Bootstrap unit root tests: a Monte Carlo investigation

CJL van Heerden

orcid.org/0000-0003-4007-8622

Dissertation submitted in partial fulfilment of the requirements for the degree *Master of Science* in Statistics at the North-West University

Supervisor: Prof FC van Graan

Graduation May 2018

22675752

NWU®
NORTH-WEST UNIVERSITY
NOORDWES-UNIVERSITEIT
UNIBESITI YA BOKONE-BOPHIRIMA

Bootstrap unit root tests: a Monte Carlo investigation

Carl J.L. van Heerden
North-West University
Potchefstroom

November 2017

Acknowledgements

Without all the extraordinary people in my life this study would not have been possible. I wish to express special appreciation and gratitude to the following individuals.

Our Heavenly Father, thank You for giving me the ability and determination to finish this study. Dr Charl Pretorius, thank you for having the answers when I did not. Your assistance in perfecting this study is invaluable and I will forever be grateful. To my colleagues and friends in the department, in particular Prof Hennie Venter, Prof James Allison and Mr Marius Smuts, our discussions and your words of encouragement are highly valued. To my family, thank you for your unwavering support in all my academic endeavours. Nataniël le Roux, your constant presence has shaped me in more ways than you could ever imagine. For this I will forever be your slave. A final word of thanks goes to the National Research Foundation of South Africa (Grant Number 108863) for generous financial support.

Abstract

In this study we investigate the finite-sample performance of some of the existing bootstrap unit root tests in terms of size and power by means of an extensive Monte Carlo study. Comparing the performance of these bootstrap tests to that of popular parametric unit root tests, such as the Dickey–Fuller and Elliott–Rothenberg–Stock tests, we find that the bootstrap-based unit root tests are less prone to size distortions frequently caused by autoregressive and moving average components in the innovations of the data generating process. Moreover, some of the bootstrap tests also seem to be more robust against forms of conditional heteroskedasticity in the innovations. In terms of power the performance of the bootstrap unit root tests are comparable to that of the well-established parametric tests. Furthermore, we discuss some practical considerations valuable in the implementation of these bootstrap unit root tests, such as an optimal method of selecting the lag parameter required in most unit root tests. We provide the reader with a practical field guide documenting in detail the most successful bootstrap unit root testing procedures, which are scattered across 20 years of statistical literature. In addition, we provide an informative review of the relevant statistical literature, which includes discussions on bootstrap procedures for independent and dependent data. The study is concluded with an application of the considered unit root tests to real-world data.

Keywords: bootstrap, dependent data, Dickey–Fuller test, stationarity, unit root tests.

Contents

Chapter 1

Introduction

Today econometrics forms an essential part of the applied statistician's toolbox. With the introduction of stricter regulations such as Basel III and IFRS 9 in the financial industry, analysts are forced to increasingly rely on analytic tools offered by econometric theory and to apply these tools responsibly. Examples of where econometric theory plays an important role are macroeconomic modelling, forecasting of macroeconomic indicators, loss forecasting and stress testing.

Owing to the dramatic development of econometric theory over the last few decades, it is not surprising that many practitioners are unaware of the numerous pitfalls often encountered. One example is the problems ecountered when working with nonstationary data. The problem with nonstationary data is that they are unpredictable and cannot be used for forecasting. In order to obtain trustworthy results, one possibility is to transform the data to be stationary. Examples of such transformations include differencing or detrending or a combination of the two, and the choice thereof depends on the nature of the individual time series model. The first objective, however, is to identify the type of nonstationarity so that an appropriate transformation can be applied.

Following the seminal article of Granger and Newbold (1974), econometricians realised that, in fact, most macroeconomic variables were nonstationary and that insufficient attention was being paid to trending behaviour (see e.g. Davidson, Hendry, Srba and Yeo, 1978; Hendry and Mizon, 1978; Plosser and Schwert, 1978; Nelson and Plosser, 1982; Bhattacharya, Gupta and Waymire, 1983; Phillips, 1986). This is also observed by Durlauf and Phillips (1988) who make the following comment:

> *"Traditional analyses of economic time series frequently rely on the assumption that the time series in question are stationary, ergodic processes [...]. However, the assumptions of the traditional theory do not provide much solace to the empirical worker. Even casual examination of such time series as GNP reveals that these series do not possess constant means."*

Nonstationary processes is a broad class containing, for example, trend stationary and unit root processes. Both these types of processes exhibit trending behaviour, the former

exhibiting a *deterministic* trend and the latter a *stochastic* trend. In the context of time series regression, Granger and Newbold (1974) show that either of these types of trending behaviour can give rise to what is referred to as *spurious* or *nonsense* regressions. The earliest indication of econometricians' awareness of spurious regression dates back to the work of Yule (1897) and Pearson (1897). Practitioners became increasingly aware that most time series are trended and hence in most cases nonstationary. This nonstationarity has an effect on standard ordinary least squares (OLS) regression procedures which can easily lead to incorrect conclusions. The term spurious regression is defined in Phillips (1986) as occurring "when a pair of independent series, but with strong temporal properties, is found apparently to be related according to standard inference in a least squares regression" (see also Asteriou and Hall, 2006, p. 293).

For the detection of spurious regressions, Granger and Newbold (1974) propose a "rule of thumb". According to them one warning sign is a very large coefficient of determination R^2 accompanied by a very small Durbin-Watson (autocorrelation) statistic, which we denote by DW. That is, if $R^2 > DW$ or if $R^2 \approx 1$ then the regression might be spurious. Ventosa-Santaulària (2009) warns, however, that when long memory is present in the variables, the regression might still be spurious even if $R^2 < DW$ and offers useful procedures of pretesting the series in order to identify the nature of the trending mechanism and potential difficulties that could be faced by the practitioner.

From the above discussion it is clear that the classification of a time series as being stationary or nonstationary is paramount, and subsequent analysis relies heavily on this classification. For the identification of stationary processes, there exists a great deal of literature. On the testing for the presence of a unit root in model-based time series, one of the most popular tests is due to David Dickey and Wayne Fuller (1979, 1981) and another popular test is due to Peter Phillips and Pierre Perron (1988). The Dickey–Fuller and Phillips–Perron *unit root tests* are used to test the null hypothesis that a time series has a unit root against the alternative of stationarity. Many other more advanced and refined tests have been developed in the literature, some of which we will discuss in this study. We note that there also exist so-called *stationarity tests* which test the null hypothesis of stationarity against an alternative of nonstationarity, of which the most commonly used is the KPSS test due to Kwiatkowski, Phillips, Schmidt and Shin (1992). For other examples, see Cappuccio and Lubian (2006). However, this type of tests will not be considered in this study as the focus is solely on unit root tests.

It is shown in various works (see e.g. De Angelis, Fachin and Young, 1997; Park, 2003; Palm, Smeekes and Urbain, 2008) that the bootstrap counterparts of the unit root tests seem to be superior to the asymptotic tests in the finite-sample setting. Their empirical sizes are closer to the nominal ones and they deliver rejection rates under the alternative hypothesis at least as high as those obtained from the asymptotic tests. Phillips and Xiao (1998) also note that the bootstrap Dickey–Fuller test has basically the same power as the

asymptotic Dickey–Fuller test, except for some non-Gaussian distribution cases where the bootstrap counterparts are slightly superior.

In this study we will investigate by means of a simulation study the exact size and power of various bootstrap unit root tests. We will use these results to compare the bootstrap unit root tests to each other and also to the most prominent asymptotic unit root tests. The precise objectives are given in the next section.

Objectives

The main objectives of this study are as follows:

- review basic time series concepts which are necessary for presenting and discussing existing asymptotic and bootstrap unit root tests;
- review the bootstrap in general and how the traditional bootstrap for independent data extends to the context of dependent data;
- provide a review of the most popular asymptotic unit root tests as well as their most important properties;
- study and document, in one chapter, the most successful bootstrap unit root tests, providing the detailed procedure of each test;
- provide practical guidance for the implementation of each of the bootstrap unit root tests considered, such as the choice of associated parameters;
- conduct a detailed Monte Carlo study to investigate and compare the finite-sample performance of the considered asymptotic and bootstrap unit root tests in terms of size and power;
- illustrate the application of all considered tests to real-world data.

Outline

In Chapter 2 we review the basic time series concepts which are necessary for presenting and discussing existing asymptotic and bootstrap unit root tests. In Chapter 3 we provide a detailed literature review of the most popular asymptotic unit root tests and discuss the most important properties associated with each of them. Chapter 4 is a review of the bootstrap applied to the case of independent data with examples of how it is frequently employed, as well as the various procedures for applying the bootstrap to dependent data. The implementation of the bootstrap to unit root testing is considered in Chapter 5. Chapter 6 contains a detailed Monte Carlo study where we investigate and compare the finite-sample performance of the considered asymptotic and bootstrap unit root tests, and Chapter 7 concludes our study with an empirical application of the considered tests to real-world data.

Frequently used notation

We define the following notation, which we will use throughout the text:

- $\mathbb{Z}$ denotes the set of all integers.
- $\mathbb{Z}^+$ denotes the set of all positive integers.
- $\mathbb{Z}^m = \{(a_1, a_2, \ldots, a_m) : a_i \in \mathbb{Z}\}$, the set of m-dimensional vectors of integers.
- $\mathbf{1}$ denotes the indicator function. For example, $\mathbf{1}(x \in A)$ is equal to 1 if $x \in A$ and 0 otherwise.
- $\lfloor x \rfloor$ denotes the largest integer less than or equal to $x \in \mathbb{R}$.
- $\lceil x \rceil$ denotes the smallest integer greater than or equal to $x \in \mathbb{R}$.

Chapter 2

Time series analysis

A *time series* is a sequence of random variables $(y_1, y_2, \ldots)$ observed over time, usually at equally spaced points. In time series analysis the observations are no longer assumed to be mutually independent and identically distributed (i.i.d.), but generated by a process describing the joint probabilistic structure of the random variables. This complicates data analysis as many traditional statistical methods are developed for the case of independent data.

Fortunately, time series analysis has a rich history which dates back to over a century.

In this chapter we review some of the theoretical aspects of time series processes which form the basis of our analysis of existing unit root testing procedures. We start by defining some basic concepts and then move on to define autoregressive moving average processes and associated properties of these stationary processes. In Section 2.3 we consider nonstationary processes, which leads to the introduction of so-called *unit root processes*, a special class of nonstationary processes.

2.1 Mean, variance and covariance functions

For a time series process $\{y_t : t = 0, \pm 1, \pm 2, \ldots\}$, we define the *mean function* by

$$\mu_t = \mathrm{E}(y_t) \quad \text{for} \quad t = 0, \pm 1, \pm 2, \ldots.$$

That is, μ_t is the expected value of the process at time t. Generally μ_t can be different at each time point t.

The *autocovariance function* of the process $\{y_t\}$ is defined as

$$\gamma_{t,s} = \mathrm{Cov}(y_t, y_s) \quad \text{for} \quad t, s = 0, \pm 1, \pm 2, \ldots,$$

where $\mathrm{Cov}(y_t, y_s) = \mathrm{E}[(y_t - \mu_t)(y_s - \mu_s)] = \mathrm{E}(y_t y_s) - \mu_t \mu_s$. Note that

$$\gamma_{t,t} = \mathrm{Cov}(y_t, y_t) = \mathrm{Var}(y_t) \quad \text{for} \quad t = 0, \pm 1, \pm 2, \ldots.$$

Similarly, the *autocorrelation function* (ACF) is defined as

$$\rho_{t,s} = \mathrm{Corr}(y_t, y_s) \quad \text{for} \quad t, s = 0, \pm 1, \pm 2, \ldots,$$

where

$$\text{Corr}(y_t, y_s) = \frac{\text{Cov}(y_t, y_s)}{\sqrt{\text{Var}(y_t)\text{Var}(y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}.$$

Both the covariance and correlation are measures of the strength of the linear relationship between two random variables, but the correlation is a standardised, unitless quantity which facilitates interpretation. The following are standard, yet important properties associated with the covariance and correlation functions. Both are symmetrical in their arguments, that is,

$$\gamma_{t,s} = \gamma_{s,t} \quad \text{and} \quad \rho_{t,s} = \rho_{s,t},$$

and, furthermore,

$$\rho_{t,t} = 1 \quad \text{and} \quad |\rho_{t,s}| \leq 1.$$

When $|\rho_{t,s}|$ is close to unity it is an indication of a *strong linear relationship*. Values close to zero indicate a *weak linear relationship*. We say y_t and y_s are *uncorrelated* if $\rho_{t,s} = 0$.

It is sometimes useful to consider the autocorrelation between $\{y_t\}$ and $\{y_{t+k}\}$ with the linear dependence of $\{y_t\}$ on $\{y_{t+1}\}, \{y_{t+2}\}, \ldots, \{y_{t+k-1}\}$ removed. This is referred to as the *partial autocorrelation function* (PACF) which is defined in Wei (2006, p. 11) as the conditional correlation

$$\phi_{kk} = \text{Corr}(y_t, y_{t+k} | y_{t+1}, \ldots, y_{t+k-1}),$$

for $t = 0, \pm 1, \pm 2, \ldots$ and $k = 1, 2, \ldots$. The exact expression for the PACF is derived in Wei (2006, p. 14) and is given by

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}},$$

where $|\mathbf{A}|$ denotes the determinant of a matrix $\mathbf{A}$.

Before proceeding, we deem it necessary to provide the definition of a stationary time series. The concept of stationarity plays a very important role in time series analysis and, as we shall see later, influences the way time series should be analysed. Patterson (2011, p. 8) defines a strictly stationary process as follows:

Definition 2.1. We say that a process $\{y_t\}$ is mth-*order strictly stationary* if

$$F(y_{r+t_1}, y_{r+t_2}, \ldots, y_{r+t_m}) = F(y_{s+t_1}, y_{s+t_2}, \ldots, y_{s+t_m}),$$

for any vector $(t_1, t_2, \ldots, t_m) \in \mathbb{Z}^m$ such that $t_1 < t_2 < \cdots < t_m$, where $r \neq s$ and $F(\cdot)$ is the joint distribution function of $(y_{t_1}, y_{t_2}, \ldots, y_{t_m})$. That is, the joint distribution function of the

process observed at times $(r+t_1, r+t_2, \ldots, r+t_m)$ is the same for any fixed length shift in the time index from r to s, and hence implies that it does not matter which portion of fixed length of the sequence we observe.

A weaker form of stationarity is weak or second-order stationarity which is defined below.

Definition 2.2. A process $\{y_t\}$ is *weakly stationary* if it satisfies the following three conditions for arbitrary $t \neq s$:

(i) $\mathrm{E}(y_t) = \mathrm{E}(y_s) = \mu$,

(ii) $\mathrm{Var}(y_t) = \mathrm{Var}(y_s) = \sigma^2$, and

(iii) $\mathrm{Cov}(y_t, y_{t+k}) = \mathrm{Cov}(y_s, y_{s+k}) = \gamma_k$.

That is, the mean and the variance of the process are constant over time (i.e. time-invariant), and the kth order autocovariance is invariant to an arbitrary shift in the time origin (i.e. depends only on the lag length k).

In practical time series analysis the conditions of weak stationarity are much easier to verify than those of strict stationarity. Hence much of the techniques developed to analyse time series are based on the concept of weak stationarity.

2.2 Models for stationary time series

This section presents the basic concepts of the most broad class of stationary time series models, the autoregressive moving average models as discussed in Cryer and Chan (2008, pp. 55-80). We first view it necessary to discuss the linear representation of weakly stationary processes from which the moving average process follows as a special case. We then move on to the autoregressive process and finally discuss the mixed autoregressive moving average model. Throughout, let $\{y_t\}$ denote the observed time series and $\{\varepsilon_t\}$ an unobserved white noise series (in the weak sense). That is, $\{\varepsilon_t\}$ is a sequence of uncorrelated random variables with $\mathrm{E}(\varepsilon_t) = 0$ and $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$.

2.2.1 Linear process representation

An important result pertaining to weakly stationary processes is the fact that any weakly stationary process can be written as a linear process. That is, a weakly stationary process $\{y_t\}$ can be written as the sum of two time series, one deterministic and one stochastic. Formally,

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \forall t \in \mathbb{Z}, \tag{2.1}$$

where μ is a constant mean, $\{\varepsilon_t\}$ is a zero-mean white noise process and $\boldsymbol{\psi}$ is the possibly infinite vector of moving average weights which are square summable, i.e. $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, and $\psi_0 = 1$ (Bühlmann, 1995b; Wei, 2006, p. 23).

The form in (2.1) is called the moving average representation of a process or Wold's representation. Using the lag operator B, that is, $B^k x_t = x_{t-k}$ and defining $\dot{y}_t = y_t - \mu$ for all t, the process in (2.1) can be written in compact form as

$$\dot{y}_t = \psi(B)\varepsilon_t,$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$.

The following results hold for the process in (2.1):

$$\mathrm{E}(y_t) = \mu,$$

$$\mathrm{Var}(y_t) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2,$$

and

$$\mathrm{Cov}(\varepsilon_t, y_{t-j}) = \mathrm{E}(\varepsilon_t y_{t-j}) = \begin{cases} \sigma_\varepsilon^2, & j = 0, \\ 0, & j > 0. \end{cases}$$

Hence,

$$\begin{aligned} \gamma_k &= \mathrm{E}(\dot{y}_t \dot{y}_{t+k}) \\ &= \mathrm{E}(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \varepsilon_{t-i} \varepsilon_{t+k-j}) \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}, \end{aligned}$$

and

$$\begin{aligned} \rho_k &= \frac{\sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sqrt{[\sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2]^2}} \\ &= \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}. \end{aligned}$$

What remains to be shown in order for (2.1) to be stationary, is that γ_k and ρ_k, since they are functions of the lag k only but involve infinite sums, are finite. This is proved by Wei (2006, p. 24) by applying the Cauchy-Schwarz inequality as follows:

$$|\gamma_k| = |\mathrm{E}(\dot{y}_t \dot{y}_{t+k})| \leq \sqrt{\mathrm{Var}(\dot{y}_t)\mathrm{Var}(\dot{y}_{t+k})} = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2,$$

and hence the assumption $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ was made in order for (2.1) to be stationary.

2.2.2 Moving average processes

Considering the moving average representation discussed in the previous subsection, when a finite number of the moving average weights ψ are nonzero, that is, when

$$\psi_k = \begin{cases} -\theta_k, & k = 1, 2, \ldots, q, \\ 0, & k > q, \end{cases}$$

and assuming a process mean $\mu = 0$, the process takes the form

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}, \tag{2.2}$$

where $(\theta_1, \theta_2, \ldots, \theta_q)$ are unknown parameters and $\{\varepsilon_t\}$ is a zero-mean white noise process. This process in (2.2) is called a *moving average process of order q* with the name abbreviated to MA(q). The process is referred to as moving average since $\{y_t\}$ is obtained by applying the weights $(1, -\theta_1, -\theta_2, \ldots, -\theta_q)$ to the variables $(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q})$ and then moving the weights to obtain $\{y_{t+1}\}$ by applying them to $(\varepsilon_{t+1}, \varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q+1})$ and so forth. Using the lag operator B the process can be written as

$$y_t = \theta_q(B)\varepsilon_t,$$

where

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q.$$

Without loss of generality a nonzero process mean may be assumed, i.e. $\mu \neq 0$, in which case y_t is replaced with $\dot{y}_t = y_t - \mu$ for all t. We, however, continue with the specification $\mu = 0$.

A finite MA process is always stationary since $1 + \theta_1^2 + \cdots + \theta_q^2 < \infty$. Such an MA process is invertible if the roots of $\theta_q(B) = 0$ lie outside the unit circle (i.e. if the q roots of $\theta_q(B) = 0$ each exceed one in absolute value).

For the general MA(q) process in (2.2) the expressions for the autocovariance and autocorrelation functions follow directly from those of the moving average representation and are given by Wei (2006, p. 52) as follows:

$$\gamma_0 = \mathrm{Var}(y_t) = \sigma_\varepsilon^2 \sum_{j=0}^{q} \theta_j^2,$$

assuming $\theta_0 = 1$,

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q) & \text{for } k = 1, 2, \ldots, q, \\ 0 & \text{for } k > q, \end{cases}$$

and, using the identity $\rho_k = \gamma_k/\gamma_0$ which holds for stationary processes,

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} & \text{for } k = 1, 2, \ldots, q, \\ 0 & \text{for } k > q. \end{cases}$$

The autocorrelation function can have any shape for lags up to q but "cuts off" after lag q. Wei (2006, p. 54) mentions that the PACF of a general MA(q) process tails off as a mixture of exponential decays and/or damped sine waves, which depends on the nature of the roots of $\theta_q(B) = 0$. The PACF contains damped sine waves if some of the roots are complex. These characteristics are important in identifying a suitable model to describe a given time series. The following model, the autoregressive process, provides models for autocorrelation patterns of alternative forms.

2.2.3 Autoregressive processes

A pth-*order autoregressive process* $\{y_t\}$, abbreviated as AR(p), satisfies the equation

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{2.3}$$

where $(\phi_1, \phi_2, \ldots, \phi_p)$ are unknown parameters and $\{\varepsilon_t\}$ is a zero-mean white noise process. The name autoregressive process is derived from the fact that the process is expressed as a regression on itself. The present value of $\{y_t\}$ is a linear combination of the p most recent past values of itself plus an innovation term ε_t which contains new information on the series at time t which was not explained by the past values. Hence, we assume that $\text{Cov}(\varepsilon_t, y_{t-k}) = 0, k > 0$. Using the lag operator B, the process can be written as

$$\phi_p(B) y_t = \varepsilon_t,$$

where

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p. \tag{2.4}$$

Using the assumption $\text{Cov}(\varepsilon_t, y_{t-k}) = 0, k > 0$, Wei (2006, p. 33) states that a stationary solution to (2.4) exists if the p roots of $\phi_p(B) = 0$ lie outside the unit circle and that the condition for invertibility of the process is that $\sum_{j=1}^{p} |\phi_j| < \infty$.

The expressions for the autocovariance and autocorrelation functions of the general AR(p) process, as given by Wei (2006, p. 45), are as follows:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, \qquad k \geq 1,$$

and

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \qquad k \geq 1. \tag{2.5}$$

An important recursive relationship is obtained from (2.5): substituting $k = 1, 2, \ldots, p$ into (2.5) and using $\rho_0 = 1$ and $\rho_{-k} = \rho_k$ we obtain the well-known general *Yule-Walker equations*

$$\begin{cases} \rho_1 = \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \cdots + \phi_p \rho_{p-1} \\ \rho_2 = \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1 + \cdots + \phi_p \rho_{p-2} \\ \vdots \\ \rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \phi_3 \rho_{p-3} + \cdots + \phi_p. \end{cases}$$

Numerical values for $(\rho_1, \rho_2, \ldots, \rho_p)$ can be obtained by solving these linear equations given numerical values for $(\phi_1, \phi_2, \ldots, \phi_p)$. Consequently, equation (2.5) can be used to obtain numerical values for ρ_k at any number of higher lags. This method can also be used to obtain method of moments estimators for $(\phi_1, \phi_2, \ldots, \phi_p)$. Noting that

$$\text{E}(\varepsilon_t y_t) = \text{E}[\varepsilon_t(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t)] = \text{E}(\varepsilon_t^2) = \sigma_\varepsilon^2,$$

equation (2.3) can be multiplied by y_t and taking expectation yields

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma_\varepsilon^2.$$

Using $\rho_k = \gamma_k/\gamma_0$, the equation above can be written as

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p},$$

in order to express the variance γ_0 in terms of the parameters $(\sigma_\varepsilon^2, \phi_1, \phi_2, \ldots, \phi_p)$ and the known values of $(\rho_1, \rho_2, \ldots, \rho_p)$. Explicit expressions for the ACF ρ_k are not possible in this setting, but it is known that ρ_k will be a linear combination of exponentially decaying terms and damped sine wave terms corresponding to the real and complex roots of the equation $\phi_p(B) = 0$, respectively (Wei, 2006, p. 47). The PACF ϕ_{kk} will fade after lag p which, together with the possible shapes of the ACF, is a property of the AR(p) model that is useful in model identification for a given time series (Wei, 2006, p. 47).

2.2.4 Mixed autoregressive moving average processes

A more general time series model is obtained when it is assumed that the series consists of a partly autoregressive and partly moving average component. In general, we say $\{y_t\}$ is a mixed *autoregressive moving average process* of orders p and q, respectively, abbreviated as ARMA(p, q), if

$$\phi_p(B)y_t = \theta_q(B)\varepsilon_t,$$

where

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p,$$

and

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q.$$

That is, the model satisfies the equation

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p} = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q},$$

or, by writing the autoregressive terms on the right side of the equation,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}, \tag{2.6}$$

where $(\phi_1, \phi_2, \ldots, \phi_p)$ and $(\theta_1, \theta_2, \ldots, \theta_q)$ are unknown parameters and $\{\varepsilon_t\}$ is a zero-mean white noise process.

Assuming that $\phi_p(B) = 0$ and $\theta_q(B) = 0$ share no common roots, the general ARMA(p, q) model in (2.6) is stationary if all the roots of $\phi_p(B) = 0$ lie outside the unit circle and (2.6) is invertible if all the roots of $\theta_q(B) = 0$ lie outside the unit circle (Wei, 2006, p. 57).

The expressions for the autocovariance and autocorrelation functions of the general ARMA(p, q) process are derived by Wei (2006, p. 58) as follows:

Multiply (2.6) by y_{t-k} on both sides

$$y_{t-k}y_t = \phi_1 y_{t-k}y_{t-1} + \cdots + \phi_p y_{t-k}y_{t-p} + y_{t-k}\varepsilon_t - \theta_1 y_{t-k}\varepsilon_{t-1} - \cdots - \theta_q y_{t-k}\varepsilon_{t-q},$$

and take the expected value to obtain

$$\gamma_k = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p} + \mathrm{E}(y_{t-k}\varepsilon_t) - \theta_1\,\mathrm{E}(y_{t-k}\varepsilon_{t-1}) - \cdots - \theta_q\,\mathrm{E}(y_{t-k}\varepsilon_{t-q}).$$

Noting that

$$\mathrm{E}(y_{t-k}\varepsilon_{t-i}) = 0, \qquad \forall k > i,$$

we have that

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p}, \qquad k \geq (q+1),$$

and, since $\rho_k = \gamma_k/\gamma_0$,

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}, \qquad k \geq (q+1).$$

Numerical computation of the autocovariance and autocorrelation functions for this model can be found in Cryer and Chan (2008, p. 85), but as an important special case used in the Monte Carlo simulations we focus on the ARMA(1,1) model next. The PACF ϕ_{kk} will be a mixture of exponential decays or damped sine waves depending on the real and complex roots of $\phi_p(B) = 0$ and $\theta_q(B) = 0$ as before (Wei, 2006, p. 59).

The ARMA(1,1) **model**

The general equation for the ARMA(1,1) model is given by

$$y_t = \phi y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, \tag{2.7}$$

where the subscripts of the coefficient parameters have been omitted for simplicity. To derive the Yule-Walker equations, note that

$$\begin{aligned} \mathrm{E}(\varepsilon_t y_t) &= \mathrm{E}[\varepsilon_t(\phi y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1})] \\ &= \sigma_\varepsilon^2, \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} \mathrm{E}(\varepsilon_{t-1} y_t) &= \mathrm{E}[\varepsilon_{t-1}(\phi y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1})] \\ &= \phi\sigma_\varepsilon^2 - \theta\sigma_\varepsilon^2 \\ &= (\phi - \theta)\sigma_\varepsilon^2. \end{aligned} \tag{2.9}$$

Multiplying (2.7) by y_{t-k} and taking the expected value yield

$$\gamma_k = \phi\gamma_{k-1} + \mathrm{E}(y_{t-k}\varepsilon_t) - \theta\,\mathrm{E}(y_{t-k}\varepsilon_{t-1}). \tag{2.10}$$

More specifically, when $k = 0$,

$$\gamma_0 = \phi\gamma_1 + \mathrm{E}(y_t\varepsilon_t) - \theta\,\mathrm{E}(y_t\varepsilon_{t-1}),$$

where expressions for the expectation terms were obtained in (2.8) and (2.9) above, so

$$\gamma_0 = \phi\gamma_1 + \sigma_\varepsilon^2 - \theta(\phi - \theta)\sigma_\varepsilon^2. \tag{2.11}$$

When $k = 1$ in (2.10), then

$$\gamma_1 = \phi\gamma_0 - \theta\sigma_\varepsilon^2. \tag{2.12}$$

Substituting (2.12) in (2.11), we obtain

$$\gamma_0 = \phi^2\gamma_0 - \phi\theta\sigma_\varepsilon^2 + \sigma_\varepsilon^2 - \phi\theta\sigma_\varepsilon^2 + \theta^2\sigma_\varepsilon^2,$$

so

$$\gamma_0 = \frac{(1 - 2\phi\theta + \theta^2)}{1 - \phi^2}\sigma_\varepsilon^2, \tag{2.13}$$

and, by substituting (2.13) in (2.12),

$$\begin{aligned} \gamma_1 &= \frac{\phi(1 - 2\phi\theta + \theta^2)}{1 - \phi^2}\sigma_\varepsilon^2 - \theta\sigma_\varepsilon^2 \\ &= \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2}\sigma_\varepsilon^2. \end{aligned}$$

When $k \geq 2$ in (2.10), then

$$\gamma_k = \phi\gamma_{k-1}, \qquad k \geq 2.$$

Summarising our results, we have

$$\gamma_k = \begin{cases} \frac{(1-2\phi\theta+\theta^2)}{1-\phi^2}\sigma_\varepsilon^2, & k = 0, \\ \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\sigma_\varepsilon^2, & k = 1, \\ \phi\gamma_{k-1}, & k \geq 2. \end{cases}$$

By using $\rho_k = \gamma_k/\gamma_0$, we have that

$$\rho_k = \begin{cases} 1, & k = 0, \\ \frac{(\phi-\theta)(1-\phi\theta)}{1-2\phi\theta+\theta^2}, & k = 1, \\ \phi\rho_{k-1}, & k \geq 2, \end{cases}$$

which can be rewritten as

$$\rho_k = \begin{cases} 1, & k = 0, \\ \frac{(\phi-\theta)(1-\phi\theta)}{1-2\phi\theta+\theta^2}\phi^{k-1}, & k \geq 1. \end{cases}$$

This autocorrelation function decays exponentially as the lag k increases. The damping factor is ϕ, but the decay starts from initial value ρ_1 which is also dependent of θ. There are several possible shapes for ρ_k which all depend on the sign of ϕ and the sign of ρ_1.

2.3 Models for nonstationary time series

Any time series with a deterministic trend is said to be nonstationary. Such models are only reasonable if it is believed that the deterministic trend is appropriate in the long run (Cryer and Chan, 2008, p. 87). For example, if a series exhibits some linear trend for an observed time window, how sure can we be that it is intrinsically part of the process and will hold

in future? Nonstationary time series can occur in numerous ways. For example, they could have time-varying means, time-varying second moments, or both of these properties (Wei, 2006, p. 68). Figure 2.1 illustrates examples of (a) a stationary series, (b) nonstationarity in the mean, (c) nonstationarity in the variance, and (d) nonstationarity in both the mean and variance. Fortunately, many stochastic trends can be modeled quite parsimoniously, as we shall see in the following discussion.

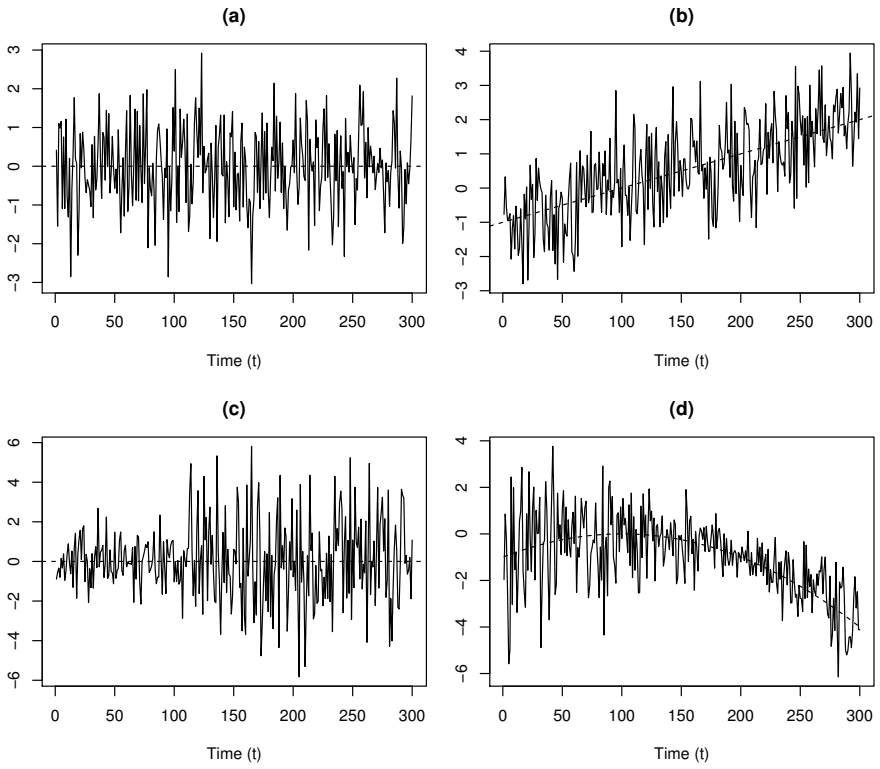

Figure 2.1: Examples of (a) a stationary series, (b) nonstationarity in the mean, (c) nonstationarity in the variance, and (d) nonstationarity in both the mean and variance.

2.3.1 Nonstationarity in the mean

A time series $\{y_t\}$ that is nonstationary in its mean implies that the mean of the series varies over time. Instead of writing, as for the stationary case,

$$\mathrm{E}(y_t) = \mu \quad \text{for} \quad t = 0, \pm 1, \pm 2, \ldots,$$

the dependence on time is indicated with a subscript t,

$$\mathrm{E}(y_t) = \mu_t \quad \text{for} \quad t = 0, \pm 1, \pm 2, \ldots.$$

Without multiple realisations of the nonstationary time series involved, the estimation of the time dependent mean can be troublesome (Wei, 2006, p. 69). There are, however, models that can be constructed from a single realisation to describe such a series. We discuss two such models next which are used to model this nonstationarity in the mean.

Deterministic trend models and detrending

In many cases the mean function of a nonstationary time series can be modelled by a deterministic function of time. That is, a standard regression model with the time-variant mean as response and time as predictor may be employed to describe such a drift. This technique is known as *detrending* the series.

In general, if μ_t follows a kth order polynomial trend, the kth *order polynomial trend model* becomes

$$x_t = \psi_0 + \psi_1 t + \cdots + \psi_k t^k + y_t,$$

where $(\psi_0, \psi_1, \ldots, \psi_k)$ are unknown parameters and $\{y_t\}$ is a zero-mean stationary process. Assuming this specification, it is easy to detrend this series to obtain the stationary series

$$\tilde{x}_t = x_t - \hat{\psi}_0 - \hat{\psi}_1 t - \cdots - \hat{\psi}_k t^k,$$

where $(\hat{\psi}_0, \hat{\psi}_1, \ldots, \hat{\psi}_k)$ are estimators for $(\psi_0, \psi_1, \ldots, \psi_k)$. These estimators may be obtained by, for example, least squares regression of y_t on $(1, t, t^2, \ldots, t^k)$.

For example, if it is reasonable to believe that the mean function μ_t follows a linear drift, then one can write

$$\mu_t = \psi_0 + \psi_1 t,$$

and the deterministic *linear trend model* becomes

$$\begin{aligned} x_t &= \mu_t + y_t \\ &= \psi_0 + \psi_1 t + y_t. \end{aligned}$$

If μ_t follows a quadratic trend, $\mu_t = \psi_0 + \psi_1 t + \psi_2 t^2$, the *quadratic trend model* can be written

$$x_t = \psi_0 + \psi_1 t + \psi_2 t^2 + y_t.$$

The deterministic trend can also follow a trigonometric sine-cosine wave which, in general, may be expressed as

$$x_t = \nu_0 + \sum_{j=1}^{m} [\eta_j \cos(\omega_j t) + \psi_j \sin(\omega_j t)] + y_t,$$

which is called the *model of hidden periodicities* (see Wei (2006, p. 70) for a discussion hereof).

All of these models are analysed using standard regression procedures. Note, however, that one may only assume such a deterministic specification if it is reasonable to believe that the series continues to behave this way outside the observed time window (this relates to the same caution needed when extrapolation in standard regression analysis is applied).

Stochastic trend models and differencing

For many nonstationary series, different parts of the series may behave very much alike except for their difference in local mean levels. This kind of nonstationary behaviour is coined by Box and Jenkins (1976, p. 85) as homogeneous nonstationarity. That is, different mean levels of a process are homogeneous in behaviour. Figure 2.2 illustrates an example of such a series with three homogeneous levels.

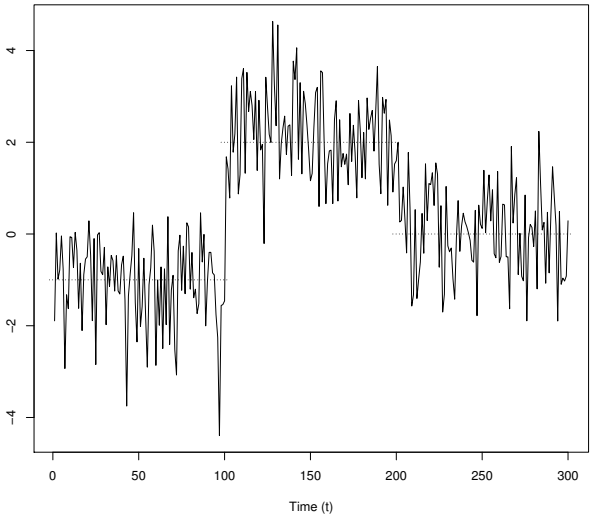

Figure 2.2: An example of a homogeneous time series with three homogeneous levels.

For ARMA models, the process is nonstationary when at least one of the roots of its AR polynomial does not lie outside the unit circle. As an example of a homogeneous nonstationary series, the local behaviour thereof is independent of its mean level. Hence, by letting $\Psi(B)$ be the generalised AR operator describing the behaviour of the process, we may write

$$\Psi(B)(y_t + c) = \Psi(B)y_t,$$

for any constant c. This representation reflects the invariance of the process' behaviour with regards to a shift in its mean level (represented by c). Note that this equation implies that $\Psi(B)$ must be of the form

$$\Psi(B) = \phi(B)(1-B)^d, \quad \text{some } d \in \mathbb{Z}^+,$$

where $\phi(B)$ is the stationary AR operator and assuming that the dth difference of $\{y_t\}$ is stationary, in order to ensure a process stationary in the mean (Wei, 2006, p. 71).

To understand why a time series with a dth order polynomial time trend is rendered stationary after taking d differences we provide a proof based on a proof by Anderson (1976, p. 111).

For some fixed $d > 0$, consider the process described by

$$x_t = \mu_t + y_t := f_0 + f_1 t + \cdots + f_d t^d + y_t,$$

where $(f_0, f_1, \ldots, f_d)$ are finite constants and $\{y_t\}$ is an unobserved error process. It is clear that this process is stationary around the deterministic trend μ_t. It can be shown that the process

$$(1-B)^d x_t = (1-B)^d(\mu_t + y_t) \tag{2.14}$$

is stationary. We now provide a proof for the case where $\{y_t\}$ is an ARIMA process, that is,

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t,$$

or, equivalently,

$$(1-B)^d y_t = \phi^{-1}(B)\theta(B)\varepsilon_t.$$

In this case (2.14) becomes

$$(1-B)^d x_t = (1-B)^d \mu_t + \phi^{-1}(B)\theta(B)\varepsilon_t.$$

Now, since $\phi^{-1}(B)\theta(B)\varepsilon_t$ is stationary, we only need to show that $(1-B)^d\mu_t$ is independent of t. Indeed, one can show that

$$(1-B)^d \mu_t = (1-B)^d(f_0 + f_1 t + \cdots + f_d t^d) = f_d d!, \qquad d \geq 1,$$

which we do with mathematical induction. For $d = 1$ one has

$$\begin{aligned} & (1-B)(f_0 + f_1 t) \\ = \; & f_0 + f_1 t - f_0 - f_1(t-1) \\ = \; & f_1 \\ = \; & f_1 1!. \end{aligned}$$

Now assume the statement holds for $d = k$, that is

$$(1-B)^k(f_0 + f_1 t + \cdots + f_k t^k) = f_k k!, \tag{2.15}$$

then we need to prove it holds for $d = k+1$. That is, we need to prove

$$(1-B)^{k+1}(f_0 + f_1 t + \cdots + f_k t^k + f_{k+1} t^{k+1}) = f_{k+1}(k+1)!.$$

Now

$$\begin{aligned} & (1-B)^{k+1}(f_0 + f_1 t + \cdots + f_k t^k + f_{k+1} t^{k+1}) \\ = \; & (1-B)^k(1-B)(f_0 + f_1 t + \cdots + f_k t^k + f_{k+1} t^{k+1}) \\ = \; & (1-B)[(1-B)^k(f_0 + f_1 t + \cdots + f_k t^k) + (1-B)^k f_{k+1} t^{k+1}] \\ = \; & (1-B) f_k k! + (1-B)^{k+1} f_{k+1} t^{k+1} \\ = \; & (1-B)^{k+1} f_{k+1} t^{k+1} \\ = \; & (1-B)^k(1-B) t^{k+1} f_{k+1} \\ = \; & (1-B)^k[t^{k+1} - (t-1)^{k+1}] f_{k+1} \\ = \; & (1-B)^k[t^{k+1} - \textstyle\sum_{j=0}^{k+1} \binom{k+1}{j} t^j (-1)^{k+1-j}] f_{k+1} \\ = \; & (1-B)^k \textstyle\sum_{j=0}^{k} \binom{k+1}{j} t^j (-1)^{k+2-j} f_{k+1} \\ = \; & \binom{k+1}{k}(-1)^2 k! f_{k+1} \\ = \; & f_{k+1}(k+1)!, \end{aligned}$$

where the second last step follows from the induction assumption in (2.15) by choosing the constants f_j as

$$f_j = \binom{k+1}{j}(-1)^{k+2-j}, \qquad j = 0, 1, \ldots, k.$$

Hence

$$(1-B)^d x_t = f_d d! + \phi^{-1}(B)\theta(B)\varepsilon_t,$$

or

$$\begin{aligned}\phi(B)(1-B)^d x_t &= \phi(B) f_d d! + \theta(B)\varepsilon_t \\ &= \phi(1) f_d d! + \theta(B)\varepsilon_t.\end{aligned}$$

By this we have shown that by taking d differences we have removed not only the dth degree deterministic trend μ_t but also the stochastic trend. □

The conclusion is thus that a homogeneous nonstationary process can be reduced to a stationary process by taking the suitable difference which yields the process stationary. That is, the series $\{y_t\}$ is nonstationary but its dth difference $\{(1-B)^d y_t\}$ is stationary for some $d \in \mathbb{Z}^+$. In other words, the mean level of $\{y_t\}$ in $\{(1-B)^d y_t\}$ changes stochastically over time, whence we characterise the process as having a stochastic trend rather than a deterministic trend which is purely a deterministic function of time.

For a discussion on spurious detrending and overdifferencing the reader is referred to Patterson (2011, p. 40). Since testing for overdifferencing is equivalent to testing for a moving average unit root, and since there is a formal correspondence between testing for a moving average unit root and testing for the null of stationarity (Zivot and Wang, 2006, p. 113), the reader may also refer to Phillips and Xiao (1998).

2.3.2 ARIMA models

The time series $\{y_t\}$ is said to be an *integrated autoregressive moving average process* if its dth difference $\{(1-B)^d y_t\}$ is a stationary ARMA process for some $d \in \mathbb{Z}^+$. We denote this by writing $\{y_t\} \sim \mathrm{I}(d)$ which reads "$\{y_t\}$ is integrated of order d". When $\{(1-B)^d y_t\}$ follows an ARMA(p,q) model, then $\{y_t\}$ is an ARIMA(p,d,q) process. It usually suffices to take $d = 1$ or $d = 2$ at most to render the process stationary.

More formally, if $\{y_t\}$ is an ARIMA(p,d,q) process, then $\{y_t\}$ will have the form

$$\phi_p(B)(1-B)^d y_t = \theta_q(B)\varepsilon_t,$$

where

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p,$$

and

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q,$$

and $\phi_p(B)$ and $\theta_q(B)$ share no common factors (Wei, 2006, p. 72). When we write

$$\phi_p(B)W_t = \theta_q(B)\varepsilon_t,$$

where $\{W_t\} = \{(1-B)^d y_t\}$, then $\{W_t\}$ is stationary for some $d \in \mathbb{Z}^+$ if all the roots of $\phi_p(B) = 0$ lie outside the unit circle and $\{W_t\}$ is invertible if all the roots of $\theta_q(B) = 0$ lie outside the unit circle (Wei, 2006, p. 57).

For all the processes discussed we assumed that the mean of the process $\mu = 0$, but for the case $\mu \neq 0$ we may write

$$\phi_p(B)(1-B)^d y_t = \theta_0 + \theta_q(B)\varepsilon_t.$$

When $d = 0$ the original process $\{y_t\}$ is stationary and θ_0 is related to the mean of the process by

$$\theta_0 = \mu(1 - \phi_1 - \cdots - \phi_p).$$

When $d \geq 1$, then θ_0 is called the *deterministic trend term* (Wei, 2006, p. 72).

It can happen that either $p = 0$ or $q = 0$, especially in time series that arise in business and economics (Cryer and Chan, 2008, p. 93). When $p = 0$, the ARIMA(p,d,q) model is called an IMA(d,q) model. When $q = 0$, the ARIMA(p,d,q) model is called an ARI(p,d) model.

2.3.3 Nonstationarity in the variance

In the previous subsection we considered the case of homogeneous nonstationary processes. Many time series, however, do not fall in this class and are called nonhomogeneous nonstationary processes (Wei, 2006, p. 80). The nonstationarity of this class of processes arises not from their time-varying means but from their time-varying variances and autocovariances. A process that is stationary in its mean is not necessarily stationary in its variance and autocovariance, but a process that is nonstationary in its mean will definitely be nonstationary in its variance and autocovariance (since they are functions of the mean) (Wei, 2006, p. 82). Differencing the series will not reduce this type of nonstationarity, and hence other transformations are required. These types of transformations are called variance stabilising transformations. One method of deriving such a variance stabilising transformation for a specific process $\{y_t\}$ is the δ-method which we discuss next.

The δ-method

It is a common occurrence for a nonstationary process to have a variance that changes with its mean level. That is, the variance of the process may be written as some function of its time-varying mean,

$$\mathrm{Var}(y_t) = cf(\mu_t),$$

for some positive constant c and function f. We need to find some function Q such that the series transformed under Q, that is, $\{Q(y_t)\}$ has a constant variance.

To illustrate the method, we approximate the desired function by a first-order Taylor expansion about μ_t as follows:

$$Q(y_t) \approx Q(\mu_t) + Q'(\mu_t)(y_t - \mu_t),$$

where $Q'(\mu_t)$ is the first derivative of $Q(y_t)$ evaluated at μ_t. Now,

$$\begin{aligned} \text{Var}[Q(y_t)] &\approx [Q'(\mu_t)]^2 \text{Var}(y_t) \\ &= c[Q'(\mu_t)]^2 f(\mu_t). \end{aligned}$$

Consequently, in order for the variance of $Q(y_t)$ to be constant, the variance stabilising transformation $Q(y_t)$ must be chosen such that

$$Q'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}}.$$

By integrating with regards to μ_t, we obtain

$$Q(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t.$$

As an example, consider $f(\mu_t) = \mu_t^2$. We then have

$$Q(\mu_t) = \int \frac{1}{\sqrt{\mu_t^2}} d\mu_t = \ln(\mu_t).$$

Hence, by transforming the series $\{y_t\}$ to $\{\ln(y_t)\}$, the transformed series will have constant variance.

A discussion of this and other variance stabilising transformations can be found in Wei (2006, p. 83) or Rice (2007, p. 161).

2.3.4 Conclusion

To summarise, we discussed models for stationary time series which included moving average processes, autoregressive processes and mixed autoregressive moving average processes. In the discussion of models for nonstationary time series we first examined nonstationarity in the mean which was divided into two classes namely deterministic trend models and stochastic trend models. A transformation technique was discussed for each of these trending behaviours; they were detrending and differencing, respectively. Stochastic trend models and differencing led to the discussion of integrated autoregressive moving average models. Lastly, we took a glance at nonstationarity in the variance and an associated variance stabilising transformation.

Chapter 3

Asymptotic unit root tests

Countless unit root tests have been developed since the first tests proposed by Dickey and Fuller (1979). Apart from these tests, the most popular tests available today are those developed by Said and Dickey (1984), Phillips and Perron (1988), Ng and Perron (1995), Elliott, Rothenberg and Stock (1996) and Phillips and Xiao (1998). Other lesser-known tests have been developed by Schmidt and Phillips (1992), Pantula, Gonzalez-Farias and Fuller (1994), Leybourne (1995), and Leybourne, McCabe and Tremayne (1996), among others.

This chapter serves as a summary of the most important unit root tests and the development thereof. As this study is mainly concerned with bootstrap-based tests, the intention of this chapter is not to provide an exhaustive discussion of all available asymptotic tests. We shall therefore focus on those asymptotic tests on which the bootstrap tests of Chapter 5 are based. Specifically we discuss the tests of Dickey and Fuller (1979), Phillips and Perron (1988), Ng and Perron (1995) and Elliott et al. (1996). We conclude the chapter with a discussion of the choice of the lag parameter required in most unit root tests.

Before moving on we first define some notation which we will be using throughout the chapter. Consider the data generating process (DGP) defined by

$$x_t = d_t + y_t, \qquad y_t = \rho y_{t-1} + u_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{3.1}$$

where d_t is a fixed deterministic component and $\{u_t\}$ is an unobserved stationary zero-mean error process. Based on a finite sample $\{x_0, x_1, \ldots, x_T\}$ generated by this DGP, our objective is to test the hypothesis

$$H_0 : \rho = 1 \quad \text{vs.} \quad H_A : |\rho| < 1. \tag{3.2}$$

If $\rho = 1$ the process in (3.1) is said to have a *unit root*.

In the literature it has become common practice to express the deterministic component d_t in (3.1) as

$$d_t = \boldsymbol{\psi}' \mathbf{z}_t,$$

where $\boldsymbol{\psi} = (\psi_0, \psi_1, \ldots, \psi_p)'$ is a set of $p+1$ finite parameters and $\mathbf{z}_t$ is a set of deterministic components. If, for example, we set $\mathbf{z}_t = (1, t, t^2, \ldots, t^p)'$, then d_t is the polynomial

$$d_t = \sum_{i=0}^{p} \psi_i t^i = \psi_0 + \psi_1 t + \psi_2 t^2 + \cdots + \psi_p t^p.$$

We will consider testing the hypothesis in (3.2) for each of the following two standard specifications of the deterministic component d_t:

- with $p = 0$ so that $d_t = \psi_0$, i.e. the process has a constant mean, and
- with $p = 1$ so that $d_t = \psi_0 + \psi_1 t$, i.e. the process follows a linear time trend.

At this point the reader should note that the process in (3.1) is often expressed in an alternative form. Observe that

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

from which it follows that

$$x_t - d_t = \rho(x_{t-1} - d_{t-1}) + u_t,$$

so that an equivalent representation of the DGP in (3.1) is given by

$$x_t = d_t^\dagger + \rho x_{t-1} + u_t = \boldsymbol{\psi}_\dagger' \mathbf{z}_t + \rho x_{t-1} + u_t, \tag{3.3}$$

where $\boldsymbol{\psi}_\dagger' \mathbf{z}_t = d_t^\dagger = d_t - \rho d_{t-1}$ represents the *drift* of the process $\{x_t\}$. For example, if $d_t = \psi_0 + \psi_1 t$, then

$$\begin{aligned} d_t^\dagger &= \psi_0 + \psi_1 t - \rho\psi_0 - \rho\psi_1(t-1) \\ &= (1-\rho)\psi_0 + \rho\psi_1 + (1-\rho)\psi_1 t. \end{aligned}$$

We now move on to discuss the asymptotic tests which are important for our purposes.

3.1 The Dickey–Fuller unit root tests

Arguably the most well known tests for testing the hypothesis in (3.2) are the classical tests developed by Dickey and Fuller (1979). They propose using either the *coefficient statistic*

$$DF_{\rho,0} := T(\hat{\beta}_0 - 1)$$

or the *t-statistic*

$$DF_{t,0} := \frac{\hat{\beta}_0 - 1}{\widehat{\mathrm{SE}}(\hat{\beta}_0)},$$

where $\hat{\beta}_0$ denotes the OLS estimator for the unknown parameter ρ in a regression of x_t on $(\mathbf{z}_t, x_{t-1})$, as in (3.3). Exact expressions for $\hat{\beta}_0$ and its estimated standard error $\widehat{\mathrm{SE}}(\hat{\beta}_0)$ depend on the specification of the deterministic component $d_t^\dagger$ and are readily available in standard texts such as Kutner, Nachtsheim, Neter and Li (2005) and Seber and Lee (2012).

Notice that $DF_{t,0}$ is the usual t-statistic used to test for significance of the slope coefficient in a simple linear regression. However, as shown in Phillips (1987), under the null hypothesis of a unit root the limiting distributions of $DF_{\rho,0}$ and $DF_{t,0}$ are not normal distributions. He shows that in the case where $d_t^\dagger = 0$, i.e. no deterministic component, and the u_t are *uncorrelated*, one has under H_0 that

$$DF_{\rho,0} \xrightarrow{\mathscr{D}} \frac{\int_0^1 W(t)dW(t)}{\int_0^1 W(t)^2 dt} \qquad \text{and} \qquad DF_{t,0} \xrightarrow{\mathscr{D}} \frac{\int_0^1 W(t)dW(t)}{\left(\int_0^1 W(t)^2 dt\right)^{1/2}}, \tag{3.4}$$

where $W(\cdot)$ denotes a standard Brownian motion process defined on the unit interval. These limiting distributions are sometimes referred to as *Dickey–Fuller distributions*.

It is well known that the distributions of $DF_{\rho,0}$ and $DF_{t,0}$ depend on the specification of the deterministic component $d_t^\dagger$ used in the test (see Dickey and Fuller, 1979; Said and Dickey, 1984; Xiao and Phillips, 1998). For example, if $d_t^\dagger$ is a linear function of time, the limiting distributions of $DF_{\rho,0}$ and $DF_{t,0}$ are the same as in (3.4), but with $W(t)$ replaced by

$$\widetilde{W}(t) = W(t) + (6t-4)\int_0^1 W(s)ds - (12t-6)\int_0^1 sW(s)ds.$$

See Paparoditis and Politis (2016, p. 6).

Quantiles of these limiting distributions have been approximated by Monte Carlo simulations and are available in texts such as Fuller (1996). In Table 3.1 we provide the asymptotic critical values as obtained using our own Monte Carlo simulations (using 3 000 000 independent iterations).

3.2 The Augmented Dickey–Fuller Unit Root Test

In addition to establishing the limiting distributions in (3.4) under the white noise assumption, Phillips (1987) also shows that if the u_t are serially correlated, the limiting null distributions of $DF_{\rho,0}$ and $DF_{t,0}$ depend on nuisance parameters. This causes practical difficulties in implementation of the tests.

For the t-statistic $DF_{t,0}$, Said and Dickey (1984) had already proposed a solution shortly after the original Dickey–Fuller tests were introduced. They show that it is possible to approximate an ARIMA$(p,1,q)$ model by an autoregressive process with order depending on the number of observations T. With this in mind they suggest using the t-statistic

$$DF_{t,k} := \frac{\hat{\beta}_0^{(k)} - 1}{\widehat{\mathrm{SE}}\left(\hat{\beta}_0^{(k)}\right)}, \tag{3.5}$$

where, for some integer $0 < k < \infty$, $\hat{\beta}_0^{(k)}$ is the OLS estimator for β_0 in the regression

$$x_t = \boldsymbol{\delta}'\mathbf{z}_t + \beta_0 x_{t-1} + \sum_{j=1}^{k} \beta_j \Delta x_{t-j} + \varepsilon_{t,k}, \tag{3.6}$$

with $\boldsymbol{\delta} = (\delta_0, \delta_1, \ldots, \delta_p)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)$ unknown parameters, $\widehat{\mathrm{SE}}(\hat{\beta}_0^{(k)})$ the estimated standard error of $\hat{\beta}_0^{(k)}$, and $\varepsilon_{t,k}$ unobserved error terms. Said and Dickey (1984) show that $DF_{t,k}$ has the same limiting distribution as $DF_{t,0}$ under the condition that $k \to \infty$ and $k = o(T^{1/3})$ as $T \to \infty$, i.e. k should increase with the sample size at a rate less than $T^{1/3}$. Chang and Park (2002) were able to establish this convergence even if this rate is relaxed to $T^{1/2}$.

Although Said and Dickey (1984) proposed a remedy for the t-statistic, they assert that even with the added lagged difference terms in the regression in (3.6) the limiting distribution of the coefficient statistic $T(\hat{\beta}_0^{(k)} - 1)$ still depends on nuisance parameters. Xiao

and Phillips (1998) demonstrate how a simple transformation using estimated parameters eliminates these nuisance parameters. They propose the *normalised* coefficient statistic

$$DF_{\rho,k} := \frac{T(\hat{\beta}_0^{(k)} - 1)}{1 - \sum_{j=1}^{k} \hat{\beta}_j}, \tag{3.7}$$

where the $\hat{\beta}_j$ are OLS estimators for the β_j in the regression in (3.6). The limiting distribution of $DF_{\rho,k}$ is the same as that of $DF_{\rho,0}$ given in (3.4). Also for this statistic Chang and Park (2002) show that the convergence is valid under the assumption that the lag length k increases with the sample size at a rate of $T^{1/2}$.

Recently Paparoditis and Politis (2016) have shown that the convergence of the ADF t-statistic $DF_{t,k}$ to its limiting distribution (stated earlier) is valid under much more general conditions that go far beyond the linear AR(∞) process assumption typically imposed. Essentially, the only requirement is that the error process $\{u_t\}$ has a continuous spectral density that is strictly positive.

An alternative solution to eliminate nuisance parameters from the asymptotic distributions of $DF_{\rho,0}$ and $DF_{t,0}$, in the case where the errors are serially correlated, was proposed by Phillips and Perron (1988) and is presented in the following section.

3.3 The Phillips–Perron unit root tests

The unit root tests developed by Phillips (1987) and Phillips and Perron (1988), henceforth referred to as the Phillips–Perron tests, differ from the ADF test mainly in how they deal with serial correlation and heteroskedasticity in the errors. Rather than using an autoregressive sieve, they eliminate nuisance parameters present in the limiting distributions of the test statistics $DF_{\rho,0}$ and $DF_{t,0}$ by transforming these statistics using consistent estimators for the nuisance parameters.

The Phillips–Perron test regression is simply the regression of x_t on $(\mathbf{z}_t, x_{t-1})$ as in (3.3). Define the residuals

$$\hat{u}_t = x_t - \hat{\beta}_0 x_{t-1} - \hat{\boldsymbol{\psi}}_\dagger' \mathbf{z}_t,$$

where $\hat{\beta}_0$ and $\hat{\boldsymbol{\psi}}_\dagger$ denote the OLS estimators for ρ and $\boldsymbol{\psi}_\dagger$ in (3.3), respectively. The transformed version of the coefficient test $DF_{\rho,0}$ proposed by Phillips and Perron (1988) is given by

$$Z_\rho = DF_{\rho,0} - T^2 \widehat{\mathrm{Var}}(\hat{\beta}_0) \left(\frac{\hat{\sigma}^2 - \hat{\sigma}_u^2}{2\hat{\sigma}_u^2} \right),$$

where $\hat{\sigma}^2$ and $\hat{\sigma}_u^2$ are consistent estimates of the parameters

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{T} \mathrm{E}\left(S_T^2\right) \qquad \text{and} \qquad \sigma_u^2 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathrm{E}\left(u_t^2\right),$$

respectively, where $S_T = \sum_{t=1}^{T} u_t$. They also propose a modified version of the $DF_{t,0}$ statistic which is defined as

$$Z_t = \frac{\hat{\sigma}_u}{\hat{\sigma}} DF_{t,0} - T \cdot \widehat{\mathrm{SE}}(\hat{\beta}_0) \left(\frac{\hat{\sigma}^2 - \hat{\sigma}_u^2}{2\hat{\sigma}\hat{\sigma}_u} \right).$$

Table 3.1: Critical values for the Dickey–Fuller tests

		Level (zero mean)				Level (constant mean)				Level (linear trend)			
Statistic	T	1%	2.5%	5%	10%	1%	2.5%	5%	10%				
$DF_{\rho,0}$	50	-12.77	-9.89	-7.71	-5.53	-18.79	-15.69	-13.24	-10.73	-25.69	-22.36	-19.68	-16.83
	100	-13.21	-10.14	-7.86	-5.62	-19.66	-16.27	-13.67	-10.98	-27.44	-23.66	-20.64	-17.49
	250	-13.52	-10.34	-7.98	-5.68	-20.28	-16.70	-13.93	-11.16	-28.57	-24.49	-21.28	-17.95
	∞	-13.68	-10.44	-8.04	-5.71	-20.62	-16.91	-14.09	-11.25	-29.35	-25.05	-21.76	-18.24
$DF_{t,0}$	50	-2.62	-2.25	-1.95	-1.61	-3.57	-3.22	-2.92	-2.60	-4.16	-3.80	-3.50	-3.18
	100	-2.59	-2.24	-1.94	-1.61	-3.50	-3.17	-2.90	-2.58	-4.05	-3.73	-3.45	-3.15
	250	-2.58	-2.24	-1.94	-1.62	-3.46	-3.14	-2.87	-2.57	-4.00	-3.69	-3.43	-3.14
	∞	-2.57	-2.23	-1.94	-1.62	-3.43	-3.12	-2.86	-2.57	-3.96	-3.66	-3.41	-3.13

As a consistent estimator for σ_u^2, Phillips and Perron (1988) prefer the mean squared residuals defined by

$$\hat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2.$$

To estimate σ^2 they suggest using the long-run estimator for the variance of u_t proposed by Newey and West (1987). This estimator is defined as

$$\hat{\sigma}_\ell^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 + \frac{2}{T}\sum_{s=1}^{\ell} w_{s\ell}\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s},$$

where $w_{s\ell}$ is the Bartlett (or triangular) window given by

$$w_{s\ell} = 1 - \frac{s}{\ell+1},$$

for some bandwidth ℓ. Other choices of the window $w_{s\ell}$ have also been proposed in the literature, such as the Parzen window. For other choices, see e.g. Newey and West (1994).

Implementation of the Phillips–Perron tests requires selecting an appropriate value for the bandwidth parameter ℓ appearing in the Bartlett window. After the numerical studies conducted by Schwert (1989) it has become common practice to use the fixed *short truncation lag*

$$\ell_4 = \left\lfloor 4\left(\frac{T}{100}\right)^{1/4}\right\rfloor$$

or the *long truncation lag*

$$\ell_{12} = \left\lfloor 12\left(\frac{T}{100}\right)^{1/4}\right\rfloor.$$

Numerical studies by Lee and Mossi (1996) show that, unlike with the ADF tests, using the optimal bandwidth selection procedure of Andrews (1991) does not offer any significant improvement over using the fixed truncation lags given above (also see Cheung and Lai, 1997). Under these considerations we will use the long truncation lag ℓ_{12} as our choice for ℓ in our Monte Carlo study.

Under H_0, the modified test statistics Z_ρ and Z_t have the same asymptotic distributions as the ADF normalised bias statistic $DF_{\rho,0}$ and the ADF t-statistic $DF_{t,0}$, respectively, given that the deterministic specification in the tests is the same. Hence, the distributional tables as reported in Fuller (1996, pp. 641–642) and Wei (2006, pp. 592–593) may be used to conduct the Phillips–Perron tests.

An advantage which makes the Phillips–Perron tests more popular than the ADF test, as stated in Zivot and Wang (2006, p. 127), is that the Phillips–Perron tests are robust to general forms of heteroskedasticity in the errors u_t. Another advantage is that the specification of a lag length for the test regression is not needed and, as discussed above, the choice of the lag truncation parameter does not have a significant impact on the performance of the tests.

3.4 The Elliott–Rothenberg–Stock unit root tests

Elliott et al. (1996) propose a family of point-optimal tests which, under very mild conditions on $\{u_t\}$, has asymptotic power curves that lie close to the Neyman–Pearson power envelope at 50% power. This family of tests, referred to by P_T, involves local-to-unity generalised least squares (GLS) detrending of $\{x_t\}$ prior to performing the usual Dickey–Fuller tests.

We now describe the testing procedure proposed by Elliott et al. (1996). Let $\bar{\alpha} = 1 + \bar{c}/T$ for some chosen scalar $\bar{c}$ and define the so-called *local GLS detrended* series version of $\{x_t\}$ by

$$\tilde{x}_t = x_t - \tilde{\boldsymbol{\psi}}' \mathbf{z}_t, \qquad t = 1, 2, \ldots, T, \tag{3.8}$$

where $\tilde{\boldsymbol{\psi}}$ is obtained by OLS regression of $x_t^{\bar{\alpha}}$ on $\mathbf{z}_t^{\bar{\alpha}}$, with

$$x_t^{\bar{\alpha}} := \begin{cases} x_1 & \text{if } t = 1, \\ x_t - \bar{\alpha} x_{t-1} & \text{otherwise,} \end{cases}$$

and

$$\mathbf{z}_t^{\bar{\alpha}} := \begin{cases} \mathbf{z}_1 & \text{if } t = 1, \\ \mathbf{z}_t - \bar{\alpha} \mathbf{z}_{t-1} & \text{otherwise.} \end{cases}$$

Based on the GLS detrended series $\{\tilde{x}_t\}$, Elliott et al. (1996) propose two test statistics: the DF_t^{GLS} statistic defined as the t-statistic for testing whether $\beta_0 - 1 = 0$ in the OLS regression

$$\Delta \tilde{x}_t = (\beta_0 - 1)\tilde{x}_{t-1} + \sum_{j=1}^{k} \beta_j \Delta \tilde{x}_{t-j} + \varepsilon_{t,k}, \tag{3.9}$$

where β_0 and the β_j are unknown parameters and $\{\varepsilon_{t,k}\}$ is an unspecified error process, and the P_T statistic defined as

$$P_T = \frac{S(\bar{\alpha}) - \bar{\alpha} S(1)}{\hat{\omega}^2}, \tag{3.10}$$

where $S(a)$ denotes the sum of squared residuals from an OLS regression of x_t^a on $\mathbf{z}_t^a$ and $\hat{\omega}^2$ is a consistent estimator for $\omega^2 = \sum_{k=-\infty}^{\infty} \mathrm{E}(u_t u_{t-k})$.

Elliott et al. (1996) suggest choosing the value of $\bar{c}$ in the above procedure such that the asymptotic power functions of the tests are tangent to the Neyman–Pearson power envelope (which they derive under normality assumptions on $\{u_t\}$). In the constant mean case (i.e. when $p = 0$) they suggest taking $\bar{c} = -7$ and in the linear trend case (i.e. when $p = 1$) taking $\bar{c} = -13.5$.

As an estimator for ω^2 they consider, among others, the autoregressive estimator

$$s_{AR}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\left(1 - \sum_{j=1}^{k} \hat{\beta}_j\right)^2}, \tag{3.11}$$

where the $\hat{\beta}_j$ are OLS estimates for the β_j' in the regression

$$\Delta x_t = c + d x_{t-1} + \sum_{j=1}^{k} \beta_j' \Delta x_{t-j} + \varepsilon_{t,k},$$

Table 3.2: Critical values for the tests proposed by Elliott et al. (1996).

		Level (constant mean)				Level (linear trend)			
Test	T	1%	2.5%	5%	10%	1%	2.5%	5%	10%
DF_t^{GLS}	50	−2.62	−2.25	−1.95	−1.61	−3.77	−3.46	−3.19	−2.89
	100	−2.60	−2.24	−1.95	−1.61	−3.58	−3.29	−3.03	−2.74
	200	−2.58	−2.23	−1.95	−1.62	−3.46	−3.18	−2.93	−2.64
	∞	−2.58	−2.23	−1.95	−1.62	−3.48	−3.15	−2.89	−2.57
P_T	50	1.87	2.39	2.97	3.91	4.22	4.94	5.72	6.77
	100	1.95	2.47	3.11	4.17	4.26	4.90	5.64	6.79
	200	1.91	2.47	3.17	4.33	4.05	4.83	5.66	6.86
	∞	1.99	2.55	3.26	4.48	3.96	4.78	5.62	6.89

with c and d unknown parameters, $\{\varepsilon_{t,k}\}$ an unspecified error process, and $\hat{\sigma}_\varepsilon^2$ denoting the mean squared residuals resulting from this regression.

Remark. Note that the autoregressive spectral density estimator s_{AR}^2 above is based on the original series $\{x_t\}$ and *not on the detrended series* $\{\tilde{x}_t\}$. Also note the presence of the constant term c regardless of whether d_t is a constant or a linear trend. Ng and Perron (2001) study the finite-sample performance of the P_T test when s_{AR}^2 is based on the GLS detrended series $\{\tilde{x}_t\}$. Despite the improvements they observe for local alternatives of the form $\rho = 1 + \bar{c}/T$, Seo (2005) shows that doing this may significantly reduce the power of these tests for *non-local* alternatives. In fact, he demonstrates that the power of these tests can even decrease as ρ moves closer to 0 (for a given sample size). Perron and Qu (2007) provide a simple solution which we discuss in the next section.

A selection of approximate critical values of the DF_t^{GLS} and P_T tests is given in Elliott et al. (1996) and is reproduced in Table 3.2.

3.5 M tests

A common drawback of standard unit root tests is the severe size distortions seen when the DGP has an AR or MA error process with a root close to the unit circle (Ng and Perron, 2001). To address this issue Perron and Ng (1996) propose modified versions of the Phillips–Perron tests described in Section 3.3. These modified tests, often termed *M tests*, are shown to have exact sizes much closer to the nominal size when used in conjunction with a particular formulation of an autoregressive spectral density estimator.

For the GLS detrended series $\{\tilde{x}_t\}$ defined in the previous section, the M tests of Ng and Perron (2001) are given by

$$MZ_\alpha = \frac{T^{-1}\tilde{x}_T^2 - s_{AR}^2}{2T^{-2}\sum_{t=1}^{T}\tilde{x}_{t-1}^2}, \qquad MSB = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{t-1}^2 / s_{AR}^2}, \qquad MZ_t = MZ_\alpha \cdot MSB,$$

with s_{AR}^2 as defined in (3.11). They also define a modification of the feasible point optimal test in (3.10), which, in the notation of Section 3.4 and, as before, p denotes the specification

of the deterministic component, is given by

$$MP_T = \left[\bar{c}^2 T^{-2} \sum_{t=1}^{T} \tilde{x}_{t-1}^2 - (\bar{c} - p) T^{-1} \tilde{x}_T^2 \right] / s_{AR}^2.$$

Ng and Perron (2001) also investigate the finite-sample performance of their modified tests when the autoregressive estimator s_{AR}^2 appearing in the tests is based on the GLS detrended series $\{\tilde{x}_t\}$ instead of the original series $\{x_t\}$. That is, they replace s_{AR}^2 by

$$\tilde{s}_{AR}^2 = \frac{\tilde{\sigma}_{\varepsilon'}^2}{\left(1 - \sum_{j=1}^{k} \tilde{\beta}_j\right)^2},$$

where the $\tilde{\beta}_j$ are OLS estimates for the β_j'' in the regression

$$\Delta \tilde{x}_t = d' \tilde{x}_{t-1} + \sum_{j=1}^{k} \beta_j'' \Delta \tilde{x}_{t-j} + \varepsilon_{t,k}',$$

with d' an unknown parameter, $\{\varepsilon_{t,k}'\}$ an unspecified error process, and $\tilde{\sigma}_{\varepsilon'}^2$ denoting the mean squared residuals resulting from this regression. However, Seo (2005) advises against this, as this may significantly reduce the power of these tests for non-local alternatives. Notwithstanding, Perron and Qu (2007) show that one may use $\tilde{s}_{AR}^2$ in these tests, but that the choice of the bandwidth parameter k should be based on OLS detrended data and not on GLS detrended data.

Remark. Taking into account the findings by Perron and Qu (2007), we select the lag parameter k (using the modified AIC described below) based on both GLS and OLS detrended data in our Monte Carlo study, but calculate the M test statistics and the autoregressive spectral density estimator based on the GLS detrended series.

3.6 Lag selection

The augmented Dickey–Fuller type tests considered in this chapter rely heavily on the choice of the lag parameter k in the ADF regression in (3.6). It is clear that when the choice of k is too small, the test will fail to account for possible autocorrelation in the error process. On the other hand, it has been established that an increased lag is associated with a loss of power (see e.g. Cheung and Lai, 1995; Paparoditis and Politis, 2016; Aylar, Smeekes and Westerlund, 2017). Also, in some cases, such as when $\{u_t\}$ is an MA process with a large negative MA coefficient, a correct choice of k is crucial to improve size and power properties of these tests (see Ng and Perron, 2001, p. 1527).

Several authors have investigated possible choices, such as the *ad hoc* choices given by Schwert (1989). Data-dependent choices include using the well-known AIC and BIC criteria, which will be discussed below, and a sequential t-test for the significance of the last lag (Ng and Perron, 1995). We will also discuss the class of modified information criteria proposed by Ng and Perron (2001), which has been shown to improve size and power properties of unit root tests.

A class of information criteria commonly found in the literature is given by

$$\mathrm{IC}(k) = \ln(\hat{\sigma}_k^2) + \frac{C_T(k+p+1)}{T-k_{\max}}, \qquad k = 0,1,\ldots,k_{\max},$$

where $\hat{\sigma}_k^2$ is the sum of the squared residuals resulting from the (ADF) regression, $C_T > 0$ is some constant such that $C_T = o(T)$ as $T \to \infty$, the constant $k_{\max}$ is an arbitrary upper bound for the considered lag parameters, and p indicates the specification of the deterministic component in the test (as defined earlier, $p = 0$ indicates a constant mean only and $p = 1$ indicates a linear time trend). Setting $C_T = 2$ one obtains the well-known *Akaike information criterion* (AIC). The *Bayesian* (or *Schwarz*) *information criterion* (BIC) also commonly found in the literature is obtained when setting $C_T = \log(T - k_{\max})$.

Using one of the criteria defined above, an optimal value for k is given by the rule

$$k_{ic} = \underset{k=0,1,\ldots,k_{\max}}{\operatorname{argmin}} \ \mathrm{IC}(k).$$

As an upper bound for the lag truncation k we will follow Schwert (1989) and use

$$k_{\max} = \ell_{12} = \left\lfloor 12\left(\frac{T}{100}\right)^{1/4} \right\rfloor.$$

It has become standard practice in the literature to use this so-called "Schwert lag" as an upper bound for k.

It is well known (see e.g. Schwert, 1989; Cavaliere, Phillips, Smeekes and Taylor, 2015) that unit root tests show large size distortions in the case where the errors follow a moving average process with a large negative coefficient. Ng and Perron (2001) show that these size distortions can be reduced significantly with a correct choice of k. In the same spirit as the above information criteria they propose a class of *modified information criteria* (MIC) defined as

$$\mathrm{MIC}(k) = \ln(\hat{\sigma}_k^2) + \frac{C_T(\tau_T(k)+k)}{T-k_{\max}},$$

with $\tau_T(k) = (\hat{\sigma}_k^2)^{-1}(\hat{\beta}_0^{(k)})^2 \sum_{t=k_{\max}+1}^{T} \tilde{x}_{t-1}^2$, where $\hat{\beta}_0^{(k)}$ is an estimator for β_0 in (3.9) which is based on the detrended data. Note that some authors suggest replacing the GLS detrended series $\{\tilde{x}_t\}$ with the OLS detrended series (Perron and Qu, 2007). As above, setting $C_T = 2$ yields a modified Akaike information criterion (MAIC) and setting $C_T = \log(T - k_{\max})$ yields a modified Bayesian criterion (MBIC).

Remarks.

(a) In the case where the error process $\{u_t\}$ follows an MA process with a large negative coefficient, the usual AIC or BIC tends to select the lag length too small, resulting in significant size distortions. As a practical solution to the problem, Elliott et al. (1996) set the lower bound of allowable values of k to 3, which reduces the size distortions. However, the modified information criteria provide a technically more satisfactory approach for choosing the lag length.

(b) Ng and Perron (2005) suggest that the effective sample size be held fixed across the candidate models.

(c) A more recent contribution by Cavaliere et al. (2015) is a *rescaled modified Akaike information criterion* (RSMAIC) which offers power gains in the case where the error process is unconditionally heteroskedastic. This addresses the issue that, for such models, the lag parameter k is selected too large. As unconditionally heteroskedastic models fall outside the scope of this study, we will employ the MAIC defined above.

(d) As mentioned earlier, in our Monte Carlo study we consider the suggestion of Perron and Qu (2007) and, for comparison, base lag selection on the MAIC *calculated from the OLS and GLS detrended data*, even if the test statistics are calculated from the GLS detrended series.

3.7 Recent developments

Hosseinkouchack and Hassler (2016) propose a variance ratio-type unit root test that is free of any tuning parameters, unlike the other tests discussed in this chapter. They further show that, apart from outperforming other tests free of tuning parameters, their test has an asymptotic power curve that lies close to that of the point optimal test by Elliott et al. (1996). Hosseinkouchack and Hassler even claim that their test may become more powerful than the point optimal test in finite samples.

Chapter 4

The bootstrap

The bootstrap is an automated, computerised resampling technique that, since its introduction by Bradley Efron in the 1970s, has proved to be successful in many problems of inference too complex to address adequately by means of traditional analytical methods. In fact, apart from being straightforward to implement, in many situations the bootstrap has been shown to produce results that are superior to results obtained by traditional methods, especially when the sample size is small or when underlying model assumptions cannot be verified. Efron and Tibshirani (1993, p. 45) state that *"The bootstrap [...] enjoys the advantage of being completely automatic. [It] requires no theoretical calculations, and is available no matter how mathematically complicated the estimator may be."*

To fully appreciate our discussion of the numerous bootstrap methods that have been developed to accommodate time series data, it is worth first reviewing the bootstrap in the context of independent data. In this chapter we describe how the bootstrap works and highlight some of the situations of estimation and inference where it is frequently employed. We also touch on the important property *bootstrap consistency* and discuss some situations in which the traditional bootstrap has been shown to work. We then move on to discuss the various procedures that have been developed for application of the bootstrap to dependent data and for each of these procedures provide a detailed algorithm along with remarks on some of their important associated properties.

4.1 Bootstrapping independent data

Let $\mathscr{X}_n = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ denote a random sample from an unknown d-variate distribution function F. In this chapter we shall give an overview of how the bootstrap can be used to draw inferences from $\mathscr{X}_n$ about a parameter vector $\theta = \theta(F)$, for some known functional $\theta(\cdot)$. The *plug-in estimator* for θ is given by $\hat{\theta}_n = \theta(\hat{F})$, with $\hat{F}$ an appropriate estimator for F.

Typically the statistician would draw inferences about θ based on the distributional properties of $\hat{\theta}_n$, which depend on F. Having only one sample at his disposal it might be a daunting task to uncover these properties, except in instances where assumptions can be made about F or where specific details (such as asymptotics) of $\hat{\theta}_n$ are known or can be derived.

The main idea behind the bootstrap is to mimic the mechanism F that generated the original sample by resampling from $\hat{F}$ to obtain what is termed a *bootstrap (re)sample*, which we denote by $\mathscr{X}_n^* = \{\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_n^*\}$. Since we have imitated the process of sampling from F by instead sampling from $\hat{F}$, the idea is that the distribution of $\hat{\theta}_n^*$ (conditional on $\hat{F}$) will be "close" to the distribution of $\hat{\theta}_n$ (conditional on F). The statistician may draw many such samples to obtain many realisations of $\hat{\theta}_n^* = \theta(\hat{F}^*)$ which can be used to obtain an approximate distribution of $\hat{\theta}_n^*$. Here $\hat{F}^*$ denotes the bootstrap estimator for $\hat{F}$.

A popular choice for $\hat{F}$ is $\hat{F}(\mathbf{x}) = F_n(\mathbf{x}) := n^{-1}\sum_{i=1}^n \mathbf{1}(\mathbf{X}_i \leq \mathbf{x})$, the empirical distribution function (EDF) of $\mathscr{X}_n$, which leads to the *nonparametric bootstrap*. Sampling from F_n is equivalent to randomly sampling with replacement from $\mathscr{X}_n$. If it is assumed that F is known up to a set of unknown parameters λ, and to show the dependence we write $F = F_{(\lambda)}$, we may choose, for example, $\hat{F} = F_{(\hat{\lambda})}$, where $\hat{\lambda}$ is some estimator for λ. Using $F_{(\hat{\lambda})}$ to generate $\mathscr{X}_n^*$ is known as the *parametric bootstrap*.

A standard introductory text on the bootstrap is Efron and Tibshirani (1993). A more advanced but still practical text is Davison and Hinkley (1997). More formal discussions of the bootstrap, along with important theoretical results and proofs, are given in texts such as Hall (1992) or Shao and Tu (1995).

The following subsections are devoted to illustrate how the bootstrap is employed in some commonly occurring applications.

4.1.1 Estimation of sampling distributions

Let $R_n(\mathscr{X}_n; F)$ denote a random variable of interest, which may depend on both the sample $\mathscr{X}_n$ and the unknown distribution function F. The sampling distribution of the random variable $R_n(\mathscr{X}_n; F)$ is given by

$$H_n(x) := \mathrm{P}(R_n(\mathscr{X}_n; F) \leq x), \qquad \forall x \in \mathbb{R}, \tag{4.1}$$

where P is the probability measure characterised by F. Replacing the distribution function F by an appropriate estimator $\hat{F}$, we obtain the traditional bootstrap estimator for $H_n(x)$:

$$\hat{H}_n(x) := \mathrm{P}\left(R_n(\mathscr{X}_n^*; \hat{F} \mid \mathscr{X}_n) \leq x\right) = \mathrm{P}^*\left(R_n(\mathscr{X}_n^*; \hat{F}) \leq x\right), \qquad \forall x \in \mathbb{R}, \tag{4.2}$$

where P^* refers to the conditional probability law of $\mathscr{X}_n^*$ given $\mathscr{X}_n$. The notation P^* will be used throughout the text to denote the probability measure in the *bootstrap world* (Efron and Tibshirani, 1986, p. 56).

In most practical settings, $\hat{H}_n(x)$ is unknown, but can be approximated by the Monte Carlo algorithm given below.

Algorithm 4.1.

1. Generate B independent bootstrap samples $\mathscr{X}_n^*(1), \mathscr{X}_n^*(2), \ldots, \mathscr{X}_n^*(B)$, each containing n values drawn randomly with replacement from the sample data $\mathscr{X}_n$.

2. For each bootstrap replication $b = 1,2,\ldots,B$, calculate the corresponding statistic

$$\hat{\theta}_n^*(b) = \theta(\hat{F}_n^b),$$

where $\hat{F}_n^b$ denotes the empirical distribution function of $\mathscr{X}_n^*(b)$. That is,

$$\begin{aligned}
\mathscr{X}_n^*(1) &= \{\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_n^*\} \quad \to \hat{\theta}_n^*(1) \\
\mathscr{X}_n^*(2) &= \{\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_n^*\} \quad \to \hat{\theta}_n^*(2) \\
&\vdots \\
\mathscr{X}_n^*(B) &= \{\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_n^*\} \quad \to \hat{\theta}_n^*(B)
\end{aligned}$$

3. Approximate $\hat{H}_n(x)$ by $\hat{H}_B(x) := \dfrac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(\hat{\theta}_n^*(b) \le x\right)$.

Bootstrap consistency

Consistency is an essential requirement for any estimator, including estimators based on the bootstrap. For the case of sampling distributions, Shao and Tu (1995) define consistency of $\hat{H}_n$ as an estimator for H_n as follows:

Definition 4.1. Let ν be a metric on $\mathscr{F}_{\mathbb{R}^s} = \{\text{all distributions on } \mathbb{R}^s\}$. $\hat{H}_n$ is (weakly) ν-consistent if $\nu(\hat{H}_n, H_n) \xrightarrow{p} 0$ as $n \to \infty$, and strongly ν-consistent if $\nu(\hat{H}_n, H_n) \xrightarrow{a.s.} 0$ as $n \to \infty$.

The metric ν in the above definition is usually chosen to be the supremum-norm $\nu_\infty(F,G) = \sup_x |F(x) - G(x)|$, but other measures such as Mallows' distance are also commonly used. In fact, in a famous paper Bickel and Freedman (1981) established consistency (under some conditions) of $\hat{H}_n$ in (4.2) as an estimator for H_n in (4.1) using Mallows' distance and its properties.

It has been shown that the traditional bootstrap is consistent (and sometimes strongly consistent) in many cases of practical interest, such as when the statistic of interest is a smooth function of the sample mean or when considering the empirical and quantile processes. For an overview of the major results on bootstrap consistency that have been established in the literature, see Section 3.2 of Shao and Tu (1995). A recent paper by Romano and Shaikh (2012) provides strengthened conditions under which the bootstrap can be used to construct estimators of the quantiles of the distribution of a root.

4.1.2 Estimation of standard error

Consider the problem of estimating the true standard error of the estimator $\hat{\theta}_n$, which we will denote by

$$\mathrm{SE}(\hat{\theta}_n) := \sqrt{\mathrm{Var}(\hat{\theta}_n)}. \tag{4.3}$$

The *(ideal) bootstrap estimate* of $\mathrm{SE}(\hat{\theta}_n)$ is given by

$$\mathrm{SE}^*(\hat{\theta}_n^*) := \sqrt{\mathrm{Var}^*(\hat{\theta}_n^*)} = \sqrt{\mathrm{E}^*\left(\hat{\theta}_n^* - \mathrm{E}^*(\hat{\theta}_n^*)\right)^2}, \tag{4.4}$$

where E^* and Var^* respectively denote the expected value and variance taken with respect to F_n. In very few cases there exist neat, explicit formulas for the ideal bootstrap estimate of $\mathrm{SE}(\hat{\theta}_n)$. For example, if we choose the parameter of interest as the population *mean*, then our estimator is the sample mean $\hat{\theta}_n = \theta(F_n) = \bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ and the bootstrap equivalent is $\hat{\theta}_n^* = \theta(\hat{F}_n) = \bar{X}_n^* = n^{-1}\sum_{i=1}^n X_i^*$. In this case (4.3) becomes

$$\mathrm{SE}(\bar{X}_n) = \sqrt{\mathrm{Var}(\bar{X}_n)} = \frac{\sigma}{\sqrt{n}},$$

where $\sigma = \sqrt{\mathrm{Var}(X)}$. Also, since $\mathrm{E}^*(X_i^*) = \bar{X}_n$ and

$$\hat{\sigma}_n^2 := \mathrm{Var}^*(X_i^*) = \int (x - \bar{X}_n)^2 dF_n = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

the bootstrap estimate in (4.4) becomes

$$\mathrm{SE}^*(\bar{X}_n^*) = \sqrt{\mathrm{Var}^*(\bar{X}_n^*)} = \frac{1}{n}\sqrt{\sum_{i=1}^n \mathrm{Var}^*(X_i^*)} = \frac{\hat{\sigma}_n}{\sqrt{n}}.$$

Note that, given the sample data $\mathscr{X}_n$, we can calculate the quantity $\hat{\sigma}_n$, and consequently the ideal bootstrap standard error $\mathrm{SE}^*(\bar{X}_n^*)$, exactly. In most cases, however, (4.4) cannot be calculated explicitly from the sample data. In many such cases one can effectively approximate the ideal bootstrap estimate of standard error by the following algorithm given in Efron and Tibshirani (1986, p. 56).

Algorithm 4.2.

Follow Algorithm 4.1, but replace step 3 by

3. Approximate the ideal bootstrap standard error $\mathrm{SE}^*(\hat{\theta}_n^*)$ by

$$\widehat{\mathrm{SE}}_B := \sqrt{\frac{1}{B-1}\sum_{b=1}^B \left(\hat{\theta}_n^*(b) - \hat{\theta}_n^*(\cdot)\right)^2},$$

where $\hat{\theta}_n^*(\cdot) = \frac{1}{B}\sum_{b=1}^B \hat{\theta}_n^*(b)$.

Note that, by the strong law of large numbers, $\widehat{\mathrm{SE}}_B \to \mathrm{SE}^*(\hat{\theta}_n^*)$ almost surely as $B \to \infty$.

4.1.3 Estimation of bias and bias reduction

The bias of an estimator $\hat{\theta}_n$ for a parameter θ is defined as

$$\mathrm{bias}(\hat{\theta}_n, \theta) = \mathrm{E}\left(\hat{\theta}_n\right) - \theta.$$

We could use this information to construct an unbiased estimate of θ, namely

$$\hat{\theta}_n - \mathrm{bias}(\hat{\theta}_n, \theta). \tag{4.5}$$

Of course, the quantity $\mathrm{bias}(\hat{\theta}_n, \theta)$ is unknown and therefore also needs to be estimated. To this end we define the *bootstrap estimator for bias* by

$$\mathrm{bias}^*(\hat{\theta}_n^*, \hat{\theta}_n) = \mathrm{E}^*\left(\hat{\theta}_n^*\right) - \hat{\theta}_n, \tag{4.6}$$

so that the new bias-corrected estimator for θ is given by

$$\tilde{\theta}_n = \hat{\theta}_n - \text{bias}^*(\hat{\theta}_n^*, \hat{\theta}_n).$$

Note that since we have replaced the exact bias appearing in (4.5) by an estimator, the new bias-corrected estimator $\tilde{\theta}_n$ will not necessarily be unbiased. However, in regular cases an estimator $\hat{\theta}_n$ will typically have a bias of $O(n^{-1})$, in which case the bias-corrected version $\tilde{\theta}_n$ will have a reduced bias of $O(n^{-2})$. See Chang and Hall (2015).

The above procedure may be repeated indefinitely on the estimator $\tilde{\theta}_n$ to obtain even less biased estimators for θ. This is known in the literature as *bootstrap iteration*. Typically, each iteration of bootstrap bias reduction will reduce the bias of an estimator by an order of n^{-1}. The interested reader is referred to Hall and Martin (1988) and Chan and Lee (2001).

Remark. The bootstrap bias in (4.6) may be approximated by the following "obvious" estimator:

$$\widehat{\text{bias}}_B = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}_n^*(b) - \hat{\theta}_n,$$

where $\hat{\theta}_n^*(b)$ is defined as in Algorithm 4.1. However, this estimator may be improved significantly by a method discussed in Chapter 23 of Efron and Tibshirani (1993).

4.2 Error reduction

The avid reader would have noticed that in bootstrap estimation there are mainly two sources of variation, namely *statistical error* (which constitutes errors such as model misspecification) and variability resulting from *simulation error* (see Davison and Hinkley, 1997). To be concrete, suppose we want to estimate $\text{SE}(S_n^2)$ using Algorithm 4.2 above, where S_n^2 is the (unbiased) sample variance calculated from $n = 50$ independent observations from an $N(0,1)$ distribution. In our example the statistical error results from *estimating* $\text{SE}(S_n^2)$ by its bootstrap counterpart $\text{SE}^*((S_n^*)^2)$, which usually is also unknown and has to be *approximated* by $\widehat{\text{SE}}_B$, resulting in simulation error.

A natural question that arises in the practical application of the bootstrap, is how many bootstrap resamples are necessary for the latter approximation to be sufficient? The choice of B, the number of bootstrap replications, ultimately depends on the nature of the underlying problem. For instance, when estimating the standard error of an estimator, Efron (1987) suggests that in many cases taking $B = 25$ already gives reasonable results (also see Efron and Tibshirani, 1993, p. 55ff.). We shall see that this choice of B is insufficient, even in the case of a simple example.

In our simple example we have by standard results that the true value of the parameter of interest is $\text{SE}(S_n^2) = \sqrt{2/(n-1)} \approx 0.202$. Figure 4.1 shows the effect that B has on the resulting bootstrap estimate of $\text{SE}(S_n^2)$ for 10 independent experiments. The erratic behaviour seen in the left of the figure stems from a combination of statistical and simulation error. Moving to the right of the figure, the estimates settle down as simulation error is gradually

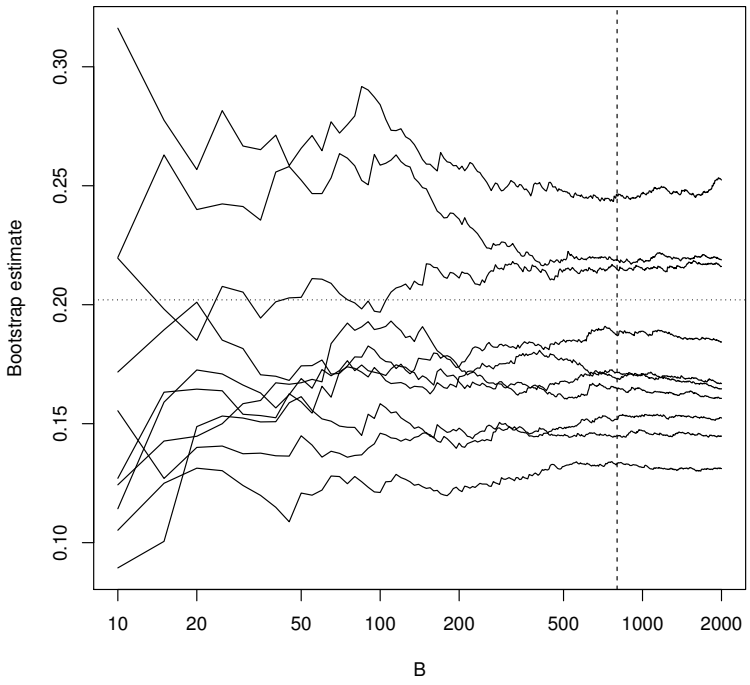

Figure 4.1: Effect of B on the bootstrap estimate of $\mathrm{SE}(S_n^2)$ for 10 independent experiments. The dotted horizontal line indicates the true theoretical value. The dashed vertical line represents Booth and Sarkar's recommended minimum choice for B.

reduced. It is clear that increasing B past a certain point has only little effect on the resulting estimate, i.e. larger choices of B do not continue to significantly reduce simulation error.

From this observation it seems reasonable that the choice of B should be based *only on resampling variability* and not on variability originating from statistical error. In agreement with this, Booth and Sarkar (1998) argue that the choice of B should be based on the coefficient of variation of $\mathrm{SE}^*(\hat{\theta}_n^*)$, conditional on the sample $\mathscr{X}_n$. This *conditional* coefficient of variation may be approximated (by means of Taylor expansion) by

$$\mathrm{CV}^*\left(\mathrm{SE}^*(\hat{\theta}_n^*)\right) \approx \sqrt{\frac{\hat{\delta}+2}{4B}},$$

where $\hat{\delta}$ denotes the kurtosis of the bootstrap distribution of $\hat{\theta}_n^*$. Under the assumption that the statistic $\hat{\theta}_n$ is approximately normal, they suggest using no less than approximately 800 bootstrap resamples. This recommendation is indicated by a dashed vertical line on Figure 4.1. Beyond this point the estimates seem to be fairly stable, in accordance with their result.

If the aim is to approximate distribution quantiles—a problem encountered when conducting hypothesis testing—a slightly different result is needed. Define the β-level bootstrap quantile of $\hat{\theta}_n^*$ by $\hat{\xi}_\beta$, i.e. $\mathrm{P}^*\left(\hat{\theta}_n^* \leq \hat{\xi}_\beta\right) = \beta$. If it is assumed that $\hat{\theta}_n^*$ is roughly normal, it

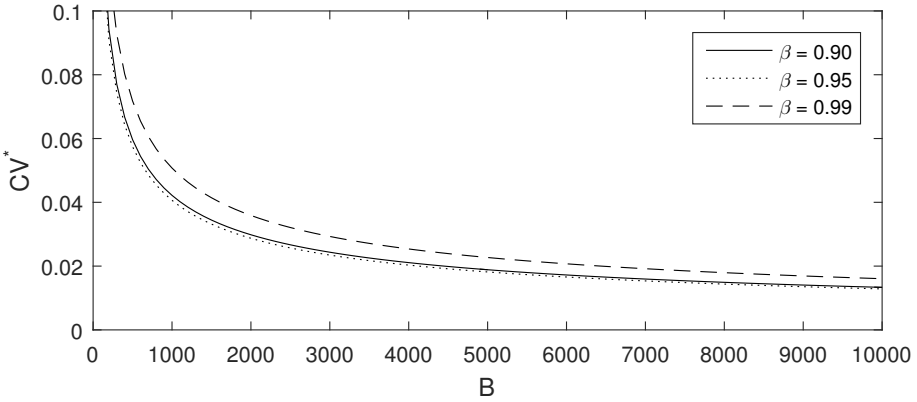

Figure 4.2: Conditional coefficient of variation (CV^*) of $\hat{\xi}_\beta$ versus resample size B for different choices of β.

follows from standard results that the conditional coefficient of variation of $\hat{\xi}_\beta$ is approximately

$$\mathrm{CV}^*\left(\hat{\xi}_\beta\right) \approx \frac{1}{z_\beta}\sqrt{\frac{\beta(1-\beta)}{B\,\phi(z_\beta)^2}},$$

as also stated in Efron (1987) and Booth and Sarkar (1998), where $z_\beta = \Phi^{-1}(\beta)$ and $\phi(x) = \Phi'(x)$, with Φ denoting the distribution function of the standard normal distribution. This relation (illustrated in Figure 4.2) shows that to obtain 95%-level quantile estimates with a (conditional) coefficient of variation of less than 2% requires approximately 4 100 bootstrap resamples. Although we use the bootstrap in the context of dependent data, this fact already provides a rough guide for choosing the minimum number of bootstrap replications we require in Chapter 7.

4.3 Hypothesis testing

Let Ω be the collection of all possible $\theta = \theta(F)$. Consider testing

$$H_0 : \theta \in \Omega_0 \qquad \text{vs.} \qquad H_A : \theta \in \Omega_A, \tag{4.7}$$

where Ω_0 and Ω_A are two disjoint subsets of $\Omega = \Omega_0 \cup \Omega_A$. Given an appropriate test statistic $T(\cdot)$, suppose the test rejects the null hypothesis whenever $T(\mathscr{X}_n) \le c_\alpha$, where c_α is some constant such that

$$\mathrm{P}_{H_0}(T(\mathscr{X}_n) \le c_\alpha) = \alpha, \tag{4.8}$$

with P_{H_0} denoting the probability measure under the null hypothesis. The p-value of a (left-sided) test is given by

$$p = \mathrm{P}_{H_0}(T(\mathscr{X}_n) \le t),$$

where $t = T(\boldsymbol{x}_n)$ denotes the value of the test statistic calculated from the realisation $\boldsymbol{x}_n$ of $\mathscr{X}_n$.

Hall and Wilson (1991) provide two guidelines for performing bootstrap-based hypothesis tests: (i) resample in a way that reflects the null hypothesis (see also Hinkley, 1988), and (ii) employ methods that are already recognised as having good features in the related problem of confidence interval construction (such as using asymptotically pivotal statistics). The first guideline has become common practice in the application of bootstrap hypothesis testing and we now discuss two methods to "mimic" the null hypothesis when resampling.

Two popular methods used to enforce the null hypothesis when resampling from $\mathscr{X}_n$ are the *transformation method* and the *exponentially tilted version of the empirical distribution function* (EDF) (Efron and Tibshirani, 1993; Davison and Hinkley, 1997; Allison and Swanepoel, 2010). The transformation method involves transforming the original sample $\mathscr{X}_n$ in such a way that $\hat{\theta}_n = \theta(\tilde{F}_n) \in \Omega_0$, where $\tilde{F}_n$ denotes the empirical distribution of the transformed data. The exponentially tilted EDF method involves keeping $\mathscr{X}_n$ fixed, but associating a probability p_i with each observation $\mathbf{X}_i$. Denote this new probability distribution by $\tilde{F}_p$. It is then possible to obtain the probabilities p_i by minimising some measure of distance (such as the Kullback-Leibler distance) between F_n and $\tilde{F}_p$ subject to the constraint $\theta(\tilde{F}_p) \in \Omega_0$. Resampling done from $\tilde{F}_p$ ensures that the null distribution is enforced.

Recently a general resampling plan for hypothesis testing in the k-sample problem has been devised by Martínez-Camblor and Corral (2012).

Other standard sources on bootstrap-based hypothesis testing are Chernick (1999) and Good (2000). Also see MacKinnon (2009).

4.4 Model-based bootstrap

The bootstrap methods discussed thus far were designed for the i.i.d. case. These methods may not work in more complex cases. Suppose we observe mutually independent pairs $\{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)\}$, with d-vectors $\mathbf{X}_i = (1, X_{i2}, \ldots, X_{id})'$, generated by the linear regression model

$$Y_i = \boldsymbol{\beta}'\mathbf{X}_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where $\boldsymbol{\beta}$ is a d-vector of real coefficients and ε_i are mutually independent random error terms, each with distribution function F_i. Also assume that $\mathrm{E}(\varepsilon_i) = 0$ and $\mathrm{E}(\varepsilon_i^2) = \sigma_i^2 > 0$. Define the residuals resulting from the model fit by $e_i = Y_i - \bar{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}_i$, where $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$ and $\hat{\boldsymbol{\beta}}$ is the ordinary least squares estimate of $\boldsymbol{\beta}$. Note that the inclusion of an intercept in the model statement ensures that $\bar{e} = n^{-1}\sum_{i=1}^n e_i = 0$.

In the literature there are two popular methods for using the bootstrap to draw inferences about $\boldsymbol{\beta}$: (i) resampling residuals and (ii) resampling cases. These two methods are discussed in the following two subsections.

Resampling errors

Many authors (see e.g. Freedman, 1981; Davison and Hinkley, 1997) suggest using the following bootstrap resampling scheme:

1. For each $i = 1,2,\ldots,n$, generate independent random variables ε_i^* from some distribution $\hat{F}_i$, which serves as an estimate of F_i.

2. Compute the bootstrap responses $Y_i^* = \hat{\boldsymbol{\beta}}'\mathbf{X}_i + \varepsilon_i^*$, $i = 1,2,\ldots,n$.

3. Calculate the ordinary least squares estimate $\hat{\boldsymbol{\beta}}^*$ from the bootstrap data $\{(\mathbf{X}_i, Y_i^*)\}_{i=1}^n$.

There have been many suggestions in the literature for the choice of $\hat{F}_i$. The first of these, due to Efron (1979), suggests simply using the traditional bootstrap and resampling randomly with replacement from $\{e_i\}_{i=1}^n$ to obtain $\{\varepsilon_i^*\}_{i=1}^n$. Although this procedure is reasonable if the error terms originate from a common distribution, i.e. if $F_i = F\ \forall i$, in the case of heteroskedastic errors this choice has been shown to yield an inconsistent estimator for the standard error of $\hat{\boldsymbol{\beta}}$ (Liu, 1988). She also shows how, in the case of *simple* linear regression, the residuals may be modified to achieve consistency. For another related example where the traditional bootstrap is inconsistent, see Härdle and Mammen (1993).

As an alternative to the traditional bootstrap, Wu (1986) proposed what is now known as the *wild bootstrap*. The idea is to generate the bootstrap error terms according to

$$\varepsilon_i^* = f(e_i)u_i^*,$$

where $f(\cdot)$ is some transformation and u_i^* is a random variable with mean 0, independent of $\{e_i\}_{i=1}^n$. Some choices for f are given in Davidson and Flachaire (2008). Usually the u_i^* are generated such that the first three moments of the bootstrap error terms correspond to those of the original error terms. Imposing the following restrictions on u_i^* achieves this:

$$\mathrm{E}^*\left(u_i^*\right) = 0, \qquad \mathrm{E}^*\left((u_i^*)^2\right) = 1, \qquad \mathrm{E}^*\left((u_i^*)^3\right) = 1, \qquad i = 1,2,\ldots,n.$$

Addition of the last restriction based on the third moment was first suggested by Liu (1988), who showed that this additional restriction improves the convergence rate of bootstrap estimates (see also Härdle and Mammen, 1993). Probably the most popular choice for u_i^* is the two-point-distribution

$$u_i^* = \begin{cases} -\frac{1}{2}\left(\sqrt{5}-1\right) & \text{with probability } p = \frac{1}{2\sqrt{5}}\left(\sqrt{5}+1\right), \\ \frac{1}{2}\left(\sqrt{5}+1\right) & \text{with probability } 1-p. \end{cases}$$

This choice was suggested by Mammen (1993), who also provides other constructions. More constructions are given in Liu (1988) and MacKinnon (2006). Davidson and Flachaire (2008) also study using Rademacher variables as choice for u_i^* and show that in one specific case it is possible to draw perfect bootstrap inference.

Examples of where the wild bootstrap is applied in *nonparametric* regression can be found in Härdle and Marron (1991), Cao-Abad (1991), Cao-Abad and González-Manteiga (1993), Härdle and Mammen (1993), Stute, González-Manteiga and Quindimil (1998), Delgado and González-Manteiga (2001), González-Manteiga, Miranda and González (2004), Racine, Hart and Li (2006), Ferraty, van Keilegom and Vieu (2010) and Raña, Aneiros,

Vilar and Vieu (2016). The wild bootstrap is also frequently applied in goodness-of-fit tests concerning the distribution of the error terms in parametric and nonparametric models (see e.g. Dette, Kusi-Appiah and Neumeyer, 2002; Neumeyer, Dette and Nagel, 2005; Neumeyer and Dette, 2007; Hušková and Meintanis, 2007, 2010, 2012; Jiménez Gamero, 2014).

Resampling cases

As an alternative to resampling errors, one may retain the relationship between the response and predictor variables by resampling pairs of observations as in the following procedure:

1. Randomly sample n pairs with replacement from $\{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)\}$ to obtain the bootstrap data $\{(\mathbf{X}_1^*, Y_1^*), \ldots, (\mathbf{X}_n^*, Y_n^*)\}$.
2. Calculate the ordinary least squares estimate $\hat{\boldsymbol{\beta}}^*$ from the bootstrap data $\{(\mathbf{X}_i^*, Y_i^*)\}_{i=1}^n$.

Like the wild bootstrap, this procedure allows for heteroskedastic errors. Moreover, the major advantage of this procedure is that it is not assumed that the conditional mean of Y_i given $\mathbf{X}_i$ is linear. However, Flachaire (2005) shows in a Monte Carlo study that one specific version of the wild bootstrap outperforms the bootstrap based on resampling pairs.

Restricted versus unrestricted estimates

When testing hypotheses placing a restriction on $\boldsymbol{\beta}$, the question arises whether the estimates of $\boldsymbol{\beta}$ used to compute the bootstrap observations should be restricted by the null hypothesis. To be concrete, suppose that we want to test

$$H_0 : \boldsymbol{\beta} = \mathbf{0}.$$

Should the bootstrap scheme be based on the *restricted* residuals $e_i = Y_i - \bar{Y}$ (assuming $\boldsymbol{\beta} = 0$) or on the *unrestricted* residuals $e_i = Y_i - \bar{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}_i$ as above? This problem has been studied by Paparoditis and Politis (2005a), who reached the following conclusions:

- Imposing the null hypothesis when using non-pivotal statistics is valid, but leads to a loss of power under the alternative. In this case they recommend using parameter estimators that converge to the true parameter value under both the null and alternative hypotheses.
- When using Studentised statistics, consistency of the test is quite robust to whether the null hypothesis is imposed on parameter estimates or not.

Also see Flachaire (2005) and MacKinnon (2006).

4.5 Bootstrapping dependent data

The original idea of the bootstrap is to mimic the complete structure of the data generating process using a single sample. When applying the bootstrap to time series data this leads to the additional challenge of estimating the dependence structure of the process. Singh (1981) demonstrated that, although the traditional i.i.d. bootstrap is superior in many settings, it is inadequate for dependent data. The residual or wild bootstrap methods discussed in the previous section may be applicable in parametric settings, but are not very useful in situations where the statistician is unable to adequately describe the data by means of an appropriate model. What followed was a multitude of tests designed specifically to deal with this challenge of reproducing the dependence structure of the data.

In this section we describe how the most significant of these bootstrap methods for dependent data work and for each method highlight the features that have led to its respective success. Throughout let $\{X_1, X_2, \ldots, X_T\}$ be a sample of dependent random variables with *joint* distribution F depending on some unknown parameter θ. Let $V_T = V(X_1, X_2, \ldots, X_T)$ be some statistic that may be used to draw inferences about θ. For each algorithm, assume that when the bootstrap sample $\{X_1^*, X_2^*, \ldots, X_T^*\}$ has been drawn, we calculate the bootstrap version of the test statistic $V_T^* = V(X_1^*, X_2^*, \ldots, X_T^*)$ as a final step in order to draw our inferences.

4.5.1 The autoregressive sieve bootstrap

The autoregressive sieve bootstrap (ASB) was first introduced by Kreiss (1988) for *linear* time series of the form (2.1) but later extended by Bühlmann (1995b) for time series of a more general form. Letting $\{X_t, t \in \mathbb{Z}\}$ be a real-valued, stationary time series with $\mathrm{E}(X_t) = \mu$, it is known by Wold's Theorem that $\{X_t - \mu, t \in \mathbb{Z}\}$ can be written as a one-sided infinite order MA process

$$X_t - \mu = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \forall t \in \mathbb{Z}, \tag{4.9}$$

where $\psi_0 = 1$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ and $\{\varepsilon_t\}$ is a zero-mean white noise process. Bühlmann (1995b) only requires *invertibility* of (4.9) as an additional assumption for $\{X_t, t \in \mathbb{Z}\}$ to be represented as a one-sided infinite order AR process

$$\sum_{j=0}^{\infty} \beta_j (X_{t-j} - \mu) = \varepsilon_t, \qquad \forall t \in \mathbb{Z},$$

where $\beta_0 = 1$, the β_j are unknown parameters such that $\sum_{j=0}^{\infty} \beta_j^2 < \infty$ and $\{\varepsilon_t\}$ is a zero-mean white noise process. This infinite order representation can now be approximated by a finite order AR process as a sieve for $\{X_t, t \in \mathbb{Z}\}$. The AR approximation is of the form

$$\sum_{j=0}^{p} \beta_j (X_{t-j} - \mu) = \varepsilon_t, \qquad \forall t \in \mathbb{Z}. \tag{4.10}$$

This approximation is preferred over an MA approximation since it is faster and more successful for general situations (cf. Berk, 1974; An, Chen and Hannan, 1982; Hannan, 1987).

We now provide the algorithm of the sieve bootstrap. Let $\{X_1, X_2, \ldots, X_T\}$ denote a sample from the process $\{X_t, t \in \mathbb{Z}\}$.

1. Fit an AR(p) model to the sample $\{X_1, X_2, \ldots, X_T\}$ with order p chosen such that $p \to \infty$ as $T \to \infty$. See remark (a) below.

2. Estimate the coefficients $(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)$ corresponding to (4.10). See remark (b) below.

3. Construct the residuals

$$\hat{\varepsilon}_t = \sum_{j=0}^{p} \hat{\beta}_j (X_{t-j} - \bar{X}) \quad \text{for} \quad t = p+1, \ldots, T,$$

   with $\hat{\beta}_0 = 1$.

4. Construct the centred residuals

$$\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\varepsilon}_t \quad \text{for} \quad t = p+1, \ldots, T.$$

5. Randomly sample with replacement $T-p$ elements from the centred residuals to form the bootstrap errors $\{\varepsilon_t^*\}_{t=p+1}^{T}$.

6. Construct the second part of the bootstrap sample, i.e. $\{X_{p+1}^*, X_{p+2}^*, \ldots, X_T^*\}$, by the recursion

$$\sum_{j=0}^{p} \hat{\beta}_j (X_{t-j}^* - \bar{X}) = \varepsilon_t^* \quad \text{for} \quad t = p+1, \ldots, T.$$

   Initial values $\{X_1^*, X_2^*, \ldots, X_p^*\}$ can be obtained in various ways, two of which are discussed in remark (c) below.

Remarks.

(a) Shibata (1980) has shown the Akaike information criterion (AIC) to be optimal for prediction in AR(∞) models, and hence an estimate $\hat{p}$ of p can be chosen by the AIC model selection procedure with Gaussian innovations.

(b) Bühlmann (1995b) recommends using the Yule-Walker estimates, although other estimates have also been used in the past.

(c) Two ways to obtain the initial values (when working with lagged variables) are as follows.

- Due to its simplicity the most popular method to obtain initial values $\{X_1^*, X_2^*, \ldots, X_p^*\}$ is to set them equal to the values in the original sample $\{X_1, X_2, \ldots, X_p\}$. That is, $X_t^* = X_t$ for $t = 1, 2, \ldots, p$.

- A more appropriate method for the case of hypothesis testing, as recommended by Bühlmann (1995b) and Allison and Swanepoel (2010), is to make use of a so-called *burn-in* period which involves generating $T+k+1$ elements to construct an original sample $\{X_{-k}, X_{-k+1}, \ldots, X_0, \ldots, X_T\}$ instead of $\{X_1, X_2, \ldots, X_T\}$. Of course, the value of p is unknown when this original sample is constructed, so k is usually chosen as a fraction of the sample size, e.g. $k = \lceil T/4 \rceil$, so that $k > p$. The algorithm is then performed using this sample containing $T+k+1$ elements to obtain an initial bootstrap sample $\{X^*_{-k}, X^*_{-k+1}, \ldots, X^*_0, \ldots, X^*_T\}$. The terms $X^*_{-k}, X^*_{-k+1}, \ldots, X^*_0$ are then discarded and one is left with a final bootstrap sample of size T, i.e. $\{X^*_1, X^*_2, \ldots, X^*_T\}$.

4.5.2 The moving block bootstrap

The class of block bootstraps is a set of resampling methods which involves splitting the sample into blocks of observations and then resampling blocks instead of resampling individual observations. Generally speaking, the blocks can be of varying (i.e. random) length and can be overlapping (i.e. if the observations are ordered sequentially then an observation may belong to more than one block simultaneously). This method of resampling dependent data ensures that the dependence structure within blocks of short-range dependent observations is preserved. To ensure that the blocking procedure asymptotically reproduces the underlying dependence structure, the block length may be chosen as an increasing function of the sample size.

The moving block bootstrap (MBB) was independently proposed by Künsch (1989) and Liu and Singh (1992) and splits the data into $T-\ell+1$ overlapping blocks of length ℓ. Block 1 will contain the observations 1 to ℓ, block 2 will contain observations 2 to $\ell+1$, and so forth. A new bootstrap series is then formed by drawing T/ℓ blocks randomly with replacement from the $T-\ell+1$ blocks and aligning them in the order they were drawn. By principle, the bootstrap series will be nonstationary since the bootstrap works with dependent data.

More precisely, the MBB algorithm works as follows:

1. Define blocks $\mathscr{B}_j = \{X_j, \ldots, X_{j+\ell-1}\}$ for $j = 1, 2, \ldots, N$, where $N = T-\ell+1$ and where $1 \leq \ell \leq T$ denotes the block size.

2. Let $b = \lfloor T/\ell \rfloor$. Select a random sample of blocks with replacement $\{\mathscr{B}^*_1, \mathscr{B}^*_2, \ldots, \mathscr{B}^*_b\}$ from $\{\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_N\}$.

3. Concatenate $\mathscr{B}^*_1, \mathscr{B}^*_2, \ldots, \mathscr{B}^*_b$ in the order they were drawn to form a sequence.

4. This yields $T_1 = b\ell$ bootstrap observations $\mathbf{X}^*_{T_1} = \{X^*_1, X^*_2, \ldots, X^*_{T_1}\}$. Note that $T_1/T \to 1$ as $T \to \infty$.

The distributional properties of V^*_T may now be used as a proxy for those of V_T.

The reason that b elements are used from $\{\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_N\}$ is that T elements in total are wanted in the bootstrap sample. But when using b blocks in the bootstrap sample it does not result in a sample of size T, but rather a sample of size $T_1 = b\ell$. Note, however, that since $T_1/T \to 1$ as $T \to \infty$, the bootstrap sample size approaches the true sample size as the true sample size increases. The choice of the block length ℓ needs careful consideration as will be discussed shortly.

Bühlmann (1995b) mentions that the MBB performs satisfactory and is robust against misspecified models, but that the bootstrap series exhibits properties which are caused by the concatenation of randomly selected blocks. The dependence between different blocks is ignored in the bootstrap series and hence it is not (conditionally) stationary. A modification to the MBB which yields a conditionally stationary bootstrap sample is the stationary block bootstrap procedure discussed shortly.

It has been shown that the MBB is valid (i.e. produces consistent estimates) in many cases, but it is typically required that

$$\ell \to \infty \quad \text{and} \quad T^{-1}\ell \to 0 \quad \text{as} \quad T \to \infty,$$

and that the data generating processes are only short-range dependent (Naik-Nimbalkar and Rajarshi, 1994; Bühlmann, 1995a; Bühlmann and Künsch, 1995; Radulović, 1996a,b). Lahiri (2003) provides rigorous proofs of the consistency of the MBB when estimating the variance or distribution function of the centred and scaled sample mean. Second-order correctness of the MBB for Studentised statistics has been established independently by Goëtze and Künsch (1996) and Lahiri (1996). Radulović (2012) establishes a bootstrap central limit theorem for the mean, along with necessary and sufficient conditions.

4.5.3 The circular block bootstrap

As a member of the class of block bootstraps, the circular block bootstrap (CBB) of Politis and Romano (1992) is just an extension of the well-known MBB. Whereas the MBB creates $T - \ell + 1$ blocks, each of length ℓ, the CBB creates T blocks, each of length ℓ, by wrapping around the time series to form the last $\ell - 1$ blocks, the last block having X_T as its first observation. Put differently, if the block formation is shifted to the end of the time series and cannot shift any further, then the block wraps around to the beginning of the series to fill the remaining positions in the block with observations. We have the collection of blocks $\{\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_T\}$, i.e. there are the same number of blocks as there are observations. We then sample $b = \lfloor T/\ell \rfloor$ blocks from this collection to obtain $\mathscr{B}_1^*, \mathscr{B}_2^*, \ldots, \mathscr{B}_b^*$ and concatenate them to obtain the bootstrap series $\{X_1^*, X_2^*, \ldots, X_{b\ell}^*\}$. Seen differently, since the first and last few observations of a series do not have equal chance of being drawn into a bootstrap sample as the observations in the middle of the series, the observations are laid in a circle and consecutive blocks of bootstrap data are generated from the circle. For example, if we choose blocks of length $\ell = 3$ from a series $\{X_1, X_2, \ldots, X_T\}$, then the blocks would be $\mathscr{B}_1 = \{X_1, X_2, X_3\}$, $\mathscr{B}_2 = \{X_2, X_3, X_4\}$, $\ldots$, $\mathscr{B}_T = \{X_T, X_1, X_2\}$.

4.5.4 Modifications and considerations of the block bootstrap procedure

Other variants of the block bootstrap exist, for example:

- the *nonoverlapping block bootstrap* (NBB) (Carlstein, 1986; Radulović, 2009), where the blocks $\{\mathscr{B}_k\}_{k=1}^{T-\ell+1}$ are chosen such that they are disjoint, i.e. that they do not overlap;

- the *stationary block bootstrap* (SBB) (Politis and Romano, 1994) where the fixed block lengths are replaced by random block lengths (drawn from a geometric distribution) which ensures that the resampled data again form a stationary time series (conditional on $\hat{F}$), which is required in some situations. Bühlmann (1995b) notes that this procedure depends on a tuning parameter p, the probability parameter of the geometric distribution, which is difficult to control. At present there exists no data-driven choice for this parameter and in practice only rules of thumb are employed;

- the *tapered block bootstrap* (TBB) (Paparoditis and Politis, 2001) which typically produces estimators with reduced bias. For certain parameter choices in the dependent wild bootstrap of Shao (2010) the TBB is obtained (see Section 2 in Shao (2010) and also p. 8 of Smeekes and Urbain (2014a)).

More recently Dudek, Leśkow, Paparoditis and Politis (2014) proposed a generalised block bootstrap method for time series containing a periodic or seasonal component.

The choice of block length ℓ is crucial to the success of the block bootstrap methods. Hall, Horowitz and Jing (1995) proposed a general method for estimating the optimal block size in the cases of variance, bias or distribution function estimation. Their method uses subsampling, another method of resampling, to approximate the mean squared error (MSE) of the estimator for different choices of ℓ. The estimated block size is that value of ℓ that minimises this MSE of the bootstrap estimator. The drawback of this method is that it requires an explicit expression for the MSE of the estimator of interest, which they derive for the cases of variance, bias and distribution function estimation. These expressions are also used by Politis and White (2004) (with corrections by Patton, Politis and White, 2009) to choose an optimal block length based on the notion of spectral estimation via flat-top lag-windows of Politis and Romano (1995). To get around this problem, Lahiri, Furukwa and Lee (2007) came up with a nonparametric plug-in method which does not require such explicit expression of the MSE. This plug-in method is applicable to problems concerning the variance, bias, distribution function and quantiles.

Another interesting choice of plug-in estimator for the block length has recently been used by Hong, Wang and Wang (2016) in a goodness-of-fit setting where the objective is to test whether a given time series was generated by a strictly stationary process. They employ Lima and Xiao's (2010) partially data-dependent estimator for lag order. Their estimator imposes an upper bound to the frequently-used AR(1) plug-in estimator originally studied

(in more generality) by Andrews (1991). Lima and Xiao's motivation behind the added upper bound is that under the alternative hypothesis the estimated lag-order may become too large and may result in a loss of power.

Theoretical comparisons in terms of the bias and variance of estimators obtained by means of the CBB, MBB, NBB and SBB are given in Lahiri (1999) and Nordman (2009). The authors show that variances of NBB estimators are larger than those of CBB, MBB and SBB estimators. The bias and variance of the latter three methods have been found to be comparable. An excellent monograph on resampling methods for dependent data is Lahiri (2003). Kreiss and Lahiri (2012) provide a comprehensive overview of bootstrap methods for time series data. Subsampling has also been generalised for dependent data (see Politis, Romano and Wolf, 1999).

4.5.5 The dependent wild bootstrap

The dependent wild bootstrap (DWB) of Shao (2010) is a resampling technique which can be used for smooth functions of the sample mean in the case where $\{X_1, X_2, \dots, X_T\}$ is a sample of real-valued weakly dependent random variables. Given this sample, the DWB constructs a bootstrap sample $\{X_1^*, X_2^*, \dots, X_T^*\}$ as follows:

$$X_t^* = \bar{X}_T + (X_t - \bar{X}_T)\varepsilon_{t,T}^*, \qquad t = 1, 2, \dots, T,$$

where $\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$ and $\{\varepsilon_{t,T}^*\}_{t=1}^{T}$ is a triangular scheme of weakly dependent random variables independent of $\{X_1, X_2, \dots, X_T\}$. More precisely, $\{\varepsilon_{t,T}^*\}$ is a stationary process with

$$\mathrm{E}(\varepsilon_{t,T}^*) = 0, \qquad \mathrm{Var}(\varepsilon_{t,T}^*) = 1, \qquad \mathrm{Cov}(\varepsilon_{s,T}^*, \varepsilon_{t,T}^*) = K\left(\frac{s-t}{\ell}\right),$$

where $K(\cdot)$ is a kernel function and ℓ is a bandwidth parameter. For the kernel we require that $K(0) = 1$, $K(x) = 0$ for $x \geq 1$, and $\int_{-\infty}^{\infty} K(u)e^{-iux}du \geq 0$ for all $x \in \mathbb{R}$.

In our Monte Carlo study we will investigate the performance of several unit root tests based on the DWB, where we use the Bartlett kernel as a choice for K and generate the $\varepsilon_{t,T}^*$ from a multivariate normal distribution.

Chapter 5

Bootstrap unit root tests

The past two decades have seen an almost bewildering array of bootstrap-based unit root tests emerging in the statistical literature. The first application of the bootstrap to unit root testing that we are aware of is by Rayner (1990), but research in this area started gaining momentum only somewhat later with the contributions of Nankervis and Savin (1996), Psaradakis (2001), Chang and Park (2003), Paparoditis and Politis (2003) and Park (2003). These applications rely on more traditional methods of bootstrapping dependent data, such as the sieve and block bootstrap procedures (and the countless variations thereof).

Owing to some of the deficiencies of these methods, such as the invalidity in the case of heteroskedastic time series data, alternative more modern methods for bootstrapping dependent data soon saw the light. An interesting application of the traditional (independent) wild bootstrap to unit root testing is bootstrap versions of Ng and Perron's M tests by Cavaliere and Taylor (2009a). Shortly thereafter Shao developed the dependent wild bootstrap, an extension of the traditional wild bootstrap to accommodate stationary time series data, which paved the way for numerous variations, such as the autoregressive and the block wild bootstrap (see e.g. Shao, 2011; Smeekes and Urbain, 2014a).

A promising recent development is the linear process bootstrap proposed by McMurry and Politis (2010), which sprouted from a new estimator for autocovariance matrices. Although little research has been done on the linear process bootstrap in the context of unit root testing, we will discuss the linear process unit root test as recently applied in this setting by Zou and Politis (2016).

Today there are many different unit root tests and almost as many bootstrap procedures. These procedures and their important implementational considerations are scattered over more than 20 years of literature, which makes it challenging for practitioners to apply these tests responsibly. Our aim in this chapter is not to describe all bootstrap unit root tests available in the literature, but to provide a concise but detailed selection of the most successful unit root tests, as well as important variations of these tests and considerations to bear in mind in implementation.

As in Chapter 3, suppose throughout that we have at our disposal a set of observations

$\{x_0, x_1, \ldots, x_T\}$ generated by the process

$$x_t = d_t + y_t, \qquad y_t = \rho y_{t-1} + u_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{5.1}$$

where d_t is a fixed deterministic component and $\{u_t\}$ is an unobserved stationary zero-mean error process. Our objective remains testing the hypothesis

$$H_0 : \rho = 1 \quad \text{vs.} \quad H_A : |\rho| < 1. \tag{5.2}$$

We now move on to describe the tests.

5.1 Test 1: Difference-based ADF sieve bootstrap test

The first bootstrap test we consider is the test proposed by Chang and Park (2003) which involves applying the sieve bootstrap to the differenced observations. The test is a generalisation of a bootstrap-based test proposed by Psaradakis (2001) in that it incorporates the use of the augmented version of the Dickey–Fuller statistics. The test is based on the observation that the process $\{y_t\}$ in (5.1) may be rewritten as

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

which under H_0 becomes

$$\Delta y_t = u_t.$$

Based on the observed series $\{x_0, x_1, \ldots, x_T\}$, Chang and Park (2003) propose the following bootstrap procedure:

1. Calculate the Dickey–Fuller statistics $DF_{\rho,q}$ and $DF_{t,q}$ based on the sample data. The series may be detrended if necessary.

2. Detrend the series $\{x_t\}$ according to the specification of the deterministic component $d_t = \boldsymbol{\psi}' \mathbf{z}_t$. Define the detrended series as

   $$\tilde{x}_t = x_t - \hat{\boldsymbol{\psi}}' \mathbf{z}_t,$$

   where $\hat{\boldsymbol{\psi}}$ is an appropriate estimate for $\boldsymbol{\psi}$ in a regression of x_t on $\mathbf{z}_t$. See remark (a) and (b) below.

3. For some fixed positive integer q', fit an AR(q') model to $\{\hat{u}_t\}$, where $\hat{u}_t = \Delta \tilde{x}_t = \tilde{x}_t - \tilde{x}_{t-1}$. This yields coefficient estimates $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{q'}$ from which we may obtain the residuals

   $$\hat{\varepsilon}_t = \hat{u}_t - \sum_{j=1}^{q'} \hat{\beta}_j \hat{u}_{t-j}, \qquad t = q', \ldots, T.$$

   See remark (c) below.

4. (i) Draw a sample of size $T - q'$ randomly with replacement from the centred residuals

$$\hat{\varepsilon}_t - \frac{1}{T-q'} \sum_{t=q'+1}^{T} \hat{\varepsilon}_t$$

to generate the bootstrap sample $\{\varepsilon_t^*\}$.

(ii) Construct bootstrap errors u_t^* using the recursion

$$u_t^* = \sum_{j=1}^{q'} \hat{\beta}_j u_{t-j}^* + \varepsilon_t^*. \tag{5.3}$$

See remark (d) below.

(iii) Recursively generate the bootstrap series $\{y_t^*\}_{t=1}^T$ by

$$y_t^* = y_{t-1}^* + u_t^*, \qquad y_0^* = 0. \tag{5.4}$$

See remark (e) below.

(iv) For some fixed positive integer q^*, calculate the normalised coefficient statistic

$$DF_\rho^* = \frac{T(\hat{\beta}_0^* - 1)}{1 - \sum_{j=1}^{q^*} \hat{\beta}_j^*} \tag{5.5}$$

and the corresponding t-statistic

$$DF_t^* = \frac{\hat{\beta}_0^* - 1}{\mathrm{SE}(\hat{\beta}_0^*)}, \tag{5.6}$$

where $\hat{\beta}_0^*$ and $\hat{\beta}_j^*$ are the OLS estimators for ρ^* and β_j^* in the ADF test regression

$$y_t^* = (\boldsymbol{\psi}^*)' \mathbf{z}_t + \beta_0^* y_{t-1}^* + \sum_{j=1}^{q^*} \beta_j^* \Delta y_{t-j}^* + \varepsilon_{t,q^*}^*.$$

5. Repeat step 4 a large number of times, say B, to obtain the ordered replications $DF^*(1) \le \ldots \le DF^*(B)$. Reject the null hypothesis whenever $DF < DF^*(\lfloor B\alpha \rfloor)$ (assuming a significance level of α). Here DF stands for either $DF_{\rho,q}$ or $DF_{t,q}$.

Remarks.

(a) Chang and Park (2003) do not recommend a specific method to be used for detrending the series $\{y_t\}$. While Cavaliere and Taylor (2009a) suggest using local GLS detrending in Chang and Park's procedure, we will consider detrending the series using both OLS and GLS in our Monte Carlo study.

(b) Smeekes (2013) shows that detrending should not only be done for the construction of the test statistic (if required), but also in step 2 of the above bootstrap algorithm. The two methods used for detrending need not be the same and, from a theoretical perspective, should be treated separately.

(c) For estimation of the AR(q') model in step 3, Psaradakis (2001) recommends making use of Yule–Walker equations. This ensures that the generated bootstrap innovations u_t^* admit a one-sided MA(∞) representation. A suitable value for the lag order q' may be chosen using a complexity-penalised likelihood criterion such as Akaike's information criterion (AIC), as recommended by Bühlmann (1997, 1998). Alternatively, sequential tests for the significance of the coefficient on the longest lag may be used (see e.g. Hall, 1994; Ng and Perron, 1995).

(d) Generating the bootstrapped error process $\{u_t^*\}$ in (5.3) requires initialisation. Following Chang and Park (2003), we make use of a burn-in period by generating a large number of u_t^* and then discarding the superfluous initial observations such that we are left with a series $\{u_t^*\}$ consisting of exactly T observations.

(e) Chang and Park (2003) show that the initial value y_0^* in (5.4) does not affect the asymptotic bootstrap distributions of the statistics DF^*, as long as it is stochastically bounded. They therefore recommend setting $y_0^* = 0$ when using demeaned or detrended data.

(f) It might feel intuitive choosing $q = q' = q^*$ in the above procedure. Although Chang and Park (2003) seem to make no distinction between these three lag lengths, Smeekes (2013) remarks that it is important to allow for lag selection of q^* in the bootstrap as this will improve the finite-sample properties of the test (also see Richard, 2009). Smeekes also allows for q and q' to be different and to be chosen independently using information criterion.

5.2 Test 2: Residual-based ADF sieve bootstrap test

Palm et al. (2008) propose a modified test based on the difference-based sieve bootstrap test by Chang and Park (2003). Their test is based on residuals and they construct it in the same way as the ADF coefficient test of Paparoditis and Politis (2005b). Palm et al. (2008) also show that the bootstrap distributions of the statistics obtained in this way converge to the same limit distributions as the test statistics.

The new algorithm differs from Test 1 only in step 3. For convenience we provide the complete algorithm of Palm et al. (2008) here:

1. Calculate the Dickey–Fuller statistics DF_ρ and DF_t based on the sample data. The series may be detrended if necessary.

2. Detrend the series $\{x_t\}$ according to the specification of the deterministic component $d_t = \boldsymbol{\psi}'\mathbf{z}_t$. Define the detrended series as

$$\tilde{x}_t = x_t - \hat{\boldsymbol{\psi}}'\mathbf{z}_t,$$

where $\hat{\boldsymbol{\psi}}$ is an appropriate estimate for $\boldsymbol{\psi}$ in a regression of x_t on $\mathbf{z}_t$.

3. For some fixed integer q', calculate the residuals

$$\hat{\varepsilon}_t = \tilde{y}_t - \hat{\beta}_0\tilde{y}_{t-1} - \sum_{j=1}^{q'}\hat{\beta}_j\Delta\tilde{y}_{t-j}, \qquad t = q', \ldots, T,$$

where $\hat{\beta}_0$ and the $\hat{\beta}_j$ are obtained by OLS in a q'th order ADF regression. In their simulation study, Palm et al. (2008) use OLS estimation to obtain $\hat{\beta}_0$.

4. (i) Draw a sample of size $T - q'$ randomly with replacement from the centred residuals

$$\hat{\varepsilon}_t - \frac{1}{T-q'}\sum_{t=q'+1}^{T}\hat{\varepsilon}_t$$

to generate the bootstrap sample $\{\varepsilon_t^*\}$.

(ii) Construct bootstrap errors u_t^* using the recursion

$$u_t^* = \sum_{j=1}^{q'}\hat{\beta}_j u_{t-j}^* + \varepsilon_t^*.$$

(iii) Recursively generate the bootstrap series $\{y_t^*\}$ by $y_t^* = y_{t-1}^* + u_t^*$, $y_0^* = 0$.

(iv) From the ADF regression

$$y_t^* = \beta_0^* y_{t-1}^* + \sum_{j=1}^{q^*}\beta_j^*\Delta y_{t-j}^* + \varepsilon_{t,q^*}^*,$$

calculate the statistics DF_ρ^* and DF_t^* as defined in (5.5) and (5.6).

5. Repeat step 4 a large number of times, say B, to obtain the ordered replications $DF^*(1) \le \ldots \le DF^*(B)$. Reject the null hypothesis whenever $DF < DF^*(\lfloor B\alpha \rfloor)$ (assuming a significance level of α). Again, DF stands for either DF_ρ or DF_t.

5.3 Test 3: Residual-based ADF block bootstrap test

The residual-based block bootstrap (RBB) of Paparoditis and Politis (2003) is a block bootstrap method applied to the residuals of a regression of y_t on y_{t-1}. Regardless of the process that generated the observations $\{x_t\}$, they start by defining a new series $\{v_t\}$ by

$$v_t = x_t - \alpha - \rho x_{t-1}, \qquad t = 1, 2, \ldots, \tag{5.7}$$

where $\alpha = \mathrm{E}(x_t - \rho x_{t-1})$.

Paparoditis and Politis (2003) propose the following bootstrap testing procedure:

1. Let $\tilde{\rho}$ be a consistent estimator for ρ in (5.7). In the context of unit root testing, such an estimator may be the OLS estimator $\hat{\beta}_0$ for β_0 in the ADF regression

$$x_t = \alpha_0 + \beta_0 x_{t-1} + \sum_{j=1}^{q'} \beta_j \Delta x_{t-j} + \varepsilon_t. \tag{5.8}$$

   Although other options exist, for our purposes we will choose $\tilde{\rho} = \hat{\beta}_0$ throughout, as suggested by Paparoditis and Politis (2003). Note the inclusion of the constant mean term α_0 *regardless of the specification of the deterministic component* of the test.

2. Obtain the centred residuals

$$\hat{v}_t = x_t - \tilde{\rho} x_{t-1} - \frac{1}{T-1} \sum_{s=2}^{T} (x_s - \tilde{\rho} x_{s-1})$$

   and the centred differences

$$D_t = x_t - x_{t-1} - \frac{1}{T-1} \sum_{s=2}^{T} (x_s - x_{s-1}).$$

3. (i) For a fixed block length b, draw $k = \lfloor (T-1)/b \rfloor$ index points $i_0, i_1, \ldots, i_{k-1}$ from the uniform distribution on the set $\{1, 2, \ldots, T-b\}$. These points will serve as starting points for the blocks of centred residuals $\hat{v}_t$.

   (ii) Set $v_1^* = D_1^* = 0$. Let $T' = kb + 1$ and, for $t = 2, 3, \ldots, T'$, define the bootstrap series

$$v_t^* = \hat{v}_{i_m + t - mb - 1} \quad \text{and} \quad D_t^* = D_{i_m + t - mb - 1},$$

   where $m = \lfloor (t-2)/b \rfloor$.

   (iii) Construct the bootstrap sample recursively by

$$x_t^* = \begin{cases} x_0 & \text{for } t = 0, \\ \tilde{\alpha} + x_{t-1}^* + v_t^* & \text{for } t = 1, 2, \ldots, T', \end{cases}$$

   where $\tilde{\alpha} \equiv 0$ in the case of a constant (possibly zero) mean or, in the case of a linear time trend, a consistent estimator for the drift parameter α in (5.7). Following Paparoditis and Politis (2003) we will, in our Monte Carlo study, take $\tilde{\alpha}$ as the OLS estimator for α_0 obtained in the regression in (5.8).

   (iv) Calculate the normalised coefficient statistic

$$DF_\rho^* = \frac{T(\hat{\beta}_0^* - 1)}{1 - \sum_{j=1}^{q^*} \hat{\beta}_j^*} \tag{5.9}$$

   and the corresponding t-statistic

$$DF_t^* = \frac{\hat{\beta}_0^* - 1}{\mathrm{SE}(\hat{\beta}_0^*)}, \tag{5.10}$$

where $\hat{\beta}_0^*$ and the $\hat{\beta}_j^*$ are the OLS estimators for β_0^* and the β_j^*, respectively, in the ADF test regression

$$x_t^* = d_t + \beta_0^* x_{t-1}^* + \sum_{j=1}^{q^*} \beta_j^* D_{t-j}^* + \varepsilon_{t,q^*}^*. \tag{5.11}$$

4. Repeat step 3 a large number of times, say B, to obtain the ordered replications $DF^*(1) \leq \ldots \leq DF^*(B)$. Reject the null hypothesis whenever $DF < DF^*(\lfloor B\alpha \rfloor)$ (assuming a significance level of α). Here, DF stands for either DF_ρ or DF_t.

Remarks.

(a) In their theoretical results, Paparoditis and Politis (2003) show that their procedure is consistent in the case where d_t, the deterministic specification in the test, is a constant (possibly zero) and in the case where d_t is a linear time trend. However, in their numerical study they only investigate the finite-sample performance of their procedure for the case where d_t is a constant. We will include both specifications in our Monte Carlo study.

(b) Note that the lagged differences D_t^* appearing in (5.11) are differences calculated from the original sample $\{x_t\}$ and *not from the bootstrap sample* $\{x_t^*\}$. The D_t^* are defined in step 3(ii) of the procedure.

An important feature of the block bootstrap is that the innovations u_t need not belong to the class of linear processes to be strong mixing, as opposed to the sieve bootstrap. Hence, the block bootstrap allows for a class of possibly non-linear processes whereas the sieve bootstrap does not. However, Palm et al. (2008) note that, in the case where both the block and sieve bootstrap are consistent, it seems that the sieve bootstrap performs better than the block bootstrap.

5.4 Test 4: Difference-based ADF block bootstrap test

Paparoditis and Politis (2003) also consider a difference-based version of the block bootstrap test described in the previous section. They base their modification on differences (DBB) instead of residuals by simply taking $\tilde{\rho} = 1$ in Test 3. The asymptotic validity of this version is established by Paparoditis and Politis (2003).

5.5 Test 5: ADF wild bootstrap test

Although already introduced in the 1980s by Wu (1986), the traditional (independent) wild bootstrap was first applied in the context of unit root testing by Cavaliere and Taylor (2008, 2009a,b). To account for serial correlation in the innovations, they advocate the introduction of a sieve, or a recolouring, into the wild bootstrap procedure when constructing the bootstrap innovations. Apart from showing that the wild bootstrap is valid in terms of first-order

asymptotics, they demonstrate that the wild bootstrap is more robust to serial correlation in the innovation process than the test based on asymptotic critical values and the block-based bootstrap methods. The improvements offered by the wild bootstrap is even more prominent in the presence of conditional heteroskedasticity in the error process.

In this section we present application of several wild bootstrap procedures as applied to the ADF unit root test by Smeekes and Urbain (2014a). As alternatives to using the sieve advocated by Cavaliere and Taylor, they consider three other versions of the wild bootstrap: the blockwise, dependent and autoregressive wild bootstrap. The bootstrap M tests originally considered by Cavaliere and Taylor will be presented in the next section.

The wild bootstrap testing procedure is as follows:

1. Detrend $\{x_t\}$ according to the specification of the deterministic component d_t to obtain the detrended series $\{\tilde{x}_t\}$.

2. For some fixed integer q', from the ADF regression

$$\tilde{x}_t = \beta_0 \tilde{x}_{t-1} + \sum_{j=1}^{q'} \beta_j \Delta \tilde{x}_{t-j} + \varepsilon_t, \qquad t = q', \dots, T,$$

obtain the residuals

$$\hat{\varepsilon}_t = \tilde{x}_t - \hat{\beta}_0 \tilde{x}_{t-1} - \sum_{j=1}^{q'} \hat{\beta}_j \Delta \tilde{x}_{t-j}, \qquad t = q', \dots, T,$$

where $\hat{\beta}_0$ and the $\hat{\beta}_j$ are OLS estimators for β_0 and the β_j, respectively.

3. (i) Let $\{w_t\}_{t=0}^T$ be a sequence of random variables generated *independently of* $\hat{\varepsilon}_t$ according to one of the wild bootstrap schemes described in the following subsections. Obtain the bootstrap errors ε_t^* according to the device

$$\varepsilon_t^* = w_t \hat{\varepsilon}_t.$$

(ii) Set $u_t^* = \varepsilon_t^*$ and recursively construct the bootstrap series $\{y_t^*\}$ as $y_t^* = y_{t-1}^* + u_t^*$ with $y_0^* = 0$.

(iii) Using the same detrending method used in the calculation of the test statistic DF, detrend the series y_t^* to obtain $\tilde{y}_t^*$.

(iv) For some fixed integer q^*, from the ADF regression

$$\tilde{y}_t^* = \beta_0^* \tilde{y}_{t-1}^* + \sum_{j=1}^{q^*} \beta_j^* \Delta \tilde{y}_{t-j} + \varepsilon_{t,q^*}^*, \qquad t = q^*, \dots, T,$$

obtain the residuals

$$\hat{\varepsilon}_{t,q^*}^* = \tilde{y}_t^* - \hat{\beta}_0^* \tilde{y}_{t-1}^* - \sum_{j=1}^{q^*} \hat{\beta}_j^* \Delta \tilde{y}_{t-j}^*, \qquad t = q^*, \dots, T,$$

where $\hat{\beta}_0^*$ and the $\hat{\beta}_j^*$ are OLS estimators for β_0^* and the β_j^*, respectively.

(v) Calculate the normalised coefficient statistic

$$DF_\rho^* = \frac{T(\hat{\beta}_0^* - 1)}{1 - \sum_{j=1}^{q^*} \hat{\beta}_j^*} \tag{5.12}$$

and the corresponding t-statistic

$$DF_t^* = \frac{\hat{\beta}_0^* - 1}{\mathrm{SE}(\hat{\beta}_0^*)}. \tag{5.13}$$

4. Repeat step 4 a large number of times, say B, to obtain the ordered replications $DF^*(1) \leq \ldots \leq DF^*(B)$. Reject the null hypothesis whenever $DF < DF^*(\lfloor B\alpha \rfloor)$ (assuming a significance level of α). Again, DF stands for either DF_ρ or DF_t.

The *traditional wild bootstrap*, originally introduced by Wu (1986), involves generating the sequence $\{w_t\}$ by randomly sampling uncorrelated random variables with zero mean and unit variance. Although many alternative suggestions exist in the literature, it is common practice to independently generate the w_t from a standard normal distribution (see e.g. Cavaliere and Taylor, 2009a; Shao, 2010). Accuracy of the wild bootstrap can be improved by generating the w_t from an asymmetric distribution with $\mathrm{E}(w_t^3) = 1$ (see Liu, 1988). Some examples are given in Section 4.4. However, Cavaliere and Taylor (2009a, p. 401) remark that in their Monte Carlo studies they found no discernible differences between using the normal, the Mammen or the Rademacher distributions (also see Gonçalves and Kilian, 2004, p. 105).

It is worth noting that, conditionally on the $\hat{\varepsilon}_t$, the bootstrap innovations u_t^* have zero mean and variance $\hat{\varepsilon}_t^2$. The wild bootstrap therefore has the desirable property that it preserves the pattern of heteroskedasticity present in the original innovations u_t.

In contrast, the independent wild bootstrap fails to replicate serial correlation from the true innovations u_t. Since $\mathrm{E}^*(w_t) = \mathrm{E}(w_t) = 0$, notice that for all t we have

$$\mathrm{E}^*(u_t^*) = \mathrm{E}^*(w_t \hat{\varepsilon}_t) = \hat{\varepsilon}_t \, \mathrm{E}^*(w_t) = 0.$$

Hence, for any $s \neq t$,

$$\mathrm{Cov}^*(u_s^*, u_t^*) = \mathrm{E}^*(u_s^* u_t^*) = \mathrm{E}^*(w_s \hat{\varepsilon}_s w_t \hat{\varepsilon}_t) = \hat{\varepsilon}_s \hat{\varepsilon}_t \, \mathrm{E}^*(w_s) \mathrm{E}^*(w_t) = 0.$$

This means that the generated bootstrap innovations are conditionally uncorrelated and any serial correlation in the u_t will not be accounted for. Therefore, in cases where the true innovations u_t are autocorrelated, the independent wild bootstrap will fail to replicate the long-run covariance matrix of the innovations and consequently will be asymptotically invalid.

One remedy of this problem involves incorporating autocorrelation into the test statistic, such as adding lagged difference terms to the Dickey–Fuller regression. Although this approach works well for our purposes, Smeekes and Urbain (2014a) note that in multivariate settings (panel data), for example, this approach is often infeasible.

Other alternatives have been proposed in the literature to address this problem, some of which we discuss in the following subsections.

5.5.1 Recolouring the wild bootstrap

Advocated by Cavaliere and Taylor (2008, 2009a,b), one can account for autocorrelation in the errors by introducing a sieve when constructing the bootstrap errors. This is achieved by replacing step 3(ii) by the following:

3. (ii) Recursively generate $\{u_t^*\}$ according to

$$u_t^* = \sum_{j=1}^{q'} \hat{\beta}_j \Delta u_{t-j}^* + \varepsilon_t^*,$$

and then generate the bootstrap sample $\{y_t^*\}$ by $y_t^* = y_{t-1}^* + u_t^*$ with $y_0^* = 0$.

Using the estimated coefficients of the lagged difference terms included in step 2, this modified step reintroduces serial correlation into the construction of the bootstrap innovations.

5.5.2 Dependent wild bootstrap

As an extension of the traditional wild bootstrap, Shao (2010) proposed the so-called *dependent wild bootstrap* which allows dependence to be reintroduced into the bootstrap series $\{u_t^*\}$. This is accomplished by generating $\{w_t\}$ as a stationary process with

$$\mathrm{E}(w_t) = 0, \quad \mathrm{Var}(w_t) = 1, \quad \mathrm{Cov}(w_s, w_t) = K\left(\frac{s-t}{\ell}\right),$$

where $K(\cdot)$ is a kernel function and ℓ is a bandwidth parameter. For the kernel we require that $K(0) = 1$, $K(x) = 0$ for $x \geq 1$, and $\int_{-\infty}^{\infty} K(u)e^{-iux}du \geq 0$ for all $x \in \mathbb{R}$.

In our Monte Carlo study we follow Smeekes and Urbain (2014a) and generate the w_t from a multivariate normal distribution with the moment properties given above. As recommended by Shao (2010), we choose

$$K(t) = \frac{\int_{-1}^{1} w_{0.43}(x) w_{0.43}(x + |t|)dx}{\int_{-1}^{1} w_{0.43}^2(x)dx},$$

where

$$w_c(x) = \begin{cases} t/c & \text{if } 0 \leq t < c, \\ 1 & \text{if } c \leq t \leq 1-c, \\ (1-t)/c & \text{if } 1-c < t \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

5.5.3 Autoregressive wild bootstrap

Another modification of the wild bootstrap considered by Smeekes and Urbain (2014a) is the autoregressive wild bootstrap. For this type of bootstrap, the w_t are generated as follows.

For some fixed $\gamma \in [0,1)$, independently generate $\xi_0, \xi_1, \ldots, \xi_T$ from an $N(0, 1-\gamma^2)$ distribution and let

$$w_t = \gamma w_{t-1} + \xi_t, \qquad t = 1, 2, \ldots, T,$$

with $w_0 \sim N(0,1)$. Following Smeekes and Urbain (2014a) we choose $\gamma = \theta^{1/\ell}$ in our Monte Carlo study, where $\theta = 0.01$ and ℓ is the same as the bandwidth parameter used for the dependent wild bootstrap.

5.6 Test 6: Wild bootstrap M tests

We now present the wild bootstrap procedure proposed by Cavaliere and Taylor (2009a) for the M tests of Ng and Perron (2001) defined in Section 3.5. Note again that the traditional and recoloured wild bootstrap procedures presented in the previous section are based on the procedure given here.

1. Detrend $\{x_t\}$ according to the specification of the deterministic component d_t to obtain the detrended series $\{\tilde{x}_t\}$.

2. For some fixed integer q', from the ADF regression

$$\tilde{x}_t = \beta_0 \tilde{x}_{t-1} + \sum_{j=1}^{q'} \beta_j \Delta \tilde{x}_{t-j} + \varepsilon_t, \qquad t = q', \ldots, T,$$

obtain the residuals

$$\hat{\varepsilon}_t = \tilde{x}_t - \hat{\beta}_0 \tilde{x}_{t-1} - \sum_{j=1}^{q'} \hat{\beta}_j \Delta \tilde{x}_{t-j}, \qquad t = q', \ldots, T,$$

where $\hat{\beta}_0$ and the $\hat{\beta}_j$ are OLS estimators for β_0 and the β_j, respectively.

3. (i) Let $\{w_t\}_{t=0}^T$ be an i.i.d. sequence of $N(0,1)$ random variables. Generate the bootstrap errors ε_t^* according to the device

$$\varepsilon_t^* = w_t \hat{\varepsilon}_t.$$

(ii) Set $u_t^* = \varepsilon_t^*$ and recursively generate the bootstrap series $\{y_t^*\}$ by $y_t^* = y_{t-1}^* + u_t^*$ with $y_0^* = 0$.

(iii) Using the procedure described in Section 3.4, detrend the series $\{y_t^*\}$ to obtain $\{\tilde{y}_t^*\}$.

(iv) For some fixed integer q^*, from the ADF regression

$$\tilde{y}_t^* = \beta_0^* \tilde{y}_{t-1}^* + \sum_{j=1}^{q^*} \beta_j^* \Delta \tilde{y}_{t-j} + \varepsilon_{t,q^*}^*, \qquad t = q^*, \ldots, T,$$

obtain the residuals

$$\hat{\varepsilon}_{t,q^*}^* = \tilde{y}_t^* - \hat{\beta}_0^* \tilde{y}_{t-1}^* - \sum_{j=1}^{q^*} \hat{\beta}_j^* \Delta \tilde{y}_{t-j}^*, \qquad t = q^*, \ldots, T,$$

where $\hat{\beta}_0^*$ and the $\hat{\beta}_j^*$ are OLS estimators for β_0^* and the β_j^*, respectively.

(v) Calculate the test statistics

$$MZ_\alpha^* = \frac{T^{-1}(\tilde{y}_T^*)^2 - (s_{AR}^*)^2}{2T^{-2}\sum_{t=1}^T (\tilde{y}_{t-1}^*)^2}, \qquad MSB^* = \sqrt{\frac{1}{T}\sum_{t=1}^T (\tilde{y}_{t-1}^*)^2 / (s_{AR}^*)^2},$$

and $MZ_t^* = MZ_\alpha^* \cdot MSB^*$, where

$$s_{AR}^* = \frac{\hat{\sigma}^*}{1 - \sum_{j=1}^{q^*} \hat{\beta}_j^*},$$

with $\hat{\sigma}^* = (T - q^* + 1)^{-1} \sum_{j=q^*}^T \hat{\varepsilon}_{t,q^*}^*$.

4. Repeat step 2 a large number of times, say B, to obtain replications $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$. An approximate critical value of the test (at a significance level of α) is given by $\hat{\theta}_{\lfloor B\alpha \rfloor}^*$.

5.6.1 Recoloured wild bootstrap M tests

As was described in the previous section, this variation on the wild bootstrap is different only in step 3(ii), which is replaced by

2. (iii) Recursively generate $\{u_t^*\}$ according to

$$u_t^* = \sum_{j=1}^{q'} \hat{\beta}_j \Delta u_{t-j}^* + \varepsilon_t^*, \qquad t = q', \ldots, T,$$

and then generate the bootstrap sample $\{y_t^*\}$ by $y_t^* = y_{t-1}^* + u_t^*$ with $y_0^* = 0$.

5.7 Test 7: Linear process bootstrap

In light of the fact that most unit root tests suffer from size distortions when moving average noise exists, Zou and Politis (2016) recently proposed a bootstrap test designed specifically to address this issue. Their approach makes use of the so-called *linear process bootstrap* first introduced by McMurry and Politis (2010). Jentsch, Politis et al. (2015) showed that, for the case of the mean, the linear process bootstrap shows good asymptotic and empirical performance.

The linear process bootstrap unit root test proposed by Zou and Politis (2016) is given below. Note that their procedure is originally proposed only for the case where the test contains no deterministic component. We have added step 1 to investigate whether their procedure might also be valid in the case where a deterministic component is specified.

1. Detrend $\{x_t\}$ according to the specification of the deterministic component $d_t^\dagger$.

2. For some fixed integer q', calculate the residuals

$$\hat{v}_t = \tilde{y}_t - \hat{\beta}_0 \tilde{y}_{t-1} - \sum_{j=1}^{q'} \hat{\beta}_j \Delta \tilde{y}_{t-j}, \qquad t = q', \ldots, T,$$

where $\hat{\beta}_0$ and the $\hat{\beta}_j$ are obtained by OLS in a q'th order ADF regression. Let $\check{\boldsymbol{v}} = (\check{v}_{q'+1}, \ldots, \check{v}_t)'$ denote the centred residuals, i.e.

$$\check{v}_t = \hat{v}_t - \frac{1}{T-q'} \sum_{t=q'+1}^{T} \hat{v}_t.$$

3. Define

$$\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_{q'+1}, \ldots, \hat{\varepsilon}_T)' = \hat{\Sigma}_{\hat{\boldsymbol{v}}}^{-1/2} \check{\boldsymbol{v}},$$

where $\hat{\Sigma}_{\hat{\boldsymbol{v}}}^{1/2}$ is a lower triangular matrix that satisfies the Cholesky decomposition $\hat{\Sigma}_{\hat{\boldsymbol{v}}}^{1/2}(\hat{\Sigma}_{\hat{\boldsymbol{v}}}^{1/2})' = \hat{\Sigma}_{\hat{\boldsymbol{v}}}$, with $\hat{\Sigma}_{\hat{\boldsymbol{v}}}$ a positive definite estimator for the covariance matrix $\Sigma = \mathrm{Var}(\hat{\boldsymbol{v}})$. Also, define

$$\check{\varepsilon}_t = \left(\hat{\varepsilon}_t - \frac{1}{T-q'} \sum_{t=q'+1}^{T} \hat{v}_t \right) \Big/ \hat{\sigma}_{\hat{\varepsilon}},$$

where $\hat{\sigma}_{\hat{\varepsilon}}^2 = T^{-1}\Sigma_{t=1}^{T}(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}_t)^2$ and $\bar{\hat{\varepsilon}}_t = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t$.

4. (i) Draw a sample of size T randomly with replacement from $\{\check{\varepsilon}_{q'+1}, \ldots, \check{\varepsilon}_T\}$ to generate the bootstrap sample $\boldsymbol{\varepsilon}^* = (\varepsilon^*_{q'+1}, \ldots, \varepsilon^*_T)'$.

   (ii) Calculate bootstrap innovations u_t^* as

$$\boldsymbol{v}^* = \hat{\Sigma}_{\hat{\boldsymbol{v}}}^{1/2} \boldsymbol{\varepsilon}^*.$$

   (iii) Recursively generate the bootstrap series $\{y_t^*\}$ by $y_t^* = y_{t-1}^* + u_t^*$ with $y_0^* = 0$.

   (iv) From the ADF regression

$$y_t^* = (\hat{\boldsymbol{\psi}}^*)' \mathbf{z}_t + \beta_0^* y_{t-1}^* + \sum_{j=1}^{q^*} \beta_j^* \Delta y_{t-j}^* + \varepsilon_{t,q^*}^*,$$

   calculate the statistics DF_ρ^* and DF_t^* as defined in (5.5) and (5.6).

5. Repeat step 4 a large number of times, say B, to obtain the ordered replications $DF^*(1) \leq \ldots \leq DF^*(B)$. Reject the null hypothesis whenever $DF < DF^*(\lfloor B\alpha \rfloor)$ (assuming a significance level of α). Again, DF stands for either DF_ρ or DF_t.

Estimation of Σ

To estimate the autocovariance matrix $\Sigma = \mathrm{Var}(\hat{\boldsymbol{v}})$, we follow Zou and Politis (2016) and use an estimator recently proposed by McMurry and Politis (2010). Their estimator is a generalisation of a banded estimator introduced by Wu and Pourahmadi (2009). Although the estimator of McMurry and Politis has a much broader application, it paved the way for the development of the linear process bootstrap.

Their estimation procedure is as follows. First, define the sample autocovariance function

$$\hat{\gamma}_{\hat{\boldsymbol{v}}}(h) = \frac{1}{T} \sum_{t=|h|+1}^{T} \hat{v}_t \hat{v}_{t-|h|}$$

and the weight function

$$K(x) = \begin{cases} 1 & \text{if } |x| \le 1, \\ 2 - |x| & \text{if } 1 < |x| \le 2, \\ 0 & \text{if } |x| > 2. \end{cases}$$

This weight function is the trapezoid kernel function of Politis and Romano (1992). Also define the tapered covariance matrix estimator

$$\tilde{\Sigma}_{\hat{\boldsymbol{v}}} = [K_h(i-j)\hat{\gamma}_{\hat{\boldsymbol{v}}}(i-j)]_{i,j=1}^{T},$$

where $K_h(x) = K(x/h)$. Other choices for the weight function $K(\cdot)$ are possible and a general form is given in McMurry and Politis (2010).

Now suppose that $\tilde{\Sigma}_{\hat{\boldsymbol{v}}} = TDT'$, where T is orthogonal and $D = \text{diag}(d_1, d_2, \ldots, d_T)$ is diagonal. Then

$$\hat{\Sigma}_{\hat{\boldsymbol{v}}} = T\hat{D}T',$$

where $\hat{D} = \text{diag}(\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_T)$ is a diagonal matrix with $\hat{d}_j = \max(d_j, T^{-1}\hat{\gamma}_{\hat{\boldsymbol{v}}}(0))$. It is shown by McMurry and Politis (2010) that $\hat{\Sigma}_{\hat{\boldsymbol{v}}}$ is a positive definite matrix.

Like some other bootstrap procedures, a major drawback encountered in the application of the procedure is that a choice of the bandwidth parameter h is required. Although Zou and Politis (2016) provide an algorithm for choosing h, more research is required to find an appropriate choice for h. Some guidance based on Monte Carlo results is given in Remark 3 of McMurry and Politis (2010).

Chapter 6

Simulation study

In this chapter we will evaluate and compare the finite-sample performance of the asymptotic and bootstrap unit root tests discussed in Chapter 3 and Chapter 5, respectively. We start by describing the Monte Carlo setup in Section 6.1. In Section 6.2 we provide simulation results for the case where the innovations are generated from an i.i.d., AR or MA process. We then extend the study in Section 6.3 by considering cases where the innovations have conditional heteroskedastic behaviour. The latter includes the popular GARCH model and a few of its many variations. We end the chapter with some general conclusions.

6.1 Monte Carlo setup

We compare the finite-sample size and power performance of the considered tests when applied to a series $\{y_0, y_1, \ldots, y_T\}$ generated by the process

$$x_t = d_t + y_t, \qquad y_t = \rho_T y_{t-1} + u_t, \tag{6.1}$$

where $\{u_t\}$ is a stationary process which will be specified later. To evaluate the exact size of each test, we set $\rho_T = 1$ in (6.1) to reflect the null hypothesis of a unit root. We evaluate the power properties at local alternatives of the form

$$\rho_T = 1 - c/T,$$

where we choose $c = 7.0$ in the case where the specification of the deterministic component in the test is a constant mean, and $c = 13.5$ in the case where the deterministic component is a linear time trend. This corresponds to the choice of ρ_T in Ng and Perron (2001) and Cavaliere and Taylor (2009a).

The process in (6.1) requires initialisation. Although it has been shown that the initial value of the process can have a large impact on the size and power properties of unit root tests (Harvey and Leybourne, 2005; Harvey, Leybourne and Taylor, 2009), this effect and remedial measures merit a separate study and fall outside the scope of this study. To simulate realistic situations where little is unknown about the initial value of the process we will simply make use of a burn-in period of length $\lceil T/2 \rceil$, which we discard before performing the tests.

For all tests that require a lag selection, we employed the modified Akaike information criterion (MAIC, refer to Section 3.6) proposed by Ng and Perron (2001). For comparison, lag selection is done on both GLS detrended data and, as recommended by Perron and Qu (2007), on OLS detrended data. Any test that requires additional configuration, was configured as described in Chapter 5.

All tests were performed on 50 000 independent realisations of $\{y_T\}$ generated according to (6.1) for sample sizes $T = 50, 100, 250$. Bootstrap tests were performed by means of the warp-speed method of Giacomini, Politis and White (2013).

The tests and notation

In this chapter we will refer to the various tests by making use of the notation given in Table 6.1. We make the following remarks:

- Whenever these symbols have the superscript PQ, it indicates that lag selection was based on OLS detrended data instead of GLS detrended data. M tests for which lag selection is based on OLS detrended data are indicated with the suffix PQ.
- Tests marked with a bar are detrended using GLS instead of OLS as proposed by the original authors.
- The M tests are all detrended using GLS.

Table 6.1: Notation for the various asymptotic and bootstrap tests

Notation	Description
DF_t	Asymptotic ADF t-test
DF_ρ	Asymptotic ADF coefficient test
Z_t	Phillips–Perron t-test
Z_ρ	Phillips–Perron coefficient test
ERS_t	GLS detrended ADF test of Elliott et al. (1996)
P_T	Efficient test proposed by Elliott et al. (1996)
CP_t	Difference-based sieve bootstrap ADF t-test of Chang and Park (2003)
CP_ρ	Difference-based sieve bootstrap ADF coefficient test of Chang and Park (2003)
PS_t	Residual-based sieve bootstrap ADF t-test of Chang and Park (2003)
PS_ρ	Residual-based sieve bootstrap ADF coefficient test of Chang and Park (2003)
PaP_t	Residual-based block bootstrap ADF t-test of Paparoditis and Politis (2003)
PaP_ρ	Residual-based block bootstrap ADF coefficient test of Paparoditis and Politis (2003)
PaP_t^{diff}	Difference-based block bootstrap ADF t-test of Paparoditis and Politis (2003)
PaP_ρ^{diff}	Difference-based block bootstrap ADF coefficient test of Paparoditis and Politis (2003)
RWB_t	Recoloured wild bootstrap ADF t-test
RWB_ρ	Recoloured wild bootstrap ADF coefficient test
DWB_t	Dependent wild bootstrap ADF t-test
DWB_ρ	Dependent wild bootstrap ADF coefficient test
AWB_t	Autoregressive wild bootstrap ADF t-test
AWB_ρ	Autoregressive wild bootstrap ADF coefficient test
LPB_t^0	Linear process bootstrap test of Zou and Politis (2016)
LPB_t	Adapted linear process bootstrap ADF t-test (allows addition of lags)
MZ_α	Asymptotic M test of Ng and Perron (2001)
MZ_t	Asymptotic M test of Ng and Perron (2001)
MSB	Asymptotic M test of Ng and Perron (2001)
MP_T	Asymptotic M test of Ng and Perron (2001)
M^{iwb}	Independent wild bootstrap implementation of the above M tests (Cavaliere and Taylor, 2009a)
M^{rwb}	Recoloured wild bootstrap implementation of the above M tests (Cavaliere and Taylor, 2009a)
M^{dwb}	Dependent wild bootstrap implementation of the above M tests
M^{awb}	Autoregressive wild bootstrap implementation of the above M tests

6.2 ARMA innovations

In this section we consider the case where the innovations $\{u_t\}$ are generated according to either the AR(1) process

$$u_t = \phi u_{t-1} + \varepsilon_t$$

or the MA(1) process

$$u_t = \varepsilon_t + \theta \varepsilon_{t-1},$$

where the ε_t are i.i.d. $N(0,1)$ random variables. Following Cavaliere and Taylor (2009a) we consider for the AR parameter ϕ the values in the set $\{0, \pm 0.8\}$ and for the MA parameter θ the values in the set $\{0, \pm 0.5, \pm 0.8\}$.

6.2.1 Size properties

We now discuss the finite-sample size results obtained for the case where data were generated according to the process in (6.1) with ARMA innovations. The results for the case where the deterministic specification of the test is a constant mean are given in Table 6.2 and the results for the case of a linear time trend are given in Table 6.3.

Independent and identically distributed innovations

In the case where the innovations u_t are i.i.d., most of the tests seem to respect the nominal level $\alpha = 5\%$ quite well. For all asymptotic and bootstrap tests the convergence to the nominal level is clear as the sample size increases.

Notice how undersized the asymptotic M tests are in the case where the deterministic component in the test is a constant mean. This is mainly due to the fact that an asymptotic critical value is used in a finite-sample setting (this is also seen in the results of Ng and Perron, 2001). Indeed, the empirical sizes seem to approach the nominal level as the sample size is increased. The bootstrap versions of the M tests have levels much closer to the significance level for all sample sizes.

MA innovations

Some severe size distortions are seen for tests in the case where the MA parameter is -0.8 or -0.5, although the distortions are clearly less severe in the latter case. Also note that the size distortions are much more serious in the case where the deterministic specification of the test is a linear time trend. Whereas a significant part of the issue has disappeared for $T = 100$ in the case of a constant mean, it is still prominent for this sample size in the case of a linear trend.

Corresponding to results already available in the literature, most of the asymptotic tests are severely affected by the large negative MA parameter and show severe size distortions. The distortion is especially severe in the case of the Phillips–Perron tests, which incorrectly rejected the null hypothesis almost all of the time. The tests based on GLS detrending,

Table 6.2: Empirical size of the considered tests for $p = 0$

	$T = 50$							$T = 100$							$T = 250$						
ϕ	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0
θ	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5
DF_t^μ	3.3	4.6	2.8	3.1	9.5	3.2	5.3	3.9	4.5	3.7	3.4	9.3	3.6	5.1	4.3	4.7	4.3	4.1	9.1	4.1	5.3
DF_ρ^μ	5.4	8.6	8.2	13.4	42.0	9.5	17.1	5.1	6.3	6.3	11.3	32.0	7.3	12.1	4.7	5.2	5.3	8.5	20.6	6.0	8.5
DF_t^{PQ}	3.7	4.8	2.9	2.6	19.7	2.5	6.8	3.9	4.5	3.4	2.8	9.3	3.4	4.9	4.1	4.5	3.9	3.5	6.1	3.6	4.6
DF_ρ^{PQ}	4.0	7.3	6.1	6.7	38.3	4.4	14.8	4.0	5.3	4.7	6.7	25.2	5.2	10.0	4.1	4.6	4.5	6.8	15.8	4.9	7.3
Z_t	5.0	4.6	81.2	2.6	99.4	2.6	56.1	5.5	3.3	79.8	2.6	99.6	2.6	56.2	5.7	2.6	74.1	3.1	99.3	3.4	50.3
Z_ρ	4.2	0.2	84.5	0.4	99.7	0.4	62.2	5.8	0.3	82.2	1.0	99.7	1.1	60.4	6.0	0.7	76.7	2.5	99.5	2.8	53.8
ERS_t	6.8	6.9	5.5	3.8	16.0	4.9	10.2	5.7	5.5	5.1	3.6	10.2	4.9	7.6	5.3	5.1	5.1	4.5	8.5	4.7	6.3
P_T	3.3	10.7	0.7	6.4	9.3	4.9	5.5	3.6	6.2	1.0	5.4	3.4	4.5	4.5	3.9	4.6	2.6	5.1	2.0	4.3	4.5
ERS_t^{PQ}	7.6	7.6	6.3	4.2	27.6	4.9	12.7	5.8	5.8	5.3	3.7	13.3	5.0	8.0	5.3	5.0	5.1	4.4	8.2	4.6	6.2
P_T^{PQ}	3.5	10.9	0.7	5.9	18.0	4.5	6.9	3.6	6.3	1.1	4.9	5.8	4.5	4.8	3.9	4.5	2.6	4.9	2.1	4.1	4.5
CP_t	4.8	5.1	4.3	3.6	11.7	3.9	5.9	4.9	5.1	5.0	4.3	7.4	4.5	5.9	5.3	5.3	4.9	4.6	8.0	4.7	5.8
CP_ρ	4.1	4.3	4.0	3.6	17.7	3.7	7.4	4.5	4.5	4.2	3.6	10.6	4.1	6.6	4.8	4.8	4.7	4.4	9.8	4.4	6.1
$\overline{CP}_t$	4.8	4.9	4.2	3.3	11.5	3.9	6.2	5.1	5.0	4.7	4.0	7.5	4.5	6.2	5.0	5.3	4.9	4.6	8.5	4.8	5.9
$\overline{CP}_\rho$	4.4	4.3	3.9	3.3	14.7	3.9	7.1	4.8	4.4	4.4	3.8	8.6	4.4	6.4	4.9	5.2	4.8	4.5	8.6	4.7	5.8
PS_t	4.4	5.3	4.4	4.2	8.5	4.5	3.7	4.7	5.1	5.3	4.8	7.9	5.1	5.4	5.1	5.1	5.3	5.0	8.8	5.2	6.0
PS_ρ	3.6	3.7	2.9	4.1	13.4	4.2	4.5	4.1	3.9	4.1	3.7	9.5	4.3	5.4	4.6	4.5	5.0	4.2	9.5	4.5	5.8
$\overline{PS}_t$	3.7	4.1	3.3	2.9	7.6	3.7	3.2	4.7	4.4	4.6	3.7	7.6	4.4	5.6	5.0	5.1	4.8	4.9	8.8	4.9	5.8
$\overline{PS}_\rho$	3.6	3.2	2.8	3.3	10.4	3.9	3.9	4.4	3.7	4.3	3.5	8.2	4.0	5.4	4.8	4.7	4.5	4.3	8.6	4.5	5.5
PaP_t	5.3	6.2	3.0	3.4	25.7	4.3	7.6	5.3	6.2	4.9	3.8	17.0	5.4	5.1	5.4	5.8	5.2	4.3	12.1	4.7	4.7
PaP_ρ	5.2	8.3	2.7	7.4	41.7	6.1	12.0	5.0	3.1	2.5	7.5	31.2	5.6	6.9	5.1	3.1	2.6	6.6	18.4	3.5	4.1
$\overline{PaP}_t$	5.2	11.1	1.9	3.9	25.2	5.1	6.7	5.3	6.1	4.5	4.5	13.3	6.3	4.4	5.4	5.9	4.9	5.3	8.6	4.8	4.8
$\overline{PaP}_\rho$	5.0	9.4	2.0	6.8	31.1	6.1	9.3	5.1	4.0	3.1	6.5	17.5	6.1	5.4	5.2	4.5	3.6	6.1	10.0	4.1	4.4
PaP_t^{diff}	5.0	5.8	4.5	4.9	19.4	4.9	6.0	5.2	4.9	5.4	4.5	9.6	5.0	5.5	5.1	4.9	5.1	4.7	6.5	4.7	5.4
PaP_ρ^{diff}	3.8	4.5	3.9	6.9	32.1	5.5	7.6	4.3	3.7	4.5	5.8	19.3	4.6	6.4	4.5	4.0	4.6	6.2	11.3	4.6	5.5
$\overline{PaP}_t^{\text{diff}}$	3.9	4.1	3.4	3.8	20.6	4.4	5.8	4.5	3.9	4.3	3.7	9.8	4.5	5.3	4.8	4.2	4.4	4.5	6.5	4.8	5.1
$\overline{PaP}_\rho^{\text{diff}}$	3.7	5.3	3.2	5.7	25.7	5.4	6.8	4.2	4.1	4.1	5.1	12.7	4.7	5.5	4.5	4.1	4.3	5.2	7.6	4.8	5.2
RWB_t	4.7	5.1	4.2	4.3	12.0	4.3	5.4	5.1	5.2	5.0	4.9	6.7	4.6	5.7	5.0	5.2	5.1	4.9	8.0	4.9	5.6
RWB_ρ	4.4	4.2	3.9	3.9	19.2	4.2	7.5	4.7	4.7	4.4	3.9	9.9	4.1	6.4	4.6	4.6	4.9	4.6	9.7	4.5	5.9
DWB_t	4.6	4.3	3.6	2.1	21.2	2.6	7.7	5.0	4.8	4.5	2.8	10.3	4.0	5.8	5.1	4.8	4.6	3.6	7.0	4.3	5.4
DWB_ρ	4.9	9.4	7.0	8.1	39.9	5.3	16.2	5.4	6.7	6.0	8.3	27.7	6.3	11.6	5.1	5.4	5.4	7.8	17.5	5.8	8.3
AWB_t	5.0	4.3	3.9	2.2	21.0	3.0	8.0	5.4	4.9	4.7	2.8	10.4	4.1	6.2	5.2	5.1	4.9	3.8	6.9	4.5	5.6
AWB_ρ	5.4	9.1	7.5	8.2	39.4	5.9	16.3	5.5	6.7	6.3	8.5	27.9	6.7	11.6	5.2	5.7	5.6	7.9	17.4	6.0	8.4
LPB_t^0	3.0	4.3	2.6	2.8	3.4	2.8	2.6	4.4	4.5	4.3	4.1	5.6	4.1	4.3	5.1	5.0	5.1	5.5	6.3	5.2	5.0
$\overline{LPB}_t^0$	1.7	2.8	1.7	1.7	2.0	1.8	1.5	1.9	1.9	1.8	1.7	2.6	1.9	1.8	2.0	2.2	2.1	2.2	2.7	2.1	2.0
LPB_t	4.7	4.4	3.7	2.6	22.3	2.9	8.7	4.9	4.7	4.5	3.2	11.1	4.1	6.3	4.6	4.8	4.4	3.8	7.2	4.0	5.4
$\overline{LPB}_t$	2.6	2.4	1.8	0.8	19.2	1.0	6.0	2.5	2.1	1.9	1.0	7.5	1.7	3.2	2.3	2.4	2.1	1.5	3.8	1.8	3.0

however, are affected to a lesser extent. Also, notice how using MAIC for lag length selection reduces the severity of the size distortions for the augmented Dickey–Fuller tests. On the other hand, despite the suggestion of Perron and Qu (2007) to base lag selection on OLS detrended data instead of GLS detrended data, the distortion seems to be aggravated when following their suggestion.

The difference-based sieve bootstrap tests of Chang and Park (2003) and its residual-based modification by Palm et al. (2008) seem to lessen the size distortion seen with the asymptotic tests. As we shall see later, this comes without a sacrifice in power. In contrast, the difference-based and residual-based tests proposed by Paparoditis and Politis (2003) do not alleviate this size distortion problem and in many of the cases considered even exacerbated the problem quite significantly. Although Paparoditis and Politis (2003) argue against enforcing the null hypothesis when obtaining the residuals used for resampling, in this case the difference-based tests fare better in preserving the nominal level of the tests.

Out of all the asymptotic tests, the M tests of Ng and Perron (2001) seem to be least affected by a large negative MA parameter. This is not surprising, as these tests were de-

Table 6.2 (continued): Empirical size of the considered tests for $p = 0$

	$T = 50$							$T = 100$							$T = 250$						
ϕ	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0
θ	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5
MZ_α	4.7	15.2	0.8	8.7	9.9	7.0	7.0	4.8	8.5	1.3	6.9	3.8	6.0	5.7	4.9	5.8	3.2	6.2	2.3	5.3	5.5
MZ_t	4.9	14.6	0.9	8.6	10.2	7.0	7.2	5.0	8.3	1.5	7.0	4.0	6.1	6.0	5.1	5.8	3.4	6.2	2.5	5.4	5.8
MSB	4.5	15.3	0.7	8.5	9.3	6.8	6.3	4.4	8.5	1.0	6.6	3.3	5.6	5.0	4.4	5.6	2.6	5.8	1.8	4.7	5.0
MP_T	4.3	13.3	0.8	7.8	9.8	6.1	6.5	4.4	7.3	1.3	6.3	3.7	5.4	5.3	4.5	5.2	3.0	5.7	2.2	4.8	5.1
$PQMZ_\alpha$	5.0	15.4	0.8	7.9	20.1	6.4	9.0	4.9	8.7	1.4	6.3	6.5	5.9	6.1	4.9	5.7	3.2	5.9	2.4	5.1	5.5
$PQMZ_t$	5.2	14.9	0.9	7.9	20.4	6.4	9.2	5.0	8.4	1.6	6.4	6.8	6.0	6.3	5.0	5.7	3.4	6.0	2.6	5.2	5.7
$PQMSB$	4.7	15.4	0.7	7.7	19.0	6.1	8.1	4.4	8.5	1.0	6.0	5.9	5.6	5.4	4.4	5.5	2.6	5.6	2.0	4.6	4.9
$PQMP_T$	4.5	13.6	0.8	7.1	19.8	5.6	8.4	4.4	7.5	1.3	5.7	6.4	5.3	5.7	4.5	5.1	3.0	5.4	2.3	4.6	5.1
MZ_α^{iwb}	4.2	15.0	0.8	9.2	10.3	7.1	7.2	4.1	8.2	1.1	7.1	3.9	5.9	5.5	4.6	5.7	3.1	6.4	2.3	5.2	5.7
MZ_t^{iwb}	4.2	14.3	0.9	8.8	10.4	6.9	7.3	4.1	7.8	1.3	7.0	4.0	5.8	5.6	4.6	5.5	3.3	6.3	2.4	5.3	5.7
MSB^{iwb}	4.2	15.7	0.8	9.1	9.7	7.3	6.8	3.9	8.5	1.0	7.2	3.6	5.9	5.3	4.5	5.8	2.8	6.5	2.1	5.1	5.6
MP_T^{iwb}	4.2	13.9	0.9	8.7	10.5	6.9	7.3	4.0	7.6	1.3	6.9	4.1	5.6	5.6	4.6	5.4	3.3	6.3	2.4	5.2	5.7
MZ_α^{rwb}	3.4	2.2	1.7	2.9	5.1	3.4	4.6	3.8	3.3	2.5	2.9	4.7	3.4	4.6	4.2	4.2	4.1	4.2	5.3	4.1	5.1
MZ_t^{rwb}	3.5	2.2	1.7	2.8	5.2	3.4	4.6	3.8	3.3	2.6	3.0	4.7	3.4	4.6	4.3	4.2	4.2	4.2	5.5	4.1	5.1
MSB^{rwb}	3.5	2.3	1.6	2.8	5.0	3.3	4.5	3.8	3.2	2.4	2.9	4.6	3.4	4.5	4.1	4.1	4.1	4.2	5.3	4.1	5.0
MP_T^{rwb}	3.5	2.2	1.7	2.8	5.2	3.5	4.6	3.8	3.3	2.6	3.1	4.7	3.5	4.6	4.3	4.2	4.2	4.2	5.5	4.2	5.1
MZ_α^{dwb}	4.2	15.2	0.8	9.2	10.3	6.9	7.0	4.2	7.9	1.2	7.1	3.9	5.9	5.5	4.8	5.5	2.9	6.3	2.4	5.2	5.5
MZ_t^{dwb}	4.2	14.6	0.9	8.9	10.5	6.7	7.1	4.2	7.6	1.3	7.0	4.0	5.8	5.6	4.8	5.5	3.0	6.3	2.5	5.2	5.6
MSB^{dwb}	4.3	15.7	0.8	9.3	9.8	7.1	6.6	4.1	8.3	1.0	7.2	3.6	5.8	5.2	4.7	5.7	2.7	6.4	2.1	5.0	5.4
MP_T^{dwb}	4.2	14.2	0.9	8.8	10.6	6.6	7.1	4.1	7.4	1.3	6.9	4.1	5.7	5.7	4.9	5.4	3.1	6.3	2.6	5.1	5.6
MZ_α^{awb}	4.4	15.0	0.8	9.4	10.3	7.5	7.0	4.2	8.3	1.3	7.5	3.8	6.2	5.3	4.5	5.6	3.0	6.4	2.2	5.3	5.3
MZ_t^{awb}	4.5	14.2	0.9	9.2	10.5	7.3	7.1	4.2	7.9	1.4	7.4	4.0	6.1	5.3	4.6	5.6	3.2	6.4	2.3	5.3	5.4
MSB^{awb}	4.5	15.5	0.8	9.4	9.7	7.5	6.5	4.1	8.5	1.0	7.6	3.5	6.1	5.1	4.5	5.8	2.8	6.5	1.9	5.1	5.1
MP_T^{awb}	4.5	13.8	0.9	9.1	10.6	7.2	7.1	4.1	7.6	1.5	7.4	4.0	6.0	5.4	4.6	5.5	3.2	6.3	2.4	5.3	5.4
$PQMZ_\alpha^{\text{iwb}}$	4.4	15.3	0.8	8.2	20.6	6.4	9.2	4.3	8.1	1.2	6.3	6.6	5.9	5.9	4.5	5.6	2.9	6.0	2.5	5.1	5.7
$PQMZ_t^{\text{iwb}}$	4.4	14.6	0.9	8.0	20.8	6.2	9.2	4.3	7.8	1.4	6.2	6.8	5.7	6.0	4.6	5.5	3.2	6.0	2.7	5.1	5.7
$PQMSB^{\text{iwb}}$	4.4	15.8	0.8	8.2	19.7	6.5	8.7	4.1	8.5	1.0	6.4	6.2	5.9	5.7	4.4	5.8	2.8	6.1	2.2	5.1	5.5
$PQMP_T^{\text{iwb}}$	4.4	14.2	0.9	7.9	20.8	6.1	9.2	4.2	7.6	1.4	6.2	6.8	5.6	6.0	4.6	5.4	3.2	6.0	2.7	5.1	5.8
$PQMZ_\alpha^{\text{rwb}}$	3.6	2.4	1.7	2.7	12.9	3.1	6.2	3.9	3.4	2.8	3.1	4.7	3.4	4.8	4.5	4.0	4.1	3.9	5.4	4.0	5.1
$PQMZ_t^{\text{rwb}}$	3.5	2.4	1.7	2.6	13.0	3.1	6.2	4.0	3.4	2.9	3.1	4.7	3.5	4.8	4.6	4.1	4.1	4.0	5.4	4.0	5.0
$PQMSB^{\text{rwb}}$	3.6	2.4	1.6	2.8	12.8	3.2	6.0	3.8	3.3	2.7	3.1	4.6	3.4	4.7	4.4	4.1	4.0	3.9	5.3	4.0	5.2
$PQMP_T^{\text{rwb}}$	3.6	2.4	1.8	2.7	13.1	3.1	6.2	3.9	3.4	2.9	3.2	4.8	3.5	4.7	4.6	4.1	4.1	4.0	5.4	4.0	5.0
$PQMZ_\alpha^{\text{dwb}}$	4.4	15.3	0.8	8.5	20.4	6.3	9.3	4.3	8.0	1.3	6.3	6.7	5.7	5.9	4.5	5.6	3.2	6.0	2.5	5.0	5.6
$PQMZ_t^{\text{dwb}}$	4.4	14.6	0.9	8.3	20.6	6.2	9.3	4.3	7.7	1.4	6.3	6.8	5.5	6.0	4.6	5.6	3.4	5.9	2.6	5.0	5.7
$PQMSB^{\text{dwb}}$	4.3	15.8	0.8	8.4	19.6	6.5	8.7	4.2	8.4	1.1	6.4	6.3	5.7	5.6	4.5	5.8	2.9	6.1	2.2	4.9	5.5
$PQMP_T^{\text{dwb}}$	4.5	14.2	0.9	8.1	20.7	6.1	9.3	4.3	7.4	1.4	6.2	6.8	5.5	6.0	4.6	5.5	3.5	6.0	2.7	5.0	5.7
$PQMZ_\alpha^{\text{awb}}$	4.7	15.1	0.9	8.5	20.5	6.7	9.2	4.3	8.1	1.4	6.7	6.6	5.8	5.8	4.6	5.5	3.0	6.0	2.3	5.3	5.3
$PQMZ_t^{\text{awb}}$	4.8	14.5	1.0	8.3	20.7	6.7	9.2	4.4	7.8	1.5	6.6	6.8	5.6	5.9	4.7	5.3	3.2	6.1	2.5	5.2	5.4
$PQMSB^{\text{awb}}$	4.7	15.5	0.8	8.5	19.5	6.9	8.6	4.2	8.5	1.2	6.9	6.3	5.7	5.5	4.6	5.6	2.7	6.2	2.1	5.1	5.3
$PQMP_T^{\text{awb}}$	4.8	14.2	1.0	8.1	20.7	6.6	9.2	4.4	7.6	1.6	6.5	6.8	5.6	5.9	4.6	5.2	3.3	6.1	2.5	5.2	5.4

signed specifically for these situations. However, as the sample size increases, it seems that the size of the tests drops significantly below the nominal level (the sizes being approximately 2% and 1.3% for $T = 250$ in the constant mean and linear trend cases, respectively). Similar behaviour is seen for the (independent) wild bootstrap M tests of Cavaliere and Taylor (2009a). Notice that the sizes of the bootstrap tests are marginally, albeit consistently, greater than those of the asymptotic tests. In cases where the test is already oversized, this is surely undesirable. However, the *recoloured* versions of the bootstrap M tests of Cavaliere and Taylor (2009a) do not exhibit this strange behaviour and seem to respect the nominal level very well. In addition to being oversized to a less serious degree for small samples, the exact sizes of these tests are very close to the nominal level for $T = 100, 250$ both in the case of a constant mean and a linear trend.

Our dependent and autoregressive wild bootstrap implementations of the M tests exhibit very similar behaviour to that seen for the non-recoloured versions of the M tests. This

Table 6.3: Empirical size of the considered tests for $p = 1$

	$T = 50$							$T = 100$							$T = 250$						
ϕ	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0
θ	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5
DF_t	3.4	4.8	3.1	3.7	9.4	3.9	7.3	3.6	4.3	3.4	3.4	11.4	3.6	6.6	4.2	4.6	4.1	3.8	12.9	4.1	6.4
DF_ρ	8.4	15.6	14.9	26.9	56.6	18.9	30.3	6.5	9.9	9.5	22.8	52.2	13.5	20.7	5.3	6.4	6.4	14.3	35.1	8.4	12.6
DF_t^{PQ}	3.4	3.8	2.5	1.3	35.1	1.2	10.9	3.2	3.8	2.7	1.6	17.6	2.2	5.9	3.7	3.9	3.3	2.4	7.6	2.8	4.9
DF_ρ^{PQ}	4.6	10.7	9.9	7.4	62.4	3.8	25.9	3.8	6.4	5.7	7.3	44.0	6.0	15.5	3.9	4.7	4.4	7.7	24.9	4.7	9.6
Z_t	3.6	2.1	95.7	1.0	99.9	1.1	73.9	4.6	1.2	97.1	0.8	100.0	0.9	82.8	6.1	1.0	96.0	1.4	100.0	1.5	79.9
Z_ρ	2.5	0.0	96.9	0.1	99.8	0.1	77.2	4.6	0.0	97.7	0.1	100.0	0.2	85.3	6.6	0.1	96.5	0.7	100.0	1.0	81.7
ERS_t	4.2	4.0	2.7	0.7	25.8	0.8	10.3	2.7	2.8	2.1	0.8	9.5	1.7	4.6	2.5	2.4	2.2	1.3	4.4	1.8	3.3
P_T	1.2	16.2	0.2	5.4	20.6	2.7	6.7	2.1	7.5	0.1	5.8	6.9	4.1	4.0	3.2	4.9	0.7	5.9	1.5	3.9	3.9
ERS_t^{PQ}	4.7	4.5	3.3	0.8	38.9	0.8	13.6	3.1	3.0	2.5	0.9	17.4	1.8	6.0	2.6	2.5	2.3	1.3	5.6	1.8	3.5
P_T^{PQ}	1.3	16.8	0.2	4.5	30.7	2.3	8.4	2.3	7.9	0.1	5.0	13.1	4.0	4.9	3.2	5.0	0.7	5.4	2.1	3.7	4.1
CP_t	4.7	5.1	3.3	2.3	25.8	2.3	7.5	4.7	4.7	4.2	2.9	7.8	3.8	5.3	5.2	5.2	4.9	4.2	8.4	4.3	6.0
CP_ρ	4.3	4.6	3.7	3.5	40.6	3.1	11.7	4.1	4.2	3.8	2.6	15.3	3.7	6.8	4.6	4.4	4.3	3.6	10.4	3.7	6.2
$\overline{CP}_t$	4.3	4.9	3.2	2.1	25.4	2.1	7.7	4.8	5.0	4.1	2.8	7.1	4.0	5.5	4.9	4.9	4.9	4.2	7.7	4.0	6.3
$\overline{CP}_\rho$	4.1	4.6	3.3	2.8	35.5	3.1	10.7	4.4	4.6	3.9	2.7	11.4	3.9	6.4	4.5	4.6	4.6	4.0	8.4	4.0	6.3
PS_t	4.0	6.2	3.3	3.7	10.9	3.4	3.4	4.6	4.7	4.7	3.5	4.8	5.5	4.8	5.1	5.4	5.4	5.3	10.3	4.7	6.9
PS_ρ	3.6	4.0	2.4	5.3	21.8	5.1	6.0	4.1	3.3	3.6	3.0	8.9	5.0	4.8	4.7	4.4	4.6	3.8	10.0	3.7	6.2
$\overline{PS}_t$	3.3	5.6	2.6	3.1	10.0	3.0	3.0	4.7	4.1	4.1	3.0	4.3	5.0	4.9	5.0	5.1	5.3	4.8	9.2	4.5	6.9
$\overline{PS}_\rho$	3.3	4.0	2.2	4.2	18.6	4.6	5.0	4.3	3.4	3.6	3.0	6.8	4.7	4.8	4.6	4.5	4.8	3.9	8.9	4.0	6.4
PaP_t	4.2	7.1	1.3	1.3	38.2	2.0	9.8	4.7	7.1	3.6	2.2	23.5	5.1	4.4	5.1	5.9	5.1	3.7	14.7	4.7	4.8
PaP_ρ	4.8	14.7	3.3	6.5	64.7	4.6	21.4	4.6	3.4	1.8	8.0	50.9	6.9	10.1	5.0	2.1	1.9	7.7	29.5	3.0	4.6
$\overline{PaP}_t$	5.4	19.7	1.0	2.1	42.4	2.8	11.3	5.4	9.8	3.2	3.8	23.9	7.9	4.9	5.6	6.4	5.3	5.4	11.5	4.9	4.6
$\overline{PaP}_\rho$	5.3	17.5	2.2	7.3	56.7	5.3	18.4	5.2	4.4	2.0	8.1	37.0	7.6	8.9	5.2	3.2	2.7	7.9	18.1	3.6	4.3
PaP_t^{diff}	3.8	4.7	3.6	3.5	32.3	3.4	7.3	4.4	3.8	5.0	3.1	17.2	4.7	6.0	4.8	4.1	5.0	4.3	7.7	4.2	5.8
PaP_ρ^{diff}	3.2	5.1	3.6	7.6	49.8	5.2	10.0	3.6	3.2	4.3	5.2	31.6	5.4	7.0	4.2	3.3	4.4	6.6	16.0	4.1	6.2
$\overline{PaP}_t^{\text{diff}}$	3.5	5.6	2.8	3.1	32.5	3.5	7.3	4.3	3.7	4.1	2.7	16.1	4.6	5.5	4.9	4.0	4.3	4.0	6.9	4.2	5.5
$\overline{PaP}_\rho^{\text{diff}}$	3.2	7.4	2.8	6.7	43.6	5.7	9.4	3.9	4.0	3.8	4.7	22.9	5.2	6.3	4.4	3.8	4.2	5.9	10.4	4.4	5.8
RWB_t	4.4	5.5	3.5	2.6	26.9	2.3	7.7	4.6	5.0	4.2	3.3	7.6	4.3	5.3	5.0	5.0	5.0	4.3	8.3	4.2	6.3
RWB_ρ	4.4	3.9	3.7	3.7	44.1	3.4	12.5	4.2	4.1	3.8	3.0	16.1	4.1	6.8	4.6	4.3	4.5	3.5	9.8	3.6	6.3
DWB_t	4.7	4.0	3.4	0.8	37.3	1.0	12.8	5.2	4.5	4.1	1.4	19.6	3.0	7.5	5.0	5.0	4.7	2.8	8.9	3.6	6.2
DWB_ρ	5.1	14.9	10.4	8.4	62.6	4.5	27.4	5.4	8.9	7.7	9.5	46.6	7.9	18.0	5.2	6.5	5.8	9.8	27.6	6.1	11.4
AWB_t	5.7	3.8	3.8	0.9	37.4	1.3	13.1	5.7	4.4	4.5	1.5	19.6	3.2	7.7	5.7	4.9	5.0	2.9	9.1	4.0	6.6
AWB_ρ	6.1	14.0	11.3	9.1	62.4	5.3	27.5	6.0	8.5	8.0	9.7	46.7	8.1	18.1	5.7	6.0	6.1	9.8	27.9	6.6	11.8
LPB_t^0	2.6	3.8	2.4	2.3	3.1	2.7	2.2	5.6	6.0	5.5	5.3	7.7	5.6	5.5	8.6	8.6	8.5	9.3	11.1	8.5	8.6
$\overline{LPB}_t^0$	4.3	5.9	3.9	3.9	4.1	4.5	3.8	7.6	8.4	7.1	7.5	8.1	7.5	7.3	8.7	9.2	8.6	8.9	9.7	8.5	8.8
LPB_t	6.9	7.2	5.9	1.9	41.8	2.3	16.9	7.3	7.4	6.4	3.3	24.0	5.6	11.3	7.4	7.5	7.1	5.5	14.3	6.4	9.9
$\overline{LPB}_t$	6.0	6.8	5.2	1.6	42.6	1.6	17.3	5.9	6.3	4.9	2.9	24.0	4.7	10.6	5.8	5.8	5.3	4.6	12.3	4.8	8.1

suggests that there is more to be gained by recolouring the bootstrap sample by means of a sieve than by reintroducing autocorrelation by means of these alternative bootstrap methods.

In contrast to the case where the MA parameter is close to -1, in the case where the MA parameter of the generated innovations is close to $+1$, the asymptotic tests are undersized or close to the nominal level, except for the Dickey–Fuller coefficient test DF_ρ. Also note that the M tests, designed specifically to address the issue of a large negative MA parameter, are somewhat oversized in this case. The excess size, however, is small when compared to the excess seen in some tests in the case of a large negative MA parameter.

AR innovations

Although not as severe as the size distortions seen with a large negative MA parameter, when the innovations follow an AR process with a parameter close to 1, some of the asymptotic tests seem to be somewhat oversized. Of all the asymptotic tests, this size distortion is most prominent in the case of the M tests. On the other hand, when the AR parameter

Table 6.3 (continued): Empirical size of the considered tests for $p = 1$

	$T = 50$							$T = 100$							$T = 250$						
ϕ	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0
θ	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5
MZ_α	0.9	17.6	0.2	5.3	19.4	2.7	5.8	1.8	8.1	0.1	5.7	6.3	3.8	3.5	2.9	4.7	0.5	5.5	1.2	3.5	3.4
MZ_t	0.9	16.6	0.2	5.1	19.5	2.4	5.8	1.8	7.6	0.1	5.5	6.4	3.8	3.6	2.9	4.5	0.5	5.4	1.3	3.5	3.4
MSB	0.9	18.9	0.2	5.8	19.2	2.9	5.8	1.9	8.9	0.1	6.0	6.2	4.1	3.5	2.9	4.9	0.5	5.7	1.2	3.6	3.4
MP_T	0.9	16.0	0.2	5.0	19.7	2.3	6.0	1.9	7.3	0.1	5.4	6.5	3.8	3.6	2.9	4.5	0.6	5.5	1.3	3.6	3.5
$PQMZ_\alpha$	0.9	18.2	0.1	4.4	29.9	2.2	7.7	2.0	8.5	0.1	4.7	12.6	3.7	4.5	2.9	4.8	0.5	5.0	1.8	3.4	3.6
$PQMZ_t$	0.9	17.2	0.1	4.2	30.0	2.1	7.7	2.0	8.0	0.1	4.6	12.6	3.6	4.5	2.9	4.6	0.6	4.9	1.9	3.4	3.6
$PQMSB$	1.0	19.4	0.1	4.7	29.8	2.5	7.8	2.0	9.2	0.1	5.0	12.5	3.9	4.5	2.9	5.0	0.5	5.2	1.8	3.5	3.6
$PQMP_T$	1.0	16.6	0.1	4.1	30.1	2.0	7.8	2.1	7.7	0.1	4.6	12.7	3.6	4.6	2.9	4.6	0.6	5.0	1.9	3.4	3.7
MZ_α^{iwb}	4.2	29.7	0.5	12.1	26.2	6.9	13.2	3.6	12.9	0.3	10.3	8.5	7.5	6.7	3.8	6.6	1.0	8.4	2.0	5.7	5.5
MZ_t^{iwb}	4.2	28.5	0.5	11.8	26.4	6.6	13.3	3.7	12.2	0.3	10.0	8.6	7.5	6.7	3.9	6.5	1.1	8.4	2.0	5.6	5.5
MSB^{iwb}	4.1	30.5	0.4	12.3	25.8	7.0	12.7	3.5	13.5	0.2	10.4	8.2	7.6	6.6	3.8	6.8	1.0	8.5	1.9	5.7	5.4
MP_T^{iwb}	4.2	28.3	0.5	11.7	26.4	6.6	13.4	3.7	12.0	0.3	9.9	8.6	7.4	6.7	3.9	6.5	1.1	8.3	2.1	5.6	5.6
MZ_α^{rwb}	3.6	3.1	0.6	2.9	16.8	3.2	6.5	3.0	2.6	1.1	1.8	3.2	2.5	3.6	3.6	3.1	2.5	2.7	3.4	3.0	4.5
MZ_t^{rwb}	3.6	3.1	0.6	2.9	16.8	3.2	6.5	3.0	2.5	1.1	1.8	3.2	2.6	3.6	3.6	3.2	2.6	2.7	3.4	3.1	4.4
MSB^{rwb}	3.3	3.1	0.6	2.9	16.8	3.2	6.4	2.9	2.7	1.0	1.9	3.2	2.5	3.6	3.6	3.1	2.4	2.7	3.3	3.0	4.5
MP_T^{rwb}	3.6	2.9	0.6	2.9	16.8	3.1	6.5	3.0	2.4	1.1	1.8	3.3	2.6	3.7	3.6	3.1	2.6	2.7	3.4	3.1	4.4
MZ_α^{dwb}	4.1	29.6	0.5	12.1	26.2	7.0	13.1	4.0	12.4	0.3	9.9	8.5	7.3	7.0	4.0	6.5	1.1	8.2	1.9	5.5	5.6
MZ_t^{dwb}	4.2	28.3	0.5	11.8	26.3	6.7	13.3	4.0	11.8	0.3	9.7	8.7	7.3	7.1	4.0	6.4	1.2	8.2	2.0	5.4	5.7
MSB^{dwb}	3.9	30.2	0.4	12.1	25.7	7.2	12.7	3.9	13.1	0.3	10.0	8.2	7.4	6.8	4.0	6.6	1.1	8.3	1.8	5.5	5.5
MP_T^{dwb}	4.2	28.1	0.5	11.8	26.3	6.7	13.3	4.0	11.6	0.3	9.6	8.7	7.2	7.1	4.0	6.3	1.2	8.1	2.1	5.4	5.7
MZ_α^{awb}	4.9	28.0	0.5	12.6	26.4	8.2	13.2	4.3	12.2	0.3	10.3	8.6	7.7	6.8	4.4	6.4	1.3	8.8	2.0	5.8	5.5
MZ_t^{awb}	5.0	26.9	0.5	12.3	26.5	8.0	13.3	4.4	11.7	0.3	10.1	8.7	7.7	6.9	4.5	6.3	1.3	8.7	2.0	5.8	5.6
MSB^{awb}	4.8	28.5	0.5	12.6	25.9	8.2	12.7	4.2	12.7	0.3	10.4	8.3	7.8	6.6	4.4	6.6	1.2	8.8	1.8	5.8	5.3
MP_T^{awb}	5.0	26.6	0.5	12.3	26.5	8.0	13.4	4.4	11.4	0.4	10.1	8.7	7.6	6.9	4.4	6.2	1.3	8.7	2.1	5.8	5.6
$PQMZ_\alpha^{\text{iwb}}$	4.4	30.2	0.5	10.4	37.9	6.0	15.8	3.9	13.2	0.2	8.4	15.4	7.2	7.7	4.1	6.8	1.1	7.6	2.6	5.3	5.7
$PQMZ_t^{\text{iwb}}$	4.4	29.0	0.5	10.1	38.0	5.7	15.9	4.0	12.6	0.3	8.3	15.6	7.1	7.7	4.1	6.7	1.1	7.5	2.7	5.2	5.8
$PQMSB^{\text{iwb}}$	4.2	30.8	0.4	10.5	37.4	6.1	15.3	3.9	13.9	0.2	8.4	15.2	7.3	7.5	4.0	7.0	1.0	7.7	2.5	5.2	5.6
$PQMP_T^{\text{iwb}}$	4.5	28.8	0.5	10.0	38.1	5.6	15.9	4.0	12.3	0.3	8.2	15.6	7.0	7.8	4.1	6.6	1.1	7.5	2.8	5.2	5.8
$PQMZ_\alpha^{\text{rwb}}$	3.5	3.1	0.6	2.4	27.0	2.7	8.5	3.2	2.8	1.1	2.2	5.1	2.9	3.8	3.7	3.2	2.6	2.4	4.6	3.1	4.4
$PQMZ_t^{\text{rwb}}$	3.6	3.1	0.6	2.5	26.9	2.7	8.5	3.2	2.9	1.1	2.2	5.1	2.9	3.8	3.7	3.3	2.6	2.5	4.6	3.1	4.5
$PQMSB^{\text{rwb}}$	3.4	3.1	0.6	2.5	27.0	2.8	8.5	3.1	2.9	1.1	2.2	5.1	2.9	3.7	3.7	3.2	2.5	2.4	4.5	3.0	4.5
$PQMP_T^{\text{rwb}}$	3.6	3.0	0.6	2.5	26.7	2.7	8.5	3.3	2.8	1.1	2.2	5.2	3.0	3.8	3.8	3.3	2.6	2.5	4.6	3.1	4.5
$PQMZ_\alpha^{\text{dwb}}$	4.5	30.0	0.4	10.4	38.0	6.1	16.0	4.1	12.8	0.3	8.1	15.3	7.1	8.0	4.3	6.5	1.1	7.5	2.7	5.5	6.0
$PQMZ_t^{\text{dwb}}$	4.5	28.9	0.5	10.1	38.2	5.9	16.2	4.2	12.2	0.3	8.0	15.4	7.0	8.0	4.3	6.4	1.2	7.3	2.8	5.4	6.0
$PQMSB^{\text{dwb}}$	4.2	30.6	0.4	10.4	37.6	6.3	15.5	4.1	13.4	0.3	8.2	15.1	7.1	7.9	4.2	6.7	1.0	7.5	2.6	5.4	5.9
$PQMP_T^{\text{dwb}}$	4.4	28.7	0.5	10.1	38.2	5.8	16.2	4.2	11.9	0.3	8.0	15.4	6.9	8.0	4.3	6.2	1.2	7.3	2.8	5.4	6.0
$PQMZ_\alpha^{\text{awb}}$	5.3	29.0	0.5	11.2	38.0	7.3	16.4	4.4	12.3	0.3	8.8	15.5	7.5	8.0	4.4	6.5	1.3	7.9	2.8	5.8	5.7
$PQMZ_t^{\text{awb}}$	5.4	28.1	0.6	10.9	38.2	7.1	16.6	4.4	11.7	0.4	8.7	15.7	7.5	8.0	4.4	6.3	1.4	7.8	2.8	5.8	5.7
$PQMSB^{\text{awb}}$	5.1	29.2	0.5	11.3	37.5	7.5	15.8	4.2	12.8	0.3	8.9	15.3	7.5	7.8	4.4	6.5	1.2	7.8	2.7	5.7	5.6
$PQMP_T^{\text{awb}}$	5.4	27.8	0.6	10.9	38.2	7.1	16.6	4.4	11.4	0.4	8.7	15.8	7.5	8.0	4.4	6.3	1.4	7.8	2.9	5.8	5.7

of the DGP is close to -1, the tests seem to be *severely* undersized. Again, the M tests show the greatest size distortion in this case. However, the distortion disappears quickly as the sample size increases and it is clear that the problem is less severe than in the case where the DGP has MA innovations with a large negative parameter.

Looking at the sieve and block bootstrap tests we see that there is some size distortion which seems to occur predominantly when using the coefficient-based tests when the AR parameter is close to 1 or -1. However, the largest size distortions are seen when using the residual-based tests of Paparoditis and Politis (2003), i.e. the tests denoted by PaP. Again, although the problem is very prominent for the small sample size $T = 50$, the exact size comes closer to the nominal size very rapidly as the sample size increases.

As with the asymptotic tests, the bootstrap versions of the M tests suffer serious size distortions, being oversized in most cases and undersized when using the recoloured versions of the bootstrap M tests.

6.2.2 Power properties

In this section we report on the finite-sample power properties of the considered tests. Each value reported in Table 6.4 and Table 6.5 represents the *size-adjusted power* of the associated test, which was calculated as

$$\hat{\beta}_{\text{adj}} = \Phi\left(\Phi^{-1}(\alpha) + \hat{\delta}\right), \qquad \hat{\delta} = \Phi^{-1}(\hat{\beta}) - \Phi^{-1}(\hat{\alpha}),$$

where Φ denotes the standard normal distribution function, α is the nominal size of the test, $\hat{\alpha}$ is the empirical (exact) size of the corresponding test and configuration (given in Table 6.2 and Table 6.3), and $\hat{\beta}$ is the (unadjusted) empirical power obtained from the simulation. See Lloyd (2005) for a discussion on other simple methods of estimating test power adjusted for empirical size.

Table 6.4: Size-adjusted power of the considered tests for $p = 0$

	$T = 50$							$T = 100$							$T = 250$						
ϕ	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0
θ	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5
DF_t	12.0	6.3	9.2	4.6	22.6	5.4	12.3	13.5	9.2	10.9	7.0	12.1	9.8	12.5	15.9	12.8	14.5	11.1	12.0	12.9	15.5
DF_ρ	19.4	12.4	18.7	10.1	17.6	11.3	19.9	20.7	16.3	19.9	14.3	20.3	18.0	21.7	23.6	20.6	22.8	20.0	23.0	21.0	24.5
DF_t^{PQ}	14.4	7.2	13.3	5.7	20.9	6.3	16.0	15.2	10.2	13.4	8.1	18.6	11.2	15.6	16.7	13.4	15.5	11.8	16.5	13.6	17.0
DF_ρ^{PQ}	23.5	13.5	23.5	13.8	25.0	13.7	23.9	23.9	18.1	23.6	18.1	27.1	21.3	24.6	25.1	21.7	24.7	21.8	27.4	22.8	26.0
ERS_t	28.2	16.0	20.8	16.6	11.3	18.4	18.8	27.7	20.3	23.3	19.7	13.6	23.0	21.8	27.5	24.6	26.1	22.3	19.6	24.9	26.0
P_T	23.7	11.0	12.7	11.6	12.7	13.8	17.8	25.0	17.5	17.7	17.3	13.8	21.5	19.7	26.3	23.5	22.2	21.9	15.9	23.9	23.9
ERS_t^{PQ}	30.1	16.4	23.0	17.6	19.8	18.8	22.0	28.8	20.8	24.5	20.4	18.2	23.7	23.4	27.6	24.8	26.0	22.4	20.0	24.9	26.0
P_T^{PQ}	24.9	11.5	14.3	12.2	19.8	13.8	20.5	25.7	18.0	18.3	17.9	18.0	21.9	21.0	26.4	23.7	22.4	21.8	16.7	24.0	23.9
CP_t	14.4	10.0	12.1	8.3	22.8	7.7	12.7	15.4	11.7	13.1	9.5	14.5	12.1	13.9	16.2	14.0	15.6	13.4	15.7	14.4	16.7
CP_ρ	19.7	12.9	16.6	12.0	28.6	13.0	17.8	21.9	16.4	19.6	14.5	21.3	18.0	18.5	23.7	20.0	23.0	19.2	21.7	21.6	23.0
$\overline{CP}_t$	22.1	15.1	16.6	14.7	22.0	14.4	14.5	25.2	19.2	22.3	17.9	13.8	21.1	19.7	26.9	22.7	26.0	21.4	17.5	23.8	25.0
$\overline{CP}_\rho$	21.9	14.5	17.1	14.8	24.5	14.6	16.9	25.0	18.9	22.1	18.3	16.7	20.8	20.3	26.2	22.6	25.4	21.7	18.1	23.2	25.1
PS_t	12.9	10.4	14.1	8.7	17.3	8.0	12.3	15.8	11.8	13.7	9.6	16.2	13.0	16.8	16.7	14.9	16.4	13.9	17.5	14.2	17.1
PS_ρ	17.2	13.3	16.8	12.0	20.3	11.8	15.3	22.3	16.3	21.0	14.7	18.6	19.0	20.2	25.5	21.1	24.0	20.2	24.2	21.2	23.5
$\overline{PS}_t$	19.5	16.2	19.5	14.8	16.2	14.4	13.8	24.6	19.4	22.7	18.2	15.7	22.1	21.2	26.9	23.1	26.7	22.7	18.5	25.2	25.1
$\overline{PS}_\rho$	18.9	14.3	18.9	14.6	18.0	14.0	15.1	24.1	18.8	22.4	16.8	17.0	21.6	21.3	26.5	22.6	26.6	22.2	19.5	24.6	25.0
PaP_t	16.2	9.8	15.0	7.7	19.3	8.1	18.6	17.3	11.5	15.0	9.5	18.9	12.7	17.9	16.6	14.5	17.5	12.6	21.3	13.9	18.6
PaP_ρ	25.1	15.4	25.5	14.6	24.9	14.5	29.2	26.5	18.6	25.3	17.4	29.7	21.8	27.9	25.3	21.3	26.3	20.5	34.0	21.1	26.8
$\overline{PaP}_t$	28.7	17.2	21.7	16.5	20.0	18.4	24.4	28.8	21.2	22.3	19.0	19.8	22.6	23.8	27.6	23.5	27.7	21.4	22.9	25.1	25.1
$\overline{PaP}_\rho$	28.7	15.6	23.7	15.5	21.6	17.3	27.1	28.6	19.6	23.5	19.0	21.0	22.3	26.2	27.5	22.2	26.3	21.9	23.8	24.2	25.7
PaP_t^{diff}	12.6	8.3	11.3	6.8	22.0	7.3	17.0	14.0	10.8	13.3	8.8	19.0	11.6	15.9	15.5	13.2	15.3	12.7	16.3	13.6	16.3
PaP_ρ^{diff}	17.2	15.0	18.4	13.3	28.3	13.5	25.1	19.3	18.2	21.0	16.6	30.0	20.0	22.4	22.0	20.7	23.0	21.4	27.4	21.5	24.1
$\overline{PaP}_t^{\text{diff}}$	20.5	17.9	19.5	16.2	22.0	16.6	20.8	21.9	20.5	23.6	18.8	20.2	21.1	20.9	25.3	22.8	25.0	22.3	19.9	23.7	23.4
$\overline{PaP}_\rho^{\text{diff}}$	20.4	14.8	19.8	15.0	23.8	15.0	23.4	21.3	19.4	22.5	18.4	22.9	20.9	22.1	25.2	22.7	24.6	22.6	21.4	23.3	23.1
RWB_t	14.2	10.4	12.5	8.5	23.2	7.1	14.2	14.9	11.5	13.2	9.3	15.1	12.5	14.3	16.7	14.3	15.8	13.0	16.5	13.6	17.3
RWB_ρ	20.1	13.5	17.2	13.1	28.8	12.2	19.0	22.1	16.3	19.8	15.0	21.8	18.5	19.1	24.4	21.0	23.1	19.0	21.3	20.8	23.7
DWB_t	15.8	9.9	15.0	8.3	21.2	8.3	17.0	15.9	12.0	14.3	10.2	18.6	12.0	16.4	17.3	14.1	16.7	13.4	16.9	14.0	17.9
DWB_ρ	23.6	12.7	23.7	13.0	24.8	13.9	24.2	24.6	18.4	25.0	17.6	27.2	21.8	25.2	25.9	21.8	25.5	23.0	27.5	23.2	27.4
AWB_t	15.2	9.5	14.1	8.2	21.1	8.4	16.4	15.5	11.9	14.4	10.3	18.8	12.5	15.5	16.8	14.3	16.4	12.9	17.5	14.4	17.4
AWB_ρ	24.2	12.9	23.3	14.4	24.9	14.5	24.3	25.1	18.4	24.5	18.8	27.3	21.7	25.0	25.6	21.6	25.4	22.2	28.0	23.6	27.2
LPB_t^0	9.4	6.9	10.0	8.7	12.8	9.6	10.5	12.2	11.0	12.3	12.3	15.9	13.2	13.0	15.7	14.0	15.6	14.5	17.6	15.0	16.1
$\overline{LPB}_t^0$	10.2	7.0	9.3	9.5	8.7	9.7	10.4	14.3	12.4	13.9	13.9	12.9	14.0	14.4	19.9	17.0	18.4	18.8	17.5	19.0	19.5
LPB_t	16.4	10.6	15.5	8.8	20.9	9.0	16.9	15.9	12.9	14.6	10.6	19.0	13.1	16.5	16.6	14.5	16.2	13.3	18.5	14.4	17.7
$\overline{LPB}_t$	24.7	13.9	20.8	15.0	21.1	15.8	20.3	24.3	19.4	22.1	18.0	18.6	21.2	21.4	25.3	21.5	23.5	20.8	18.2	23.2	22.0

Independent and identically distributed innovations

The first striking observation from the power results is the low power of the unit root tests, especially in cases where the innovations are not i.i.d. This might not be surprising, as unit

Table 6.4 (continued): Size-adjusted power of the considered tests for $p = 0$

	$T = 50$							$T = 100$							$T = 250$						
ϕ	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0
θ	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5
MZ_α	24.7	10.5	12.3	11.3	12.5	13.5	17.4	26.2	17.5	17.8	17.5	13.5	22.0	19.9	27.2	23.9	23.0	22.7	15.9	24.7	24.7
MZ_t	25.1	11.1	12.8	11.9	12.4	14.3	17.7	26.7	18.4	18.1	18.0	13.5	22.7	20.3	27.5	24.5	23.5	23.2	16.0	25.0	24.8
MSB	21.7	9.3	10.4	9.6	12.6	11.7	16.5	24.3	15.7	16.5	15.8	13.6	20.2	18.7	25.6	22.0	21.7	21.0	15.3	23.4	23.1
MP_T	24.8	11.1	12.4	11.7	12.5	14.2	17.7	26.2	18.3	17.9	17.9	13.7	22.5	20.2	27.1	24.3	23.0	22.8	15.9	24.6	24.6
$PQMZ_\alpha$	26.7	11.2	14.3	12.1	19.8	13.7	20.4	27.3	18.0	18.9	18.2	18.0	22.6	21.5	27.3	24.1	23.0	22.5	16.6	24.5	24.6
$PQMZ_t$	27.2	11.9	14.9	12.8	19.8	14.5	20.8	27.7	18.9	19.1	18.7	17.9	23.3	21.9	27.6	24.7	23.5	23.0	16.8	24.9	24.8
$PQMSB$	23.7	10.0	12.7	10.3	19.8	11.9	19.6	25.4	16.3	17.6	16.4	18.1	20.7	20.3	25.8	22.2	21.8	20.7	16.1	23.3	23.1
$PQMP_T$	26.8	11.8	14.6	12.6	19.8	14.3	20.8	27.2	18.8	18.9	18.5	18.0	23.0	21.7	27.2	24.5	23.0	22.5	16.8	24.6	24.6
MZ_α^{iwb}	24.6	10.7	12.8	11.6	12.2	13.2	17.8	26.0	17.8	18.3	17.7	13.4	22.4	20.3	27.3	23.6	22.9	22.5	16.3	23.9	24.2
MZ_t^{iwb}	25.0	11.2	13.2	12.4	12.2	14.1	18.1	26.5	18.5	18.6	18.2	13.4	23.0	20.8	27.4	24.1	23.5	23.4	16.5	24.3	24.5
MSB^{iwb}	22.0	9.6	10.9	10.5	12.3	11.6	16.8	24.9	16.0	17.1	16.1	13.3	20.6	18.8	26.3	21.6	22.5	20.8	15.7	23.4	22.9
MP_T^{iwb}	25.1	11.4	13.5	12.4	12.1	14.3	18.1	26.8	18.7	18.9	18.4	13.4	23.4	21.0	27.4	24.3	23.4	23.4	16.7	24.6	24.5
MZ_α^{rwb}	19.5	12.7	9.9	10.7	12.3	14.1	12.3	23.8	15.5	17.5	16.6	10.9	19.5	18.5	26.9	22.1	23.2	20.1	15.8	23.5	23.7
MZ_t^{rwb}	20.2	13.3	10.2	11.2	12.3	14.8	12.6	24.1	16.2	18.0	16.9	11.0	20.1	18.9	27.3	22.5	23.2	20.5	15.8	23.8	24.1
MSB^{rwb}	17.4	11.6	9.2	9.9	12.4	12.6	12.1	22.5	14.6	16.5	15.6	10.7	18.3	17.5	26.1	21.2	21.8	19.4	15.2	22.2	22.9
MP_T^{rwb}	20.4	14.6	10.2	11.7	12.0	14.7	12.9	24.4	17.0	18.1	17.1	11.0	20.1	19.0	27.3	22.5	23.3	20.8	15.8	23.8	24.3
MZ_α^{dwb}	24.1	10.5	12.7	11.3	12.2	13.9	17.6	26.2	17.6	18.7	17.7	13.4	22.6	20.8	26.8	24.0	24.4	23.1	15.6	24.7	25.2
MZ_t^{dwb}	24.8	10.9	13.1	12.1	12.1	14.6	17.9	26.9	18.1	18.9	18.4	13.3	23.2	20.9	27.2	24.5	24.8	23.5	15.6	24.8	25.2
MSB^{dwb}	21.5	9.5	10.8	10.2	12.3	12.0	16.7	25.0	16.1	17.4	16.2	13.3	21.1	19.7	25.7	22.1	23.5	21.4	15.1	23.5	24.0
MP_T^{dwb}	24.9	11.1	13.4	12.2	12.1	14.9	17.9	27.2	18.3	18.9	18.4	13.3	23.6	21.0	27.0	24.5	24.7	23.5	15.8	25.1	25.4
MZ_α^{awb}	24.8	10.4	13.1	11.4	12.3	15.0	18.1	26.4	16.3	18.0	17.3	13.6	21.6	21.0	27.6	24.0	22.9	23.2	16.6	24.6	25.2
MZ_t^{awb}	25.0	11.1	13.6	12.1	12.2	16.0	18.4	27.3	17.2	17.9	17.5	13.5	22.2	21.7	27.7	24.0	23.4	23.7	16.5	25.3	25.1
MSB^{awb}	21.9	9.4	11.0	10.4	12.4	13.3	17.3	25.1	15.1	17.1	15.5	13.5	20.0	19.5	26.3	22.0	22.3	21.5	16.2	23.8	24.2
MP_T^{awb}	25.3	11.2	14.0	12.3	12.2	16.2	18.4	27.7	17.5	18.1	17.6	13.5	22.5	21.7	27.7	24.3	23.4	23.7	16.4	25.6	25.1
$PQMZ_\alpha^{\text{iwb}}$	25.5	11.3	14.4	12.4	19.4	13.2	20.6	25.7	18.0	18.3	18.6	17.9	21.7	21.8	26.8	24.1	23.3	22.6	16.8	24.3	23.9
$PQMZ_t^{\text{iwb}}$	26.3	11.8	14.9	13.0	19.4	14.3	21.0	26.4	18.5	18.7	19.1	17.8	22.4	21.9	27.1	24.2	23.6	23.2	16.7	24.8	23.9
$PQMSB^{\text{iwb}}$	23.1	10.2	12.9	10.9	19.4	11.7	19.4	24.7	16.2	17.9	17.1	18.0	20.4	20.5	25.3	22.5	22.3	21.4	16.4	23.1	22.9
$PQMP_T^{\text{iwb}}$	26.4	12.0	15.4	13.3	19.4	14.5	21.0	26.8	18.8	18.9	19.1	18.0	22.6	22.1	26.9	24.3	23.7	23.2	16.8	24.9	23.9
$PQMZ_\alpha^{\text{rwb}}$	23.0	13.5	12.0	11.9	20.9	13.9	17.0	23.4	16.6	16.4	16.0	13.8	19.9	18.2	25.4	22.9	23.6	20.3	16.8	23.4	23.4
$PQMZ_t^{\text{rwb}}$	23.5	14.2	12.4	12.9	21.0	14.5	17.2	23.9	17.1	16.7	16.5	13.8	20.2	18.5	25.6	23.5	23.8	20.5	16.6	23.7	24.0
$PQMSB^{\text{rwb}}$	21.2	12.3	11.2	10.7	20.7	12.3	16.3	22.4	15.7	15.0	15.3	13.7	19.1	17.6	24.7	21.6	22.3	19.5	16.2	22.0	21.6
$PQMP_T^{\text{rwb}}$	23.6	15.2	12.4	13.3	20.9	14.8	17.4	24.1	17.7	16.8	16.6	13.9	20.5	18.6	25.8	23.6	23.9	20.7	16.6	23.7	24.1
$PQMZ_\alpha^{\text{dwb}}$	26.5	11.3	15.0	11.8	19.6	12.9	20.3	27.1	17.7	19.4	18.3	17.8	23.0	22.6	27.4	23.4	22.2	23.0	16.9	24.5	24.5
$PQMZ_t^{\text{dwb}}$	27.1	11.8	15.4	12.5	19.5	13.6	20.7	28.0	18.4	19.8	18.7	17.7	23.7	22.7	27.5	24.2	22.6	23.6	16.7	24.7	24.6
$PQMSB^{\text{dwb}}$	23.7	10.2	13.2	10.5	19.5	11.5	19.3	26.3	16.3	18.8	16.6	17.7	21.3	21.8	25.8	21.8	21.5	21.4	16.8	23.9	23.0
$PQMP_T^{\text{dwb}}$	27.1	12.1	15.6	12.7	19.4	13.8	20.7	28.1	18.8	20.0	18.8	17.9	23.8	22.6	27.5	24.3	22.6	23.5	16.8	24.9	24.6
$PQMZ_\alpha^{\text{awb}}$	26.7	11.3	14.9	13.2	19.5	16.0	20.8	27.9	17.9	18.6	18.6	17.8	23.8	22.6	27.4	23.8	23.8	23.2	17.5	25.3	25.2
$PQMZ_t^{\text{awb}}$	27.1	11.9	15.2	14.0	19.5	16.7	21.3	28.5	18.6	19.2	19.0	17.8	24.7	22.7	27.6	24.3	24.0	23.4	17.4	25.6	25.4
$PQMSB^{\text{awb}}$	23.5	10.2	13.4	11.4	19.7	13.7	19.9	26.5	16.3	17.9	16.5	17.7	21.9	21.5	26.1	22.5	23.2	21.8	16.9	24.3	23.6
$PQMP_T^{\text{awb}}$	27.4	12.0	15.4	14.3	19.4	16.8	21.4	28.5	19.0	19.0	19.2	17.8	24.9	22.7	27.9	24.7	24.0	23.5	17.6	25.9	25.6

root tests are known to have "notoriously low power" (Cochrane, 1991). As expected, the power of the tests increases with the sample size and one notices that the power becomes comparable across the different choices of the process generating $\{u_t\}$.

It is interesting to note that the unit root tests exhibit higher power in cases where the deterministic specification in the test is a linear time trend when compared to the case of a constant mean. This corresponds to results already available in the literature (see e.g. Cavaliere and Taylor, 2009a; Smeekes, 2013).

In the case of i.i.d. innovations it seems like the tests that yield the highest power are the tests $\overline{PaP}_t$ and $\overline{PaP}_\rho$, the GLS versions of the residual-based block bootstrap tests of Paparoditis and Politis (2003), and our autoregressive wild bootstrap adaptations (denoted by PQM^{awb}) of the bootstrap M tests of Cavaliere and Taylor (2009a), using OLS detrending for lag selection as suggested by Perron and Qu (2007). Note that the recoloured versions of the bootstrap M tests have slightly lower power than the non-recoloured versions, especially

Table 6.5: Size-adjusted power of the considered tests for $p = 1$

	$T = 50$							$T = 100$							$T = 250$						
ϕ	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0
θ	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5
DF_t	15.0	3.4	9.5	2.0	45.5	2.4	14.1	17.5	7.3	12.2	4.4	18.3	7.8	10.9	22.1	14.6	19.5	11.3	10.4	14.6	16.5
DF_ρ	21.5	6.8	19.8	4.5	20.3	4.6	18.7	23.9	11.6	22.9	8.9	18.8	14.0	22.8	28.6	19.9	27.9	18.4	24.9	21.6	27.6
DF_t^{PQ}	26.1	4.5	24.0	5.5	29.8	9.0	23.8	26.2	11.4	23.7	12.1	26.7	12.9	21.9	26.5	18.0	24.9	16.1	22.0	20.0	22.7
DF_ρ^{PQ}	34.1	7.1	34.0	9.4	31.0	9.1	31.4	34.1	16.6	33.8	20.0	32.8	21.0	31.4	34.1	24.4	34.0	25.8	35.3	29.5	32.8
ERS_t	34.2	7.8	25.3	12.8	16.7	18.4	21.7	33.9	16.9	27.1	20.9	15.3	21.9	22.8	33.4	24.6	30.2	24.7	19.0	27.8	27.7
P_T	28.6	4.1	11.4	4.4	17.5	4.4	22.0	32.9	14.0	17.6	14.0	16.1	20.1	22.2	34.5	25.3	25.9	23.6	18.1	29.1	26.9
ERS_t^{PQ}	36.4	7.3	28.9	13.2	29.2	20.0	25.6	35.4	17.5	29.6	21.9	26.1	21.4	25.9	34.5	25.1	31.4	25.5	22.5	28.5	28.9
P_T^{PQ}	30.1	3.9	13.8	4.5	27.5	4.3	25.8	34.2	14.6	19.1	15.4	25.5	20.0	24.9	35.3	25.7	27.0	24.3	20.9	29.9	28.0
CP_t	24.9	7.7	18.8	9.1	37.0	10.9	24.2	24.6	12.9	22.2	14.5	30.3	15.1	16.3	26.0	18.4	25.3	17.2	19.0	20.0	22.6
CP_ρ	28.9	9.3	24.7	10.4	44.4	9.7	32.5	29.3	14.4	26.6	17.5	39.3	19.7	23.4	31.3	22.6	30.8	21.8	27.4	25.3	27.9
$\overline{CP}_t$	31.4	9.6	19.3	11.9	37.0	14.1	24.0	31.8	16.0	25.4	18.4	29.6	19.9	17.5	35.2	25.3	31.4	22.9	19.0	29.6	26.2
$\overline{CP}_\rho$	32.5	10.2	23.4	13.2	40.3	11.6	30.4	32.3	15.9	25.8	18.9	33.0	21.1	22.3	35.2	25.2	31.6	23.3	22.5	29.0	27.6
PS_t	11.1	10.0	11.8	5.8	16.2	6.5	14.0	17.9	12.7	18.8	13.0	18.8	13.2	11.9	25.4	18.2	25.4	16.4	20.7	20.5	22.8
PS_ρ	14.0	12.6	13.1	9.7	26.5	6.1	18.2	19.9	13.3	20.2	13.7	23.9	17.1	14.5	29.8	21.0	29.3	18.8	24.0	24.3	24.8
$\overline{PS}_t$	12.2	13.3	13.3	6.5	16.5	6.8	14.7	20.2	14.7	21.7	15.1	19.0	17.0	11.4	33.5	24.0	31.0	22.8	20.5	28.0	26.9
$\overline{PS}_\rho$	14.9	12.9	13.7	11.2	24.7	6.7	18.4	20.6	14.1	21.3	15.2	22.0	18.4	13.2	33.0	22.9	31.2	22.2	21.7	25.7	27.1
PaP_t	31.3	7.3	34.2	9.6	29.9	13.4	31.1	28.9	13.6	25.0	15.0	27.1	15.2	31.5	28.9	20.0	30.8	17.0	33.2	20.4	26.2
PaP_ρ	37.2	8.9	41.9	10.3	29.6	10.6	38.9	36.9	18.6	39.1	18.5	33.3	22.4	41.4	35.9	25.0	38.8	22.3	46.7	26.1	37.5
$\overline{PaP}_t$	37.2	6.6	34.7	11.2	26.8	17.2	30.7	36.6	18.9	26.7	18.3	24.7	19.6	31.3	37.4	27.9	36.1	20.7	28.1	29.4	30.0
$\overline{PaP}_\rho$	38.8	7.9	38.2	9.7	26.7	11.2	36.1	37.9	18.8	33.2	18.7	24.5	22.8	35.9	38.6	26.6	36.9	21.6	29.4	28.8	36.0
PaP_t^{diff}	16.6	9.2	19.3	6.4	32.1	7.4	28.3	17.6	12.1	19.4	12.7	28.6	12.9	20.3	20.8	17.7	22.9	15.9	23.2	18.6	20.6
PaP_ρ^{diff}	16.7	13.5	25.4	8.6	37.0	6.7	41.0	19.7	15.8	23.8	15.9	42.2	16.8	28.4	24.5	23.0	29.2	23.2	38.6	24.0	26.7
$\overline{PaP}_t^{\text{diff}}$	19.6	13.4	22.0	9.6	32.4	9.3	29.5	20.8	17.8	25.8	18.4	29.0	18.4	23.2	27.6	25.0	31.8	24.0	23.4	27.4	25.9
$\overline{PaP}_\rho^{\text{diff}}$	19.7	11.7	25.0	10.0	33.8	7.0	38.1	20.6	16.3	25.1	17.4	33.2	19.0	27.3	28.2	24.9	31.1	24.7	28.4	26.9	27.7
RWB_t	26.9	7.6	18.5	8.9	36.5	10.7	24.6	25.3	12.9	23.6	13.9	31.0	14.5	17.6	26.4	18.7	24.7	18.2	19.6	21.2	22.0
RWB_ρ	31.0	10.7	25.2	11.2	42.3	10.0	33.9	29.8	15.0	27.6	16.2	39.1	19.3	25.7	31.6	22.7	30.6	22.7	28.9	27.2	27.1
DWB_t	28.3	5.7	26.5	10.4	29.9	14.8	24.2	27.4	13.4	25.4	17.3	26.7	15.8	22.3	28.3	18.5	25.6	18.5	22.8	22.2	24.6
DWB_ρ	35.1	5.8	34.9	8.8	30.0	8.8	30.8	34.6	15.7	34.7	19.4	33.0	21.7	31.6	35.4	23.3	35.2	26.1	35.8	30.1	34.6
AWB_t	27.7	6.6	26.5	11.8	29.6	15.6	24.0	27.1	14.1	25.1	19.0	26.4	16.9	23.0	28.4	20.0	26.8	18.5	23.1	22.8	23.8
AWB_ρ	34.7	6.7	34.7	10.3	29.9	10.9	30.9	34.4	16.5	34.4	20.2	32.3	22.5	32.2	35.6	25.3	36.1	26.5	36.5	30.5	33.6
LPB_t^0	7.3	4.3	7.7	6.8	10.1	7.1	9.5	12.3	8.2	13.0	11.6	18.0	11.9	14.3	19.4	15.3	19.7	18.3	25.8	18.4	20.0
$\overline{LPB}_t^0$	8.0	5.0	7.1	7.2	7.2	7.4	7.9	13.1	9.2	12.5	12.1	12.4	13.6	13.6	22.1	17.2	20.8	21.8	20.5	22.1	22.3
LPB_t	30.5	6.2	27.8	11.8	28.7	14.0	23.6	29.6	15.1	27.6	17.7	25.6	18.2	23.7	29.9	21.5	28.5	20.7	23.5	23.6	26.3
$\overline{LPB}_t$	38.3	7.7	29.5	14.1	27.5	18.7	24.7	37.4	18.2	31.7	22.5	23.8	22.6	25.9	36.5	25.8	34.8	26.4	22.6	31.7	31.0

for smaller sample sizes.

The wild bootstrap tests RWB_ρ, DWB_ρ and AWB_ρ also yielded relatively high power with the recoloured wild bootstrap having the lowest power of the three, as was the case for the M tests. Note that the implementations based on the ADF t-statistic do not perform as well as the implementations based on the ADF normalised coefficient statistic.

Notice also the good performance of the ADF LPB test that makes use of GLS detrending rather than OLS detrending. Keep in mind that the LPB requires a choice of a kernel function and a bandwidth and although some guidance is given by McMurry and Politis (2010), more research needs to be done on this relatively new bootstrap procedure. Also, the original implementation of the LPB by Zou and Politis (2016) was proposed only for the case where the test contains no deterministic component. As described in Section 5.7 we have adapted the procedure to allow for specification of a deterministic component.

MA innovations

In the case of a large positive MA parameter, most bootstrap tests seem to have comparable power, with the sieve-based procedures CP and PS and the block-based procedures PaP

Table 6.5 (continued): Size-adjusted power of the considered tests for $p = 1$

	$T = 50$							$T = 100$							$T = 250$						
ϕ	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0	0.0	0.8	-0.8	0.0	0.0	0.0	0.0
θ	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5	0.0	0.0	0.0	0.8	-0.8	0.5	-0.5
MZ_α	26.6	3.4	9.3	3.6	17.9	3.2	22.2	32.8	12.8	14.9	12.8	16.5	19.3	22.2	33.9	24.3	25.6	23.3	18.1	28.4	26.8
MZ_t	27.0	3.6	9.8	3.8	17.9	3.5	22.4	32.8	13.3	15.0	13.2	16.5	19.5	22.2	34.1	24.8	25.7	23.6	18.1	28.7	26.8
MSB	26.1	3.2	9.1	3.3	17.9	3.0	22.1	32.1	12.0	13.2	12.2	16.6	18.4	22.0	33.6	23.7	25.5	22.8	18.2	27.9	26.3
MP_T	27.4	3.8	10.2	3.9	17.8	3.9	22.4	33.0	13.8	16.6	13.6	16.4	19.9	22.2	34.4	25.1	25.5	23.6	18.0	28.7	26.9
$PQMZ_\alpha$	28.4	3.3	12.0	3.5	27.7	3.1	25.8	34.2	13.4	17.7	14.4	25.9	19.3	24.9	34.9	24.8	26.8	24.0	20.8	29.2	27.9
$PQMZ_t$	28.9	3.5	12.1	3.7	27.7	3.4	26.0	34.1	13.9	18.3	14.8	25.8	19.6	24.9	35.1	25.3	26.9	24.3	20.8	29.4	27.9
$PQMSB$	27.6	3.1	11.3	3.2	27.7	2.9	25.6	33.6	12.6	16.8	13.8	25.9	18.3	24.6	34.5	24.3	26.7	23.5	20.8	28.7	27.4
$PQMP_T$	29.1	3.6	12.0	3.9	27.8	3.8	26.1	34.1	14.3	18.1	15.2	25.8	19.8	25.0	35.3	25.6	26.6	24.4	20.6	29.6	28.0
MZ_α^{iwb}	34.6	4.2	14.3	5.4	15.5	5.7	19.4	35.3	14.4	19.6	14.4	15.2	21.1	22.0	36.0	25.2	26.3	23.9	17.3	29.4	27.6
MZ_t^{iwb}	35.2	4.4	14.7	5.6	15.5	6.2	19.5	35.4	15.0	19.6	14.8	15.2	21.3	22.3	36.1	25.6	26.7	24.1	17.4	30.0	27.6
MSB^{iwb}	33.2	3.9	13.8	4.8	15.7	5.0	19.4	34.8	13.4	18.4	13.9	15.3	20.4	21.6	35.3	24.5	26.3	23.1	17.3	28.6	27.0
MP_T^{iwb}	35.2	4.5	14.7	5.7	15.5	6.3	19.5	35.3	15.2	20.0	14.9	15.2	21.6	22.4	36.0	25.6	26.8	24.3	17.3	30.1	27.7
MZ_α^{rwb}	28.3	8.7	7.4	8.8	18.9	8.8	19.7	29.2	14.5	13.0	14.4	15.1	20.1	15.4	33.0	24.8	26.1	22.1	14.3	26.4	23.8
MZ_t^{rwb}	28.8	9.1	7.6	9.1	18.9	9.3	19.8	29.4	15.2	13.3	14.7	15.1	20.3	15.5	33.0	24.9	26.1	22.2	14.3	26.4	24.0
MSB^{rwb}	28.1	8.2	7.0	8.1	18.8	8.1	19.8	29.2	13.7	13.0	13.8	15.2	19.3	15.4	32.7	24.1	26.4	22.0	14.5	26.2	23.3
MP_T^{rwb}	28.7	10.0	7.7	9.2	18.9	9.5	19.7	29.7	15.9	13.4	15.0	15.1	20.6	15.6	33.0	25.3	26.3	22.7	14.4	26.5	24.1
MZ_α^{dwb}	34.9	4.2	13.8	5.5	15.6	5.7	19.6	36.0	13.5	20.3	15.0	15.1	21.6	21.6	37.2	26.1	26.1	24.5	18.2	30.4	27.4
MZ_t^{dwb}	35.0	4.5	14.2	5.7	15.5	6.4	19.5	35.9	14.1	21.0	15.3	15.0	21.9	21.6	37.4	26.6	26.7	24.8	17.8	30.7	27.7
MSB^{dwb}	33.7	3.9	13.4	5.0	15.8	5.0	19.5	35.8	12.6	19.3	14.3	15.3	20.8	21.4	36.7	25.1	26.4	23.8	18.2	29.7	26.8
MP_T^{dwb}	35.0	4.5	14.3	5.8	15.5	6.6	19.5	36.1	14.4	21.1	15.5	15.0	22.1	21.6	37.5	26.8	26.7	25.1	17.8	30.6	27.8
MZ_α^{awb}	34.5	4.7	14.2	6.8	15.2	8.8	19.6	35.2	14.1	20.0	15.1	15.1	22.4	22.1	36.5	25.5	28.0	24.2	17.9	30.9	28.1
MZ_t^{awb}	34.5	4.9	14.6	7.1	15.2	9.4	19.5	35.2	14.7	20.5	15.5	15.0	22.8	22.2	36.9	25.7	28.1	24.7	17.7	31.2	28.3
MSB^{awb}	33.4	4.4	13.4	6.2	15.4	8.0	19.6	34.7	13.3	19.2	14.4	15.2	21.5	21.9	35.8	24.8	27.4	23.8	18.1	30.2	27.8
MP_T^{awb}	34.5	5.0	14.5	7.2	15.2	9.5	19.5	35.3	15.1	20.5	15.6	15.0	22.9	22.3	37.2	26.1	28.1	24.8	17.6	31.3	28.4
$PQMZ_\alpha^{\text{iwb}}$	37.3	4.0	15.7	5.4	26.7	6.1	23.8	36.1	14.5	21.7	15.9	24.9	20.1	25.3	36.5	26.4	28.1	25.4	20.7	31.0	28.7
$PQMZ_t^{\text{iwb}}$	37.2	4.3	15.9	5.9	26.7	6.8	23.8	36.4	15.1	22.2	16.3	24.8	20.5	25.4	36.8	26.9	28.2	25.7	20.6	31.5	28.8
$PQMSB^{\text{iwb}}$	35.8	3.7	15.6	5.0	26.8	5.4	23.5	35.1	13.5	20.5	15.4	25.0	19.2	25.0	36.1	25.4	27.9	24.8	20.8	30.3	28.3
$PQMP_T^{\text{iwb}}$	37.1	4.3	15.9	5.9	26.7	6.9	23.8	36.5	15.4	22.2	16.6	24.8	20.8	25.4	37.0	27.0	28.4	25.9	20.6	31.3	28.8
$PQMZ_\alpha^{\text{rwb}}$	32.2	9.1	9.2	9.3	31.8	9.5	24.1	30.9	14.6	14.0	15.0	27.4	20.3	17.7	33.9	24.7	27.8	24.2	17.2	27.3	27.0
$PQMZ_t^{\text{rwb}}$	32.2	9.7	9.3	9.7	31.8	10.0	24.1	31.5	15.0	14.1	15.4	27.4	20.3	17.8	33.9	24.8	27.4	24.3	17.2	27.6	26.8
$PQMSB^{\text{rwb}}$	31.3	8.5	9.0	8.5	31.6	8.6	23.9	31.0	13.8	13.5	14.2	27.4	19.6	17.7	33.0	23.6	27.0	23.8	17.3	26.6	26.4
$PQMP_T^{\text{rwb}}$	32.0	10.5	9.4	9.9	32.1	10.2	24.1	31.5	15.9	14.2	15.9	27.3	20.5	18.0	34.0	25.2	27.5	24.4	17.1	27.8	27.1
$PQMZ_\alpha^{\text{dwb}}$	36.7	4.0	15.8	5.5	26.6	5.9	23.5	37.2	14.4	21.7	16.6	25.1	21.8	24.9	37.2	25.7	29.4	24.9	20.9	28.8	27.6
$PQMZ_t^{\text{dwb}}$	37.1	4.2	15.8	5.8	26.6	6.5	23.6	37.6	15.1	22.2	17.0	25.0	22.2	25.1	37.6	26.3	30.3	25.5	20.8	29.5	28.1
$PQMSB^{\text{dwb}}$	35.8	3.7	15.2	5.1	26.7	5.1	23.5	36.8	13.4	20.6	16.1	25.2	21.1	24.5	36.7	24.7	29.1	24.4	21.0	28.0	27.2
$PQMP_T^{\text{dwb}}$	37.3	4.2	15.8	5.8	26.6	6.6	23.5	37.7	15.4	22.6	17.1	25.0	22.5	25.1	37.7	26.5	30.3	25.7	20.7	29.5	28.2
$PQMZ_\alpha^{\text{awb}}$	36.1	4.6	15.2	7.2	26.3	9.4	23.1	36.7	15.0	21.5	16.9	24.8	22.3	25.3	37.0	25.9	27.9	25.8	20.4	31.3	29.9
$PQMZ_t^{\text{awb}}$	35.8	4.7	15.5	7.7	26.2	9.9	23.2	37.1	15.7	21.9	17.3	24.7	22.8	25.3	37.4	26.4	28.2	26.0	20.3	31.6	30.3
$PQMSB^{\text{awb}}$	35.5	4.4	14.8	6.5	26.4	8.2	23.3	36.6	14.0	20.4	16.4	24.9	21.8	25.0	36.7	25.3	27.4	25.4	20.3	30.8	29.5
$PQMP_T^{\text{awb}}$	35.9	4.8	15.5	7.7	26.2	9.9	23.2	36.9	16.0	21.6	17.3	24.6	22.9	25.5	37.3	26.5	28.4	26.2	20.4	31.6	30.3

having somewhat better power in some cases. This difference disappears as the sample size increases. The same is true in the case of a large negative MA parameter.

Notice the exceptionally low power of the M tests in the case of a large negative MA parameter, even for large sample sizes. The power of these tests is improved slightly if lag selection is based on OLS detrended data. In these cases it is interesting to note that the performance of the M tests closely resembles that of the efficient ADF test of Elliott et al. (1996).

AR innovations

As was the case for MA innovations with a large positive or negative parameter, in the case of AR innovations with a large positive or negative parameter the M tests have rather low power for small samples. The power of the tests seems to increase quite substantially as the sample size increases, unlike in the case of innovations with a large positive or negative MA

parameter.

6.3 Conditionally heteroskedastic innovations

We now move on to the case where the error process $\{u_t\}$ exhibits some form of conditional heteroskedasticity. The models we consider in this section are the same models as those considered by Gonçalves and Kilian (2004) and Cavaliere and Taylor (2009a), which are as follows:

Model A. A standard GARCH(1,1) process defined by

$$u_t = \sqrt{h_t} v_t, \quad t = 1, 2, \ldots, T,$$

where v_t are i.i.d. $N(0,1)$ random variables and

$$h_t = \mu + \alpha u_{t-1}^2 + \beta h_{t-1}.$$

For all simulations we took $\mu = 1$ and considered the parameter configurations $(\alpha, \beta) \in \{(0.5, 0.0), (0.3, 0.65), (0.2, 0.79), (0.05, 0.94)\}$, which we refer to as Model A1 to A4.

Model B. As in Model A, but with v_t i.i.d. t_5 random variables (standardised to have unit variance). We refer to these models as Model B1 to B4.

Model C. EGARCH(1,1) or exponential GARCH of Nelson (1991) with

$$u_t = \sqrt{h_t} v_t, \quad t = 1, 2, \ldots, T,$$

where $\ln(h_t) = -0.23 + 0.9 \ln(h_{t-1}) + 0.25(v_{t-1}^2 - 0.3 v_{t-1})$, with v_t i.i.d. $N(0,1)$ random variables.

Model D. AGARCH(1,1) or asymmetric GARCH of Engle (1990) with

$$u_t = \sqrt{h_t} v_t, \quad t = 1, 2, \ldots, T,$$

where $h_t = 0.0216 + 0.6896 h_{t-1} + 0.3174(u_{t-1} - 0.1108)^2$, with v_t i.i.d. $N(0,1)$ random variables.

Model E. GJR–GARCH(1,1) of Glosten, Jaganathan and Runkle (1993) with

$$u_t = \sqrt{h_t} v_t, \quad t = 1, 2, \ldots, T,$$

where $h_t = 0.005 + 0.7 h_{t-1} + 0.28(|u_{t-1}| - 0.23 u_{t-1})^2$, with v_t i.i.d. $N(0,1)$ random variables.

Remark. For all the above models the processes $\{h_t\}$ and $\{v_t\}$ were initialised by means of a burn-in period of length $T/4$.

We now move on to make a few comments on the Monte Carlo results based on these DGPs. Throughout, for tests based on the M tests we only report results for tests based on MZ_α as the results for all M tests are very similar.

6.3.1 Size properties

First observe that in most cases the asymptotic and bootstrap tests seem to preserve the nominal level of the test fairly well, with some mild distortions for some of the models for some parameter configurations. The size distortions are certainly much more severe than those seen in the case of very persistent AR processes or MA processes with a parameter close to 1 or -1. Not surprisingly, the most prominent size distortions are seen with the asymptotic tests, and especially when the data are generated by Model B to E, with Model B causing the largest size distortions.

Clearly the sieve bootstrap tests CP and PS and the block bootstrap tests PaP are mildly oversized for some of the models. This suggests that the sieve and block bootstrap procedures do not replicate the pattern of conditional heteroskedasticity seen in the data very well. In contrast, the tests based on the wild bootstrap seem to do much better in preserving the level, attesting to the wild bootstrap's ability to cope with conditional heteroskedasticity. These aspects are especially visible in the case where the deterministic component of the test is a constant mean.

Note that the linear process bootstrap of Zou and Politis (2016) is also mildly oversized for some models. However, when allowing lag selection in the test (i.e. not the standard Dickey–Fuller test proposed by Zou and Politis), the test is not really distorted by the presence of conditional heteroskedasticity.

6.3.2 Power properties

In terms of power, most tests do fairly well and are comparable in power. The sieve, block and wild bootstrap tests based on the ADF t-statistic have lower power than those based on the coefficient statistic in the case of a constant mean. However, the power of these tests is comparable in the case where the deterministic component is a linear time trend.

It is worth noting that our implementation of the LPB of Zou and Politis (2016) which makes use of GLS detrending outperforms the bootstrap-based M tests in the linear trend case. In the constant case the LPB test that makes use of OLS detrending has power comparable to that of the M tests.

6.4 Conclusions

We conclude the chapter with the following remarks:

- We have seen that, in the case of small samples, following the suggestion of Perron and Qu (2007) to base lag selection on OLS detrended data instead of GLS detrended data might aggravate size distortions in the presence of innovations which follow an MA process with a parameter close to -1.
- It is clear from the Monte Carlo study that most of the procedures based on any of the variations of the wild bootstrap cope very well when there is conditional heteroskedas-

ticity present in the innovations. In contrast, some size distortions are seen for some parameter configurations when using the sieve and block bootstrap tests.

- The Monte Carlo results suggest that recolouring the bootstrap sample (generated by the independent wild bootstrap) by means of a sieve contributes more towards respecting the nominal size of the test than reintroducing autocorrelation by means of the autoregressive or dependent wild bootstrap procedures.
- We have seen that the linear process bootstrap delivered promising results in some cases. We note again that this bootstrap method is fairly new and more research is necessary to determine how it can be optimally implemented for unit root testing.

Table 6.6: Size of the considered tests for $p = 0$ (conditional heteroskedastic DGPs)

	$T = 100$											$T = 250$										
Model	A1	A2	A3	A4	B1	B2	B3	B4	C	D	E	A1	A2	A3	A4	B1	B2	B3	B4	C	D	E
DF_t	4.4	5.0	3.7	2.1	5.4	5.0	2.1	1.5	5.5	4.5	4.7	4.5	5.6	5.7	3.4	5.5	8.2	4.2	2.3	5.6	6.7	6.5
DF_ρ	5.6	7.2	6.2	4.4	6.6	9.3	8.1	5.4	7.6	7.6	7.6	5.1	7.0	7.6	4.7	5.8	13.5	14.4	8.0	6.9	9.4	9.4
DF_t^{PQ}	4.3	5.0	3.7	2.2	5.3	5.1	2.3	1.8	5.4	4.6	4.6	4.2	5.1	5.2	3.2	5.1	7.5	3.9	2.1	5.2	6.1	6.1
DF_ρ^{PQ}	4.3	5.3	4.5	3.5	5.4	7.0	6.0	4.1	5.6	5.4	5.5	4.4	5.6	6.1	4.1	5.0	11.7	12.9	6.7	5.6	7.6	7.6
Z_t	6.8	7.7	5.9	3.5	9.6	10.0	7.0	3.5	8.8	7.5	7.6	6.5	8.1	8.2	5.0	8.7	15.2	13.6	5.7	8.1	10.3	9.7
Z_ρ	7.1	7.8	6.8	5.2	9.7	11.2	10.4	6.1	8.8	8.3	8.3	6.8	8.2	8.6	6.1	8.6	16.4	19.1	9.4	8.3	10.6	10.3
ERS_t	5.3	5.8	5.7	5.7	5.4	6.2	6.9	6.4	5.6	5.9	6.1	5.0	5.2	5.6	5.5	4.7	7.7	11.8	7.8	5.0	5.7	5.8
P_T	3.5	4.4	4.5	3.8	3.8	6.8	9.0	5.6	4.1	5.1	5.1	4.0	4.5	5.1	4.3	3.8	8.8	15.9	9.2	4.3	5.8	5.6
ERS_t^{PQ}	5.5	6.1	6.1	6.1	5.7	6.8	7.5	6.9	5.9	6.4	6.5	5.0	5.1	5.6	5.5	4.7	7.6	11.5	7.8	5.0	5.7	5.7
P_T^{PQ}	3.6	4.4	4.5	3.9	3.9	6.8	8.9	5.6	4.2	5.1	5.1	4.0	4.4	5.0	4.3	3.8	8.4	15.1	8.7	4.2	5.5	5.4
CP_t	5.6	6.1	4.6	3.0	6.3	5.5	3.0	2.4	6.4	5.4	5.7	5.2	6.1	5.9	4.0	5.9	7.9	3.8	2.7	6.3	6.8	6.9
CP_ρ	4.9	5.2	4.5	3.7	5.4	5.6	4.5	3.8	5.4	4.9	5.1	4.8	5.1	5.4	4.4	5.1	8.2	8.0	5.3	5.3	6.2	6.3
$\overline{CP}_t$	5.0	5.2	5.5	5.4	5.0	5.7	6.0	5.7	5.7	5.5	5.8	4.9	5.0	5.6	5.7	4.6	7.5	9.1	8.2	5.0	6.0	5.8
$\overline{CP}_\rho$	4.7	5.2	5.4	5.1	4.8	6.2	7.3	5.9	5.4	5.4	5.6	4.8	4.8	5.6	5.5	4.5	8.4	12.6	9.5	5.1	6.0	5.8
PS_t	5.2	5.6	4.3	3.0	6.2	5.4	2.8	2.3	6.0	5.1	5.3	5.3	6.2	6.3	4.0	5.9	8.7	5.1	2.8	6.2	7.3	7.4
PS_ρ	4.5	4.8	4.2	3.8	5.4	5.8	5.3	4.0	5.2	4.8	5.1	4.9	5.4	5.9	4.5	5.1	9.1	10.6	6.1	5.6	6.6	7.0
$\overline{PS}_t$	5.0	4.5	4.9	4.9	4.7	5.1	5.4	5.2	4.8	5.0	4.9	4.8	5.0	5.7	5.4	4.7	8.1	11.9	8.3	5.1	6.2	6.1
$\overline{PS}_\rho$	4.7	4.5	5.0	4.7	4.5	5.9	7.5	5.7	4.6	5.1	4.9	4.6	4.8	5.7	5.3	4.5	8.8	16.0	9.8	5.1	6.1	6.0
PaP_t	5.7	6.3	5.5	4.1	6.4	6.8	5.0	4.2	6.5	6.4	6.1	5.5	6.1	6.4	4.7	5.9	8.7	7.0	4.3	6.2	7.0	6.9
PaP_ρ	5.3	6.1	5.9	4.9	5.7	7.7	7.9	5.9	6.3	6.4	6.3	5.0	5.9	6.6	5.1	5.0	11.2	14.3	8.2	5.9	7.6	7.7
$\overline{PaP}_t$	5.5	5.4	5.8	5.7	5.2	6.0	6.7	6.1	5.7	5.8	5.9	5.1	5.1	5.9	5.7	4.7	7.8	11.9	8.4	5.2	6.1	5.9
$\overline{PaP}_\rho$	5.3	5.6	6.0	5.7	5.1	7.4	9.4	6.7	5.7	6.2	6.2	5.0	5.3	6.1	5.6	4.6	9.9	17.7	10.7	5.4	6.9	6.5
PaP_t^{diff}	5.3	6.0	5.2	4.0	6.2	6.7	4.5	3.7	6.2	6.1	5.8	5.2	5.9	6.2	4.6	5.8	8.5	5.7	3.7	5.8	6.9	6.7
PaP_ρ^{diff}	4.5	5.1	4.9	4.1	5.2	6.9	6.8	5.0	5.4	5.6	5.5	4.8	5.5	6.1	4.6	5.1	11.0	13.3	7.5	5.5	7.3	7.2
$\overline{PaP}_t^{\text{diff}}$	4.7	4.6	5.1	5.0	4.8	5.6	6.2	5.3	4.9	5.2	5.3	4.6	4.8	5.6	5.3	4.7	8.0	12.1	8.1	5.0	6.0	5.6
$\overline{PaP}_\rho^{\text{diff}}$	4.4	4.8	5.1	4.8	4.9	6.8	8.7	5.8	4.8	5.6	5.4	4.5	5.0	5.9	5.2	4.7	9.9	17.6	10.2	5.1	6.6	6.0
RWB_t	5.3	5.3	5.1	4.8	5.5	5.6	5.0	5.1	5.5	5.0	5.2	5.0	5.2	5.3	5.1	5.3	5.2	4.9	5.1	5.3	5.4	5.5
RWB_ρ	4.8	4.5	4.3	4.3	5.1	4.7	4.2	4.3	4.6	4.2	4.6	4.8	4.2	4.3	4.7	4.8	4.5	4.2	3.8	4.5	4.3	4.6
DWB_t	5.0	5.0	4.9	4.9	5.2	5.2	4.9	5.1	5.2	4.8	4.8	4.8	4.8	5.0	5.2	4.9	5.4	5.4	5.1	5.1	4.9	5.2
DWB_ρ	5.3	5.2	5.4	5.5	5.7	6.5	6.8	5.8	5.6	5.6	5.6	5.1	5.1	5.5	5.2	5.1	6.8	7.7	6.2	5.3	5.7	6.0
AWB_t	5.4	5.4	5.0	5.0	5.5	5.4	4.8	5.3	5.4	5.1	5.5	5.2	5.3	5.2	5.4	5.4	5.7	5.8	5.2	5.4	5.2	5.3
AWB_ρ	5.4	5.9	5.7	5.6	5.8	6.8	6.7	6.1	5.8	5.8	6.2	5.2	5.6	5.7	5.4	5.4	7.2	7.9	6.4	5.5	6.1	6.5
LPB_t^0	4.3	4.7	4.5	3.5	4.7	6.0	5.1	3.7	4.9	4.9	4.9	5.1	6.0	6.7	4.6	5.6	10.4	10.5	5.6	5.9	7.6	7.3
$\overline{LPB}_t^0$	1.8	1.8	2.1	2.0	1.6	3.1	5.0	2.9	1.8	2.3	2.1	1.9	2.1	2.9	2.3	1.8	5.5	11.5	5.6	2.0	3.0	2.7
LPB_t	5.2	5.7	4.6	2.8	6.3	5.6	3.1	2.5	6.1	5.4	5.3	4.8	5.6	5.7	3.7	5.8	8.0	4.5	2.6	5.6	6.5	6.5
$\overline{LPB}_t$	2.5	2.5	2.7	2.5	2.5	2.9	3.3	2.9	2.7	2.9	2.9	2.4	2.3	2.7	2.7	2.1	4.1	6.9	4.1	2.4	2.8	2.7
MZ_α	4.7	5.8	6.0	5.3	5.1	8.9	11.7	7.7	5.6	6.7	6.6	4.9	5.7	6.2	5.4	4.8	10.8	20.0	11.9	5.3	7.0	6.8
$PQMZ_\alpha$	4.8	5.9	6.0	5.5	5.3	8.9	11.6	7.7	5.6	6.8	6.7	4.9	5.4	6.1	5.3	4.7	10.3	19.0	11.3	5.2	6.8	6.5
MZ_α^{iwb}	3.9	4.3	4.0	3.8	3.7	5.3	6.9	5.0	4.0	4.6	4.6	4.4	4.4	4.7	4.5	4.2	5.7	8.3	6.2	4.2	5.0	4.8
MZ_α^{rwb}	3.3	3.4	3.3	3.4	3.2	3.9	4.6	3.9	3.2	3.4	3.6	4.2	3.8	3.8	4.2	4.0	4.4	5.9	4.4	3.6	4.0	4.0
MZ_α^{dwb}	3.9	4.3	4.3	4.0	3.9	5.6	7.0	5.2	4.0	4.6	4.8	4.6	4.5	4.6	4.6	4.2	5.8	8.5	6.5	4.3	5.0	4.9
MZ_α^{awb}	4.0	4.5	4.4	4.0	4.1	5.5	7.1	5.4	4.1	4.8	4.9	4.6	4.6	4.8	4.6	4.3	6.1	8.5	6.4	4.3	5.0	5.1
$PQMZ_\alpha^{\text{iwb}}$	4.0	4.3	4.3	4.0	4.1	5.3	6.6	5.0	4.0	4.7	4.7	4.2	4.3	4.5	4.5	4.1	5.6	7.7	6.0	4.2	4.7	4.7
$PQMZ_\alpha^{\text{rwb}}$	3.4	3.4	3.4	3.5	3.6	3.8	4.5	3.9	3.3	3.6	3.6	4.2	3.8	3.8	4.1	3.8	4.1	5.6	4.1	3.6	3.8	3.8
$PQMZ_\alpha^{\text{dwb}}$	3.9	4.4	4.4	4.0	4.0	5.5	6.8	5.3	4.0	4.7	4.8	4.2	4.2	4.3	4.5	4.2	5.5	8.0	6.1	4.2	4.8	4.8
$PQMZ_\alpha^{\text{awb}}$	4.1	4.5	4.6	4.3	4.5	5.6	7.0	5.4	4.4	4.9	4.9	4.4	4.3	4.5	4.6	4.2	5.7	8.1	6.1	4.4	4.8	4.9

Table 6.7: Size of the considered tests for $p = 1$ (conditional heteroskedastic DGPs)

	$T = 100$											$T = 250$										
Model	A1	A2	A3	A4	B1	B2	B3	B4	C	D	E	A1	A2	A3	A4	B1	B2	B3	B4	C	D	E
DF_t	4.3	4.8	3.5	2.4	5.2	4.3	2.3	2.0	5.3	4.3	4.5	4.5	5.6	5.4	3.5	5.2	6.9	3.9	2.7	5.6	6.3	6.7
DF_ρ	7.7	10.8	9.5	6.5	9.2	13.6	12.3	8.6	11.4	11.5	11.6	6.0	9.8	10.1	5.8	7.0	17.2	17.3	10.9	9.2	13.0	12.8
DF_t^{PQ}	3.8	4.4	3.4	2.2	4.9	4.4	2.6	2.0	4.9	4.0	4.1	3.8	4.6	4.3	3.0	4.4	5.9	3.1	2.1	4.5	5.2	5.5
DF_ρ^{PQ}	4.5	5.3	4.4	3.4	5.8	6.3	5.5	3.8	5.7	5.2	5.3	4.1	5.6	5.4	3.8	4.9	10.9	11.5	5.9	5.4	7.4	7.5
Z_t	7.2	7.6	5.6	3.3	11.2	10.0	7.3	3.8	9.0	7.6	7.7	7.4	9.4	8.8	5.5	10.9	16.0	14.1	6.6	9.8	11.4	11.4
Z_ρ	7.2	7.5	6.1	4.0	11.3	10.7	9.2	5.1	8.7	7.8	7.9	7.9	9.9	9.4	6.5	11.0	17.2	18.6	9.4	10.1	11.7	12.1
ERS_t	2.8	3.0	3.0	2.9	2.9	3.1	3.0	2.9	3.1	2.9	3.2	2.4	2.5	2.6	2.6	2.2	3.3	4.0	3.0	2.3	2.7	3.0
P_T	2.5	3.0	3.0	2.4	2.9	5.4	6.9	3.5	3.1	3.5	3.7	3.2	4.3	4.7	3.5	3.3	9.5	15.7	7.9	3.9	5.7	5.7
ERS_t^{PQ}	3.2	3.5	3.5	3.3	3.4	3.9	3.8	3.4	3.6	3.4	3.7	2.4	2.7	2.9	2.7	2.4	3.7	4.3	3.3	2.6	3.0	3.3
P_T^{PQ}	2.7	3.2	3.1	2.6	3.0	5.3	6.5	3.5	3.2	3.5	3.8	3.2	4.2	4.5	3.5	3.4	8.7	13.8	7.2	4.0	5.5	5.5
CP_t	4.8	5.5	4.4	3.4	6.1	4.9	3.3	3.0	5.8	4.9	5.1	5.2	5.7	5.6	4.1	5.5	6.9	4.1	3.1	5.9	6.3	6.9
CP_ρ	4.5	4.8	4.1	3.5	5.5	4.6	3.9	3.6	5.3	4.5	4.7	4.5	4.7	4.6	4.2	4.9	6.1	6.0	4.1	5.0	5.2	5.6
$\overline{CP}_t$	4.5	5.0	4.9	4.7	5.0	5.0	5.0	5.0	5.1	4.8	5.0	5.2	5.2	5.4	5.2	4.5	6.6	7.0	6.1	4.9	5.5	5.9
$\overline{CP}_\rho$	4.4	4.7	4.6	4.4	4.9	5.1	5.2	4.9	4.9	4.6	4.7	4.8	5.0	5.0	5.0	4.4	6.5	8.0	6.5	4.7	5.3	5.5
PS_t	4.9	5.3	4.3	3.3	5.9	4.3	2.9	2.7	5.6	4.4	4.7	5.2	6.0	5.6	4.2	6.1	7.1	4.2	3.1	6.2	6.7	6.7
PS_ρ	4.3	4.7	4.1	3.5	5.4	4.7	4.6	3.5	4.9	4.2	4.3	4.6	5.1	4.7	4.3	5.3	7.1	8.5	5.1	5.3	5.7	5.7
$\overline{PS}_t$	4.5	4.2	4.3	4.5	4.5	4.0	4.1	4.5	4.2	4.2	4.3	5.0	5.3	5.2	5.5	4.7	6.7	7.4	6.0	5.1	5.6	5.7
$\overline{PS}_\rho$	4.4	4.2	4.3	4.2	4.5	4.6	5.6	4.7	4.2	4.2	4.3	4.8	5.1	4.9	5.3	4.6	7.1	10.1	7.1	4.7	5.6	5.8
PaP_t	4.8	5.2	5.0	4.4	5.5	5.2	4.6	4.6	5.6	5.0	5.1	5.2	5.3	5.5	4.8	5.5	6.2	5.6	4.5	5.5	5.7	6.0
PaP_ρ	5.1	5.6	5.4	4.8	5.5	6.3	6.7	5.5	5.9	5.6	5.7	4.7	5.9	6.0	5.0	4.9	9.5	11.7	7.2	5.5	6.9	7.2
$\overline{PaP}_t$	5.5	6.1	6.0	5.7	5.2	5.8	6.2	6.2	6.0	5.8	6.1	5.7	5.8	5.9	5.7	5.1	6.8	7.4	6.0	5.5	5.9	6.1
$\overline{PaP}_\rho$	5.4	6.1	6.1	5.5	5.2	6.9	7.9	6.6	5.9	6.1	6.3	5.3	6.4	6.3	5.5	4.7	9.6	12.8	8.3	5.8	7.2	7.2
PaP_t^{diff}	4.4	5.0	4.6	4.0	4.8	5.0	4.1	3.9	5.1	4.7	4.7	5.0	5.2	5.5	4.5	5.2	6.6	4.5	4.0	5.4	6.0	6.1
PaP_ρ^{diff}	4.1	4.4	4.1	3.6	4.3	4.9	5.0	4.1	4.5	4.4	4.3	4.5	5.2	5.5	4.3	4.8	8.9	10.4	6.3	5.2	6.5	6.4
$\overline{PaP}_t^{\text{diff}}$	4.2	4.5	4.6	4.2	4.3	4.7	4.5	4.5	4.6	4.4	4.6	4.9	5.3	5.4	5.0	4.7	6.5	7.0	5.6	4.9	5.5	5.7
$\overline{PaP}_\rho^{\text{diff}}$	3.8	4.3	4.4	3.9	4.2	5.3	5.9	4.6	4.6	4.5	4.5	4.6	5.4	5.5	4.7	4.6	8.7	11.8	7.6	5.0	6.1	6.4
RWB_t	4.7	5.3	5.0	4.6	5.2	5.1	4.7	4.5	5.2	4.9	5.2	4.9	5.3	5.3	4.9	5.1	5.5	5.2	5.2	5.3	5.6	5.9
RWB_ρ	4.4	4.4	4.2	4.1	4.6	4.1	3.9	4.1	4.4	4.0	4.2	4.3	4.0	3.6	4.2	4.4	3.6	3.6	3.2	3.9	3.6	3.9
DWB_t	5.2	5.4	5.3	5.2	5.6	5.2	5.2	5.2	5.5	5.1	5.1	5.1	4.8	4.9	5.2	5.1	5.3	5.2	5.0	5.1	5.2	5.3
DWB_ρ	5.8	5.9	5.6	5.5	6.4	6.4	7.0	5.9	6.3	6.0	5.9	5.3	5.5	5.4	5.2	5.4	7.0	7.8	6.4	5.5	6.2	6.5
AWB_t	5.7	5.7	5.9	5.9	6.1	5.8	5.4	5.7	5.7	5.4	5.8	5.5	5.6	5.6	5.8	5.7	5.7	5.6	5.5	5.5	5.7	6.0
AWB_ρ	6.3	6.2	6.2	6.1	6.6	6.8	7.1	6.3	6.4	6.2	6.6	5.5	6.2	5.9	5.7	5.8	7.3	8.0	6.9	5.8	6.6	7.1
LPB_t^0	5.2	5.5	5.7	5.3	4.9	5.5	6.6	6.1	5.1	5.4	5.8	8.7	9.1	9.9	8.2	8.1	12.6	14.0	9.8	9.1	10.3	10.4
$\overline{LPB}_t^0$	7.3	7.5	7.6	7.7	6.4	7.7	9.1	8.2	7.2	7.4	7.7	9.0	9.3	10.0	9.3	7.7	11.8	16.6	13.0	8.3	9.9	10.2
LPB_t	7.3	8.1	6.9	5.2	8.9	7.1	5.2	4.8	8.8	7.4	7.9	7.8	8.3	8.0	6.4	8.1	9.0	5.7	4.8	8.2	8.4	9.2
$\overline{LPB}_t$	5.9	6.2	6.3	6.0	6.2	6.8	6.8	6.4	6.6	6.5	6.8	5.9	6.2	6.1	5.9	5.6	7.8	8.7	6.9	5.7	6.4	6.9
MZ_α	2.2	2.8	2.7	2.1	2.6	5.3	7.2	3.4	2.9	3.3	3.4	2.9	3.9	4.3	3.1	3.0	9.3	16.8	8.3	3.5	5.3	5.3
$PQMZ_\alpha$	2.4	3.0	2.9	2.3	2.8	5.3	6.8	3.4	3.0	3.3	3.5	2.9	3.9	4.2	3.1	3.0	8.6	15.0	7.5	3.6	5.2	5.1
MZ_α^{iwb}	3.4	3.9	3.8	3.5	3.5	5.6	7.1	4.4	3.9	4.2	4.3	3.9	4.2	4.5	3.7	3.5	6.2	9.6	6.7	3.9	4.9	5.0
MZ_α^{rwb}	2.8	2.6	2.6	2.8	2.6	3.2	4.1	2.9	2.5	2.5	2.9	3.2	2.7	2.7	3.3	3.0	3.3	5.1	3.5	2.5	2.7	3.1
MZ_α^{dwb}	3.8	4.2	4.1	4.0	3.8	6.0	7.6	4.7	4.2	4.6	4.7	3.8	4.5	4.6	4.1	3.5	6.4	10.1	6.9	4.0	5.0	5.2
MZ_α^{awb}	4.2	4.7	4.7	4.4	4.5	6.4	7.8	5.0	4.7	4.9	5.1	4.2	4.8	5.1	4.3	3.9	7.0	10.4	7.3	4.4	5.3	5.3
$PQMZ_\alpha^{\text{iwb}}$	3.9	3.9	4.0	3.7	3.5	5.2	6.5	4.3	3.9	4.1	4.4	3.8	4.1	4.3	3.9	3.6	5.6	8.0	6.0	3.7	4.5	4.7
$PQMZ_\alpha^{\text{rwb}}$	3.0	2.9	2.8	3.1	2.8	3.2	4.0	3.1	2.9	2.8	2.9	3.2	2.8	2.5	3.2	2.8	2.8	4.2	3.0	2.6	2.6	2.7
$PQMZ_\alpha^{\text{dwb}}$	4.1	4.4	4.5	4.1	4.1	5.8	7.1	4.9	4.3	4.6	4.8	3.9	4.3	4.3	3.9	3.6	5.7	8.3	6.1	3.8	4.6	4.9
$PQMZ_\alpha^{\text{awb}}$	4.6	4.9	5.0	4.6	4.5	6.3	7.4	5.4	4.8	5.0	5.2	4.3	4.6	4.6	4.5	3.9	6.0	8.8	6.4	4.3	5.1	5.5

Table 6.8: Size-adjusted power of the considered tests for $p = 0$ (conditional heteroskedastic DGPs)

	$T = 100$											$T = 250$										
Model	A1	A2	A3	A4	B1	B2	B3	B4	C	D	E	A1	A2	A3	A4	B1	B2	B3	B4	C	D	E
DF_t^c	12.0	9.9	11.9	16.8	10.5	8.3	13.3	17.4	10.0	10.0	10.5	14.1	11.8	11.1	16.5	12.2	7.7	11.0	16.6	11.9	10.0	10.9
DF_ρ	19.4	16.6	18.6	22.4	17.8	13.8	14.4	18.9	16.5	16.5	16.9	22.3	19.4	18.7	23.8	20.3	14.4	13.6	18.5	19.2	17.3	17.8
DF_t^{PQ}	13.6	11.5	13.9	18.7	11.8	10.0	14.9	19.2	11.7	12.0	12.5	14.9	12.9	12.1	17.5	12.9	8.6	11.8	17.9	12.8	10.9	11.7
DF_ρ^{PQ}	22.3	19.8	22.3	25.5	19.6	16.1	17.0	22.3	19.5	20.0	20.4	23.6	21.4	20.6	25.3	21.5	15.1	14.0	20.0	20.9	19.1	19.5
ERS_t	26.0	25.7	27.6	30.2	22.8	23.0	24.0	29.3	24.3	26.1	26.9	26.8	26.2	25.8	29.2	24.8	20.2	19.4	27.0	25.9	24.4	25.6
P_T	18.7	17.5	19.4	23.6	17.3	16.7	19.5	24.1	17.4	18.5	18.8	19.3	19.1	18.6	21.9	18.7	15.6	16.9	23.5	19.3	17.9	19.0
ERS_t^{PQ}	27.8	26.1	28.4	32.6	27.8	24.7	23.1	29.6	25.2	26.4	26.9	27.8	26.0	26.5	31.9	29.2	23.0	17.8	26.9	26.6	25.9	26.5
P_T^{PQ}	25.3	23.3	25.0	29.7	25.3	20.2	16.7	24.9	23.1	22.6	23.3	25.9	24.7	25.2	30.1	27.7	21.9	15.2	22.8	25.1	24.3	25.0
CP_t	28.6	27.3	29.5	33.5	28.2	25.5	23.7	30.7	26.3	27.2	28.0	27.7	26.2	26.7	32.1	29.1	23.1	17.9	27.2	26.6	26.1	26.8
CP_ρ	26.0	24.3	25.8	30.4	25.6	21.2	17.9	26.2	24.1	23.7	24.4	26.0	24.9	25.5	30.2	27.6	21.9	15.3	23.4	25.3	24.5	25.2
$\overline{CP}_t$	13.2	11.6	13.8	18.6	12.0	10.5	13.3	18.7	11.9	12.1	12.5	15.0	13.5	13.0	17.6	13.6	9.4	11.7	17.9	13.3	11.8	12.3
$\overline{CP}_\rho$	19.5	17.6	19.7	23.7	18.0	13.6	13.4	19.4	17.3	17.5	18.2	22.3	19.9	19.8	24.3	20.3	12.7	11.5	18.2	19.9	17.2	18.0
PS_t	24.0	23.6	25.5	29.5	25.6	21.5	19.8	27.1	20.9	23.5	23.8	27.7	25.7	25.1	29.9	28.0	21.6	16.3	24.9	25.2	24.4	25.6
PS_ρ	23.7	23.2	25.0	29.5	24.8	19.9	18.2	26.9	20.9	22.8	23.9	27.0	26.0	24.5	29.6	27.0	20.5	14.9	24.2	24.3	24.2	24.9
$\overline{PS}_t$	13.9	12.1	14.0	17.3	12.3	9.7	13.2	17.1	11.9	11.9	12.2	15.5	13.3	12.9	18.0	13.9	9.4	11.9	17.6	13.3	11.5	11.8
$\overline{PS}_\rho$	20.1	17.7	18.5	21.0	17.3	11.9	11.9	17.7	16.2	16.0	16.3	22.6	19.8	18.4	24.6	20.6	12.6	11.3	17.1	18.6	16.3	16.3
PaP_t	22.8	22.7	23.4	28.2	24.8	19.9	18.8	25.0	21.2	21.8	22.2	28.1	25.8	25.3	31.3	27.9	22.2	17.7	24.7	24.7	23.4	24.1
PaP_ρ	22.1	21.7	22.1	27.9	23.8	18.1	15.9	23.5	20.6	20.6	21.5	27.4	25.6	24.1	30.7	27.1	20.5	15.9	23.8	23.5	22.5	23.1
$\overline{PaP}_t$	14.8	13.4	15.9	21.6	12.9	11.5	16.3	21.8	13.0	13.6	13.8	15.9	13.5	13.2	18.4	14.0	9.2	12.8	19.5	13.5	11.7	12.5
$\overline{PaP}_\rho$	23.6	21.6	23.5	28.4	20.5	17.4	18.8	25.1	20.6	21.5	21.6	24.7	22.1	22.0	26.4	22.2	14.9	14.3	21.3	21.6	19.1	19.5
PaP_t^{diff}	27.2	26.7	29.3	32.8	28.8	25.7	23.6	30.6	24.9	26.6	28.4	28.0	27.4	26.4	32.0	28.5	23.3	17.9	26.2	26.5	25.1	26.3
PaP_ρ^{diff}	26.6	26.2	28.7	33.0	28.5	24.0	21.4	30.6	24.9	26.4	27.7	27.5	26.8	26.4	31.8	27.6	22.8	16.8	25.8	25.6	24.4	25.9
$\overline{PaP}_t^{\text{diff}}$	13.1	11.1	13.5	17.8	10.9	9.6	13.9	17.7	11.0	11.0	12.1	13.9	12.9	12.2	17.1	13.4	8.3	11.6	17.0	13.2	10.9	11.9
$\overline{PaP}_\rho^{\text{diff}}$	18.4	17.0	18.6	21.2	16.1	13.7	14.7	18.8	15.9	16.2	17.0	21.3	20.4	19.1	22.7	20.2	13.3	13.0	18.6	19.6	17.4	18.1
RWB_t	21.9	23.5	23.5	25.6	22.8	21.5	20.3	24.8	21.3	22.7	21.8	25.8	24.8	24.8	29.0	25.3	22.3	17.8	24.9	24.2	23.3	25.0
RWB_ρ	21.3	22.3	22.8	25.5	21.7	19.2	17.7	23.5	21.3	21.3	21.8	24.8	24.7	24.4	28.7	24.8	21.4	16.8	24.3	23.9	22.8	24.2
DWB_t	12.7	11.3	13.7	18.6	11.4	9.7	13.9	17.8	10.9	11.8	12.0	14.4	12.7	12.1	17.9	13.5	7.9	11.7	18.6	12.5	10.9	11.4
DWB_ρ	19.3	17.5	19.3	24.2	17.2	12.7	13.0	19.7	16.2	17.0	17.0	22.1	18.4	18.0	25.1	21.2	10.5	8.8	17.6	18.8	15.9	16.3
AWB_t	14.2	12.2	15.8	20.4	12.3	11.3	16.5	20.0	11.7	13.3	13.7	15.6	13.5	12.5	18.5	13.8	8.1	13.2	20.5	13.2	11.5	11.7
AWB_ρ	23.0	20.6	22.6	26.4	20.1	15.5	16.4	23.8	18.7	19.6	20.5	24.5	21.3	19.5	26.6	22.0	11.8	9.9	19.8	21.2	17.6	18.5
LPB_t^0	13.7	12.3	15.0	19.8	12.2	11.7	16.1	20.0	12.0	13.0	12.4	15.2	13.1	12.0	18.6	13.3	8.5	12.8	20.2	12.9	10.9	11.7
$\overline{LPB}_t^0$	22.8	19.3	21.8	26.6	20.0	15.5	16.4	22.9	19.4	19.3	19.2	24.8	20.7	20.0	27.1	22.2	12.2	10.4	20.0	20.8	16.8	17.9
LPB_t	12.1	11.5	12.3	13.8	10.3	9.5	11.7	13.2	11.2	11.1	11.5	15.0	13.8	13.4	15.9	13.1	11.0	11.6	15.8	13.9	12.7	12.7
$\overline{LPB}_t$	14.5	15.9	15.4	16.2	15.7	14.5	12.4	15.1	15.8	15.1	16.0	19.8	21.3	19.8	21.6	21.3	18.7	14.2	19.4	20.4	19.7	20.6
MZ_α	14.9	13.3	15.0	19.9	13.2	11.7	15.7	19.5	13.2	13.6	14.2	15.7	13.5	13.3	17.6	13.9	9.6	12.6	18.0	13.8	11.9	12.9
$PQMZ_\alpha$	23.9	21.4	23.1	28.4	22.9	17.1	15.5	23.6	21.1	20.6	21.3	25.0	23.0	23.4	28.8	25.8	18.8	14.5	22.9	23.6	22.0	22.9
MZ_α^{iwb}	24.7	22.6	24.2	29.3	23.3	18.1	16.8	25.2	22.3	21.8	22.4	25.1	23.2	23.7	29.0	25.8	18.8	14.8	23.6	23.8	22.3	23.2
MZ_α^{rwb}	23.9	20.8	23.3	26.8	22.2	14.1	10.4	20.8	19.8	19.1	19.7	25.8	22.2	22.1	28.1	24.8	13.1	7.8	16.5	23.2	19.7	21.9
MZ_α^{dwb}	22.2	19.0	20.2	25.2	21.0	12.5	9.4	18.7	18.7	16.5	18.3	24.1	21.2	21.3	27.2	24.3	12.7	7.4	14.6	22.2	18.9	20.8
MZ_α^{awb}	24.7	20.9	22.6	28.7	22.9	13.9	11.4	21.3	21.0	19.4	20.3	24.7	22.8	22.3	28.3	25.0	13.1	7.9	17.4	23.6	20.1	22.2
$PQMZ_\alpha^{\text{iwb}}$	24.7	21.5	22.8	28.0	23.7	15.2	11.4	22.6	21.2	19.8	21.0	26.3	22.8	23.1	30.3	26.7	13.6	8.1	17.3	24.3	21.1	22.3
$PQMZ_\alpha^{\text{rwb}}$	23.7	20.8	22.9	28.5	22.0	14.6	11.3	21.8	21.1	19.4	20.6	25.0	22.2	21.8	29.0	26.1	12.7	7.8	17.1	22.3	20.0	21.7
$PQMZ_\alpha^{\text{dwb}}$	23.0	20.8	21.0	26.3	20.2	13.3	10.2	20.4	19.9	18.8	19.5	24.4	21.5	20.8	27.5	24.7	13.0	7.4	16.6	22.9	19.1	21.4
$PQMZ_\alpha^{\text{awb}}$	25.5	22.5	23.8	29.5	24.6	15.3	12.1	23.3	21.7	20.1	21.8	26.6	22.4	23.1	28.8	25.9	13.3	8.1	17.3	23.3	20.3	22.1

Table 6.9: Size-adjusted power of the considered tests for $p = 1$ (conditional heteroskedastic DGPs)

	$T = 100$											$T = 250$										
Model	A1	A2	A3	A4	B1	B2	B3	B4	C	D	E	A1	A2	A3	A4	B1	B2	B3	B4	C	D	E
DF_t^c	14.8	12.8	14.7	18.6	13.0	11.8	14.8	18.2	12.3	12.9	13.0	19.6	14.7	14.8	21.9	16.9	10.0	12.6	18.6	14.6	12.1	13.1
DF_ρ	21.4	16.7	17.2	22.2	19.4	13.1	13.0	17.8	16.1	15.1	15.6	26.3	19.3	19.0	26.4	24.6	14.7	13.6	18.3	20.4	16.5	17.7
DF_t^{PQ}	22.7	20.1	22.9	29.1	18.7	17.9	22.7	28.6	18.9	20.6	21.0	24.0	19.4	20.3	27.4	20.4	13.2	17.3	26.2	19.5	16.7	17.7
DF_ρ^{PQ}	30.9	27.9	29.6	34.3	26.7	23.3	23.9	31.6	26.8	26.9	28.0	32.0	26.7	27.6	33.9	29.3	18.9	16.8	26.4	27.5	23.3	24.6
Z_t	27.9	26.8	29.3	33.9	23.7	23.9	26.9	33.0	25.3	26.6	26.8	31.7	29.3	29.8	34.5	27.8	23.7	24.4	34.2	29.0	27.5	28.7
Z_ρ	32.1	31.3	33.6	38.4	26.5	26.9	29.4	36.4	29.8	30.8	31.3	36.9	34.2	34.9	39.7	32.2	27.1	27.1	37.9	34.3	32.7	33.5
ERS_t	30.7	28.9	30.7	35.4	29.4	26.3	27.5	34.4	27.5	29.3	28.7	32.4	28.0	28.0	35.4	31.5	21.9	19.8	30.5	28.6	25.6	26.2
P_T	29.2	26.3	28.5	34.5	28.6	18.8	16.7	29.2	25.5	25.3	24.7	33.5	27.4	26.8	36.2	33.5	18.9	13.3	24.0	28.7	23.8	25.1
ERS_t^{PQ}	32.1	30.3	32.1	36.9	30.2	27.1	28.4	35.6	29.0	31.0	30.4	33.3	28.8	29.0	36.4	31.9	22.4	20.8	31.6	29.3	26.6	27.2
P_T^{PQ}	30.4	28.1	30.4	36.3	29.6	21.2	19.4	31.7	27.5	27.7	27.0	34.4	28.6	28.3	37.2	33.7	19.9	14.5	25.9	29.4	25.0	26.5
CP_t	23.1	19.4	21.3	26.0	18.6	17.1	19.8	25.1	19.1	19.2	20.4	23.7	20.0	20.6	26.7	21.1	14.0	17.1	25.1	19.2	17.3	18.6
CP_ρ	27.5	24.2	25.7	28.9	23.3	19.7	18.9	26.2	23.1	23.2	23.8	29.4	25.0	25.0	30.2	26.7	16.4	13.8	22.7	24.5	20.3	22.7
$\overline{CP}_t$	30.1	26.5	27.5	32.3	26.9	23.6	22.4	29.9	25.2	26.2	26.6	31.9	28.7	28.4	35.9	34.1	21.6	19.7	29.4	29.3	26.3	27.0
$\overline{CP}_\rho$	30.1	27.5	28.6	33.0	28.4	23.4	22.1	30.0	26.3	26.6	27.3	32.7	28.3	28.5	35.7	33.8	21.9	18.2	27.8	29.0	26.1	27.2
PS_t	16.2	12.3	13.4	17.6	14.3	10.8	13.0	16.0	12.1	12.3	12.9	23.5	17.4	18.3	25.8	20.4	11.9	14.7	21.3	17.8	15.0	16.3
PS_ρ	18.9	14.0	14.9	18.1	16.3	11.1	10.7	15.5	14.1	13.7	14.1	27.7	19.9	20.7	27.9	24.1	13.1	11.0	17.6	20.8	16.6	17.8
$\overline{PS}_t$	19.7	17.1	17.3	20.1	21.0	14.0	13.2	17.8	15.4	15.5	15.8	32.2	24.2	25.5	33.1	32.1	17.9	16.1	24.9	25.5	22.2	24.2
$\overline{PS}_\rho$	20.0	17.1	17.8	21.1	21.0	14.0	11.7	18.1	16.0	16.2	16.3	31.7	22.5	24.7	31.9	30.8	17.4	14.4	22.4	25.2	20.7	22.1
PaP_t	26.3	23.5	25.9	32.1	20.2	21.1	25.5	31.7	21.2	23.8	24.3	26.0	22.3	23.1	31.0	22.4	15.1	18.0	27.9	21.3	19.1	20.3
PaP_ρ	33.1	29.6	31.0	36.8	27.7	24.1	25.0	34.2	28.3	28.6	29.4	33.8	27.6	29.4	36.6	30.1	18.7	16.6	28.2	28.1	23.9	25.3
$\overline{PaP}_t$	33.3	30.2	33.1	38.2	32.9	27.4	27.1	34.1	28.6	30.3	30.5	35.1	28.7	30.4	38.3	34.8	22.0	20.5	32.6	30.6	26.4	28.8
$\overline{PaP}_\rho$	34.6	31.3	34.3	40.7	34.2	26.5	25.3	35.3	30.3	30.8	31.5	35.6	28.3	30.4	39.3	35.3	21.5	18.0	31.0	30.4	26.2	28.1
PaP_t^{diff}	16.6	14.9	16.9	19.3	14.5	13.7	17.0	20.3	13.8	15.6	15.6	20.1	17.0	18.0	23.5	17.2	11.9	16.0	22.9	17.2	15.0	15.6
PaP_ρ^{diff}	18.0	17.3	18.3	20.2	17.2	15.2	15.7	19.6	16.5	17.0	17.3	24.4	21.2	21.6	26.0	22.3	14.9	13.9	21.5	22.0	18.7	20.0
$\overline{PaP}_t^{\text{diff}}$	21.2	20.3	20.1	22.3	20.1	18.7	19.0	22.1	19.2	20.4	20.5	27.7	24.8	25.0	29.7	27.8	19.3	18.2	27.0	25.8	23.1	23.8
$\overline{PaP}_\rho^{\text{diff}}$	21.2	20.0	20.1	22.7	20.2	17.3	15.7	21.1	18.7	19.7	19.7	27.3	24.0	24.6	29.4	27.4	18.9	15.7	24.2	25.2	22.8	23.3
RWB_t	22.4	18.6	21.6	25.3	18.3	17.1	19.1	25.5	18.6	19.5	19.8	25.6	19.0	20.0	28.3	21.3	13.9	16.7	24.2	19.7	16.5	18.0
RWB_ρ	26.1	23.2	25.6	29.7	23.1	18.2	17.9	26.9	22.4	23.5	23.2	30.6	23.2	24.5	32.4	26.5	14.1	10.6	22.3	25.2	19.3	21.1
DWB_t	24.8	21.8	25.5	30.6	19.7	21.0	24.2	30.8	21.4	22.9	23.5	25.7	21.8	21.9	29.5	23.2	15.0	19.0	27.6	20.8	18.7	20.0
DWB_ρ	31.9	27.4	31.3	36.4	25.7	22.3	21.6	34.1	26.0	26.4	27.5	33.2	26.7	27.2	36.0	30.4	16.2	12.4	25.6	27.3	21.9	23.6
AWB_t	23.6	22.9	25.6	29.1	20.7	20.4	24.8	30.0	21.8	23.0	23.2	25.8	21.0	21.9	28.7	22.9	14.9	19.2	26.6	21.7	18.6	19.7
AWB_ρ	29.9	29.1	30.7	34.7	27.2	21.8	22.4	32.9	27.1	26.4	27.8	32.8	26.4	27.5	35.3	31.3	16.3	12.8	25.0	28.5	22.9	23.3
LPB_t^0	12.8	12.2	12.6	12.4	12.4	12.6	11.9	11.9	13.5	13.2	12.2	18.5	18.9	18.8	19.7	19.5	16.4	15.3	18.5	18.4	19.2	18.9
$\overline{LPB}_t^0$	13.7	13.5	14.0	13.2	14.9	14.8	12.7	13.7	14.4	14.4	13.5	21.7	22.0	22.0	23.8	23.8	22.0	17.3	21.9	23.4	22.4	22.7
LPB_t	27.5	24.2	26.3	31.7	22.5	21.4	24.2	30.8	22.4	23.6	24.3	26.9	21.8	22.8	29.8	25.5	16.2	18.8	26.4	22.5	20.6	21.3
$\overline{LPB}_t$	34.6	32.9	34.0	39.8	32.6	28.8	28.7	36.2	30.8	31.7	31.8	35.4	30.4	31.7	39.8	35.5	23.3	22.1	33.1	32.2	28.6	29.6
MZ_α	28.5	25.0	27.4	33.8	27.3	17.4	15.2	27.8	24.1	23.7	23.5	33.0	26.7	25.8	36.0	32.3	17.8	12.5	23.3	28.1	22.8	24.3
$PQMZ_\alpha$	29.9	27.0	29.5	35.6	28.4	19.9	18.2	30.5	26.4	26.4	26.3	33.7	27.9	27.5	37.1	32.6	18.8	13.6	25.3	28.8	24.1	25.7
MZ_α^{iwb}	31.5	26.6	28.9	37.0	28.1	17.1	15.1	29.9	25.2	24.7	24.9	32.9	26.3	25.5	38.2	33.1	14.7	7.2	20.4	28.3	21.1	23.2
MZ_α^{rwb}	26.5	23.7	25.4	30.0	24.8	14.9	12.4	24.7	22.7	22.0	21.8	31.4	25.9	24.4	33.5	30.1	12.5	6.3	17.8	27.1	20.5	20.9
MZ_α^{dwb}	32.2	27.4	30.2	37.7	29.7	18.2	15.8	30.9	26.7	25.6	26.3	35.3	26.4	26.5	37.2	34.8	14.9	7.4	21.7	28.0	22.8	24.1
MZ_α^{awb}	31.9	27.2	29.2	37.0	28.8	17.8	16.1	31.9	27.2	26.2	26.5	35.5	26.9	26.8	37.7	34.5	15.2	8.1	22.1	28.6	22.6	26.3
$PQMZ_\alpha^{\text{iwb}}$	31.5	29.2	31.6	38.8	29.3	19.6	17.2	32.5	27.5	27.5	27.0	34.0	26.9	26.5	38.2	33.1	14.5	8.0	22.7	28.8	22.7	25.4
$PQMZ_\alpha^{\text{rwb}}$	29.0	26.0	28.6	33.1	26.3	17.9	15.0	28.6	24.8	24.8	26.0	33.0	26.7	27.8	36.1	30.0	14.8	6.9	21.9	28.0	21.9	23.9
$PQMZ_\alpha^{\text{dwb}}$	32.9	29.8	32.2	39.6	30.9	19.7	18.2	33.8	28.5	28.3	28.7	35.1	28.5	27.7	39.3	34.2	15.2	8.2	23.2	29.8	23.2	25.3
$PQMZ_\alpha^{\text{awb}}$	32.9	28.2	31.3	38.2	30.5	20.6	18.4	32.5	28.5	27.3	29.2	36.7	28.6	28.6	38.7	35.3	16.8	9.4	24.9	29.9	24.3	24.9

Chapter 7

Empirical study: is there a unit root in the inflation rate?

In this chapter we revisit a question that has enjoyed much attention in the literature in recent years: is the inflation rate stationary or is it integrated of order one? The answer to this question has far-reaching implications, as it affects the validity of many macroeconomic models. Some examples are the Phillips curves of Calvo (1983) and Ball (1993), which are based on the premise that the inflation rate is stationary (also see Dornbusch, 1976; Taylor, 1979). Recent work on the subject of stationarity of the inflation rate includes Holmes (2002), Henry and Shields (2004), Österholm (2004), Charemza, Hristova and Burridge (2005), Cook (2005), Narayan and Popp (2011) and Lee and Tsong (2013).

Empirical results from different authors are rather mixed, but it has been shown that the logarithm of prices is I(2) and consequently that the inflation rate is I(1), i.e. integrated of order one (see e.g. Nelson and Schwert, 1977; Baillie, 1989; Ball, Cecchetti and Gordon, 1990; Johansen, 1992). In contrast, confirming the findings of Culver and Papell (1997), Basher and Westerlund (2008) find strong evidence for the stationarity of inflation by applying an array of *panel* unit root tests.

We applied all bootstrap unit root test procedures considered in the simulation study of the previous chapter to the inflation rate of the G7 countries, Australia and South Africa. Quarterly data for the consumer price indices (CPI) of these countries spanning the period from the first quarter of 1957 until the second quarter of 2017 were obtained from the International Monetary Fund's *International Financial Statistics* database[1]. In agreement with the fact that price levels are I(2), each inflation index series was converted to an annualised quarterly inflation rate series x_t as 400 times the difference of the log-ratio of two consecutive quarters:

$$x_t = 400\Delta \ln C_t = 400 \ln\left(\frac{C_t}{C_{t-1}}\right),$$

where C_t represents the consumer price inflation *index*. Here, ln stands for the natural logarithm. This transformation is common practice in the literature (see e.g. Culver and

[1] Data for Germany and the United Kingdom are only available from 1991 onward and 1988 onward, respectively.

Papell, 1997; Ng and Perron, 2001).

We now move on to the results obtained when applying the considered unit root tests to these real-world data. Note that the configurations of all tests were exactly as specified in Chapter 5 and therefore correspond to the Monte Carlo results of the previous chapter. To be consistent with a long-run positive but non-accelerating inflation, we follow Culver and Papell (1997) and do not include a time trend in the model specification, but only a constant mean component.

The bootstrapped p-values for all series and all tests are given in Table 7.1. Clearly, at the 5% level of significance, the only two series showing any signs of not containing a unit root are the inflation rate series of Australia and Germany. Initially this suggests that most of the considered inflation rate series each contains at least one unit root, i.e. is nonstationary. This is not surprising and is supported by numerous results published in the literature, such as the examples given in the introduction.

From the Monte Carlo results in the previous chapter, it is clear that highly persistent AR processes or processes with strong MA components tend to distort the sizes of these unit root tests. Looking at the estimated AR and MA parameters supplied in Table 7.1, one sees that some of these processes exhibit both a strong AR component and a strong MA component. This prompts a more cautious interpretation of the results. Although we have not investigated the combined effects of a strong AR and a strong MA component, we have seen that, especially for large sample sizes, the bootstrap tests seem to preserve the nominal size well, which gives some comfort in the interpretation of the results.

For completeness, approximate critical values of the tests are provided in Table 7.2.

Future research

Although ample research has already been done to establish the nature of the inflation rate in terms of stationarity, recent advances in the bootstrap for dependent data and specifically its application to unit root testing have not been applied to this problem. These developments may be employed to help address some of the concerns raised by Basher and Westerlund (2008):

1. *Cross-sectional inflation rates are likely not independent of each other and should rather be analysed jointly as panel data.* Recently proposed procedures such as the modified wild bootstrap procedure of Smeekes and Urbain (2014a) can take dependence between multiple series into account. Also see Palm, Smeekes and Urbain (2011), Smeekes (2011) and Smeekes and Urbain (2014b).

2. *Inflation rates are prone to structural change if observed over long time periods and the standard unit root tests do not allow for the possibility of such change.*

3. *Standard unit root tests are not applicable in the presence of unconditional heteroskedasticity.* The bootstrap tests proposed by Cavaliere and Taylor (2008, 2009b)

Table 7.1: Estimated ARMA(1,1) parameters for all series and p-values for the tests.

	Australia	Canada	Germany	France	Italy	Japan	S. Africa	UK	USA
T	246	302	168	302	302	302	302	178	302
AR	0.96	0.98	0.99	0.98	0.97	0.98	-0.38	-0.99	0.97
MA	-0.64	-0.85	-0.92	-0.70	-0.55	-0.84	0.23	0.95	-0.81
DF_t	.199	.521	.005	.718	.349	.360	.361	.146	.499
DF_ρ	.076	.327	.007	.643	.210	.142	.229	.098	.301
DF_t^{PQ}	.199	.521	.014	.718	.525	.360	.532	.187	.499
DF_ρ^{PQ}	.076	.327	.007	.643	.377	.142	.379	.155	.301
ERS_t	.037	.124	.270	.294	.217	.142	.250	.636	.123
ERS_t^{PQ}	.037	.124	.465	.294	.273	.142	.250	.636	.123
CP_t	.188	.529	.026	.749	.514	.313	.514	.132	.483
CP_ρ	.119	.446	.048	.737	.451	.234	.454	.207	.409
$\overline{CP}_t$	.037	.159	.567	.343	.299	.149	.267	.690	.150
$\overline{CP}_\rho$	.035	.144	.569	.459	.340	.130	.255	.699	.137
PS_t	.176	.492	.015	.702	.434	.269	.466	.130	.468
PS_ρ	.120	.345	.034	.664	.390	.191	.394	.214	.308
$\overline{PS}_t$	.043	.127	.526	.333	.260	.134	.250	.701	.126
$\overline{PS}_\rho$	.045	.100	.508	.447	.313	.116	.226	.710	.096
PaP_t	.192	.389	.000	.631	.539	.386	.454	.109	.341
PaP_ρ	.098	.206	.002	.547	.403	.138	.316	.128	.181
$\overline{PaP}_t$	.051	.096	.460	.225	.333	.095	.266	.694	.083
$\overline{PaP}_\rho$	.038	.072	.454	.299	.362	.073	.280	.703	.068
PaP_t^{diff}	.198	.384	.001	.636	.537	.366	.435	.112	.351
PaP_ρ^{diff}	.104	.207	.003	.546	.413	.135	.297	.136	.184
$\overline{PaP}_t^{\text{diff}}$	.047	.096	.463	.216	.336	.089	.265	.714	.084
$\overline{PaP}_\rho^{\text{diff}}$	.035	.073	.455	.295	.368	.065	.278	.723	.071
RWB_t	.102	.447	.017	.765	.704	.404	.351	.118	.434
RWB_ρ	.080	.393	.047	.775	.569	.256	.382	.215	.405
DWB_t	.138	.472	.009	.831	.728	.482	.367	.250	.422
DWB_ρ	.056	.264	.006	.736	.502	.192	.325	.233	.238
AWB_t	.148	.556	.005	.854	.756	.554	.406	.289	.516
AWB_ρ	.061	.298	.003	.749	.498	.189	.334	.253	.272
LPB_t^0	.456	.738	.050	.973	.591	.453	.760	.001	.691
$\overline{LPB}_t^0$	.213	.336	1.000	.789	.625	.383	.682	1.000	.322
LPB_t	.223	.736	.012	.978	.746	.467	.769	.208	.700
$\overline{LPB}_t$	.086	.321	.999	.793	.732	.380	.684	1.000	.334
MZ_α	.053	.100	.100	.100	.100	.100	.100	.100	.100
$PQMZ_\alpha$	.053	.100	.100	.100	.100	.100	.100	.100	.100
MZ_α^{iwb}	.065	.310	.580	.504	.251	.221	.624	.656	.358
MZ_α^{rwb}	.057	.221	.403	.403	.236	.144	.300	.602	.268
MZ_α^{dwb}	.065	.297	.616	.460	.270	.207	.605	.683	.339
MZ_α^{awb}	.060	.289	.583	.448	.253	.204	.606	.677	.329
$PQMZ_\alpha^{\text{iwb}}$	.060	.305	.657	.507	.353	.216	.613	.662	.357
$PQMZ_\alpha^{\text{rwb}}$	.055	.227	.546	.416	.320	.145	.313	.617	.278
$PQMZ_\alpha^{\text{dwb}}$	.067	.289	.689	.469	.343	.218	.598	.680	.340
$PQMZ_\alpha^{\text{awb}}$	.067	.286	.677	.454	.339	.205	.619	.678	.321

Table 7.2: Approximate critical values of the tests obtained for each series.

	Australia	Canada	Germany	France	Italy	Japan	S. Africa	UK	USA
T	246	302	168	302	302	302	302	178	302
DF_t	-2.87	-2.87	-2.88	-2.87	-2.87	-2.87	-2.87	-2.88	-2.87
DF_ρ	-13.84	-13.90	-13.73	-13.90	-13.90	-13.90	-13.90	-13.75	-13.90
DF_t^{PQ}	-2.87	-2.87	-2.88	-2.87	-2.87	-2.87	-2.87	-2.88	-2.87
DF_ρ^{PQ}	-13.84	-13.90	-13.73	-13.90	-13.90	-13.90	-13.90	-13.75	-13.90
ERS_t	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94
ERS_t^{PQ}	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94
CP_t	-2.81	-2.80	-2.84	-2.83	-2.69	-2.69	-2.73	-2.65	-2.92
CP_ρ	-16.01	-18.39	-19.67	-20.65	-16.37	-18.69	-17.84	-17.39	-22.75
$\overline{CP}_t$	-1.95	-2.05	-2.10	-2.02	-1.95	-1.92	-1.93	-1.90	-2.07
$\overline{CP}_\rho$	-9.10	-10.38	-10.89	-10.85	-9.57	-9.87	-8.32	-8.46	-11.42
PS_t	-2.82	-2.73	-2.61	-2.64	-2.55	-2.58	-2.66	-2.63	-2.70
PS_ρ	-17.56	-14.69	-17.24	-15.16	-14.82	-17.26	-15.65	-18.77	-13.89
$\overline{PS}_t$	-1.99	-1.89	-1.93	-1.88	-1.75	-1.85	-1.92	-1.98	-1.90
$\overline{PS}_\rho$	-9.86	-7.92	-8.66	-8.92	-8.11	-8.80	-7.79	-9.23	-8.18
PaP_t	-2.75	-2.36	-2.37	-2.52	-2.65	-2.76	-2.43	-2.55	-2.27
PaP_ρ	-14.81	-10.79	-10.58	-11.91	-13.52	-13.60	-11.99	-12.92	-10.43
$\overline{PaP}_t$	-2.08	-1.74	-1.68	-1.77	-1.96	-1.70	-1.88	-1.89	-1.70
$\overline{PaP}_\rho$	-9.63	-7.00	-6.85	-6.75	-8.79	-6.92	-8.35	-8.52	-7.10
PaP_t^{diff}	-2.79	-2.34	-2.35	-2.57	-2.74	-2.78	-2.40	-2.53	-2.30
PaP_ρ^{diff}	-15.47	-10.35	-10.37	-12.46	-14.38	-13.49	-11.37	-12.88	-10.84
$\overline{PaP}_t^{\text{diff}}$	-2.05	-1.77	-1.71	-1.78	-2.02	-1.65	-1.89	-1.93	-1.74
$\overline{PaP}_\rho^{\text{diff}}$	-9.37	-7.10	-6.99	-6.82	-9.21	-6.63	-8.52	-8.82	-7.26
RWB_t	-2.48	-2.55	-2.59	-2.76	-2.95	-2.73	-2.52	-2.72	-2.73
RWB_ρ	-14.30	-16.96	-19.38	-20.56	-20.69	-19.43	-19.03	-20.42	-22.22
DWB_t	-2.64	-2.50	-2.52	-2.94	-2.99	-3.20	-2.40	-3.44	-2.43
DWB_ρ	-12.64	-12.19	-12.17	-14.23	-16.43	-15.74	-13.24	-16.77	-12.13
AWB_t	-2.69	-2.53	-2.51	-2.90	-3.04	-3.12	-2.42	-3.68	-2.53
AWB_ρ	-12.89	-12.54	-11.86	-13.59	-16.29	-15.47	-13.24	-17.53	-12.78
LPB_t^0	-2.72	-2.76	-2.70	-2.75	-2.78	-2.76	-2.77	-2.74	-2.73
$\overline{LPB}_t^0$	-2.30	-2.29	-2.35	-2.26	-2.31	-2.32	-2.26	-2.33	-2.32
LPB_t	-2.79	-2.76	-2.72	-2.76	-2.79	-2.77	-2.76	-2.78	-2.74
$\overline{LPB}_t$	-2.27	-2.29	-2.35	-2.32	-2.26	-2.31	-2.30	-2.35	-2.31
MZ_α	-8.10	-8.10	-8.10	-8.10	-8.10	-8.10	-8.10	-8.10	-8.10
$PQMZ_\alpha$	-8.10	-8.10	-8.10	-8.10	-8.10	-8.10	-8.10	-8.10	-8.10
MZ_α^{iwb}	-8.82	-10.90	-8.94	-7.70	-9.04	-8.42	-25.18	-8.06	-12.46
MZ_α^{rwb}	-8.48	-7.15	-3.05	-4.74	-8.18	-5.68	-2.86	-2.22	-9.44
MZ_α^{dwb}	-8.80	-10.09	-9.34	-7.39	-9.16	-7.93	-22.23	-8.15	-10.90
MZ_α^{awb}	-8.54	-9.45	-9.25	-7.41	-9.03	-7.79	-20.77	-8.40	-10.58
$PQMZ_\alpha^{\text{iwb}}$	-8.53	-10.82	-8.67	-7.80	-9.17	-8.61	-23.95	-8.13	-12.61
$PQMZ_\alpha^{\text{rwb}}$	-8.27	-7.10	-2.68	-5.28	-8.56	-5.75	-2.96	-2.15	-9.40
$PQMZ_\alpha^{\text{dwb}}$	-8.73	-10.09	-8.79	-7.37	-9.28	-7.88	-20.34	-8.22	-11.39
$PQMZ_\alpha^{\text{awb}}$	-9.19	-9.64	-8.51	-7.18	-9.33	-8.30	-19.65	-8.45	-10.74

are designed to allow for the presence of nonstationary volatility. Cavaliere et al. (2015) also propose a rescaled modified information criterion (RSMIC) which is shown to avoid significant power losses seen when using standard lag selection methods. Other possibilities are given by Smeekes and Taylor (2012), Smeekes and Urbain (2014a) and Westerlund (2014)

These points, however, fall outside the scope of this study and are possible avenues for future research.

Bibliography

Allison, J. S. and Swanepoel, J. W. H. (2010). Three methods to apply the bootstrap correctly to hypothesis testing, *South African Statistical Journal* **44(2)**: 123–143.

An, H., Chen, Z. and Hannan, E. J. (1982). Autocorrelation, autoregression and autoregressive approximation, *Annals of Statistics* **10**: 926–936.

Anderson, O. D. (1976). *Time Series Analysis and Forecasting: The Box-Jenkins Approach*, London: Butterworth & Co.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59**: 817–858.

Asteriou, D. and Hall, S. G. (2006). *Applied Econometrics*, New York: Palgrave Macmillan.

Aylar, E., Smeekes, S. and Westerlund, J. (2017). Lag truncation and the local asymptotic distribution of the ADF test for a unit root, *Statistical Papers* pp. 1–10.

Baillie, R. T. (1989). Commodity prices and aggregate inflation: Would a commodity price rule be worthwhile?, *Carnegie-Rochester Conference Series on Public Policy*, Vol. 31, Elsevier, pp. 185–240.

Ball, L. (1993). The dynamics of high inflation, *Technical report*, National Bureau of Economic Research.

Ball, L., Cecchetti, S. G. and Gordon, R. J. (1990). Inflation and uncertainty at short and long horizons, *Brookings Papers on Economic Activity* **1990**: 215–254.

Basher, S. A. and Westerlund, J. (2008). Is there really a unit root in the inflation rate? more evidence from panel data models, *Applied Economics Letters* **15**: 161–164.

Berk, K. N. (1974). Consistent autoregressive spectral estimates, *Annals of Statistics* **2**: 489–502.

Bhattacharya, R. N., Gupta, V. K. and Waymire, E. C. (1983). The Hurst effect under trends, *Journal of Applied Probability* **20(3)**: 649–662.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap, *Ann. Statist.* **9**: 1196–1217.

Booth, J. G. and Sarkar, S. (1998). Monte carlo approximation of bootstrap variances, *Am. Stat.* **52**: 354–357.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis Forecasting and Control*, 2nd ed., San Franscisco: Holden-Day.

Bühlmann, P. (1995a). The blockwise bootstrap for general empirical processes of stationary sequences, *Stochastic Processes and their Applications* **58**: 247–265.

Bühlmann, P. (1995b). Sieve bootstrap for time series, *Technical Report 431*, Department of Statistics, University of California.

Bühlmann, P. (1997). Sieve bootstrap for time series, *Bernoulli* **3**: 123–148.

Bühlmann, P. (1998). Sieve bootstrap for smoothing in non-stationary time series, *Annals of Statistics* **26**: 48–83.

Bühlmann, P. and Künsch, H. R. (1995). The blockwise bootstrap for general parameters of a stationary time series, *Scand. J. Stat.* **22**: 35–54.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework, *Journal of monetary Economics* **12**: 383–398.

Cao-Abad, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression, *Ann. Statist.* **19**: 2226–2231.

Cao-Abad, R. and González-Manteiga, W. (1993). Bootstrap methods in regression smoothing, *J. Nonparametr. Stat.* **2**: 379–388.

Cappuccio, N. and Lubian, D. (2006). Local asymptotic distributions of stationarity tests, *Journal of Time Series Analysis* **27**: 323–345.

Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series, *Ann. Statist.* **14**: 1171–1179.

Cavaliere, G., Phillips, P. C. B., Smeekes, S. and Taylor, A. M. R. (2015). Lag length selection for unit root tests in the presence of nonstationary volatility, *Econometric Reviews* **34**(4): 512–536.

Cavaliere, G. and Taylor, A. M. R. (2008). Bootstrap unit root tests for time series with nonstationary volatility, *Econometric Theory* **24**: 43–71.

Cavaliere, G. and Taylor, A. M. R. (2009a). Bootstrap M unit root tests, *Econometric Reviews* **28**(5): 393–421.

Cavaliere, G. and Taylor, A. M. R. (2009b). Heteroskedastic time series with a unit root, *Econometric Theory* **25**: 1228–1276.

Chan, K. Y. F. and Lee, S. M. S. (2001). An exact iterated bootstrap algorithm for small-sample bias reduction, *Comput. Stat. Data Anal.* **36**: 1–13.

Chang, J. and Hall, P. (2015). Double-bootstrap methods that use a single double-bootstrap simulation, *Biometrika* **102**: 203–214.

Chang, Y. and Park, J. Y. (2002). On the asymptotics of ADF tests for unit roots, *Econometric Reviews* **21**: 431–447.

Chang, Y. and Park, J. Y. (2003). A sieve bootstrap for the test of a unit root, *Journal of Time Series Analysis* **24**: 379–400.

Charemza, W. W., Hristova, D. and Burridge, P. (2005). Is inflation stationary?, *Applied Economics* **37**: 901–903.

Chernick, M. (1999). *Bootstrap Methods: A Practitioner's Guide*, Wiley.

Cheung, Y.-W. and Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test, *Journal of Business & Economic Statistics* **13**: 277–280.

Cheung, Y.-W. and Lai, K. S. (1997). Bandwidth selection, prewhitening, and the power of the Phillips–Perron test, *Econometric Theory* **13**: 679–691.

Cochrane, J. H. (1991). A critique of the application of unit root tests, *Journal of Economic Dynamics and Control* **15**: 275–284.

Cook, S. (2005). Rank-based unit root testing in the presence of structural change under the null: simulation results and an application to us inflation, *Applied Economics* **37**: 607–617.

Cryer, J. D. and Chan, K. S. (2008). *Time Series Analysis*, 2nd ed., Springer.

Culver, S. E. and Papell, D. H. (1997). Is there a unit root in the inflation rate? evidence from sequential break and panel data models, *Journal of Applied Econometrics* **12**: 435–444.

Davidson, J., Hendry, D., Srba, F. and Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom, *Economic Journal* **88**: 661–692.

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last, *J. Econometrics* **146**: 162–169.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Applications*, Cambridge.

De Angelis, D., Fachin, S. and Young, G. A. (1997). Bootstrapping unit root tests, *Applied Economics* **29**: 1155–1161.

Delgado, M. A. and González-Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap, *Ann. Statist.* **29**: 1469–1507.

Dette, H., Kusi-Appiah, S. and Neumeyer, N. (2002). Testing symmetry in nonparametric regression models, *J. Nonparametr. Stat.* **14**: 477–494.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* **74**: 427–431.

Dornbusch, R. (1976). Expectations and exchange rate dynamics, *Journal of political Economy* **84**: 1161–1176.

Dudek, A. E., Leśkow, J., Paparoditis, E. and Politis, D. N. (2014). A generalized block bootstrap for seasonal time series, *J. Time Ser. Anal.* **35**: 89–114.

Durlauf, S. N. and Phillips, P. C. B. (1988). Trends versus random walks in time series analysis, *Econometrica* **56(6)**: 1333–1354.

Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Ann. Statist.* **7**: 1–26.

Efron, B. (1987). Better bootstrap confidence intervals, *J. Stat. Plan. Infer.* **82**: 171–185.

Efron, B. and Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Stat. Sci.* **1**: 54–77.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall/CRC.

Elliott, G., Rothenberg, T. J. and Stock, J. H. (1996). Efficient tests for an autoregressive unit root, *Econometrica* **64(4)**: 813–836.

Engle, R. F. (1990). Discussion: stock market volatility and the crash of '87, *Review of Financial Studies* **3**: 103–106.

Ferraty, F., van Keilegom, I. and Vieu, P. (2010). On the validity of the bootstrap in nonparametric functional regression, *Scand. J. Stat.* **37**: 286–306.

Flachaire, E. (2005). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap, *Comput. Stat. Data Anal.* **49**: 361–376.

Freedman, D. A. (1981). Bootstrapping regression models, *Ann. Statist.* **9**: 1218–1228.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd ed., New York: John Wiley.

Giacomini, R., Politis, D. N. and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators, *Econometric Theory* **29**: 567–589.

Glosten, L. R., Jaganathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of nominal excess returns on stocks, *Journal of Finance* **48**: 1779–1801.

Goëtze, F. and Künsch, H. R. (1996). Second-order correctness of the blockwise bootstrap for stationary observations, *Ann. Statist.* **24**: 1914–1933.

Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **123**(1): 89–120.

González-Manteiga, W., Miranda, M. D. M. and González, A. P. (2004). The choice of smoothing parameter in nonparametric regression through wild bootstrap, *Comput. Stat. Data Anal.* **47**: 487–515.

Good, P. (2000). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer.

Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics, *Journal of Econometrics* **2**: 111–120.

Hall, A. (1994). Testing for a unit root in time series with pretest data-based model selection, *Journal of Business and Economic Statistics* **12**: 461–470.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

Hall, P., Horowitz, J. L. and Jing, B. Y. (1995). On blocking rules for the bootstrap with dependent data, *Biometrika* **82**: 561–574.

Hall, P. and Martin, M. A. (1988). On bootstrap resampling and iteration, *Biometrika* **75**: 661–671.

Hall, P. and Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing, *Biometrics* **47**: 757–762.

Hannan, E. J. (1987). Rational transfer function approximation, *Statistical Science* **5**: 105–138.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *Ann. Statist.* **21**: 1926–1947.

Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression, *Ann. Statist.* **19**: 778–796.

Harvey, D. I. and Leybourne, S. J. (2005). On testing for unit roots and the initial observation, *The Econometrics Journal* **8**: 97–111.

Harvey, D. I., Leybourne, S. J. and Taylor, A. R. (2009). Unit root testing in practice: dealing with uncertainty over the trend and initial condition, *Econometric theory* **25**: 587–636.

Hendry, D. and Mizon, G. (1978). Serial correlation as convenience simplification, not a nuisance: a comment on a study of the demand for money by the bank of England, *Economic Journal* **88**: 549–563.

Henry, O. T. and Shields, K. (2004). Is there a unit root in inflation?, *Journal of Macroeconomics* **26**: 481–500.

Hinkley, D. V. (1988). Bootstrap methods, *J. R. Statist. Soc. B* **50**: 321–337.

Holmes, M. J. (2002). Panel data evidence on inflation convergence in the european union, *Applied Economics Letters* **9**: 155–158.

Hong, Y., Wang, X. and Wang, S. (2016). Testing strict stationarity with applications to macroeconomic time series. Forthcoming.

Hosseinkouchack, M. and Hassler, U. (2016). Powerful unit root tests free of nuisance parameters, *Journal of Time Series Analysis* **37**: 533–554.

Hušková, M. and Meintanis, S. G. (2007). Omnibus tests for the error distribution in the linear regression model, *Statistics* **41**: 363–376.

Hušková, M. and Meintanis, S. G. (2010). Tests for the error distribution in nonparametric possibly heteroscedastic regression models, *Test* **19**: 92–112.

Hušková, M. and Meintanis, S. G. (2012). Tests for symmetric error distribution in linear and nonparametric regression models, *Commun. Statist. – Simula. Comput.* **41**: 833–851.

Jentsch, C., Politis, D. N. et al. (2015). Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension, *The Annals of Statistics* **43**: 1117–1140.

Jiménez Gamero, M. D. (2014). On the empirical characteristic function process of the residuals in GARCH models and applications, *Test* **23**: 409–432.

Johansen, S. (1992). Testing weak exogeneity and the order of cointegration in uk money demand data, *Journal of Policy modeling* **14**: 313–334.

Kreiss, J.-P. (1988). *Asymptotic Statistical Inference for a Class of Stochastic Processes*, Habilitationsschrift, Universität Hamburg.

Kreiss, J.-P. and Lahiri, S. N. (2012). Bootstrap methods for time series, *in* T. S. Rao, S. S. Rao and C. R. Rao (eds), *Time Series Analysis: Methods and Applications*, Vol. 30 of *Handbook of Statistics*, Elsevier, chapter 1, pp. 3–26.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations, *Annals of Statistics* **17**: 1217–1241.

Kutner, M. H., Nachtsheim, C. J., Neter, J. and Li, W. (2005). *Applied Linear Statistical Models*, New York: McGraw-Hill.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?, *Journal of Econometrics* **54**: 159–178.

Lahiri, S. N. (1996). On Edgeworth expansions and the moving block bootstrap for studentized M-estimators in multiple linear regression models, *J. Multivar. Anal.* **56**: 42–59.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods, *Ann. Statist.* **27**: 386–404.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*, Springer.

Lahiri, S. N., Furukwa, K. and Lee, Y. D. (2007). A nonparametric plug-in rule for selecting the optimal block length for block bootstrap methods, *Statistical Methodology* **4**: 292–321.

Lee, C.-F. and Tsong, C.-C. (2013). Bootstrapping covariate unit root tests: an application to inflation rates, *Bulletin of Economic Research* **65**.

Lee, J. and Mossi, D. (1996). On improvements of Phillips–Perron unit root tests using optimal bandwidth estimates, *Applied Economics Letters* **3**: 197–200.

Leybourne, S. J. (1995). Testing for unit roots using forward and reverse dickey-fuller regressions, *Oxford Bulletin of Economics and Statistics* **57**: 559–571.

Leybourne, S. J., McCabe, B. P. and Tremayne, A. R. (1996). Can economic time series be differenced to stationarity?, *Journal of Business & Economic Statistics* **14**: 435–446.

Lima, L. R. and Xiao, Z. (2010). Is there long memory in financial time series?, *Appl. Finan. Econ.* **20**: 487–500.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models, *Ann. Statist.* **16**: 1696–1708.

Liu, R. Y. and Singh, K. (1992). Moving block jackknife and bootstrap capture weak dependence, *in* R. Lepage and L. Billard (eds), *Exploring the Limits of the Bootstrap*, Wiley, pp. 225–248.

Lloyd, C. J. (2005). Estimating test power adjusted for size, *Journal of Statistical Computation and Simulation* **75**: 921–933.

MacKinnon, J. G. (2006). Bootstrap methods in econometrics, *Econ. Rec.* **82**: S2–S18.

MacKinnon, J. G. (2009). Bootstrap hypothesis testing, *in* D. A. Belsley and E. J. Kontoghiorghes (eds), *Handbook of Computational Econometrics*, Wiley, chapter 6, pp. 183–214.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models, *Ann. Statist.* **21**: 255–285.

Martínez-Camblor, P. and Corral, N. (2012). A general bootstrap algorithm for hypothesis testing, *J. Stat. Plan. Infer.* **142**: 589–600.

McMurry, T. L. and Politis, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap, *Journal of Time Series Analysis* **31**: 471–482.

Naik-Nimbalkar, U. V. and Rajarshi, M. B. (1994). Validity of blockwise bootstrap for empirical processes with stationary observations, *Ann. Statist.* **22**: 980–994.

Nankervis, J. C. and Savin, N. (1996). The level and power of the bootstrap t test in the ar (1) model with trend, *Journal of Business & Economic Statistics* **14**: 161–168.

Narayan, P. K. and Popp, S. (2011). An application of a new seasonal unit root test to inflation, *International Review of Economics & Finance* **20**: 707–716.

Nelson, C. R. and Plosser, C. R. (1982). Trends and random walks in macroeconomic time series: some evidence and implications, *Journal of Monetary Economics* **10(2)**: 139–162.

Nelson, C. R. and Schwert, G. W. (1977). Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest is constant, *The American Economic Review* **67**: 478–486.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach, *Econometrica* **59**: 347–369.

Neumeyer, N. and Dette, H. (2007). Testing for symmetric error distribution in nonparametric regression models, *Statist. Sinica* **17**: 775–795.

Neumeyer, N., Dette, H. and Nagel, E.-R. (2005). A note on testing symmetry of the error distribution in linear regression models, *J. Nonparametr. Stat.* **17**: 697–715.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**: 703–708.

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* **61**: 631–653.

Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association* **90**: 268–281.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power, *Econometrica* **69**(6): 1519–1554.

Ng, S. and Perron, P. (2005). A note on the selection of time series models, *Oxford Bulletin of Economics and Statistics* **67**: 115–134.

Nordman, D. J. (2009). A note on the stationary bootstrap's variance, *Ann. Statist.* **37**: 359–370.

Österholm, P. (2004). Killing four unit root birds in the us economy with three panel unit root test stones, *Applied Economics Letters* **11**(4): 213–216.

Palm, F. C., Smeekes, S. and Urbain, J.-P. (2008). Bootstrap unit root tests: comparison and extensions, *Journal of Time Series Analysis* **29(2)**: 371–401.

Palm, F. C., Smeekes, S. and Urbain, J.-P. (2011). Cross-sectional dependence robust block bootstrap panel unit root tests, *Journal of Econometrics* **163**(1): 85–104.

Pantula, S. G., Gonzalez-Farias, G. and Fuller, W. A. (1994). A comparison of unit-root test criteria, *Journal of Business & Economic Statistics* **12**: 449–459.

Paparoditis, E. and Politis, D. (2001). Tapered block bootstrap, *Biometrika* **88**: 1105–1119.

Paparoditis, E. and Politis, D. N. (2003). Residual-based block bootstrap for unit root testing, *Econometrica* **71**: 813–855.

Paparoditis, E. and Politis, D. N. (2005a). Bootstrap hypothesis testing in regression models, *Stat. Probabil. Lett.* **74**: 356–365.

Paparoditis, E. and Politis, D. N. (2005b). Bootstrapping unit root tests for autoregressive time series, *Journal of the American Statistical Association* **100**: 545–553.

Paparoditis, E. and Politis, D. N. (2016). The asymptotic size and power of the augmented Dickey–Fuller test for a unit root, *Econometric Reviews* pp. 1–19.

Park, J. Y. (2003). Bootstrap unit root tests, *Econometrica* **71**: 1845–1895.

Patterson, K. (2011). *Unit Root Tests in Time Series: Volume 1*, New York: Palgrave Macmillan.

Patton, A., Politis, D. and White, H. (2009). Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White, *Econometric Reviews* **28**: 372–375.

Pearson, K. (1897). On a form of spurious correlation which may arise when indices are used in the measurement of organs, *Proceedings of the Royal Society of London* **60**: 489–498.

Perron, P. and Ng, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties, *The Review of Economic Studies* **63**(3): 435–463.

Perron, P. and Qu, Z. (2007). A simple modification to improve the finite sample properties of Ng and Perron's unit root tests, *Economics letters* **94**: 12–19.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics, *Journal of Econometrics* **33(3)**: 311–340.

Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica* **55**: 227–301.

Phillips, P. C. B. and Perron, P. (1988). Testing for unit roots in time series regression, *Biometrika* **75**: 335–346.

Phillips, P. C. B. and Xiao, Z. (1998). A primer on unit root testing, *Journal of Economic Surveys* **12(5)**: 423–469.

Plosser, C. I. and Schwert, G. W. (1978). Money, income, and sunspots: measuring economic relationships and the effects of differencing, *Journal of Monetary Economics* **4(4)**: 637–660.

Politis, D. N. and Romano, J. P. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation, *Annals of Statistics* **20**: 1985–2007.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap, *Am. Stat.* **89**: 1303–1313.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*, Springer.

Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap, *Econometric Reviews* **23**: 53–70.

Politis, D. and Romano, J. (1995). Bias-corrected nonparametric spectral estimation, *J. Time Ser. Anal.* **16**: 67–103.

Psaradakis, Z. (2001). Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors, *Journal of Time Series Analysis* **22**: 577–594.

Racine, J. S., Hart, J. and Li, Q. (2006). Testing the significance of categorical predictor variables in nonparametric regression models, *Econometric Reviews* **25**: 523–544.

Radulović, D. (1996a). The bootstrap for empirical processes based on stationary observations, *Stochastic Processes and their Applications* **65**: 259–279.

Radulović, D. (1996b). The bootstrap of the mean for strong mixing sequences under minimal conditions, *Stat. Probabil. Lett.* **28**: 65–72.

Radulović, D. (2009). Another look at the disjoint blocks bootstrap, *Test* **18**: 195–212.

Radulović, D. (2012). Necessary and sufficient conditions for the moving blocks bootstrap central limit theorem of the mean, *J. Nonparametr. Stat.* **24**: 343–357.

Raña, P., Aneiros, G., Vilar, J. and Vieu, P. (2016). Bootstrap confidence intervals in functional nonparametric regression under dependence, *Electron. J. Statist.* **10**: 1973–1999.

Rayner, R. K. (1990). Bootstrapping p values and power in the first-order autoregression: A monte carlo investigation, *Journal of Business & Economic Statistics* **8**: 251–263.

Rice, J. A. (2007). *Mathematical Statistics and Data Analysis*, 3rd ed., Belmont: Brooks/Cole.

Richard, P. (2009). Modified fast double sieve bootstraps for ADF tests, *Computational Statistics & Data Analysis* **53**: 4490–4499.

Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap, *Ann. Statist.* **40**: 2798–2822.

Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive moving average models with unknown order, *Biometrika* **71**: 599–607.

Schmidt, P. and Phillips, P. C. (1992). Lm tests for a unit root in the presence of deterministic trends, *Oxford Bulletin of Economics and Statistics* **54**: 257–287.

Schwert, G. W. (1989). Test for unit roots: A monte carlo investigation, *Journal of Business and Economic Statistics* **7**: 147–159.

Seber, G. A. and Lee, A. J. (2012). *Linear regression analysis*, Vol. 936, John Wiley & Sons.

Seo, M. (2005). Improving unit root tests by a generalization of the autoregressive spectral density estimator at frequency zero, *Unpublished Manuscript, Department of Economics, London School of Economics* .

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer.

Shao, X. (2010). The dependent wild bootstrap, *Journal of the American Statistical Association* **105**: 218–235.

Shao, X. (2011). A bootstrap-assisted spectral test of white noise under unknown dependence, *Journal of Econometrics* **162**: 213–224.

Shibata, R. (1980). An optimal autoregressive spectral estimate, *Annals of Statistics* **9**: 300–306.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap, *Ann. Statist.* **9**: 1187–1195.

Smeekes, S. (2011). Bootstrap sequential tests to determine the stationary units in a panel.

Smeekes, S. (2013). Detrending bootstrap unit root tests, *Econometric Reviews* **32**: 869–891.

Smeekes, S. and Taylor, A. M. R. (2012). Bootstrap union tests for unit roots in the presence of nonstationary volatility, *Econometric Theory* **28**: 422–456.

Smeekes, S. and Urbain, J.-P. (2014a). A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing, *GSBE Research Memorandum RM/14/008, Maastricht University* .

Smeekes, S. and Urbain, J.-P. (2014b). On the applicability of the sieve bootstrap in time series panels, *Oxford Bulletin of Economics and Statistics* **76**: 139–151.

Stute, W., González-Manteiga, W. and Quindimil, M. P. (1998). Bootstrap approximations in model checks for regression, *J. Am. Stat. Assoc.* **93**: 141–149.

Taylor, J. B. (1979). Staggered wage setting in a macro model, *The American Economic Review* **69**(2): 108–113.

Ventosa-Santaulària, D. (2009). Spurious regression, *Journal of Probability and Statistics* **2009**: 1–27.

Wei, W. W. S. (2006). *Time Series Analysis*, Pearson.

Westerlund, J. (2014). On the asymptotic distribution of the Dickey Fuller-gls test statistic, *Statistics* **48**: 1233–1253.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis, *Ann. Statist.* **14**: 1261–1295.

Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes, *Statistica Sinica* pp. 1755–1768.

Xiao, Z. and Phillips, P. C. (1998). An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the us economy, *The Econometrics Journal* **1**: 27–43.

Yule, G. (1897). On the theory of correlation, *Journal of the Royal Statistical Society* **60(4)**: 812–854.

Zivot, E. and Wang, J. (2006). *Modelling Financial Time Series with S-PLUS®*, Springer Science & Business Media.

Zou, N. and Politis, D. (2016). Linear process bootstrap unit root test, *arXiv preprint arXiv:1610.06279* .