*$</th>
<th>$T_{a,n}^*$</th>
<th>$T_{a,a}$</th>
<th>$T_{nc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.042</td>
<td>0.042</td>
<td>0.043</td>
<td>0.040</td>
<td>0.072</td>
</tr>
<tr>
<td>-5</td>
<td>0.102</td>
<td>0.105</td>
<td>0.103</td>
<td>0.163</td>
<td>0.121</td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td>0.281</td>
<td>0.284</td>
<td>0.283</td>
<td>0.414</td>
<td>0.322</td>
<td></td>
</tr>
<tr>
<td>-20</td>
<td>0.831</td>
<td>0.839</td>
<td>0.829</td>
<td>0.918</td>
<td>0.866</td>
<td></td>
</tr>
<tr>
<td>$\sqrt{0.3}$</td>
<td>0</td>
<td>0.052</td>
<td>0.052</td>
<td>0.052</td>
<td>0.051</td>
<td>0.085</td>
</tr>
<tr>
<td>-5</td>
<td>0.163</td>
<td>0.160</td>
<td>0.162</td>
<td>0.243</td>
<td>0.163</td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td>0.521</td>
<td>0.522</td>
<td>0.522</td>
<td>0.661</td>
<td>0.524</td>
<td></td>
</tr>
<tr>
<td>-20</td>
<td>0.949</td>
<td>0.950</td>
<td>0.960</td>
<td>0.984</td>
<td>0.964</td>
<td></td>
</tr>
<tr>
<td>$\sqrt{0.7}$</td>
<td>0</td>
<td>0.052</td>
<td>0.051</td>
<td>0.051</td>
<td>0.052</td>
<td>0.079</td>
</tr>
<tr>
<td>-5</td>
<td>0.501</td>
<td>0.498</td>
<td>0.503</td>
<td>0.618</td>
<td>0.488</td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td>0.898</td>
<td>0.898</td>
<td>0.903</td>
<td>0.935</td>
<td>0.898</td>
<td></td>
</tr>
<tr>
<td>-20</td>
<td>0.983</td>
<td>0.985</td>
<td>0.997</td>
<td>0.999</td>
<td>0.998</td>
<td></td>
</tr>
</tbody>
</table>

$n = 50$

<table>
<thead>
<tr>
<th>r</th>
<th>c</th>
<th>$T_{c,n}^*$</th>
<th>$T_{c,a}^*$</th>
<th>$T_{a,n}^*$</th>
<th>$T_{a,a}$</th>
<th>$T_{nc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.048</td>
<td>0.048</td>
<td>0.048</td>
<td>0.049</td>
<td>0.059</td>
</tr>
<tr>
<td>-5</td>
<td>0.108</td>
<td>0.108</td>
<td>0.109</td>
<td>0.133</td>
<td>0.113</td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td>0.317</td>
<td>0.315</td>
<td>0.314</td>
<td>0.381</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td>-20</td>
<td>0.868</td>
<td>0.865</td>
<td>0.861</td>
<td>0.906</td>
<td>0.875</td>
<td></td>
</tr>
<tr>
<td>$\sqrt{0.3}$</td>
<td>0</td>
<td>0.061</td>
<td>0.061</td>
<td>0.061</td>
<td>0.059</td>
<td>0.072</td>
</tr>
<tr>
<td>-5</td>
<td>0.180</td>
<td>0.181</td>
<td>0.185</td>
<td>0.225</td>
<td>0.160</td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td>0.541</td>
<td>0.535</td>
<td>0.541</td>
<td>0.601</td>
<td>0.501</td>
<td></td>
</tr>
<tr>
<td>-20</td>
<td>0.960</td>
<td>0.963</td>
<td>0.963</td>
<td>0.978</td>
<td>0.952</td>
<td></td>
</tr>
<tr>
<td>$\sqrt{0.7}$</td>
<td>0</td>
<td>0.056</td>
<td>0.056</td>
<td>0.053</td>
<td>0.058</td>
<td>0.071</td>
</tr>
<tr>
<td>-5</td>
<td>0.545</td>
<td>0.538</td>
<td>0.539</td>
<td>0.597</td>
<td>0.524</td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td>0.936</td>
<td>0.938</td>
<td>0.933</td>
<td>0.949</td>
<td>0.933</td>
<td></td>
</tr>
<tr>
<td>-20</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

$n = 100$
Table 4.3: Size for serially correlated errors

<table>
<thead>
<tr>
<th>Φ</th>
<th>Θ</th>
<th>( T_{c,u}^* )</th>
<th>( T_{c,a}^* )</th>
<th>( T_{c,n}^* )</th>
<th>( T_{a,s}^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( n = 50 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ 0.2 \ 0 ]</td>
<td>[ 0 \ 0 ]</td>
<td>0.060</td>
<td>0.057</td>
<td>0.059</td>
<td>0.058</td>
</tr>
<tr>
<td>[ 0.8 \ 0 ]</td>
<td>[ 0 \ 0 ]</td>
<td>0.049</td>
<td>0.045</td>
<td>0.045</td>
<td>0.046</td>
</tr>
<tr>
<td>[ 0.2 \ 0.5 ]</td>
<td>[ 0 \ 0 ]</td>
<td>0.059</td>
<td>0.088</td>
<td>0.055</td>
<td>0.088</td>
</tr>
<tr>
<td>[ 0.5 \ 0.2 ]</td>
<td>[ 0 \ 0 ]</td>
<td>0.063</td>
<td>0.061</td>
<td>0.065</td>
<td>0.058</td>
</tr>
<tr>
<td>[ 0 \ 0 ]</td>
<td>[ 0.2 \ 0 ]</td>
<td>0.050</td>
<td>0.055</td>
<td>0.050</td>
<td>0.055</td>
</tr>
<tr>
<td>[ 0 \ 0 ]</td>
<td>[ 0 \ 0.8 ]</td>
<td>0.075</td>
<td>0.092</td>
<td>0.075</td>
<td>0.095</td>
</tr>
<tr>
<td>[ 0 \ 0 ]</td>
<td>[ -0.8 \ 0 ]</td>
<td>0.459</td>
<td>0.625</td>
<td>0.453</td>
<td>0.625</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( n = 100 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ 0.2 \ 0 ]</td>
<td>[ 0 \ 0 ]</td>
<td>0.063</td>
<td>0.065</td>
<td>0.059</td>
<td>0.064</td>
</tr>
<tr>
<td>[ 0.8 \ 0 ]</td>
<td>[ 0 \ 0 ]</td>
<td>0.051</td>
<td>0.047</td>
<td>0.050</td>
<td>0.049</td>
</tr>
<tr>
<td>[ 0.2 \ 0.5 ]</td>
<td>[ 0 \ 0 ]</td>
<td>0.056</td>
<td>0.056</td>
<td>0.056</td>
<td>0.052</td>
</tr>
<tr>
<td>[ 0.5 \ 0.2 ]</td>
<td>[ 0 \ 0 ]</td>
<td>0.056</td>
<td>0.059</td>
<td>0.059</td>
<td>0.058</td>
</tr>
<tr>
<td>[ 0 \ 0 ]</td>
<td>[ 0 \ 0.8 ]</td>
<td>0.057</td>
<td>0.067</td>
<td>0.058</td>
<td>0.066</td>
</tr>
<tr>
<td>[ 0 \ 0 ]</td>
<td>[ 0 \ 0.5 ]</td>
<td>0.070</td>
<td>0.075</td>
<td>0.067</td>
<td>0.078</td>
</tr>
<tr>
<td>[ 0 \ 0 ]</td>
<td>[ -0.8 \ 0 ]</td>
<td>0.485</td>
<td>0.518</td>
<td>0.481</td>
<td>0.525</td>
</tr>
</tbody>
</table>
Table 4.4: Power for serially correlated errors

<table>
<thead>
<tr>
<th>dynamics</th>
<th>( \Phi = \begin{bmatrix} 0.2 &amp; 0.2 \ 0.5 &amp; 0.2 \end{bmatrix} )</th>
<th>( \Theta = \begin{bmatrix} 0.2 &amp; 0.2 \ 0.5 &amp; 0.2 \end{bmatrix} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 50 )</td>
<td>( \Phi )</td>
<td>( \Theta )</td>
</tr>
<tr>
<td>-5</td>
<td>0.624</td>
<td>0.401</td>
</tr>
<tr>
<td>-10</td>
<td>0.857</td>
<td>0.810</td>
</tr>
<tr>
<td>-20</td>
<td>0.929</td>
<td>0.940</td>
</tr>
<tr>
<td>( \Phi = \begin{bmatrix} 0.2 &amp; 0.2 \ 0.5 &amp; 0.2 \end{bmatrix} )</td>
<td>( \Theta = \begin{bmatrix} 0.2 &amp; 0.2 \ 0.5 &amp; 0.2 \end{bmatrix} )</td>
<td></td>
</tr>
<tr>
<td>( n = 100 )</td>
<td>( \Phi )</td>
<td>( \Theta )</td>
</tr>
<tr>
<td>-5</td>
<td>0.893</td>
<td>0.493</td>
</tr>
<tr>
<td>-10</td>
<td>0.993</td>
<td>0.907</td>
</tr>
<tr>
<td>-20</td>
<td>0.999</td>
<td>0.998</td>
</tr>
</tbody>
</table>

4.4 Simulations

power than the residual-based tests (especially for \( n = 50 \) and for alternatives close to the null). This is quite surprising, as it is exactly the opposite of what Paparoditis and Politis (2005) found for unit root tests. This may possibly be a small sample phenomenon reflecting the fact that very often imposing invalid restrictions may lead to improved finite sample statistical inference by reducing the effect of sampling errors.

These results show that the bootstrap tests all offer significant size improvements over the asymptotic test, while retaining quite good power properties. Note that the four bootstrap tests perform similarly, with a small advantage for the difference-based tests, both in terms of size and power. The bootstrap tests based on the conditional-marginal representation perform as their counterparts based on the vector representation, thus giving no reason to prefer the conditional representation over the more straightforward vector representation.

As suggested by a referee, the similar performances of the bootstrap tests based on the vector representation and the conditional-marginal representation may be due to the normality of the innovations in our DGP. In order address this
issue, we also performed simulations where the ε_t’s are generated from non-normal distributions, in particular central χ²- and t-distributions. The simulation results not reported here show that the two representations also lead to very similar results if the variables are not normal.

We also investigated the sensitivity to the form of vt. In the first analysis we generate the innovations ε_t as multivariate GARCH errors, which fall outside the class of processes defined by Assumption 4.2. In the second analysis we consider a Markov-switching model in which the parameters of the short-run dynamics are generated by a Markov process. The results show that the bootstrap tests are robust against both types of processes.

Finally, we also performed simulations with the original DGP using AIC in- stead of BIC to select lag lengths. The results show that the bootstrap tests are somewhat undersized. The only notable improvement of the size of the bootstrap tests with respect to lag length selection by BIC occurs in the case of the large negative MA parameters. The power of the bootstrap tests is adversely affected by the use of AIC. Surprisingly, the asymptotic test has larger size distortions using AIC than BIC.⁹

### 4.5 Deterministic components

In this section we will discuss how to include deterministic components in the tests. Deterministic components have to be included both in the test regression (Dt and Dr_t in equation (4.6) and their bootstrap counterparts in equation (4.11)) and in Step 1 of the bootstrap procedure (Ds_t in equation (4.8)). We consider the five different options proposed by Boswijk (1994).

The first option is to simply leave out all deterministic components, which is the case we analyzed before in the chapter. Obviously this is only valid if both µ and τ in equation (4.1) are equal to zero.

The second and third options (Boswijk’s ξ∗µ and ξµ) arise if there is no drift in z_t (τ = 0). In this case we include an intercept in regression (4.6) and its bootstrap equivalent (4.11). The intercept can but need not be restricted to zero under the null of no cointegration. In the first case Dt = 0 and Dr_t = 1, in the second case Dt = 1 and Dr_t = 0. As in both cases z_t does not have a drift, there is no need to include any deterministic components in Step 1 of the bootstrap procedure, hence Ds_t = 0.

If the variables are generated by a process with drift, we have to include a linear trend as well as an intercept in equations (4.6) and (4.11) (Boswijk’s ξ∗τ and ξτ). Again we can either restrict the trend to be equal to zero under the null, in which case Dt = 1 and Dr_t = t, or we leave it unrestricted, in which case Dt = (1, t)′ and Dr_t = 0. As Δz_t now has a nonzero mean, we include a constant term in equation (4.8) in Step 1 of the bootstrap procedure, i.e. we set Ds_t = 1 in both cases.

---

⁹The results of all the additional simulations discussed above can be found on the website http://www.personeel.unimaas.nl/s.smeekes/research.htm.
4.5 Deterministic components

Remark 4.8. While it is possible to account for the presence of deterministic components in Step 3 of the bootstrap algorithm as well, it is not necessary. By specifying the tests as above, the tests are similar, i.e. their asymptotic distributions do not depend on the true value of the deterministic components. Therefore, building the bootstrap process with or without deterministic components will both lead to the correct limiting distribution, as long as the deterministic specification in the bootstrap test regression (4.11) is the same as the specification in the original regression (4.6), i.e. \( D_t^* = D_t \) and \( D_t^* = D_t \).\(^{10}\)

Remark 4.9. One might want to use the test with the unrestricted constant term to deal with the situation where the variables have a drift, but the drift does not lead to a time trend in the cointegrating relation (\( \beta' \tau = 0 \)). However, Boswijk (1994) stresses that in this case the asymptotic distribution of the test will not be similar and depend on whether the drift is zero or not. Therefore we do not consider this to be a viable option.

Remark 4.10. One can also adapt the bootstrap procedure mentioned in Remark 4.2 to the inclusion of deterministic components. As estimation in Step 1 is done under the alternative hypothesis, the inclusion of deterministic components is slightly different. If we only include a constant term in the regression, then a constant term must be included in equation (4.13) as well, hence \( D_t^{*,e} = 1 \). If the variables are generated by a drift, and a trend is added to the regression, \( D_t^{*,e} = (1, t)' \).

To illustrate the tests with deterministic components, we perform a small simulation study. The DGP used for the simulations corresponds to the DGP used in Section 4.4, except that we now add deterministic components to the triangular system as follows.

\[
\begin{align*}
y_t &= \mu_1 + \tau_1 t + \gamma x_t + w_t, \\
w_t &= \rho w_{t-1} + v_{1t}, \\
\Delta(x_t - \mu_2 - \tau_2 t) &= v_{2t}.
\end{align*}
\]

(4.16)

Note that \( \mu_1 \) and \( \tau_1 \) correspond to \( \beta' \mu \) and \( \beta' \tau \) respectively in equation (4.4). To reduce the size of the experiment we only report simulations for \( n = 50 \), and for \( c = 0 \) and \( c = -10 \). Also, we only consider three combinations of \( \Phi \) and \( \Theta \): \( \Phi = \Theta = 0; \Phi = [0.2 \ 0.5 \ 0.2] \) and \( \Theta = 0; \) and \( \Phi = 0 \) and \( \Theta = [0.2 \ 0.5 \ 0.2] \). We restrict our attention to the two bootstrap variants \( T_{v,n}^* \) and \( T_{v,a}^* \) and the asymptotic test \( T_{as} \).

We consider two models without a drift, and two where a drift is present. For the models without drift, a DGP with no deterministic components and one with just a constant term are chosen. For the models with drift, we select one DGP where the drift cancels out in the direction of the cointegrating vector (i.e. \( \tau_1 = 0 \)), and one where it does not. For each model we perform the tests with every deterministic specification that is appropriate for that specific model. The specific

\(^{10}\)Unreported simulation results, which can also be found on the website mentioned above, show that in finite samples the tests perform the same whether or not deterministic components are included in Step 3.
values used and the corresponding empirical rejection frequencies can be found in Table 4.5.

It can be seen from the table that the size of the bootstrap test is satisfactory for all settings considered. As in Section 4.4, the null-based test has slightly better size than the alternative-based test in the presence of serial correlation. The asymptotic test has again large size distortions almost everywhere. In terms of power the conclusions are similar to those drawn in Section 4.4 as well. Also, both in terms of size and power, the rejection frequencies for a particular deterministic specification of the tests \(D_t^{(r)}\), are comparable across different specifications for the trends in the DGP \((\mu_t, \tau_t)\), confirming the similarity of the tests.

Noticeable is that the bootstrap tests lose power if deterministic components are included unnecessarily. This is very much a small sample effect, unreported simulations for \(n = 100\) show that this effect, although still present, is less pronounced there. The asymptotic test does not seem to lose as much power. This can be explained by the fact that (contrary to the bootstrap tests) the size distortions of the asymptotic test increase when deterministic components are added unnecessarily. It also appears that the tests with unrestricted deterministic components are slightly more powerful than their restricted counterparts.

### 4.6 Conclusion

In this chapter we present a bootstrap version of the Wald test for cointegration in a conditional single-equation ECM originally proposed by Boswijk (1994) and also considered by Pesavento (2004). A multivariate sieve bootstrap method is used to deal with dependence in the data, and shown to be asymptotically valid. We also consider several alternative bootstrap tests, for which the asymptotic validity can be established in a similar fashion, and show how deterministic components can be included in the test.

The small sample properties of our bootstrap tests are studied by simulation, and compared to those of the asymptotic test and several alternative bootstrap tests. All bootstrap tests clearly outperform the asymptotic test in terms of size, while retaining good power. Our bootstrap test based on the null hypothesis performs slightly better in terms of size and power than the bootstrap test based on the alternative, while the performance of the tests based on the vector representation is very similar to that of the tests based on the conditional representation. The bootstrap tests with deterministic components retain excellent size properties and are insensitive to the true value of the trends in the model as long as sufficient deterministic components are included.

The results show that our bootstrap version of the Wald ECM test is worth being considered in empirical research, as our test can be seen to improve upon the original Wald test considered by Boswijk (1994) and Pesavento (2004). The Wald ECM test easily allows for other bootstrap variants as well, such as those considered in the simulation study, or block bootstrap methods, which account for somewhat more general DGP’s. Such tests could easily be placed in the framework
Table 4.5: Size and power for tests with deterministic trends

<table>
<thead>
<tr>
<th>$\mu_1 \tau_1$</th>
<th>$\mu_2 \tau_2$</th>
<th>$D_t^{(r)}$</th>
<th>$T_{\nu,n}^*$</th>
<th>$T_{\nu,a}^*$</th>
<th>$T_{\nu,n}^{*s}$</th>
<th>$T_{\nu,a}^{*s}$</th>
<th>$T_{\nu,n}^{*s}$</th>
<th>$T_{\nu,a}^{*s}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi = \Theta = 0$</td>
<td>$\Phi = \begin{bmatrix} 0.2 &amp; 0.5 \ 0.5 &amp; 0.2 \end{bmatrix}$</td>
<td>$\Theta = \begin{bmatrix} 0.2 &amp; 0.5 \ 0.5 &amp; 0.2 \end{bmatrix}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0</td>
<td>$D_t = 1$</td>
<td>0.051</td>
<td>0.051</td>
<td>0.106</td>
<td>0.050</td>
<td>0.104</td>
<td>0.291</td>
<td>0.059</td>
</tr>
<tr>
<td>0 0</td>
<td>$D_t = t$</td>
<td>0.046</td>
<td>0.043</td>
<td>0.133</td>
<td>0.050</td>
<td>0.135</td>
<td>0.372</td>
<td>0.064</td>
</tr>
<tr>
<td>1 0</td>
<td>$D_t = 1$</td>
<td>0.046</td>
<td>0.049</td>
<td>0.102</td>
<td>0.047</td>
<td>0.108</td>
<td>0.282</td>
<td>0.061</td>
</tr>
<tr>
<td>1 0</td>
<td>$D_t = t$</td>
<td>0.059</td>
<td>0.051</td>
<td>0.134</td>
<td>0.042</td>
<td>0.131</td>
<td>0.373</td>
<td>0.063</td>
</tr>
<tr>
<td>1 0</td>
<td>$D_t = 1$, $t$</td>
<td>0.035</td>
<td>0.034</td>
<td>0.114</td>
<td>0.031</td>
<td>0.114</td>
<td>0.346</td>
<td>0.075</td>
</tr>
<tr>
<td>1 1</td>
<td>$D_t = t$</td>
<td>0.051</td>
<td>0.053</td>
<td>0.154</td>
<td>0.034</td>
<td>0.116</td>
<td>0.358</td>
<td>0.057</td>
</tr>
<tr>
<td>1 1</td>
<td>$D_t = 1$, $t$</td>
<td>0.050</td>
<td>0.053</td>
<td>0.125</td>
<td>0.029</td>
<td>0.121</td>
<td>0.333</td>
<td>0.065</td>
</tr>
<tr>
<td>$c = -10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0</td>
<td>$D_t = 1$</td>
<td>0.244</td>
<td>0.249</td>
<td>0.420</td>
<td>0.667</td>
<td>0.612</td>
<td>0.902</td>
<td>0.552</td>
</tr>
<tr>
<td>0 0</td>
<td>$D_t = t$</td>
<td>0.166</td>
<td>0.159</td>
<td>0.387</td>
<td>0.514</td>
<td>0.356</td>
<td>0.859</td>
<td>0.363</td>
</tr>
<tr>
<td>1 0</td>
<td>$D_t = 1$</td>
<td>0.173</td>
<td>0.170</td>
<td>0.384</td>
<td>0.497</td>
<td>0.344</td>
<td>0.801</td>
<td>0.360</td>
</tr>
<tr>
<td>1 0</td>
<td>$D_t = t$</td>
<td>0.247</td>
<td>0.245</td>
<td>0.428</td>
<td>0.682</td>
<td>0.624</td>
<td>0.906</td>
<td>0.548</td>
</tr>
<tr>
<td>1 1</td>
<td>$D_t = 1$, $t$</td>
<td>0.133</td>
<td>0.140</td>
<td>0.355</td>
<td>0.508</td>
<td>0.334</td>
<td>0.855</td>
<td>0.356</td>
</tr>
<tr>
<td>$c = -15$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0</td>
<td>$D_t = 1$</td>
<td>0.156</td>
<td>0.159</td>
<td>0.391</td>
<td>0.484</td>
<td>0.332</td>
<td>0.819</td>
<td>0.369</td>
</tr>
<tr>
<td>1 0</td>
<td>$D_t = t$</td>
<td>0.156</td>
<td>0.153</td>
<td>0.370</td>
<td>0.501</td>
<td>0.327</td>
<td>0.843</td>
<td>0.364</td>
</tr>
<tr>
<td>1 1</td>
<td>$D_t = 1$, $t$</td>
<td>0.162</td>
<td>0.170</td>
<td>0.363</td>
<td>0.481</td>
<td>0.318</td>
<td>0.809</td>
<td>0.366</td>
</tr>
</tbody>
</table>
4 Sieve Bootstrap ECM Cointegration Test

presented here.

4.A Appendix: Proofs

All bootstrap weak convergence results that we present in the following are in probability. We do not add this explicitly to every result in order to simplify the notation.

Also note that we define bootstrap stochastic order symbols $O_p^s(\cdot)$ and $o_p^s(\cdot)$ in the same way as $O_p(\cdot)$ and $o_p(\cdot)$ for the original sample (see Chang and Park, 2003, Remark 1).

In order to prove Theorem 4.1, we first need the following lemma.

Lemma 4.A.1. Under Assumptions 4.2 and 4.4 we have for any $2 < a \leq 4$

$$\mathbb{E}^* |\varepsilon_t^*|^a = O_p(1).$$


\[
\mathbb{E}^* |\varepsilon_t^*|^a = \frac{1}{n} \sum_{t=1}^n |\varepsilon_{q,t} - \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{q,t}|^a
\]

\[
= \frac{1}{n} \sum_{t=1}^n \varepsilon_{q,t} - \epsilon_t + \varepsilon_t - \epsilon_t - \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{q,t} |^a
\]

\[
\leq 4^{a-1} \frac{1}{n} \sum_{t=1}^n \left( |\varepsilon_{q,t} - \varepsilon_{q,t}|^a + |\varepsilon_{q,t} - \varepsilon_t|^a + |\varepsilon_t|^a + \frac{1}{n} \sum_{t=1}^n |\varepsilon_{q,t}|^a \right)
\]

\[
= c(A_n + B_n + C_n + D_n)
\]

where

\[
A_n = \frac{1}{n} \sum_{t=1}^n |\varepsilon_t|^a
\]

\[
B_n = \frac{1}{n} \sum_{t=1}^n |\varepsilon_{q,t} - \varepsilon_t|^a
\]

\[
C_n = \frac{1}{n} \sum_{t=1}^n |\varepsilon_{q,t} - \varepsilon_{q,t}|^a
\]

\[
D_n = \frac{1}{n} \sum_{t=1}^n |\varepsilon_{q,t}|^a
\]

and $c = 4^{a-1}$ is a constant not depending on $n$. Note that $\varepsilon_{q,t}$ is defined as

\[
\varepsilon_{q,t} = u_t - \sum_{j=1}^q \Phi_j u_{t-j} = \varepsilon_t + \sum_{j=q+1}^\infty \Phi_j u_{t-j}.
\]

We first look at $A_n$. By the weak law of large numbers, $\frac{1}{n} \sum_{t=1}^n |\varepsilon_t|^a \overset{p}{\to} \mathbb{E} |\varepsilon_t|^a$. As by Assumption 4.2 $\mathbb{E} |\varepsilon_t|^a = O(1)$, we have that $A_n = O_p(1)$.

\footnote{Every convex function $f(x)$ has the property that $f(\sum_{i=1}^k x_i/k) \leq \sum_{i=1}^k f(x_i)/k$. Applying this to the function $f(x) = |x|^a$, we have

\[
\left| \sum_{i=1}^k x_i \right|^a / k \leq k^a \sum_{i=1}^k |x_i|^a / k = k^{a-1} \sum_{i=1}^k |x_i|^a.
\]}

116
For $B_n$, we wish to show that $E(|\varepsilon_{q,t} - \varepsilon_{t}|^a) = o(q^{-a})$ as $\frac{1}{n} \sum_{t=1}^{n} |\varepsilon_{q,t} - \varepsilon_{t}|^a \rightarrow E|\varepsilon_{q,t} - \varepsilon_{t}|^a$. Using Minkowski’s inequality we have

$$E|\varepsilon_{q,t} - \varepsilon_{t}|^a = E \left( \sum_{j=q+1}^{\infty} \Phi_j u_{t-j} \right)^a \leq \left( \sum_{j=q+1}^{\infty} (E|\Phi_j u_{t-j}|^a)^{1/a} \right)^a \leq \left( \sum_{j=q+1}^{\infty} |\Phi_j| (E|u_{t-j}|)^{1/a} \right)^a = \left( (E|u_t|^a)^{1/a} \sum_{j=q+1}^{\infty} |\Phi_j| \right)^a = E|u_t|^a \left( \sum_{j=q+1}^{\infty} |\Phi_j| \right)^a = o(q^{-a}).$$

The final step comes from Bühlmann (1995), where it is shown in Lemma 2.1 that Assumption 4.2 implies that $\sum_{j=q+1}^{\infty} |\Phi_j| < \infty$. It is also shown (in the proof of Theorem 3.1) that $\sum_{j=q+1}^{\infty} |\Phi_j| = o(1)$ if $\sum_{j=0}^{\infty} |\Phi_j| < \infty$. Consequently $\sum_{j=q+1}^{\infty} |\Phi_j| = o(q^{-1})$ as $q \sum_{j=q+1}^{\infty} |\Phi_j| \leq \sum_{j=q+1}^{\infty} |\Phi_j|$. Next we turn to $C_n$. We can write

$$\hat{\varepsilon}_{q,t} = u_t - \sum_{j=1}^{q} \hat{\Phi}_j u_{t-j} = \varepsilon_{q,t} + \sum_{j=1}^{q} \hat{\Phi}_j u_{t-j} - \sum_{j=1}^{q} \Phi_j u_{t-j}$$

(4.18) where $\Phi_{q,j}$ is defined as the coefficient of $y_{t-j}$ in the best linear predictor of $y_t$ in terms of $y_{t-1}, \ldots, y_{t-q}$. Then

$$|\hat{\varepsilon}_{q,t} - \varepsilon_{q,t}|^a \leq 2^{a-1} \left( \sum_{j=1}^{q} \left| \hat{\Phi}_j - \Phi_{q,j} \right| u_{t-j} \right)^a + \sum_{j=1}^{q} \left| \Phi_{q,j} - \Phi_j \right| u_{t-j} \right)^a.$$

We define

$$C_{1n} = \frac{1}{n} \sum_{t=1}^{n} \left| \sum_{j=1}^{q} \left( \hat{\Phi}_j - \Phi_{q,j} \right) u_{t-j} \right|^a, \quad C_{2n} = \frac{1}{n} \sum_{t=1}^{n} \left| \sum_{j=1}^{q} \left( \Phi_{q,j} - \Phi_j \right) u_{t-j} \right|^a$$

and show that $C_{1n}, C_{2n} = o_p(1)$. Then we have that

$$C_{1n} = \frac{1}{n} \sum_{t=1}^{n} \left| \sum_{j=1}^{q} \left( \hat{\Phi}_j - \Phi_{q,j} \right) u_{t-j} \right|^a \leq q^{a-1} \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{q} \left| \Phi_j - \Phi_{q,j} \right|^a |u_{t-j}|^a \leq q^{a-1} \left( \max_{1 \leq j \leq q} \left| \Phi_j - \Phi_{q,j} \right|^a \right)^{1/a} \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{q} |u_{t-j}|^a,$$

As every value of $|u_{t-j}|^a$ for $j = 1 - q, \ldots, n-1$ occurs at most $q$ times in the double sum $\sum_{t=1}^{n} \sum_{j=1}^{q} |u_{t-j}|^a$, we have that

$$C_{1n} \leq q^{a-1} \left( \max_{1 \leq j \leq q} \left| \Phi_j - \Phi_{q,j} \right|^a \right)^{1/a} \frac{1}{n} \sum_{t=1}^{n-1} \sum_{j=1}^{q} |u_{t-j}|^a \leq q^{a} \left( \max_{1 \leq j \leq q} \left| \Phi_j - \Phi_{q,j} \right|^a \right)^{1/a} \left( \frac{1}{n} \sum_{t=0}^{n-1} |u_t|^a + \frac{1-q}{n} |u_t|^a \right) = o_p((\ln n/n)^{a/2}) \left( (a'/n)O_p(n) = O_p((\ln n/n)^{a/2}),

117
where we use that
\[
\max_{1 \leq j \leq q} |\hat{\Phi}_j - \Phi_{q,j}| = O_p((\ln n/n)^{1/2}) \tag{4.19}
\]
uniformly in \( q < Q_n \), where \( Q_n = o((n/\ln n)^{1/2}) \), from Hannan and Kavalieris (1986). Note that while Hannan and Kavalieris (1986) show their result for the Yule-Walker estimator, (4.19) is valid for OLS as well by Theorem 1 of Poskitt (1994). To conclude this part of the proof, note that \( C_{1n} = o_p(1) \) as \( q = o((n/\ln n)^{1/2}) \).

For \( C_{2n} \), note that by Markov’s inequality for any \( \epsilon > 0 \)
\[
P(|C_{2n}| > \epsilon) \leq \epsilon^{-1} \mathbb{E}\left[ \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j)u_{t-j} \right]^{a}
\]
Then, using Minkowski’s inequality and the stationarity of \( u_t \), we have
\[
\mathbb{E}\left[ \sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j)u_{t-j} \right]^{a} \leq \left( \sum_{j=1}^{q} (\mathbb{E}|(\Phi_{q,j} - \Phi_j)u_{t-j}|^{a})^{1/a} \right)^{a}
\leq \left( \mathbb{E}|u_t|^{a} \sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| \right)^{a}
= \mathbb{E}|u_t|^{a} \left( \sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| \right)^{a}.
\]

Again from Bühlmann (1995, p. 337), we have that
\[
\sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| \leq c \sum_{j=q+1}^{\infty} |\Phi_j| = o(q^{-a}) \tag{4.20}
\]
with \( c \) some constant. Hence, \( C_{2n} = o(q^{-a}) \) which completes the proof for \( C_n \).

Finally, we look at \( D_n \). We want to show that
\[
\frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{q,t} = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_{q,t} + o_p(1) = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t + o_p(1). \tag{4.17}
\]

Using equations (4.17) and (4.18) we can write
\[
\hat{\varepsilon}_{q,t} = \varepsilon_t + \sum_{j=q+1}^{\infty} \Phi_j u_{t-j} - \sum_{j=1}^{q} (\Phi_j - \Phi_{q,j}) u_{t-j} - \sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j) u_{t-j}.
\]

Hence, what we need to show is that
\[
\frac{1}{n} \sum_{t=1}^{n} \sum_{j=q+1}^{\infty} \Phi_j u_{t-j} \xrightarrow{p} 0 \tag{4.21}
\]
\[
\frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j) u_{t-j} \xrightarrow{p} 0 \tag{4.22}
\]
\[
\frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{q} (\Phi_j - \Phi_{q,j}) u_{t-j} \xrightarrow{p} 0. \tag{4.23}
\]
Note that, using Markov’s inequality, we have for any $\varepsilon > 0$
\[
P\left(\frac{1}{n} \sum_{t=1}^{n} \sum_{j=q+1}^{\infty} \Phi_{j} u_{t,j} > \varepsilon\right) \leq \varepsilon^{-a} E \left|\frac{1}{n} \sum_{t=1}^{n} \sum_{j=q+1}^{\infty} \Phi_{j} u_{t,j}\right|^a = o(q^{-a})
\]
for any $2 < a \leq 4$ which follows from the proof for $B_n$. This shows (4.21). To show (4.22), we can use the proof of $C_{2n}$ to show that
\[
P\left(\frac{1}{n} \sum_{t=1}^{n} \sum_{j=q+1}^{\infty} (\Phi_{q,j} - \Phi_{q,j}) u_{t,j} > \varepsilon\right) \leq \varepsilon^{-a} E \left|\frac{1}{n} \sum_{t=1}^{n} \sum_{j=q+1}^{\infty} (\Phi_{q,j} - \Phi_{q,j}) u_{t,j}\right|^a = o(q^{-a}).
\]
Finally, to prove (4.23), note that
\[
\left|\frac{1}{n} \sum_{t=1}^{n} \sum_{j=q+1}^{\infty} (\Phi_{q,j} - \Phi_{q,j}) u_{t,j}\right| \leq q \left(\max_{1 \leq j \leq q} |\Phi_{j} - \Phi_{q,j}|\right) \frac{1}{n} \left(\sum_{t=0}^{n-1} |u_t| + \sum_{t=-1}^{1-q} |u_t|\right)
\]
which follows exactly as in the proof of $C_{1n}$. This shows that $D_n = o_p(1)$, and the proof is complete.

Before proceeding with the proof of Theorem 4.1, we need one additional lemma to ensure that the covariance matrix of the bootstrap errors correctly mimics that of the original errors.

**Lemma 4.A.2.** Under Assumptions 4.2 and 4.4 we have that
\[
\Sigma^* = E^*(\varepsilon_t^* \varepsilon_t^*) = \Sigma + o_p(1).
\]

**Proof of Lemma 4.A.2.** This proof follows Paparoditis (1996, Proof of Theorem 2.5, p. 288). First note that $E^*(\varepsilon_t^* \varepsilon_t^*) = n^{-1} \sum_{t=1}^{n} \tilde{\varepsilon}_{q,t} \tilde{\varepsilon}_{q,t}$. Then
\[
|\Sigma^* - \Sigma| = \left|n^{-1} \sum_{t=1}^{n} \tilde{\varepsilon}_{q,t} \tilde{\varepsilon}_{q,t} - \Sigma\right| = \left|n^{-1} \sum_{t=1}^{n} (\tilde{\varepsilon}_{q,t} \tilde{\varepsilon}_{q,t} - \varepsilon_t \varepsilon_t)\right| + \left|n^{-1} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t - \Sigma\right|
\]
\[
= n^{-1} \sum_{t=1}^{n} |\tilde{\varepsilon}_{q,t} \tilde{\varepsilon}_{q,t} - \varepsilon_t \varepsilon_t| + o_p(1)
\]
\[
= n^{-1} \sum_{t=1}^{n} (|\tilde{\varepsilon}_{q,t} | - |\varepsilon_t|) |\tilde{\varepsilon}_{q,t}| + o_p(1)
\]
\[
\leq n^{-1} \sum_{t=1}^{n} |\tilde{\varepsilon}_{q,t} | n^{-1} \sum_{t=1}^{n} |\tilde{\varepsilon}_{q,t} | + n^{-1} \sum_{t=1}^{n} |\varepsilon_t| |\tilde{\varepsilon}_{q,t} | + o_p(1)
\]
\[
\leq \max_{1 \leq t \leq n} |\tilde{\varepsilon}_{q,t}| n^{-1} \sum_{t=1}^{n} |\tilde{\varepsilon}_{q,t} | + \max_{1 \leq t \leq n} |\varepsilon_t| n^{-1} \sum_{t=1}^{n} |\tilde{\varepsilon}_{q,t} | + o_p(1),
\]
as $n^{-1} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t - \Sigma$ $p$–$0$ by the weak law of large numbers.

Note that
\[
|\tilde{\varepsilon}_{q,t}| \leq |\tilde{\varepsilon}_{q,t}| + n^{-1} \sum_{t=1}^{n} |\tilde{\varepsilon}_t| \leq |\tilde{\varepsilon}_{q,t} - \varepsilon_t| + |\varepsilon_t - \varepsilon_t| + |\varepsilon_t| + |n^{-1} \sum_{t=1}^{n} \tilde{\varepsilon}_{q,t}|
\]

119
and
\[
|\tilde{\varepsilon}_{q,t} - \varepsilon_t| = \left| \tilde{\varepsilon}_{q,t} - n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_{q,t} - \varepsilon_t \right| \leq |\tilde{\varepsilon}_{tq} - \varepsilon_t| + |\varepsilon_{q,t} - \varepsilon_t| + n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_i.
\]

It now follows that \(|\Sigma^* - \Sigma| = o_p(1)\) by applying the methods from the proof of Lemma 4.A.1 (with \(a = 1\)).

**Proof of Theorem 4.1.** In this proof we draw heavily on results by Einmahl (1987), as in Chang et al. (2006). Therefore, we first need to introduce notation used by Einmahl (1987). Let \(\mathbb{R}^d\) denote the \(d\)-dimensional Euclidean space. Let \(C_d[0,1]\) be the space of all continuous \(\mathbb{K}\)-valued functions on \([0,1]\) endowed with the sup-norm \(|| \cdot ||\).

Let \(\lambda(Q_1, Q_2, \delta)\) denote the \(\delta\)-distance of two measures \(Q_1\) and \(Q_2\), that is
\[
\lambda(Q_1, Q_2, \delta) = \sup\{Q_1(A) - Q_2(A') : A \subseteq C_d[0,1] \text{ closed}\}
\]
where \(A' = \{g \in C_d[0,1] : 3f \in A, ||f - g|| < \delta\}\). Then Einmahl (1987) shows that
\[
\lambda(W^*_n, W, \delta) \leq c\delta^{-a} K_{an},
\]
holds for all \(\delta\) if \(2 < a < 4\), and \(\delta > K_{an}^*\) if \(a \geq 4\); where \(K_{an}^* = \sum_{i=1}^{n} E^* |\varepsilon_i|^a\), \(c\) is a positive constant depending only on \(a, l\) and \(\gamma\), and \(0 < \gamma < 1/(2a - 4)\).

By noting that \(K_{an}^* = \sum_{i=1}^{n} E^* |\varepsilon_i|^a = n E^* |\varepsilon|^a\), we can, as in Chang et al. (2006), transform this into the following condition:
\[
P^* \left\{ \sup_{0 \leq r \leq 1} |W^*_n(r) - W(r)| > n^{-1/2} c_n \right\} \leq K n a^{-a} E^* |\varepsilon|^a
\]
for any sequence \(\{c_n\}\), \(c_n = n^{1/a + \delta_2}\) for any \(\delta_2 > 0\), where \(K\) is an absolute constant depending only on \(a, l\).

Once we have the result in (4.25), we can take \(0 < \delta_2 < 1/2 - 1/a\), or alternatively, \(\delta_2 = 1/2 - 1/a - \epsilon\), where \(0 < \epsilon < 1/2 - 1/a\). Then on the left-hand side we have \(n^{-1/2} c_n = n^{-1/2} (1/2 - 1/a + \epsilon) = n^{-1/2} \epsilon^{-1/2} = n^{-1/2} \epsilon^{-1/2}\), to show that
\[
P^* \left\{ \sup_{0 \leq r \leq 1} |W^*_n(r) - W(r)| > n^{-1} \epsilon^{-1/2} \right\} \leq K n^{-1/(1 + \delta_2)} E^* |\varepsilon|^a,
\]
from which we can deduce that, as \(n \to \infty\),
\[
\sup_{0 \leq r \leq 1} |W^*_n(r) - W(r)| = o_p(1).
\]

**Proof of Theorem 4.2.** Using the Beveridge-Nelson decomposition, we can write
\[
\tilde{\varepsilon}_{t} = u_t^* - \sum_{j=1}^{q} \Phi_j u_{t-j}^*
\]
\[
= (I - \sum_{j=1}^{q} \Phi_j)u_t^* + \sum_{i=1}^{q} \sum_{j=1}^{q} \Phi_j (u_{t-i+1}^* - u_{t-i}^*)
\]
\[
= \Phi(1)u_t^* - \sum_{i=1}^{q} \sum_{j=1}^{q} \Phi_j (u_{t-i}^* - u_{t-i+1}^*)
\]
and hence

\[ u^*_t = \Psi(1)\varepsilon_t^* + \Psi(1) \sum_{j=1}^{q} \left( \sum_{j=1}^{q} \Phi_j \right) (u^*_{t-i} - u^*_{t-i+1}) = \Psi(1)\varepsilon_t^* + (\bar{\Phi}_{1} - \bar{u}_t^*), \]

where \( \bar{\Phi}_{1} = \Psi(1) \sum_{j=1}^{q} (\sum_{j=q+1}^{r} \Phi_j) u^*_{t-i} \) and \( \Psi(1) = \Phi(1)^{-1} \). Then

\[ B_n^*(r) = n^{-1/2} \sum_{t=1}^{[nr]} u^*_t = n^{-1/2} \sum_{t=1}^{[nr]} \Psi(1)\varepsilon_t^* + n^{-1/2} \sum_{t=1}^{[nr]} (\bar{u}_{t-1}^* - \bar{u}_t^*) \]

\[ = \Psi(1) W_n^*(r) + n^{-1/2} (\bar{u}_0^* - \bar{u}_{[nr]}) \]

Hence, we need to show that

\[ \Phi(1) \rightsquigarrow \Phi(1) \]  \hspace{1cm} (4.26)

\[ P^* \left\{ \max_{1 \leq i \leq n} |n^{-1/2} \bar{u}_i^*| > \epsilon \right\} = o_p(1) \]  \hspace{1cm} (4.27)

We can follow Chang et al. (2006, Proof of Theorem 3.3) for the proofs of these results.

We first show (4.26). Using equations (4.19) and (4.20) we have that

\[ \left| \hat{\Phi}(1) - \Phi(1) \right| \leq \sum_{j=1}^{q} |\hat{\Phi}_j - \Phi_{q,j}| + \sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| + \sum_{j=q+1}^{\infty} |\Phi_j| = O_p(q \sqrt{n} (\ln n/n)^{1/2}) + o(q^{-1}). \]

Hence \( \hat{\Phi}(1) = \Phi(1) + o_p(1) \). This proves (4.26).

To prove (4.27), we have as in Park (2002)

\[ P^* \left\{ \max_{1 \leq i \leq n} |n^{-1/2} \bar{u}_i^*| > \epsilon \right\} \leq n P^* \left\{ |n^{-1/2} \bar{u}_t^*| > \epsilon \right\} \leq (1/\epsilon^a)n^{1-a/2} E^* |\bar{u}_t^*|^{\alpha} \]

The second inequality follows from Markov’s inequality. Hence, we have to show that

\[ n^{1-a/2} E^* |\bar{u}_t^*|^{\alpha} = o_p(1), \] \hspace{1cm} (4.28)

which for \( a > 2 \) implies that we have to show that \( E^* |\bar{u}_t^*|^{\alpha} = O_p(1) \). If the Yule-Walker method is used to estimate (4.8), the estimated autoregression is always invertible. Although invertibility of the estimated autoregression is not guaranteed for finite samples using OLS, the asymptotic equivalence of OLS to Yule-Walker (Poskitt, 1994, Theorem 1) implies that for large \( n \) we can write \( u^*_t = \sum_{j=0}^{\infty} \psi_j \varepsilon^*_{t-j} \) and furthermore \( \bar{u}_t^* = \sum_{j=0}^{\infty} \psi_j \varepsilon^*_{t-j} \), where \( \psi_j = \sum_{j=0}^{\infty} \psi_j \). Let \( \bar{u}_{(k),t}^* \) be the \( k \)-th element of \( \bar{u}_t^* \) and let \( \bar{\psi}_{(k),j} \) be the \( k \)-th row of \( \bar{\psi}_j \). By successive application of the Marcinkiewicz-Zygmund
for some constant $c_a$ not depending on $n$. Phillips and Solo (1992, p. 973) show that a sufficient condition for $\sum_{j=1}^{\infty} |\Psi_j|^2 = O_p(1)$ is
\begin{equation}
\sum_{j=1}^{\infty} j^{1/2} |\Psi_j| = O_p(1).
\end{equation}

This will in turn hold if (Hannan and Kavalieris, 1986)
\begin{equation}
\sum_{j=1}^{q} j^{1/2} |\Phi_j| = O_p(1).
\end{equation}

We have
\begin{align*}
\sum_{j=1}^{q} j^{1/2} |\Phi_j| &= \sum_{j=1}^{q} j^{1/2} |\Phi_j - \Phi_{q,j} + \Phi_{q,j} - \Phi_j + \Phi_j| \\
&\leq \sum_{j=1}^{q} j^{1/2} |\Phi_j - \Phi_{q,j}| + \sum_{j=1}^{q} j^{1/2} |\Phi_{q,j} - \Phi_j| + \sum_{j=1}^{q} j^{1/2} |\Phi_j| \\
&\leq q^{1/2} \sum_{j=1}^{q} |\Phi_j - \Phi_{q,j}| + q^{1/2} \sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| + \sum_{j=1}^{q} j^{1/2} |\Phi_j| \\
&\leq q^{3/2} \max_{1 \leq j \leq q} |\Phi_j - \Phi_{q,j}| + q^{1/2} \sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| + \sum_{j=1}^{q} j^{1/2} |\Phi_j| \\
&= O_p(q^{3/2}(\ln n/n)^{1/2}) + o(q^{-1/2}) + O(1) = O_p(1),
\end{align*}
by (4.19), (4.20) and Assumption 4.4'. Together with Lemma 4.A.1 this shows that
\begin{equation}
E^* |\bar{\mathbf{u}}_t|^n = O_p(1).
\end{equation}

This concludes the proof of this theorem.
Next we need several lemmas in order to show the limiting distribution of the bootstrap test statistic.

**Lemma 4.A.3.** Let \( \xi^*_t \) be the bootstrap equivalent of \( \xi \) defined in equation (4.5), i.e.

\[
y^*_t = \pi^*_0 \Delta x^*_t + \sum_{j=1}^{\infty} \pi^*_j \Delta z^*_{t-j} + \xi^*_t.
\]  
(4.32)

Then, if Assumptions 4.2 and 4.4 hold,

\[
n^{-1/2} \sum_{t=1}^{[nr]} \xi^*_t \overset{d}{\to} B_\xi(r),
\]

where \( B_\xi(r) \) is a scalar Brownian motion with variance \( \omega^2 \), i.e

\[
B_\xi(r) = \omega W_1(r),
\]

where \( W_1(r) \) is the first element of the standard Brownian motion \( W(r) \).

**Proof of Lemma 4.A.3.** Follows immediately from Theorem 4.1.

**Lemma 4.A.4.** Let \( f^* \) denote the spectral density and \( \Gamma^*(k) \) the autocovariance function of \( u^*_t \). Under Assumptions 4.2 and 4.4,

\[
\sup_{\lambda} |f^*(\lambda) - f(\lambda)| = o_p(1)
\]  
(4.33)

and

\[
\sum_{k=-\infty}^{\infty} \Gamma^*(k) = \sum_{k=-\infty}^{\infty} \Gamma(k) + o_p(1).
\]  
(4.34)

**Proof of Lemma 4.A.4.** The spectral density \( f^*(\lambda) \) of \( u^*_t \) is

\[
f^*(\lambda) = \frac{1}{2\pi} (I - \sum_{j=1}^{q} \hat{\Phi}_j e^{-i j \lambda})^{-1} \Sigma^* (I - \sum_{j=1}^{q} \hat{\Phi}_j e^{i j \lambda})^{-1}.
\]

Note that by Lemma 4.A.2 \( \Sigma^* \overset{d}{=} \Sigma \). Furthermore,

\[
\sum_{j=1}^{q} \left| \hat{\Phi}_j - \Phi_j \right| e^{-i j \lambda} \leq \sum_{j=1}^{q} \left| \hat{\Phi}_j - \Phi_{q,j} \right| e^{-i j \nu} + \sum_{j=1}^{q} \left| \Phi_{q,j} - \Phi_j \right| e^{-i j \nu} \leq q \max_{1 \leq j \leq q} \left| \hat{\Phi}_j - \Phi_{q,j} \right| + \sum_{j=1}^{q} \left| \Phi_{q,j} - \Phi_j \right| = o_p(1)
\]

by (4.19) and (4.20). Now the result in (4.33) follows straightforwardly.

The result in (4.34) follows trivially by noting that \( \sum_{k=-\infty}^{\infty} \Gamma(k) = 2\pi f(0) \) and correspondingly \( \sum_{k=-\infty}^{\infty} \Gamma^*(k) = 2\pi f^*(0) \).
4 Sieve Bootstrap ECM Cointegration Test

Lemma 4.A.5. Under Assumptions 4.2, 4.3, 4.4’ and 4.5 we have

\[ a) \ n^{-2}Z^*_1 Z^*_1 = n^{-2} \sum_{t=1}^{n} z^*_{t-1} z^*_{t-1} \frac{d^*}{d^*} \int_0^1 B(r)B(r)^t dr \] (4.35)

\[ b) \ n^{-1}Z^*_1 \xi^*_r = n^{-1} \sum_{t=1}^{n} z^*_{t-1} \xi^*_r \frac{d^*}{d^*} \int_0^1 B(r)dB(r) \] (4.36)

\[ c) \ |(n^{-1}W^*_{p,r}W^*_{p,r})^{-1}| = \left( \frac{1}{n} \sum_{t=1}^{n} w^*_{p,r} w^*_{p,r} \right)^{-1} = O_p(1) \] (4.37)

\[ d) \ |Z^*_1 W^*_{p,r}| = \sum_{t=1}^{n} z^*_{t-1} w^*_{p,r} = O_p(np^{1/2}) \] (4.38)

\[ e) \ |W^*_{p,r}| = \sum_{t=1}^{n} w^*_{p,r} = O_p(np^{1/2}) \] (4.39)

Proof of Lemma 4.A.5. First we look at a). As we set \( z^*_0 = 0 \), we have

\[ z^*_1 = \sum_{i=1}^{1} u^*_i \]

and therefore

\[ B^*_r = n^{-1/2} z^*_1. \]

Then by Theorem 4.2 and the continuous mapping theorem we have

\[ n^{-2} \sum_{t=1}^{n} z^*_{t-1} z^*_{t-1} = n^{-1} \sum_{t=1}^{t/n} z^*_1 z^*_1 \frac{1}{t/n} \int_0^{t/n} B(r)B(r)^t dr \]

\[ = \int_0^1 B^*_r B^*_r dr \rightarrow \int_0^1 B(r)B(r)^t dr \]

as in Chang et al. (2006, Proof of Lemma 3.4).

Next we look at b). We have

\[ |n^{-1} \sum_{t=1}^{n} z^*_{t-1} \xi^*_r| \leq |n^{-1} \sum_{t=1}^{n} z^*_{t-1} \xi^*_r| + |n^{-1} \sum_{t=1}^{n} z^*_{t-1} (\xi^*_r - \xi^*_r)|. \]

Hence, we first have to show that \( n^{-1} \sum_{t=1}^{n} z^*_{t-1} (\xi^*_r - \xi^*_r) = o_p(1) \). We can follow Chang et al. (2006, Proof of Lemma A.6) for the proof.

Note that \( \xi^*_r = \sum_{k=p+1}^{q} \pi^*_k u^*_{r-k} + \xi^*_r \), where

\[ \pi^*_k = \Phi_{1,k} - \Sigma_{12} \Sigma_{22}^{-1} \Phi_{2,k} \]

and \( \Phi_k = (\Phi_{1,k}^*, \Phi_{2,k}^*)^t \). As \( \Phi_k = 0 \) for \( k > q \) and using Assumption 4.5, we have that

\[ \sum_{j=p+1}^{q} \pi^*_j = \sum_{k=p+1}^{q} \pi^*_k = o_p(1). \]
Then define $\hat{\Psi}_{p,j}$ such that

$$\xi_{p,t}^* - \xi_{p,t}^* = \sum_{k=p+1}^{\infty} \pi_k^* u_{i-k}^* = \sum_{j=p+1}^{\infty} \sum_{k=p+1}^{\infty} \pi_k^* \hat{\Psi}_{j-k} \bar{\xi}_{i-j}^* = \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} \bar{\xi}_{i-j}^*, \quad (4.41)$$

We then have that

$$\sum_{j=p+1}^{\infty} |\hat{\Psi}_{p,j}| \leq \sum_{j=p+1}^{\infty} \sum_{i=0}^1 |\hat{\Psi}_t| \leq \left( \sum_{j=p+1}^{\infty} |\pi_j^*| \right) (\sum_{j=p+1}^{\infty} |\pi_j^*|) O_p(1).$$

Define $\eta_t^* = \sum_{i=1}^t \bar{\varepsilon}_i^*$, such that we can write

$$z_t^* = \hat{\Psi}(1)\eta_t^* + (\bar{u}_0 - \bar{u}_t^*).$$

Then we have

$$\sum_{t=1}^n z_{t-1}^* (\xi_{p,t}^* - \xi_t^*)' \leq \sum_{t=1}^n \hat{\Psi}_t (1) \eta_{t-1}^* (\xi_{p,t}^* - \xi_t^*)' + \sum_{t=1}^n \bar{u}_0^* (\xi_{p,t}^* - \xi_t^*)' - \sum_{t=1}^n \bar{u}_{t-1}^* (\xi_{p,t}^* - \xi_t^*)' = R_{1n}^* + R_{2n}^* + R_{3n}^*.$$

We first want to show that $R_{1n}^* = o_p(n)$. Let $\delta_{ij}$ be the Kronecker delta. We have

$$\left| \sum_{t=1}^n \eta_{t-1}^* (\xi_{p,t}^* - \xi_t^*)' \right| = \sum_{t=1}^n \left| \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} \bar{\varepsilon}_{t-j}^* \right| \leq \sum_{j=p+1}^{\infty} \left| \sum_{t=1}^n \left( (n-j) \sum_{j=p+1}^{\infty} \xi_{t-i}^* - \delta_{ij} \xi_t^* \right) \hat{\Psi}_{p,j} \right| = \sum_{j=p+1}^{\infty} \left| \sum_{t=1}^n \hat{\Psi}_{p,j} O_p^*(n) + \sum_{j=p+1}^{\infty} |\hat{\Psi}_{p,j}| O_p^*(n^{1/2}) \right| \leq \left( \sum_{j=p+1}^{\infty} \sum_{t=1}^n |\pi_j^*| \right) O_p^*(n) = o_p^*(n).$$

Next we turn to $R_{2n}^*$. We have

$$\sum_{t=1}^n (\xi_{p,t}^* - \xi_t^*) = \sum_{t=1}^n \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} \bar{\varepsilon}_{t-j}^* = \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} \sum_{t=1}^n \bar{\varepsilon}_{t-j}^* = \left( \sum_{j=p+1}^{\infty} |\hat{\Psi}_{p,j}| \right) O_p^*(n^{1/2}) = \left( \sum_{j=p+1}^{\infty} \sum_{t=1}^n |\pi_j^*| \right) O_p^*(n^{1/2}),$$

from which we can easily see that $R_{2n}^* = o_p^*(n^{1/2})$.

Finally we look at $R_{3n}^*$. By applying the Beveridge-Nelson decomposition in a slightly different way than before, we can derive that $\bar{u}_t^* = \sum_{j=0}^{\infty} \sum_{t=j+1}^{\infty} \hat{\Psi}_t \bar{\xi}_{t-j-i}$. Then, using
Therefore, following Chang and Park (2003, Proof of Lemma 3) we first want to show that
\[ \sum_{t=1}^{n} u_{t-1}^{\ast} (\xi_{p,t}^{\ast} - \xi^{\ast}) = \sum_{t=1}^{n} u_{t-1} \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} \varepsilon_{t-j}^{\ast} \]
\[ = \sum_{t=1}^{n} \sum_{i=0}^{\infty} \sum_{k=1}^{\infty} \hat{\Psi}_{i} \sum_{t=1}^{\infty} \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} \varepsilon_{t-j}^{\ast} \]
\[ = \sum_{t=1}^{n} \sum_{i=0}^{\infty} \sum_{k=1}^{\infty} \hat{\Psi}_{i} \sum_{j=p+1}^{\infty} (\varepsilon_{t-i-1}^{\ast} - \delta_{i+1,j} \Sigma) \hat{\Psi}_{p,j} \]
\[ = \sum_{j=p+1}^{\infty} |\hat{\Psi}_{p,j}| O_{p}^{\ast}(n) + \sum_{i=0}^{\infty} \sum_{k=1}^{\infty} \sum_{j=p+1}^{\infty} |\hat{\Psi}_{j}| O_{p}^{\ast}(n^{1/2}) \]
\[ = \sum_{j=p+1}^{\infty} |\hat{\Psi}_{p,j}| O_{p}^{\ast}(n) + \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} O_{p}^{\ast}(n^{1/2}) \]
\[ = \left( \sum_{j=p+1}^{\infty} |\pi_{j}^{\ast}| \right) O_{p}^{\ast}(n). \]

Therefore, \( R_{1n}^{\ast} = o_{p}^{\ast}(n) \), and hence
\[ n^{-1} \sum_{t=1}^{n} z_{t-1}^{\ast} (\xi_{p,t}^{\ast} - \xi^{\ast}) = n^{-1} (R_{1n}^{\ast} + R_{2n}^{\ast} + R_{3n}^{\ast}) = o_{p}^{\ast}(1). \]

Then
\[ n^{-1} \sum_{t=1}^{n} z_{t-1}^{\ast} \xi_{p,t}^{\ast} = n^{-1} \sum_{t=1}^{n} z_{t-1}^{\ast} \xi^{\ast} + o_{p}^{\ast}(1), \]
while by Park and Phillips (1989, Lemma 2.1), Theorem 4.2 and Lemma 4.A.3, we have that
\[ n^{-1} \sum_{t=1}^{n} z_{t-1}^{\ast} \xi^{\ast} \xrightarrow{d} \int_{0}^{1} B(r) dB_{C}(r). \]

This completes the proof of part b).

For c), we want to show that
\[ E^{*} \left( \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^{\ast} w_{p,t}^{\ast'} \right)^{-1} = O_{p}(1). \] (4.42)

Following Chang and Park (2003, Proof of Lemma 3) we first want to show that
\[ E^{*} \left| \sum_{t=1}^{n} [u_{t-1}^{\ast} u_{t-j}^{\ast} - \Gamma^{\ast}(i-j)] \right|^{2} = O_{p}(n). \] (4.43)
For this to hold it is sufficient to show that
\[ E^* \left( \sum_{t=1}^{n} [u_{t(a',t)} u_{t(b',t)} - \Gamma_{(ab)}(i-j)] \right)^2 = O_p(n), \tag{4.44} \]
for all \(1 \leq a, b \leq 1 + l\), where \(u_{t(a',t)}\) is the \(a\)-th element of \(u_{t}^*\), and similarly \(\Gamma_{(ab)}(i-j)\) is the \((a,b)\)-th element of \(\Gamma^*(i,j)\).

Analogous to the case for univariate time series models discussed in Berk (1974, eqs (2.10) and (2.11), p. 491), we have that
\[
E^* \left( \sum_{t=1}^{n} [u_{t(a',t)} u_{t(b',t)} - \Gamma_{(ab)}(i-j)] \right)^2 \leq 2n \sum_{k=-\infty}^{\infty} \Gamma_{(ab)}(k)^2 + n \sum_{c,d,e,f} |\kappa_{def}^*|,
\]
with \(\kappa_{def}^* = E^* (\varepsilon_{i,c}^* \varepsilon_{j,d}^* \varepsilon_{k,e}^* \varepsilon_{l,f}^*) - \sigma_{cd} \sigma_{ef} - \sigma_{ce} \sigma_{df} - \sigma_{cf} \sigma_{de} \) and \(\sigma_{cd} = E^* (\varepsilon_{i,c}^* \varepsilon_{j,d}^*)\). Note that \(|\kappa_{def}^*| = O_p(1)\) as \(E^* |\varepsilon|^4 = O_p(1)\) (take \(a = 4\) in Lemma 4.A.1). Furthermore, \(\sum_{k=-\infty}^{\infty} \Gamma_{(ab)}(k)^2 = O_p(1)\) through Lemma 4.A.4 and \(\sum_{k=0}^{\infty} \psi_{(ae),k}^2)^1/2 = O_p(1)\) as \(\sum_{k=0}^{\infty} k^{1/2} |\psi_k| = O_p(1)\), which we demonstrated in the proof of Theorem 4.2, equation (4.30). Now equation (4.44) follows straightforwardly.

Next, partition \(\Gamma^*(k)\) conformably with \(y_t\) and \(x_t\) as
\[
\Gamma^*(k) = \begin{bmatrix} \Gamma_{11}^*(k) & \Gamma_{12}^*(k) \\ \Gamma_{21}^*(k) & \Gamma_{22}^*(k) \end{bmatrix}
\]
and define \(\Gamma^*(k) = [\Gamma_{11}^*(k), \Gamma_{12}^*(k)]\) and \(\Gamma^*(k) = [\Gamma_{21}^*(k), \Gamma_{22}^*(k)]^t\). Then define \(\Omega_{pp}^*\) as
\[
\Omega_{pp}^* = \begin{bmatrix} \Gamma_{11}^*(0) & \Gamma_{11}^*(1) & \cdots & \Gamma_{11}^*(p-1) & \Gamma_{11}^*(p) \\ \Gamma_{12}^*(0) & \Gamma_{12}^*(1) & \cdots & \Gamma_{12}^*(p-1) & \Gamma_{12}^*(p) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Gamma_{21}^*(0) & \Gamma_{21}^*(1) & \cdots & \Gamma_{21}^*(p-1) & \Gamma_{21}^*(p) \\ \Gamma_{22}^*(0) & \Gamma_{22}^*(1) & \cdots & \Gamma_{22}^*(p-1) & \Gamma_{22}^*(p) \end{bmatrix}.
\]

As for any matrix \(M\), \(\|M\| \leq 1, \|M_i\|_i \leq \sum_{i,j} M_{ij} \leq \sum_{i,j} M_{ij}^2\), we can write
\[
E^* \left( \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{u_{t,p,t} u_{t,p,t}^* - \Omega_{pp}^*}{n} \right\| \right) \leq E^* \left( \left\| \frac{1}{n} \sum_{t=1}^{n} u_{t,p,t} u_{t,p,t}^* - \Gamma_{*p}^*(0) \right\| \right)^2
\]
\[
+ \sum_{j=1}^{p} \left\| \frac{1}{n} \sum_{t=1}^{n} u_{t,j,t} u_{t,j,t}^* - \Gamma_{*j}^*(0) \right\| \right)^2
\]
\[
+ \sum_{(i,j=1)}^{(p,p)} \left\| \frac{1}{n} \sum_{t=1}^{n} u_{t,i,t} u_{t,j,t} - \Gamma^*(j-i) \right\| \right)^2
\]
\[
= O_p(n^{-1}) + O_p(n^{-1}p) + O_p(n^{-1}p) + O_p(n^{-1}p^2) = O_p(n^{-1}p^2).
\]

Next we need to show that
\[
\|\Omega_{pp}^{-1}\| \leq 2\pi \left( \inf_{\Lambda} f^*(\Lambda) \right)^{-1} = O_p(1).
\]

\[\text{We let } M_{ij} \text{ denote submatrices into which one can partition } M.\]
Let us consider an “extended” $\Omega_{pp}^*$ matrix, i.e.

$$
\Omega_{pp}^* = \begin{bmatrix}
\Gamma_{11}^*(0) & \Gamma_{12}^*(0) & \Gamma_{1}^*(-1) & \ldots & \Gamma_{1}^*(-p) \\
\Gamma_{21}^*(0) & \Gamma_{22}^*(0) & \Gamma_{2}^*(-1) & \ldots & \Gamma_{2}^*(-p) \\
\Gamma_{1}^*(1) & \Omega_{pp} & \Gamma_{1}^*(2) & \ldots & \Gamma_{1}^*(p) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\Gamma_{1}^*(p) & \Gamma_{2}^*(p-1) & \ldots & \Gamma_{2}^*(0) & 1
\end{bmatrix}.
$$

Let $\lambda' = (\lambda_1^*, \ldots, \lambda_{(1+p)}^*)'$ be the eigenvalues of $\Omega_{pp}^*$, and define $0 < F_1^* = \inf_{\lambda'} ||f'(\lambda)||$. Then as a direct consequence of Lemma A.2 of Chang et al. (2006) we have that

$$
||\Omega_{pp}^{*-1}|| \leq (2\pi F_1^*)^{-1} = O_p(1).
$$

As $||\Omega_{pp}^{*-1}|| \leq ||\Omega_{pp}^{-1}||$, we know that $||\Omega_{pp}^{*-1}|| = O_p(1)$ as well. Then

$$
\left|\left| \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* \right| \right| = \left|\left| \Omega_{pp}^{*-1} \right| \right| + \left|\left| \Omega_{pp}^{*-1} \left( \Omega_{pp} - \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* \right) \left( \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* \right)^{-1} \Omega_{pp}^{*-1} \right| \right| \leq \left|\left| \Omega_{pp}^{*-1} \right| \right| + \left|\left| \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* - \Omega_{pp}^{*-1} \right| \right| \left|\left| \Omega_{pp}^{*-1} \right| \right| \left|\left| \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* - \Omega_{pp}^{*-1} \right| \right| \left|\left| \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* - \Omega_{pp}^{*-1} \right| \right|
$$

which implies that

$$
\left|\left| \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* \right| \right| \leq 1 - \left|\left| \Omega_{pp}^{*-1} \right| \right| \left|\left| \Omega_{pp}^{*-1} \right| \right| \left|\left| \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* - \Omega_{pp}^{*-1} \right| \right|
$$

holds for large $n$ with probability 1 as $\left|\left| \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* - \Omega_{pp}^{*-1} \right| \right| \left|\left| \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^* - \Omega_{pp}^{*-1} \right| \right| \leq O_p(n^{-1/2})$. As $\left|\left| \Omega_{pp}^{*-1} \right| \right| = O_p(1)$, the result in (4.42) follows.

For d) we want to show that

$$
E^* \sum_{t=1}^{n} z_{t-1}^* w_{p,t}^* = O_{p}(np^{1/2}).
$$

13Suppose we have a matrix $M$ and a vector $v$ that we can write as

$$
M = \begin{bmatrix}
M_{11} & M_{12} \\
M_{21} & M_{22}
\end{bmatrix}, \quad v = \begin{bmatrix}
v_1 \\
v_2
\end{bmatrix}.
$$

Then we have that

$$
||M||^2 = \max_{v} \frac{|Mv|^2}{|v|^2} = \max_{v_1, v_2} \frac{|M_{11}v_1 + M_{12}v_2|^2}{|v_1|^2} = \max_{v_1, v_2} \frac{|M_{11}v_1 + M_{12}v_2|^2}{|v_1|^2} + \frac{|M_{21}v_1 + M_{22}v_2|^2}{|v_2|^2} \geq \max_{v_1, v_2} \frac{|M_{11}v_1|^2 + |M_{12}v_2|^2}{|v_1|^2} + \frac{|M_{21}v_1|^2 + |M_{22}v_2|^2}{|v_2|^2} \geq \max_{v_2(v_1=0)} \frac{|M_{12}v_2|^2 + |M_{22}v_2|^2}{|v_2|^2} \geq \max_{v_2} \frac{|M_{22}|^2}{|v_2|^2} = ||M_{22}||^2.
$$
Following Chang and Park (2002, Proof of Lemma 3.2), we write
\[
\sum_{t=1}^{n} z_{t-1} u_{t-j} = \sum_{t=1}^{n-1} z_{t-1} u_{t-j} + \sum_{t=n-j+1}^{n} z_{t-1} u_{t-j} - \sum_{t=1}^{n-1} z_{t-1} u_{t-j} = \sum_{t=1}^{n} z_{t-1} u_{t-j} + R'_n
\]
and we want to show that \( R'_n = O_p(n) \) uniformly in \( j = 1, \ldots, p \).

First, write
\[
\sum_{t=1}^{n} z_{t-1} u_{t-j} = \sum_{t=1}^{n-1} z_{t-1} u_{t-j} + \sum_{t=n-j+1}^{n} z_{t-1} u_{t-j} = \sum_{t=1}^{n} z_{t-1} u_{t-j} + \sum_{t=n-j+1}^{n} z_{t-1} u_{t-j}
\]
(as \( u_t = 0 \) for \( t = 0, \ldots, p \)) and rewrite \( R'_n \) as
\[
\sum_{t=1}^{n} (z_{t-1} - z_{t-j-1}) u_{t-j} - \sum_{t=n-j+1}^{n} z_{t-1} u_{t-j} = R_{1n} - R_{2n}.
\]
Then
\[
R_{1n} = \sum_{t=1}^{n} (z_{t-1} - z_{t-j-1}) u_{t-j} = \sum_{t=1}^{n} \left( \sum_{t=1}^{j} u_{t-1} \right) u_{t-j} = \sum_{t=1}^{n} \sum_{t=1}^{j} u_{t-1} u_{t-j}
\]
\[
= n \sum_{i=1}^{j} \Gamma(i-j) + \sum_{i=1}^{j} \left( \sum_{t=1}^{n} (u_{t-1} - u_{t-j} - \Gamma(i-j)) \right)
\]
\[
= O_p(n) + O_p(n^{1/2})
\]
as \( 1 \leq j \leq p \), the result in equation (4.43), and the fact that \( \sum_{k=1}^{\infty} \Gamma(k) = O_p(1) \) by Assumption 4.2 and Lemma 4.4. We can write \( R_{2n} \) as
\[
R_{2n} = \sum_{t=1}^{n} \sum_{t=n-j+1}^{t-1} u_{t-1} u_{t-j} = \sum_{t=n-j+1}^{n-1} \left( \sum_{t=1}^{n-j} u_{t-1} + \sum_{t=n-j+1}^{t-1} u_{t-1} \right) u_{t-j}
\]
\[
= \sum_{t=n-j+1}^{n} \sum_{t=1}^{n-j} u_{t-1} u_{t-j} + \sum_{t=n-j+1}^{n} \sum_{t=n-j+1}^{t-1} u_{t-1} u_{t-j} = R_{2n} + R_{2n}^{\text{est}}.
\]
Then we have
\[
R_{2n}^{\text{est}} = j \sum_{i=1}^{n-j} \Gamma(i) + \sum_{i=1}^{n-j} \left[ \sum_{t=1}^{n-j} (u_{t-1} - \Gamma(i)) \right] = O_p(p) + O_p(n^{1/2})
\]
and
\[
R_{2n}^{\text{est}} = j \sum_{i=1}^{n-j} \Gamma(i) + \sum_{i=1}^{n-j} \left[ \sum_{t=n-j+1}^{t-1} (u_{t-1} - \Gamma(i)) \right] = O_p(p) + O_p(p^{3/2}),
\]
as
\[
\sum_{t=n-j+1}^{t-1} (u_{t-1} - \Gamma(i)) = \sum_{t=n-j+1}^{t-(n-j)-1} (u_{t-1} - \Gamma(i)) = \sum_{t=n-j+1}^{t-1} O_p((t-n+j)^{1/2}) = \sum_{t=n-j+1}^{t-1} O_p(j^{1/2})
\]
\[
= \sum_{t=n-j+1}^{t-1} O_p(p^{1/2}) = jO_p(p^{1/2}) = O_p(p^{3/2}).
\]
Hence,
\[ \sum_{t=1}^{n} z_{t-1}^{*} u_{t-j}^{*} = \sum_{t=1}^{n} z_{t-1}^{*} u_{t}^{*} + R_{n}^{*} = \sum_{t=1}^{n} z_{t-1}^{*} u_{t}^{*} + O_{p}(n) + O_{p}^{*}(n^{1/2}p). \]

Note that
\[ \left| \sum_{t=1}^{n} z_{t-1}^{*} u_{t}^{*} \right| = O_{p}^{*}(n), \]
by Phillips (1988), and
\[ \left| \sum_{t=1}^{n} z_{t-1}^{*} u_{t}^{*} \right| = \sum_{t=1}^{n} z_{t-1}^{*} u_{t}^{*} + O_{p}^{*}(n). \]

Then
\[
E^{*} \left| \sum_{t=1}^{n} z_{t-1}^{*} w_{p,t}^{*} \right| = E^{*} \left| \sum_{t=1}^{n} \left[ z_{t-1}^{*} u_{2,t}^{*} z_{t-1}^{*} u_{t-1}^{*} \ldots z_{t-1}^{*} u_{t-p}^{*} \right] \right|
\]
\[
= E^{*} \left( \sum_{t=1}^{n} \left| z_{t-1}^{*} u_{2,t}^{*} \right|^2 + \sum_{j=1}^{p} \sum_{t=1}^{n} \left| z_{t-1}^{*} u_{t-j}^{*} \right|^2 \right)^{1/2}
\]
\[
= E^{*} \left( \sum_{t=1}^{n} \left| z_{t-1}^{*} u_{2,t}^{*} \right|^2 + \sum_{j=1}^{p} \left( \sum_{t=1}^{n} \left| z_{t-1}^{*} u_{t-j}^{*} \right|^2 + O_{p}(n) + O_{p}^{*}(n^{1/2}p) \right)^{2} \right)^{1/2}
\]
\[
= E^{*} \left( \sum_{t=1}^{n} \left| z_{t-1}^{*} u_{2,t}^{*} \right|^2 + p \left( \sum_{t=1}^{n} \left| z_{t-1}^{*} u_{t}^{*} \right|^2 + O_{p}(n) + O_{p}^{*}(n^{1/2}p) \right)^{2} \right)^{1/2}
\]
\[
= O_{p}(np^{1/2}).
\]

Finally, we look at e). We want to show that
\[ \left| \sum_{t=1}^{n} w_{p,t}^{*} \xi_{p,t}^{*} \right| = O_{p}^{*}(n^{1/2}p^{1/2}). \]

Write
\[ \sum_{t=1}^{n} w_{p,t}^{*} \xi_{p,t}^{*} = \sum_{t=1}^{n} w_{p,t}^{*} \xi_{t}^{*} + \sum_{t=1}^{n} w_{p,t}^{*} (\xi_{p,t}^{*} - \xi_{t}^{*}) . \]

We first show that
\[ \sum_{t=1}^{n} u_{t-j}^{*} (\xi_{p,t}^{*} - \xi_{t}^{*}) = o_{p}^{*}(n^{1/2}) \]
uniformly in $1 \leq j \leq p$. We have that
\[
\sum_{t=1}^{n} u_{t-j}^* (\xi_{t,j}^* - \xi_t^*) = \sum_{t=1}^{n} u_{t-j}^* \left( \sum_{k=p+1}^{\infty} \Psi_{p,k} \xi_{t-k}^* \right)
= \sum_{t=1}^{n} \left( \sum_{m=0}^{\infty} \Psi_{m,t} \xi_{t-m}^* \right) \left( \sum_{k=p+1}^{\infty} \Psi_{p,k} \xi_{t-k}^* \right)
= \sum_{k=p+1}^{\infty} \sum_{t=1}^{n} \sum_{m=0}^{\infty} \Psi_{m,t} \xi_{t-m}^* \xi_{t-k}^* \Psi_{p,k} \xi_{t-k}^*
= n \sum_{k=p+1}^{\infty} \Psi_k \Sigma \Psi_{p,k} + \sum_{k=p+1}^{\infty} \sum_{m=0}^{\infty} \Psi_m \Psi_{k,m} \Psi_{p,k} \xi_{t-k}^* \xi_{t-k}^*
\leq \sum_{k=p+1}^{\infty} \sum_{m=0}^{\infty} \Psi_m \Psi_{k,m} \Psi_{p,k} \xi_{t-k}^* \xi_{t-k}^* = O_p(n^{1/2})
= a_p(n^{1/2}),
\]
as in Chang et al. (2006, Proof of Lemma A.6), such that
\[
\sum_{t=1}^{n} u_{t-j}^* \xi_{p,t}^* = \sum_{t=1}^{n} u_{t-j}^* \xi_t^* + O_p(n^{1/2}).
\]

Furthermore,
\[
E^* \left( \sum_{t=1}^{n} u_{t-j}^* \xi_t^* \right)^2 = E^* \left( \sum_{s=1}^{n} u_{s-j}^* \xi_t^* \right)' \left( \sum_{t=1}^{n} u_{t-j}^* \xi_t^* \right)
= \sum_{s=1}^{n} E^* u_{s-j}^* u_{s-j}^* \xi_t^* \xi_t^* = \sum_{s=1}^{n} E^* u_{s-j}^* u_{s-j}^* E^* \xi_t^* \xi_t^*
= \sum_{t=1}^{n} E^* u_{s-j}^* u_{s-j}^* E^* \xi_t^* \xi_t^* = O_p(n).
\]

Then
\[
\sum_{t=1}^{n} w_{p,t}^* \xi_{p,t}^* = \left( \sum_{j=1}^{p} \sum_{t=1}^{n} u_{t-j}^* \xi_{p,t}^* \right)^2 = O_p(n^{1/2} \rho^{1/2}),
\]
which concludes the proof.

The following lemma shows the consistency of the bootstrap variance estimator.

**Lemma 4.A.6.** Let $\hat{\omega}^2$ be the estimator of the variance of the bootstrap errors $\xi_{p,t}^*$ in regression (4.11), i.e.
\[
\hat{\omega}^2 = \frac{1}{n} \left( \Delta y^* - Z_{-1}^* \hat{\delta}^* \right)' (I - W_p^* (W_p^* W_p^*)^{-1} W_p^*) (\Delta y^* - Z_{-1}^* \hat{\delta}^*).
\]
Then \( \hat{\omega}^2 \xrightarrow{p} \omega^2 \) under Assumptions 4.2, 4.3, 4.4’ and 4.5.

Proof of Lemma 4.A.6. Note that

\[
\begin{align*}
n \hat{\omega}^2 &= (\Delta y^* - Z_{-1}^* \delta^*)'(I - W_p^*(W_p^* W_p^*)^{-1} W_p^*) (\Delta y^* - Z_{-1}^* \delta^*) \\
&= \Delta y^*'(I - W_p^*(W_p^* W_p^*)^{-1} W_p^*) \Delta y^* - \Delta y^*'(I - W_p^*(W_p^* W_p^*)^{-1} W_p^*) Z_{-1}^* \delta^* \\
&\quad - \hat{\delta}^* Z_{-1}^* (I - W_p^*(W_p^* W_p^*)^{-1} W_p^*) \Delta y^* + \hat{\delta}^* Z_{-1}^* (I - W_p^*(W_p^* W_p^*)^{-1} W_p^*) Z_{-1}^* \delta^* \\
&= Z_{-1}^*(I - W_p^*(W_p^* W_p^*)^{-1} W_p^*) \hat{\Xi} - Z_{-1}^*(I - W_p^*(W_p^* W_p^*)^{-1} W_p^*) \delta^* \\
&\quad - \hat{\delta}^* Z_{-1}^* (I - W_p^*(W_p^* W_p^*)^{-1} W_p^*) \Xi + \hat{\delta}^* Z_{-1}^* (I - W_p^*(W_p^* W_p^*)^{-1} W_p^*) Z_{-1}^* \delta^*.
\end{align*}
\]

which we write as

\[
\hat{\omega}^2 = C_n^* - 2D_n^* + E_n^*.
\]

We first look at \( C_n^* \). Write

\[
C_n^* = n^{-1} \Xi_p^* \Xi_p^* - n^{-1} \Xi_p^* W_p^* (W_p^* W_p^*)^{-1} W_p^* \Xi_p^*.
\]

Using that \( \hat{\delta}^* = O_p(n^{-1}) \) and the results from Lemma 4.A.5, we have that

\[
n^{-1} |Z_{-1}^* W_p^* (W_p^* W_p^*)^{-1} W_p^* \Xi_p^*| \leq n^{-1} |n^{-1} \Xi_p^* W_p^*| \left| (W_p^* W_p^*)^{-1} \right| \left| W_p^* \Xi_p^* \right| = n^{-1} O_p(n^{1/2} p^{1/2}) O_p(n^{-1}) O_p(n^{1/2} p^{1/2}) = O_p(n^{-1/2}).
\]

Hence,

\[
C_n^* = n^{-1} \Xi_p^* \Xi_p^* + o_p(1).
\]

Next we turn to \( D_n^* \). We can write \( D_n^* \) as

\[
D_n^* = n^{-1} \Xi_p^* Z_{-1}^* \delta^* - n^{-1} \Xi_p^* W_p^* (W_p^* W_p^*)^{-1} W_p^* Z_{-1}^* \delta^*.
\]

Again using Lemma 4.A.5 and \( \delta^* = O_p(n^{-1}) \), we have

\[
|D_n^*| \leq n^{-1} |\Xi_p^* Z_{-1}^*| |\delta^*| + n^{-1} |\Xi_p^* W_p^*| \left| (W_p^* W_p^*)^{-1} \right| \left| W_p^* Z_{-1}^* \right| |\delta^*| = n^{-1} O_p(n) O_p(n^{-1}) + n^{-1} O_p(n^{1/2} p^{1/2}) O_p(n^{-1}) O_p(n^{1/2} p^{1/2}) = O_p(n^{-1}).
\]

Finally we look at \( E_n^* \).

\[
E_n^* = n^{-1} \hat{\delta}^* Z_{-1}^* \delta^* - n^{-1} \hat{\delta}^* Z_{-1}^* W_p^* (W_p^* W_p^*)^{-1} W_p^* Z_{-1}^* \delta^*.
\]

As before, we use the results from Lemma 4.A.5 and \( \delta^* = O_p(n^{-1}) \) to obtain

\[
|E_n^*| \leq n^{-1} |\hat{\delta}^*| |Z_{-1}^*| |\delta^*| + n^{-1} |\hat{\delta}^*| |Z_{-1}^* W_p^*| \left| (W_p^* W_p^*)^{-1} \right| \left| W_p^* Z_{-1}^* \right| |\delta^*| = n^{-1} O_p(n^{-1}) O_p(n^2) O_p(n^{-1}) + n^{-1} O_p(n^{-1}) O_p(n^{1/2} p^{1/2}) O_p(n^{-1}) O_p(n^{1/2} p^{1/2}) = O_p(n^{-1}).
\]

Therefore, we have that

\[
\hat{\omega}^2 = \frac{1}{n} \sum_{t=1}^{n} \xi_{p,t}^2 + o_p(1).
\]
Next we wish to show that \( \hat{\psi} \sum_{t=1}^{n} \xi_{p,t}^2 = \hat{\psi} \sum_{t=1}^{n} \xi^2 + o_p(1) \), for which our proof is similar as Chang and Park (2002, Proof of Lemma 3.1(c)). Note that

\[
\frac{1}{n} \sum_{t=1}^{n} (\xi_{p,t} - \xi_t)^2 = \frac{1}{n} \sum_{t=1}^{n} \left( \sum_{j=p+1}^{\infty} \hat{\psi}_{p,j} \xi_{t-j}^* \right)^2 = \sum_{j=p+1}^{\infty} \hat{\psi}_{p,j} \sum_{t=1}^{n} \left( \frac{1}{n} \sum_{l=1}^{\infty} \xi_{t-l}^* \xi_{t-l}^* \right) \hat{\psi}_{p,j}
\]

\[
= \sum_{j=p+1}^{\infty} \hat{\psi}_{p,j} \Sigma^* \hat{\psi}_j^* + \sum_{j=p+1}^{\infty} \sum_{t=1}^{n} \left( \hat{\psi}_{p,j} \left( \frac{1}{n} \sum_{l=1}^{\infty} \xi_{t-l}^* \xi_{t-l}^* \right) \hat{\psi}_j - \delta_{ij} \Sigma \right)
\]

\[
= \sum_{j=p+1}^{\infty} \| \hat{\psi}_{p,j} \|^2 O_p(1) + \sum_{j=p+1}^{\infty} \sum_{t=1}^{n} \left( | \hat{\psi}_{p,j} | \| \hat{\psi}_{p,j} \| O_p(n^{-1/2}) = o_p(1).
\]

Then, as

\[
\left| \frac{1}{n} \sum_{t=1}^{n} (\xi_{p,t}^2 - \xi_t^2) \right|^{1/2} \leq \left( \frac{1}{n} \sum_{t=1}^{n} (\xi_{p,t}^2 - \xi_t^2) \right)^{1/2}
\]

as a consequence from the triangle inequality it follows that

\[
\frac{1}{n} \sum_{t=1}^{n} \xi_{p,t}^2 = \frac{1}{n} \sum_{t=1}^{n} \xi_t^2 + o_p(1)
\]

which concludes this step.

For the final step we show that

\[
\left| \frac{1}{n} \sum_{t=1}^{n} \xi_t^2 - \omega \right| \leq \left| \frac{1}{n} \sum_{t=1}^{n} \xi_t^2 - \omega^2 \right| + \left| \omega^2 - \omega \right| = o_p(1),
\]

where \( \omega^2 = E^*(\xi_t^2) \). First, we show that \( \frac{1}{n} \sum_{t=1}^{n} \xi_t^2 - \omega^2 = o_p(1) \). Note that

\[
P^* \left( \left| \frac{1}{n} \sum_{t=1}^{n} \xi_t^2 - \omega^2 \right| > \epsilon \right) \leq \epsilon^{-2} E^* \left( \frac{1}{n} \sum_{t=1}^{n} \xi_t^2 - \omega^2 \right)^2
\]

\[
= \epsilon^{-2} E^* \left( \frac{1}{n} \sum_{t=1}^{n} (\xi_t^2 - E^*(\xi_t^2)) \right)^2
\]

\[
= \epsilon^{-2} \frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} (E^*(\xi_t^2 - E^*(\xi_t^2)) (\xi_t^2 - E^*(\xi_t^2)))
\]

\[
= \epsilon^{-2} \frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} (E^*(\xi_t^2) - E^*(\xi_t^2)) (\xi_t^2 - E^*(\xi_t^2))
\]

\[
= \epsilon^{-2} \frac{1}{n^2} \sum_{t=1}^{n} (E^*(\xi_t^4) - (E^*(\xi_t^2))^2)
\]

Next, we next show that \( \omega^2 \overset{p}{\rightarrow} \omega^2 \). As

\[
\omega^2 = \sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad \text{and} \quad \omega = \sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21},
\]

and \( \Sigma \overset{p}{\rightarrow} \Sigma \) by Lemma 4.A.2, the result follows. This completes the proof. \( \square \)
Proof of Theorem 4.3. Note that

\[ T_{\text{wald}}^* = n\delta^* \left(n^2 \text{Var}\left(\hat{\delta}^*\right)\right)^{-1} n\delta^*. \]

We first look at \( \delta^* \). We can write \( n\delta^* \) as

\[
n\delta^* = n[Z_{-1}^\prime (I - W_p^* (W_p^* W_p^*)^{-1} W_p^*) Z_{-1}]^{-1} Z_{-1}^\prime (I - W_p^* (W_p^* W_p^*)^{-1} W_p^*) \Xi_p^* = \]

\[
[n^{-2} Z_{-1}^\prime Z_{-1} - A_n^*)]^{-1} (n^{-1} Z_{-1}^\prime \Xi_p^* - B_n^*),
\]

where

\[
A_n^* = n^{-2} Z_{-1}^\prime W_p^* (W_p^* W_p^*)^{-1} W_p^* Z_{-1}^\prime \quad \text{and} \quad B_n^* = n^{-1} Z_{-1}^\prime W_p^* (W_p^* W_p^*)^{-1} W_p^* \Xi_p^*.
\]

Using Lemma 4.A.5 c), d) and e), we have

\[
|A_n^*| \leq n^{-3} |Z_{-1}^\prime W_p^*| \left|(n^{-1} W_p^* W_p^*)^{-1}\right| |W_p^* Z_{-1}^\prime| = n^{-3} O_p^*(np^{1/2})O_p^*(1)O_p^*(np^{1/2}) = O_p^*(n^{-1}p)
\]

and

\[
|B_n^*| \leq n^{-2} |Z_{-1}^\prime W_p^*| \left|(n^{-1} W_p^* W_p^*)^{-1}\right| |W_p^* \Xi_p^*| = n^{-2} O_p^*(np^{1/2})O_p^*(1)O_p^*(n^{1/2}p^{1/2}) = O_p^*(n^{-1}p).
\]

Hence, as \( p = o(n^{1/2}) \) (Assumption 4.3), we have that

\[
A_n^* = o_p^*(1) \quad \text{and} \quad B_n^* = o_p^*(1).
\]

Then by Lemma 4.A.5 a) and b), we have

\[
n\delta^* \to n^{-2} Z_{-1}^\prime Z_{-1}^{-1} n^{-1} Z_{-1}^\prime \Xi_p^* + a_p^*(1) \quad \overset{d^*}{\to} \quad \int_0^1 B(r)B(r)'dr \int_0^1 B(r)dB(r).
\]

The estimated variance of \( \delta^* \), is defined as

\[
\text{Var}\left(\hat{\delta}^*\right) = \omega^2 [Z_{-1}^\prime (I - W_p^* (W_p^* W_p^*)^{-1} W_p^*) Z_{-1}]^{-1}.
\]

Using Lemma 4.A.5 and Lemma 4.A.6, we have that

\[
n^2 \text{Var}\left(\hat{\delta}^*\right) \overset{d^*}{\to} \omega^2 \left[ \int_0^1 B(r)B(r)'dr \right]^{-1}.
\]
Finally, using equations (4.45) and (4.46) we can derive that

\[
T_{\text{wald}} \overset{d^*}{=} \left( \int_0^1 dB_z(r) B(r) \left[ \int_0^1 B(r) B(r)' dr \right]^{-1} \right) \left( \omega^2 \left[ \int_0^1 B(r) B(r)' dr \right]^{-1} \right)^{-1} \\
\times \left( \left[ \int_0^1 B(r) B(r)' dr \right]^{-1} \int_0^1 B(r) dB_z(r) \right) \\
= \omega^{-2} \int_0^1 dB_z(r) B(r) \left[ \int_0^1 B(r) B(r)' dr \right]^{-1} \int_0^1 B(r) dB_z(r) \\
= \omega^{-2} \int_0^1 dW_1(r) \omega W(r)' L \Psi(1)' \left[ \int_0^1 \Psi(1) LW(r) W(r)' L \Psi(1)' dr \right]^{-1} \\
\times \int_0^1 \Psi(1) LW(r) dW_1(r) \omega \\
= \int_0^1 dW_1(r) W(r)' \left[ \int_0^1 W(r) W(r)' dr \right]^{-1} \int_0^1 W(r) dW_1(r)
\]

as \( B(r) = \Psi(1) LW(r) \) and \( B_z(r) = \omega W_1(r) \). This completes the proof. \( \square \)
Chapter 5

Cross-Sectional Dependence
Robust Block Bootstrap
Panel Unit Root Tests

In this chapter we consider the issue of unit root testing in cross-sectionally dependent panels. We consider panels that may be characterized by various forms of cross-sectional dependence including (but not exclusive to) the popular common factor framework. We consider block bootstrap versions of the group-mean Im, Pesaran, and Shin (2003) and the pooled Levin, Lin, and Chu (2002) unit root coefficient DF-tests for panel data, originally proposed for a setting of no cross-sectional dependence beyond a common time effect. The tests, suited for testing for unit roots in the observed data, can be easily implemented as no specification or estimation of the dependence structure is required. Asymptotic properties of the tests are derived for $T$ going to infinity and $N$ finite. Asymptotic validity of the bootstrap tests is established in very general settings, including the presence of common factors and even cointegration across units. Properties under the alternative hypothesis are also considered. In a Monte Carlo simulation, the bootstrap tests are found to have rejection frequencies that are much closer to nominal size than the rejection frequencies for the corresponding asymptotic tests. The power properties of the bootstrap tests appear to be similar to those of the asymptotic tests.\footnote{This chapter is based on the paper Palm, Smeekes, and Urbain (2008b).}

5.1 Introduction

The use of panel data to test for unit roots and cointegration has become very popular recently. A major problem with tests for unit roots (and cointegration) in univariate time series is that they lack power for small sample sizes. Therefore one
of the reasons people have turned to panel data, is to utilize the cross-sectional dimension to increase power. Another reason to use panel data is that one might be interested in testing a joint unit root hypothesis for $N$ entities. The so-called first-generation panel unit root tests such as the tests proposed by Levin et al. (2002) and Im et al. (2003) are examples where the cross-sectional dimension is used to construct tests that have higher power than individual unit root tests. However, all the first-generation tests rely on independence along the cross-sectional dimension.

It was soon realized that cross-sectional independence is a highly unrealistic assumption for most settings encountered in practice, and it has been shown that the first-generation tests exhibit large size distortions in the presence of cross-sectional dependence (e.g. O’Connell, 1998). Therefore, so called second-generation panel unit root tests have been constructed to take the cross-sectional dependence into account in some way. These second-generation tests assume specific forms of the cross-sectional dependence as their application depends on modelling the structure of the dependence. Most tests model the cross-sectional dependence in the form of common factors, although the way the common factors are dealt with differs for each test. Examples of second-generation panel unit root tests are Bai and Ng (2004), Moon and Perron (2004), and Pesaran (2007). An extensive Monte Carlo comparison of these tests can be found in Gengenbach, Palm, and Urbain (2008). Breitung and Das (2008) provide an analytical comparison of several first- and second-generation tests in the presence of factor structures.

While the second-generation panel unit root tests can deal with common factor structures and contemporaneous dependence, they cannot deal with other forms of cross-sectional dependence, with the exception of Pedroni, Vogelsang, Wagner, and Westerlund (2008). Of particular interest for practical applications are dynamic interrelationships (an example of which is Granger causality). Our goal in this paper is to present panel unit root tests that can deal not only with common factors, but also with a wide range of other plausible dynamic dependencies.

The tool we use to achieve this is the bootstrap, and in particular the block bootstrap method. Two very useful features of the block bootstrap are that one does not have to model the dependence (both temporal and cross-sectional) in order to apply it, and that it is valid to use under a wide range of possible data generating processes (DGPs). This makes it an appropriate tool to use in this setting with $N$ fixed, possibly large, and large $T$ asymptotics.

Of course, the idea to use the bootstrap in cross-sectionally dependent panels is not new and has already been proposed by Maddala and Wu (1999), but so far no one has considered the theoretical properties of the block bootstrap in this setup. There are theoretical results available for other bootstrap and related resampling methods. Chang (2004) considers sieve bootstrap unit root tests, but the sieve bootstrap can only be applied in panels under restrictive assumptions on the cross-sectional dependence. Kapetanios (2008) proposes a bootstrap resampling scheme which resamples in the cross-sectional dimension instead of the usual time dimension, but this is based on cross-sectional independence. Choi and Chue
5.2 Cross-sectionally dependent panels

(2007) consider subsampling, which does allow for more general dependence, but as the authors themselves state (p. 235) “Notwithstanding these nice features of the subsampling approach, depending on the nature of the problem at hand, other methods like bootstrapping may work better in finite samples.”

Hence, the properties of the block bootstrap are still largely unknown in this setting, while in fact the block bootstrap is quite popular among practitioners. We try to fill this gap by providing theoretical results, mainly about asymptotic validity of block bootstrap panel unit root tests. The block bootstrap method we consider here is the moving-blocks bootstrap (Künsch, 1989), and is an extension of the univariate bootstrap unit root test proposed by Paparoditis and Politis (2003). We will consider a very general DGP that can capture many different interesting and relevant forms of cross-sectional and time dependence.

Our results provide the theoretical justification, supported by Monte Carlo evidence, for the use of the proposed panel unit root tests in applications where one is interested in testing for a unit root in the observed data, and where cross-sectional dependence of possibly unknown form might be present in the data. The tests can be easily implemented, as they do not require the specification and estimation of the cross-sectional dependence structure. For example, it is not necessary to know the number of common factors, nor to estimate these factors. It is not even necessary to know whether common factors are present in the data at all.

The structure of the paper is as follows. Section 5.2 explains the model and assumptions. The test statistics and the construction of the bootstrap versions are discussed in Section 5.3. We establish the asymptotic validity of the bootstrap tests (for \( T \to \infty \) and \( N \) fixed) for various settings in Section 5.4. Finite sample performance, including block length selection, is investigated in Section 5.5. Section 5.6 concludes. All proofs and preliminary results are contained in the Appendix.

Finally, a word on notation. We use \( | \cdot | \) to denote the Euclidean norm for vectors and matrices, i.e. \( |v| = (v'v)^{1/2} \) for a vector \( v \) and \( |M| = (\text{tr} M'M)^{1/2} \) for a matrix \( M \). \( \lfloor x \rfloor \) is the largest integer smaller than or equal to \( x \). Convergence in distribution (probability) is denoted by \( \sim_d \) (\( \sim_p \)). Bootstrap quantities (conditional on the original sample) are indicated by appending a superscript \( * \) to the standard notation.

5.2 Cross-sectionally dependent panels

Let us first describe the model that we use for panels with possible unit roots and that allows for various types of cross-sectional and temporal dependence.

Let 
\[
y_t = (y_{1,t}, \ldots, y_{N,t})' \quad (t = 1, \ldots, T)
\]
be generated as
\[
y_t = \Lambda F_t + \omega_t,
\]
where \( \Lambda = (\lambda_1, \ldots, \lambda_N)' \), \( F_t = (F_{1,t}, \ldots, F_{d,t})' \) and \( \omega_t = (\omega_{1,t}, \ldots, \omega_{N,t})' \). Hence, \( f_t \) are common factors (\( d \) in total), \( \Lambda \) are the factor loadings, and \( \omega_t \) are the
We let the factors and the idiosyncratic components be generated by

\[ F_t = \Phi F_{t-1} + f_t, \]
\[ w_t = \Theta w_{t-1} + v_t, \]

where \( \Phi = \text{diag}(\phi_1, \ldots, \phi_d) \) and \( \Theta = \text{diag}(\theta_1, \ldots, \theta_N) \).

Furthermore we let \( f_t \) and \( v_t \) be constructed as

\[ \begin{bmatrix} v_t \\ f_t \end{bmatrix} = \Psi(L) \varepsilon_t = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{v,t} \\ \varepsilon_{f,t} \end{bmatrix}, \]

where \( \Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j \) (with \( \Psi_0 = I \)). We also divide \( \Psi(z) \) as \( \Psi(z) = (\Psi_1(z), \Psi_2(z))' \) where \( \Psi_i(z) = (\Psi_{i1}(z), \Psi_{i2}(z)) \), \( i = 1, 2 \).

We only need some mild conditions on \( \Psi(z) \) and \( \varepsilon_t \).

**Assumption 5.1.**

(i) \( \det(\Psi(z)) \neq 0 \) for all \( \{z \in \mathbb{C} : |z| = 1\} \) and \( \sum_{j=0}^{\infty} j|\Psi_j| < \infty \).

(ii) \( \varepsilon_t \) is i.i.d. with \( E\varepsilon_t = 0, E\varepsilon_t\varepsilon_t' = \Sigma \) and \( E|\varepsilon_t|^{2+\epsilon} < \infty \) for some \( \epsilon > 0 \).

Our null hypothesis is \( H_0: y_{i,t} \) has a unit root for all \( i = 1, \ldots, N \). As in Bai and Ng (2004) and Breitung and Das (2008), we can discern three different settings under which this can occur.

(A) \( \theta_i = \phi_j = 1 \) for all \( i = 1, \ldots, N \) and \( j = 1, \ldots, d \): both the common factors and the idiosyncratic components have a unit root. This is our first main setting.

(B) \( |\theta_i| < 1 \) for all \( i = 1, \ldots, N \), \( \phi_j = 1 \) for all \( j = 1, \ldots, d \): the common factors have a unit root while the idiosyncratic components are stationary. This is the setting where the units are cross-sectionally cointegrated. In accordance with most of the literature we shall call this *cross-unit cointegration*. We also discuss this case in detail.\(^3\)

(C) \( \theta_i = 1 \) for all \( i = 1, \ldots, N \), \( |\phi_j| < 1 \) for all \( j = 1, \ldots, d \): the common factors are stationary while the idiosyncratic components have a unit root. We shall not discuss this case in detail in Section 5.4 but its properties can easily be derived from the previous two cases.

Note that we are not interested in which of the three settings occur, instead we simply want to test if \( y_{i,t} \) has a unit root for all \( i \).

\(^3\)We could also easily think of a setting in between setting A and B, i.e. one where \( |\theta_i| < 1 \) for all \( i \in I_1 \) and \( \theta_i = 1 \) for all \( i \in I_2 \) (with \( I_1 \cup I_2 = \{1, \ldots, N\} \)). In other words, where part of the units are cointegrated and others are not. We will not analyze this setting in detail as it is basically contained in the analysis of settings (A) and (B).
We can discern different alternative hypotheses. The following two are of interest to us.

- **$H_1^a$:** $y_{i,t}$ is stationary for all $i = 1, \ldots, N$. This implies that $|\theta_i| < 1$ for all $i = 1, \ldots, N$ and $|\phi_j| < 1$ for all $j = 1, \ldots, d$.

- **$H_1^b$:** $y_{i,t}$ is stationary for a (significant) portion of the units. This implies that $|\phi_j| < 1$ for all $j = 1, \ldots, d$; while $|\theta_i| < 1$ for all $i \in I_1$ and $\theta_i = 1$ for all $i \in I_2$, with $I_1 \cup I_2 = \{1, \ldots, N\}$ and $n_1/N = \kappa > 0$, where $n_1$ is the number of elements of $I_1$.

**Remark 5.1.** Note that while the setting we adopt is fairly comparable to factor models such as those considered in Bai and Ng (2004) and Breitung and Das (2008), it is more general in several ways. First, it is very common to assume $\Psi_{12}(z) = \Psi_{21}(z) = 0$ and $\Sigma_{12} = \Sigma_{21} = 0$ such that the factors are independent of the idiosyncratic components. There is however no need to do so in order to obtain our theoretical results, and therefore we will not make this assumption in general. Whenever this assumption is made, this will be explicitly mentioned.

Moreover, and more importantly, in most common factor models only weak dependence between the idiosyncratic components is allowed. We do not make this assumption; instead we allow for a wide array of possible dependencies between the idiosyncratic components, both through $\Sigma$ and $\Psi(z)$. Especially the lag polynomial allows for a wide range of dependencies, including all sorts of dynamic dependencies.

It is therefore that setting (A) is our main setting of interest, as simply setting $\lambda_i = 0$ for all $i = 1, \ldots, N$ results in a model without common factors, where the cross-sectional dependence is completely generated by $\Sigma$ and $\Psi(z)$. This setting is therefore the most general. We also analyze setting (B) as it has generated a lot of attention in the literature (mainly due to Bai and Ng, 2004), but it is in fact a very specialized setting that lacks the generality of setting (A).

**Remark 5.2.** One might wonder if we can actually call $\omega_i$ idiosyncratic components given the degree of interdependence that we allow for, as $\Sigma_{11}$ and $\Psi_{11}(z)$ might be non-diagonal without restrictions beyond full rank and $\Psi_{12}(z)$ might be non-zero. The reason why we keep doing so however is that we would like our setup to encompass two types of models. The first is the traditional approximate factor model, for which one would place additional conditions on the DGP to ensure that the idiosyncratic components would only be weakly dependent. The second is the multivariate time series model where we allow for common components as well as for dependence through a VARMA structure (and where the term idiosyncratic components is rather meaningless).

---

4 Di Iorio and Fachin (2008) discuss several alternative hypotheses that are relevant when testing for the null of no panel cointegration. They also argue that the choice of the test statistic should depend on the alternative hypothesis. Their arguments are valid for the unit root setting as well.

5 In principle we could also let some of the factors be $I(1)$ provided they have zero loadings on the units in $I_1$. We do not consider this however.
Hence, while we formulate our setup as a multivariate time series model, we retain the terminology belonging to the factor model to emphasize that such a model is covered by our setup as well. Note that in our simulations in Section 5.5 we will restrict the dependence between the idiosyncratic components to be weak.

5.3 Bootstrap unit root tests in panels

5.3.1 Test statistics

We will consider bootstrapping simplified versions of the Levin, Lin, and Chu (2002) [LLC] and Im, Pesaran, and Shin (2003) [IPS] test statistics. The first simplification is that we take the test statistics before corrections for mean and variance. The reason is that adding or multiplying the original test statistic and the bootstrap test statistic with the same number will obviously not have an effect on the performance of the tests. This is therefore a completely harmless simplification.

The second simplification is that we consider DF instead of ADF tests. Usually, the main reason to use ADF type of tests is to obtain asymptotically pivotal statistics. However, in the presence of complicated cross-sectional dependence it is often not possible to obtain asymptotically pivotal statistics anyway. There is therefore little reason (at least asymptotically) to use ADF instead of DF tests.

The third simplification is that we look at the DF coefficient test rather than the t-test. The main reason for this is that block bootstrapping naively studentized statistics leads to serious problems in terms of accuracy of the tests as discussed for example in Section 3.1.2 of Härdle et al. (2003). As this is a second order problem, it does not lead to invalidity of the bootstrap, but it does cause the bootstrap to converge at a slower rate than the standard asymptotic approximation.6

Given all these modifications, we prefer to call our test statistics “pooled” and “group-mean” instead of LLC and IPS, respectively. Note though that the essence of the LLC and IPS tests remains in our tests and that our methods can be trivially extended to the original LLC and IPS statistics if one so desires.7

Consider the pooled regression

$$\Delta y_{i,t} = \beta y_{i,t-1} + u_{i,t}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \quad (5.4)$$

for which we define the pooled statistic as

$$\tau_p = T \hat{\beta} \quad \text{where} \quad \hat{\beta} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1} \Delta y_{i,t}}{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1}^2}, \quad (5.5)$$

6While this is a well known result in the statistics literature, it seems to have been widely ignored in the (applied) econometrics literature.

7Note that these tests could also be implemented when we have an unbalanced panel with different numbers of observations $T_i$ over time, provided of course the number of observations increase. The implementation of the block bootstrap in such a setting would, while possible, become considerably more complicated.

142
5.3 Bootstrap unit root tests in panels

Also consider the individual regressions for \( i = 1, \ldots, N \)

\[
\Delta y_{i,t} = \beta_i y_{i,t-1} + u_{i,t}, \quad t = 1, \ldots, T.
\] (5.6)

Then we define our group-mean statistic as

\[
\tau_{gm} = \frac{1}{N} \sum_{i=1}^{N} T \hat{\beta}_i \text{ where } \hat{\beta}_i = \frac{\sum_{t=2}^{T} y_{i,t-1} \Delta y_{i,t}}{\sum_{t=2}^{T} y_{i,t-1}^2}.
\] (5.7)

### 5.3.2 Bootstrap algorithm

We employ the following block bootstrap algorithm, which is a multivariate extension of the algorithm proposed by Paparoditis and Politis (2003) to test for unit roots in univariate time series.

**Bootstrap Algorithm 5.1.**

1. For \( i = 1, \ldots, N \) estimate

\[
y_{i,t} = \rho_i y_{i,t-1} + u_{i,t}
\] (5.8)

consistently by OLS and calculate

\[
\hat{u}_{i,t} = y_{i,t} - \hat{\rho}_i y_{i,t-1} - \frac{1}{T-1} \sum_{t=2}^{T} (y_{i,t} - \hat{\rho}_i y_{i,t-1}).
\] (5.9)

Let \( \hat{u}_t = (\hat{u}_{1,t}, \ldots, \hat{u}_{N,t})' \).

2. Choose a block length \( b \) (smaller than \( T \)). Draw \( i_1, \ldots, i_{k-1} \) i.i.d. from the uniform distribution on \( \{1, 2, \ldots, T-b\} \), where \( k = \lfloor (T - 2)/b \rfloor + 1 \) is the number of blocks.

3. Construct the bootstrap errors \( u^*_1, \ldots, u^*_T \) as follows. Let \( u^*_1 = y_1 \). For \( t > 1 \), let

\[
u^*_t = \hat{u}_{m+s},
\] (5.10)

where \( m = \lfloor (t - 2)/b \rfloor \) and \( s = t - mb - 1 \).

4. Construct \( y^*_t \) recursively as

\[
y^*_t = y^*_{t-1} + u^*_t.
\] (5.11)

5. Calculate the bootstrap versions of the group-mean and pooled statistics.

Using the regression

\[
\Delta y^*_{i,t} = \beta y^*_{i,t-1} + u^*_{i,t}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T
\] (5.12)
calculate

\[ \tau_p^* = T \hat{\beta}^*, \quad \text{where} \quad \hat{\beta}^* = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1}^* \Delta y_{i,t}^*}{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1}^*}, \]  

(5.13)

and using the regressions for \( i = 1, \ldots, N, \)

\[ \Delta y_{i,t}^* = \beta_i^* y_{i,t-1}^* + u_{i,t}, \quad t = 1, \ldots, T \]  

(5.14)
calculate

\[ \tau_{gm}^* = \frac{1}{N} \sum_{i=1}^{N} T \hat{\beta}_i^*, \quad \text{where} \quad \hat{\beta}_i^* = \frac{\sum_{t=2}^{T} y_{i,t-1}^* \Delta y_{i,t}^*}{\sum_{t=2}^{T} y_{i,t-1}^*}, \]  

(5.15)

6. Repeat Steps 2 to 5 \( B \) times, obtaining bootstrap test statistics \( \tau_{\kappa}^{*b}, b = 1, \ldots, B, \kappa = p, gm, \) and select the bootstrap critical value \( c_{\alpha}^* \) as

\[ c_{\alpha}^* = \max \{ c : \sum_{b=1}^{B} I(\tau_{\kappa}^{*b} < c) \leq \alpha \}, \]  

or equivalently as the \( \alpha \)-quantile of the ordered \( \tau_{\kappa}^{*b} \) statistics. Reject the null of a unit root if \( \tau_{\kappa}^{*b} \), calculated from equation (5.5) if \( \kappa = p \) or equation (5.7) if \( \kappa = gm \), is smaller than \( c_{\alpha}^* \), where \( \alpha \) is the nominal level of the test.

Note that a crucial role in the analysis of our block bootstrap method will be played by the series

\[ u_{i,t} = y_{i,t} - \rho_i y_{i,t-1}. \]  

(5.16)

As in Paparoditis and Politis (2003), \( \rho_i = 1 \) should correspond to a unit root in \( y_{i,t} \), while \( \rho_i < 1 \) should correspond to \( y_{i,t} \) being stationary. Given our estimation of \( \rho_i \) in step 1, \( \rho_i \) is implicitly defined as

\[ \rho_i = \lim_{t \to \infty} \frac{E(y_{i,t-1} y_{i,t})}{E(y_{i,t}^2)}, \]  

(5.17)

which fulfills these correspondences (Paparoditis and Politis, 2003, Example 2.1).\(^8\)

Note that under \( H_0 \) we simply have that \( u_{i,t} = y_{i,t} - y_{i,t-1} \) for all \( i = 1, \ldots, N \) or in vector notation \( u_t = \Delta y_t \).

We need that the estimator in step 1 satisfies the properties \( \hat{\rho}_i - \rho_i = O_p(T^{-1}) \) if \( \rho_i = 1 \) and \( \hat{\rho}_i - \rho_i = o_p(1) \) if \( \rho_i < 1 \). Our OLS estimator satisfies these properties (Paparoditis and Politis, 2003, Remark 2.3).

We also need the following assumption on the block length.

**Assumption 5.2.** Let \( b \to \infty \) and \( b = o(T^{1/2}) \) as \( T \to \infty \).

---

\(^8\)Given our definition of \( \rho_i \) it is clear that under stationarity we will always have \( |\rho_i| < 1 \). Paparoditis and Politis (2003, Example 2.2) show that if one estimates and hence implicitly defines \( \rho_i \) differently, for example through an ADF regression, it is not always the case that \( \rho_i > -1 \).
Remark 5.3. While we do not consider deterministic components, our tests can be modified to account for them in the same way as discussed by Levin et al. (2002) and Im et al. (2003). The crucial issue regarding the bootstrap tests is to implement exactly the same deterministic specification in the calculation of the test statistic on the bootstrap sample as in the calculation of the test statistic on the original sample. The only further modification of the bootstrap algorithm would be to include the appropriate deterministic components in step 1 as well.

We will not discuss deterministic components in detail in this paper as it would detract from our main objective to deal with cross-sectional dependence. There is a large literature on deterministic components and their impact. Part of the literature, for example on the local power of panel unit root tests in the case of incidental trends (Moon, Perron, and Phillips, 2007), depends on $N \to \infty$ and will therefore not apply here, although in finite samples these results will most likely have an impact on our tests as well. We would like to stress that the bootstrap will not solve any problems that arise due to the implementation of deterministic components. In order to avoid shifting the focus from dealing with the cross-sectional dependence to dealing with deterministic components, we do not consider them in this paper and refer to the existing literature instead (cf. Breitung and Das, 2005; Moon et al., 2007).

Remark 5.4. Unlike the methods considered by Moon and Perron (2004) and Pesaran (2007), which are essentially tests on the presence of a unit root in the idiosyncratic components as pointed out by Bai and Ng (2007), our methods are tests on the presence of a unit root in the observed data. Therefore in our setup there is no need to consider the properties of the common factors separately.

5.4 Asymptotic properties

In this section we will investigate the asymptotic properties of our (bootstrap) test statistics by letting $T$ go to infinity while keeping $N$ fixed. We study only $T$ asymptotics for two reasons. First, it is standard practice in studies on resampling methods; see for example Chang (2004) and Choi and Chue (2007). Second, it is very difficult to obtain meaningful results for infinite $N$ with our general model without making several stringent additional assumptions. However, as neither our bootstrap method nor our proofs of asymptotic validity depend on the finiteness of $N$, there is no reason to expect that asymptotic validity breaks down with joint $T$ and $N$ asymptotics.

5.4.1 Asymptotic properties under the main null hypothesis

In this section we investigate the validity of the bootstrap procedure proposed above in setting (A), i.e. where $\phi_i = 1$ for all $j = 1, \ldots, d$ and $\theta_i = 1$ for all $i = 1, \ldots, N$ or equivalently $\Phi = I_d$ and $\Theta = I_N$.

Note that under this null hypothesis we can write

$$u_t = \Delta y_t = \Gamma' x_t,$$

(5.18)
where $\Gamma = (I_N, \Lambda)'$, and

$$x_t = (v_t', f_t')' = \Psi(L) \varepsilon_t. \tag{5.19}$$

### Asymptotic properties of the test statistics

We start by presenting the asymptotic distributions for the original series. After all, the bootstrap test statistics should mimic these distributions. The first step is the invariance principle, or functional central limit theorem.

**Lemma 5.1.** Let $y_t$ be generated under $H_0$ setting (A) and let Assumption 5.1 hold. Then, as $T \to \infty$,

$$S_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]} u_t \xrightarrow{d} B(r),$$

where $B(r) = \Gamma' \Psi(1) \Sigma^{1/2} W(r)$ and $W(r)$ denotes a $(N+d)$-dimensional standard Brownian motion.

Next define

$$\Omega = \lim_{T \to \infty} T^{-1} E \left( \sum_{t=1}^{T} u_t \right) \left( \sum_{t=1}^{T} u_t \right)' \quad \text{and} \quad \Omega_0 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E(u_t u_t').$$

The limiting distributions now follow straightforwardly.

**Theorem 5.1.** Let $y_t$ be generated under $H_0$ setting (A) and let Assumption 5.1 hold. Then, as $T \to \infty$,

$$\tau_p \xrightarrow{d} \frac{\sum_{i=1}^{N} \int_{0}^{1} B_i(r) dB_i(r) + \frac{1}{2} (\omega_i - \omega_{0,i})}{\sum_{i=1}^{N} \int_{0}^{1} B_i(r)^2 dr}$$

and

$$\tau_{gm} \xrightarrow{d} \frac{1}{N} \sum_{i=1}^{N} \int_{0}^{1} B_i(r) dB_i(r) + \frac{1}{2} (\omega_i - \omega_{0,i})}{\int_{0}^{1} B_i(r)^2 dr},$$

where $B_i(r)$ is the $i$-th element of $B(r) = \Gamma' \Psi(1) \Sigma^{1/2} W(r)$ and $\omega_i$ ($\omega_{0,i}$) is the $(i,i)$-th element of $\Omega$ ($\Omega_0$).

**Remark 5.5.** To see how the Brownian motion $B(r)$ depends on the idiosyncratic components and on the factors, let $B_v(r) = \Psi_1(1) \Sigma^{1/2} W_v(r)$ be the Brownian motion generated by the idiosyncratic components and $B_f(r) = \Psi_2(1) \Sigma^{1/2} W_f(r)$ the Brownian motion generated by the common factors. With this definition $B(r) = B_v(r) + \Lambda B_f(r)$. Note that if $\Psi_{12}(L) = \Psi_{21}(L) = 0$ and $\Sigma_{12} = \Sigma_{21} = 0$ we can write $B_v(r) = \Psi_{11}(1) \Sigma_{11}^{1/2} W_1(r)$ and $B_f(r) = \Psi_{22}(1) \Sigma_{22}^{1/2} W_2(r)$ where $W_1(r)$ is of dimension $N$ and $W_2(r)$ is of dimension $d$. For the $i$-th element of $B(r)$, $B_i(r)$, we can then write $B_i(r) = B_{v,i}(r) + \lambda_i B_f(r)$.
Asymptotic properties of the bootstrap test statistics

Next we turn to the bootstrap test statistics. The first step is the bootstrap invariance principle.

**Lemma 5.2.** Let $y_t$ be generated under $H_0$ setting (A). Let Assumptions 5.1 and 5.2 hold. Then, as $T \to \infty$,

$$S^*_T(r) = T^{-1/2} \sum_{t=1}^{[T r]} u^*_t \xrightarrow{d'} B(r) \text{ in probability.}$$

Lemma 5.2 shows that the bootstrap partial sum process correctly mimics the original partial sum process. The limiting distributions of the bootstrap test statistics now follow as given below.

**Theorem 5.2.** Let $y_t$ be generated under $H_0$ setting (A). Let Assumptions 5.1 and 5.2 hold. Then, as $T \to \infty$,

$$\tau^*_p \xrightarrow{d'} \frac{1}{N} \sum_{i=1}^N \int_0^1 B_i(r) dB_i(r) + \frac{1}{2} (\omega_i - \omega_{0,i})$$

in probability

and

$$\tau^*_gm \xrightarrow{d'} \frac{1}{N} \sum_{i=1}^N \int_0^1 B_i(r) dB_i(r) + \frac{1}{2} (\omega_i - \omega_{0,i})$$

in probability.

Theorem 5.2 establishes the asymptotic validity of the proposed tests.

5.4.2 Asymptotic properties of the tests under cross-unit cointegration

In this section we look at setting (B), i.e. where $\Phi = I_N$ and $\theta_i < 1$ for all $i = 1, \ldots, N$ in 5.2. Note that in this case we may write

$$u_t = \Delta y_t = \Lambda f_t + \Delta w_t = \Lambda f_t + (1 - L)(1 - \theta L)^{-1} v_t.$$  (5.20)

Now let

$$\Psi(z) = \begin{bmatrix} (1 - \theta L)^{-1} \Psi_1(z) \\ \Psi_2(z) \end{bmatrix},$$  (5.21)

such that

$$\begin{bmatrix} u_t \\ f_t \end{bmatrix} = \Psi(L) \varepsilon_t.$$  (5.22)

Note that $\Psi(z)$ satisfies Assumption 5.1 just as $\Psi(z)$. 

147
Asymptotic properties of the test statistics

We start again by presenting the invariance principle for the original series.

**Lemma 5.3.** Let $y_t$ be generated under $H_0$ setting (B). Let Assumption 5.1 hold. Then, as $T \to \infty$,

$$S_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]} u_t \overset{d}{\to} \bar{B}(r),$$

where $\bar{B}(r) = \Lambda B_f(r)$ and $B_f(r) = \Psi_2(1)^{1/2}W(r)$.

Note that the resulting Brownian motion $\bar{B}(r)$ has reduced rank as it is only generated by the factors and not the idiosyncratic components.

Define

$$\bar{\Omega} = \lim_{T \to \infty} T^{-1} \mathbb{E} \left( \sum_{t=1}^{T} u_t \right) \left( \sum_{t=1}^{T} u_t \right)'$$

and

$$\bar{\Omega}_0 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathbb{E} \left( u_t u_t' \right).$$

Now we can derive the asymptotic distributions.

**Theorem 5.3.** Let $y_t$ be generated under $H_0$ setting (B). Let Assumption 5.1 hold. Then, as $T \to \infty$,

$$\tau_p \overset{d}{\to} \frac{\sum_{i=1}^{N} \int_{0}^{1} \bar{B}_i(r) d\bar{B}_i(r) + \frac{1}{2}(\bar{\omega}_i - \bar{\omega}_0,i)}{\sum_{i=1}^{N} \int_{0}^{1} \bar{B}_i(r)^2 dr},$$

and

$$\tau_{gm} \overset{d}{\to} \frac{1}{N} \sum_{i=1}^{N} \frac{\int_{0}^{1} \bar{B}_i(r) d\bar{B}_i(r) + \frac{1}{2}(\bar{\omega}_i - \bar{\omega}_0,i)}{\int_{0}^{1} \bar{B}_i(r)^2 dr},$$

where $\bar{B}_i(r)$ is the $i$-th element of $\bar{B}(r)$ and $\bar{\omega}_i (\bar{\omega}_0,i)$ is the $(i,i)$-th element of $\bar{\Omega}$ ($\bar{\Omega}_0$).

Asymptotic properties of the bootstrap test statistics

Next we turn to the bootstrap series. Before presenting the bootstrap invariance principle, some discussion is in order.

As can be seen in Lemma 5.3, the Brownian motion generated by the partial sum process has reduced rank as it is only driven by the factors. In order to properly replicate the structure of the original series, the same should be true for the bootstrap partial sum process.

In the proof of Lemma 5.2 it is shown that the bootstrap series $u_t^*$ behaves approximately like $u_{im+s}$, ignoring centering for the moment. Summing over the variables within one block, we obtain

$$\sum_{s=1}^{b} u_{im+s} = \sum_{s=1}^{b} (\Lambda f_{im+s} + \Delta w_{im+s}) = \sum_{s=1}^{b} \Lambda f_{im+s} + w_{im+b} - w_{im},$$

148
as all intermediate terms cancel against each other. This is also what happens in the partial sum of the original series and what explains why only the factors contribute to the Brownian motion.

However, summing both over the blocks and within the blocks, we obtain
\[
\sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^b u_{im+s} = \sum_{m=0}^{\lfloor (k-1)r \rfloor} \left( \sum_{s=1}^b \Lambda f_{im+s} + w_{im+b} - w_{im} \right)
\]
\[
= \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^b \Lambda f_{im+s} + \sum_{m=0}^{\lfloor (k-1)r \rfloor} (w_{im+b} - w_{im}),
\]
where now the endpoints of the blocks do not cancel against each other as the blocks are randomly selected. The first term in this sum is the partial sum process of the factors, which generates the Brownian motion in Lemma 5.3 if we divide by \(T^{1/2}\).

The second part is the partial sum process of the idiosyncratic components which generates an (unwanted) Brownian motion by dividing by \(k^{1/2}\). As this rate is slower than \(T^{1/2}\) by Assumption 5.2, the second part will vanish at rate \(T^{1/2}/k^{1/2}\), so at rate \(b^{1/2}\). Therefore, an increasing block length is crucial to make the second part vanish. In finite samples however one will always have a non-zero partial sum of the idiosyncratic components, although the magnitude will depend on both the sample size and the actual block length. Due to this, the covariance matrix of the resulting Brownian motion will always be of full rank in finite samples instead of reduced rank as in Lemma 5.3. It might therefore be expected that in this setting the block bootstrap might not work optimally in finite samples, although it is also clear that large block lengths should improve the performance of the tests in this case.

Remark 5.6. This result is closely related to the result obtained by Paparoditis and Politis (2003, Lemma 8.5) in their discussion about the difference-based block bootstrap (DBB), in which one also bootstraps an over-differenced series. However, where the different bootstrap stochastic order leads to serious (power) problems for the DBB, it is what preserves the validity of the bootstrap tests in the case of cross-unit cointegration. The result described above is formalized in Lemma 5.A.9 in the Appendix.

Given the discussion above, it is clear that the bootstrap validity is preserved in this setting, giving rise to the following bootstrap invariance principle.

Lemma 5.4. Let \(y_t\) be generated under \(H_0\) setting (B). Let Assumptions 5.1 and 5.2 hold. Then, as \(T \to \infty\),
\[
S_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t^* d_{t} \to \bar{B}(r) \quad \text{in probability.}
\]

Finally we derive the limiting distributions of the test statistics, again establishing asymptotic validity of the bootstrap tests.
Theorem 5.4. Let $y_t$ be generated under $H_0$ setting (B). Let Assumptions 5.1 and 5.2 hold. Then, as $T \to \infty$, 
\[
\tau_p^* \xrightarrow{d^*} \frac{\sum_{i=1}^N \int_0^1 B_i(r)dB_i(r) + \frac{1}{2}(\bar{\omega}_i - \bar{\omega}_{0,i})}{\sum_{i=1}^N \int_0^1 B_i(r)^2dr} \quad \text{in probability}
\]
and
\[
\tau_{gm}^* \xrightarrow{d^*} \frac{1}{N} \sum_{i=1}^N \frac{\int_0^1 B_i(r)dB_i(r) + \frac{1}{2}(\bar{\omega}_i - \bar{\omega}_{0,i})}{\int_0^1 B_i(r)^2dr} \quad \text{in probability.}
\]

5.4.3 Asymptotic properties under the alternative hypothesis

Let us start by considering the alternative $H_1^a$ (stationarity for all $y_{i,t}$).

Let us define
\[
y_t = \Lambda(I_d - \Phi L)^{-1} f_t + (I_N - \Theta L)^{-1} v_t = \Gamma' \Psi^+(L) \varepsilon_t,
\]
where
\[
\Psi^+(L) = \begin{bmatrix} (I_N - \Theta L)^{-1} \Psi_1(L) \\ (I_d - \Phi L)^{-1} \Psi_2(L) \end{bmatrix}.
\]

Note that the lag polynomial $\Psi^+(z)$ meets the conditions in Assumption 5.1.

We start by describing the asymptotic properties of our test statistics.

Lemma 5.5. Let $y_t$ be generated under $H_1^a$. Let Assumption 5.1 hold. Then, as $T \to \infty$, 
\[
T^{-1-\gamma} \xrightarrow{P} \frac{\sum_{i=1}^N (\gamma_i(1) - \gamma_i(0))}{\sum_{i=1}^N \gamma_i(0)}
\]
and
\[
T^{-1-\gamma} \xrightarrow{P} N^{-1} \sum_{i=1}^N \frac{\gamma_i(1) - \gamma_i(0)}{\gamma_i(0)},
\]
where $\gamma_i(j) = E(y_{i,t-j}y_{i,t})$.

Lemma 5.5 shows that both test statistics diverge to $-\infty$ under $H_1^a$ as $\gamma_i(1) < \gamma_i(0)$ for all $i = 1, \ldots, N$. This is a necessary, but for bootstrap tests not sufficient step in showing consistency of the tests. The second step that is needed is to show that the bootstrap tests, and correspondingly the bootstrap critical values, do not diverge under $H_1^a$.

To that end, let $P = \text{diag}(\rho_1, \ldots, \rho_N)$ and consequently $u_t = (I_N - PL)y_t$. Then
\[
u_t = (I_N - PL)\Gamma' \Psi^+(L) \varepsilon_t = \Psi^+(L) \varepsilon_t \tag{5.25}
\]
where $\Psi^{++}(L) = (I_N - PL)\Gamma'\Psi^+$. Note that the summability condition from Assumption 5.1 still holds for this lag polynomial. Therefore we can give the following theorem.

**Theorem 5.5.** Let $y_t$ be generated under $H^a_1$. Let Assumptions 5.1 and 5.2 hold. Then, as $T \to \infty$,

$$
\tau^*_p \xrightarrow{d} \frac{\sum_{i=1}^{N} \int_0^1 B_i^+(r) dB_i^+(r) + \frac{1}{2}(\omega_i^+ - \omega_{0,i}^+)}{\sum_{i=1}^{N} \int_0^1 B_i^+(r)^2 dr} \text{ in probability,}
$$

and

$$
\tau^*_{gm} \xrightarrow{d} \frac{1}{N} \sum_{i=1}^{N} \frac{\int_0^1 B_i^+(r) dB_i^+(r) + \frac{1}{2}(\omega_i^+ - \omega_{0,i}^+)}{\int_0^1 B_i^+(r)^2 dr} \text{ in probability},
$$

where $B_i^+(r)$ is the $i$-th element of $B^+(r) = \Psi^{++}(1)\Sigma^{1/2}W(r)$ and $\omega_i^+$ and $\omega_{0,i}^+$ are the $(i,i)$-th elements of $\Omega^+ = \Psi^{++}(1)\Sigma^{+1/2}$ and $\Omega_0^+ = \sum_{j=0}^{\infty} \Psi_j^+\Sigma^{+1/2}$, respectively.

Note that Lemma 5.5 and Theorem 5.5 jointly establish the consistency of our tests.

Let us now consider $H^b_1$. Again we first look at the properties of the test statistics. Let us first without loss of generality assume that the first $n_1$ units are I(0), while the rest is I(1). Hence, $\rho_i < 1$ for $i = 1, \ldots, n_1$ and $\rho_i = 1$ for $i = n_1 + 1, \ldots, N$.

**Lemma 5.6.** Let $y_t$ be generated under $H^b_1$. Let Assumption 5.1 hold. Then, as $T \to \infty$,

$$
\tau^*_p \xrightarrow{d} \frac{\sum_{i=1}^{n_1} (\gamma_i(1) - \gamma_i(0)) + \sum_{i=n_1+1}^{N} \left( \int_0^1 B_i(r) dB_i(r) + \frac{1}{2}(\omega_i^+ - \omega_{0,i}^+) \right)}{\sum_{i=n_1+1}^{N} \int_0^1 B_i(r)^2 dr} \text{ in probability,}
$$

and

$$
T^{-1/2} \tau^*_{gm} \xrightarrow{p} N^{-1} \sum_{i=1}^{n_1} \frac{\gamma_i(1) - \gamma_i(0)}{\gamma_i(0)},
$$

where $\gamma_i(j) = E(y_{i,t-j}y_{i,t})$.

We see that the group-mean statistic diverges to $-\infty$ as it should. The pooled statistic does not diverge however, which means it is not consistent against this alternative. This is in fact not surprising, given that the pooled test is designed as a large $N$-test for homogeneous alternatives (also see Remark 5.8).

Let us turn to the bootstrap series and define $u_t = y_t - Py_{t-1}$, where now part of the $\rho_i$ are equal to one and the rest is smaller than one. We may then write that

$$
u_t = \Psi^#(L)\varepsilon_t, \quad (5.26)$$
where the values for $\Psi^\#(L)$ for the I(1) components are determined as in the analysis under the null, and for the I(0) components as in the analysis above. The summability condition will obviously still hold and therefore we can directly state the limiting distributions as a corollary.

**Corollary 5.1.** Let $y_t$ be generated under $H^b_1$. Let Assumptions 5.1 and 5.2 hold. Then, as $T \to \infty$,

$$
\tau_p^* \overset{d^*}{\to} \frac{\sum_{i=1}^N \int_0^1 B^\#_i(r) dB^\#_i(r) + \frac{1}{2}(\omega^\# - \omega^\#_{0,i})}{\sum_{i=1}^N \int_0^1 B^\#_i(r)^2 dr}
$$

in probability, and

$$
\tau_{gm}^* \overset{d^*}{\to} \frac{1}{N} \sum_{i=1}^N \int_0^1 B^\#_i(r) dB^\#_i(r) + \frac{1}{2}(\omega^\# - \omega^\#_{0,i})}{\int_0^1 B^\#_i(r)^2 dr}
$$

in probability,

where $B^\#_i(r)$ is the $i$-th element of $B^\#(r) = \Psi^\#(1)\Sigma^{1/2}W(r)$ and $\omega^\#_i$ and $\omega^\#_{0,i}$ are the $(i,i)$-th elements of $\Omega^\# = \Psi^\#(1)\Sigma\Psi^\#(1)'$ and $\Omega^\#_0 = \sum_{j=0}^{\infty} \Psi^\#_j \Sigma \Psi^\#_j'$, respectively.

Note that Lemma 5.6 and Corollary 5.1 jointly establish the consistency of the bootstrap group-mean test. Also note that the inconsistency of the pooled test does not depend on the bootstrap distribution, but purely on the original test statistic.

**Remark 5.7.** It might seem that our bootstrap method does not correctly reproduce the asymptotic null distribution if the alternative is true as the nuisance parameters are different than for example in Theorem 5.2, but this is not so straightforward. It all depends on how exactly the alternative is formulated related to the null. Had we formulated our alternative as $y_t = Py_{t-1} + u_t$ where $u_t = \Gamma^\prime \Psi(L) \epsilon_t$, the nuisance parameters would have been the same. The key to understanding this is that the process under the null corresponding to the process in (5.1) and (5.2) with $\Phi$ and $\Theta$ implying stationarity is not necessarily the same process with $\Phi = I_d$ and $\Theta = I_N$.

**Remark 5.8.** A few qualifications are in order regarding the inconsistency of the pooled test. First, the actual location of the pooled test can be seen to depend on both the proportion of stationary units (through $n_1$ in the sums) and the distance from the null (through the quantity $\gamma_i(1) - \gamma_i(0)$). If either becomes larger, the statistic will become more negative. Second, if $T$ increases, the denominator will become smaller as the sum over the stationary units disappears (the $b_{iT}$ part in the proof). Hence the test statistic will grow larger with increasing $T$, but the denominator will not go to zero as the nonstationary part does not vanish. Both factors imply that the actual power of the test can still be non-trivial and even reach 1.
5.5 Small sample performance

In this section we will investigate the small sample properties of our tests using Monte Carlo simulations. First we perform a simulation study to investigate the properties of our tests while fixing the block length to be a function of $T$ only. Next we will perform a separate and smaller simulation study to investigate the selection of block lengths.

5.5.1 Monte Carlo design

We consider the following DGP for the simulation study.

$$y_t = \Lambda F_t + w_t,$$

where

$$F_t = \phi F_{t-1} + f_t,$$
$$w_{i,t} = \theta_i w_{i,t-1} + v_{i,t}.$$

Furthermore,

$$v_t = A_1 v_{t-1} + \varepsilon_{1,t} + B_1 \varepsilon_{1,t-1},$$
$$f_t = \alpha_2 f_{t-1} + \varepsilon_{2,t} + \beta_2 \varepsilon_{2,t-1},$$

where $\varepsilon_{2,t} \sim N(0,1)$ and $\varepsilon_{1,t} \sim N(0,\Sigma)$.

We consider $\Sigma$ is generated as in Chang (2004):

1. Generate an $N \times N$ matrix $U \sim U[0,1]$. Construct $H = U(U'U)^{-1/2}$.
2. Generate $N$ eigenvalues $\lambda_1, \ldots, \lambda_N$ with $\lambda_1 = r$, $\lambda_N = 1$ and $\lambda_i \sim U[r,1]$ for $i = 2, \ldots, N-1$.
3. Let $\Lambda = diag(\lambda_1, \ldots, \lambda_N)$. Then let $\Sigma = H\Lambda H'$.

We consider both $r = 1$ (no cross-sectional dependence) and $r = 0.1$.

We consider five settings regarding the parameters in equations (5.27) and (5.29) in accordance with Gengenbach et al. (2008).

I No common factor, unit root for all idiosyncratic components: $\lambda_i = 0$, $\theta_i = 1$ for all $i = 1, \ldots, N$.

II Unit root in common factor and idiosyncratic components: $\phi = 1$, $\theta_i = 1$ for all $i = 1, \ldots, N$ and $\lambda_i \sim U[-1,3]$.

III Unit root in common factor, stationary idiosyncratic components: $\phi = 1$, $\theta_i \sim U[0.8,1]$ and $\lambda_i \sim U[-1,3]$. This is the setting of cross-unit cointegration.
IV No common factor, stationary idiosyncratic component: \( \theta_i \sim U[0.8, 1] \) and \( \lambda_i = 0 \) for all \( i = 1, \ldots, N \). This is under the alternative hypothesis.\(^9\)

V Stationary common factor and idiosyncratic component: \( \phi = 0.95, \theta_i \sim U[0.8, 1] \) and \( \lambda_i \sim U[-1, 3] \). This is also under the alternative hypothesis.\(^9\)

We consider two different options for the parameters \( A_1 \) and \( B_1 \):

1. No dynamic dependence: \( A_1 = B_1 = 0 \).

2. Dynamic autoregressive moving-average cross-sectional dependence: \( A_1 \) and \( B_1 \) are non-diagonal.

   We let \( A_1 = \Xi \), where
   \[
   \Xi = \begin{bmatrix}
   \xi_1 & \xi_1 \eta_1 & \xi_1 \eta_1^2 & \cdots & \xi_1 \eta_1^{N-1} \\
   \xi_2 \eta_2 & \xi_2 & \xi_2 \eta_2^2 & \cdots & \xi_2 \eta_2^{N-2} \\
   \vdots & \ddots & \ddots & \ddots & \vdots \\
   \xi_N \eta_N^{N-1} & \cdots & \xi_N \eta_N^2 & \xi_N \eta_N & \xi_N 
\end{bmatrix},
   \]
   \[(5.30)\]

   where \( \xi_i, \eta_i \sim U[-0.5, 0.5] \). To ensure stationarity and invertibility we impose that \( \det(I_N - A_1 z) \neq 0 \) for \( \{ z \in \mathbb{C} : |z| \leq 1.2 \} \).

   Furthermore we let \( B_1 = \Omega \). We construct \( \Omega \) in much the same way as \( \Sigma \). Let \( M = HLH' \) where \( H = U(U'U)^{-1/2} \), with \( U \) a \( N \times 1 \)-vector of \( U[0,1] \)-variables, and \( L \) is a diagonal matrix with on the diagonal \( L_1, \ldots, L_N \) where \( L_1 = 0.1, L_N = 1 \) and \( L_2, \ldots, L_{N-1} \sim U[0,1,1] \). We then let \( \Omega = 2M - I_N \).

   By generating \( \Omega \) this way we assure that \( I_N + \Omega \) is of full rank. Note that invertibility is not guaranteed (on purpose).

   The parameters of the common factor in (5.29), \( \alpha_2 \) and \( \beta_2 \), are taken in accordance with the setting for the idiosyncratic components, so if the dependence for the idiosyncratic components is of the ARMA type, then the same will hold for the common factor. Note that for both \( \Sigma \) and the \( \Psi(1) \) matrix derived from \( A \) and \( B \) the eigenvalues are bounded if \( N \to \infty \); as such these parameters can be regarded as weak dependence parameters.

   For all combinations of the parameters described above we consider all combinations of \( T = 25, 50, 100 \) and \( N = 5, 25, 50 \). As several parameters in our DGP are chosen randomly, we repeat the simulations for each setting ten times, and store the mean, median, minimum and maximum. We only report results for the mean here. The mean is representative as in general there is little dispersion between the simulation results. The other results are available upon request. The results are based on 2000 simulations and the Warp-Speed bootstrap (Giacomini et al., 2007) is used to obtain estimates for the rejection frequencies of the bootstrap tests.

   In our simulation study we consider the LLC and IPS tests (with lag lengths selected by BIC), denoted by \( \tau_{llc} \) and \( \tau_{ips} \) respectively, and the bootstrap pooled

\(^9\)The reported power estimates are not size adjusted.
and group-mean tests, denoted by $\tau_p$ and $\tau_{gm}$. We also consider a bootstrap test based on the median of the individual test statistics, denoted by $\tau_{med}$. This test might be more robust to outlying units than the test based on the mean (also see the discussion in Di Iorio and Fachin, 2008). While we do not consider this test explicitly in our theoretical analysis as the median presents difficulties for asymptotic analysis, it is clear that a median based test will be valid as well as we can show that the joint bootstrap distribution of the individual DF statistics is asymptotically valid. Block lengths of the bootstrap tests were taken as $b = 1.75 T^{1/3}$, which amounts to blocks of length 6, 7 and 9 for sample size 25, 50 and 100 respectively, which is within the range usually considered in the literature. We return to the issue of block length selection in Section 5.5.3.

5.5.2 Monte Carlo results

Table 5.1 presents results for the setting without common factors. It can be noted in general that the asymptotic tests have poor size for $T = 25$, which is mainly caused by the performance of the BIC, as this tends to select too large lag lengths for $T = 25$. From $T = 50$ on this does not happen anymore. The first part of the table presents results for the setting without any dependence (both temporal and cross-sectional). It can be seen that the asymptotic tests have good size properties for $T = 50$ and $T = 100$, while the bootstrap tests are undersized increasing in $N$. The second part lists results for the setting where there is only contemporaneous correlation. The asymptotic tests have slight positive size distortions here, while the bootstrap tests are somewhat undersized. The third and fourth part of the table give results for the model with autoregressive moving-average errors. It is clear here that the asymptotic tests are quite oversized, while the bootstrap tests perform well although there is some undersize for large $N$. There is little difference between the three bootstrap tests.

Table 5.2 present the results for the model with a nonstationary common factor and nonstationary idiosyncratic components. For all three settings considered the table shows that the bootstrap tests have good size properties, while the asymptotic tests have large size distortions increasing with $N$. The bootstrap tests again perform very similarly.

Table 5.3 gives the results for the model with cross-unit cointegration, i.e. with a nonstationary common factor and stationary idiosyncratic components. The asymptotic tests have very large size distortions, and while the size distortions of the bootstrap tests are significantly less, they are still large. As expected it indeed seems that the bootstrap tests do not perform very well in this setting.

The problem partly arises, especially for the group-mean test, because for some units the loadings will be very close to zero, thereby making that unit effectively stationary and hence inflating the test statistic. In such a situation we may expect the median-based test to be more robust, and it indeed seems to perform somewhat better than the group-mean test although it still suffers from considerable size distortions.

\footnote{A similar result was obtained by Hlouskova and Wagner (2006).}
Table 5.1: Size properties without common factors (setting I)

<table>
<thead>
<tr>
<th>$A_1$, $B_1$</th>
<th>$\Sigma$</th>
<th>$r$</th>
<th>$T$</th>
<th>$N$</th>
<th>$\tau_{llc}$</th>
<th>$\tau_p$</th>
<th>$\tau_{ips}$</th>
<th>$\tau_{gm}$</th>
<th>$\tau_{med}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td>$r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.140</td>
<td>0.024</td>
<td>0.141</td>
<td>0.020</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.211</td>
<td>0.001</td>
<td>0.183</td>
<td>0.005</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.260</td>
<td>0.000</td>
<td>0.207</td>
<td>0.001</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.076</td>
<td>0.031</td>
<td>0.051</td>
<td>0.024</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.063</td>
<td>0.004</td>
<td>0.060</td>
<td>0.011</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>50</td>
<td>0.055</td>
<td>0.000</td>
<td>0.056</td>
<td>0.003</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>5</td>
<td>0.077</td>
<td>0.032</td>
<td>0.049</td>
<td>0.032</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>25</td>
<td>0.062</td>
<td>0.009</td>
<td>0.051</td>
<td>0.014</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>50</td>
<td>0.056</td>
<td>0.001</td>
<td>0.051</td>
<td>0.005</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.151</td>
<td>0.026</td>
<td>0.159</td>
<td>0.022</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.200</td>
<td>0.003</td>
<td>0.197</td>
<td>0.006</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.236</td>
<td>0.000</td>
<td>0.215</td>
<td>0.001</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.111</td>
<td>0.033</td>
<td>0.070</td>
<td>0.028</td>
<td>0.037</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.077</td>
<td>0.006</td>
<td>0.072</td>
<td>0.009</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>50</td>
<td>0.067</td>
<td>0.001</td>
<td>0.069</td>
<td>0.004</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>5</td>
<td>0.113</td>
<td>0.040</td>
<td>0.066</td>
<td>0.031</td>
<td>0.039</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>25</td>
<td>0.084</td>
<td>0.013</td>
<td>0.064</td>
<td>0.015</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>50</td>
<td>0.073</td>
<td>0.003</td>
<td>0.067</td>
<td>0.007</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td>$r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.215</td>
<td>0.054</td>
<td>0.207</td>
<td>0.082</td>
<td>0.055</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.235</td>
<td>0.004</td>
<td>0.198</td>
<td>0.032</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.280</td>
<td>0.000</td>
<td>0.237</td>
<td>0.010</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.154</td>
<td>0.052</td>
<td>0.123</td>
<td>0.097</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.113</td>
<td>0.008</td>
<td>0.097</td>
<td>0.032</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>50</td>
<td>0.109</td>
<td>0.001</td>
<td>0.106</td>
<td>0.023</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>5</td>
<td>0.152</td>
<td>0.067</td>
<td>0.110</td>
<td>0.099</td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>25</td>
<td>0.130</td>
<td>0.013</td>
<td>0.108</td>
<td>0.028</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>50</td>
<td>0.117</td>
<td>0.004</td>
<td>0.096</td>
<td>0.026</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.222</td>
<td>0.056</td>
<td>0.212</td>
<td>0.063</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.252</td>
<td>0.003</td>
<td>0.238</td>
<td>0.020</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.265</td>
<td>0.000</td>
<td>0.214</td>
<td>0.007</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.197</td>
<td>0.049</td>
<td>0.146</td>
<td>0.059</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.144</td>
<td>0.011</td>
<td>0.131</td>
<td>0.052</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>50</td>
<td>0.127</td>
<td>0.001</td>
<td>0.119</td>
<td>0.018</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>5</td>
<td>0.187</td>
<td>0.055</td>
<td>0.129</td>
<td>0.094</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>25</td>
<td>0.158</td>
<td>0.012</td>
<td>0.129</td>
<td>0.029</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>50</td>
<td>0.143</td>
<td>0.004</td>
<td>0.119</td>
<td>0.027</td>
<td>0.015</td>
<td></td>
</tr>
</tbody>
</table>
## 5.5 Small Sample Performance

Table 5.2: Size properties with common factors (setting II)

<table>
<thead>
<tr>
<th>$A_1, B_1$</th>
<th>$\Sigma$</th>
<th>$T$</th>
<th>$N$</th>
<th>$\tau_{llc}$</th>
<th>$\tau_p$</th>
<th>$\tau_{ps}$</th>
<th>$\tau_{gm}$</th>
<th>$\tau_{med}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1 = 0,\ r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.212</td>
<td>0.029</td>
<td>0.188</td>
<td>0.024</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td>$B_1 = 0$</td>
<td>25</td>
<td>25</td>
<td>0.288</td>
<td>0.014</td>
<td>0.337</td>
<td>0.015</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>50</td>
<td>0.352</td>
<td>0.009</td>
<td>0.425</td>
<td>0.013</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>5</td>
<td>0.165</td>
<td>0.036</td>
<td>0.104</td>
<td>0.030</td>
<td>0.037</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.218</td>
<td>0.021</td>
<td>0.280</td>
<td>0.022</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.263</td>
<td>0.018</td>
<td>0.362</td>
<td>0.020</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>5</td>
<td>0.160</td>
<td>0.039</td>
<td>0.095</td>
<td>0.030</td>
<td>0.038</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>25</td>
<td>0.213</td>
<td>0.030</td>
<td>0.258</td>
<td>0.025</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.253</td>
<td>0.020</td>
<td>0.342</td>
<td>0.022</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.229</td>
<td>0.030</td>
<td>0.204</td>
<td>0.029</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>25</td>
<td>0.314</td>
<td>0.025</td>
<td>0.389</td>
<td>0.021</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>50</td>
<td>0.359</td>
<td>0.023</td>
<td>0.467</td>
<td>0.021</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>5</td>
<td>0.193</td>
<td>0.036</td>
<td>0.122</td>
<td>0.031</td>
<td>0.036</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.283</td>
<td>0.032</td>
<td>0.332</td>
<td>0.026</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.296</td>
<td>0.027</td>
<td>0.393</td>
<td>0.025</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>5</td>
<td>0.182</td>
<td>0.040</td>
<td>0.114</td>
<td>0.030</td>
<td>0.036</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>25</td>
<td>0.277</td>
<td>0.034</td>
<td>0.316</td>
<td>0.030</td>
<td>0.038</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.315</td>
<td>0.031</td>
<td>0.390</td>
<td>0.031</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>$A_1 = \Xi,\ r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.269</td>
<td>0.022</td>
<td>0.243</td>
<td>0.023</td>
<td>0.022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>25</td>
<td>0.351</td>
<td>0.012</td>
<td>0.381</td>
<td>0.013</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>50</td>
<td>0.406</td>
<td>0.007</td>
<td>0.448</td>
<td>0.010</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>5</td>
<td>0.253</td>
<td>0.025</td>
<td>0.177</td>
<td>0.050</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.348</td>
<td>0.015</td>
<td>0.358</td>
<td>0.018</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.378</td>
<td>0.013</td>
<td>0.411</td>
<td>0.016</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>5</td>
<td>0.252</td>
<td>0.032</td>
<td>0.172</td>
<td>0.032</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>25</td>
<td>0.373</td>
<td>0.023</td>
<td>0.362</td>
<td>0.025</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.420</td>
<td>0.023</td>
<td>0.425</td>
<td>0.022</td>
<td>0.028</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.3: Size properties with cross-unit cointegration (setting III)

<table>
<thead>
<tr>
<th>$A_1$, $B_1$</th>
<th>$\Sigma$</th>
<th>$T$</th>
<th>$N$</th>
<th>$\tau_{llc}$</th>
<th>$\tau_p$</th>
<th>$\tau_{ps}$</th>
<th>$\tau_{gm}$</th>
<th>$\tau_{med}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td>$r = 1$</td>
<td>25 5 25 5 25 50 50 50 50 100 100 100 100</td>
<td>0.410 0.629 0.698 0.612 0.782 0.816 0.674 0.798 0.845</td>
<td>0.129 0.198 0.224 0.200 0.267 0.281 0.240 0.303</td>
<td>0.332 0.585 0.643 0.463 0.620 0.671 0.550 0.642</td>
<td>0.103 0.173 0.173 0.169 0.259 0.282 0.333</td>
<td>0.102 0.198 0.216 0.170 0.251 0.283</td>
<td>0.198 0.257</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td>$r = 0.1$</td>
<td>25 5 25 5 25 50 50 50 50 100 100 100 100</td>
<td>0.461 0.612 0.667 0.554 0.742 0.764 0.651 0.700</td>
<td>0.096 0.128 0.166 0.141 0.179 0.184 0.171</td>
<td>0.356 0.561 0.598 0.362 0.552 0.591</td>
<td>0.086 0.111 0.148 0.136</td>
<td>0.084 0.126 0.177 0.137</td>
<td>0.184 0.183 0.175</td>
</tr>
<tr>
<td>$A_1 = \Omega$, $B_1 = \Xi$</td>
<td>$r = 0.1$</td>
<td>25 5 25 5 25 50 50 50 50 100 100 100 100</td>
<td>0.427 0.549 0.597 0.466 0.670 0.698 0.651</td>
<td>0.047 0.062 0.067 0.085 0.098</td>
<td>0.357 0.492 0.522 0.309</td>
<td>0.049 0.071 0.065 0.125</td>
<td>0.041 0.064 0.068 0.089</td>
<td>0.092 0.122 0.087</td>
</tr>
</tbody>
</table>

158
Table 5.4 presents results for the model under the alternative without a common factor. The power of the bootstrap tests is satisfactory, and as expected, increases with both $T$ and $N$. The only setting in which we can directly compare the power of the asymptotic and the bootstrap tests is the setting of no dependence, and here power results are very similar. Given that the bootstrap tests are somewhat undersized, this shows that the power of the bootstrap tests is good. In the other settings the power of the bootstrap test is somewhat less than the power of the asymptotic tests, which can be explained by the size distortions of the asymptotic tests. Note that the bootstrap tests perform similarly.

Table 5.5 gives results for power with a common factor. It can be seen that the power of the bootstrap tests still increases with $T$ and $N$, although power is less than in Table 5.4 and especially the increase in power with $N$ is less. This is not surprising as the common factor which is present in every unit ensures that the information on the order of integration is not increased by much by the addition of units in the panel. The fact that the power of the asymptotic tests is higher than the power of the bootstrap tests can be explained by the large size distortions of the asymptotic tests in this case. The bootstrap tests all have similar power properties, although the median-based test seems to be somewhat less powerful than the group-mean test.

### 5.5.3 Block length selection

The Monte Carlo experiment in the previous section was done with fixed block lengths. It is well known from the literature on block bootstrap that the block length selected can have a large effect on the performance of any kind of application of the block bootstrap. That is of course valid here as well. Added to the usual issues relating to the structure of the temporal dependence, block length selection is also important in our setting in the case of cross-unit cointegration, where one can expect that large blocks are needed based on the discussion in Section 5.4.2. Our discussion here mirrors the discussion in Paparoditis and Politis (2003, Section 6.1), who discuss the selection of block lengths for univariate unit root tests.

Quite some research has been done on optimal block length selection in the framework of stationary time series. As noted in Paparoditis and Politis (2003) in order to talk about optimality one needs to set a criterion that is to be optimized. This criterion will depend on the type of application of the bootstrap (variance estimation, confidence intervals, hypothesis tests, etc.). Using higher order asymptotics, it has been found for stationary series that an optimal block length $b_{\text{opt}}$ is of the form

$$b_{\text{opt}} = CT^{1/\kappa},$$

where $\kappa$ is a known integer depending on the type of application and $C$ is usually unknown and depends on the data. Härdle et al. (2003) and Lahiri (2003) give an overview on optimal block lengths in stationary time series.
### Table 5.4: Power properties without common factors (setting IV)

<table>
<thead>
<tr>
<th>$A_1, B_1$</th>
<th>$\Sigma$</th>
<th>$T$</th>
<th>$N$</th>
<th>$\tau_{llc}$</th>
<th>$\tau_p$</th>
<th>$\tau_{ips}$</th>
<th>$\tau_{gm}$</th>
<th>$\tau_{med}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.607</td>
<td>0.507</td>
<td>0.851</td>
<td>0.354</td>
<td>0.337</td>
</tr>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.829</td>
<td>0.866</td>
<td>0.980</td>
<td>0.894</td>
<td>0.892</td>
</tr>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.875</td>
<td>0.958</td>
<td>0.999</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.754</td>
<td>0.757</td>
<td>0.829</td>
<td>0.810</td>
<td>0.773</td>
</tr>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.995</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td></td>
<td>50</td>
<td>50</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td></td>
<td>100</td>
<td>5</td>
<td>0.905</td>
<td>0.929</td>
<td>0.989</td>
<td>0.974</td>
<td>0.946</td>
</tr>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td></td>
<td>100</td>
<td>25</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td></td>
<td>100</td>
<td>50</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.553</td>
<td>0.508</td>
<td>0.608</td>
<td>0.357</td>
<td>0.361</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.832</td>
<td>0.827</td>
<td>0.985</td>
<td>0.850</td>
<td>0.854</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.878</td>
<td>0.919</td>
<td>1.000</td>
<td>0.981</td>
<td>0.980</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.856</td>
<td>0.630</td>
<td>0.887</td>
<td>0.648</td>
<td>0.633</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.998</td>
<td>0.989</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>50</td>
<td>50</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>100</td>
<td>5</td>
<td>0.928</td>
<td>0.943</td>
<td>0.996</td>
<td>0.985</td>
<td>0.950</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>100</td>
<td>25</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>100</td>
<td>50</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.573</td>
<td>0.491</td>
<td>0.625</td>
<td>0.498</td>
<td>0.391</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.798</td>
<td>0.618</td>
<td>0.972</td>
<td>0.862</td>
<td>0.768</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.866</td>
<td>0.711</td>
<td>0.998</td>
<td>0.984</td>
<td>0.934</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.816</td>
<td>0.700</td>
<td>0.865</td>
<td>0.779</td>
<td>0.675</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.991</td>
<td>0.955</td>
<td>1.000</td>
<td>1.000</td>
<td>0.997</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>50</td>
<td>50</td>
<td>1.000</td>
<td>0.983</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>100</td>
<td>5</td>
<td>0.896</td>
<td>0.891</td>
<td>0.975</td>
<td>0.950</td>
<td>0.955</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>100</td>
<td>25</td>
<td>1.000</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>100</td>
<td>50</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.575</td>
<td>0.354</td>
<td>0.609</td>
<td>0.413</td>
<td>0.361</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.767</td>
<td>0.754</td>
<td>0.949</td>
<td>0.860</td>
<td>0.820</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.841</td>
<td>0.729</td>
<td>0.997</td>
<td>0.981</td>
<td>0.941</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.755</td>
<td>0.693</td>
<td>0.814</td>
<td>0.729</td>
<td>0.607</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.989</td>
<td>0.934</td>
<td>1.000</td>
<td>0.999</td>
<td>0.996</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>50</td>
<td>50</td>
<td>0.999</td>
<td>0.989</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>100</td>
<td>5</td>
<td>0.985</td>
<td>0.762</td>
<td>0.994</td>
<td>0.884</td>
<td>0.830</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>100</td>
<td>25</td>
<td>0.999</td>
<td>0.995</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td></td>
<td>100</td>
<td>50</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Table 5.5: Power properties with common factors (setting V)

<table>
<thead>
<tr>
<th>$A_1, B_1$</th>
<th>$\Sigma$</th>
<th>$T$</th>
<th>$N$</th>
<th>$\tau_{llc}$</th>
<th>$\tau_p$</th>
<th>$\tau_{gps}$</th>
<th>$\tau_{gm}$</th>
<th>$\tau_{med}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1 = 0$, $B_1 = 0$</td>
<td>$r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.597</td>
<td>0.312</td>
<td>0.313</td>
<td>0.177</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.784</td>
<td>0.405</td>
<td>0.306</td>
<td>0.344</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.840</td>
<td>0.858</td>
<td>0.326</td>
<td>0.365</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.883</td>
<td>0.505</td>
<td>0.504</td>
<td>0.474</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.989</td>
<td>0.672</td>
<td>0.672</td>
<td>0.658</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>50</td>
<td>0.997</td>
<td>0.723</td>
<td>0.740</td>
<td>0.722</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>5</td>
<td>0.978</td>
<td>0.970</td>
<td>0.830</td>
<td>0.802</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>25</td>
<td>1.000</td>
<td>0.964</td>
<td>0.933</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>50</td>
<td>1.000</td>
<td>0.981</td>
<td>0.961</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = 0$</td>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.590</td>
<td>0.506</td>
<td>0.140</td>
<td>0.148</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.767</td>
<td>0.772</td>
<td>0.236</td>
<td>0.263</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.822</td>
<td>0.812</td>
<td>0.234</td>
<td>0.266</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.828</td>
<td>0.682</td>
<td>0.423</td>
<td>0.399</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.981</td>
<td>0.935</td>
<td>0.487</td>
<td>0.497</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>50</td>
<td>0.994</td>
<td>0.935</td>
<td>0.509</td>
<td>0.508</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>5</td>
<td>0.963</td>
<td>0.928</td>
<td>0.792</td>
<td>0.731</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>25</td>
<td>1.000</td>
<td>0.986</td>
<td>0.834</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>50</td>
<td>1.000</td>
<td>0.998</td>
<td>0.850</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_1 = \Xi$, $B_1 = \Omega$</td>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.534</td>
<td>0.480</td>
<td>0.087</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.707</td>
<td>0.674</td>
<td>0.085</td>
<td>0.100</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
<td>0.753</td>
<td>0.720</td>
<td>0.126</td>
<td>0.124</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.801</td>
<td>0.669</td>
<td>0.244</td>
<td>0.190</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.953</td>
<td>0.876</td>
<td>0.329</td>
<td>0.253</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>50</td>
<td>0.972</td>
<td>0.903</td>
<td>0.328</td>
<td>0.276</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>5</td>
<td>0.967</td>
<td>0.904</td>
<td>0.619</td>
<td>0.454</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>25</td>
<td>0.998</td>
<td>0.987</td>
<td>0.712</td>
<td>0.494</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>50</td>
<td>0.999</td>
<td>0.989</td>
<td>0.756</td>
<td>0.564</td>
<td></td>
</tr>
</tbody>
</table>
Several methods have been proposed in the setting where one can describe \( b_{\text{opt}} \)
as in (5.31). Some are based on estimation of \( C \) by exploiting the dependence of
\( C \) on certain quantities that can be estimated. Bühlmann and Künsch (1999) and
Politis and White (2004) are examples of such methods that are applicable for
variance estimation. Lahiri, Furukawa, and Lee (2007) propose a plug-in method,
based on the jackknife-after-bootstrap, that is also applicable for confidence inter-
vals and hypothesis test.

A different method is the subsampling approach by Hall, Horowitz, and Jing
(1995). The attractive feature of this method is that it avoids estimation of \( C \).
This feature, as well as the ease of its implementation, has made this method a
popular choice among practitioners. It does however require knowledge of \( \kappa \) to
implement it.

The problem with nonstationary time series is that \( \kappa \) is unknown here, as the
required asymptotic expansions have not been developed yet. This makes it very
difficult to implement any of the methods discussed above using a well funded
choice of \( \kappa \). Paparoditis and Politis (2003) discuss this issue and propose some
heuristic ideas to determine \( \kappa \).

An alternative strategy to the methods discussed above is provided by the
minimum volatility method and calibration method proposed by Politis, Romano,
and Wolf (1999). These methods do not require knowledge of \( \kappa \). The minimum
volatility method involves calculating critical values using a range of block lengths
and selecting the optimal one in the region where the critical values have the lowest
volatility.

We will focus here on the calibration method,\(^{11}\) which we will describe be-
low. In particular, we will consider the Warp-Speed calibration method, which
was considered as a modification of the original calibration method by Giacomini
et al. (2007) for the purpose of constructing confidence intervals. We present the
procedure for hypothesis tests below for completeness.

**Block length selection by Warp-Speed calibration.**

1. Choose a starting value \( b_0 \) for the block length. Using this value, generate \( K \)
bootstrap samples: \( \{(y^1_t), \ldots, (y^K_t)\} \). Calculate the statistic of interest for each
bootstrap sample, say \( \hat{\theta}^k(b_0) \) for \( k = 1, \ldots, K \). Using the empirical distribution
of the statistics, calculate the bootstrap critical value \( c(b_0) \).

2. Let \( (b_1, \ldots, b_M) \) be the candidate block lengths. For each \( i = 1, \ldots, M \) and
\( k = 1, \ldots, K \), construct one bootstrap resample from the bootstrap sample
\( \{y^k_t\} \) using block length \( b_i \), call this \( \{y^k_{i}(i)\} \). Using each resample calculate the
statistic of interest, say \( \hat{\theta}^{*k}(b_i) \).

3. Using the distribution of \( \hat{\theta}^{*k}(b_i) \) for \( k = 1, \ldots, K \), calculate the bootstrap
resample critical value \( c^*(b_i) \) for all \( i = 1, \ldots, M \).

\(^{11}\)We also considered the minimum volatility method, the subsampling method by Hall et al.
(1995) and the plug-in method by Lahiri et al. (2007), the latter two with the value for \( \kappa \) based
on the results for stationary time series, but all these methods were inferior to the calibration
method; see Remark 5.9.
4. Select the optimal block length $b_{\text{opt}}$ such that
\[ b_{\text{opt}} = \arg \min_{b_i, i=1, \ldots, M} |c^*(b_i) - c(b_0)|. \] (5.32)

To reduce the dependence on $b_0$ one can apply this algorithm iteratively, by using $b_{\text{opt}}$ as the starting block length in the next iteration and continuing until convergence.

To analyze the performance of the method, we performed a small Monte Carlo experiment using the same DGP as in Section 5.5.1 applying the tests $\tau_p$ and $\tau_{gm}$. Based on 500 simulations, we let the block length be selected by the Warp-Speed calibration method, and using the same seed, we run the tests for a wide range of fixed block lengths (up to 0.75 times the sample size) to determine the optimal block length. As starting block lengths we take the fixed block lengths from the previous section, while we take $K = 199$. Due to computational costs we do not iterate the algorithm.

Results for size are given in Table 5.6 and 5.7. Optimal block lengths are determined as that block length which gives an empirical rejection frequency the closest to the nominal level (5%). It can be seen that while the optimal rejection frequencies are not obtained using the block length selection method, the rejection frequencies for setting I and II are reasonably close. However, while the selected block lengths do increase for setting III, they do not increase sufficiently compared to the optimal block lengths and size distortions persist.

Results for power are presented in Table 5.8 and 5.9. Optimal block lengths here are selected as the block lengths that give the highest possible power. One should regard this with caution, as optimal block lengths under the alternative hypothesis are difficult to define, as higher power could come at the expense of good size properties under the null. It is therefore not clear that high power is the criterion that should be optimized.\(^\text{12}\) What is clear though, is that choosing an unnecessarily large block length will decrease power. The results show that the calibration method performs reasonably satisfactorily.

To conclude, using the calibration method improves on using a fixed block length, but it is not optimal. It is clear that a lot of work still needs to be done on this topic, especially from a theoretical perspective.

Remark 5.9. As mentioned before, we compared the calibration method to the subsampling approach of Hall et al. (1995), the plug-in method of Lahiri et al. (2007) and the minimum volatility method. The subsampling method tends to select block lengths in a somewhat unpredictable way, although the obtained rejection frequencies are reasonably close (but somewhat inferior) to those obtained with the calibration method. The plug-in method generally favors too small block lengths, regardless of the underlying DGP. The minimum volatility method selects block lengths almost uniformly over the range of allowed lengths, thereby selecting too large block lengths in general. The results are available on request.

\(^{12}\)Note that even when using size-adjusted power this problem would still be present.
Table 5.6: Size properties of $\tau_p$ with block length selection

<table>
<thead>
<tr>
<th>Set.</th>
<th>$A_1$, $B_1$</th>
<th>$\Sigma$</th>
<th>$T$</th>
<th>$N$</th>
<th>RF</th>
<th>AvB</th>
<th>OpB</th>
<th>OpRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>I A1 = $0$, $r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.022</td>
<td>3.992</td>
<td>1</td>
<td>0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 = 0</td>
<td>25</td>
<td>25</td>
<td>0.002</td>
<td>2.778</td>
<td>1</td>
<td>0.012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.028</td>
<td>6.472</td>
<td>5</td>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.012</td>
<td>3.610</td>
<td>1</td>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.044</td>
<td>4.414</td>
<td>4</td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>0.010</td>
<td>2.944</td>
<td>1</td>
<td>0.014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.034</td>
<td>6.990</td>
<td>1</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.016</td>
<td>3.658</td>
<td>1</td>
<td>0.024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 = $\Xi$, $r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.032</td>
<td>4.056</td>
<td>2</td>
<td>0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 = $\Omega$</td>
<td>25</td>
<td>25</td>
<td>0.002</td>
<td>2.304</td>
<td>1</td>
<td>0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.042</td>
<td>7.968</td>
<td>6</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.006</td>
<td>2.918</td>
<td>3</td>
<td>0.010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.024</td>
<td>4.502</td>
<td>4</td>
<td>0.048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>0.008</td>
<td>2.578</td>
<td>1</td>
<td>0.024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.042</td>
<td>7.212</td>
<td>6</td>
<td>0.054</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.008</td>
<td>3.314</td>
<td>5</td>
<td>0.012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II A1 = $0$, $r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.022</td>
<td>4.878</td>
<td>2</td>
<td>0.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 = 0</td>
<td>25</td>
<td>25</td>
<td>0.018</td>
<td>3.510</td>
<td>2</td>
<td>0.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.018</td>
<td>8.850</td>
<td>12</td>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.018</td>
<td>5.566</td>
<td>2</td>
<td>0.046</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.028</td>
<td>5.494</td>
<td>6</td>
<td>0.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>0.016</td>
<td>3.978</td>
<td>1</td>
<td>0.038</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.020</td>
<td>10.292</td>
<td>5</td>
<td>0.048</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.044</td>
<td>7.154</td>
<td>4</td>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 = $\Xi$, $r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.006</td>
<td>4.946</td>
<td>3</td>
<td>0.020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 = $\Omega$</td>
<td>25</td>
<td>25</td>
<td>0.000</td>
<td>4.572</td>
<td>3</td>
<td>0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.010</td>
<td>8.606</td>
<td>2</td>
<td>0.036</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.014</td>
<td>6.270</td>
<td>3</td>
<td>0.018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III A1 = $0$, $r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.098</td>
<td>3.442</td>
<td>16</td>
<td>0.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 = 0</td>
<td>25</td>
<td>25</td>
<td>0.192</td>
<td>4.208</td>
<td>17</td>
<td>0.078</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.126</td>
<td>9.266</td>
<td>30</td>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.230</td>
<td>5.984</td>
<td>37</td>
<td>0.090</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.024</td>
<td>6.626</td>
<td>5</td>
<td>0.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>0.152</td>
<td>4.296</td>
<td>15</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.044</td>
<td>11.222</td>
<td>10</td>
<td>0.048</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.180</td>
<td>7.512</td>
<td>35</td>
<td>0.062</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 = $\Xi$, $r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.046</td>
<td>5.140</td>
<td>5</td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 = $\Omega$</td>
<td>25</td>
<td>25</td>
<td>0.082</td>
<td>4.044</td>
<td>10</td>
<td>0.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.032</td>
<td>9.536</td>
<td>6</td>
<td>0.048</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.070</td>
<td>7.192</td>
<td>1</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RF = rejection frequency with block length selection; AvB = average block length selected; OpB = optimal block length (such that the corresponding rejection frequency is as close as possible to 0.05); OpRF = rejection frequency corresponding to the optimal block length.
5.5 Small sample performance

Table 5.7: Size properties of $\tau_{gm}$ with block length selection

<table>
<thead>
<tr>
<th>Set.</th>
<th>$A_1$, $B_1$</th>
<th>$\Sigma$</th>
<th>$T$</th>
<th>$N$</th>
<th>RF</th>
<th>AvB</th>
<th>OpB</th>
<th>OpRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$A_1 = 0$,</td>
<td>$r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.018</td>
<td>4.172</td>
<td>1</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>$B_1 = 0$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.012</td>
<td>2.762</td>
<td>2</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.020</td>
<td>6.468</td>
<td>2</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.014</td>
<td>3.760</td>
<td>1</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.014</td>
<td>4.378</td>
<td>2</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.012</td>
<td>2.912</td>
<td>1</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.034</td>
<td>6.482</td>
<td>5</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.014</td>
<td>3.672</td>
<td>3</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>$A_1 = \Xi$,</td>
<td>$r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.012</td>
<td>4.504</td>
<td>2</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>$B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.008</td>
<td>2.332</td>
<td>2</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.024</td>
<td>9.906</td>
<td>13</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.046</td>
<td>4.144</td>
<td>5</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.042</td>
<td>5.066</td>
<td>10</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.028</td>
<td>2.968</td>
<td>3</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.036</td>
<td>7.092</td>
<td>4</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.010</td>
<td>3.434</td>
<td>2</td>
<td>0.020</td>
</tr>
<tr>
<td>II</td>
<td>$A_1 = 0$,</td>
<td>$r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.020</td>
<td>4.550</td>
<td>2</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>$B_1 = 0$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.008</td>
<td>3.286</td>
<td>1</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.034</td>
<td>7.086</td>
<td>5</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.024</td>
<td>5.064</td>
<td>1</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.020</td>
<td>4.938</td>
<td>1</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.004</td>
<td>4.154</td>
<td>2</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.032</td>
<td>8.892</td>
<td>2</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.038</td>
<td>6.098</td>
<td>4</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>$A_1 = \Xi$,</td>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.004</td>
<td>5.128</td>
<td>5</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>$B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.006</td>
<td>4.558</td>
<td>6</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.012</td>
<td>8.802</td>
<td>7</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.016</td>
<td>5.532</td>
<td>6</td>
<td>0.024</td>
</tr>
<tr>
<td>III</td>
<td>$A_1 = 0$,</td>
<td>$r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.128</td>
<td>3.638</td>
<td>16</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>$B_1 = 0$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.144</td>
<td>3.976</td>
<td>19</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.108</td>
<td>8.056</td>
<td>27</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.266</td>
<td>5.140</td>
<td>33</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.018</td>
<td>6.394</td>
<td>1</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.102</td>
<td>4.122</td>
<td>15</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.038</td>
<td>9.448</td>
<td>26</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.132</td>
<td>6.572</td>
<td>24</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>$A_1 = \Xi$,</td>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.044</td>
<td>4.756</td>
<td>8</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>$B_1 = \Omega$</td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.114</td>
<td>3.734</td>
<td>14</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.018</td>
<td>8.686</td>
<td>3</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.086</td>
<td>6.266</td>
<td>20</td>
<td>0.052</td>
</tr>
</tbody>
</table>

RF = rejection frequency with block length selection; AvB = average block length selected; OpB = optimal block length (such that the corresponding rejection frequency is as close as possible to 0.05); OpRF = rejection frequency corresponding to the optimal block length.
Table 5.8: Power properties of $\tau_p$ with block length selection

<table>
<thead>
<tr>
<th>Set</th>
<th>$A_1$, $B_1$</th>
<th>$r$</th>
<th>$\Sigma$</th>
<th>T</th>
<th>N</th>
<th>RF</th>
<th>AvB</th>
<th>OpB</th>
<th>OpRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV</td>
<td>$A_1 = 0$,</td>
<td>1</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.218</td>
<td>1.088</td>
<td>2</td>
<td>0.436</td>
</tr>
<tr>
<td></td>
<td>$B_1 = 0$</td>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.916</td>
<td>2.928</td>
<td>1</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.740</td>
<td>5.974</td>
<td>5</td>
<td>0.856</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>1.000</td>
<td>3.732</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td></td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.686</td>
<td>4.530</td>
<td>4</td>
<td>0.816</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.984</td>
<td>2.868</td>
<td>1</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.652</td>
<td>6.592</td>
<td>5</td>
<td>0.792</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>1.000</td>
<td>3.766</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>$A_1 = \Xi$,</td>
<td>1</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.107</td>
<td>3.788</td>
<td>6</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>$B_1 = \Omega$</td>
<td>25</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.476</td>
<td>2.458</td>
<td>4</td>
<td>0.428</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.582</td>
<td>7.586</td>
<td>4</td>
<td>0.682</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.988</td>
<td>3.192</td>
<td>2</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td></td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.234</td>
<td>4.076</td>
<td>4</td>
<td>0.330</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.642</td>
<td>2.414</td>
<td>2</td>
<td>0.704</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.264</td>
<td>6.976</td>
<td>3</td>
<td>0.410</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.998</td>
<td>3.362</td>
<td>2</td>
<td>1.000</td>
</tr>
<tr>
<td>V</td>
<td>$A_1 = 0$,</td>
<td>1</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.118</td>
<td>4.712</td>
<td>1</td>
<td>0.300</td>
</tr>
<tr>
<td></td>
<td>$B_1 = 0$</td>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.328</td>
<td>3.802</td>
<td>1</td>
<td>0.550</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.340</td>
<td>8.188</td>
<td>3</td>
<td>0.486</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.538</td>
<td>6.338</td>
<td>1</td>
<td>0.682</td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td></td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.226</td>
<td>4.652</td>
<td>2</td>
<td>0.406</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.182</td>
<td>4.342</td>
<td>1</td>
<td>0.250</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.182</td>
<td>9.964</td>
<td>3</td>
<td>0.292</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.414</td>
<td>7.096</td>
<td>1</td>
<td>0.508</td>
</tr>
<tr>
<td></td>
<td>$A_1 = \Xi$,</td>
<td>0.1</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.064</td>
<td>5.120</td>
<td>8</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>$B_1 = \Omega$</td>
<td>25</td>
<td></td>
<td>25</td>
<td>5</td>
<td>0.058</td>
<td>4.638</td>
<td>3</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.110</td>
<td>8.774</td>
<td>2</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.200</td>
<td>6.424</td>
<td>7</td>
<td>0.286</td>
</tr>
</tbody>
</table>

RF = rejection frequency with block length selection; AvB = average block length selected; OpB = optimal block length (such that the corresponding rejection frequency is as close as possible to 1); OpRF = rejection frequency corresponding to the optimal block length.
### 5.5 Small Sample Performance

Table 5.9: Power properties of $\tau_{gm}$ with block length selection

<table>
<thead>
<tr>
<th>Set.</th>
<th>$A_1$, $B_1$</th>
<th>Block Length Selection</th>
<th>T</th>
<th>N</th>
<th>RF</th>
<th>AvB</th>
<th>OpB</th>
<th>OpRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV</td>
<td>$A_1 = 0$, $r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.186</td>
<td>4.060</td>
<td>2</td>
<td>0.244</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$B_1 = 0$</td>
<td>25</td>
<td>25</td>
<td>0.986</td>
<td>2.596</td>
<td>1</td>
<td>0.990</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.884</td>
<td>5.438</td>
<td>4</td>
<td>0.948</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>1.000</td>
<td>3.168</td>
<td>1</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.340</td>
<td>4.194</td>
<td>1</td>
<td>0.636</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.966</td>
<td>2.606</td>
<td>2</td>
<td>0.976</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.748</td>
<td>5.730</td>
<td>3</td>
<td>0.850</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>1.000</td>
<td>3.118</td>
<td>1</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$A_1 = \Xi$, $r = 1$</td>
<td>25</td>
<td>5</td>
<td>0.192</td>
<td>3.908</td>
<td>4</td>
<td>0.270</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$B_1 = \Omega$</td>
<td>25</td>
<td>25</td>
<td>0.756</td>
<td>2.390</td>
<td>3</td>
<td>0.860</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.570</td>
<td>7.412</td>
<td>1</td>
<td>0.798</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>1.000</td>
<td>2.850</td>
<td>1</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.250</td>
<td>3.966</td>
<td>3</td>
<td>0.400</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.784</td>
<td>2.390</td>
<td>3</td>
<td>0.880</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.198</td>
<td>6.702</td>
<td>4</td>
<td>0.354</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>1.000</td>
<td>2.994</td>
<td>1</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$A_1 = \Xi$, $r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.138</td>
<td>4.386</td>
<td>1</td>
<td>0.254</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$B_1 = \Omega$</td>
<td>25</td>
<td>25</td>
<td>0.242</td>
<td>3.608</td>
<td>2</td>
<td>0.402</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.372</td>
<td>6.868</td>
<td>3</td>
<td>0.438</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.484</td>
<td>5.638</td>
<td>2</td>
<td>0.684</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.160</td>
<td>4.484</td>
<td>1</td>
<td>0.268</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>0.100</td>
<td>4.276</td>
<td>2</td>
<td>0.240</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.214</td>
<td>8.812</td>
<td>1</td>
<td>0.364</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.580</td>
<td>5.858</td>
<td>1</td>
<td>0.752</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$A_1 = \Xi$, $r = 0.1$</td>
<td>25</td>
<td>5</td>
<td>0.058</td>
<td>4.574</td>
<td>2</td>
<td>0.136</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$B_1 = \Omega$</td>
<td>25</td>
<td>25</td>
<td>0.042</td>
<td>4.108</td>
<td>5</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>5</td>
<td>0.098</td>
<td>8.166</td>
<td>3</td>
<td>0.158</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>0.258</td>
<td>5.508</td>
<td>8</td>
<td>0.288</td>
<td></td>
</tr>
</tbody>
</table>

RF = rejection frequency with block length selection; AvB = average block length selected; OpB = optimal block length (such that the corresponding rejection frequency is as close as possible to 1); OpRF = rejection frequency corresponding to the optimal block length.
5.6 Conclusion

We have established the asymptotic validity of two block bootstrap panel unit root tests for a model that includes various kinds of cross-sectional and temporal dependence. This includes a common factor structure and possibly cross-unit cointegration. The tests are very simple pooled and group-mean tests based on the popular LLC and IPS tests. The finite sample properties of our test statistics have also been investigated and shown to be satisfactory in general. There also seems to be little difference between the bootstrap tests considered.

While for most specific settings (in particular cross-unit cointegration) some tests can be found that perform better for that particular setting, it is a lot more difficult to find a test that is valid for all the settings for which our bootstrap tests are valid. Moreover, there are currently very few tests that are valid in the empirically relevant case of dynamic cross-sectional dependence, while our tests are valid even in that setting. Our tests are very easy to implement as no specification and estimation of the dependence structure is necessary, and will therefore be very useful for practice when the true form of the cross-sectional (and temporal) dependence is not known and robustness to the unknown cross-sectional dependence matters. In fact, quite a lot of practitioners already use the bootstrap to account for cross-sectional dependence for the reasons listed above. Hence, this work provides the necessary theoretical justification.

On the basis of the theoretical and simulation results in this paper, we conclude that it is legitimate to use the proposed tests in practice when testing for unit roots in the observed data of a panel of fixed $N$ entities, in the presence of various forms of cross-sectional dependence. The block bootstrap algorithm described in Section 5.3 can be straightforwardly implemented whereby block lengths can be selected using the Warp-Speed calibration method.

This study still leaves several ends open. First, while we briefly considered the subject of block length selection, much still needs to be done as at the moment there does not exist a fully satisfactory method to select block lengths. Second, while our derivations do not depend on small $N$ in any way, it will be interesting to see what happens if $N \to \infty$. As explained, such a theoretical analysis is very difficult in our setting but it is certainly worth further research. Third, the specification of deterministic components remains an open issue. While a “naive” implementation of deterministic components is quite straightforward, and can even be seen to be valid without too much difficulty, experience has shown that including “naive” deterministic terms in panels is hardly ever a good solution. Thus, further investigation of this issue is also merited.

5.A Appendix: Proofs

Proof of Lemma 5.1. Note that by Assumption 5.1 we have $W_T(r) = T^{-1/2} \sum_{t=1}^{T} \epsilon_t \Sigma^{1/2} W(r)$. Then it follows from standard asymptotic theory for linear processes (see for
5.A Appendix: Proofs

example Phillips and Solo, 1992) that, uniformly in \( r \),

\[
T^{-1/2} \sum_{t=1}^{[Tr]} x_t = \Psi(1)W_T(r) + o_p(1),
\]

and consequently \( T^{-1/2} \sum_{t=1}^{[Tr]} x_t \xrightarrow{d} \Psi(1)\Sigma^{1/2}W(r) \). The result then follows straightforwardly by the continuous mapping theorem.

To prove Theorem 5.1 we need some moments that appear in the asymptotic distributions.

Lemma 5.A.1. Let \( y_t \) be generated under \( H_0 \) setting (A). Let Assumption 5.1 hold. Then

(i) \( \Omega = \lim_{T \to \infty} T^{-1} \mathbb{E} \left( \sum_{t=1}^{T} u_t \right)\left( \sum_{t=1}^{T} u_t \right)' = \Gamma'\Psi(1)\Sigma\Psi(1)'\Gamma \),

(ii) \( \Omega_0 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathbb{E}(u_t u_t') = \sum_{j=0}^{\infty} \Gamma'\Psi_j\Sigma\Psi_j'\Gamma \).

Proof of Lemma 5.A.1. For part (i), note that

\[
\Omega = \lim_{T \to \infty} T^{-1} \mathbb{E} \left( \sum_{t=1}^{T} u_t \right)\left( \sum_{t=1}^{T} u_t \right)' = \lim_{T \to \infty} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E} u_s u_t'
\]

\[
= \lim_{T \to \infty} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E} \left( \sum_{j=0}^{\infty} \Gamma'\Psi_j \varepsilon_{s-j} \right)\left( \sum_{j=0}^{\infty} \Gamma'\Psi_j \varepsilon_{t-j} \right)'
\]

\[
= \lim_{T \to \infty} T^{-1} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma'\Psi_i \left( \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E} \varepsilon_{s-i} \varepsilon_{t-j}' \right) \Psi_j'\Gamma
\]

\[
+ \lim_{T \to \infty} T^{-1} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma'\Psi_i \left( \sum_{s=1}^{T} \mathbb{E} \varepsilon_{s-i} \varepsilon_{t-j}' \right) \Psi_j'\Gamma
\]

\[
= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma'\Psi_i \Sigma \Psi_j'\Gamma = \Gamma'\Psi(1)\Sigma\Psi(1)'\Gamma.
\]

For part (ii) we have

\[
\Omega_0 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathbb{E}(u_t u_t') = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathbb{E} \left( \sum_{j=0}^{\infty} \Gamma'\Psi_j \varepsilon_{t-j} \right)\left( \sum_{j=0}^{\infty} \Gamma'\Psi_j \varepsilon_{t-j} \right)'
\]

\[
= \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma'\Psi_i \mathbb{E} \varepsilon_{t-i} \varepsilon_{t-j}' \Psi_j'\Gamma = \sum_{i=0}^{\infty} \Gamma'\Psi_i \Sigma \Psi_i'\Gamma.
\]

This completes the proof.

Lemma 5.A.2. Let \( y_t \) be generated under \( H_0 \) setting (A). Let Assumption 5.1 hold. Then, as \( T \to \infty \), we have for \( i = 1, \ldots, N \),

169
5 Block Bootstrap Panel Unit Root Tests

(i) \( T^{-1} \sum_{t=1}^{T} y_{t,t-1} \Delta y_{t,t} \xrightarrow{d} \int_{0}^{1} B_{t}(r) dB_{t}(r) + \frac{1}{2} (\omega_{t} - \omega_{0,t}) \),

(ii) \( T^{-2} \sum_{t=1}^{T} y_{t,t-1}^{2} \xrightarrow{d} \int_{0}^{1} B_{t}(r)^{2} dr \).

Proof of Lemma 5.A.3. The proof follows straightforwardly from Lemma 5.1, Lemma
5.A.1 and the continuous mapping theorem.

Proof of Theorem 5.1. The proof follows directly from Lemma 5.A.2.

In order to derive the bootstrap invariance principle we need three preliminary lemmas that build on each other. We use the fact that we have linear processes in our

The proof follows straightforwardly from Lemma 5.1, Lemma 5.1, Lemma 5.A.1 and the continuous mapping theorem.

Lemma 5.A.3. Define \( H_{m}^{*} = b^{-1/2} \sum_{s=1}^{b} (\varepsilon_{m+s} - E^* \varepsilon_{m+s}) \). If Assumptions 5.1 and
5.2 hold, we have

(i) \( E^* H_{m}^{*} = 0 \),

(ii) \( E^* H_{m}^{*} H_{m}^{*} = \Sigma + o_{p}(1) \).

Proof of Lemma 5.A.3. Statement (i) is trivial. To prove statement (ii), write

\[
E^* H_{m}^{*} H_{m}^{*} = E^* \left[ b^{-1} \left( \sum_{s=1}^{b} (\varepsilon_{m+s} - E^* \varepsilon_{m+s}) \right) \left( \sum_{s=1}^{b} (\varepsilon_{m+s} - E^* \varepsilon_{m+s}) \right) \right]
\]

\[
= b^{-1} \sum_{s=1}^{b} \sum_{s=2}^{b} \left( E^* \varepsilon_{m+s} \varepsilon_{m+s} - E^* \varepsilon_{m+s} E^* \varepsilon_{m+s} \right)
\]

\[
= b^{-1} \sum_{s=1}^{b} \sum_{s=2}^{b} \left[ \frac{1}{T-b} \sum_{t=1}^{T-b} \varepsilon_{t+s} \varepsilon_{t+s} - \frac{1}{(T-b)^{2}} \left( \sum_{t=1}^{T-b} \varepsilon_{t+s} \right) \left( \sum_{t=1}^{T-b} \varepsilon_{t+s} \right) \right]
\]

\[
= \frac{1}{b(T-b)} \sum_{s=1}^{b} \sum_{s=2}^{b} \left( \frac{1}{T-b} \sum_{t=1}^{T-b} \varepsilon_{t+s} \right) \left( \frac{1}{T-b} \sum_{t=1}^{T-b} \varepsilon_{t+s} \right)
\]

\[
= A_{T} + B_{T}.
\]

Let us first look at \( B_{T} \). Note that

\[
\frac{1}{T-b} \sum_{s=1}^{b} \sum_{t=1}^{T-b} \varepsilon_{t+s} = \frac{b}{T} \sum_{t=1}^{T} \varepsilon_{t} + \frac{b}{T(T-b)} \sum_{t=1}^{T} \varepsilon_{t} - \frac{1}{T-b} \sum_{s=1}^{b} \sum_{t=1}^{T-b} \varepsilon_{t}
\]

\[
- \frac{1}{T-b} \sum_{s=1}^{b} \sum_{t=T-b-s+1}^{T} \varepsilon_{t}
\]

\[
= O_{p}(bT^{-1/2}) + O_{p}(bT^{-3/2}) + O_{p}(b^{3/2}T^{-1}) + O_{p}(b^{3/2}T^{-1}),
\]

from which we can conclude that \( B_{T} = O_{p}(bT^{-1}) \).
Next we look at the first term. We have

\[
A_T = \frac{1}{b(T-b)} \sum_{s=1}^{b} \sum_{t=1}^{T-b} \varepsilon_{t+s} \varepsilon'_{t+s} + \frac{1}{b(T-b)} \sum_{s=1}^{b} \sum_{l=1, l \neq s}^{b} \sum_{t=1}^{T-b} \varepsilon_{t+s} \varepsilon'_{t+s}
\]

\[
= \frac{1}{b(T-b)} \sum_{t=1}^{T-b} \frac{1}{b(T-b)} \sum_{s=1}^{b} \sum_{l=1}^{b} \left( \sum_{t=1}^{T-b} \varepsilon_{t+s} \varepsilon'_{t+s} \right)
\]

\[
= \frac{1}{b(T-b)} \sum_{s=1}^{b} \sum_{t=1}^{T-b} \sum_{l=1}^{b} \varepsilon_{t+s} \varepsilon'_{t+s} + \frac{1}{b(T-b)} \sum_{s=1}^{b} \sum_{t=1}^{T-b} \sum_{l=1}^{b} \varepsilon_{t+s} \varepsilon'_{t+s}
\]

\[
= T^{-1} \sum_{t=1}^{T} \varepsilon_{t} \varepsilon'_{t} + O_p(bT^{-1}) + O_p(bT^{-1}) + O_p(bT^{-1/2}) = \Sigma + o_p(1).
\]

This concludes the proof of part (ii).

**Lemma 5.A.4.** Let Assumptions 5.1 and 5.2 hold. Then, as \( T \to \infty \),

\[
W_T(r) = T^{-1/2} \sum_{k=0}^{[(k-1)r]} b \sum_{s=1}^{b} (\varepsilon_{im+s} - E^* \varepsilon_{im+s}) \overset{d}{\longrightarrow} \Sigma^{1/2} W(r) \quad \text{in probability.}
\]

**Proof of Lemma 5.A.4.** First note that

\[
T^{-1/2} \sum_{m=0}^{[(k-1)r]} \sum_{s=1}^{b} (\varepsilon_{im+s} - E^* \varepsilon_{im+s}) = k^{-1/2} \sum_{m=0}^{[(k-1)r]} b^{-1/2} \sum_{s=1}^{b} (\varepsilon_{im+s} - E^* \varepsilon_{im+s})
\]

\[
= k^{-1/2} \sum_{m=0}^{[(k-1)r]} H_m^*.
\]

We check the conditions of Corollary 2.2 of Phillips and Durlauf (1986) for the \( H_m^* \) terms. Weak stationarity follows straightforwardly by the definition of the block bootstrap. The moment condition (a), that \( E^* |H_{i,m}|^\beta = O_p(1) \) for some \( 2 \leq \beta < \infty \) is fulfilled with \( \beta = 2 \) by Lemma 5.A.3.

By construction, each \( H_m^* \) is independent, thus fulfilling the mixing condition (b). Then the result follows from Corollary 2.2.

**Lemma 5.A.5.** Let \( y_t \) be generated under \( H_0 \) setting (A). Let Assumptions 5.1 and 5.2 hold. Then, as \( T \to \infty \),

\[
T^{-1/2} \sum_{m=0}^{[(k-1)r]} \sum_{s=1}^{b} (u_{im+s} - E^* u_{im+s}) \overset{d}{\longrightarrow} \Gamma^* \Psi(1) \Sigma^{1/2} W(r).
\]
Proof of Lemma 5.A.5. As

\[ T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (u_{im+s} - E^* u_{im+s}) \]

\[ = T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (\Gamma' x_{im+s} - E^* \Gamma' x_{im+s}) \]

\[ = \Gamma' \left( T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (x_{im+s} - E^* x_{im+s}) \right). \]

we focus on \( T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (x_{im+s} - E^* x_{im+s}) \).

Using the Beveridge-Nelson decomposition we can write

\[ x_{im+s} = \Psi(L) \varepsilon_{im+s} = \Psi(1) \varepsilon_{im+s} - \tilde{\Psi}(L) (\varepsilon_{im+s} - \varepsilon_{im+s-1}), \]

where \( \tilde{\Psi}(z) = \sum_{j=0}^{\infty} \tilde{\Psi}_j z^j \), \( \tilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i \). Then

\[ T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (x_{im+s} - E^* x_{im+s}) \]

\[ = T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} \Psi(1) (\varepsilon_{im+s} - E^* \varepsilon_{im+s}) \]

\[ - T^{-1/2} \sum_{m=0}^{(k-1)r} \left( (\tilde{\Psi}(L) \varepsilon_{im+h} - E^* \tilde{\Psi}(L) \varepsilon_{im+h}) \right) \]

\[ - (\tilde{\Psi}(L) \varepsilon_{im} - E^* \tilde{\Psi}(L) \varepsilon_{im}) \].

We will show that \( T^{-1/2} \sum_{m=0}^{(k-1)r} (\tilde{\Psi}(L) \varepsilon_{im+h} - E^* \tilde{\Psi}(L) \varepsilon_{im+h}) = o_p^*(1) \). First note that

\[ \mathbb{P}^* \left[ T^{-1/2} \sum_{m=0}^{(k-1)r} \left( (\Psi(L) \varepsilon_{im+s} - E^* \tilde{\Psi}(L) \varepsilon_{im+s}) \right) > \epsilon \right] \]

\[ \leq \frac{1}{\epsilon^2} \mathbb{E}^* \left[ T^{-1/2} \sum_{m=0}^{(k-1)r} \left( (\Psi(L) \varepsilon_{im+s} - E^* \tilde{\Psi}(L) \varepsilon_{im+s}) \right)^2 \right] = \frac{1}{\epsilon^2} \mathbb{E}^* |G_{T,s}^*|^2 \]
for \( s = 0, b \) by the Markov inequality. Then, letting \( \xi^*_t = \varepsilon_t - E^* \varepsilon_t \),

\[
E^* G^*_{T,s} = T^{-1} E^* \left( \sum_{m=0}^{[k-1]r} \sum_{j=0}^{\infty} \Psi_j \xi^*_{m+s-j} \right) \left( \sum_{m=0}^{[k-1]r} \sum_{j=0}^{\infty} \tilde{\Psi}_j \xi^*_{m+s-j} \right)
\]

\[
= T^{-1} \sum_{m=0}^{[k-1]r} \sum_{j=0}^{\infty} E^* \left( \sum_{m=0}^{[k-1]r} \sum_{j=0}^{\infty} \Psi_j \xi^*_{m+s-j} \right) \left( \sum_{j=0}^{\infty} \tilde{\Psi}_j \xi^*_{m+s-j} \right)
\]

\[
+ T^{-1} \sum_{m=0}^{[k-1]r} \sum_{m_1=0}^{[k-1]r} \sum_{m_2=0}^{[k-1]r} E^* \left( \sum_{j=0}^{\infty} \Psi_j \xi^*_{m_1+s-j} \right) E^* \left( \sum_{j=0}^{\infty} \tilde{\Psi}_j \xi^*_{m_2+s-j} \right)
\]

\[
= T^{-1} \sum_{m=0}^{[k-1]r} \sum_{j=0}^{\infty} E^* \left( \sum_{j=0}^{\infty} \tilde{\Psi}_j \xi^*_{m+s-j} \right) \left( \sum_{j=0}^{\infty} \tilde{\Psi}_j \xi^*_{m+s-j} \right)
\]

as

\[
E^* \left( \sum_{j=0}^{\infty} \tilde{\Psi}_j (\varepsilon_{m+s-j} - E^* \varepsilon_{m+s-j}) \right) = 0.
\]

Now, by Minkowski’s inequality, we have uniformly in \( r \),

\[
E^* G^*_{T,s}^2 \leq T^{-1} \sum_{m=0}^{[k-1]r} \left[ \sum_{j=0}^{\infty} \left| \tilde{\Psi}_j \right|^2 \left( E^* |\varepsilon_{m+s-j} - E^* \varepsilon_{m+s-j}|^2 \right)^{1/2} \right]^2
\]

\[
\leq 4kT^{-1} \left( \sum_{j=0}^{\infty} \left| \tilde{\Psi}_j \right|^2 \left( \frac{1}{T-b} \right)^{1/2} \right)^2
\]

\[
\leq 4kT^{-1} \left( \sum_{j=0}^{\infty} \left| \tilde{\Psi}_j \right|^2 \right) \max \frac{1}{T-b} \sum_{t=1}^{T-b} |\varepsilon_{t+s-j}|^2.
\]

A sufficient condition for

\[
\sum_{j=0}^{\infty} \left| \tilde{\Psi}_j \right| < \infty
\]

is that

\[
\sum_{j=0}^{\infty} j \left| \Psi_j \right| < \infty,
\]

see Phillips and Solo (1992, Lemma 2.1). This holds by Assumption 5.1. We also have that

\[
\frac{1}{T-b} \sum_{t=1}^{T-b} |\varepsilon_{t+s-j}| = O_p(1)
\]
by the moment conditions in Assumption 5.1. Therefore \( \mathbb{E}^* [G_{T,s}^r]^2 = O_p(b^{-1}) \) for \( s = 0, b \) from which it follows that, uniformly in \( r \),

\[
T^{-1/2} \sum_{m=0}^{[T(b-1)r]} \sum_{s=1}^{b} (x_{im+s} - \mathbb{E}^* x_{im+s}) = \Psi(1) W_T^r(r) + o_p(1)
\]

and therefore

\[
T^{-1/2} \sum_{m=0}^{[T(b-1)r]} \sum_{s=1}^{b} (x_{im+s} - \mathbb{E}^* x_{im+s}) \stackrel{d}{\rightarrow} \Psi(1) \Sigma^{1/2} W(r) \quad \text{in probability}
\]

by Lemma 5.A.4. The proof is concluded by referring to (5.33) and applying the continuous mapping theorem.

**Proof of Lemma 5.2.** Our proof is similar to Paparoditis and Politis (2003, Proof of Theorem 3.1). Note that

\[
S_T^r(r) = T^{-1/2} y_1 + T^{-1/2} \sum_{m=0}^{M_r-1} \sum_{s=1}^{b} \hat{u}_{im+s} + T^{-1/2} \sum_{s=1}^{N_r} \hat{u}_{iM_r+s},
\]

where \( M_r = [(Tr) - 2]/b \) and \( N_r = [Tr] - M_r b - 1 \). As \( T^{-1/2} y_1 = O_p(T^{-1/2}) \), we write

\[
S_T^r(r) = T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \hat{u}_{im+s} - T^{-1/2} \sum_{s=1}^{N_r+1} \hat{u}_{iM_r+s} + O_p(T^{-1/2}).
\]

Now, for the \( i \)-th component of \( S_T^r(r) \), \( S_{T,i}(r) \) \((i = 1, \ldots, N)\), we can write

\[
S_{T,i}(r) = T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( \hat{u}_{i,im+s} - \frac{1}{T-1} \sum_{t=2}^{T} u_{i,t} \right)
- T^{-1/2}(\hat{\rho}_1 - \rho_1) \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( y_{i,im+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right)
- T^{-1/2} \sum_{s=N_r+1}^{b} \left( \hat{u}_{i,im+s} - \frac{1}{T-1} \sum_{t=2}^{T} u_{i,t} \right)
+ T^{-1/2}(\hat{\rho}_1 - \rho_1) \sum_{s=N_r+1}^{b} \left( y_{i,im+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) + O_p(T^{-1/2})
= A_T^r - B_T^r - C_T^r + D_T^r,
\]

which follows from the fact that

\[
\hat{u}_{i,im+s} = y_{i,im+s} - \hat{\rho}_1 y_{i,im+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} (y_{i,t} - \hat{\rho}_1 y_{i,im+s-1})
= y_{i,im+s} - \hat{\rho}_1 y_{i,im+s-1} + \rho_1 y_{i,im+s-1} - \hat{\rho}_1 y_{i,im+s-1}
- \frac{1}{T-1} \sum_{t=2}^{T} (y_{i,t} - \rho_1 y_{i,im+s-1} + \rho_1 y_{i,im+s-1} - \hat{\rho}_1 y_{i,im+s-1})
= u_{i,im+s} - (\hat{\rho}_1 - \rho_1) y_{i,im+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} (u_{i,t} - (\hat{\rho}_1 - \rho_1) y_{i,t-1}).
\]
Similarly, which is of order $O_u$ by the stationarity of $	ext{E}_y$.

Next we turn to $D_T$. We have that

\[ D_T = T^{-1/2}(\hat{\rho}_i - \rho_i) \sum_{s=N+1}^b y_{i,m+s} - \frac{b - N}{T^{1/2}(T - 1)} \sum_{t=2}^T u_{i,t} \]

which is of order $O_u(T^{-1})$ uniformly in $r$ as $\hat{\rho}_i - \rho_i = O_p(T^{-1})$.

Next we turn to $B_T$. Consider

\[
\text{E}^* \left[ \sum_{s=1}^b \left( y_{i,m+s} - \frac{1}{T-1} \sum_{t=2}^T y_{i,t-1} \right) \right] = \frac{1}{T-b} \sum_{t=2}^T \sum_{s=1}^b \left( y_{i,s+t-1} - \frac{1}{T-1} \sum_{t=2}^T y_{i,t-1} \right)
\]

Similarly,

\[
\text{E}^* \left[ \sum_{s=1}^b \left( y_{i,m+s} - \frac{1}{T-1} \sum_{t=2}^T y_{i,t-1} \right) \right]^2 = \frac{1}{T-b} \sum_{t=2}^T \sum_{s=1}^b \left( y_{i,s+t-1} - \frac{1}{T-1} \sum_{t=2}^T y_{i,t-1} \right)^2
\]
Therefore we have, uniformly in \( r \),

\[
E^* \left[ T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( y_{i,m+s} - \frac{1}{T} \sum_{t=2}^{T} y_{i,t-1} \right) \right]^2 \\
= T^{-1} E^* \left\{ \sum_{m_1=0}^{M_r} \sum_{m_2=0}^{M_r} \sum_{s=1}^{b} \left( y_{i,m_1+s} - \frac{1}{T} \sum_{t=2}^{T} y_{i,t-1} \right) \right\} \\
= T^{-1} \sum_{m=0}^{M_r} E^* \left[ \sum_{s=1}^{b} \left( y_{i,m+s} - \frac{1}{T} \sum_{t=2}^{T} y_{i,t-1} \right) \right]^2 \\
+ T^{-1} \sum_{m_1=0}^{M_r} \sum_{m_2=0}^{M_r} \sum_{s=1}^{b} E^* \left[ \left( y_{i,m_1+s} - \frac{1}{T} \sum_{t=2}^{T} y_{i,t-1} \right) \right] \times E^* \left[ \left( y_{i,m_2+s} - \frac{1}{T} \sum_{t=2}^{T} y_{i,t-1} \right) \right] \\
= T^{-1} O(k) O_p(b^2 T) + T^{-1} O(k^2) O_p(b^2 T^{-1}) = O_p(b^2 T) + O_p(b^2),
\]

where we use that the blocks are independent. From this and the fact that \( \hat{\rho}_i - \rho_i = O_p(T^{-1}) \) as we are under the null hypothesis, it follows that

\[
B^*_T = O_p(b^{1/2} T^{-1/2}) = O_p(k^{-1/2}).
\]

Finally we look at \( A^*_T \). We have that

\[
A^*_T = T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( u_{i,m+s} - \frac{1}{T} \sum_{t=2}^{T} u_{i,t} \right) \\
= T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( u_{i,m+s} - E^* u_{i,m+s} + E^* u_{i,m+s} - \frac{1}{T} \sum_{t=2}^{T} u_{i,t} \right) \\
= T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( u_{i,m+s} - E^* u_{i,m+s} \right) \\
+ T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( \frac{1}{T-b} \sum_{t=1}^{T-b} u_{i,t+s} - \frac{1}{T} \sum_{t=2}^{T} u_{i,t} \right).
\]

Note that, uniformly in \( r \),

\[
T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( \frac{1}{T-b} \sum_{t=1}^{T-b} u_{i,t+s} - \frac{1}{T} \sum_{t=2}^{T} u_{i,t} \right) \\
= T^{-1/2} \sum_{m=0}^{M_r} \left[ \frac{1}{T-b} \sum_{t=1}^{b} \left( \sum_{t=2}^{T} u_{i,t} - \sum_{t=T-b+2}^{T} u_{i,t} \right) - \frac{b}{T} \sum_{t=2}^{T} u_{i,t} \right] \\
= T^{-1/2} M_r \left[ \frac{b(b-1)}{T-b} \sum_{t=2}^{T} u_{i,t} - \frac{1}{T-b} \sum_{s=1}^{b} \left( \sum_{t=2}^{T} u_{i,t} + \sum_{t=T-b+2}^{T} u_{i,t} \right) \right] \\
= T^{-1/2} O(k) O_p(b^2 T^{-3/2}) + T^{-1/2} O(k) O_p(b^{3/2} T^{-1}) = O_p(k^{-1/2}).
\]
Combining all the previous results, and realizing they hold for all \( i = 1, \ldots, N \), we have that, uniformly in \( r \),

\[
S^*_T(r) = T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^b (u_{i,m+s} - E^* u_{i,m+s}) + o_p(1),
\]

(5.37)

where we also take the sum up to \( \lfloor (k-1)r \rfloor \) instead of \( M_r \) which is the same asymptotically. The proof is then concluded by applying Lemma 5.A.5.

The next step is to determine the moments of the bootstrap series corresponding to the moments in Lemma 5.A.1.

**Lemma 5.A.6.** Let \( y_t \) be generated under \( H_0 \) setting (A). Let Assumptions 5.1 and 5.2 hold. Then, as \( T \to \infty \),

\begin{enumerate}
\item \( \Omega^* = T^{-1} \left[ E^* \left( \sum_{t=1}^T u_t^* \right) \left( \sum_{t=1}^T u_t^* \right)' - E^* \left( \sum_{t=1}^T u_t^* \right) E^* \left( \sum_{t=1}^T u_t^* \right)' \right] \)
\end{enumerate}

\[
= \Gamma^* \Psi(1) \Sigma \Psi(1)' \Gamma + o_p(1),
\]

\[
(i) \quad \Omega^* = T^{-1} \sum_{t=1}^T [E^* (u_t^* u_t'^*) - E^* u_t^* E^* u_t'^*] = \sum_{t=1}^\infty \Gamma^* \Psi_1 \Sigma \Psi_1' \Gamma + o_p(1).
\]

Proof of Lemma 5.A.6. We start with part (i). Using the arguments in the proof of Lemma 5.2 (take \( r = 1 \)) we can show that

\[
T^{-1/2} \sum_{t=1}^T u_t^* = T^{-1/2} \sum_{m=0}^{k-1} \sum_{s=1}^b (u_{i,m+s} - E^* u_{i,m+s}) + o_p(1).
\]

(5.38)

Therefore

\[
E^* \left[ T^{-1/2} \sum_{t=1}^T u_t^* \right] = o_p(1)
\]

and

\[
\Omega^* = T^{-1} E^* \left( \sum_{m=0}^{k-1} \sum_{s=1}^b (u_{i,m+s} - E^* u_{i,m+s}) \right) \times \left( \sum_{m=0}^{k-1} \sum_{s=1}^b (u_{i,m+s} - E^* u_{i,m+s}) \right) + o_p(1).
\]

Using the Beveridge-Nelson decomposition, we can show, as in the proof of Lemma 5.A.5, that

\[
T^{-1/2} \sum_{m=0}^{k-1} \sum_{s=1}^b (u_{i,m+s} - E^* u_{i,m+s})
\]

\[
= T^{-1/2} \Gamma^* \Psi(1) \sum_{m=0}^{k-1} \sum_{s=1}^b (\varepsilon_{i,m+s} - E^* \varepsilon_{i,m+s}) + o_p(1).
\]

177
5 Block Bootstrap Panel Unit Root Tests

Consequently

$$
\Omega^* = T^{-1} \Gamma' \Psi(1) \mathbb{E} \left[ \left( \sum_{m=0}^{k-1} \sum_{s=1}^b (\varepsilon_{im+s} - E^{*} \varepsilon_{im+s}) \right)^{\prime} \right] \Psi(1) \Gamma + o_p(1)
$$

$$
= k^{-1} \Gamma' \Psi(1) \left( \sum_{m=0}^{k-1} b^{-1/2} \sum_{s=1}^b (\varepsilon_{im+s} - E^{*} \varepsilon_{im+s}) \right)^{\prime} \Psi(1) \Gamma + o_p(1)
$$

$$
= \Gamma' \Psi(1) \Sigma \Psi(1)^{\prime} \Gamma + o_p(1),
$$

where we use the independence of the blocks and the last line follows from Lemma 5.A.3. This concludes the proof of part (i).

The proof of part (ii) is similar to part (i). By (5.36) and the arguments used in the proof of Lemma 5.2, we can straightforwardly show that

$$
\Omega_0 = T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^b E^{*} (u_{im+s} - E^{*} u_{im+s}) (u_{im+s} - E^{*} u_{im+s})^{\prime} + o_p(1).
$$

Then, using that $u_{im+s} = \sum_{j=0}^{\infty} \Gamma^{j} \varepsilon_{im+s-j}$, we can write

$$
\Omega_0^* = T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^b E^{*} \left( \sum_{j=0}^{\infty} \Gamma^{j} \varepsilon_{im+s-j} - E^{*} \varepsilon_{im+s-j} \right)^{\prime} + o_p(1)
$$

$$
= T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^b \sum_{j=0}^{\infty} \Gamma^{j} \varepsilon_{im+s-j} \left( \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma^{i} \Psi(1) \left( \sum_{t=1}^{T-b} (\varepsilon_{t+s-i} - E^{*} \varepsilon_{t+s-i}) \right)^{\prime} \Psi(1) \Gamma + o_p(1) \right)
$$

$$
= A_T + B_T + o_p(1).
$$

Note that

$$
|B_T| \leq b \max_{1 \leq s \leq b} \left| \frac{1}{T-b} \sum_{t=1}^{T-b} \varepsilon_{t+s} \right|^2 |\Gamma|^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} |\Psi| |\Psi_j| = bO_p(T^{-1})O(1) = O_p(k^{-1}).
$$
Then
\[ A_T = \frac{1}{(T-b)b} \sum_{t=1}^{T-b} \sum_{s=1}^{b} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma' \Psi_i (\varepsilon_{t+s-j} \varepsilon'_{t+s-j}) \Psi'_j \Gamma \]
\[ = \frac{1}{(T-b)b} \sum_{t=1}^{T-b} \sum_{s=1}^{b} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma' \Psi_i (\varepsilon_{t+s-i} \varepsilon'_{t+s-i}) \Psi'_j \Gamma \]
\[ + \frac{1}{(T-b)b} \sum_{t=1}^{T-b} \sum_{s=1}^{b} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma' \Psi_i (\varepsilon_{t+s-j} \varepsilon'_{t+s-j}) \Psi'_j \Gamma \]
\[ = A_{1,T} + A_{2,T} \]
and furthermore
\[ |A_{2,T}| \leq \max_{1 \leq r, s \leq 5, r \neq s} \left| \frac{1}{T-b} \sum_{t=1}^{T-b} \varepsilon_{t+r} \varepsilon'_{t+s} \right| |\Gamma|^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} |\Psi_i||\Psi_j| = O_p(T^{-1/2}). \]

For \( A_{1,T} \) note that it is easy to see that (see for example the proof of Lemma 5.A.3)
\[ \frac{1}{(T-b)b} \sum_{t=1}^{T-b} \sum_{s=1}^{b} \varepsilon_{t+s-i} \varepsilon'_{t+s-i} = \Sigma + o_p(1), \]
by which we can conclude that \( \Omega^*_b = \sum_{i=0}^{\infty} \Gamma \Psi_i \Sigma \Psi'_i \Gamma + o_p(1). \)

**Lemma 5.A.7.** Let \( y_t \) be generated under \( H_0 \) setting (A). Let Assumptions 5.1 and 5.2 hold. Then, as \( T \to \infty \), we have for \( i = 1, \ldots, N \),
(i) \( T^{-1} \sum_{t=1}^{T} y^2_{t-i} \Delta y^2_{t-i} \overset{d}{\to} \int_0^1 B_i(r) dB_1(r) + \int_0^1 \omega_i - \omega_{0,i} \),
(ii) \( T^{-2} \sum_{t=1}^{T} y^2_{t-i} d \overset{d}{\to} \int_0^1 B_i(r)^2 dr. \)

**Proof of Lemma 5.A.7.** The result follows straightforwardly from Lemma 5.2, Lemma 5.A.6 and the continuous mapping theorem.

**Proof of Theorem 5.2.** The result follows directly from Lemma 5.A.7.

**Proof of Lemma 5.3.** As in Lemma 5.1, \( W_T(r) = T^{-1/2} \sum_{t=1}^{[T_r]} \varepsilon_t \overset{d}{\to} \Sigma^{1/2} W(r) \) by Assumption 5.1. Then it follows that
\[ T^{-1/2} \sum_{t=1}^{[T_r]} f_t = \Psi_2(1) W_T(r) + o_p(1), \]
uniformly in \( r \), and consequently \( T^{-1/2} \sum_{t=1}^{[T_r]} f_t \overset{d}{\to} \Psi_2(1) \Sigma^{1/2} W(r) \).

Now
\[ T^{-1/2} \sum_{t=1}^{[T_r]} u_t = T^{-1/2} \sum_{t=1}^{[T_r]} \Delta f_t + T^{-1/2} \sum_{t=1}^{[T_r]} \Delta u_t \]
\[ = T^{-1/2} \sum_{t=1}^{[T_r]} \Delta f_t + v_{[T_r]} - v_0 = T^{-1/2} \sum_{t=1}^{[T_r]} \Delta f_t + O_p(T^{-1/2}) \]
uniformly in \( r \) and \( T^{-1/2} \sum_{t=1}^{[T_r]} A f_t \overset{d}{\to} \Lambda \Psi_2(1) \Sigma^{1/2} W(r). \)
The next lemma is the counterpart of Lemma 5.A.1.

**Lemma 5.A.8.** Let $y_t$ be generated under $H_0$ setting (B). Let Assumption 5.1 hold. Then

(i) $\Omega = \lim_{T \to \infty} T^{-1} E \left( \sum_{t=1}^{T} u_t \right) \left( \sum_{t=1}^{T} u_t \right)' = \Lambda \Psi_2 (1) \Sigma \Psi_2 (1)' \Lambda'$,

(ii) $\Omega_0 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E (u_t u_t') = \sum_{j=0}^{\infty} \Lambda \Psi_2 (j) \Sigma \Psi_2 (j)' \Lambda' + \Lambda \Psi_2 (j) \Sigma (\bar{\Psi}_{1,j} - \bar{\Psi}_{1,j+1})' + 2 \Psi_1 (j) \Sigma \Psi_1 (j+1) - \Psi_1 (j) \Sigma \Psi_1 (j)

**Proof of Lemma 5.A.8.** For part (i), note that

$$
\Omega = \lim_{T \to \infty} T^{-1} E \left( \sum_{t=1}^{T} (f_t + \Delta v_t) \right) \left( \sum_{t=1}^{T} (f_t + \Delta v_t) \right)'
$$

$$
= \lim_{T \to \infty} T^{-1} E \left( \sum_{t=1}^{T} f_t + v_T - v_0 \right) \left( \sum_{t=1}^{T} f_t + v_T - v_0 \right)'
$$

$$
= \lim_{T \to \infty} \left\{ T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} E f_s f_t' + T^{-1} E \left( \sum_{t=1}^{T} f_t (v_T - v_0)' \right) + T^{-1} E (v_T - v_0) (v_T - v_0) \right\}
$$

$$
= \lim_{T \to \infty} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} E f_s f_t' = \Lambda \Psi_2 (1) \Sigma \Psi_2 (1)' \Lambda',
$$

where the last step follows straightforwardly as in the proof of Lemma 5.A.1 part (i).

For part (ii) we have

$$
\Omega_0 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E ((f_t + \Delta v_t) (f_t + \Delta v_t)')
$$

$$
= \lim_{T \to \infty} \left\{ T^{-1} \sum_{t=1}^{T} E f_t f_t' + T^{-1} \sum_{t=1}^{T} E \Delta v_t f_t' + T^{-1} \sum_{t=1}^{T} E f_t \Delta v_t' + T^{-1} \sum_{t=1}^{T} E \Delta v_t \Delta v_t' \right\}
$$

$$
= A_T + B_T + C_T + D_T.
$$

Now

$$
A_T = \sum_{j=0}^{\infty} \Lambda \Psi_2 (j) \Sigma \Psi_2 (j)' \Lambda',
$$

analogous to the proof of Lemma 5.A.1 part (ii). Similarly

$$
B_T = \sum_{j=0}^{\infty} (\bar{\Psi}_{1,j} - \bar{\Psi}_{1,j+1}) \Sigma \Psi_2 (j)' \Lambda'
$$

and

$$
C_T = \sum_{j=0}^{\infty} (2 \bar{\Psi}_{1,j} \Sigma \bar{\Psi}_{1,j} - \Psi_{1,j}) \Sigma \bar{\Psi}_{1,j+1} - \Psi_{1,j} \Sigma \bar{\Psi}_{1,j}).
$$

This completes the proof. \qed

180
Proof of Theorem 5.3. Using Lemma 5.3, Lemma 5.A.8 and the continuous mapping theorem we can construct the counterpart of Lemma 5.A.2. The result then follows. □

Lemma 5.A.9. Let \( y_t \) be generated under \( H_0 \) setting (A). Let Assumptions 5.1 and 5.2 hold. Then, as \( T \to \infty \),

\[
T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (u_{i_{m+s}} - E^* u_{i_{m+s}}) \overset{d^*}{\longrightarrow} A\Psi_2(1)\Sigma^{1/2}W(r).
\]

Proof of Lemma 5.A.9. Note that

\[
T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (u_{i_{m+s}} - E^* u_{i_{m+s}}) = T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (Af_{i_{m+s}} + \Delta w_{i_{m+s}}) - E^*(Af_{i_{m+s}} + \Delta w_{i_{m+s}})
\]

\[
= \Lambda \left( T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (f_{i_{m+s}} - E^* f_{i_{m+s}}) + T^{-1/2} \sum_{m=0}^{(k-1)r} (w_{i_{m+s}} - E^* w_{i_{m+s}}) - T^{-1/2} \sum_{m=0}^{(k-1)r} (w_{i_{m}} - E^* w_{i_{m}}) \right)
\]

\[
= A^*_T + B^*_T, 0 + B^*_T, b.
\]

(5.39)

We want to show that \( B^*_T,s = O_p(b^{-1/2}) \) uniformly in \( r \) for \( s = 0, b \). First note that by equation (5.34)

\[
B^*_T,s = T^{-1/2} \sum_{m=0}^{(k-1)r} (\bar{\Psi}_1(L)\varepsilon_{i_{m+s}} - E^* \bar{\Psi}_1(L)\varepsilon_{i_{m+s}}) + o_p(1).
\]

As

\[
k^{-1/2} \sum_{m=0}^{(k-1)r} (\bar{\Psi}_1(L)\varepsilon_{i_{m+s}} - E^* \bar{\Psi}_1(L)\varepsilon_{i_{m+s}}) = O_p(1),
\]

it follows that \( B^*_T,s = O_p(b^{-1/2}) \) uniformly in \( r \) for \( s = 0, b \).

Now we can show in exactly the same way as in the proof of Lemma 5.A.5 that

\[
T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (f_{i_{m+s}} - E^* f_{i_{m+s}}) = \Psi_2(1)W_T(r) + o_p(1)
\]

uniformly in \( r \) and consequently that

\[
T^{-1/2} \sum_{m=0}^{(k-1)r} \sum_{s=1}^{b} (u_{i_{m+s}} - E^* u_{i_{m+s}}) \overset{d^*}{\longrightarrow} A\Psi_2(1)W(r) \quad \text{in probability.} \quad \square
\]
5 Block Bootstrap Panel Unit Root Tests

Proof of Lemma 5.4. As the order of the $f_t$ determines the order of $u_t$, we can show in exactly the same way as in the proof of Lemma 5.2 that, uniformly in $r$,

$$S^*_T(r) = T^{-1/2} \sum_{m=0}^{[k-1]r} \sum_{s=1}^b (u_{t,m+s} - E^* u_{t,m+s}) + o_p^*(1). \quad (5.40)$$

The proof is then concluded by applying Lemma 5.A.9.

We consider the bootstrap moments in the following lemma.

**Lemma 5.A.10.** Let $y_t$ be generated under $H_0$ setting (B). Let Assumptions 5.1 and 5.2 hold. Then, as $T \to \infty$,

(i) $\bar{\Omega}^* = T^{-1} E \left( \sum_{t=1}^T u_t^* \right) \left( \sum_{t=1}^T u_t^* \right)' = \Lambda \Psi_2(1) \Sigma \Psi_2(1)' \Lambda' + o_p(1)$,

(ii) $\bar{\Omega}_0^* = T^{-1} \sum_{t=1}^T E(u_t^* u_t^*) = \sum_{j=0}^\infty \Lambda \Psi_2(1) \Sigma \Psi_2(1)' \Lambda' + \Lambda \Psi_2(1) \Sigma \Psi_2(1)' \Lambda' + 2 \Psi_1(1) \Sigma \Psi_1(1)' \Psi_1(1) - \Psi_1(1) \Sigma \Psi_1(1) + o_p(1)$.

**Proof of Lemma 5.A.10.** We start with part (i). Using the arguments from the proofs of Lemma 5.2 and Lemma 5.4 (take $r = 1$) we can again show that

$$T^{-1/2} \sum_{t=1}^T u_t^* = T^{-1/2} \sum_{m=0}^{k-1} \sum_{s=1}^b (u_{t,m+s} - E^* u_{t,m+s}) + o_p(1). \quad (5.41)$$

from which it follows that

$$\bar{\Omega}^* = T^{-1} E^* \left( \sum_{m=0}^{k-1} \sum_{s=1}^b (u_{t,m+s} - E^* u_{t,m+s}) \right) \left( \sum_{m=0}^{k-1} \sum_{s=1}^b (u_{t,m+s} - E^* u_{t,m+s}) \right)' + o_p(1).$$

Combining the proof of Lemma 5.A.5 and 5.A.9 we can show that

$$T^{-1/2} \sum_{m=0}^{k-1} \sum_{s=1}^b (u_{t,m+s} - E^* u_{t,m+s})$$

$$= T^{-1/2} \sum_{m=0}^{k-1} \sum_{s=1}^b (\varepsilon_{t,m+s} - E^* \varepsilon_{t,m+s}) + o_p(1).$$

Consequently

$$\bar{\Omega}^* = \Lambda \Psi_2(1) \Sigma \Psi_2(1)' \Lambda' + o_p(1),$$

which follows in exactly the same way as in the proof of Lemma 5.A.6. This concludes the proof of part (i).

Next we consider part (ii). As in the proof of Lemma 5.A.6 we can show that

$$\bar{\Omega}_0^* = T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^b (E^* (u_{t,m+s} - E^* u_{t,m+s}) (u_{t,m+s} - E^* u_{t,m+s})' + o_p(1).$$

182
Then we can write
\[
\hat{\Omega}_0^* = T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} E^* ((f_{m+s-j} - E^* f_{m+s-j}) + (\Delta v_{m+s-j} - E^* \Delta v_{m+s-j})) \\
\times ((f_{m+s-j} - E^* f_{m+s-j}) + (\Delta v_{m+s-j} - E^* \Delta v_{m+s-j}))^t + o_p(1)
\]
\[
= T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} E^* (f_{m+s-j} - E^* f_{m+s-j}) (f_{m+s-j} - E^* f_{m+s-j})^t \\
+ T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} E^* (\Delta v_{m+s-j} - E^* \Delta v_{m+s-j}) (f_{m+s-j} - E^* f_{m+s-j})^t \\
+ T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} E^* (\Delta v_{m+s-j} - E^* \Delta v_{m+s-j}) (\Delta v_{m+s-j} - E^* \Delta v_{m+s-j})^t
\]
\[
= A_T + B_T + B_T^* + C_T.
\]
Then we can easily show, in the same way as in the proof of Lemma 5.A.6 part (ii), that

\[
A_T^* = \sum_{j=0}^{\infty} \Lambda \Psi_2,j \Sigma \Psi_2,j \Lambda' + o_p(1),
\]
as well as

\[
B_T = \sum_{j=0}^{\infty} (\tilde{\Psi}_{1,j} - \tilde{\Psi}_{1,j+1}) \Sigma \Psi_2,j \Lambda' + o_p(1)
\]
and

\[
C_T = \sum_{j=0}^{\infty} (2\tilde{\Psi}_{1,j} \Sigma \tilde{\Psi}_{1,j} - \tilde{\Psi}_{1,j} \Sigma \tilde{\Psi}_{1,j+1} - \tilde{\Psi}_{1,j} \Sigma \tilde{\Psi}_{1,j}^t) + o_p(1).
\]
This completes the proof.

**Proof of Theorem 5.4.** As for Theorem 5.3 we can construct the counterpart of Lemma 5.A.7 using Lemma 5.4, Lemma 5.A.10 and the continuous mapping theorem. The result then follows.

**Proof of Lemma 5.5.** We can write

\[
T^{-1} \tau_p = \frac{\sum_{i=1}^{N} T^{-1} \sum_{t=2}^{N} \gamma_{i,t-1} \Delta y_{i,t}}{\sum_{i=1}^{N} T^{-1} \sum_{t=2}^{N} y_{i,t-1}^2} = \frac{\sum_{i=1}^{N} a_{i,T}}{\sum_{i=1}^{N} b_{i,T}}.
\]
Now as \(y_{i,t}\) is a stationary process for all \(i = 1, \ldots, N\), we have that

\[
a_{i,T} = T^{-1} \sum_{t=2}^{T} y_{i,t-1} \gamma_{i,t} - T^{-1} \sum_{t=2}^{T} y_{i,t-1}^2 \overset{P}{\rightarrow} \gamma_i(1) - \gamma_i(0)
\]
183
and
\[ b_T = T^{-1} \sum_{t=2}^{T} y_{t-1}^2 \overset{p}{\to} \gamma(0). \]

Similarly,
\[ T^{-1} \tau_m = N^{-1} \sum_{i=1}^{N} \frac{q(T)}{n T}, \]

from which the result follows.

**Lemma 5.A.11.** Let \( y_t \) be generated under \( H_1^a \). Let Assumptions 5.1 and 5.2 hold. Then, as \( T \to \infty \),
\[ T^{-1/2} \sum_{m=0}^{[(k-1)r]} \sum_{s=1}^{b} (u_{im+s} - E^* u_{im+s}) d_{i,m}^T \overset{d}{\rightarrow} \Psi^{++}(1) \Sigma_{1/2} W(r). \]

**Proof of Lemma 5.A.11.** Using the Beveridge-Nelson decomposition we can write
\[ u_{im+s} = \Psi^{++}(L) \varepsilon_{im+s} = \Psi(1)^{++} \varepsilon_{im+s} - \tilde{\Psi}^{++}(L)(\varepsilon_{im+s} - \varepsilon_{im+s-1}), \]
where \( \Psi^{++}(z) = \sum_{j=0}^{\infty} \Psi^{++}_{j} z^{j}, \Psi^{++}_{j} = \sum_{i=j+1}^{\infty} \Psi^{++}_{i}. \) Then
\[ T^{-1/2} \sum_{m=0}^{[(k-1)r]} \sum_{s=1}^{b} (u_{im+s} - E^* u_{im+s}) \]
\[ = T^{-1/2} \sum_{m=0}^{[(k-1)r]} \sum_{s=1}^{b} \Psi^{++}(1)(\varepsilon_{im+s} - E^* \varepsilon_{im+s}) \]
\[ - T^{-1/2} \sum_{m=0}^{[(k-1)r]} \left( (\tilde{\Psi}^{++}(L)\varepsilon_{im+b} - E^* \tilde{\Psi}^{++}(L)\varepsilon_{im+b}) \right) \]
\[ - (\tilde{\Psi}^{++}(L)\varepsilon_{im} - E^* \tilde{\Psi}(L)\varepsilon_{im}) \].

We need to show that \( T^{-1/2} \sum_{m=0}^{[(k-1)r]} (\tilde{\Psi}^{++}(L)\varepsilon_{im+b} - E^* \tilde{\Psi}(L)^{++} \varepsilon_{im+b}) = o_p^*(1) \), uniformly in \( r \). Completely analogous to the proof of Lemma 5.A.5 this means showing that
\[ \sum_{j=0}^{\infty} |\Psi^{++}_{j}| < \infty \]
or equivalently
\[ \sum_{j=0}^{\infty} j |\Psi^{++}_{j}| < \infty. \]

This holds as we remarked that the summability condition continues to hold.
Proof of Theorem 5.5. We start by showing that the invariance principle holds. As in the proof of Lemma 5.2 we have that

\[ S_T^*(r) = T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( u_{i,m+s} - \frac{1}{T-1} \sum_{t=2}^{T} u_{i,t} \right) - T^{-1/2} \left( \hat{\rho}_i - \rho_i \right) \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left( y_{i,m+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \]

- \[ - T^{-1/2} \sum_{s=N_r+1}^{b} \left( u_{i,m+s} - \frac{1}{T-1} \sum_{t=2}^{T} u_{i,t} \right) + T^{-1/2} \left( \hat{\rho}_i - \rho_i \right) \sum_{s=N_r+1}^{b} \left( y_{i,m+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) + O_p(T^{-1/2}) \]

= \[ A_T^* - B_T^* - C_T^* + D_T^*. \]

As before, we have that, uniformly in \( r \),

\[ C_T^* = T^{-1/2} \sum_{s=N_r+1}^{b} u_{i,m+s} - \frac{1}{T^{1/2}(T-1)} (b - N_r) \sum_{t=2}^{T} u_{i,t} \]

= \[ O_p(b^{1/2}T^{-1/2}) + O_p(bT^{-1}) = O_p(k^{-1/2}). \]

by the stationarity of \( u_i \).

Turning to \( D_T^* \) we have that

\[ D_T^* = T^{-1/2} \left( \hat{\rho}_i - \rho_i \right) \sum_{s=N_r+1}^{b} \sum_{m=0}^{b} y_{i,m+s-1} - \frac{1}{T^{1/2}(T-1)} (\hat{\rho}_i - \rho_i) (b - N_r) \sum_{t=2}^{T} y_{i,t-1} \]

which is of order \( O_p(b^{1/2}T^{-1/2}) \), uniformly in \( r \), by the stationarity of \( y_{i,t} \).

Next we turn to \( B_T^* \). We can write

\[ B_T^* = T^{-1/2} \left( \hat{\rho}_i - \rho_i \right) \sum_{m=0}^{M_r} \sum_{s=1}^{b} y_{i,m+s} - \frac{1}{T^{1/2}(T-1)} (\hat{\rho}_i - \rho_i) M_r b \sum_{t=2}^{T} y_{i,t-1} \]

from which we can conclude that, uniformly in \( r \), \( B_T^* = o_p(1) \) as \( \sum_{m=0}^{M_r} \sum_{s=1}^{b} y_{i,m+s} = O_p(T^{1/2}) \) by the stationarity of \( y_{i,t} \).

The results for \( A_T^* \) remain the same as in the proof of Lemma 5.2 from which we can conclude that, uniformly in \( r \),

\[ S_T^*(r) = T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (u_{i,m+s} - E* u_{i,m+s}) + o_p(1). \]

(5.42)

The result now follows trivially by applying Lemma 5.A.11. \( \square \)

185
Proof of Lemma 5.6. We write
\[
\tau_p \equiv \sum_{i=1}^{N} \frac{T^{-1} \sum_{t=3}^{T} y_{i,t-1} \Delta y_{i,t}}{\sum_{i=1}^{N} \frac{T^{-2} \sum_{t=2}^{T} y_{i,t}^2}{T^{-1} \sum_{i=1}^{N} T^{-1} \sum_{t=2}^{T} y_{i,t-1} + \sum_{i=n_1+1}^{N} T^{-1} \sum_{t=2}^{T} y_{i,t}} + \sum_{i=n_1+1}^{N} T^{-2} \sum_{t=2}^{T} y_{i,t-1}^2}
\]
\[
= \sum_{i=1}^{n_1} \frac{a_i T}{b_i T} + \sum_{i=n_1+1}^{N} \frac{c_i T}{d_i T}.
\]
The convergence of \(a_i T\) and \(b_i T\) follow from the proof of Lemma 5.5. Furthermore, as in Lemma 5.A.2, we have that
\[
c_i T \xrightarrow{d} \int_0^1 B_i(r) dB_i(r),
\]
\[
d_i T \xrightarrow{d} \int_0^1 B_i(r)^2 dr,
\]
from which the result for \(\tau_p\) follows.

For \(\tau_{gm}\) we can write
\[
T^{-1} \tau_{gm} = \sum_{i=1}^{n_1} \frac{a_i T}{b_i T} + \sum_{i=n_1+1}^{N} \frac{c_i T}{d_i T} = \sum_{i=1}^{n_1} \frac{a_i T}{b_i T} + O_p(T^{-1}).
\]

Proof of Corollary 5.1. The proof is immediate by combining the proofs of Theorems 5.2 and 5.5.
Chapter 6

Conclusion

This chapter concludes the thesis. We will first identify the main overall conclusions that can be drawn from this work. As each chapter contains its own conclusion, we will not go into the details here but rather focus on the overall conclusion. Next we will identify some limitations of this work and discuss how this work could be extended.

The general conclusion that can be drawn from this thesis is that it is worthwhile to apply the bootstrap to the analysis of nonstationary time series. Throughout the thesis it has been demonstrated, both theoretically and through simulations, that the bootstrap is a valid tool to use for unit root and cointegration testing. Moreover, bootstrap methods often outperform their asymptotic counterparts.

This conclusion does not imply that the bootstrap can be used without any regard. Instead, the bootstrap should always be used carefully. As illustrated in this thesis, there are many different ways in which one can apply the bootstrap. Not all work equally well, and some will not work at all. Therefore, as done in this thesis, any proposed bootstrap method should always be accompanied by theoretical results demonstrating its validity. Moreover, finite sample results (obtained through simulation studies) are important tools to discern successful methods from less successful ones.

In Chapter 2, it is demonstrated that bootstrapping ADF unit root tests is clearly preferable to bootstrapping DF tests, while it is also shown that (for linear processes) the sieve bootstrap outperforms the block bootstrap. Moreover, a sieve bootstrap test based on residuals is introduced and shown to be asymptotically valid. This test also has good finite sample performance.

Chapter 3 highlights that the method of detrending is an important aspect of bootstrap unit root testing and should be carefully considered. Theoretical results show the validity of several methods of detrending, while the simulation study gives recommendations to practitioners regarding the detrending method.

Chapter 4 again demonstrates the good performance of the sieve bootstrap, now in the multivariate setting of cointegration testing, as we propose a sieve bootstrap
test for cointegration in a conditional error correction model. The proposed test is shown to be asymptotically valid, and simulations confirm its good finite sample performance, especially relative to the asymptotic test it is based on.

In Chapter 5 we advocate the use of the block bootstrap in panel data, despite the good performance of the sieve bootstrap in previous chapters. This is again illustrative of the fact that the bootstrap should not be applied blindly and that it depends on the setting which method should be used. Using theoretical arguments we show that the block bootstrap is valid for a wide range of data generating processes allowing for various forms of cross-sectional dependence. A simulation study confirms that our block bootstrap method performs adequately in finite samples for most settings considered.

In all the chapters it is the combination of theoretical and simulation results that allows us to draw conclusions and give recommendations to practitioners. Therefore, aside from their theoretical importance, the results contained within this thesis will also be useful to practitioners who wish to apply these methods in empirical applications.

While this thesis contributes to the understanding of the use of bootstrap methods in the analysis of nonstationary time series, it is far from complete. Here we will explore some of the limitations and indicate how the work in this thesis can be extended.

At this moment there are hardly any results on asymptotic refinements for bootstrap methods applied to nonstationary time series.\(^1\) This is the main difference with the bootstrap for stationary time series, where the theory of asymptotic refinements is well developed. Asymptotic refinements are interesting by themselves, but are especially useful in the comparison of different bootstrap methods. They could give us insight on the comparison of sieve versus block bootstrap methods, residuals versus differences and so on. The reason for this lack of results is that the asymptotic expansions needed for this kind of analysis are extremely difficult to obtain for nonstationary time series.

Another area where these asymptotic expansions would be useful is the theoretical analysis of block length selection for block bootstrap methods. As explained in Chapter 5, theoretical results that have led to selection methods for stationary time series, are lacking for nonstationary time series. In order to analyze block length selection properly these results are needed.

There are other unsolved issues. One of these issues is whether bootstrapping should be done under the null hypothesis (which often translates into “difference-based”) or under the alternative hypothesis (“residual-based”). This is an issue that cannot be solved with this thesis. Theoretical arguments favor bootstrapping under the alternative hypothesis, but small sample results presented in Chapter 2 and 4 do not seem to agree fully with the theoretical arguments. As such this remains an open question.

The work contained within this thesis can be extended in several other directions. For example, in Chapter 2 we restrict ourselves to Dickey-Fuller type unit

\(^1\)As mentioned before, the exception is Park (2003).
root tests. There are many other unit root tests that one could consider bootstrap versions of (Cavaliere and Taylor, 2009, provide one example). Moreover, one could look at tests that have been proposed for more general settings, such as tests that allow for general forms of heteroscedasticity or nonlinearities.

Also, our analysis of detrending in the context of bootstrap unit root testing in 3 can be extended to the bootstrap cointegration test discussed in Chapter 4 and (even more interesting) to the panel data setup of Chapter 5. It is known from the literature that the inclusion of deterministic components in nonstationary panel data complicates the analysis (see Chapter 5, Remark 5.3). It would therefore be interesting to see how this affects the bootstrap tests.

Moreover, the bootstrap techniques considered within this thesis can be extended to other settings within the framework of nonstationary time series. One interesting extension is to consider inference on cointegrating parameters. Some work has already been done on inference on cointegrating regressions, but on inference in (conditional or vector) error correction models little is known regarding the bootstrap. The framework considered in Chapter 4 could be extended to allow for inference.

In the panel setting of Chapter 5 many extensions are possible. First, the use of the bootstrap allows for many test statistics that are unfeasible with asymptotic inference. The bootstrap test based on the median of the individual unit root statistics that we considered in the simulation study of Chapter 5 is an example of a test statistic that is feasible due to the bootstrap. This holds in general for order and quantile statistics of the individual unit root statistics. Such test statistics could be used to obtain a more complete picture of the structure of the panel. One could for example compare the mean and median of the individual statistics to analyze heterogeneity. Similarly one could look at the difference between the maximum and the minimum, or at the difference between a high and low quantile. All these could provide additional information that would be extremely difficult to analyze statistically using asymptotic methods.

Another extension in the panel direction would be to extend the methods to the analysis of panel cointegration, which includes both testing for cointegration and doing inference on cointegration parameters. For example, one could extend the conditional error correction model of 4 to a panel setup, in order to test for cointegration. As the cross-sectional dependence becomes even more crucial and complex here, the bootstrap is again the ideal tool to deal with it in this setting.

An extension that can be considered in all the settings discussed above, is the analysis of structural breaks, nonlinear models and long memory. If structural breaks, other nonlinearities or long memory are present in the data, limit distributions become distorted. The bootstrap can be used to correct for these. Alternatively, the bootstrap could be used to test for structural breaks, nonlinearities or long memory in (or versus) nonstationary time series. This has been done in applications before, and appears to work quite successfully. It could therefore be interesting to develop a theoretical framework for this kind of analysis.

Concluding, we can say that there is still much more that we do not know than that we know about the application of bootstrap methods to nonstationary time
series. Because of the recent successful applications, among which those in this thesis can be counted, there is little doubt that this field will keep evolving and expanding rapidly, both in theoretical and applied research. Hopefully the results in this thesis can contribute to both.
Bibliography


Bibliography


BIBLIOGRAPHY


De analyse van niet-stationaire tijdreeksen is een van de belangrijkste onderzoeksgebieden binnen de tijdreekseconometrie. Een tijdreeks is stationair als de eigenschappen van de tijdreeks niet veranderen over de tijd. Voor veel belangrijke (macro-)economische tijdreeksen geldt dit niet, omdat deze vaak bijvoorbeeld een trend of cyclisch gedrag vertonen. Tijdreeksen voor variabelen als het bruto binnenlands product, nationaal inkomen, inflatie, wisselkoersen en aandelenmarkten zijn niet stationair en kunnen beschreven worden aan de hand van een specifiek type niet-stationariteit. Voor deze reeksen geldt dat het verschil van het ene jaar ten opzichte van het vorige jaar, of hun groei ten opzichte van het vorige jaar, stationair is. Zulke tijdreeksen worden ook wel geïntegreerde tijdreeksen genoemd. Men zegt ook dat zulke reeksen een eenheidswortel bevatten.

Geïntegreerde tijdreeksen bevatten een trend die verantwoordelijk is voor het verloop van de reeks op lange termijn. Daarom is dit een natuurlijk concept om veel economische variabelen mee te beschrijven. De analyse van geïntegreerde tijdreeksen vergt andere technieken en methoden dan de analyse van “standaard” stationaire tijdreeksen. Een belangrijk concept in dit geheel is cointegratie. Cointegratie tussen twee tijdreeksen vindt plaats als de tijdreeksen dezelfde trend delen. Het gevolg hiervan is dat de reeksen met elkaar verbonden zijn op lange termijn, maar niet noodzakelijk op korte termijn. Dit soort verband tussen tijdreeksen is een uitstekende beschrijving van de verbanden tussen veel economische tijdreeksen, zoals bijvoorbeeld inkomen en consumptie. De Nobel Prijs voor de Economie die Clive Granger (samen met Robert Engle) in 2003 ontving voor zijn werk over de analyse van niet-stationaire tijdreeksen, onderstreept het belang van dit onderwerp voor de economische wetenschappen.

In dit proefschrift worden methodes bekeken voor het toetsen op eenheidswortels en cointegratie. De bijdrage van het proefschrift ligt in de ontwikkeling en verbetering van methodes voor de analyse van niet-stationaire tijdreeksen door het gebruik van een andere statistische techniek dan die normaal gebruikt wordt. Deze techniek wordt de bootstrap genoemd, wat letterlijk laarzenstrop betekent. De naam is afgeleid van de Engelse uitdrukking “to pull one self up by his own bootstraps”, en uiteindelijk van de beroemde verhalen van Baron von Münchhausen.
die beweerde dat hij zichzelf uit een moeras omhoog trok aan zijn eigen laarzen.

De bootstrap kan gebruikt worden als alternatief voor standaard asymptotische statistische analyse. In asymptotische analyse wordt de onbekende verdeling van een toetsingsgrootte benaderd door de grootte van de steekproef naar oneindig te laten gaan. Op die manier kan men, vaak door middel van een vorm van de centrale limietstelling, de asymptotische verdeling van de toetsingsgrootte afleiden.

Het principe achter de bootstrap is dat de relatie tussen de (onbekende) populatie en de steekproef wordt benaderd door de relatie tussen de steekproef en nieuwe steekproeven die kunnen worden getrokken uit de oorspronkelijke steekproef. In principe beschouwt men de oorspronkelijke steekproef dan als de populatie, waaruit men door middel van trekken met terugleggen nieuwe steekproeven kan maken. De toetsingsgrootte waarin men geïnteresseerd is kan dan worden berekend voor elk van deze nieuwe steekproeven. Door dit voor een groot aantal nieuwe steekproeven te doen, kan men de bootstrap verdeling verkrijgen, waarmee de verdeling van de toetsingsgrootte benaderd kan worden. De bootstrap is echter niet altijd geldig, en daarom moet deze geldigheid eerst aangetoond worden voordat de bootstrap gebruikt kan worden.

De bootstrap heeft twee voordelen ten opzichte van asymptotische analyse. Ten eerste leidt de bootstrap verdeling, onder bepaalde voorwaarden, tot een preciesere benadering van de verdeling dan de asymptotische verdeling. Ten tweede kan de bootstrap gebruikt worden als de asymptotische verdeling niet gebruikt kan worden omdat deze afhangt van onbekende parameters. In dit geval is analyse door middel van de bootstrap robuuster dan asymptotische analyse. Beide punten worden in dit proefschrift benut.

De bootstrap is oorspronkelijk niet bedoeld voor de analyse van tijdreeksen. Er zit een logische structuur in een steekproef die is samengesteld uit waarnemingen in verschillende tijdsperiodes, en bovendien hangen de waarnemingen in opeenvolgende periodes vaak van elkaar af. Deze afhankelijkheid maakt het ongeldig om nieuwe steekproeven te trekken met terugleggen uit de originele steekproef. Er zijn echter varianten bedacht die geldig zijn voor tijdreeksen. De twee die in dit proefschrift (en ook algemeen het meest) gebruikt worden zijn de block bootstrap en sieve bootstrap. De block bootstrap methode trekt niet individuele waarnemingen met terugleggen, maar trekt blokken van opeenvolgende waarnemingen met terugleggen om de nieuwe steekproef te maken. In zo’n blok blijft de structuur van de tijdreeks behouden. De sieve bootstrap filtert de afhankelijkheid uit de tijdreeks door een autoregressief model van de afhankelijkheid te schatten. Op de residuen van dit model wordt dan de standaard bootstrap gebruikt. Door het model dan weer te gebruiken om de nieuwe steekproef te bouwen, wordt de afhankelijkheid behouden.

Beide methodes zijn echter ontworpen voor het gebruik met stationaire tijdreeksen, en kunnen niet rechtstreeks gebruikt worden voor de analyse van niet-stationaire tijdreeksen als die waarover dit proefschrift gaat. In dit proefschrift worden deze methodes zodanig aangepast dat ze gebruikt kunnen worden voor de analyse van tijdreeksen met eenheidswortels en coïntegratie.

200
Hierin levert het proefschrift drie belangrijke bijdragen. Ten eerste, voor een aantal specifieke gevallen wordt de geldigheid van de bootstrap door middel van theoretische resultaten aangetoond. Ten tweede, de methodologie die in dit proefschrift toegepast wordt, creëert een algemeen kader waarin de bootstrap voor niet-stationaire tijdreeksen geanalyseerd kan worden. Ten derde, simulaties geven een beeld van de prestaties van de bootstrap methodes in kleine steekproeven. Deze bijdragen zijn niet alleen van belang voor theoretisch onderzoek, maar ook voor toegepast onderzoek, omdat dit proefschrift houvast en aanbevelingen bevat voor onderzoekers die deze methodes willen toepassen in de praktijk.

In Hoofdstuk 2 worden verschillende bootstrap methodes geanalyseerd die zijn ontworpen voor het toetsen op eenheidswortels. De toetsen verschillen op een aantal punten, en voor elk van deze verschillen wordt er bekeken welke mogelijkheden het best functioneren. Een van de conclusies is dat de sieve bootstrap over het algemeen te prefereren is boven de block bootstrap. Om deze reden wordt de sieve bootstrap gebruikt in Hoofdstuk 3 en 4, maar niet in Hoofdstuk 5 omdat deze conclusie niet geldig is in de opzet van dat hoofdstuk.

Hoofdstuk 3 bouwt verder op de resultaten van Hoofdstuk 2. Aan de hand van een van de toetsen die het best presteerde in Hoofdstuk 2 wordt bekeken hoe men het best kan omgaan met deterministische trends in toetsen voor eenheidswortels, en dan vooral in relatie tot de bootstrap. Dit aspect van de toetsen, dat in Hoofdstuk 2 buiten beschouwing werd gelaten, is erg belangrijk voor toepassingen in de praktijk aangezien veel economische tijdreeksen naast een eenheidswortel ook een deterministische trend bevatten. Aan de hand van zowel theoretische als simulatieresultaten wordt geanalyseerd hoe bootstrap toetsen het best ontworpen kunnen worden als deterministische trends aanwezig zijn.

Toetsen op coïntegratie is het onderwerp van Hoofdstuk 4. In dit hoofdstuk wordt een sieve bootstrap toets op coïntegratie ontwikkeld en de geldigheid van deze bootstrap toets wordt theoretisch aangetoond. Ook worden door middel van simulaties de eigenschappen van de toets in kleine steekproeven bekeken.

Toetsen op eenheidswortels is wederom het onderwerp van Hoofdstuk 5. Het verschil met de eerdere hoofdstukken is dat dit nu niet in een context van een enkele tijdreeks bekeken wordt, maar in de context van tijdreeksen voor meerdere eenheden, oftewel panel data. In dit hoofdstuk wordt aangetoond dat voor panel data, waar er zowel afhankelijkheid tussen de eenheden is als afhankelijkheid over de tijd, de block bootstrap methode de beste bootstrap techniek is. De geldigheid van de block bootstrap methode wordt theoretisch bewezen voor een groot aantal processen die afhankelijkheid in beide richtingen kunnen genereren, en aan de hand van simulaties worden de eigenschappen in kleine steekproeven geanalyseerd.

Hoofdstuk 6 bevat de conclusie van het proefschrift. De algemene conclusie van het proefschrift is dat voor de analyse van niet-stationaire tijdreeksen met eenheidswortels en coïntegratie, bootstrap technieken over het algemeen beter werken dan asymptotische technieken en veel potentieel hebben. Het blijft echter wel uiterst belangrijk om de bootstrap zorgvuldig toe te passen en de geldigheid ervan aan te tonen.
Curriculum Vitae

Stephan Smeekes was born on January 5, 1983 in Maastricht, The Netherlands. He attended high school (Gymnasium) between 1994 and 2000 at the Jeanne d’Arc College in Maastricht. Subsequently, he studied Econometrics at Maastricht University. In August 2004 he obtained his Master’s degree with distinction (cum laude).

After graduation, Stephan joined the Department of Quantitative Economics as a Ph.D. candidate in September 2004, under the supervision of Prof. dr. Jean-Pierre Urbain and Prof. dr. Franz C. Palm. The results of his research are presented in this thesis. Stephan presented his work at various international conferences and parts of this thesis are published or forthcoming in international refereed academic journals.