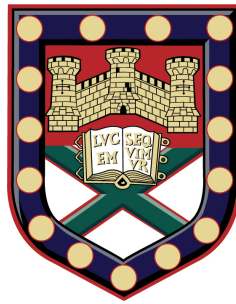# Bootstrap Methods for Heavy-Tail or Autocorrelated Distributions with an Empirical Application

Rami Chehab

Department of Economics
University of Exeter

Submitted for the degree of *Doctor of Philosophy in Economics*

August 2017

I would like to dedicate this thesis to my loving family ...

Declaration

Thesis Title: Bootstrap Methods for Heavy-Tail or Autocorrelated Distributions with an Empirical Application

Submitted by Rami Chehab to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Economics, August 2017

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

(Signature) _____

Rami Chehab

February 2018

Acknowledgements

Abstract

Chapter One: *The Truncated Wild Bootstrap for the Asymmetric Infinite Variance Case*
* Advisers *Dr Adriana Cornea-Madeira* and *Professor James Davidson*

The wild bootstrap method proposed by Cavaliere et al. (2013) to perform hypothesis testing for the location parameter in the location model, with errors in the domain of attraction of asymmetric stable law, is inappropriate. Hence, we are introducing a new bootstrap test procedure that overcomes the failure of Efron's (1979) resampling bootstrap. This bootstrap test exploits the Wild Bootstrap of Cavaliere et al. (2013) and the central limit theorem of trimmed variables of Berkes et al. (2012) to deliver confidence sets with correct asymptotic coverage probabilities for asymmetric heavy-tailed data. The methodology of this bootstrap method entails locating cut-off values such that all data between these two values satisfy the central limit theorem conditions. Therefore, the proposed bootstrap will be termed the Truncated Wild Bootstrap (TWB) since it takes advantage of both findings.

Simulation evidence to assess the quality of inference of available bootstrap tests for this particular model reveals that, on most occasions, the TWB performs better than the Parametric bootstrap (PB) of Cornea-Madeira & Davidson (2015). In addition, TWB test scheme is superior to the PB because this procedure can test the location parameter when the index of stability is below one, whereas the PB has no power in such a case. Moreover, the TWB is also superior to the PB when the tail index is close to 1 and the distribution is heavily skewed, unless the tail index is exactly 1 and the scale parameter is very high.

Chapter Two: *A frequency domain wild bootstrap for dependent data*
* Co-authored and jointly written with *Professor James Davidson*[1]

In this chapter a resampling method is proposed for a stationary dependent time series, based on Rademacher wild bootstrap draws from the Fourier transform of the data. The main distinguishing feature of our method is that the bootstrap draws share their periodogram identically with the sample, implying sound properties under dependence of arbitrary form.

---

[1]Professor Davidson has also done most of the coding using Ox programming language and his TSM software.

A drawback of the basic procedure is that the bootstrap distribution of the mean is degenerate. We show that a simple Gaussian augmentation overcomes this difficulty. Monte Carlo evidence indicates a favourable comparison with alternative methods in tests of location and significance in a regression model with autocorrelated shocks, and also of unit roots.

Chapter Three: *Frequency-based Bootstrap Methods for DC Pension Plan Strategy Evaluation*
* Advisers *Professor Richard D F Harris* and *Professor James Davidson*

The use of conventional bootstrap methods, such as Standard Bootstrap and Moving Block Bootstrap, to produce long run returns to rank one strategy over the others based on its associated reward and risk, might be misleading. Therefore, in this chapter, we will use a simple pension model that is mainly concerned with long-term accumulation wealth to assess, for the first time in pension literature, different bootstrap methods. We find that the Multivariate Fourier Bootstrap gives the most satisfactory result in its ability to mimic the true distribution using Cramér-von-mises statistics. We also address the disagreement in the pension literature on selecting the best pension plan strategy. We present a comprehensive study to compare different strategies using a different bootstrap procedures with different Cash-flow performance measures across a range of countries. We find that bootstrap methods play a critical role in determining the optimal strategy. Additionally, different CFP measures rank pension plans differently across countries and bootstrap methods.

Contents

List of Figures

List of Tables

Nomenclature

$\mathbb{P}(A)$ Probability of the random phenomenon A

$\mathbb{E}(A)$ The expectation of the random variable A

$\mathcal{N}(\mu, \sigma^2)$ Normal Distribution with mean μ and variance σ^2

$\overset{d}{=}$ Equal in distribution

$\overset{d}{\to}$ Convergence in distribution

$\overset{p}{\to}$ Convergence in probability

$\to$ Asymptotic Arrow

$\overset{w}{\to}$ weak convergence

$\mathbb{N}$ The (positive) natural numbers

$\mathbb{S}(\alpha)$ Stable random variable with index α

$S(\alpha, \beta, 1, 0)$ Stable random variable with index α and skewness parameter β

$\mathbb{R}$ The real numbers

$\hat{A}$ The estimator of the random variable A

$\mathbb{L}$ Class L of all infinitely divisible distributions

$\text{Re}(A)$ The real part of complex number A

$\text{Im}(A)$ The Imaginary part of complex number A

Chapter 1

The Truncated Wild Bootstrap for the Asymmetric Infinite Variance Case

1.1 Introduction

Simple linear regression models are often employed to study economic relationships, hence the connection between the dependent and independent variables. Therefore, the coefficients of such linear models are of interest, since if those regression coefficients are zero, then there exists no relation between both variables of interest. In Econometrics and Statistics, applied researchers are usually interested in constructing hypotheses tests to determine whether such a link exists or not.

It is often the case when utilizing a statistical software package, that practitioners fail to observe critical assumptions in their statistical analysis. In particular, some econometric tests require strong auxiliary assumptions on the distribution of the data generating process (DGP) of the regression model. For instance, the error terms in a regression model may need to be independently and identically distributed (i.i.d.) (or white noise) with finite variances (Homoscedastic). These assumptions are important because they enable practitioners to construct a finite sample *test statistic*[1] to test coefficients in linear models which are believed, under such strong auxiliary assumptions, to have known distribution under the true null hypothesis[2].

---

[1] A random variable that has a known distribution when the null hypothesis is true and some other distribution when the null hypothesis is false. It measures discrepancy between the data and the null hypothesis

[2] A Hypothesis, null or alternative, can always be represented by a set of DGPs. Null hypothesis refer to the hypothesis of interest for testing in case the observed data show significance to likely follow the distribution of a specific DGP.

Hypothesis testing is related to constructing confidence sets. Actually, confidence sets require finding a rejection region such that if the test statistic belongs there, the null hypothesis is rejected (see, Section 4.5, Shao & Tu 1995, p. 129). One approach to finding this rejection region is based on the *test statistic*, hence the approach's name "test statistic" (see, Section 4.5, Shao & Tu 1995, p. 176).

A *test* is a combination of a test statistic and a rejection rule. In case the distribution of a test statistic is exactly known, the test is called an *exact test*. In such a construction, if the test incorrectly leads to the rejection of the null hypothesis that is true, then, a Type I error is committed. The probability under the null hypothesis that the observed sample falls into the rejection region is called the *level of significance*, or, simply, the *level* of the test (Davidson & MacKinnon 2004, p. 126).

In general, the test statistic relies on random variables that may not follow an exactly known distribution. In particular, in finite samples, the distribution of the test statistic is usually unknown. However, as the sample size grows, the distribution of the random test statistic converges, under mild assumptions, to a known distribution such as the t and F distribution (see, Davidson & MacKinnon 2004, p. 147). Thus, applied researchers, in their hypothesis testing, rely on critical values of the constructed test statistic to decide whether to reject or not the null hypothesis.

In this regard, one needs to draw a distinction between a *nominal level* and an *actual rejection probability* of a test in hypothesis testing. The nominal level of the test is the probability of making a Type I error according to whatever approximate distribution of a test statistic a researcher is using to determine the rejection region. Actual rejection probability is the true rejection of the null hypothesis when it is valid. Usually, in practice, the actual rejection probability is unknown due to the uncertainty of the DGP of the observed data. Therefore, the actual rejection probability and the nominal level may differ greatly from one another.

The usage of the finite sample of the collected data has serious consequences. Notably, there may be serious approximation errors (Godfrey 2009, p. XI). Consider, for example, the well-known test statistical choices used for asymptotic hypothesis testing: the t and F statistics. Both statistics may depend on a regression model's error distribution in finite samples. Hence, such test statistics are not *pivotal test statistics*[3] in finite sample but in an infinite sample, ie. asymptotically. Hence, both test statistics are *asymptotically pivotal*. Therefore, some empirical researchers, more often than not, rely on asymptotic rather than finite sample theory to carry out tests on coefficients in their regression models.

---

[3]A test statistic, which is a random variable, is called pivotal or a pivot if its distribution is the same for any DGPs of its input random variables in a given model (see, Davison & Hinkley 1997, p. 138-139)

Asymptotic theory sometimes provides a poor approximation of the actual distribution of the test statistic, however. Such tests may fail when classical assumptions are further relaxed (Godfrey 2009, p. 2). Moreover, statisticians may have no control over the real nominal level of the asymptotic level test on the null hypothesis for a fixed sample size (see, Janssen & Pauls 2003, p. 769). Therefore, resampling methods for the testing hypothesis are often preferred alternatives to asymptotic tests (Shao & Tu 1995, p. 130).

The bootstrap method is one of the most popular resampling methods and proves to be useful in the context of hypothesis testing (see, Beran 1988, p. 691). In fact, the bootstrap test provides a better finite sample approximation than the asymptotic test (Godfrey 2009, p. XI). In its simplest version, the bootstrap is a method that re-samples randomly, with replacement, from an observed sample to obtain a *bootstrap sample* (Efron 1979). Based on the obtained bootstrap samples, one may compute for each bootstrap sample a test statistic called the *bootstrap test statistic*. The generation of several bootstrap test statistics enables the practitioner to compute *bootstrap critical value(s)*[4]. The *Bootstrap test* may be easily performed once the rejection region is determined by the bootstrap critical values. Such a process to test hypothesis is known as *bootstrap hypothesis test* (Shao & Tu 1995, p. 179).

Mammen (1992) and Janssen & Pauls (2003) have shown that in the case that the random variables used in the computation of the test statistic are i.i.d., then the bootstrap works if central limit theorem holds for the test statistic. In general, for the bootstrap to provide correct coverage probabilities, it critically depends on the asymptotic behaviour of the conditional distributions of the test statistic. Therefore, it is important, before utilizing the bootstrap, to test if the observed data follow a practitioner's set of assumptions, particularly if the observations belong to the set of DGP that satisfies the central limit theorem conditions. For further discussions on the theory of bootstrap hypothesis testing and the usage of bootstrap in hypothesis testing see, Bickel & Ren (2001), Hinkley (1988, 1989), Mammen (1992), and Janssen & Pauls (2003) among others and the references therein.

There exist cases where Efron's (1979) bootstrap (also known as the *usual bootstrap* or *standard bootstrap*) fails to mimic the distribution of the original DGP of the observations (Efron 1979). For example, it is generally known that the bootstrap does not work when the observations are originally generated from time-dependent (Singh 1981), heteroscedastic (Wu 1986), or heavy-tailed (Athreya 1987, Knight 1989) distributions. Moreover, it turns out that such properties of DGPs, in which Efron's bootstrap fails, are indispensable assumptions that need to be made about the observations.

---

[4]A bootstrap critical value are critical values computed based on bootstrap approach rather than asymptotic theory approach. This usually entails computational quantile calculation.

According to many empirical results in the literature, especially in physical and social sciences, it is hard to reject that observations do not originate from time-dependent, heteroscedastic, or heavy-tailed series (see, for example, Fama 1965, Mandelbrot 1997, Ibragimov 2014). In particular, Ibragimov et al. (2013) found that exchange rates, in emerging and developing countries, are subject to more external shocks than that of the developed markets. For example, discussions in Ibragimov et al. (2015) mention that financial returns and exchange rates in developed markets typically follow power-law distributions with tail indices that lie in the interval $(2, 4)$ and, therefore, have finite variances but their fourth moments are infinite. In addition, infinite variances and infinite first moments with asymmetric tails are exhibited by important insurance and financial risks and economic and financial variables such as economic losses from natural disasters, catastrophe risks, operational risks and, according to preliminary empirical results, returns of cryptocurrencies (see, Ibragimov et al. (2009); Ibragimov et al. (2015) and references and discussion therein). Therefore, as a result, a number of theoretical statisticians and econometricians start modifying the original bootstrap method to enable it to work for a different spectrum of DGPs.

1.1.1   Bootstrap Methods in the context of Heavy-tails Phenomenon

Athreya (1987) was the first to show that the usual bootstrap simulation-based test is not valid when the DGP of the observation belongs to a sub-class of infinitely divisible distributions (including the infinite variance case). In his argument, Athreya (1987) has shown that as the sample size increases the bootstrap distribution of the sample mean does not converge to a deterministic distribution. In fact, Athreya (1987) had proved that the bootstrap measure of the bootstrap characteristic function as in the Lévy-Khintchine representation (see, Definition A.2.3) converges to a Poisson measure and not the Lévy measure outlined in the Lévy-Khintchine representation (see, Theorem 17.1.4, Athreya & Lahiri 2006, p. 541).

In their proof of the same result, Knight (1989) and Kinateder (1992) have utilized the fact that the bootstrap version of the statistic could be expressed by a dot product of two vectors: a random and deterministic vector. The random vector consists of multinomial distributed entries while the deterministic vector consists of the deterministic observed series. It is well-known that the limiting distribution of the multinomial random vector of size n converges, as the sample size n grows, to an infinite vector of independent Poisson random variables with mean one. Therefore, the limiting distribution of the bootstrap version of the normalized sum of i.i.d. series, S_n (see equation (A.25)), is different from that of the real distribution of the statistic S_n which was represented by LePage et al. (1981) (see Theorem

A.2.11). Thus, the bootstrap fails to mimic the distribution of observations generated from distributions that belong to the domain of attraction of a stable law.

Many bootstrap algorithm modifications have been introduced to test a composite null hypothesis[5] consistently. For example, Arcones & Giné (1989) and Athreya (1987) proposed the *m* out of *n* bootstrap as a remedy to the conventional bootstrap, which yields a consistent limiting distribution of the bootstrap mean. Similarly, Bertail et al. (1999) have introduced the subsampling, which is similar to the *m* out of *n* bootstrap, based on resampling a smaller sample from a bigger sample. However, different from the *m* out of *n* bootstrap, the selection occurs without resampling.

Although the *m* out of *n* bootstrap (also known as the *low intensity bootstrap*) has some advantages over the usual bootstrap when it comes to generating bootstrap confidence interval for the sample mean (Athreya 1987), such a bootstrap remedy does not solve the problem without further assumptions on the DGPs of the series of interest. In general, the bootstrap has to be applied with care (Bickel & Sakov 2008). For stable limit laws, for instance, the low intensity bootstrap can be consistent but the size *m* may depend on the index of stability (del Barrio et al. 2013). Nonetheless, regardless of the reduced resampling size *m*, under mild conditions the conditional limit laws for the *m* out of *n* bootstrap are infinitely divisible[6] (Janssen & Pauls 2003).

Recently, Cornea-Madeira & Davidson (2015) introduced the *Parametric bootstrap* (PB) that is capable of testing, under mild assumptions, the hypothesis for the mean for heavy-tailed distributions. Moreover, they show that their new bootstrap modification gives better results than that of the *m* out of *n* bootstrap. Furthermore, they discovered, through Monte Carlo testing, that the *m* out of *n* bootstrap does not work well in finite samples. Additionally, subsampling does not always provide reliable inferences if the sample size is not very large.

Also lately, Cavaliere et al. (2013) have discovered that the Liu (1988)'s Wild Bootstrap (WB) procedure may be used as a new bootstrap method for a different purpose than what was initially introduced. Indeed, Cavaliere et al. (2013) showed that the WB may also be utilized to mimic the distribution of the sample mean for a set of observations whose distribution belongs to the domain of attraction of a symmetric stable law. Although their bootstrap modification works for the testing hypothesis, they focused on showing that their bootstrap version delivers correct coverage probabilities that have narrower asymptotic confidence than the standard bootstrap.

Despite the fact that the PB and the WB of the mean are consistent algorithms for testing the location parameter in a location model with infinite variance (α-stable) innovations, both

---

[5]A composite null hypothesis leaves some parameters unknown; therefore, does not completely specify the distribution of the DGP when the null hypothesis is true (Davison & Hinkley 1997, p. 136-137)

[6]Section A.2 gives an overview of infinitely divisible distributions and their relation with stable laws

algorithms have drawbacks. For example, the PB is incapable of testing for the location when the index of stability α is below unity. However, the WB is only applicable when the observations belong to the domain of attraction of a symmetric stable law.

It should be emphasized that when introducing new bootstrap algorithms for hypothesis testing, econometric theorists and statisticians often use Monte Carlo (MC) experiments to evaluate the proposed bootstrap method. MC simulations are ideal for evaluating a bootstrap method because the observations are being artificially generated from a predefined density function. In this case, the DGP and all other features of the data are known. Hence, under such conditions, it is easy to approximate, using numerical algorithms, the actual rejection probability of a bootstrap test under the null. If the simulated bootstrap algorithm results in a bootstrap P value (*actual P values*[7]) such that they are close or below the nominal level when the null hypothesis is true, then the bootstrap method is consistent.

Having said that, MC simulations for bootstrap method evaluation act as a powerful tool to compare different bootstrap methods (Giacomini et al. 2013). For example, Cornea-Madeira & Davidson (2015) used MC simulations to show that their proposed bootstrap method outperforms other available bootstrap methods when the null hypothesis is true. In particular, Cornea-Madeira & Davidson (2015) showed that the PB has the lowest error in rejection probability (ERP[8]) in comparison with the *m* out of *n* bootstrap and subsampling when the null hypothesis is true. Moreover, Cornea-Madeira & Davidson (2015) pointed that it is not enough to study only low ERP. The power of the bootstrap test should also be considered.

We should recall that the power of a test is generally linked to the probability of rejecting the null. Therefore, if the alternative hypothesis is true, and the power is not high (i.e. high probability of rejecting the null), then the bootstrap proposed is not fit to be utilized even if the ERP is low under the null. Therefore, Cornea-Madeira & Davidson (2015) also showed that the PB is the most powerful among other bootstrap algorithms when the DGP is generated under the alternative hypothesis.

Davidson (2008) had noted that in MC simulation experiments, in a variety of circumstances, the bootstrap tests perform better than asymptotic theory predicts. In particular, he mentioned that the *bootstrap discrepancy*[9], under the null, appears to be smaller than what was previously expected by the Edgeworth expansion theory.

---

[7]The P value (the marginal significance level for a test statistic) is the greatest level for which a test based on a value of the test statistics fails to reject the null. It is the significance probability that measures the level of evidence against the null hypothesis (see, Davison & Hinkley 1997, p. 136-137)

[8]The Error in Rejection Probability (ERP) is the difference between the nominal (asymptotically achieved) significance level and the actual significance level (Godfrey 2009, p. 70)

[9]Bootstrap discrepancy is the differences between the actual rejection probability and the nominal level

In this chapter, we will introduce the *Truncated Wild bootstrap* (TWB) algorithm that is capable of overcoming the aforementioned shortcomings of the previous PB and WB methods. It will be shown, using MC simulations, that the TWB algorithm is capable of testing for the null hypothesis for the location parameter when the index of stability is below one, or for asymmetric distributions that belong to the domain of attraction of a stable law.

1.1.2 Organization of Chapter 1

Chapter 1 includes three sections; each section plays a role in justifying the results in the simulation section in chronological order. Section 1.2 presents the location model which we are interested in, making inferences upon its parameter. Furthermore, this section develops a framework for the newly devised bootstrap to consistently test a hypothesis on the concerned parameter. This bootstrap framework depends profoundly on Theorem 1.2.1 where its proof is based on earlier literature summarized in Appendix section A.2[10].

In Section 1.3, we obtain simulation results based on MC procedure for comparing different available bootstrap tests. The results show that the newly devised bootstrap test is superior to available bootstrap tests in the literature. Section 1.4 proposes possible extensions of the results presented in Section 1.3 in both domains practical and theoretical econometrics, followed by a conclusion to this chapter.

1.2 Main Results

All distributions that belong to the domain of attraction of a stable law converge, when suitably scaled and normalized, converge, as the sample size grows, to a stably distributed random variable (see, Section A.2). Therefore, asymptotically, even if the distribution of the i.i.d. summand, say X, is non-unimodal, such that $X \in \mathscr{D}\mathscr{A}(\alpha)$, its sum, when suitably scaled and normalized, will converge to a unimodal distribution (see, Remarks of Lemma A.2.9). Hence, in the limit, since stable distributions are continuous and differentiable, the axis of symmetry of the distribution over a particular small interval would be the mode of such a distribution.

---

[10]In general, Appendix A presents the mathematical background and an in-depth analysis of properties of Stable distributions. In particular, Section A.2 shed lights on the relationship between the family of infinitely divisible distributions and the family of stable distributions. In addition, this section focuses on restating previous theoretical findings on the asymptotic behaviour of the central part of stable and infinitely divisible distributions.

Janssen (1989) proved that the central part of the sum, $\Sigma_n(r_n, s_n)$, defined as

$$\Sigma_n(r_n, s_n) = \sum_{i=r_n+1}^{n-s_n} X_{i,n}{}^{11}$$

when suitably selected and normalized converges in probability to zero (see, Lemma A.2.12). His selection of the central part is based on the magnitude of the observed series. Moreover, the number of summands in this central part of the sum grows as the sample size n grows. Therefore, this growth in the number of summands as the sample size grows is due to the fact that the majority of the realized summands are observed around the mode of the asymptotic distribution, the value that appears most often in the dataset.

Additionally, Berkes et al. (2012) and Csörgő et al. (1986) have proven that there exist some normalizing and centring constants, such that the middle part of the sum converges to a normal distribution as n approaches infinity (see, Section A.2). Therefore, combining the two findings, it is possible to conclude that the centring constant represents an axis that centres the distribution such that the density is convex below this axis and concave after it. The only property that follows such intuition is that this axis of symmetry would be the mode of the unimodal stable distributions.

Consider the *location model* defined as

$$X_i = \theta + \varepsilon_i \tag{1.1}$$

where the sequence $\{\varepsilon_i\}_{i=1}^n$ of error terms is assumed to be i.i.d. such that $\varepsilon_1 \in \mathscr{DA}(\alpha)$ such that $\alpha \in (0,2)$. It is well-known that if $X \in \mathscr{DA}(\alpha)$ then

$$S_n := \frac{1}{a_n} \sum_{i=1}^n (X_i - \theta) \xrightarrow{d} Y \qquad \text{and} \qquad \mathbb{P}(|X| \geq x) = 1 - F_{|X|}(x) = x^{-\alpha} L(x) \tag{1.2}$$

where $Y \in \mathbb{S}(\alpha)$ as defined in A.2.2, where α is the parameter of stability, $L(x)$ is a slow varying function, θ is the location parameter of X, and the constant a_n is defined as in A.27 that is

$$a_n = (1 - F_{|X|})^{-1}(1/n)$$

Moreover, suppose that the CDF of X, F_X, satisfies the following equation

$$\lim_{x \to \infty} \frac{F_X(x)}{1 - F_X(x) + F_X(-x)} = p \tag{1.3}$$

---

[11] see equation (A.34) in appendix A for further details.

where $p, q \in [0, 1]$ such that $p + q = 1$ (see, Theorem A.2.8). Observe that

$$\frac{n}{a_n} \to O(1), \qquad \text{as} \qquad n \to \infty \tag{1.4}$$

Now, let $\{X_i^*\}$ be the bootstrap sample generated as

$$X_i^* = [(X_i - m)\tau_i + m]1(|X_i| \leq v_n) + X_i 1(|X_i| > v_n) \tag{1.5}$$

where the i.i.d. sequence $\{\tau_i\}_{i=1}^n$ consists of Rademacher distributed random variables such that $\mathbb{E}(\tau_1) = 0$ and $\mathbb{E}(\tau_1^2) = 1$, and m is a constant that consists of the location θ and some axis of symmetry. In fact, $m = \mathbb{E}(X_i 1(|X_i| < v_n))$. The bootstrap analogue of S_n is given by

$$S_n^* := \frac{1}{a_n} \sum_{i=1}^n (X_i^* - \theta) \tag{1.6}$$

Our main interest concerns inference on θ based on the distribution of S_n^* under the conditional probability $\mathbb{P}^*$ of the sequence $\{|X_i|\}_{i=1}^n$.

Theorem 1.2.1. *There exists a v_n such that the limiting unconditional distribution of S_n^* and S_n is the same and*

$$P^*(S_n^* \leq S_n) = F_{S_n^*}^*(S_n) \xrightarrow{d} U \tag{1.7}$$

where U is uniformly distributed random variable that is $U \sim Uniform(0, 1)$. Consequently, for $\eta \in (0, 1)$, test of the null hypothesis

$$\boldsymbol{H}_0 : \theta = \theta_0$$

against either (1) $\boldsymbol{H}_1 : \theta \neq \theta_0$, (2) $\boldsymbol{H}_2 : \theta > \theta_0$, (3) $\boldsymbol{H}_3 : \theta < \theta_0$, when based on the statistic $|X_n - \theta_0|$ for $\boldsymbol{H}_1$, or the statistic $X_n - \theta_0$ for $\boldsymbol{H}_2$ and $\boldsymbol{H}_3$, and using the critical values $a_n n^{-1} F^{*-1}(1 - \eta/2)$ for $\boldsymbol{H}_1$, or $-a_n n^{-1} F^{*-1}(\eta)$ for $\boldsymbol{H}_2$, or $a_n n^{-1} F^{*-1}(\eta)$ for $\boldsymbol{H}_3$, each have asymptotic size η. Where F^* is the distribution function of the limit of S_n^* under P^* and $F^{*-1}(x) := \inf\{y : F^*(y) \geq x\}$. Moreover, the confidence intervals for θ is of the form

$$\left[\bar{X}_n - a_n n^{-1} F^{*-1}(1 - \eta + \zeta), \bar{X}_n - a_n n^{-1} F^{*-1}(\zeta) \right]$$

have asymptotic coverage probability $1 - \eta$ for any $\zeta \in (0, \eta)$.

Proof. Let v_n be the dth largest element of $\{|X_i|\}_{i=1}^n$ and observe that

$$
\begin{aligned}
S_n^* &= \frac{1}{a_n} \sum_{i=1}^n (X_i^* - \theta) \\
&= \frac{1}{a_n} \sum_{i=1}^n X_i^* - \frac{n}{a_n} \theta \\
&= \frac{1}{a_n} \sum_{i=1}^n X_i^* 1(|X_i| \geq v_n) + \frac{1}{a_n} \sum_{i=1}^n X_i^* 1(|X_i| < v_n) - \frac{n}{a_n} \theta \\
&= \frac{1}{a_n} \sum_{i=1}^n X_i 1(|X_i| \geq v_n) + \frac{1}{a_n} \sum_{i=1}^n (X_i - m) \tau_i 1(|X_i| \leq v_n) + m\left(\frac{n-d}{a_n}\right) - \frac{n}{a_n} \theta \\
&= \frac{1}{a_n} \sum_{i=1}^n (X_i - \theta) 1(|X_i| \geq v_n) + \frac{1}{a_n} \sum_{i=1}^n (X_i - m) \tau_i 1(|X_i| \leq v_n) + \frac{(n-d)}{a_n} (m - \theta)
\end{aligned}
\tag{1.8}
$$

Additionally, observe that

$$
\begin{aligned}
S_n &= \frac{1}{a_n} \sum_{i=1}^n (X_i - \theta) \\
&= \frac{1}{a_n} \sum_{i=1}^n X_i - \frac{n}{a_n} \theta \\
&= \frac{1}{a_n} \sum_{i=1}^n (X_i - \theta) 1(|X_i| \geq v_n) + \frac{1}{a_n} \sum_{i=1}^n X_i 1(|X_i| < v_n) - \frac{n-d}{a_n} \theta \\
&= \frac{1}{a_n} \sum_{i=1}^n (X_i - \theta) 1(|X_i| \geq v_n) + \frac{1}{a_n} \sum_{i=1}^n (X_i - m) 1(|X_i| \leq v_n) + \frac{n-d}{a_n} (m - \theta)
\end{aligned}
\tag{1.9}
$$

where in the fourth line in equation (1.8) we have used the fact that

$$
\begin{aligned}
\sum_{i=1}^n X_i^* 1(|X_i| < v_n) &= \sum_{i=1}^n [(X_i - m)\tau_i + m] 1(|X_i| \leq v_n) \\
&= \sum_{i=1}^n (X_i - m)\tau_i 1(|X_i| \leq v_n) + m(n - d)
\end{aligned}
$$

using part b of (Lemma 2.1 Arcones & Giné 1989, p. 207) (see also Lemma A.2.12), it may be concluded that the central part of the sum,

$$
\frac{1}{a_n} \sum_{i=1}^n (X_i - m)\tau_i 1(|X_i| \leq v_n) \xrightarrow{P} 0 \qquad \text{and} \qquad \frac{1}{a_n} \sum_{i=1}^n (X_i - m) 1(|X_i| \leq v_n) \xrightarrow{P} 0
$$

Therefore, the limit law of both S_n and S_n^* is the same since equation (1.4) holds. Moreover, observe that

$$P^*(S_n^* \leq S_n) = P^*(\frac{1}{n^{1/2}} \sum_{i=1}^{n} (\frac{X_i - m}{A_n}) \tau_i 1(|X_i| \leq v_n) \leq \frac{1}{n^{1/2}} \sum_{i=1}^{n} (\frac{X_i - m}{A_n}) 1(|X_i| \leq v_n)) \quad (1.10)$$

It can be concluded from equation (A.47) in Lemma A.2.15 that the random variable

$$R_i := \frac{1}{A_n} (X_i - m) \, 1 \, (|X_i| \leq v_n)$$

converges to a normal random variable in the limit. Observe that in this case, there exists a sequence of i.i.d. normal random variables, $\{N_i\}_{i=1}^{n}$ such that

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} R_i \sim \frac{1}{n^{1/2}} \sum_{i=1}^{n} N_i$$

Therefore, using the same ideas as LePage et al. (1981) (see, Theorem A.2.11), which are mentioned in Appendix A (see, propositions A.1.2 and A.1.3), the above sum may be written as

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} N_i \xrightarrow{d} \sum_{i=1}^{\infty} \delta_i Z_i$$

where $Z_k = (\sum_{i=1}^{k} E_i)^{-1/2}$ and $\{E_i\}_{i=1}^{k}$ is a sequence of i.i.d. exponential r.v.s with mean one and $\delta_1 \sim Bin(1, -1; 1/2)$. In this case, equation (1.10) turns out to be

$$P^*(S_n^* \leq S_n) \to P^* \left(\sum_{i=1}^{\infty} \delta_i \tau_i Z_i \leq \sum_{i=1}^{\infty} \delta_i Z_i \right) \quad (1.11)$$

The proof follows the same technical details as in (Cavaliere et al. 2013, p. 215-217). $\square$

Remark: The proof of the above also holds if $1(|X_i| \geq v_n)$ and $1(|X_i| \leq v_n)$ are replaced by $1(|X_i - m| \geq v_n)$ and $1(|X_i - m| \leq v_n)$ respectively for some constant m.

Theorem 1.2.1 shows that there exists a numerical value v_n such that if this value is known, one may create a consistent bootstrap algorithm, to test the null hypothesis of the location model when $X \in \mathscr{DA}(\alpha)$, such that $\alpha \in (0,2)$, and the distribution F_X is asymmetric. The difficulty lies in determining this value, however.

It is well known that all distributions that belong to the domain of attraction of a stable law may be standardized (see Definition A.2.1), hence converging to a standardized stable random variable. In this case, one may study the behaviour of v_n with standardized stable random variables $S(\alpha, \beta, 1, 0)$ (see Definition A.2.6). Therefore, in the subsequent sections,

we will study the behaviour of the value v_n in the standardized stable random distribution for various index of stability α and skewness parameter β.

1.2.1   On the Selection of the Critical Value of the Normal Portion

Berkes et al. (2012) have shown that there exists a range of values such that a particular way of trimming based upon it would enable the sum to converge to a normal distribution. However, they were not able to demonstrate a numerical analysis to determine this 'cut-off' critical value that identifies such a portion of the sum that would allow it to converge to a normal distribution. This difficulty is due to the fact that there is a great number of distributions that belong to the domain of attraction of a stable law; hence, there exist various cut-off critical values depending on such distribution which are critical in determining the norming constant A_n (see, Theorem A.2.14). However, the fact that all distributions that belong to the domain of attraction of a stable law may be normalised, such that they converge to a standard stable random variable, helps us to narrow our attention to distributions of standardised stable random variable cases only.

Even though narrowing the focus to distributions of standardised stable random variables leads to a significant ease, the selection of the cut-off critical value v_n in equation (A.47) would still be difficult without the help of Theorem 1.2.1. In particular, taking advantage of Theorem 1.2.1 is a must in order to determine the critical value v_n of a standardized stable distribution. Moreover, one should note that there exist more than one critical value v_n that lead to Theorem 1.2.1. In fact, as we will see in our Monte Carlo procedure 1.2.2, there exist more than one cut off critical value that would lead Theorem 1.2.1 to be valid.

Theorem 1.2.1 claims that as $n \to \infty$, the random measure $F^*_{S^*_n}(S_n)$ converges to a uniform random variable. Therefore, we conjecture that for a finitely large n, there exists a critical value v_n such that the random measure $F^*_{S^*_n}(S_n)$ behaves like a uniform distribution. In this case, Monte Carlo simulations provide an indispensable tool that enables econometric theorists to determine the set of plausible ranges of v_n that satisfies such a criterion by taking advantage of Theorem 1.2.1. Observe that

$$F^*_{S^*_n}(S_n) \in \left[\zeta, 1 - \eta + \zeta \right] \tag{1.12}$$

if and only if

$$\theta \in \left[\bar{X}_n - a_n n^{-1} F^{*-1}(1 - \eta + \zeta), \bar{X}_n - a_n n^{-1} F^{*-1}(\zeta) \right] \tag{1.13}$$

As Cavaliere et al. (2013) pointed, the determination of the confidence interval in equation (1.13) requires no knowledge of a_n since

$$
\begin{aligned}
a_n n^{-1} F^{*-1}(x) &= a_n n^{-1} \inf\{y : F^*(y) \geq x\} \\
&= \inf\{a_n n^{-1} y : F^*(y) \geq x\} \\
&= \inf\{r : F^*(\frac{r}{a_n n^{-1}}) \geq x\} \\
&= \inf\{r : P^*(a_n n^{-1} S_n^* \leq r) \geq x\} \\
&= \inf\{r : P^*(n^{-1}(\bar{X}_n - \theta_0) \leq r) \geq x\} \\
&= \inf\{r : F_{n^{-1}(\bar{X}_n - \theta_0)}(r) \geq x\}
\end{aligned}
$$

Therefore, $a_n n^{-1} F^{*-1}$ is simply the quantile function of $n^{-1}(\bar{X}_n - \theta_0)$.

Having said this, the following Monte Carlo procedure will be used to determine the critical value v_n for various sample sizes $n = 50, 400, 1000$ and for standardized stable distributions with various index of stability α and skewness parameter β.

Algorithm Procedure 1.2.2 (Monte Carlo procedure for determining the critical value v_n)**.**

1. *Generate a sample $\mathscr{H} := \{\varepsilon_i\}_{i=1}^n$ of i.i.d random variables such that $\varepsilon_1 \sim S(\alpha, \beta, 1, 0)$ where $\alpha \in (0, 2)$ and $\beta \in (0, 1)$.*

2. *Choose a percentage of data, in absolute value, to be truncated from $\{\varepsilon_i\}_{i=1}^n$. That is, $F_{|\varepsilon_1|}(v_n) := \pi$ where $\pi \in [[0, 1]]$.*

3. *Generate, based on the sample $\mathscr{H}$, bootstrap samples $\{\{X_{i,j}^*\}_{i=1}^n\}_{j=1}^B$ as described in equation (1.5) where B represents the number of bootstrap samples created based on a single sample $\mathscr{H}$.*

4. *Compute the sequence $\{r_i\}_{i=1}^B$ such that $r_j := n^{-1}(\bar{X}_{n,j}^* - \theta_0)$ where $j = 1, 2, \ldots, B$.*

5. *Compute the $100(\zeta)\%$ and $100(1 - \eta + \zeta)\%$ quantiles of the sequence $\{r_i\}_{i=1}^B$ in order to get the $100(1 - \eta)\%$ coverage probability of θ defined in equation (1.13).*

6. *Repeat the previous steps 10000 times for different values of π.*

7. *Display the coverage probabilities versus the truncated value for various sample sizes*

Our simulation study is based on samples of size 50 and 400 observations from a standardized stable distribution. We selected the level $\eta = 0.05$ and ζ to be half of it. Observe that when π

is chosen to be one, then the bootstrap method described becomes the Cavaliere et al. (2013) WB method. In this case, the above Monte Carlo procedure will assess the WB method rather than the TWB.

Figures 1.1, 1.2, and 1.3 show the Coverage probabilities versus the percentages of data being truncated from standardized random variables with index of stability α of 1.1, 1.8, and 1.5. The skewness parameter β ranging from 0 to 1 of sample size $T = 50$. Similarly Figure 1.4 shows the coverage probabilities versus the percentages of data being truncated from standardized random variables with an index of stability $\alpha = 1.8$ with a skewness parameter β ranging from 0 to 1 of sample size $T = 400$[12].
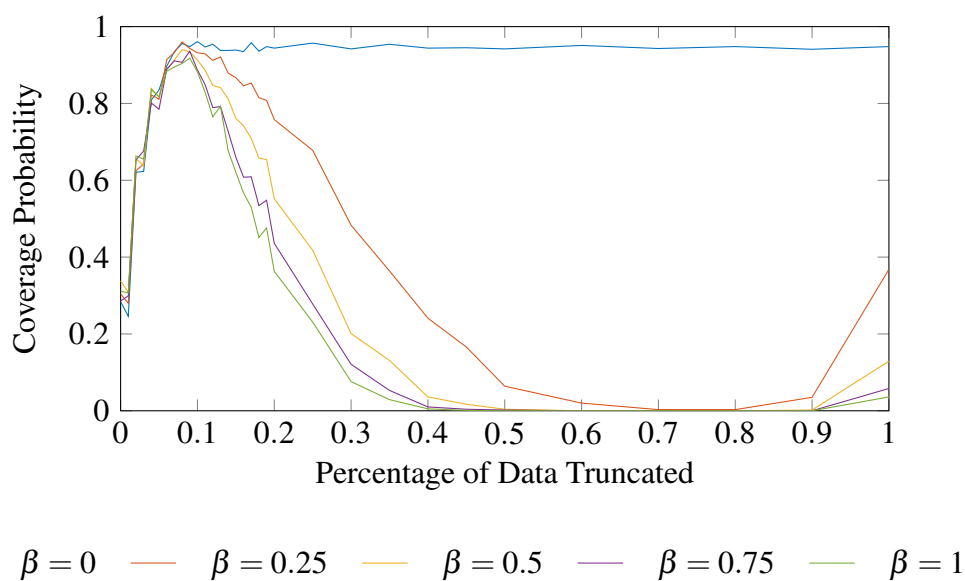
Figure 1.1 P values versus the quantile of the absolute value of the data being truncated for $(\alpha, n) = (1.1, 50)$

---

[12]Various values of truncation has also been conducted but it is suffice to show only these figures
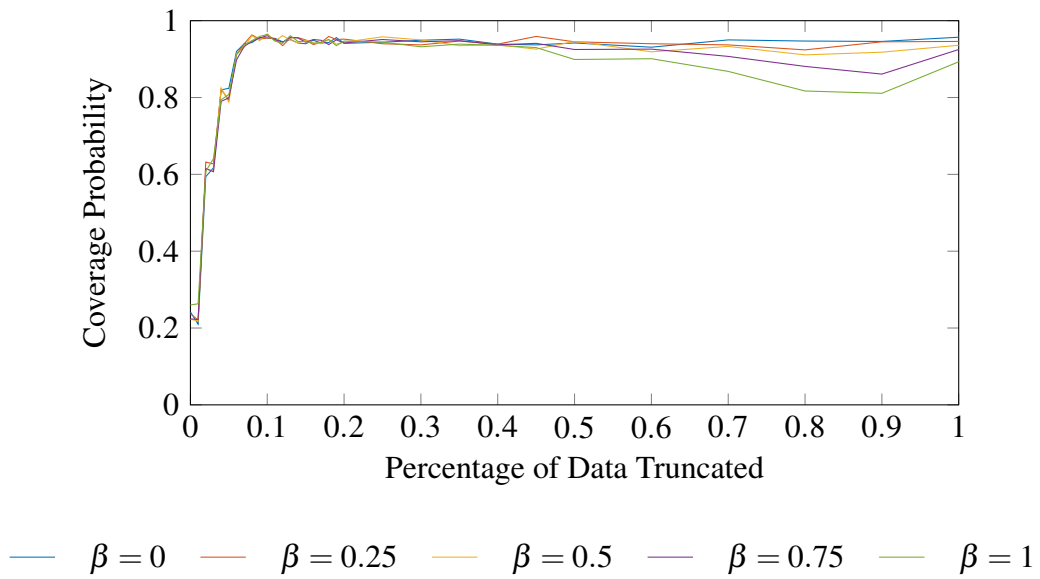
Figure 1.2 *P* values versus the quantile of the absolute value of the data being truncated for $(\alpha, n) = (1.8, 50)$
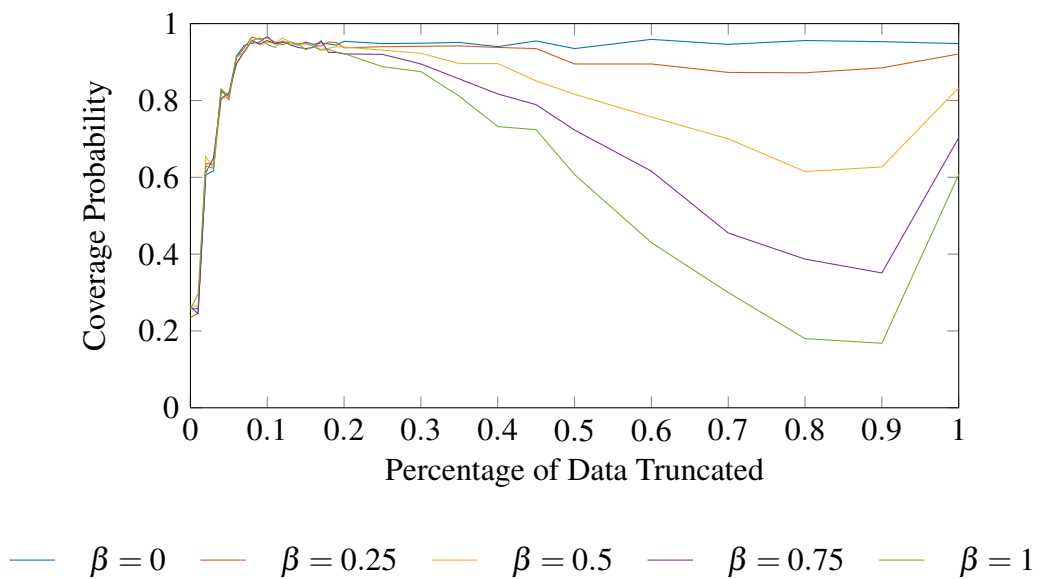
Figure 1.3 *P* values versus the quantile of the absolute value of the data being truncated for $(\alpha, n) = (1.5, 50)$
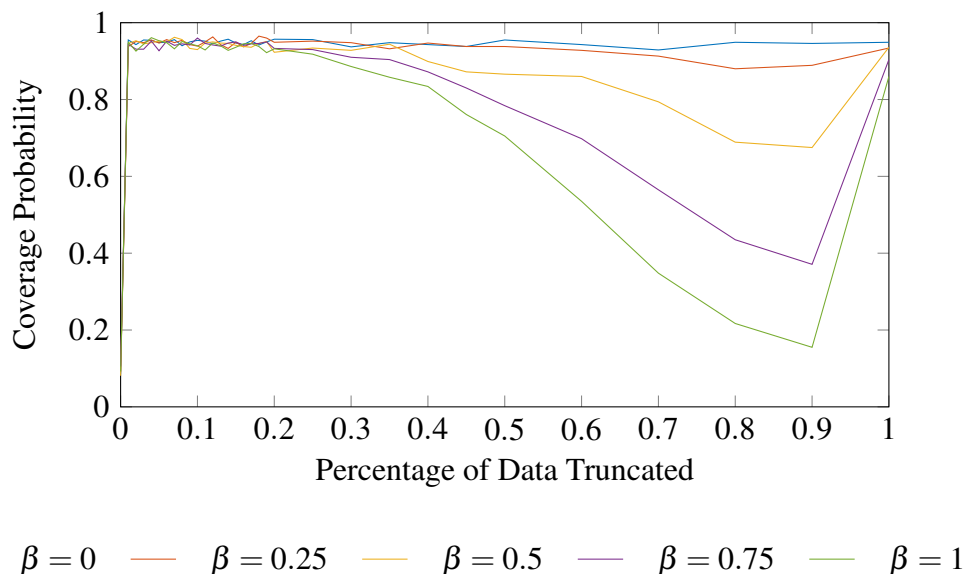
Figure 1.4 *P* values versus the quantile of the absolute value of the data being truncated for $(\alpha, n) = (1.8, 400)$

The above figures reach the same conclusions of Berkes et al. (2012). In particular, it seems evident that the validity of the central limit theorem depends sensitively on the behaviour of the tail ratio of the underlying variable X. This may be demonstrated by observing that the coverage probabilities become worse as the skewness parameter β approaches one or the index of stability becomes closer to one or even below it. However, the coverage probabilities become closer to the nominal levels as the skewness parameter goes closer to zero or the index of stability approaches two.

Moreover, the above figures also demonstrate the second conclusion of Berkes et al. (2012). That is, increasing the number of trimmed elements and thereby decreasing the critical value v_n, does not generally improve the partial sum behaviour of the central limit theorem. To Berkes et al. (2012), this is a paradox, because, intuitively, excluding more relatively large terms of a sum should make the behaviour of the trimmed sum more "normal" which is not the case, however. This could be observed by acknowledging that the coverage probability is not close to the intended (associate) nominal level when the percentage of the data truncated is small, and it becomes better after a certain proportion of data is truncated.

Truncation based on a percentage of data being assessed by the Monte Carlo procedure 1.2.2 is problematic when it comes to simulations. The problem is due to the fact that the critical value changes with the sample, given the fact that the sample originates from the same distribution. This is because such a procedure of selecting the critical value would require it to be dependent on the perceived observations. Therefore, it is vital to determine the critical

value based on the distribution but not on the perceived observations. This entails finding critical values around the axis of relative symmetry, such that the Monte Carlo procedure described above results in the "correct" nominal level of the test.

1.2.2 Symmetry around the mode of the stable distribution

As observed in subsection A.2.5 and particularly in Table A.1, there is a unique connection between the mode m of a stable distribution F_X and the joint variables (α, β) which characterise the normalized stable distribution. Therefore, if the values α and β are known, it is easy to find m, using the iterative computational method as displayed on Table A.1. It is equally easy to note, using the inverse Fourier transformation on the stable characteristic function, that the probability density function (PDF), f_X, of the stable distribution is smooth and differentiable everywhere, especially around the mode (Fofack & Nolan 1999).

Therefore, stable distributions are smooth around their unique mode which represents a *relative axis of symmetry* in such PDF. It is called the relative axis of symmetry because there exists an interval $I = [-v, v]$ around the mode m, such that the density function f_{X-m} is symmetric. The level of symmetry around the mode m in the unimodal skewed stable distributions has not been studied in the earlier literature. Therefore, to explore this, we introduce the function g defined as

$$g(v) = |f_{X-m}(v) - f_{X-m}(-v)|, \qquad \text{where} \qquad v \in \mathbb{R}^+$$

For each index of stability α and skewness parameter β of the standardized stable random variable X, we recorded the value v, such that the actual rejection probability is close to the nominal level and the value of the function g is as close to zero as possible. On the one hand, if the value of v is too small, fewer observations will be bootstrapped and the worse the central limit behaviour may become. On the other hand, the bigger the value v is, the more observations will be bootstrapped and the further the function g deviates from zero; hence leading to bootstrapping asymmetric distribution which worsens the central limit theorem behaviour (see, Berkes et al. 2012). Moreover, as a consequence of Lemma 1.2.3, applying the wild bootstrap on skewed stable distributions results in a symmetric stable distribution. Therefore, the critical value v must be chosen appropriately.

Lemma 1.2.3. *Let $\{X_i - \theta_0\}_{i=1}^n$ be an i.i.d. sequence such that $X_1 \sim \mathbf{S}(\alpha, \beta, 1, \theta_0)$ where $\alpha \in (0, 2)$ and $\beta \in (0, 1)$ then for any probability value $p_1 \in (0, 1)$ the following will hold*

$$V_T := a_T^{-1} \sum_{i=1}^{T} (X_i - \theta_0) \omega_i \xrightarrow{d} \mathbf{S}(\alpha, \beta(p_1 - q_1), 1, 0)$$

where $q_1 = 1 - p_1$. $\{\omega_i\}_{i=1}^n$ is a sequence of random variables such that $\omega_1 \sim Bin(1, -1; p_1)$.

Proof. Define $Y := X - \theta_0$ to be the random variable associated with $\mathbf{S}(\alpha, \beta, \sigma, 0)$ distribution. Using equation (1.8) in (Nolan 2015, p. 19-20) and taking into account that $\omega_i Y_i \sim \mathbf{S}(\alpha, \omega_i \beta, \sigma, 0)$ for $i = 1, 2, 3, \cdots, n$. Therefore, the sum

$$\sum_{i=1}^n \omega_i Y_i \sim \mathbf{S}(\alpha, \beta_1, \sigma_1, 0)$$

where $\sigma_1^\alpha = n\sigma^\alpha$, $\beta_1 = \beta\sigma_1^\alpha(\sum \omega_i)/\sigma^\alpha = \frac{\beta}{n}\sum_{i=1}^n \omega_i \to \beta(p - q)$. Therefore,

$$\frac{1}{n^{1/\alpha}} \sum_{i=1}^n \omega_i Y_i \xrightarrow{d} \mathbf{S}(\alpha, \beta(p - q), \sigma_1, 0)$$

$\square$

As mentioned earlier, there exist more than one critical value v_n that lead to the result of Theorem 1.2.1 for the same distribution, say $\mathbb{S}(\alpha)$. Therefore, in Table 1.1, we only displayed the critical values, such that for a small change in either the index of stability α or skewness parameter β, the critical values may exhibit the least change. The reason for this, as we will see in section 1.3, is because the new proposed bootstrap procedure 1.2.4 will rely on estimating the parameters α and β, which would vary significantly depending on the observations and the sample size. Since the selection of the critical value v_n would be dependent on the estimation of these two parameters, we would like to reduce the variability by reducing the difference between the critical values v_n, such that for a small change in either of the two parametric variables there is almost no much change in the critical values. The values that are recorded in Table 1.1 have been based on the value of the function calculated g and the observed coverage rates of the Monte Carlo procedure, described in 1.2.2, but with a minor alteration and with a sample size of $n = 400$.

	α								
β	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1
0.1	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
0.2	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
0.3	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
0.4	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
0.5	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
0.6	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
0.7	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
0.8	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
0.9	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
1	0.006	0.03	0.07	0.15	0.2	0.3	0.5	0.5	0.6
β	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
0.1	0.7	0.7	0.7	0.8	0.8	1	1.4	1.5	2
0.2	0.7	0.7	0.7	0.8	0.8	1	1.4	1.5	2
0.3	0.7	0.7	0.7	0.8	0.8	1	1.4	1.5	2
0.4	0.7	0.7	0.7	0.8	0.8	1	1.4	1.5	2
0.5	0.7	0.7	0.7	0.8	0.8	1	1.4	1.5	2
0.6	0.7	0.7	0.7	0.8	0.8	1	1.4	1.5	2
0.7	0.7	0.7	0.7	0.8	0.8	1	1.4	1.5	2
0.8	0.7	0.7	0.7	0.8	0.8	1	1.4	1.5	2
0.9	0.6	0.7	0.7	0.8	0.8	1	1.4	1.5	2
1	0.6	0.7	0.7	0.8	0.8	1	1.4	1.5	2

Table 1.1 The Location of the symmetry point v around the Mode

Using a similar approach of constructing Table 1.1, Table 1.2 depicts the percentages of observations being truncated such that the MC procedure described in 1.2.2 results in an actual rejection probability close to the nominal level. It should be noted that there exists a trade-off when utilizing each table. For example, using the values in Table 1.2, one may witness a different critical value v every time the observed sample changes. Therefore, even if the DGP of the sample is the same, the critical value v varies depending on the observed sample. Therefore, as will be shown in the simulation section 1.3, the values in this table will work well when testing the hypothesis of observed sample sizes similar or close to 400, the sample size used in the constructing table.

					α				
β	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1
0.1	0.030	0.057	0.083	0.133	0.152	0.203	0.302	0.429	0.339
0.2	0.029	0.056	0.082	0.132	0.151	0.201	0.300	0.428	0.339
0.3	0.029	0.055	0.081	0.130	0.149	0.200	0.299	0.426	0.337
0.4	0.028	0.054	0.079	0.128	0.147	0.198	0.296	0.424	0.336
0.5	0.027	0.052	0.077	0.125	0.145	0.195	0.294	0.422	0.333
0.6	0.026	0.051	0.075	0.122	0.142	0.192	0.290	0.420	0.331
0.7	0.025	0.049	0.072	0.119	0.138	0.188	0.287	0.417	0.328
0.8	0.024	0.047	0.069	0.116	0.135	0.185	0.283	0.414	0.325
0.9	0.022	0.044	0.067	0.112	0.131	0.181	0.279	0.410	0.322
1	0.021	0.042	0.064	0.109	0.128	0.177	0.275	0.404	0.318
β	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
0.1	0.383	0.382	0.380	0.426	0.426	0.516	0.598	0.706	0.843
0.2	0.383	0.381	0.380	0.426	0.426	0.516	0.598	0.706	0.843
0.3	0.382	0.380	0.379	0.426	0.426	0.516	0.598	0.706	0.843
0.4	0.380	0.379	0.378	0.425	0.426	0.515	0.598	0.706	0.843
0.5	0.378	0.378	0.377	0.424	0.425	0.515	0.598	0.706	0.843
0.6	0.376	0.376	0.376	0.423	0.424	0.514	0.598	0.706	0.843
0.7	0.374	0.374	0.374	0.422	0.424	0.514	0.597	0.706	0.843
0.8	0.371	0.372	0.373	0.421	0.423	0.513	0.597	0.706	0.843
0.9	0.368	0.369	0.371	0.419	0.422	0.513	0.597	0.706	0.843
1	0.365	0.367	0.369	0.418	0.421	0.512	0.596	0.706	0.843

Table 1.2 The Probability around the Mode on the Interval I

In general, Table 1.1, is superior to Table 1.2 when the sample size is larger or even smaller than 400. The reason for this superiority is because if the index of stability α and skewness parameter β are known, the critical value will always be the same, regardless of the change of the observed sample.

It seems evident from the previous discussion, that one may utilise equation (1.7) in Theorem 1.2.1 along with tables 1.1 and 1.2 to create a new bootstrap procedure. The new bootstrap procedure is superior to other available bootstrap methods like the PB and the WB. This superiority is due to the fact that it is capable in testing the hypothesis of location parameter when the index of stability is below one and when the skewness parameter is not zero.

1.2.3 Truncated Wild Bootstrap

The Truncated Wild Bootstrap (TWB) is a bootstrap algorithm that is similar in construction to the Wild Bootstrap (WB) of Cavaliere et al. (2013) but is only limited to a number of observations that are truncated based on some critical values around an axis of symmetry, hence its name. The TWB is more general than the WB because it accounts for asymmetric distributions. Moreover, under further assumptions, namely if the observations belong to the domain of attraction of a symmetric stable law, the TWB is the WB of Cavaliere et al. (2013).

The TWB for testing the null hypothesis 1.2.1 that we propose here is described by the following steps.

Bootstrap Procedure 1.2.4 (Truncated Wild Bootstrap)**.**

1. *Generate a sample $\mathscr{H} := \{X_i\}_{i=1}^n$ of i.i.d random variables from equation (1.1) such that $F_{\varepsilon_1} \in \mathscr{DA}(\alpha)$ where $\alpha \in (0,2)$.*

2. *Estimate α, β, and scale σ consistently from the sample $\mathscr{H}$, and denote them now, and hereafter, by $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}$ respectively.*

3. *Select the a percentage of data to be truncated from Table 1.2 based on the estimated α and β; hence select v_n.*

4. *Choose the mode m from Table A.1 based on α and β.*

5. *Generate, based on the sample $\mathscr{H}$, bootstrap samples $\{\{X_{i,j}^*\}_{i=1}^n\}_{j=1}^B$ as follow*

$$X_{i,j}^* = \left[\left((\kappa_i - m)\tau_j + m\right)\hat{\sigma} + \theta_0\right]1(|\kappa_i - m| \le v_n) + X_i 1(|\kappa_i - m| > v_n) \quad (1.14)$$

where $\kappa_i := \hat{\sigma}^{-}1(X_i - \theta_0)$ and B represents the number of bootstrap samples created based on a single sample $\mathscr{H}$.

6. *Compute the bootstrap P value which is equal to the proportion of bootstrap statistics S_n^* more extreme than S_n*

$$P_{B,n,\hat{\alpha}_n,\hat{\beta}_n}^* = \frac{1}{B}\sum_{i=1}^B 1(S_{n,i}^* \le S_n) \quad (1.15)$$

Where $S_{n,j}^* := a_n^{-1}\sum_{i=1}^n (X_{i,j}^* - \theta_0)$, $1(.)$ is an indicator function, and $S_n := a_n^{-1}\sum_{i=1}^n (X_i - \theta_0)$.

Observe that as $B \to \infty$, the bootstrap P value $P^*_{B,n,\hat{\alpha}_n,\hat{\beta}_n}$ converges almost surely to $\mathbb{P}(S^*_{n,\hat{\alpha}_n,\hat{\beta}_n,\hat{\sigma}} \leq S_n)$ by law of large number.

This is based on the discussion of section A.2.6 on the estimation methods available on stable parameters and due to assumption 2 in (Cornea-Madeira & Davidson 2015, p. 455), as $n \to \infty$, $\mathbb{P}(S^*_{n,\hat{\alpha}_n,\hat{\beta}_n,\hat{\sigma}} \leq S_n)$ converges almost surely to the same limit distribution of the probability measure $P^*(S^*_n \leq S_n)$ defined in Theorem 1.2.1.

1.3   Simulation Evidence

In this section, the performance of the TWB is investigated and compared with its main competitors: the PB and the WB. It should be noted that the PB and the WB have not been compared in the literature previously, due to the fact that the WB is applicable only for observations that emerge from distributions that belong to the domain of attraction of a symmetric stable law. However, such a piece of information is too difficult to be categorically verified. Therefore, in reality, estimation of the index of stability and skewness parameters is indispensable.

In general, practitioners may apply the WB if the estimate of the skewness parameter is zero, but it might be the case that the estimated skewness parameter is not zero yet the original DGP belong to symmetric distribution. Therefore, in this case, we would be interested in understanding the performance of the WB under the null hypothesis when the estimated skewness parameter is not zero. Therefore, in our simulations, regardless of the estimated skewness parameter, the WB would be applied, and the discrepancy between the actual rejection probability of the bootstrap test and the nominal level would be computed.

In his paper, Davidson (2008) mentions that the current asymptotic theory predicts the bootstrap test performance is worse than what simulation experiments reveal. Therefore, simulations would be the best way of measuring the performance of a bootstrap method. Thus, following these remarks, we resorted to simulation experiments to show the performance of the new bootstrap procedure with respect to the PB and WB.

Our simulation study is based on sample sizes of 100, 400, and 1000. Note that tables 1.1 and 1.2 were constructed based on a sample size 400, however[13]. The DGP considered in generating the sample $\mathscr{H}$ are t distribution with degrees of freedom 1.1, 1.5, 1.9, stable distribution with $\alpha = 0.6, 1.3, 1.5, 1.9$, $\beta = 0, 0.5, 1$, and scale $\sigma = 1$.
The sequential steps we followed to perform this comparison are described as follow

---

[13]Constructing tables 1.1 and 1.2 with a sample size of 50 would not produce result in good coverage as those tabulated; however, for larger sample sizes the procedure works as well

Algorithm Procedure 1.3.1.

1. *Generate a sample $\mathscr{H} := \{X_i\}_{i=1}^n$ from t distribution or stable distribution*

2. *Estimate α, β, and scale σ consistently from the sample $\mathscr{H}$*

3. *Generate Bootstrap sample $\mathscr{H}^* := \{X_i^*\}_{i=1}^n$ based on the bootstrap in question.*

    - *For the TWB 1, simulate the bootstrap observations as described in equation (1.14) using the cut-off values in Table 1.2*

    - *For the TWB 2, simulate the bootstrap observations as described in equation (1.14) using the cut-off values in Table 1.1*

    - *For the WB, simulate the bootstrap by simply multiplying by a radamacher distributed random variable*

    - *For the PB, simulate the bootstrap by a stable random variables such that $X_1^* \sim S(\hat{\alpha}, \hat{\beta}, 1, 0)$*

4. *Compute the bootstrap P value for each numerical simulation*

    - *For the TWB, compute the estimated bootstrap P value P_1 as described in equation (1.15)*

    - *For the WB, compute the estimated bootstrap P value P_2 as described in equation (1.15) but with*

    $$S_{n,j}^* = \frac{1}{a_n} \sum_{i=1}^n (X_i - \theta_0)\tau_{i,j}, \qquad where \qquad j = 1, 2, \ldots, B$$

    where $\{\{\tau_{i,j}\}_{i=1}^n\}_{j=1}^B$ is a sequence of Rademacher distributed random variables.

    - *For the PB, compute the estimated bootstrap P value P_3 described as*

    $$P_3 = \frac{1}{B} \sum_{i=1}^B 1(T_{n,i}^* \le T_n)$$

    where $T_{n,j}^*$ is defined as

    $$T_{n,j}^* = \frac{\sqrt{n}(\bar{X}_{n,j}^* - \theta_0)}{((n-1)^{-1} \sum_{i=1}^n (X_{i,j}^* - \bar{X}_n^*))}$$

    and T_n has the same structure of $T_{n,j}^*$ but without the asterisk and the j component.

5. *Repeat the previous steps with $N := 10,000$ replications and with $B = 400$ bootstrap repetitions to get a sequence of N bootstrap P values $\{P_{1,i}, P_{2,i}, P_{3,i}\}_{i=1}^{N}$.*

6. *Under the true null hypothesis, compute the ERP which is defined by the estimated actual significance level of each bootstrap P Value $\{P_{1,i}, P_{2,i}, P_{3,i}\}_{i=1}^{N}$ and the nominal significance level λ which is asymptotically achieved only. The ERP is given by*

$$ERP_k(\lambda) := \frac{1}{N} \sum_{i=1}^{N} (P_{k,i} \leq \lambda) - \lambda, \qquad where \qquad k = 1, 2, 3 \qquad and \qquad \lambda \in (0, 1)$$

As discussed previously, we are interested in comparing our simulations to Cornea-Madeira & Davidson (2015). Accordingly, we will follow the same approach, as well as the parameters used in the simulation section of Cornea-Madeira & Davidson (2015). Therefore, we considered the case when the null hypothesis $\theta = 0$ in equation (1.1) is true and the case in which the alternative hypothesis $\theta = -0.5$ is true. We opted to display, under the true null, the bootstrap P value discrepancy plots which are represented by the ERP versus nominal level. Recall that the best performance of the tests is achieved when the error in rejection probability (ERP) is close to zero. However, under the true alternative, we will plot the estimated power function $\pi(\theta, \lambda)$ versus the nominal level λ.

The explanation of the power function $\pi(\theta, \lambda)$ is of necessity. The power function at $\theta = -0.5$, $\pi(-0.5, \lambda)$, is justified as that if the true parameter θ in equation (1.1) is equal to -0.5 then there is $100\pi(-0.5, \lambda)\%$ probability chance that the test will reject the (false) null hypothesis that the parameter $\theta = 0$. Therefore, a bootstrap test is more powerful than another if the power function is much different from the nominal level. In our simulations, we plot the power function versus nominal level. Note that the further the power deviates from the nominal the more powerful the bootstrap method is in testing hypothesis. Graphically, this is represented by the deviation from the 45-degree line.
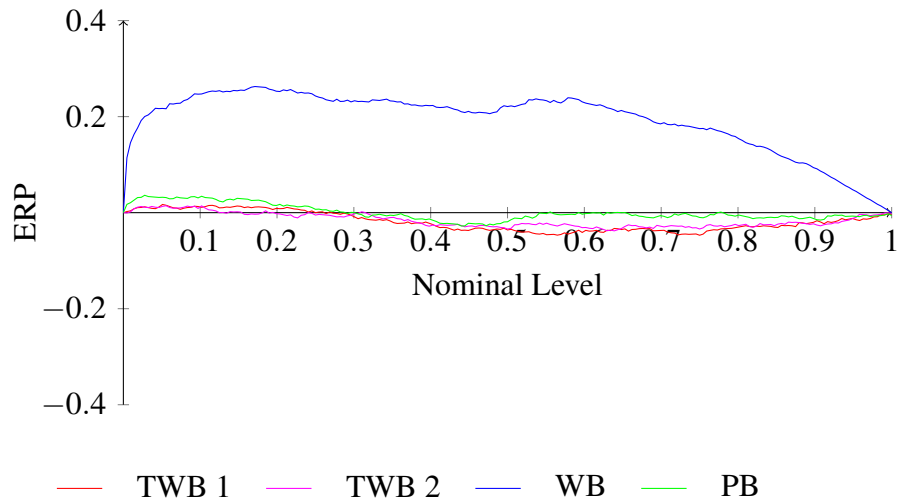
(a)
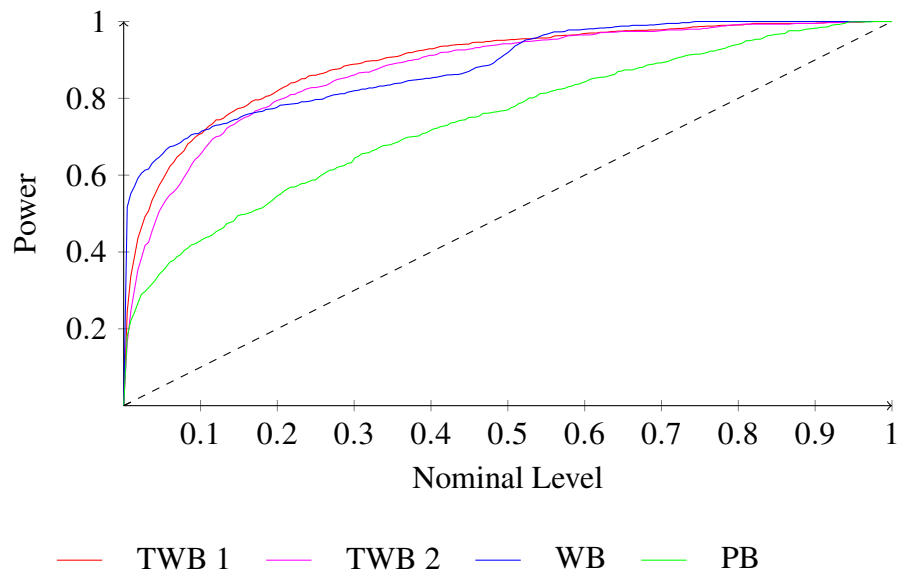
(b)

Figure 1.5 The TWB 1, TWB 2, WB, PB algorithms; data from stable law with $\alpha = 1.3$, $\beta = 0.5$, scale $\sigma = 1$ sample size $n = 400$; (a) P value discrepancy plots (b) Power

Part (a) of all figures represents the P value discrepancy plots under the null hypothesis $\theta = 0$ is true. While part (b) of all of them represents the power plots that is the probability when the DGP of the observations is generated from $\theta = -0.5$ and the null hypothesis tested is $\theta_0 = 0$.
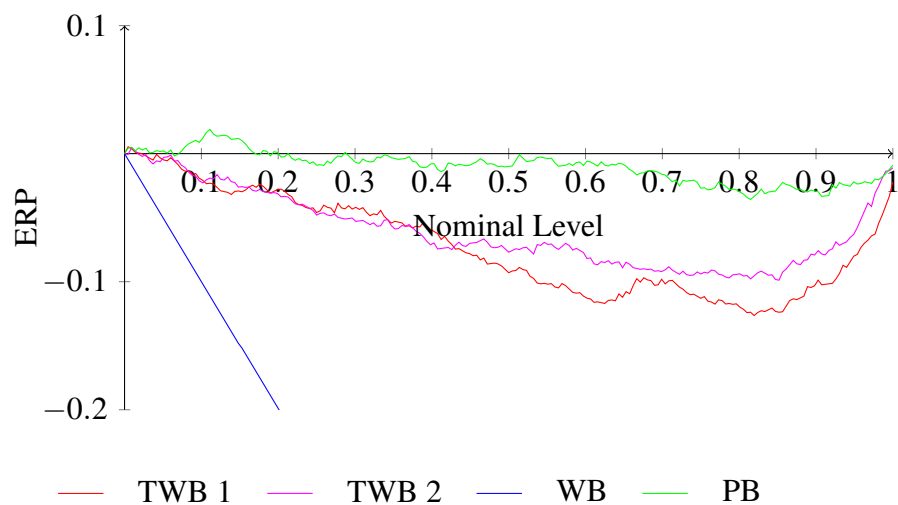
(a)
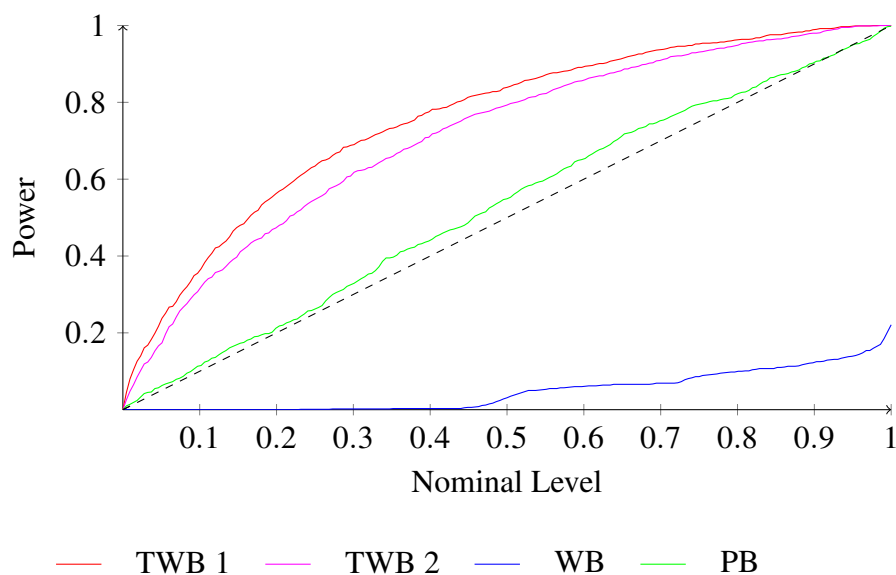
(b)

Figure 1.6 The TWB 1, TWB 2, WB, PB algorithms; data from stable law with $\alpha = 0.8$, $\beta = 0.7$, scale $\sigma = 1$ sample size $n = 400$; (a) P value discrepancy plots (b) Power

It can be seen in part (b) of figures 1.6 and 1.9 that the PB is powerless. In fact, the PB is powerless in all cases when $\alpha \in (0,1)$. Hence, it is not a suitable bootstrap method for testing the null hypothesis of location parameter in these circumstances.
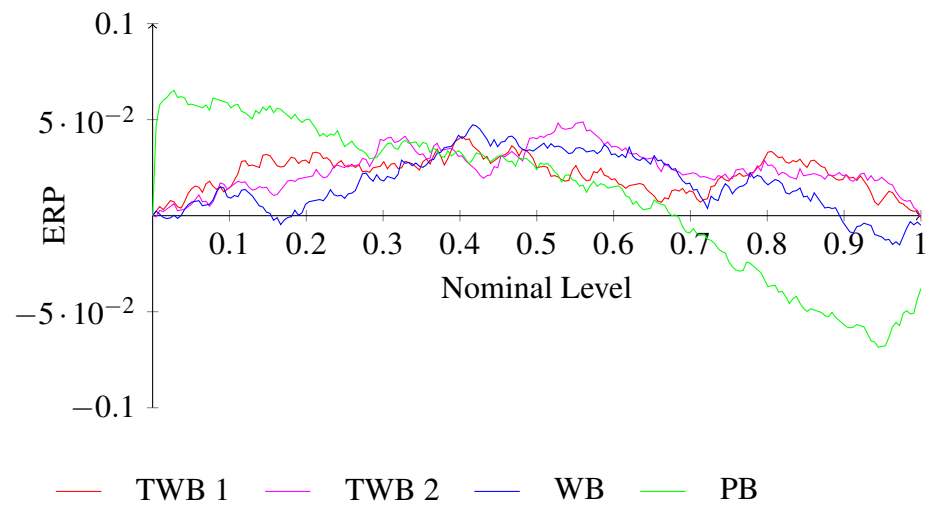
(a)
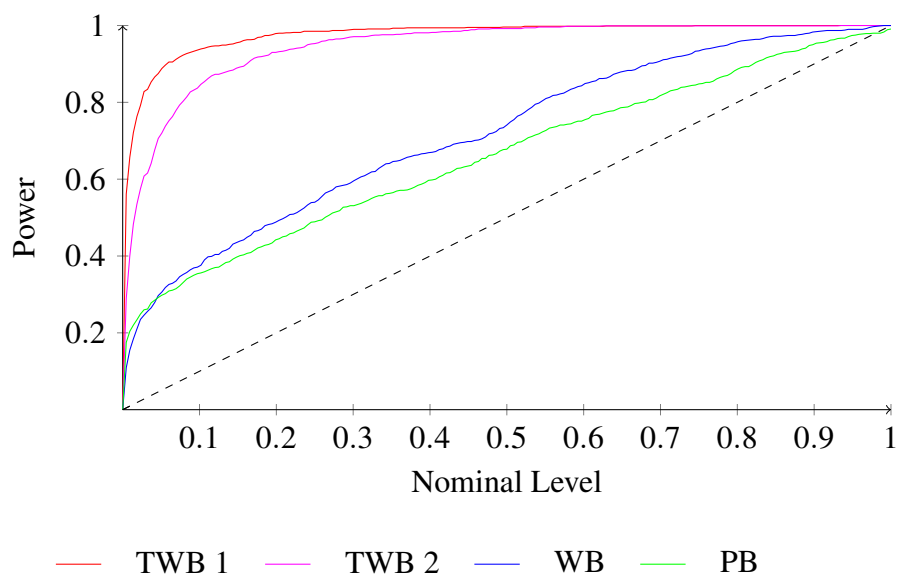
(b)

Figure 1.7 The TWB 1, TWB 2, WB, PB algorithms; data from stable law with $\alpha = 1.1$, $\beta = 0$, scale $\sigma = 1$ sample size $n = 400$; (a) P value discrepancy plots (b) Power

Figure 1.7 shows a symmetric distribution but with heavy-tails which indicates that the power is not satisfactory for the PB test, while it is for all other studied bootstraps that have a much higher power. Figures not displayed here indicate that the same conclusions hold for the t distribution.
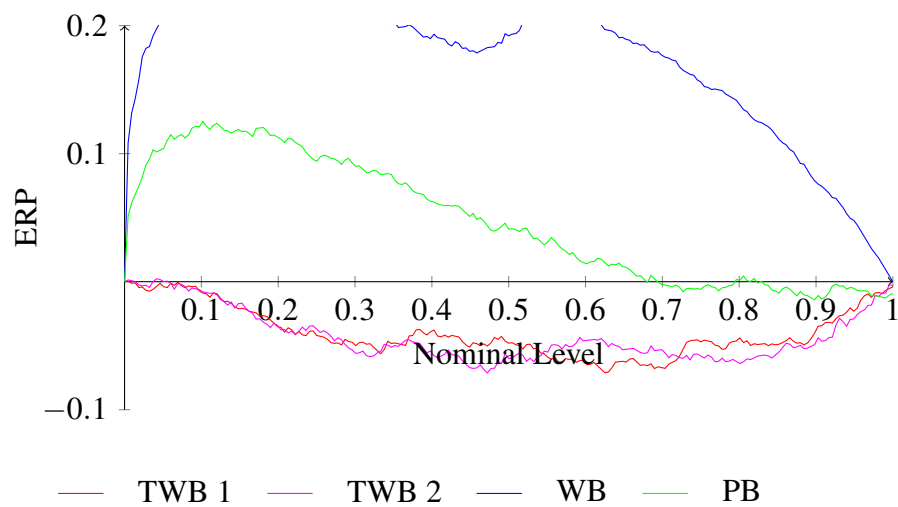
(a)
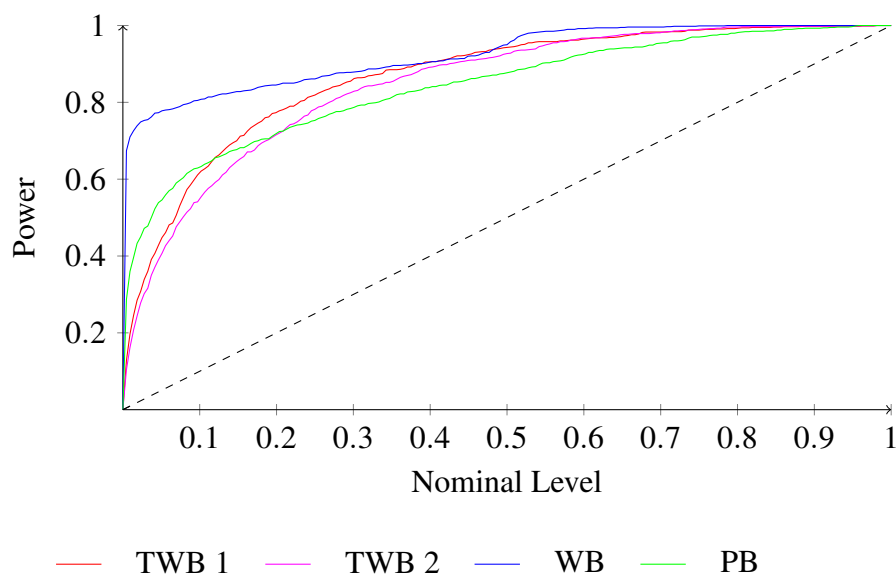
(b)

Figure 1.8 The TWB 1, TWB 2, WB, PB algorithms; data from stable law with $\alpha = 1.5$, $\beta = 1$, scale $\sigma = 1$ sample size $n = 50$; (a) P value discrepancy plots (b) Power
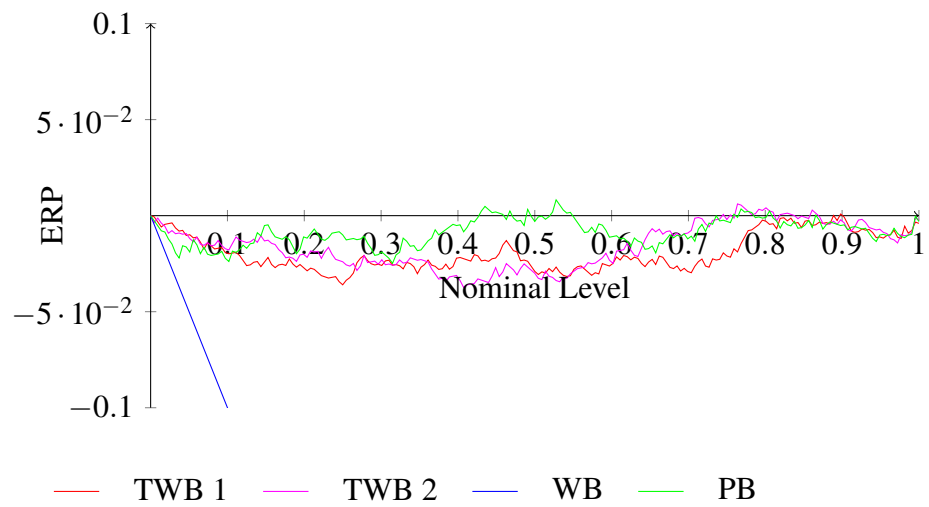
(a)
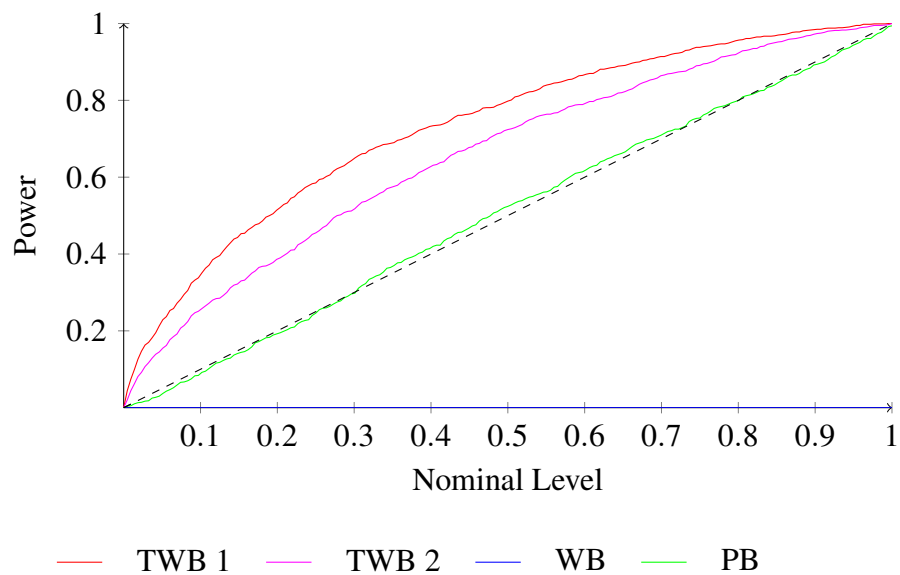
(b)

Figure 1.9 The TWB 1, TWB 2, WB, PB algorithms; data from stable law with $\alpha = 0.6$, $\beta = 1$, scale $\sigma = 1$ sample size $n = 1000$; (a) P value discrepancy plots (b) Power
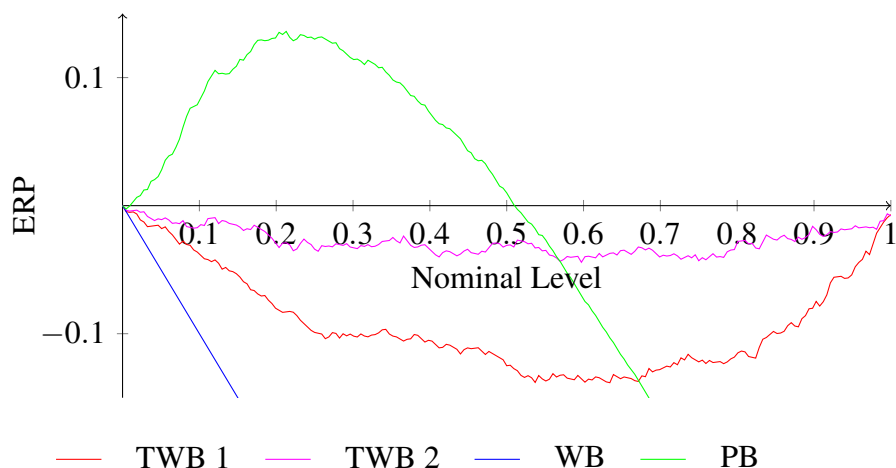
(a)
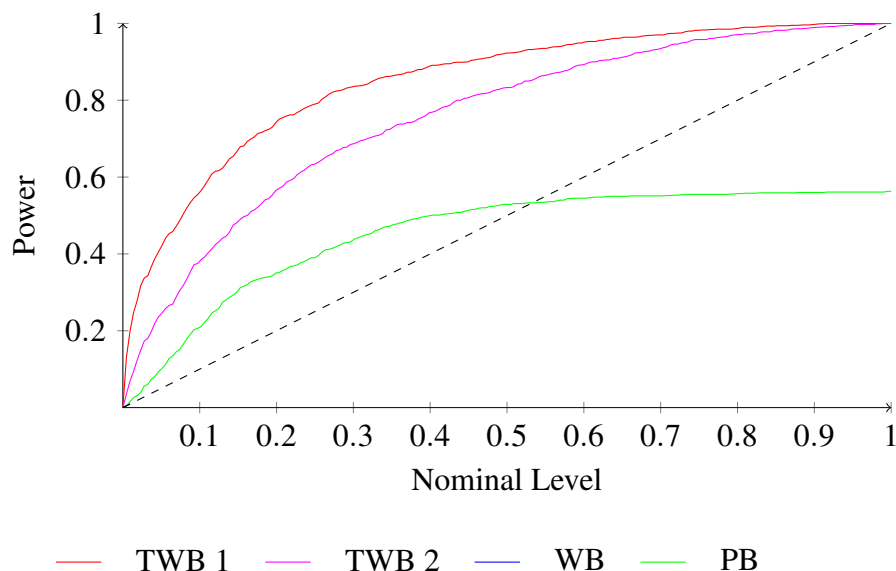
(b)

Figure 1.10 The TWB 1, TWB 2, WB, PB algorithms; data from stable law with $\alpha = 1$, $\beta = 1$, scale $\sigma = 1$ sample size $n = 400$; (a) P value discrepancy plots (b) Power

The rest of the figures, like 1.8, 1.5, and 1.10 where α is greater than one and where there is some mild to extreme skewness level β, show that the TWB 1 and 2 are better than the available bootstrap test for various sample sizes. Moreover, in general, the power of each of the TWB 1 and 2 is always higher than the power of the other bootstrap tests. Results not included here indicate that the same conclusion holds for samples smaller than the one considered in the figure and also for the stable distribution.
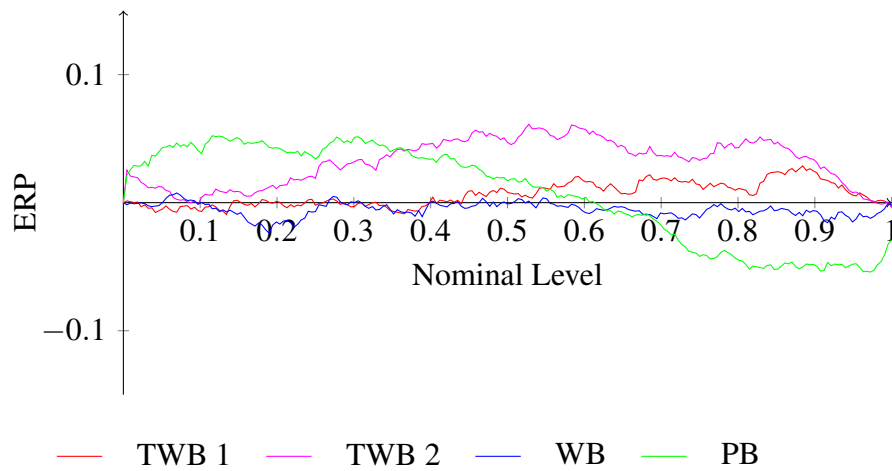
(a)
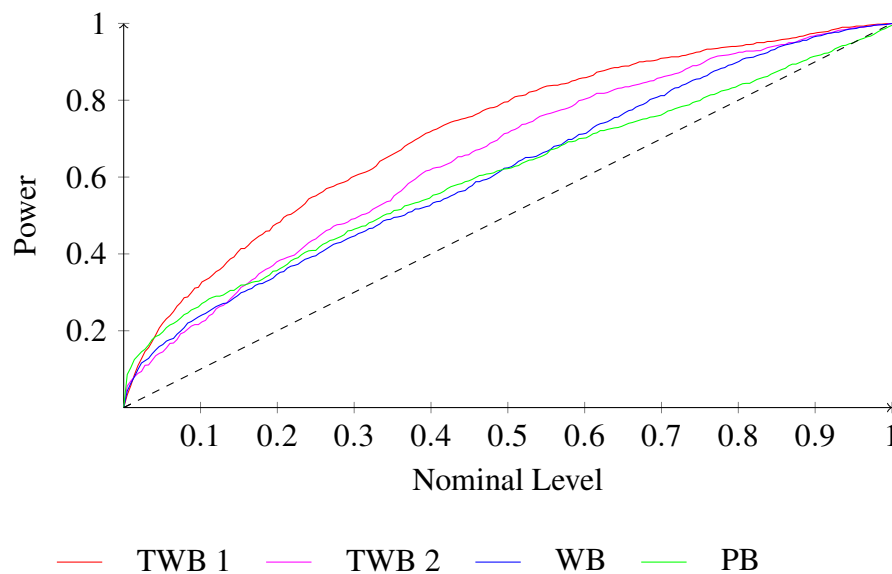
(b)

Figure 1.11 The TWB 1, TWB 2, WB, PB algorithms; data from t distribution with degrees of freedom $\alpha = 0.9$, sample size $n = 50$; (a) P value discrepancy plots (b) Power

There are some extreme situations where all tested bootstrap tests do not perform well, especially when α is equal to 1 and the scale is very high. However, when α is not identically 1, the TWB 1 and 2 give satisfactory results even when sigma is large enough, and the sample size is as small as 50. In general, the closer α is to 1, the closer β is to 1 or -1, and the bigger the scale sigma is, the lower the power of all bootstrap tests.

It can be noted that the TWB 1 and 2 have a faster rate of convergence than the other bootstrap methods and perform very well. In conclusion, the figures indicate that the TWB 1 and 2 work better than other bootstrap tests available in the literature.

1.4   Extensions and Conclusion

In this chapter, a new bootstrap procedure was proposed for the purposes of inference on the location parameter of an i.i.d. heavy-tailed distributed sample. The Truncated Wild Bootstrap is based on the fact that some parts of the normalised sum behave asymptotically like a normal distribution; hence, the central-limit argument holds for partial sums. However, the validity of the central limit theorem depends sensitively on the behaviour of the tail ratio which is influenced by both parameters, the index of stability α and the skewness parameter β. Therefore, to determine the region of the normal part and that of the Poisson part, the cut-off values that characterise the normal part are required. It was stated that for the bootstrap to be estimated consistently, the location of the normal part needs to be identified and the only way to identify it is through the estimation of the parameter α and β.

Our results show that our bootstrap is effective in testing a hypothesis as long as the estimation of α and β is reliably precise. Moreover, the two versions of the TWB, TWB 1 and TWB 2, were able to perform reliably in testing the null hypothesis of zero location parameter with a sample size as small as 50 and when the index of stability is either below or above unity with extreme skewness level[14].

Further research of this study may be split into two major branches in econometrics: theoretical and empirical. In the theoretical econometric branch, it would be of interest to study the following research questions: 1) Would a better estimation of the stable parameters leads to a better result of the proposed bootstrap method? 2) Would it be possible to generalize the TWB for testing the location parameter when the innovations originate from time-dependent heavy-tailed and skewed distributions? 3) Could the proposed idea of the TWB method help in amending the robust inference of the t-statistic approach of Ibragimov & Müller (2010) in a way that their approach may be applicable not only to approximately symmetric stable group estimators but also to asymmetric ones? 4) Furthermore, would the proposed idea of the TWB help in developing inferences on regression coefficients other than the intercept? A recent study developed by Cavaliere & Georgiev (2013) shows that by dummying out "large" innovations that exceed a certain threshold, an approach commonly

---

[14]However, the performance deteriorates when the index of stability is 1, and the scale parameter grows. This drawback is due to inadequate alternative for estimating the scale parameter and the re-scaling of the mode location when the scale becomes high.

implemented in applied econometric works, may help in testing finite-order autoregressive provided that the innovations are symmetric.

In the empirical econometric branch, a recent study by Ibragimov et al. (2013) shows that the tail indices of currency exchange rates in emerging and developing countries are considerably smaller than those of developed countries. In particular, they found that several emerging countries' exchange rates have infinite variances. Furthermore, they discovered that most developed countries have infinite fourth moments. According to Cont (2001) marginal distribution of the returns (or exchange currency) has an infinite fourth moment, then the sample autocorrelation functions (ACF) remains a consistent estimator of the theoretical ACF but with wider asymptotic confidence bands than those of the classical case whereby fourth moments are finite. Therefore, a natural question to ask is whether this current TWB algorithm may help in narrowing the bands of the asymptotic confidence bands of the ACF. Furthermore, it is interesting to device new inferential method to analyze autocorrelation functions for dependent and heavy-tailed time series (Mikosch & Starica 1998, Davis & Mikosch 1999). These applications are important not only for the infinite variance case but also for the case of infinite fourth moments as the standard asymptotic theory with normal convergence does not for autocorrelation functions of squared variables (e.g., returns) with tail indices smaller than 4 and infinite fourth moments.

In conclusion, the truncated wild bootstrap performs better in terms of both size and power than its main competitor the Parametric Bootstrap and the Wild bootstrap. In particular, when α is below one, unlike the parametric bootstrap, the truncated wild bootstrap sustain the power of the test. Furthermore, unlike the Wild bootstrap, the truncated wild bootstrap works for asymmetric distributions. Hence, it may be deduced based on Cornea-Madeira & Davidson's (2015) simulations experiments that the truncated wild bootstrap is better than subsampling and the *m* out of *n* bootstrap.

Chapter 2

A Frequency Domain Wild Bootstrap for Dependent Data

2.1 Introduction

In econometric applications, particularly in the time-series context, when a model is misspecified[1], the popular asymptotic theory tests are incapable of providing a useful tool for applied econometricians to test whether there exists a genuine or spurious link between variables[2]. Therefore, bootstrap inference in dependent data is typically suggested as a remedy for an incorrect or incomplete model specification. In particular, under such conditions, the bootstrap principle can be viewed as a vehicle for estimating the missing model components.

In general, bootstrap methods are designed to approximate moments or the distribution of a properly standardized estimator depending on observations from a time series model. Therefore, since observations are acknowledged to be generated from a temporal dependence structure, the object is to model not merely the marginal distribution of the sample data, but also its joint distribution. Such an application requires hard compromises. Unless the distribution possesses some basic regularities at a minimum stationarity, the sample reduces in effect to a single observation. Even in a nonparametric context, it is inevitably necessary to assume the joint distribution has sufficient structure to allow it to be reconstructed from the sample data.

---

[1]A model is misspecified when its functional form is incorrect, or some important variables are excluded.

[2]For example, Granger and Newbold (1974) regressed two sets of independent series such that one of these series follow an time-dependent linear process. The authors found that 93% of the time, using the usual F statistics, the test show a "relationship" exist between the independent series (Davidson 2000, p. 153-154). Although lots of asymptotic tests have been suggests in time series analysis, it is limited to the hypothetical structure of the time series model.

The various forms of block bootstrap, one form of the time-domain bootstrap procedures, represent one approach to performing this reconstruction. They are nonparametric in character, and mimic the sample dependence of the whole by the joint distributions of short segments. However, since these segments must be independently drawn with replacement and concatenated, the "joins" problem places a limitation on the effectiveness of the mimicry. Furthermore, as argued by Andrews (2002), the block bootstrap is much less effective than the nonparametric iid bootstrap. A leading alternative to the blocks method which is among the time-domain bootstrap procedures is the autoregressive sieve bootstrap, which is evidently well-adapted to capturing certain types of dependence, but equally fails to capture other types, as discussed in Kreiss et al. (2011), Bühlmann (1997). A third approach is to use an estimate of the spectral density of the series which involves the use of frequency domain to model the dependence of the observed time series observations, and a number of such methods are recounted in Kirch & Politis (2011) and references therein. These latter bootstrap methods have the advantage of not being reliant on a parametric dependence model.

The bootstrap method we propose in this chapter is related to the last-mentioned procedures, but has distinctive features of its own (see section 2.2, for comparison). While the bootstrap resampling takes place in the frequency domain in its basic form, it does not depend on the estimation of dependence features (see outline of bootstrap procedure 2.3.1), whether parametric or nonparametric. In this sense, it has more in common with the classic Efron (1979) approach for independent and identically distributed (i.i.d.) data. Moreover, while the bootstrap draws are conditionally independent, they have the special property that the periodogram is numerically identical to that of the original sample. Provided the data dependence is fully embodied in second moments, it is hard to see how the bootstrap could mimic the sample distribution any more precisely than this.

By the same token, the method is not applicable to inference problems relating to the periodogram points themselves or functions thereof, autocovariance and autoregressive parameters in particular; see Proposition 2.1 of Kirch & Politis (2011) for a related result. Our method stands in contrast to the extensive literature on inference on the periodogram and so-called ratio statistics[3], as surveyed in Lahiri (2003) Chapter 9, for example. The survey of Kreiss & Paparoditis (2011) likewise focuses chiefly on resampling of the periodogram, a procedure our method avoids. This class of procedures has a different motivation to our

---

[3]Since bootstrap periodogram ordinates is not finitely independent, the bootstrap method suggested by Franke & Härdle (1992) was ambiguous as to which class of ratio statistic it may be applied in general and specifically for processes that are non-Gaussian. Therefore, Dahlhaus & Janas (1996) have clarified which class of estimators makes such a periodogram ordinate bootstrap procedure valid. According to their paper, this class of statistics is called the *ratio-statistic*. In particular, the ratio-statistic class consists of ratios of spectral-mean estimates and the integrated periodogram. The usual moment estimator for autocorrelation is an appropriate example that lies in this class.

own. The method of Hidalgo (2003) shares our aim of making inferences on regression coefficients, but it involves casting the entire regression model to the frequency domain (see, for details, section 2.2). The so-called TFT methods of Kirch & Politis (2011) are closely related to ours, but entail kernel estimation of the spectral density. The method of Theiler et al. (1992) is also closely related, but resamples the phase-components of the distribution while holding the frequency magnitudes fixed, as estimated by the periodogram.

2.1.1   Organization of Chapter 2

Our chapter is organized as follows. Section 2.2 gives a general review of the literature for bootstrap methods in the context of hypothesis-testing in a time series analysis. This section gives more technical details than those outlined in the introduction and helps facilitate the arguments of the coming sections. Section 2.3 describes the proposed bootstrap algorithm and derives key statistical properties. Among these is the peculiarity that the bootstrap distribution of the series mean is degenerate. While this poses no problem for significance tests on slope coefficients in regressions with autocorrelated disturbances, for example, the method is unsuitable in its basic form for tests of location and tests for unit roots.

Section 2.4 describes a modified algorithm, which we call the augmented Fourier bootstrap (AFB), one that overcomes this limitation. Our implementation of the AFB entails fitting a response-surface to Monte Carlo-estimated performance measures and, as a preliminary, Section 2.5 describes and justifies the "warp-speed" Monte Carlo method which is used to accelerate the simulation of the bootstrap tests. The augmentation method itself is explained in Section 2.6. Then, Section 2.7 reports Monte Carlo evaluations of the proposed methods. We compare across various bootstrap methods along with a representative TFT method, with the moving block and stationary block bootstraps, with the Auto-regressive sieve bootstrap, and also with asymptotic inference. We also check on our method's performance in non-Gaussian data. Results are given for two cases, respectively, a static regression model with two regressors and autocorrelated disturbances, and tests for a unit root in data with correlated increments. Section 2.8 considers the multivariate case, and shows that the cross-periodogram and cross-autocovariance properties extend the univariate case as expected. Simulation of a bi-variate location test when the data are generated by a $VAR(1)$ with correlated shocks indicates that our simple augmentation formula can render good-size properties. Section 2.9 concludes, and Monte Carlo Tables are gathered in the section 2.10.

2.2   Literature Review

Any stationary process $\mathbf{X} = (X_t : t \in \mathbb{Z})$ may be decomposed into a sum of sinusoidal components (Chapter 10, Brockwell & Davis 2006). The analysis of stationary processes using their spectral representations is often regarded as the "frequency domain" analysis of time series. In cases the underlying process (or signal) is absolutely continuous, the power spectral density (or simply spectral density), $f(\lambda_j)$, is defined as the derivative of the spectrum distribution function (SDF), $F(\lambda)$. The SDF of an absolutely continuous time series spectrum $\mathbf{X}$ describes the amount of variance contributed to $\mathbf{X}$ by frequencies lying within a specified interval. To see this, let us introduce some mathematical preliminaries.

A spectral density can be derived from the spectral representation of a covariance function, and is mathematically represented by

$$f_{\mathbf{X}}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\omega}, \ \omega \in [0, 2\pi], \qquad \text{where} \qquad \gamma(k) = cov(X_0, X_k) \qquad (2.1)$$

(cf. Naidu 1995, p. 16-18).

In general, covariances and variances of observations are not known, due to the uncertainty in the data generating process (DGP) of the observations. Consequently, in the time-domain settings, practitioners often resort to estimation methods of such values to help them create inferences. Similarly, in the frequency-domain configuration, since spectral density is not known due to the unknown nature of the DGP, practitioners often estimate spectral density using periodograms. Thus, a periodogram, in the time series context, is a classical nonparametric estimator of the spectral density. For concreteness, it is common to define the periodogram $I(\lambda_j)$ at a Fourier frequency $\lambda_j := 2\pi j/n$ of the process $\mathbf{X}$, as the squared value of a discrete Fourier transform. Mathematically, it is given by

$$I_n(\lambda_{j-1}) := |a_{\mathbf{x}}(j)|^2 = \frac{1}{n} \left| \sum_{t=1}^{n} X_t e^{-i(t-1)\lambda_{j-1}} \right|^2$$

where $a_{\mathbf{x}}(j)$ is the *Discrete Fourier transform* (DFT) of the vector $\mathbf{x} := (X_1, X_2, \ldots, X_n)\prime$ and is equal to

$$a_{\mathbf{x}}(j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t e^{-i(t-1)\lambda_{j-1}} \qquad (2.2)$$

Observe that, under such a definition, the relationship between periodograms and variances is obvious. In particular, the sum of periodograms $\{I_n(\lambda_{j-1})\}_{j=1}^{n}$ is equal to the squared sum

of the process **X**

$$\sum_{j=1}^{n} I_n(\lambda_j) = \sum_{t=1}^{n} X_t^2 \qquad \text{and} \qquad I_n(0) = n\bar{X}_n^2$$

Based on the above discussion, it seems obvious that the inference based on spectral densities in the frequency-domain time series analysis is equivalent to the inferences of autocovariance in the time-domain time series analysis. This is because both the autocovariance function and spectral density contain the same information. Therefore, depending on the statistician's exercise of inference, frequency-based representation of the time series observations proves to be more illuminating than its counter-part. In particular, the bootstrap inference in mimicking distribution of temporal-dependence time series is a case in point.

The fact is that, under the assumption of a linear process, the sequence of periodograms[4] $\{I(\lambda_j)\}_{j=1}^{n}$ evaluated at a set of frequencies $\{\lambda_j\}_{j=1}^{n}$ asymptotically behaves as a sequence of independent and exponential random variables with a mean of spectral densities $\{f(\lambda_j)\}_{j=1}^{n}$ associated with the designated set of frequencies $\{\lambda_j\}_{j=1}^{n}$. Therefore, in the spectral representation, the periodogram may be observed as a series of independent but not identically distributed (i.n.i.d.) random variables. However, the periodograms have the same asymptotic form as exponential random variables but with different means. Therefore, the periodogram variables are considered to be a heteroscedastic sequence of random variables. This conclusion has motivated bootstrap theorists to apply classical bootstrap algorithms based on the periodograms (Kreiss & Paparoditis 2011, p. 365).

Inspired by the time series spectral analysis work of Hannan (1970) and Brillinger (1981), and by the simulation experiments of Hurvich & Zeger (1987), Franke & Härdle (1992) have exploited the asymptotic independence but heteroscedastic property of the periodograms to create an i.i.d. resampling bootstrap procedure. Motivated by the estimation of the spectral density function, Franke & Härdle (1992) devise confidence intervals for the spectral density based on appropriately defined frequency domain residuals. Hence, the bootstrap procedure's name, the *multiplicative residual bootstrap*. In this case, the multiplicative residual bootstrap is simply the classical bootstrap method applied to observations that are asymptotically independent. Therefore, for such a bootstrap method to be valid asymptotically, requires consistent estimates of the spectral density. Accordingly, Franke & Härdle (1992) rely on nonparametric kernel spectral density estimates[5].

The multiplicative residual bootstrap procedure is applied in the frequency-domain. Therefore, in this chapter and hereafter, all bootstrap procedures that are applied in the frequency-domain will be set to belong to the *frequency domain bootstrap* (FDB) class.

---

[4]A periodogram, in the time series context, is a classical nonparametric estimator of the spectral density.

[5]Although the periodogram is asymptotically unbiased, it is an inconsistent estimate of the spectral density, because its variance at each spectral frequency is proportional to the square of the density at these frequencies.

Moreover, since this bootstrap procedure is applied on the periodograms, this procedure belongs to another class, along with the FDB, known by *periodogram bootstrap* class (see, discussion Kreiss & Paparoditis 2011, p. 365).

The bootstrap procedure of Franke & Härdle (1992) has motivated different applications of inferences. For example, Hidalgo (2003) has used such a bootstrap algorithm for hypothesis testing in regression models. In particular, Hidalgo (2003) was interested in testing the null hypothesis:

$$\mathcal{H}_0 : \beta_1 = \beta_2 = 0 \tag{2.3}$$

of the linear model represented by:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + U_t, , \qquad \text{where} \qquad t = 1, \ldots, n \tag{2.4}$$

Hidalgo (2003) exploits the fact that the OLS estimator has both representation in the Time Domain and the Frequency Domain to perform inferences on the coefficients β_1 and β_2 in model 2.3. To see this, denote by $\boldsymbol{\beta}$ the column vector represented by the coefficients of the explanatory variables; hence, in terms of model 2.4 $\boldsymbol{\beta} := (\beta_1, \beta_2)'$. In this case, the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, in the frequency domain can be written as:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{a_X}^H \mathbf{a_X})^{-1} \mathbf{a_X}^H \mathbf{a_y} \\
&= \left(\sum_{j=1}^n a_\mathbf{x}(j)^H a_\mathbf{x}(j) \right)^{-1} \left(\sum_{j=1}^n a_\mathbf{x}(j)^H a_\mathbf{y}(j) \right)
\end{aligned}$$

where the superscript, H, stands for the conjugate transpose. $a_\mathbf{x}(j)$ is defined in 2.2, the ith row of $\mathbf{X}$ and $\mathbf{y}$ is simply the vector $[X_{1t}, X_{2t}]$ and the scalar y_t respectively, for $t = 1, \ldots, n$. $\mathbf{a_y}(n \times 1)$ is a column vector represented as

$$\{\mathbf{a_y}\}_j := a_\mathbf{y}(j), \qquad \text{where} \qquad j = 1, 2, \ldots, n \tag{2.5}$$

Remark: Observe that $a_\mathbf{x}(j)^H a_\mathbf{x}(j)$ and $a_\mathbf{x}(j)^H a_\mathbf{y}(j)$ represent the periodogram ordinate $I_n(\lambda_{j-1})$ and the cross-periodogram of the series $\{X_j\}_{i=1}^n$ and $\{y_j\}_{i=1}^n$ respectively.

Based on the frequency domain representation of the OLS estimator, Hidalgo (2003) shows that the bootstrap distribution of the F statistic for testing the null hypothesis 2.3

asymptotically works[6]. Therefore, using a different set-up of the multiplicative residual bootstrap, most periodogram bootstrap procedures are capable of testing null hypotheses of the form 2.3.

In general, periodogram bootstrap methods miss important features of the DGP; hence, restricting the classes of statistics to which it may be applied on successfully, however. For example, the asymptotic distributional characteristics of the sample-mean statistic cannot be written themselves by means of periodograms. Therefore, all available periodogram bootstrap procedures are incapable of testing the other null hypothesis, $\mathscr{H}_1$, defined by:

$$\mathscr{H}_1 : \beta_0 = 0 \tag{2.6}$$

Moreover, the mentioned methods are only applicable asymptotically for a few number of classes of periodogram statistic. Therefore, Kreiss & Paparoditis (2003) have linked a time-domain parametric and a frequency-domain nonparametric bootstrap to produce the so-called Auto-regressive-Aided Periodogram Bootstrap (AAPB). Unlike the multiplicative residual bootstrap, the AAPB mimics the complete covariance structure of the underlying time series. Since it involves both domains (time and frequency), the AAPB belongs to the class of bootstrap procedures known as *Hybrid Bootstrap Class*. Again, however, like the periodogram bootstrap, the AAPB is only applicable to statistics which may be written as functions of a periodogram (Kreiss & Paparoditis 2011, p. 367).

The time-domain steps of the AAPB procedure coincide with the time-domain bootstrap procedural steps of Kreiss (1988), known as the Auto-regressive sieve (AR-sieve) bootstrap. Therefore, the AAPB can be understood as an extension of AR-sieve bootstrap. The superiority of the AAPB over the AR-sieve and the available periodogram bootstrap can be summarized by the following two points. Firstly, the AAPB is able to correct features of the data which may not have been accounted for by the AR-sieve. In particular, if the underlying data do not stem from an auto-regressive model of an order less than the fixed fitted order p, the parametric AR-sieve would not be valid asymptotically. Secondly, the AAPB does not neglect the dependence-structure of the periodogram ordinates as the available periodogram bootstrap procedures (Kreiss & Paparoditis 2011, p. 367).

Preliminary simulation evidence not reported here, shows that there are not many differences between the AAPB and the AR-sieve when it comes to the hypothesis testing of coefficients in the form of 2.3. In fact, both the AAPB and AR-sieve have a very similar error in rejection probability (ERP) and power. Moreover, the AAPB requires much more

---

[6]The *bootstrap asymptotically works* means that the approximation error of the bootstrap distribution for the standardized distribution of the estimator converges to zero (in some meaningful sense) as the sample size increases to infinity.

computational time due to the additional steps of the frequency-domain algorithmic parts. Additionally, unlike the AAPB, and superior to other available time-domain bootstrap procedures (Bühlmann 1997), the AR-sieve bootstrap is capable of testing the null hypothesis of the form 2.6. Therefore, in our simulation experiments in section 2.7, we will only compare our proposed bootstrap with the AR-sieve, rather than the AAPB, because the former resembles the latter.

For concreteness, we will restate the AR-sieve bootstrap procedure for testing hypothesis 2.3 of model 2.4 by the following description.

Bootstrap Procedure 2.2.1 (AR-sieve).

1. *Compute the OLS residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$ of the model of interest*

2. *Specify the dependency structure among the computed OLS residuals by choosing a value for the lag length $\hat{p}(n)$ of the OLS residuals using the Akaike information criterion (AIC).*

3. *Fit a pth-order autoregressive model*

4. *Compute the OLS coefficients $\{\hat{\phi}_i\}_{i=1}^{\hat{p}(n)}$ using the Yule-Walker Equations by running the following regression model using the following*

$$\hat{\varepsilon}_t = \phi_1 \hat{\varepsilon}_{t-1} + \phi_2 \hat{\varepsilon}_{t-2} + \cdots + \phi_{\hat{p}(n)} \hat{\varepsilon}_{t-\hat{p}(n)} + u_t \tag{2.7}$$

   where $t = \hat{p}(n) + 1, \cdots, n$ and u_t are i.i.d. error terms.

5. *Compute the OLS residuals of the model $\{\hat{u}_t\}_{t=1}^{n-\hat{p}(n)}$.*

6. *Resample from $\{\hat{u}_t - \overline{\hat{u}_{n-\hat{p}(n)}}\}$ where $\overline{\hat{u}_f} = (f)^{-1} \sum_{t=1}^{f(n)} \hat{u}_t$, denote the bootstrapped sample by $\{u_t^*\}_{t=1}^{n-\hat{p}(n)+m}$ where m is only additional generated observations that are introduced for the sole purpose of removing start-up effects.*

7. *Generate the AR-Sieve ordinate of the errors as*

$$\varepsilon_t^* = \hat{\phi}_1 \varepsilon_{t-1}^* + \hat{\phi}_2 \varepsilon_{t-2}^* + \cdots + \hat{\phi}_{\hat{p}(n)} \varepsilon_{t-\hat{p}(n)}^* + u_{t-\hat{p}(n)}^* \tag{2.8}$$

   where $t = \hat{p}(n) + 1, \ldots, n + m$ and $\varepsilon_t^* = 0$ for $t = 1, 2, \ldots, \hat{p}(n)$).

8. *Compute the AR-Sieve bootstrap sample given by*

$$Y_t^* = \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \varepsilon_{t+m}^* \tag{2.9}$$

where $t = 1, 2, \ldots, n$.

In recent years, two new bootstrap procedures were introduced: the hybrid bootstrap of Jentsch & Kreiss (2010) and the time frequency toggle (TFT) bootstrap of Kirch & Politis (2011). Both procedures are superior to other available bootstrap methods in the time-domain context, because they are capable of producing a pseudo-series in the time domain that mimics the original dependence-structure as well as moments of the observed sample.

In this chapter, to distinguish between the hybrid bootstrap class and the hybrid bootstrap procedure of Jentsch & Kreiss (2010), we will refer to this particular bootstrap procedure as the Modified Auto-regressive Aided Periodogram bootstrap (MAAPB). This is because Jentsch & Kreiss's (2010) bootstrap, in the univariate case, involves the same sequential steps of the AAPB, but rather than applying the nonparametric pre-whitening correction function on the periodogram, it is applied on the Discrete Fourier Transform (DFT) of the series of interest. To generate the pseudo-series in the time domain, the MAAPB requires an additional step (not required by the AAPB) of applying the inverse-DFT on the corrected DFT (see, Jentsch & Kreiss 2010, p. 2326).

Although this bootstrap method is better in replicating time-series observations, in our unreported preliminary simulations, we notice two findings. Firstly, this bootstrap procedure is capable of making inferences on the coefficient β_1 and β_2 better than its predecessors' versions (AAPB and AR-sieve). Secondly, this bootstrap procedure does not yield a lower ERP or a higher power when testing the null hypothesis defined in 2.6. In addition to these findings, we found at a later stage and in other simulation experiments that testing the null hypothesis of the form described in 2.3 is almost equivalent to the residual-based bootstrap type in the time frequency toggle Bootstrap method (RB-TFT)[7]. Therefore, in our simulations in section 2.7, we will not consider the MAAPB in our simulation study.

Note that unlike the computation of the DFT, the computation of the periodogram causes an irretrievable loss of information. This makes any bootstrap method that involves the computation of a periodogram unsuitable for conducting inferences on a model's intercept coefficient, due to the lack of information the periodogram holds of a data-set of interest. This information has motivated Kirch & Politis (2011) to find a theoretical justification to the simulation experiments of Hurvich & Zeger (1987). Indeed, Kirch & Politis (2011) were able to find this explanation along with other important findings which create a new perspective of the frequency-domain bootstrap procedures.

Although represented as a bootstrap method, the TFT, in our view, describes a class of bootstrap procedures. In particular, the basic definition of this class is that any bootstrap

---

[7]This term will be explained later in the chapter

procedure that belongs to it—to generate pseudo-series— entails steps requiring back and forth jumps between time-domain and frequency-domain via the aid of DFT and its inverse[8].

Previous periodogram bootstrap theorists have employed in their proofs the well-known fact that the periodogram ordinates are asymptotically independent and exponentially distributed. However, in the TFT class of bootstrap procedure, Kirch & Politis (2011) exploited the fact that the *Fourier coefficients* defined by

$$\zeta_X(j) = T^{-1/2} \sum_{i=1}^{n} X_i \cos(-\lambda_j i) \tag{2.10}$$

$$\kappa_X(j) = T^{-1/2} \sum_{i=1}^{n} X_i \sin(-\lambda_j i) \tag{2.11}$$

are asymptotically independent and normally distributed, to formulate three different types of consistent bootstrap procedures: the residual-based bootstrap (RB-TFT), the wild bootstrap (WB-TFT), and the local bootstrap (LB-TFT). These bootstrap procedures are modifications of available periodogram bootstrap procedures. For example, the RB-TFT is very similar to the multiplicative residual bootstrap of Franke & Härdle (1992) but rather than bootstrapping the periodograms, Kirch & Politis (2011) bootstrap the DFT coefficients.

The LB-TFT is similar to the local bootstrap structure of Paparoditis & Politis (1999) whereby this bootstrap method entails no initial estimation of the spectral density. The local bootstrap of Paparoditis & Politis (1999) exploits the fact that the spectral density is smooth enough when the periodogram ordinates are asymptotically independent and exponentially distributed. Hence, the differences between two distinct periodogram ordinates (or the Fourier coefficients) with associated distinct frequencies are dependent on how far the two different frequencies are from one another. Therefore, bootstrapping based on frequencies that are locally close to the main frequency plays a key role in causing this bootstrap procedure to be consistent (Kreiss & Paparoditis 2011, p. 366).

The WB-TFT is identical to the parametric frequency-domain bootstrap proposal of Hurvich & Zeger (1987). This particular bootstrap, like the RB-TFT, requires estimation of the spectral density. This is because both bootstrap variants, the RB-TFT and WB-TFT, require normalization of the DFT points, so that asymptotically they behave as Gaussian i.i.d. random variables.

The currently available bootstrap procedures in the TFT class introduced by Kirch & Politis (2011) are considered to be an extension (or modification) of the periodogram bootstrap methods available for two reasons. Firstly, the three available bootstrap procedures

---

[8]For further description of the TFT class, please refer to Section 2.3.3

of the TFT class and the available periodogram bootstrap yield almost identical results when applied to statistics based on periodograms. Secondly, the bootstrap methods proposed by Kirch & Politis (2011) are applicable to statistics of interest that are not expressible by periodograms.

For the sake of clarity in formulating the algorithmic steps of these procedures in the context of hypothesis testing, we will restate the RB-TFT as well as the WB-TFT procedures in the following sequential steps:

Bootstrap Procedure 2.2.2 (RB-TFT and WB-TFT)**.**

1. *Compute the OLS residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$ of the model of interest.*

2. *Estimate the spectral density $\hat{f}_{\hat{\varepsilon}}(\omega)$ described by:*

$$\hat{f}_{\hat{\varepsilon}}(\omega) := \left(2\pi \sum_{j\in\mathbb{Z}} K\big(h^{-1}\omega j\big)\right)^{-1} \left(\sum_{j\in\mathbb{Z}} K\big(h^{-1}(\omega-\omega_j)\big)|a_{\hat{\varepsilon}}(j)|^2\right)$$

where $K(.)$ is a kernel function, $a_{\hat{\varepsilon}}$ is the DFT of the vector of residuals $\hat{\varepsilon}$ as described in equation (2.2), h is the bandwidth which usually depends on the sample size (usually estimated using Newey-West estimator), ωj is the Fourier frequency.

3. *Calculate the Fourier coefficients $\zeta_{\hat{\varepsilon}}(j)$ and $\kappa_{\hat{\varepsilon}}(j)$ using the fast Fourier transform (FFT) algorithm as described in equations (2.10) and (2.11), respectively.*

4. *Estimate the residuals of the real, as well as imaginary, part of the Fourier coefficients and put them together into a vector $\{\widetilde{s}_j\}_{j=1}^{2r}$ such that*

$$\widetilde{s}_j = \left(\pi\hat{f}_{\hat{\varepsilon}}(\omega_j)\right)^{-1/2}\zeta_{\hat{\varepsilon}}(j), \qquad and \qquad \widetilde{s}_{r+j} = \left(\pi\hat{f}_{\hat{\varepsilon}}(\omega_j)\right)^{-1/2}\kappa_{\hat{\varepsilon}}(j) \qquad (2.12)$$

where $j = 1,2,\ldots,r$ and $r = \lfloor(n-1)/2\rfloor$. Then standardize each of the values in $\{\widetilde{s}_j\}_{j=1}^{2r}$ and store them into a vector $\{s_j\}_{j=1}^{2r}$.

***Remark**: This step is not required by the WB-TFT. This is because this bootstrap procedure will resample from a standardized normal.*

5. *Compute the bootstrapped Fourier coefficients given by*

$$\zeta_{\hat{\varepsilon}}^*(j) = \left(\pi\hat{f}_{\hat{\varepsilon}}(\omega_j)\right)s_j^*, \qquad and \qquad \kappa_{\hat{\varepsilon}}^*(j) = \left(\pi\hat{f}_{\hat{\varepsilon}}(\omega_j)\right)s_{r+j}^* \qquad (2.13)$$

RB-TFT: s_j^* represents an i.i.d. sample drawn randomly with replacement from the sequence $\{s_j\}_{j=1}^{2r}$

WB-TFT: s_j^* represents an i.i.d. draw from a standard normal distribution

6. *Let $\zeta_{\hat{\varepsilon}}^*(n) = \kappa_{\hat{\varepsilon}}^*(n) = 0$; moreover, if n is even, let $\zeta_{\hat{\varepsilon}}^*(n/2) = \kappa_{\hat{\varepsilon}}^*(n/2) = 0$*

7. *Obtain a bootstrap sequence $\{\zeta_X^*(j)\}_{j=1}^r$ and $\{\kappa_X^*(j)\}_{j=1}^r$*

8. *Set the remaining bootstrap Fourier coefficients as*

$$\zeta_{\hat{\varepsilon}}^*(n-j) = \zeta_{\hat{\varepsilon}}^*(j), \qquad and \qquad \kappa_{\hat{\varepsilon}}^*(n-j) = -\kappa_{\hat{\varepsilon}}^*(j)$$

9. *Use the inverse FFT algorithm to transform the bootstrap Fourier coefficients $\zeta_{\hat{\varepsilon}}^*(j) + i\kappa_{\hat{\varepsilon}}^*(j)$ back into the time domain to receive the sequence $\{\hat{\varepsilon}_t^*\}_{t=1}^n$.*

10. *Compute the RB-TFT or WB-TFT bootstrap sample given by*

$$Y_t^* = \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \hat{\varepsilon}_t^* \qquad (2.14)$$

where $t = 1, 2, \ldots, n$.

In the above bootstrap procedure, it should be emphasised that the sequence $\{\hat{\varepsilon}_t^*\}_{t=1}^n$ has a mean identical to zero. Therefore, there will be no differences in the OLS estimation of the intercept coefficient β_0 across the bootstrap samples. Hence, this bootstrap method as is, is not capable of testing the null hypothesis in the form of 2.6. Kirch & Politis (2011) have already mentioned this drawback for such a bootstrap procedure to be qualified in producing a pseudo-series in the time domain. Therefore, as a remedy, they proposed the use of augmenting the missing information from another bootstrap procedure. In particular, they have suggested practitioners use different bootstrap algorithms, such as the AR-sieve and MBB, to mimic the sample mean of the time series and augment this result to the yielded result of the TFT-bootstrap sample.

To the best of our knowledge, we are not aware of any study that focuses on studying the proposal of Kirch & Politis (2011). Therefore, in this chapter, we will address the idea of augmentation by comparing the results of our bootstrap procedure inferences on testing the hypothesis 2.6 with that of the AR-sieve and the MBB. This comparison of the intercept parameter gives an indication of whether their suggestion of such a procedure works or not. Furthermore, in our simulation experiment in section 2.7, we opt to compare our new proposed bootstrap method with only the RB-TFT type out of the other available TFT

methods. The reason for this is based on the intuition that the RB-TFT and the LB-TFT share the common underlying idea that the FFT ordinates are i.i.d.; hence, they are similar. Additionally, although we believe that the Fourier coefficients are i.i.d., it is hard to believe in the finite sense ("non-asymptotically") that the Fourier coefficients are Gaussian random variables. Therefore, the RB-TFT outperform the WB-TFT since it does not force the time-domain bootstrap sample paths to be Gaussian (see, comparison sections, Kirch & Politis 2011, p. 10).

2.3 The Fourier wild bootstrap

2.3.1 The algorithm

Let $x = (x_1, \ldots, x_n)'$ denote the random n-vector whose joint distribution is to be modelled by the bootstrap. We assume this to be a finite realization of a covariance stationary process with population mean of zero. The procedure we propose, which we refer to by the acronym FWB, consists of three basic steps:

1. Compute the discrete Fourier transform (DFT)

$$z = Qx \tag{2.15}$$

   where Q $(n \times n)$ denotes the unitary symmetric Fourier matrix with elements

   $$q_{jt} = n^{-1/2} e^{-2\pi \mathrm{i} j(t-1)/n}$$

   for $t = 1, \ldots, n$ and $j = 0, \ldots, n-1$, where $\mathrm{i} = \sqrt{-1}$.

Then, for replications $b = 1, \ldots, B$:

2. Apply the Rademacher wild bootstrap to z, switching signs (of the real and imaginary parts together) with probability $\frac{1}{2}$. Symbolically,

   $$z_b^* = W_b z$$

   where $W_b = \mathrm{diag}(w_b)$ and $w_b = (w_{b0}, \ldots, w_{b,n-1})'$ is a n-vector of independent random draws with

   $$P(w_{bj} = 1) = P(w_{bj} = -1) = \frac{1}{2}.$$

   Remark: To simulate bootstrap sample in the frequency domain, each of the bootstrap methods used in earlier research have focused on simulating only the first half of the

DFT of the original sample. Observe that since x is real-valued, a well-known property of the DFT z is that $\mathrm{Re}(z_{n-j}) = \mathrm{Re}(z_j)$ and $\mathrm{Im}(z_{n-j}) = -\mathrm{Im}(z_j)$ for $j = 1, \ldots, n/2 - 1$ (n even) or $j = 1, \ldots, (n-1)/2$ (n odd). In this case, the inverse DFT $Q^\dagger z$ is real valued for any z of this form. If one wishes to preserve this mathematical characteristic, it would be better to use a special case of W_b, $\tilde{W}_b$ constructed so that $w_{b,n-j} = w_{bj}$, and hence the sign pattern is preserved. However, as will be discussed later in the chapter, due to the construction of this bootstrap algorithm, there is no superiority in using the special case ($\tilde{W}_b$) over the general one (W_b).

3. Create the bootstrap draw in the time domain by

$$x_b^* = \mathrm{Re}(Q^\dagger z_b^*) + \mathrm{Im}(Q^\dagger z_b^*) \tag{2.16}$$

where $Q^\dagger$ denotes the conjugate transpose of Q.

Notes:

- While formulae (2.15) and (2.16) show the exact form of the Fourier transformation and its inverse, in practice a fast Fourier transform (FFT) algorithm can be used to obtain z and the corresponding inverse[9].

- It may not be obvious why we adopt the third step of using the sum of the real and imaginary parts of the inverse DFT to define the draw. However, as we explain in the following section, this tactic is crucial to the properties of the method.

- Following our remark in Step 2, observe that the inverse transformation of the DFT of the bootstrap sample z_b^* does not yield, in general, a real-valued vector. In particular, $Q^\dagger z_b^*$ is, in general, complex valued. Therefore, to show that there is no much gain of using the special case over the general case, we have also considered the alternative bootstrap draw procedure defined by the form

$$\tilde{x}_b^* = Q^\dagger \tilde{z}_b^* \tag{2.17}$$

where $\tilde{z}_b^*$ is defined as

$$\tilde{z}_b^* = \tilde{W}_b z \tag{2.18}$$

In this case, equation (2.17) represents the special case of (2.16) when the imaginary component is zero.

---

[9]The reason for using the FFT rather than the DFT lies in the circularity property of the linear transformation this will be discussed later in the chapter.

Although the conditional distributions of (2.16) and (2.17) are necessarily different, since they depend on different numbers of Rademacher draws, the following section shows that all the relevant characteristics of the first form are shared by the second, which is merely a special case. The restricted distribution $\tilde{W}_b$ confers no apparent advantage over W_b in terms of the properties of the algorithm, and additional bootstrap draws are evidently required to attain the same precision. Moreover, (2.17) is relevant exclusively to real-valued data and, while this is the usual case in the econometric context, the unrestricted formulation (2.16) is easily generalized to resample a complex-valued series. It is this form which is used in our simulation experiments reported in Section 2.7.

2.3.2 Properties of the FWB Algorithm

Begin by writing

$$Q = A + iB, \qquad \text{and} \qquad Q^\dagger = A - iB$$

where A and B are both $n \times n$ symmetric matrices defined element-wise by

$$\{A\}_{ij} = n^{-1/2} \cos(2\pi ij/n), \qquad \text{and} \qquad \{B\}_{ij} = n^{-1/2} \sin(2\pi ij/n)$$

for $i, j = 0, \ldots, n-1$, letting $i = t - 1$ for notational convenience. A and B are mutually orthogonal, with $AB = BA = 0$, since for each pair $i, j = 0, \ldots, n-1$. That is, that

$$\begin{aligned}
\{AB\}_{ij} &= \frac{1}{n} \sum_{k=0}^{n-1} \cos(2\pi ik/n) \sin(2\pi jk/n) \\
&= \frac{1}{2n} \sum_{k=0}^{n-1} \sin(2\pi(i+j)k/n) + \frac{1}{2n} \sum_{k=0}^{n-1} \sin(2\pi(i-j)k/n) \\
&= 0.
\end{aligned}$$

As in a number of similar identities to be determined subsequently, the sums vanish here since the terms are either zero, or equal and opposite in pairs. Note that the unitary property of Q implies

$$QQ^\dagger = Q^\dagger Q = AA + BB = I. \tag{2.19}$$

Consider the structure of these matrices in more detail. Note that if $t \in \mathbb{Z}$, then

$$\sum_{k=0}^{n-1} \cos\left(\frac{2tk\pi}{n}\right) = \begin{cases} n, & \text{if } t = mod(n) \\ 0, & \text{otherwise} \end{cases}$$

and

$$\sum_{k=0}^{n-1} sin(\frac{2tk\pi}{n}) = 0$$

holds. Furthermore, note that

$$\{AA\}_{ij} = \frac{1}{n}\sum_{k=0}^{n-1} cos(\frac{2ik\pi}{n})cos(\frac{2jk\pi}{n})$$

$$= \frac{1}{2n}\sum_{k=0}^{n-1}\left(cos(\frac{2(i-j)k\pi}{n}) + cos(\frac{2(i+j)k\pi}{n})\right).$$

These facts and standard trigonometric identities imply in particular that

$$\{AA\}_{ij} = \begin{cases} 1, & \text{if } i = j = 0 \\ 0.5, & \text{if } i = j \text{ and } i = n-j, \ j > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (2.20)$$

for n odd, and

$$\{AA\}_{ij} = \begin{cases} 1, & \text{if } i = j = 0, \text{ and } i = j = n/2 \\ 0.5, & \text{if } i = j \text{ and } i = n-j, \ j \neq 0 \text{ and } j \neq n/2 \\ 0, & \text{otherwise} \end{cases} \qquad (2.21)$$

for n even. Note that, in either case, $AAA = A$. B has a structure comparable to A except where the rows (columns) labelled 0 and $n/2$ (n even) are zero, and where the pairs of rows (columns) labelled j and $n-j$ are equal in magnitude but opposite in sign. The rank of B is $(n-1)/2$ (n odd) or $n/2-1$ (n even). For n odd,

$$\{BB\}_{ij} = \begin{cases} 0, & \text{if } i = j = 0 \\ 0.5, & \text{if } i = j > 0 \\ -0.5, & \text{if } i = n-j, \ j > 0 \\ 0, & \text{otherwise,} \end{cases} \qquad (2.22)$$

for n even,

$$\{BB\}_{ij} = \begin{cases} 0, & i = j = 0 \text{ and } i = j = n/2 \\ 0.5, & i = j, \, j \neq 0 \text{ and } j \neq n/2 \\ -0.5, & i = n - j, \, j \neq 0 \text{ and } j \neq n/2 \\ 0, & \text{otherwise,} \end{cases} \tag{2.23}$$

and $BBB = B$. Also note the direct confirmation of (2.19) from the pairs (2.20)+(2.22), or (2.21)+(2.23), as the case may be.

In the light of these facts, consider the steps of the proposed bootstrap draw. For ease of notation, let the subscript b denoting the draw be henceforth implicit. The first Fourier transform yields

$$z = Ax + iBx \tag{2.24}$$

and

$$z^* = WAx + iWBx \tag{2.25}$$

The inverse Fourier transform leads to

$$Q^\dagger z^* = A(WAx + iWBx) - iB(WAx + iWBx) = (U + iV)x \tag{2.26}$$

Hence, U and V are defined as

$$U := AWA + BWB \qquad \text{and} \qquad V := AWB - BWA \tag{2.27}$$

Let R be defined as

$$R := U + V \tag{2.28}$$

the FWB draw has the form

$$x^* = Rx. \tag{2.29}$$

We may compare (2.29) with other bootstrap formulae, having this form with different randomly drawn matrices. In the standard Efron's (1979) bootstrap, R is constructed with n columns randomly drawn from the identity matrix, with replacement. In the wild bootstrap R is diagonal, with randomly drawn diagonal elements. The various block bootstrap schemes construct R from random blocks of consecutive columns of the identity matrix, while the sieve autoregressive method forms R as an Efron's (1979) matrix post-multiplied by an upper triangular matrix of moving-average weights.

For $i, j = 0, \dots, n-1$ the elements of U and V take the form

$$\{U\}_{ij} = \sum_{k=0}^{n-1} w_k \Phi_{ijk} \qquad \text{and} \qquad \{V\}_{ij} = \sum_{k=0}^{n-1} w_k \Psi_{ijk}$$

where

$$
\begin{aligned}
\Phi_{ijk} &= n^{-1}\left[\cos(2\pi ik/n)\cos(2\pi jk/n) + \sin(2\pi ik/n)\sin(2\pi jk/n)\right] \\
&= n^{-1}\cos(2\pi(i-j)k/n)
\end{aligned}
\tag{2.30}
$$

and

$$
\begin{aligned}
\Psi_{ijk} &= n^{-1}\left[\cos(2\pi ik/n)\sin(2\pi jk/n) - \sin(2\pi ik/n)\cos(2\pi jk/n)\right] \\
&= -n^{-1}\sin(2\pi(i-j)k/n).
\end{aligned}
\tag{2.31}
$$

Thus,

$$\Phi_{i+1,j+1,k} = \Phi_{ijk}, \qquad \text{and} \qquad \Psi_{i+1,j+1,k} = \Psi_{ijk}, \tag{2.32}$$

for $0 \leq i, j < n-1$ and each $k = 0, \dots, n-1$. Since it is also the case that $\cos(2\pi mk/n) = \cos(2\pi(m-n)k/n)$ and $\sin(2\pi mk/n) = \sin(2\pi(m-n)k/n)$ for $0 \leq m \leq n$, we also find

$$\Phi_{i+1,0,k} = \Phi_{i,n-1,k}, \qquad \text{and} \qquad \Psi_{i+1,0,k} = \Psi_{i,n-1,k} \tag{2.33}$$

for $0 \leq i < n-1$. In other words, U and V are both Toeplitz matrices[10], having the circulant property[11] that each row reproduces the one above with a shift of one place to the right, with the last column entry wrapping around to the first position. Therefore, the coordinates of the bootstrap draws are linear combinations of the sample series having random weights with this circulant form. This makes it easy to see how the bootstrap draws inherit the autocorrelation structure of the sample.

Lemma 2.3.1. *Let ι ($n \times 1$) denote the column of ones; and let U and V be as defined in equation (2.27), then the following properties hold:*
1) $U\iota = w_0\iota$; 2) $V\iota = 0$; 3) $U'V = 0$; 4) $U'U + V'V = I$.

---

[10]A Toeplitz matrix, in linear algebra, is a square matrix in which each descending diagonal from left to right is constant. The general form of an $n \times n$ Toeplitz matrix is represented by $T_n = [r_{k,j}; k, j = 0, 1, \dots, n-1]$ where $r_{k,j} = r_{k-j}$. For more details consult Gray (2006)

[11]A Circulant matrix is a special case of a Toeplitz matrix whereby every row of the matrix is a right cyclic shift of the row above it. "In numerical analysis, circulant matrices are important because they are diagonalized by a discrete Fourier transform, and hence linear equations that contain them may be quickly solved using a fast Fourier transform." (see, Davis 2012).

Proof. Parts (1) and (2) follow directly from the facts that $A\iota = (\sqrt{n}, 0, \ldots, 0)'$ and $B\iota = 0$. For Part (3), using the facts that A and B are symmetric and orthogonal, we obtain

$$U'V = AWAAWB - BWBBWA.$$

Observe that WA is a matrix obtained from A by changing the signs of certain rows, and AW is its transpose, with certain columns having a changed sign. $WAAW$ is therefore a matrix defined in the same way as AA in (2.20) or (2.21), except that a number of the non-diagonal elements may have changed signs. Specifically, since $w_i^2 = 1$,

$$\{WAAW\}_{ij} = \begin{cases} 1, & i = j = 0 \text{ and } i = j = n/2 \ (n \text{ even}) \\ 0.5, & i = j, \ j \neq 0 \text{ and } j \neq n/2 \ (n \text{ even}) \\ 0.5 w_i w_j, & i = n - j \\ 0, & \text{otherwise} \end{cases}$$

It follows that the matrix $AWAAW$ is defined in the same manner as A, except that certain pairs of columns may be replaced by zero columns. Thus, for $j \neq 0$ and $j \neq n/2$ (n even), $\{A\}_{ij} = \{A\}_{i,n-j}$ and hence

$$\begin{aligned} \{AWAAW\}_{ij} &= \sum_{k=0}^{n-1} \{A\}_{ik} \{WAAW\}_{kj} \\ &= \{A\}_{ij} \{WAAW\}_{jj} + \{A\}_{i,n-j} \{WAAW\}_{n-j,j} \\ &= \{A\}_{ij} \left[\{WAAW\}_{jj} + \{WAAW\}_{n-j,j} \right] \\ &= \tfrac{1}{2} \left(1 + w_{n-j} w_j \right) \{A\}_{ij}. \end{aligned}$$

The rule is that the replacement by zeros of the jth and $(n-j)$th columns occurs if $w_j w_{n-j} = -1$. Further,

$$\begin{aligned} \{AWAAWB\}_{ij} &= \frac{1}{2} \sum_{k=0}^{n-1} (1 + w_k w_{n-k}) \{A\}_{ik} \{B\}_{kj} \\ &= \frac{1}{2n} \sum_{k=0}^{n-1} (1 + w_k w_{n-k}) \cos(2\pi i k / n) \sin(2\pi j k / n) \\ &= \frac{1}{4n} \sum_{k=0}^{n-1} \left[\sin\left(2\pi k \left(i + j \right) / n \right) + \sin\left(2\pi k \left(i - j \right) / n \right) \right] \\ &\quad + \frac{1}{4n} \sum_{k=0}^{n-1} w_k w_{n-k} \left[\sin\left(2\pi k \left(i + j \right) / n \right) + \sin\left(2\pi k \left(i - j \right) / n \right) \right]. \qquad (2.34) \end{aligned}$$

Note that the first sum of terms in the last member of (2.34) vanishes since the summands are equal and opposite in pairs, with the kth term cancelling the $(n-k)^{th}$ for $k \neq 0$ (n odd) and for $k \neq 0, n/2$ (n even). However, the sign of $w_k w_{n-k}$ is invariant under replacement of k by $n-k$. Hence, the second set of summands also cancels in pairs, in the identical manner. We conclude that $AWAAWB = 0$. The same form of argument shows that $BWBBW$ matches B apart from some column replacements by zero columns, according to the same rule, and that $BWBBWA = 0$. This concludes the proof of (3).

To prove part (4), note that the identities $AB = 0$, $AA + BB = I$ and $WW = I$ imply respectively

$$U'U = AWAAWA + BWBBWB$$
$$= I - AWBBWA - BWAAWB$$

and

$$V'V = AWBBWA + BWAAWB.$$

$\square$

Based on the above lemma, one may infer the following remarks:

1. The matrix R in equation (2.29) has two important properties.

   - The sum of the bootstrap series matches the sum either x or $-x$ with equal probability $1/2$, depending on the value of w_0.

     $$\iota' x^* = \iota' R x = w_0 \iota' x \qquad (2.35)$$

   - The matrix R is an Orthogonal matrix, i.e. $R'R = I_n$.

2. Based on the properties of the matrix R, it is easy to see that the sum of the square of the bootstrap series generated by the FWB (see, equation (2.29)) is equivalent to the sum of square of the original series. That is, that,

   $$x^{*'} x^* = x' x. \qquad (2.36)$$

   Hence, both the bootstrap series and the original series share the same sum of periodograms.

3. Observe that if the sample data are expressed in mean deviation form such that $\iota' x = 0$, as it is usually the case in applications (residuals have zero sum), the conditional

distributions of the mean and variance of the bootstrap draws are therefore both degenerate, being equal with probability one to the sample mean and sample variance respectively. We explore the implications of these facts in the sequel.

4. If we replace W by the special case $\tilde{W}$, as defined following equation (2.18), and so define $\tilde{U}$ and $\tilde{V}$ as the obvious variants of (2.27), it is easy to verify from the forms of A and B that $A\tilde{W}B = B\tilde{W}A = 0$ and hence $\tilde{V} = 0$. This verifies the earlier assertion that the imaginary component of the inverse DFT vanishes, and $\tilde{x}^* = \tilde{U}x$. The following bootstrap properties depend on the forms of A and B and hold for arbitrary choices of W, including $\tilde{W}$ in particular. Since $\tilde{V} = 0$ is merely a special case, they hold for $\tilde{x}^*$ just as for x^*, and therefore do not require separate statements. For compactness we leave the alternative cases implicit and refer only to W in formulae, noting that the result of substituting $\tilde{W}$ is to eliminate the imaginary terms.

Theorem 2.3.2. *The periodogram of a FWB draw is identical with that of the sample series.*

Proof. For any vector a, let the notation $a^{\circ 2}$ and $|a|$ denote the element-wise square (Hadamard product with itself) and absolute value vectors with the same dimension as a, respectively. Then the periodogram points of x can be written as

$$\frac{|z|^{\circ 2}}{2\pi} = \frac{(Ax)^{\circ 2} + (Bx)^{\circ 2}}{2\pi}$$

The periodogram points of the bootstrap draw (2.28) take the corresponding form

$$\frac{(Ax^*)^{\circ 2} + (Bx^*)^{\circ 2}}{2\pi}$$

where

$$Ax^* = AAW(A+B)x = P_A x \qquad (2.37)$$

and

$$Bx^* = BBW(B-A)x = P_B x \qquad (2.38)$$

the second equalities defining P_A and P_B. Replacing W by $\tilde{W}$, note that these formulae become $\tilde{P}_A = AA\tilde{W}A$ and $\tilde{P}_B = BB\tilde{W}B$. Referring to the formulae in (2.20) and (2.21), we deduce that, except for the cases $j = 0$ and $j = n/2$ (n even), the rows of the matrix $AAWA$ have three possible forms. If $w_j = w_{n-j} = 1$, rows j and $n-j$ match rows j and $n-j$ of A. If $w_j = w_{n-j} = -1$, rows j and $n-j$ match the corresponding rows of $-A$. And, if the signs of w_j and w_{n-j} are different, the corresponding pairs of rows are zero. Rows 0 and $n/2$ (n

even) match the corresponding rows of A or $-A$, with the signs taken from w_0 and $w_{n/2}$ (n even), respectively.

On the other hand, the matrix $AAWB$ has zeros for rows j and $n-j$ if $w_j = w_{n-j} = \pm 1$. It has jth and $n-j$th rows matching those of B, if $w_j = 1$ and $w_{n-j} = -1$, and matching those of $-B$ if $w_j = -1$ and $w_{n-j} = 1$. Rows 0 and $n/2$ (n even) are zero, matching those of B. Notice how the zero rows arise in complementary positions in the two matrices.

Putting these results together, we can set out the following table where the notation identifies the jth rows of the matrices in question, for $j \neq 0$ and $j \neq n/2$ (n even), depending on the signs of the Rademacher pairs:

$$\{P_A\}_{j\cdot} = \begin{cases} \{A\}_{j\cdot} & \text{if } w_j = w_{n-j} = 1, \\ \{-A\}_{j\cdot} & \text{if } w_j = w_{n-j} = -1, \\ \{B\}_{j\cdot} & \text{if } w_j = 1 \text{ and } w_{n-j} = -1, \\ \{-B\}_{j\cdot} & \text{if } w_j = -1 \text{ and } w_{n-j} = 1. \end{cases} \tag{2.39}$$

The last two of these cases do not arise when w_j is replaced by $\tilde{w}_j$. Also, $\{P_A\}_{0\cdot} = w_0\{A\}_{0\cdot}$, and $\{P_A\}_{n/2\cdot} = w_{n/2}\{A\}_{n/2\cdot}$ (n even). The corresponding analysis of the matrices $BBWA$ and $BBWB$, making use of the formulae in (2.22) and (2.23), yields

$$\{P_B\}_{j\cdot} = \begin{cases} \{B\}_{j\cdot} & \text{if } w_j = w_{n-j} = 1, \\ \{-B\}_{j\cdot} & \text{if } w_j = w_{n-j} = -1, \\ \{-A\}_{j\cdot} & \text{if } w_j = 1 \text{ and } w_{n-j} = -1, \\ \{A\}_{j\cdot} & \text{if } w_j = -1 \text{ and } w_{n-j} = 1 \end{cases} \tag{2.40}$$

and also $\{P_B\}_{0\cdot} = 0'$, and $\{P_B\}_{n/2\cdot} = 0'$ (n even).

After squaring has eliminated the negative signs, the implication of (2.39) is that the elements of $(P_A x)^{\circ 2}$ match either corresponding elements of $(Ax)^{\circ 2}$ or the corresponding elements of $(Bx)^{\circ 2}$, depending on whether the Rademacher pairs match or differ. (They of course cannot differ for $\tilde{P}_A$ and $\tilde{P}_B$.) The elements of $(P_B x)^{\circ 2}$ on the other hand, according to (2.40), are the complementary cases from $(Bx)^{\circ 2}$ and $(Ax)^{\circ 2}$. In other words, if an element of $(P_A x)^{\circ 2}$ matches that of $(Ax)^{\circ 2}$, the corresponding element of $(P_B x)^{\circ 2}$ matches that of $(Bx)^{\circ 2}$. If an element of $(P_A x)^{\circ 2}$ matches that of $(Bx)^{\circ 2}$, the corresponding element of $(P_B x)^{\circ 2}$ matches that of $(Ax)^{\circ 2}$. It follows that

$$(P_A x)^{\circ 2} + (P_B x)^{\circ 2} = (Ax)^{\circ 2} + (Bx)^{\circ 2}$$

and the theorem follows directly. $\qquad \square$

Corollary 2.3.2.1. *The periodogram of x^* in (2.29) matches that of the inverse transform in (2.26).*

Proof. The DFT of (2.26) is identical with z^* in (2.25), noting that

$$
\begin{aligned}
(A + \mathrm{i}B) & [(AWA + BWB) + \mathrm{i}(AWB - BWA)] \\
& = (AA + BB)WA + \mathrm{i}(AA + BB)WB \\
& = W(A + \mathrm{i}B)
\end{aligned}
\tag{2.41}
$$

$\square$

Next, let $\mathbb{E}^*(.)$ denote the expected value under the bootstrap distribution of the random elements $w_0, \ldots, w_{n-1}$, conditional on the sample.

Theorem 2.3.3. *Let x_i denote the ith element of x, for $i = 0, \ldots, n-1$, and let x_i^* denote the corresponding element of the FWB draw. For $m \geq 0$ and $i \geq m$,*

$$
\mathbb{E}^*(x_i^* x_{i-m}^*) = \frac{1}{n} \sum_{j=m}^{n-1} x_j x_{j-m}
$$

Proof. Let $\Upsilon_{ijk} = \Phi_{ijk} + \Psi_{ijk}$ from (2.30) and (2.31), so that

$$
x_i^* = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} w_k \Upsilon_{ijk} x_j
$$

and for $i \geq m$,

$$
x_i^* x_{i-m}^* = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \sum_{j'=0}^{n-1} \sum_{k'=0}^{n-1} w_k w_{k'} \Upsilon_{ijk} \Upsilon_{i-m,j'k'} x_j x_{j'}.
$$

Since

$$
\mathbb{E}^*(w_k w_{k'}) = \begin{cases} 1, & \text{if } k = k' \\ 0, & \text{if } k \neq k' \end{cases}
$$

we have

$$
\begin{aligned}
\mathbb{E}^*(x_i^* x_{i-m}^* | x) &= \sum_{j=0}^{n-1} \sum_{j'=0}^{n-1} x_j x_{j'} \sum_{k=0}^{n-1} \Upsilon_{ijk} \Upsilon_{i-m,j',k} \\
&= \sum_{j=m}^{n-1} x_j x_{j-m} \sum_{k=0}^{n-1} \Upsilon_{ijk} \Upsilon_{i-m,j-m,k} + \sum_{j=0}^{n-1} \sum_{j' \neq j-m} x_j x_{j'} \sum_{k=0}^{n-1} \Upsilon_{ijk} \Upsilon_{i-m,j',k} \\
&= \frac{1}{n} \sum_{j=m}^{n-1} x_j x_{j-m}
\end{aligned}
$$

Here, the third equality makes use of the facts that

$$
\begin{aligned}
\sum_{k=0}^{n-1} \Upsilon_{ijk} \Upsilon_{i-m,j-m,k} &= \frac{1}{n^2} \sum_{k=0}^{n-1} \left[\cos^2(2\pi(i-j)k/n) + \sin^2(2\pi(i-j)k/n) \right] \\
&\quad - 2 \frac{1}{n^2} \sum_{k=0}^{n-1} \cos(2\pi(i-j)k/n) \sin(2\pi(i-j)k/n) \\
&= \frac{1}{n}
\end{aligned}
\tag{2.42}
$$

in view of (2.32) and $\cos^2 + \sin^2 = 1$, and also that for each i and $j' \neq j - m$,

$$
\begin{aligned}
\sum_{k=0}^{n-1} \Upsilon_{ijk} \Upsilon_{i-m,j'k} &= \sum_{k=0}^{n-1} \cos(2\pi(i-j)k/n) \cos(2\pi(i-m-j')k/n) \\
&\quad + \sum_{k=0}^{n-1} \sin(2\pi(i-j)k/n) \sin(2\pi(i-m-j')k/n) \\
&\quad - \sum_{k=0}^{n-1} \cos(2\pi(i-j)k/n) \sin(2\pi(i-m-j')k/n) \\
&\quad - \sum_{k=0}^{n-1} \sin(2\pi(i-j)k/n) \cos(2\pi(i-m-j')k/n) \\
&= \sum_{k=0}^{n-1} \cos(2\pi(m-j+j')k/n) - \sum_{k=0}^{n-1} \sin(2\pi(2i-m-j-j')k/n) \\
&= 0,
\end{aligned}
\tag{2.43}
$$

noting that $m - j + j'$ and $2i - m - j - j'$ are always nonzero integers.   □

Note how the bootstrap draw must be defined by (2.28) to achieve this result. The real and imaginary parts in (2.26) both have the circulant property producing the correct autocorrela-

tion structure of a draw, but the associated scale parameters are individually too small. The one exception is of course the case $W = \tilde{W}$ so that the imaginary part is zero. If the sample data are expressed in mean deviation form, Theorem 2.3.3 gives the autocovariance function of the bootstrap.

These results show that the bootstrap distribution is closely affiliated with the sample distribution and, for some purposes, too closely affiliated to be useful. As has been noted by Kreiss & Paparoditis (2011), it is clearly not feasible to use this bootstrap to analyse the distribution of functionals of the periodogram points. This is an issue to which we return in the next section. However, although their conditional distribution has restricted rank, there is no implication that the bootstrap draws are statistically dependent. In particular,

Theorem 2.3.4. *If W_a and W_b are independent Rademacher draws, defining R_p in (2.28) and $x_p^* = R_p x$ for $p = a, b$ then*

$$\mathbb{E}^*(x_a^* x_b^{*\prime}) = 0$$

The same result holds for $\tilde{R}_p$.

Proof. Rearranging (2.29) as

$$x^* = Rx = (A - B)WAx + (A + B)WBx$$

shows that $x_a^* x_b^{*\prime}$ is the sum of four rank 1 matrices, whose elements are sums of conditionally fixed terms weighted by random coefficients of the form $w_{aj} w_{bk}$ for $j, k = 1, \ldots, n$. Since w_{aj} and w_{bk} are independent random variables with zero means, the theorem follows directly. $\square$

Since all the w_k have even moments equal to 1 and odd moments equal to 0, similar reasoning shows that all higher-order co-moments also vanish. Moreover, Theorem 2.3.3 and the FWB sample (represented by equation (2.29)) have the following implication.

Theorem 2.3.5. *If the process generating the sample x satisfies the conditions of the central limit theorem, the FWB draws are asymptotically jointly Gaussian series, with an autocovariance function*

$$\gamma_m = \operatorname*{plim}_{n \to \infty} \frac{1}{n} \sum_{t=m+1}^{n} (x_t - \bar{x})(x_{t-m} - \bar{x}), , \qquad where \qquad m \geq 0.$$

Proof. Conditional on W, the bootstrap draw is a series composed of linear combinations of the sample data with weights (the rows of R in (2.28)) summing to a fixed value, either $+1$ or -1, and which are trigonometric functions of the time index bounded absolutely by

$n^{-1/2}$. This is therefore a straightforward application of the cited central limit theorem for dependent processes.   □

Discussion:

Theorem 2.3.5 recommends the FWB method to be applied to samples that are stationary with Gaussian characteristics (not necessarily Gaussian themselves), particularly when higher moments of the data, as well as the autocorrelation structure, influence the distribution of sample statistics. Recall that, in regression models, the OLS estimators are still consistent and unbiased even when the assumption of no autocorrelation is violated. This is because these properties depend on other assumptions that are still not violated; namely: 1) the error terms are homoskedastic and have a mean of zero; 2) the explanatory variables are nonstochastic, i.e. fixed in repeated samples (cf. Baltagi 2011, p. 112). Since this is the case in every bootstrap sample drawn, the FWB method should be robust to the form of the shock distribution, subject only to the existence of the variance.

Consider, for example, the properties of this method in the context of significance tests in a regression model, where the disturbances are serially dependent while distributed independently of the regressors. As noted, under the usual conditions of regularity, the bootstrap draws are jointly Gaussian in large samples. The bootstrap t statistics are likewise asymptotically normal, although if the conventional variance formula is used in the presence of residual autocorrelation, these distributions are not asymptotically pivotal[12]. The preferred procedure is, therefore, to use a heteroskedasticity and autocorrelation-consistent (HAC) variance estimator in the construction of test statistics. The correct asymptotic size and consistency of the bootstrap t-tests then follow by standard reasoning from the coincidence of the sample and bootstrap pivotal distributions.

Originally, Newey & West (1987) introduced the HAC method to obtain standard errors of OLS estimators that are corrected for autocorrelated or heteroscedastic error terms. Usually the HAC method is used when the model, say, for example, model 2.4 is misspecified. Moreover, it should be emphasized that Newey & West's (1987) method is only valid in large samples and is not appropriate in smaller ones. Additionally, in smaller samples, the usual OLS procedure might do better than their procedure (Section 12.11, Gujarati 2009). However, the HAC in our augmented Fourier bootstrap (AFB) algorithm will be used for a different purpose than its original version. As will be explained in later sections, the HAC is used to add the correct 'scaled' noise for making inferences on the intercept coefficient.

---

[12]For more details and clarification on definitions, please refer to the introduction section 1.1 of chapter 1.

2.3.3   The TFT class

The "time frequency toggle" procedures proposed by Kirch & Politis (2011), share with our proposed method the core idea of drawing bootstrap observations from the DFT of the sample, followed by inverse transformation to transform the pseudo generated observations back to the time domain. Based on our definition of this class (see, Section 2.2), the FWB satisfies this condition; therefore, it belongs to the TFT class of bootstrap procedures. However, unlike Kirch & Politis's (2011) three basic variants in the TFT class (RB-TFT, WB-TFT, LB-TFT), the FWB differs in its resampling method.

The RB-TFT normalizes the DFT points to equal variance, dividing by the square roots of kernel estimates of the spectral density for the relevant frequency (see, bootstrap procedure 2.2.2). The real and imaginary DFT points are treated equivalently and combined into a single vector. Following standardization to zero mean and unit variance, these points are randomly resampled with equal probability by the Efron's (1979) method. Lastly, the resampled points are assigned frequencies by multiplication by the normalization weights, and randomly allocated to the real and imaginary parts of the bootstrap DFT before inversion back to the time domain. The WB-TFT[13] uses standard Gaussian drawings in place of the DFT residuals, while re-normalizing these using the kernel spectral density estimator in the same manner (cf. bootstrap procedure 2.2.2). The LB-TFT uses normalized kernel weights as probabilities to resample DFT points, at random, but with a higher probability of being drawn from the same frequency range as the sample points they replace.

Each of the RB-TFT, WB-TFT, and LB-TFT methods are dependent on a choice of kernel function and bandwidth. Due to the availability of many choices of kernel functions and different bandwidth, many combinations and hence variants are available for each of those bootstrap methods. In particular, they represent a large class of alternative algorithms. For this reason, a systematic evaluation of these methods is not possible unless these choices of kernel function and bandwidth selection are specified. Given 'reasonable' choices[14], they should behave in a similar way to our proposed bootstrap, at least in large samples. As weighted averages of sample data points, they will share the asymptotic Gaussianity property, and a version of Theorem 2.3.5 should apply, although the autocovariance functions will

---

[13]Hence, take care to note that the designation "wild" is used in a somewhat different sense by these authors to the one we use here. Also, note that this method is an exception to the rule of construction by formula (2.29). The original sample is used only for estimation of the periodogram and is not itself resampled.

[14]'Reasonable' choices in the sense that the chosen kernel function and bandwidth selection criteria are capable of estimating the spectral density $f_{\mathbf{X}}(\cdot)$ of the process $\mathbf{X}$ by $\hat{f}_{\mathbf{X}}(\cdot)$ such that

$$\sup_{\omega \in [0,\pi]} |\hat{f}_{\mathbf{X}}(\omega) - f_{\mathbf{X}}(\omega)| \xrightarrow{p} 0.$$

need to be established differently. The bootstrap draws have a spectral density matching that estimated from the sample by construction, and the bootstrap periodograms accordingly have a conditional distribution centred on this. However, there are no counterparts of Theorem 2.3.2 and Corollary 2.3.2.1. Like all of the suggested bootstrap methods of Kirch & Politis (2011), the FWB method, as defined, is incapable of identifying the distribution of the sample mean due to its degeneracy in producing the bootstrap mean as was stated in the remarks following lemma 2.3.1.

However, in comparing these methods as they are, the key point that needs to be highlighted in this chapter is that our Fourier bootstrap (the FWB) provides the closest possible match between sample and bootstrap autocovariances. This is due to the fact that the bootstrap samples share the exact periodogram of the original sample (see, theorem 2.3.2). Indeed, our method is not suitable for making inferences on the dependence structure of the process itself, but, as we show, is very well adapted to applications such as significance tests in regression.

In what follows, for clarity purposes, we will use the respective acronyms TFT and FWB to distinguish Kirch & Politis's (2011) procedures from our own.

2.4 Tests of Location

The noted degeneracy of the sample mean following from lemma 2.3.1 is one important respect in which the conditional distribution of the FWB fails to match the joint distribution of the sample. If the sample series is centred, as is typically the case with regression residuals, the bootstrap series will likewise sum identically to zero in every case. As a result, certain tests will fail. The FWB is a valid procedure for tests of significance of regressors and restrictions on slope coefficients, since these statistics are formulated in mean deviations and hence are unaffected by this quirk. However, the phenomenon is important in the context of tests of location based on the intercept of a regression, where the mean of the disturbances plays a role in the distribution of the test statistic.

For these applications, like Kirch & Politis (2011), we propose an augmentation of the basic draw. However, unlike them, we do not suggest to augment with our FWB draw a sample mean which was simulated based on a different bootstrap method like the MBB or the AR-sieve bootstrap. To modify our bootstrap procedure so that it does not generate a degenerate sample mean, our bootstrap sample x_b^* drawn from the centred sequence x needs to exhibit a randomly distributed sample means with central tendency zero. For this reason, our procedure involves adding, element-wise, an independent zero-mean random variable to

the generated sample x_b^*. This additional random variable will create some noise that will lead to the desired modification of our bootstrap algorithm.

Although adding a random variable will modify our bootstrap and make it suitable to test a hypothesis of form 2.6, this addition requires 'caution'. In particular, adding a random variable that does not satisfy the CLT conditions will eventually invalidate our bootstrap procedure. Furthermore, even if this additional random variable possesses the properties that meet the CLT conditions, if it is not well adjusted for the variance of the sample at hand, it will cause this bootstrap procedure to mimic the wrong distribution.

Having said this, the natural choice is a standard Gaussian drawing multiplied by a scale factor $\hat{S}_n/\sqrt{n}$ where $\hat{S}_n^2$ estimates the long-run variance[15] of x. If centring means that $x = y - \iota \bar{y}$ where y is a drawing from a distribution with zero mean, then what is wanted is an estimate of $E(\bar{y}^2)$. However, kernel estimators applied to x and y are fortunately identical, and implicitly estimate the latter quantity.

In what follows, we will call this random drawing scaled by $\hat{S}_n/\sqrt{n}$ the 'surrogate mean' of the process. Under weak dependence, the added term is of $O_p(n^{-1/2})$ and it vanishes in the calculation of any statistic that only depends on x_b^* in mean-deviation form and hence cannot affect the distribution of such statistics. The Gaussianity likewise entails no loss of generality in large samples thanks to Theorem 2.3.5. We will refer to the bootstrap procedure based on resamples with added surrogate means as the "augmented Fourier bootstrap" (AFB). To implement the AFB, the remaining problem is to choose $\hat{S}_n^2$. We describe the simulation-based approach we have adopted in Section 2.6, but for clarity, we first need to outline the Monte Carlo technique underlying it.

2.5   Monte Carlo Methodology

Monte Carlo calibration and evaluation of the AFB have been performed using the 'warp speed' technique proposed by Giacomini et al. (2013). In this approach, a single bootstrap distribution is estimated for the whole experiment, with an order-of-magnitude reduction in computational cost relative to running an actual bootstrap test at each replication.

Algorithm Procedure 2.5.1. *The experimental procedure is as follows. For each of K Monte Carlo replications:*

---

[15]The long-run variance of a covariance stationary process $\{y_t\}_{i=1}^n$ is a measure of standard error of the sample mean, i.e.,

$$\lim_{n\to\infty} Var\left(\sqrt{n}(\bar{y}_n - \mathbb{E}\,y_t)\right) = \lim_{n\to\infty} n\,\mathbb{E}(\bar{y}_n - \mathbb{E}\,y_t)^2 = \sum_{j=-\infty}^{+\infty} Cov(y_t, y_{t-j}).$$

1. *Generate a sample data-set, calculate estimates, residuals and test statistics from these data, and store the latter.*

2. *Take a single bootstrap draw from the residuals from Step 1, compute matching test statistics from the bootstrap sample and also store these.*

At the termination of the experiments, we therefore have (for each test) two empirical distributions for comparison. We refer to the 'sample statistics' and 'bootstrap statistics', respectively. We sorted the K bootstrap statistics $\{t_i^*\}_{i=1}^K$ and the sample statistics $\{t_i\}_{i=1}^K$ into rank order, from smallest to largest and, for each $k = 1, \ldots, K$, as

$$t_{(1)}^* \leq t_{(2)}^* \leq \cdots \leq t_{(K)}^*$$

$$t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(K)}$$

The *p*-value $p_k \in [0,1]$ for the kth experiment is estimated by one of $K+1$ equally spaced points on the unit interval, representing the position of the kth sample statistic in the bootstrap distribution. For a 1-tailed test with rejection in the upper tail, p_k is the value such that $K(1 - p_k)$ of the bootstrap statistics fall below or at most equal the kth sample statistic. That is, that

$$p_k = \frac{1}{K} \sum_{i=1}^{K} 1\{t_i^* < t_{(k)}\}$$

The bootstrap test, therefore, results in rejection at significance levels greater than or equal to p_k. For a two-tail test based on a signed statistic, there are two options: (a) a 1-tailed test of the absolute statistics (discarding the sign, assuming symmetry of the test distribution about zero) and (b) the equal tails procedure, which was the one adopted for these experiments. In this case, the *p*-values are estimated as

$$p_k = \min(1, 2m/K)$$

where $m \geq 0$ is the smallest integer, r, such that the kth sample statistic either equals or lies below the bootstrap draw ranked $r+1$ in increasing order, or equals or lies above the bootstrap draw ranked $K - r$. In a mathematical form, m is defined as

$$m := \inf_{r=1,2\ldots,K} \left\{ r : \{t_{(r)}^* \geq t_k\} \cup \{t_{(K-r)}^* \leq t_k\} \right\}$$

Letting $p_{(k)}$ denote the kth-ranked empirical p-value such that $0 \leq p_{(1)} \leq \cdots \leq p_{(K)} \leq 1$, if the bootstrap and sample distributions match and are continuous, we expect $p_{(k)}$ to be close to k/K for each k when the number of Monte Carlo replications is large enough.

The validity of the warp speed method depends on the fact that while the bootstrap draws are conditional on the sample data in replication k, they have a common unconditional distribution in the experiment. If and, in general, only if the bootstrap is valid, this distribution must match the distribution of the sample replicates. Formally, let ξ represent the Monte Carlo replicate obtained at Step 1, and let ζ represent the random drawing that conditionally generates the bootstrap data at Step 2. Then, let $t = t(\xi)$ denote the sample statistic with distribution $F(x) = P(t \leq x)$, which we assume to be continuous, and let $t^* = t^*(\xi, \zeta)$ denote the bootstrap statistic. The conventional bootstrap exercise estimates the random measure $F_\xi(x) = P(t^* \leq x | \xi)$. Locating t in this distribution estimates the p-value, which defined either as

$$g(t) = 1 - F_\xi(t) \tag{2.44}$$

in the case of a 1-tailed test, or as

$$g(t) = 2\min(F_\xi(t), 1 - F_\xi(t)) \tag{2.45}$$

in the case of a two-tailed test with equal tails. To the extent that the error in rejection probability (ERP) of the test at any significance level is small, the distributions F and F_ξ match, and hence the latter does not in practice depend on ξ. If the distributions match and are continuous, $g(t)$ is distributed uniformly on $[0, 1]$. How far this property obtains is what the conventional Monte Carlo exercise seeks to measure.

Now consider the warp-speed Monte Carlo procedure. This estimates directly, and compares, the distributions F and F^*, where $F^*(x) = E_\xi(F_\xi(x))$ and E_ξ represents the mean with respect to the distribution of the replicated samples. The warp-speed p-values are $g^*(t)$, defined by (2.44) or (2.45) with F^* replacing F_ξ in the formulae to the extent that the bootstrap is valid and F_ξ does not depend on ξ, F^* and F_ξ match and hence F and F^* match. In particular, if $F_\xi = F$ with probability 1 and F is continuous, then g^* is uniformly distributed on $[0, 1]$. It is not impossible to have F matching F^* in spite of bootstrap failure with positive probability, if the deviations of F_ξ from F average to zero, so "only if" can not strictly be asserted here, as it can in a conventional experiment. For example, it's well-known that the bootstrap fails when the sample data have no variance, the bootstrap distribution depending on ξ even in the limit (Athreya 1987). In such circumstances, we cannot predict how the warp-speed method might perform. Excepting such cases, however, warp-speed experiments can be proposed as a reliable technique of bootstrap evaluation. As a check,

we ran one of our experiments conventionally (1000 replications with 399 bootstrap draws in each) but found no difference in the results that could not be accounted for by the usual margin of experimental error.

2.6   Constructing the Surrogate Mean

In our simulation experiments, we find that a surrogate mean based on a simple kernel HAC variance estimator of S_n^2 greatly improves the size characteristics of tests of location, relative to the uncorrected bootstrap. However, a fairly substantial error in rejection probability (ERP) remains. The variance estimator applied to fitted regression residuals is downward-biased, and the resulting ERP appears greater as the sample is small, and the degree of autocorrelation is large.

Therefore, to correct for the ERP, especially when the sample is small or the degree of autocorrelation is large, we introduced a response-surface correction as a function of the initial variance estimate and the sample size. The procedure we utilize for constructing the response-surface is summarized in the following sequential algorithmic steps.

Algorithm Procedure 2.6.1.

1. *Generate artificial data from the model*

$$Y_t = \mu + V_t, \ V_t = \rho V_{t-1} + U_t, \ U_t \sim \text{NI}(0,1), \ t = 1, \ldots, n \qquad (2.46)$$

   where the null hypothesis $\mu = 0$ is true, and various values are assigned to ρ and n; such that $n = \{50, 100, 200, 400, 800, 1600\}$ and $\rho = \{0, 0.3, 0.6, 0.9\}$.

2. *Compute from the mean deviations $Y_t - \bar{Y}$ the kernel long-run variance estimator, $\hat{\omega}^2 = \hat{\sigma}^2 + 2\hat{\lambda}$ where $\hat{\lambda}$ denotes the appropriately weighted sum of autocovariances.*

   Remark: The Bartlett kernel[16] was used with bandwidth chosen automatically by the Newey & West (1994) plug-in method.

3. *Perform bootstrap tests of the null hypothesis using the AFB as described in Section 2.4 with an estimated "long-run variance" of the form*

$$\hat{S}_n^2 = \hat{\omega}^2 R_i \qquad (2.47)$$

---

[16]The Bartlett kernel belongs to the set of kernels that satisfy the "reasonable" choices for the RB-TFT bootstrap to work (Kirch & Politis 2011). According to Kirch & Politis's (2011) simulation experiments, the Bartlett kernel works better than the uniform kernel in cases such as the *change-point test* and *Unit root testing*. However, in the former case, the Bartlett kernel exhibits more loss of power.

where $R_0 = 0.5$, and $R_i = 1.1R_{i-1}$; $i = 1, 2, \ldots$

4. *Apply, for each i, and for the same values of ρ and n, the previous steps with $K := 5000$ warp-speed Monte Carlo experiments run.*

5. *Compute the bootstrap p-value p_k: the proportion of the bootstrap statistics exceeding the kth sample statistic, as described in Section 2.4, for each warp-speed Monte Carlo experiment $k = 1, 2, \ldots, K$.*

6. *Estimate the empirical rejection rate in an α-level test as*

$$\hat{P}(\alpha) = K^{-1} \sum_{k=1}^{K} I(p_k \leq \alpha) \tag{2.48}$$

where $I(.)$ denotes the indicator function.

7. *Compute, for each i, the Cramer-von Mises (CvM) goodness-of-fit criterion for the p-value distribution as*

$$CvM_i = K^{-1} \sum_{k=1}^{K} (\hat{P}(k/K) - k/K)^2. \tag{2.49}$$

If correct experimental test size at any chosen α is the case $|\hat{P}(\alpha) - \alpha| \leq 1/K$, we expect CvM to approach zero for large K in the best case.[17]

8. *Record the value of $(R_i, \hat{\sigma}^2)$ that minimizes the CvM_i for the specific combination of (n, ρ).*

As discussed in the above algorithm, trial surrogate-means were created using the variance formula $\hat{S}_n^2 = \hat{\omega}^2 \hat{R}_i$ over a grid of values defined by $R_i = 1.1R_{i-1}$, starting with $R_0 = 0.5$. The sequence of experiments was terminated as soon as CvM was observed to increase at three successive grid points. The value of R yielding the minimum of CvM over the grid was then recorded. These experiments yielded 24 triples $(R, n, \hat{\sigma}^2)$ where $\hat{\sigma}^2$ is computed as the mean of the estimated sample variances over the 5000 replications.[18] A trans-log regression

---

[17]The Kolmogorov-Smirnov criterion

$$KS = \max_{1 \leq k \leq K} \{|P(k/K) - k/K|\}$$

yielded the same result as CvM in nearly every case, in our experiments.

[18]We also tried fitting the response-surface to the average of the long run variances $\hat{\omega}^2$. However, the scheme described gives a better fit to the data.

(the form selected as giving a better least squares fit than the corresponding linear form) was fitted to these data points yielding:

$$\hat{R}(n, \hat{\sigma}^2) = \exp\{4.33337 - 2.07486 \log n + 0.3395(\log n)^2 - 0.01868(\log n)^3$$
$$+ 4.05463 \log \hat{\sigma}^2 + 0.56462(\log \hat{\sigma}^2)^2 - 0.30931(\log \hat{\sigma}^2)^3$$
$$- 1.08039 \log n \log \hat{\sigma}^2 + 0.01368 \log n (\log \hat{\sigma}^2)^2$$
$$+ 0.07224(\log n)^2 \log \hat{\sigma}^2\}.$$

(2.50)

The experiments reported below use the scale factor $\hat{S}_n/\sqrt{n}$ given by (2.50) in (2.47) where $\hat{\sigma}^2$ is the residual variance in the sample regression and $\hat{\omega}^2$ is the kernel estimator of the long-run variance. Notwithstanding the experimental context in which (2.50) was created, note that the ingredients of the correction are available from the sample data (in this case letting $\hat{\sigma}^2$ denote the sample variance) so that this procedure is fully operational for practical applications. The experimental setup leading to (2.50) is the simple test of location in a univariate model, but our experiments suggest that the resulting correction has quite a general application.

2.7 Monte Carlo Evidence

We have run a number of experiments, designed to evaluate the performance of different test procedures in the presence of neglected autocorrelation (see, section 2.10). We set $K = 50,000$ replications, a feasible choice for the warp speed method. Since, in all cases, our test is based on HAC variance estimates, the obtained test statistics, in all cases, are asymptotically pivotal. A feature of bootstrap tests that our experiments do not capture is the approximation due to a feasibly small number of bootstrap draws. Using the warp-speed method, by contrast, we can claim to approach the asymptote in measuring bootstrap rejection rates.

We compared the performance of the AFB, implemented as described, with alternatives including the TFT residual bootstrap, with spectral density estimated according to the Parzen window and a bandwidth of $\lfloor n^{1/3} \rfloor$; the moving blocks method of Künsch (1989) with block length $\lceil n^{1/3} \rceil$; the stationary block bootstrap method of Politis & Romano (1994) with mean block length $\lceil n^{1/3} \rceil$; and the AR-sieve bootstrap of Bühlmann (1997) with lag length chosen by the Akaike criterion up to a maximum of $\lfloor 0.6n^{1/3} \rfloor$. All these bootstrap methods were applied subject to a pretest for autocorrelation. If the absolute value of the first-order autocorrelation coefficient is less than $2/\sqrt{n}$, the standard Efron's (1979) bootstrap

with random sampling of the sequence coordinates is used.[19] To provide a baseline for the comparisons, we also performed asymptotic tests, using critical values from the Student t tabulation with $n - 3$ degrees of freedom.

Test performance under the null hypothesis is evaluated by computing bootstrap p-values using the warp-speed method, as explained in Section 2.5, and comparing the experimental distribution with the uniform distribution. For the hypothesis test in question, let p_k denote the estimated p-value in the kth Monte Carlo replication, calculated as in Section 2.5. Our experiments compare three measures of bootstrap performance. Two measures of size distortion are $\hat{P}(\alpha) - \alpha$ for a chosen α, where $\hat{P}$ denotes the empirical rejection rate (2.48), and the Cramer-von Mises criterion (CvM) (see, 2.49), which summarizes the null distribution as a whole. The third item reported is the rate of rejection of the null hypothesis at the 5% level, for specific cases of the alternative. These alternatives are chosen in relation to sample size to ensure rejection rates larger than 5% but also below 100%, to ensure a useful comparison. Here, 'rejection' is defined as the case of the p-value lying below the previously estimated null rejection rate in the nominal 5% tests; hence the tables compare true powers in correctly sized tests, not nominal rejection rates.

Systematic comparisons must entail looking at a range of cases, and the burden of both reporting and absorbing the resulting plethora of results threatens to become excessive. The full set of experimental results is reported in the section 2.10. In the subsection below, we summarize the comparisons by reporting numerical results for important cases. Specifically, we will report the size distortions[20] and power of different bootstrap methods for a particular dependence structure and sample sizes.

The following subsections describe the models reported in detail.

2.7.1   Significance tests in a regression model

Tables 2.1, 2.2 and 2.3 show the results of experiments with a regression model with intercept and two exogenous regressors,

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + U_t,, \qquad \text{where} \qquad t = 1, \ldots, n$$

in samples of size $n = 50$, 200 and 800. Regressor X_1 is a serially independent standard normal (N(0,1)) sequence, and regressor X_2 is an AR(1) process with coefficient equal to

---

[19]The justification for this is simply that in preliminary experiments it appeared to improve performance in several contexts. There are many ways to implement such a pre-test, and in a practical setting a Q test (Box and Pierce 1970) might also be considered for this purpose.

[20]All distortion sizes of the tables are of 5% critical levels.

0.5, driven by N(0,1) shocks. Regressors and disturbances U_t were all sampled afresh at each replication, so the experimental results are not contingent on a fixed regressor set, but on the specified distributions alone. The test statistics are robust t ratios with standard errors computed by the HAC estimator with Parzen kernel and Newey & West (1994) plug-in bandwidth. The bootstrap significance tests use the signed statistics with equal-tailed rejection regions.

The alternative hypotheses used to evaluate power are constructed by setting $\beta_j = 3/\sqrt{n}$, for $j = 1, 2, 3$, the scale factor '3' being selected on the basis of preliminary experiments to yield powers exceeding size, but at the same time not too close to unity. Thus, their role is solely to rank the relative performances of the different tests applied to an identical set of models, and is otherwise arbitrary.

The three tables report experimental comparisons with different models of residual dependence. Table 2.1 shows the averaged results for four different cases of the AR(1) model $U_t = \rho U_{t-1} + E_t$, with E_t an i.i.d. shock and coefficients $\rho = 0, 0.3, 0.6$ and 0.9. Combined with the three sample sizes, these table entries, therefore, show the average performance over 12 different cases. The rows AFB, MBB and SBB refer to the augmented Fourier, the moving blocks, and the stationary blocks bootstraps, respectively. The MBB and SMM (mean) block lengths are respectively 4 ($n = 50$), 6 ($n = 200$) and 10 ($n = 800$). Do not overlook the fact that since test statistics relating to the slope coefficients are computed solely in mean deviations, the so-called AFB tests of β_1 and β_2 correspond to the basic FWB of Section 2.3. However, the t-statistics for β_0 are functions of the disturbance mean, and these depend critically on the augmentation.

	Size Distortion			Power		
	β_0	β_1	β_2	β_0	β_1	β_2
AFB (N)	0.0749	0.0498	0.0519	0.1100	0.3098	0.1852
AFB (χ_1^2)	0.0840	0.0481	0.0474	0.1187	0.316	0.1793
AFB (t_3)	0.0835	0.0471	0.0482	0.1398	0.3435	0.1956
RB-TFT (N)	0.3380	0.0531	0.0550	0.7037	0.3273	0.1960
MBB (N)	0.1820	0.0522	0.0532	0.3920	0.3247	0.19
SBB (N)	0.1416	0.0529	0.0555	0.3008	0.3262	0.18772
Asy (N)	0.2853	0.0636	0.0934	0.5738	0.3154	0.3426
Asy (t_3)	0.2924	0.0619	0.0937	0.5907	0.4151	0.3693

Table 2.1 The regression model with correlated disturbances - robust t-ratios with HAC std. errors for AR dependence structure $\rho = 0.9$ and sample size $n = 200$

The symbols in parentheses show how the shocks were generated in each case. Since we have emphasised the role of Gaussianity in validating the match of the bootstrap and sample distributions, it is important to have evidence of how much this influences the outcomes in practice. The AFB method was therefore compared in three cases, the standard normal N(0,1), the normalized chi-squared $(\chi_1^2 - 1)/\sqrt{2}$ supplying skewness, and the normalized Student-t, $t_3/3$, supplying excess kurtosis. To provide a baseline comparison, the final two rows of the table show the outcomes of conventional asymptotic tests.

In Table 2.2, the residual dependence is generated by finite-order moving-averages, with the form $U_t = E_t + E_{t-m}$ for the three cases, $m = 1$, $m = 2$ and $m = 4$, so giving a total of nine variations composing the average. Here the block bootstrap methods are replaced in the comparison by the AR-sieve, denoted SAR in the table. Since interesting comparisons need to be model-free as far as possible, we omit this method from Table 2.1. The AR-sieve can only be a generally useful alternative if it performs well outside the autoregressive context.

	Size Distortion			Power		
	β_0	β_1	β_2	β_0	β_1	β_2
AFB (N)	0.0796	0.0488	0.0593	0.4615	0.5281	0.6202
RB-TFT (N)	0.1612	0.0483	0.0557	0.7108	0.5185	0.5973
SAR (N)	0.0465	0.0485	0.0520	0.2694	0.5229	0.5922
Asy (N)	0.0796	0.0488	0.0593	0.4615	0.5281	0.6202

Table 2.2 The regression model with MA disturbances with $m = 2$ and sample size $n = 200$

Table 2.3, comparing the same alternatives as in Table 2.2, shows results with stationary fractionally integrated dependence[21]. The cases simulated have the form $U_t = (1 - L)_+^{-d} E_t$ where the fractional operator $(1 - L)_+^{-d}$ truncates lags to positive values of t . We set $d = 0.1$ and $d = 0.3$, so the averages in the table are over six model variants in this case.

---

[21]Note that in some finance and economics time series data, observations, taken some distance apart, show signs of dependence. This phenomenon is known as *long memory*. The fractionally integrated model is one of the many ways used to represent long memory process. Under fractionally integrated models, the corresponding autocorrelation function will decline hyperbolically to zero.

A fractionally integrated process is stationary only if the process may be represented by a MA(∞). This happens when the order of integration $d < 1/2$ (Hamilton 1994, p. 448-449).

	Size Distortion			Power		
	β_0	β_1	β_2	β_0	β_1	β_2
AFB (N)	0.1094	0.0468	0.0505	0.7619	0.7144	0.7775
RB-TFT (N)	0.2314	0.0519	0.0525	0.9113	0.7400	0.7817
SAR (N)	0.0973	0.0438	0.0493	0.7175	0.6900	0.7663
Asy (N)	0.16586	0.0938	0.1068	0.16586	0.8591	0.9144

Table 2.3 The regression model with fractional disturbances with integration parameter $d = 0.1$ and sample size $n = 50$

We emphasize that the numerical figures of the tables reported here that represents some of our findings are intended only to highlight, and are not a substitute for the full sets of results given in section 2.10. We can, however, note the key findings. Looking first at the tests of significance (β_1 and β_2) we see that the AFB (equivalently, FWB) is generally either superior to or at least as good as its rivals. This method is unique among the bootstrap alternatives in not depending on a choice of bandwidth, block length or lag length. While it is not impossible that different choices of these parameters could change the rankings, the need to make a choice, typically poorly informed, is an significant drawback of those methods. In the test of location (β_0), it should again be pointed out that no claims are made for the TFT method in this context, but nonetheless, it is helpful to see how it behaves in practice. In our implementation, the AFB response-surface has been fitted to models in the autoregressive class and so inevitably performs less well in Table 2.2 than in Table 2.1. However, it performs as well as the AR-sieve, which is under the same constraint there and is its only serious competitor.

The stationary fractional models of Table 2.3 are in a different category, because there are no grounds here for assuming the tests are consistent. Under strong dependence, the statistics are not asymptotically pivotal, and conditions for the central-limit theorem are violated. Although the spectrum diverges at the origin, the question is whether in these circumstances either the frequency domain bootstraps or the AR-sieve can provide better size characteristics than the asymptotic test. Tests of location fail unequivocally, in every case, an unsurprising result noting that the sample-means converge only at the rate $O(n^{d-1/2})$. On the other hand, the bootstrap tests of significance perform well, although without an unambiguous ranking of alternatives, whereas the asymptotic tests are conspicuously poor.

2.7.2   Unit Root Tests

It is well-known that there are differences between unit root and stationary processes when it comes to asymptotic distribution and rate of convergence of the estimated coefficients (Hamilton 1994, p. 475). It is recognised that both Said & Dickey (1984) augmented Dickey–Fuller test and Phillips & Perron (1988) modified t-ratio tests of unit root have considerable size distortion in finite samples. Therefore, several authors have investigated the consequence of applying a bootstrap method in such situation, i.e., unit root tests relying on the bootstrap critical values. In particular, they found that the bootstrap method, when applied suitably, on the finite sample reduces the size distortion of the asymptotic theory. See, for example, Chang & Park (2003), Ferretti & Romo (1996), Park (2003), Section 6.3 in Kirch & Politis (2011) and Paparoditis & Politis (2005), and also Park (2002) and Kreiss & Paparoditis (2003) for related research on these methods.

Consider the model

$$y_t = \beta_1 y_{t-1} + U_t \tag{2.51}$$

Our interest lies in acknowledging whether the Fourier bootstrap is capable of testing the unit root null hypothesis defined by

$$\mathscr{H}_0 : \beta_1 = 1 \tag{2.52}$$

when the error terms $\{U_t\}_{t=1}^n$ suffer from serial-correlation.

Remark: Model 2.51 is referred to case 1 in (Hamilton 1994, p. 487) which stands for no constant or time trend in the regression.

Since Phillips & Perron (1988) modified t-ratio offers a nonparametric treatment of dependence of the error terms $\{U_t\}_{t=1}^n$ in model 2.51 and our object here is simply to compare the performance of alternative bootstraps in a consistent framework rather than finding the most powerful or best-sized test, in our experiments, we opt for using the Phillips & Perron (1988) modified t-ratio rather than the augmented Dickey–Fuller test. There are, of course, a number of other unit-root tests that may perform better in certain contexts, see Elliott et al. (1996) and Ng & Perron (2001) inter alia.

In this framework, under the null hypothesis, the data-sets are generated as in the form

$$y_t = y_{t-1} + U_t, \ U_t = \rho U_{t-1} + E_t, \ t = 1,\ldots,n$$

where $y_0 = U_0 = 0$, and E_t is variously generated.

To obtain bootstrap samples, we first fit, using the usual OLS regression, the differenced data-set $\Delta y_t := y_t - y_{t-1}$ by the following model

$$\Delta y_t = \phi y_{t-1} + \varepsilon_t.$$

Then, obtain the residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$. This unrestricted filtering step is necessary to ensure the tests have power, undoing the effect of over-differencing under the alternative, although it does increase size distortion in small samples. In particular, observe that the residual series $\{\hat{\varepsilon}_t\}_{t=1}^n$ is stationary under the null and alternative and therefore the AFB is applicable.

The modified t-test statistics suggested by Phillips & Perron (1988) requires the estimation of the short-run variance and the long-run variance of the residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$. Therefore, denote by $\hat{\gamma}_\varepsilon$ and $\hat{J}_\varepsilon$ the estimated short-run variance and the estimated long-run variance of the residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$, respectively. The estimated short-run variance, $\hat{\gamma}_\varepsilon$, have been estimated using the standard variance estimator and the estimated long-run variance, $\hat{J}_\varepsilon$, have been estimated non-parametrically using the bartlett kernel and Newey & West (1994) plug-in bandwidth. Mathematically, the modified Phillips & Perron's (1988) t-test statistics for the no deterministic components case is given by:

$$t_{PP} := \sqrt{\frac{\hat{\gamma}_\varepsilon}{\hat{J}_\varepsilon}} t_{\hat{\phi}} - \frac{1}{2}(\hat{J}_\varepsilon - \hat{\gamma}_\varepsilon)\left(\frac{\hat{J}_\varepsilon}{n^2}\sum_{t=1}^n y_{t-1}^2\right)^{-1/2} \tag{2.53}$$

Where

$$t_{\hat{\phi}} := \hat{\phi}/sd(\hat{\phi}) \qquad \text{and} \qquad sd(\hat{\phi}) := \left(\frac{\frac{1}{n-2}\sum_{t=2}^n(\Delta y_t - \hat{\phi}y_{t-1})^2}{\sum_{t=1}^n y_{t-1}^2}\right)$$

Afterwards, obtain the order-one integrated bootstrap samples as:

$$y_t^* = y_0 + \sum_{i=1}^t \varepsilon_i^*$$

where the series $\{\varepsilon_t^*\}_{t=1}^n$ are the AFB resampled series from the residual series $\{\hat{\varepsilon}_t\}_{t=1}^n$ and y_0 is an initial value set to zero[22]Observe that the generation of the bootstrap sample fulfills the null hypothesis 2.52 for consistency reasons[23].

---

[22]The choice of the initial value y_0 does not affect the asymptotic as long as this value is stochastically bounded (See, Sec 4, Chang & Park 2003). To the best of our knowledge, we believe that the authors mean that the random variable is bounded.

[23]Basawa et al. (1991*b*,*a*) have shown that in the generation of the bootstrap sample, for the bootstrap unit root tests, the unit root must be imposed to achieve consistency.

Remark: The usage of the AFB rather than the WFB is compulsory for this bootstrap procedure to be consistent. In particular, the augmentation of the Fourier bootstrap is critical to the correct performance of the test, since otherwise, the cumulation of the bootstrap draws would yield a Brownian bridge in the limit, not the required Brownian motion.

To construct the bootstrap critical values, the bootstrap counterparts of the modified Phillips & Perron's (1988) t-statistic defined in equation (2.53) is required. This bootstrap test statistic is computed with the same sequential steps described above but with OLS residuals received by regressing the differenced bootstrap data-set on the lagged level of the bootstrap observation.

Observe that the Phillips & Perron's (1988) modified t-test statistic defined in equation (2.53) converges, as n goes to infinity, to the Dickey-Fuller test statistic defined by

$$\frac{\frac{1}{2}\left(W^2(1)-1\right)}{\left(\int_0^1 W^2(r)dr\right)^{1/2}}, \qquad \text{where} \qquad W(r) \text{ is the Wiener process} \tag{2.54}$$

(eq. 17.4.12 Hamilton 1994, p. 489). Therefore, based on theorems 2.3.2 and 2.3.3, it can be shown that the bootstrap version of the modified t-statistic converges to the same limit described in equation (2.54). Accordingly, consistency of AFB is acknowledge for unit root hypothesis testing.

Cases of the alternative hypothesis, the data-set are generated by simply setting $y_t = U_t$. There is no need for an arbitrary alternative, as in the regression example, and the rates of rejection are, in the obvious way, related directly to the degrees of induced autocorrelation in the series. These features of the experiment can be clearly seen in the tables of individual model results in section 2.10.

Table 2.4 shows the results obtained under null 2.52 and alternative in bootstrap unit root tests. The entries are the averages of the same 12 combinations of autocorrelation and sample size, as in Table 2.1.

	Size Distortion		Power
	5%	CvM	
AFB (N)	9.1	8.5	90.9
AFB (χ_1^2)	9.6	10.0	90.8
AFB (t_3)	7.6	11.4	91.2
MBB1 (N)	10.3	16.8	88.5
MBB2 (N)	9.7	14.9	89.2
SBB1 (N)	10.5	16.1	88.9
SBB2 (N)	18.8	8.6	90.0
Asy (N)	39.1	115.5	86.9
Asy (t_3)	39.8	113.6	87.8

Table 2.4 Unit root tests with correlated increments using the Phillips-Perron statistic.

2.8 The Multivariate Case

For simplicity's sake, the analysis has thus far dealt with the case of a single, autocorrelated series. In the obvious extension to multiple time series that may be cross-autocorrelated, the essential feature of the basic multivariate draw, to preserve the required characteristics, is that the same Rademacher variates must be used for each element of the process. If X ($n \times r$) denotes the matrix whose columns are the time series of each of r variables, the FWB draw takes the form

$$X^* = RX \tag{2.55}$$

where R is defined as before by (2.28).

Without loss of generality, it suffices to consider the bivariate case with $r = 2$. Two additional theorems suffice to establish the requisite properties, as follows. The cross-periodogram of $X = (x_1, x_2)$ is in general a complex-valued process.

Theorem 2.8.1. *The cross-periodogram of $X^* = (x_1^*, x_2^*)$ is identical with that of X.*

Proof. Similarly to the proof of Theorem 2.3.2, the cross-periodogram points of $X = (x_1, x_2)$ can be written as the pair of vectors $(2\pi)^{-1} z_1 \circ z_2^\dagger$ and $(2\pi)^{-1} z_2 \circ z_1^\dagger$ where the z_j are defined by (2.24) with appropriate substitutions, where '$\dagger$' denotes the complex conjugate and '$\circ$' denotes the element-wise Hadamard product of the two vectors. Thus,

$$z_1 \circ z_2^\dagger = Ax_1 \circ Ax_2 + Bx_1 \circ Bx_2 + \mathrm{i}(Bx_1 \circ Ax_2 - Ax_1 \circ Bx_2) \tag{2.56}$$

and $z_2 \circ z_1^\dagger$ is the corresponding expression with reversed sign of the imaginary component. Then, from (2.55) using (2.37) and (2.38), the cross-periodogram points of the bootstrap draw take the forms of

$$P_A x_1 \circ P_A x_2 + P_B x_1 \circ P_B x_2 + \mathrm{i}(P_B x_1 \circ P_A x_2 - P_A x_1 \circ P_B x_2) \qquad (2.57)$$

and of its complex conjugate.

We now show, reprising the arguments in the proof of Theorem 2.3.2 , that both the real and the imaginary parts of (2.57) are identical with those of (2.56). Let us consider the real terms first, recalling that the rows of P_A follow the scheme in (2.39) and the rows of P_B follow (2.40) similarly. The sign changes disappear in the Hadamard products, similarly to the cases of the squares, except that here the signs are not always positive. What matters is that the signs of the elements of $A x_1 \circ A x_2$ must match those of $(-A) x_1 \circ (-A) x_2$, and similarly for B. We therefore conclude that

$$P_A x_1 \circ P_A x_2 + P_B x_1 \circ P_B x_2 = A x_1 \circ A x_2 + B x_1 \circ B x_2$$

This result holds whether or not $x_1 = x_2$, which was the case shown previously.

Now consider the imaginary parts in the light of tables (2.39) and (2.40). We find that the elements of the vectors $P_B x_1 \circ P_A x_2$ and $P_A x_1 \circ P_B x_2$ match those of the respective vectors $B x_1 \circ A x_2$ and $A x_1 \circ B x_2$ when the signs of the Rademacher pairs agree. When the Rademacher pairs take opposite signs, the corresponding elements of $P_B x_1 \circ P_A x_2$ and $P_A x_1 \circ P_B x_2$ match, respectively, those of $-(A x_2 \circ B x_1)$ and $-(B x_1 \circ A x_2)$. It follows that

$$P_B x_1 \circ P_A x_2 - P_A x_1 \circ P_B x_2 = B x_1 \circ A x_2 - A x_1 \circ B x_2. \qquad (2.58)$$

Since the same equalities hold for the complex conjugate, which merely changes the sign on both sides of (2.58), the proof is complete. $\qquad \square$

We can calculate the means under the bootstrap distribution of the cross-autocovariances via the following corollary of Theorem 2.3.3 .

Theorem 2.8.2. *Let x_{pi} denote the ith element of vector x_p, for $i = 0, \ldots, n-1$, and $p = 1, 2$, and let x_{pi}^* denote the corresponding element of the bootstrap draw. For $m \geq 0$ and $i \geq m$,*

$$\mathbb{E}^*(x_{1i}^* x_{2,i-m}^*) = \frac{1}{n} \sum_{j=m}^{n-1} x_{1j} x_{2,j-m}$$

$$\mathbb{E}^*(x^*_{2i}x^*_{1,i-m}) = \frac{1}{n}\sum_{j=m}^{n-1}x_{1j}x_{2,j-m}$$

Proof. This follows immediately from the arguments of Theorem 2.3.3. Simply note that

$$x^*_{pi} = \sum_{j=0}^{n-1}\sum_{k=0}^{n-1}w_k\Upsilon_{ijk}x_{pj}$$

for $p = 1$ or $p = 2$. Accordingly, replace $x_jx_{j'}$ by $x_{1j}x_{2j'}$ in all the expressions where the latter product appears, and likewise replace $x^*_jx^*_{j-m}$ by $x^*_{1j}x^*_{2,j-m}$ and x_jx_{j-m} by $x_{1j}x_{2,j-m}$. With these amendments, the derivation proceeds unchanged. Then, interchange subscripts 1 and 2 for the second part of the theorem.   □

These results establish the theoretical properties of the basic multivariate bootstrap, on the same basis as the univariate case. It remains to consider how best to implement the AFB procedure of Section 2.4. Let $\hat{\Omega} = \hat{\Sigma} + \hat{\Lambda} + \hat{\Lambda}'$ denote an empirical HAC covariance matrix of X, and let lower triangular $\hat{P}$ denote its Choleski decomposition, such that $\hat{\Omega} = \hat{P}\hat{P}'$. Then, letting Z ($r \times 1$) denote a standard Gaussian vector, the vector of surrogate means might be computed as $\hat{S}_n Z$ where

$$\hat{S}_n = n^{-1/2}\hat{P}\hat{R}(n,\hat{\Sigma})^{1/2}$$

and $\hat{R}(n,\hat{\Sigma})$ denotes a $r \times r$ diagonal matrix of weights. A low-cost implementation of this method would be to use formula (2.50), with the appropriate diagonal element of $\hat{\Sigma}$ inserted in the formula, to generate each of the diagonal elements of $\hat{R}$. Systematic comparisons are not attempted in this instance. We simply report some evidence on the adequacy of this augmentation scheme, in particular, for testing locations in a bivariate model. Experimental series of length $n = 500$ were generated by the VAR(1) model

$$Y_{1t} = 0.2Y_{1,t-1} + 0.2Y_{2,t-1} + E_{1t}$$
$$Y_{2t} = 0.3Y_{1,t-1} + 0.1Y_{2,t-1} + E_{2t}$$

where (E_{1t}, E_{2t}) are serially independent Gaussian with unit variances and contemporaneous covariance $\gamma_{12} = 0.3$. Table 2.5 shows the variances and zero-order and first-order covariances of the resulting series, solved from the Yule-Walker equations.

	Y_{1t}	Y_{2t}	$Y_{1,t-1}$	$Y_{2,t-1}$
Y_{1t}	1.124	0.424	0.309	0.379
Y_{2t}	0.424	1.138	0.312	0.241

Table 2.5 VAR covariances

The estimation stage of the experiment tests the hypothesis that the series means are jointly zero, using the asymptotically pivotal Wald statistic

$$\mathrm{W} = n\hat{\mu}'\hat{\Omega}^{-1}\hat{\mu}$$

where $\hat{\mu}$ is the vector of sample means and $\hat{\Omega}$ an HAC estimate of the data variance matrix. We implement the latter with the Parzen kernel and Newey & West (1994) plug-in bandwidth. Table 2.6 compares null rejection rates at various nominal significance levels, in $50,000$ replications of the experiment, for two cases: the asymptotic test using the $\chi^2(2)$ critical values, and the AFB procedure as described. As before, the warp-speed Monte Carlo method has been used in the bootstrap case.

Nominal significance level	0.1	0.05	0.025	0.01
Rejection rates, $\chi^2(2)$ criterion	0.1256	0.0692	0.0400	0.0194
Rejection rates, AFB criterion	0.1024	0.0511	0.0264	0.0109

Table 2.6 Wald test of joint significance of series means - results from 50,000 replications

2.9   Concluding Remarks

We have proposed a simple bootstrap for dependent processes based on a Rademacher wild bootstrap draw from the discrete Fourier transform. While similar methods have been proposed previously (the so-called TFT class), our variant has the special virtue of exactly reproducing the sample periodogram in the bootstrap draws. A close correspondence between sample and bootstrap distributions should be achievable in a wide range of cases. The basic FWB algorithm is unique in not requiring a choice of bandwidth or other arbitrary parameterization. Being purely nonparametric, its performance should not depend on the actual form of the autocorrelation, subject to weak dependence. While the bootstrap distribution of the sample mean is degenerate, this quirk is irrelevant in tests depending only on the data in

mean deviation form, whereas tests of location and unit roots have been successfully handled by a simple augmentation.

We provide Monte Carlo evidence showing that our method compares well with popular alternatives, is robust to failures of Gaussianity, and performs best out of the alternatives studied in tests of location. Examining the full set of rejection frequencies for different cases, as reported in section 2.10, it can be noted that the AFB has relatively good-size properties when the autocorrelation is extreme. It does not always have the best power against the chosen alternatives, but interestingly enough, the asymptotic tests prove to be more powerful than any of the bootstrap tests, once size-corrected, although this is not a result that can be exploited in practice, since it depends on the true null distributions being estimated via the Monte Carlo experiments. In unit root tests, on the other hand, the AFB appears to dominate over all the other procedures studied.

Our present implementation of the AFB can doubtless be refined, but we note that the method achieves a very reasonable match of sample mean distributions. In the experiments reported in Table 2.1, we recorded the means of the generated samples and also the surrogate means of the AFB draws. To check on the match of distributions, the variances of these two quantities were calculated for a representative experiment: They are tabulated in section 2.10. The average over the 12 models of their ratios (bootstrap mean variance over sample mean variance) is 1.067. While the discrepancies appear linked to sample size, they are scarcely larger overall than what experimental error might produce, increasing confidence in our response-surface variance estimate.

We have considered alternatives to the response-surface approach to bias correction, in particular the application of pre-whitening[24] to the kernel estimator using an AR(1) filter, as suggested by Andrews & Monahan (1992). In trials, this method did appear to reduce the ERP, although not as effectively as the response-surface. The response-surface is calibrated using autoregressive dependence, and there is certainly evidence in the tables that it performs better in this framework that when faced with the moving-average cases. An improved calibration method is one avenue for further research.

In particular, it was reported in a number of studies that HAC methods often have poor finite sample properties. Recent papers provide robust inference methods for dependent and heterogeneous observations that are based on estimating the parameter of interest for each of several groups of data and using conservativeness results for t-statistics in asymptotically normal group estimates. The robust inference approaches are alternatives to HAC methods that do not require estimation of unknown (long-run) asymptotic variances and have appealing

---

[24]Pre-whitening is a technique used to reduce bias and over-rejection of t statistics (constructed under the kernel HAC estimator) or to improve coverage probabilities. Ever since their introduction to the pre-whitening technique in the HAC covariance matrix estimator, practitioners became aware of this method's effectiveness.

finite sample properties in different dependence and heterogeneity settings. Therefore, it would be interesting to compare finite sample properties of this proposed bootstrap methods with the robust approaches discussed in Ibragimov & Müller (2010) and Ibragimov et al. (2015).

Additionally, it would be of interest to extend this bootstrap architecturally to incorporate more time series structures. For example, it would be interesting if one may devise a new bootstrap approach to perform inferences on parameters of a linear time series model in which the data set originates from a non-stationary and possibly long-memory characteristics such as fractionally integrated and others more sophisticate processes. In the current literature, several papers have used discrete Fourier transform to perform inference in such settings see, Phillips (1999, 2009) and the references therein.

A feature of the FWB draw not so far discussed in detail is the degeneracy of the distribution of the series sums of squares according to (2.36). This fact is irrelevant to the applications studied in this paper, but there could be situations in which it assumes importance. In such cases, a 'surrogate variance' might be implemented, along the lines of our surrogate mean, by multiplying the bootstrap series by a randomly drawn scale factor centred on 1. A squared normal variate could serve this purpose, for example. However, investigation of this extension must also be left for future work.

2.10   Appendix: Tables

Tables of Monte Carlo relative frequencies (50,000 replications) of rejection at the 5% level. In this section, the values $\hat{P}(0.05)$, which denote the empirical rejection rate at 5%, are tabulated.

In the tables of previous sections, the columns headed "5%" under "Size Distortion" show $1000\times$ the average absolute deviations from 0.05 of the left-hand block entries. Under "Power", the columns show the average of the entries in the right-hand blocks, in percentage form.

The case-by-case Cramer-von Mises (CvM) criteria are not reported here.

2.10.1   Regression model with two regressors

The left-hand tables show the case of the null hypothesis of zero coefficients. The right-hand tables show cases of the alternative hypothesis, with each coefficient equal to $3/\sqrt{n}$ in the DGP.

β_0 = intercept, β_1= serially independent regressor, β_2 = AR(1) regressor.

	n	Size Distortion (ρ)				Power (ρ)			
		0	0.3	0.6	0.9	0	0.3	0.6	0.9
β_0	50	0.0501	0.0625	0.0557	0.0801	0.7483	0.5526	0.2311	0.1460
	200	0.0498	0.0577	0.0553	0.0749	0.8346	0.5804	0.2164	0.1100
	800	0.0462	0.0551	0.0502	0.0548	0.8430	0.5676	0.2089	0.0711
β_1	50	0.0515	0.0510	0.0532	0.0538	0.7343	0.7151	0.6407	0.4154
	200	0.0501	0.0509	0.0484	0.0498	0.8234	0.7969	0.6613	0.3098
	800	0.0510	0.0517	0.0514	0.0501	0.8433	0.8158	0.6685	0.2754
β_2	50	0.0531	0.0533	0.0570	0.0595	0.7945	0.6967	0.5050	0.2548
	200	0.0524	0.0528	0.0514	0.0519	0.9057	0.7687	0.5127	0.1852
	800	0.0513	0.0488	0.0508	0.0508	0.9280	0.8011	0.5221	0.1617

Table 2.7 AFB with normal shocks

	n	Size Distortion (ρ)				Power (ρ)			
		0	0.3	0.6	0.9	0	0.3	0.6	0.9
β_0	50	0.0481	0.0643	0.0723	0.1041	0.9522	0.6699	0.1933	0.1702
	200	0.0500	0.0643	0.0587	0.0840	0.9372	0.6084	0.1928	0.1187
	800	0.0495	0.0566	0.0515	0.0595	0.8960	0.5773	0.1945	0.072
β_1	50	0.0464	0.0416	0.0421	0.0455	0.7898	0.7553	0.6744	0.4386
	200	0.0467	0.0456	0.0454	0.0481	0.8317	0.7914	0.6662	0.316
	800	0.0498	0.0477	0.0482	0.0476	0.8465	0.8085	0.6706	0.2734
β_2	50	0.0447	0.0500	0.0477	0.0545	0.8181	0.7533	0.5304	0.2690
	200	0.0487	0.0498	0.0497	0.0474	0.8975	0.7915	0.5326	0.1793
	800	0.0494	0.0467	0.0488	0.0500	0.9210	0.7982	0.5145	0.1647

Table 2.8 AFB with χ_1^2 shocks

| | Size Distortion (ρ) | | | | Power (ρ) | | | |
	n	0	0.3	0.6	0.9	0	0.3	0.6	0.9
β_0	50	0.0395	0.0544	0.0556	0.0975	0.7465	0.6066	0.3031	0.2070
	200	0.0449	0.0537	0.0562	0.0835	0.8279	0.6103	0.2701	0.1398
	800	0.0453	0.0524	0.0504	0.0586	0.8386	0.5929	0.2449	0.0835
β_1	50	0.0458	0.0423	0.0442	0.0462	0.8082	0.7739	0.7073	0.4691
	200	0.0487	0.0481	0.0455	0.0471	0.8580	0.8225	0.6969	0.3435
	800	0.0518	0.0502	0.0498	0.0479	0.8655	0.8225	0.6993	0.2878
β_2	50	0.0491	0.0497	0.0513	0.0550	0.8621	0.7706	0.5804	0.2916
	200	0.0481	0.0458	0.0470	0.0482	0.9153	0.8070	0.5493	0.1956
	800	0.0495	0.0484	0.0468	0.0471	0.9279	0.8186	0.5439	0.1620

Table 2.9 AFB with $t(3)$ shocks

| | Size Distortion (ρ) | | | | Power (ρ) | | | |
	n	0	0.3	0.6	0.9	0	0.3	0.6	0.9
β_0	50	0.1331	0.1632	0.2218	0.4561	0.9560	0.8565	0.7174	0.8414
	200	0.1593	0.1743	0.1931	0.3380	0.9748	0.8737	0.6533	0.7037
	800	0.1656	0.1680	0.1687	0.2466	0.9774	0.8679	0.6040	0.5435
β_1	50	0.0515	0.0495	0.0437	0.0381	0.7543	0.7175	0.5967	0.3247
	200	0.0506	0.0511	0.0523	0.0531	0.8274	0.7936	0.6700	0.3273
	800	0.0492	0.0525	0.0508	0.0522	0.8387	0.8096	0.6609	0.2839
β_2	50	0.0530	0.0579	0.0621	0.0731	0.8154	0.7171	0.5399	0.3260
	200	0.0534	0.0543	0.0571	0.0550	0.9134	0.7884	0.5332	0.1960
	800	0.0503	0.0510	0.0505	0.0506	0.9267	0.8069	0.5199	0.1672

Table 2.10 RB-TFT with Normal shocks

	n	Size Distortion (ρ)				Power (ρ)			
		0	0.3	0.6	0.9	0	0.3	0.6	0.9
β_0	50	0.0505	0.0638	0.0856	0.2790	0.7609	0.563	0.371	0.6065
	200	0.0478	0.0538	0.0640	0.1820	0.8176	0.5568	0.2814	0.3920
	800	0.0501	0.0488	0.0483	0.1068	0.8479	0.5457	0.2102	0.2045
β_1	50	0.0508	0.0447	0.0460	0.0469	0.7271	0.6883	0.5983	0.3951
	200	0.0496	0.0502	0.0493	0.0522	0.82552	0.7935	0.6612	0.3247
	800	0.0506	0.0513	0.0499	0.0526	0.8421	0.8169	0.6643	0.2874
β_2	50	0.0510	0.0552	0.0599	0.0695	0.8031	0.7063	0.5323	0.3158
	200	0.0504	0.0539	0.0559	0.0532	0.9007	0.7909	0.5345	0.19
	800	0.0485	0.0533	0.0505	0.0500	0.9244	0.8092	0.5162	0.159

Table 2.11 MBB with block length $\lceil n^{1/3} \rceil$

	n	Size Distortion (ρ)				Power (ρ)			
		0	0.3	0.6	0.9	0	0.3	0.6	0.9
β_0	50	0.0523	0.0683	0.0963	0.2694	0.7563	0.5901	0.4150	0.5898
	200	0.0475	0.0599	0.0661	0.1416	0.8246	0.5900	0.2850	0.3008
	800	0.0506	0.0528	0.0532	0.0890	0.8450	0.5808	0.2251	0.1572
β_1	50	0.0498	0.0479	0.0472	0.0461	0.7436	0.7100	0.6122	0.3874
	200	0.0496	0.0503	0.0490	0.0529	0.8246	0.7897	0.6631	0.3262
	800	0.0504	0.0476	0.0501	0.0502	0.8384	0.8095	0.6676	0.2732
β_2	50	0.0520	0.0548	0.0584	0.0686	0.8057	0.7063	0.5381	0.3095
	200	0.0492	0.0512	0.0537	0.0555	0.90492	0.79076	0.52392	0.18772
	800	0.0516	0.0491	0.0509	0.0511	0.93138	0.79528	0.51948	0.163

Table 2.12 Stationary blocks bootstrap (SBB) with mean block length $\lceil n^{1/3} \rceil$

		Size Distortion (ρ)				Power (ρ)			
	n	0	0.3	0.6	0.9	0	0.3	0.6	0.9
β_0	50	0.0894	0.1215	0.1981	0.4741	0.8927	0.7790	0.654	0.8043
	200	0.0579	0.0851	0.1234	0.2853	0.8628	0.7087	0.4944	0.5738
	800	0.0523	0.0706	0.0852	0.1582	0.8546	0.6600	0.3673	0.3282
β_1	50	0.0970	0.0864	0.0820	0.0808	0.7398	0.6854	0.6244	0.3950
	200	0.0614	0.0595	0.0612	0.0636	0.8269	0.7990	0.6528	0.3154
	800	0.0541	0.0526	0.0546	0.0564	0.844	0.8101	0.6660	0.2807
β_2	50	0.1037	0.1142	0.1385	0.1660	0.9409	0.8954	0.7942	0.5945
	200	0.0648	0.0746	0.0864	0.0934	0.9382	0.8638	0.6815	0.3426
	800	0.0555	0.0629	0.0640	0.0705	0.9358	0.8480	0.6054	0.2362

Table 2.13 Asymptotic test with normal shocks

		Size Distortion (ρ)				Power (ρ)			
	n	0	0.3	0.6	0.9	0	0.3	0.6	0.9
β_0	50	0.0830	0.1180	0.1975	0.4817	0.9035	0.8180	0.6913	0.8178
	200	0.0571	0.0822	0.1223	0.2924	0.875	0.735	0.5235	0.5907
	800	0.0536	0.0696	0.0847	0.1608	0.8672	0.6750	0.3859	0.3330
β_1	50	0.0853	0.0794	0.0732	0.0762	0.9099	0.8938	0.8198	0.6088
	200	0.0570	0.0553	0.0570	0.0619	0.8811	0.8551	0.7554	0.4151
	800	0.0528	0.0524	0.0524	0.0537	0.8682	0.8396	0.7060	0.3237
β_2	50	0.0922	0.1059	0.1282	0.1601	0.9391	0.9093	0.8266	0.6343
	200	0.0595	0.0711	0.0813	0.0937	0.93534	0.8817	0.7210	0.3693
	800	0.0538	0.0593	0.0642	0.0652	0.9338	0.8569	0.6388	0.2393

Table 2.14 Asymptotic test with t3 shocks

2.10.2 Moving Average Dependence

Single lag coefficient of unity with lag m indicated.

	n	Size Distortion (m)			Power (m)		
		1	2	4	1	2	4
	50	0.0435	0.0840	0.1314	0.2123	0.4455	0.6169
β_0	200	0.0416	0.0796	0.1029	0.2581	0.4615	0.5449
	800	0.0439	0.0692	0.0784	0.2772	0.4146	0.4689
	50	0.0508	0.0427	0.0571	0.4976	0.4018	0.5001
β_1	200	0.0499	0.0488	0.0474	0.5363	0.5281	0.5276
	800	0.0499	0.0495	0.0518	0.5620	0.5463	0.5604
	50	0.0544	0.0544	0.0611	0.4378	0.5290	0.5883
β_2	200	0.0515	0.0593	0.0544	0.4976	0.6202	0.6578
	800	0.0531	0.0558	0.0528	0.5168	0.6189	0.6637

Table 2.15 AFB with normal shocks

	n	Size Distortion (m)			Power (m)		
		1	2	4	1	2	4
	50	0.1393	0.1760	0.2358	0.6397	0.7308	0.8166
β_0	200	0.1469	0.1612	0.2088	0.6776	0.7108	0.7934
	800	0.1588	0.1516	0.1519	0.6979	0.6937	0.7051
	50	0.0481	0.0431	0.0578	0.4665	0.4238	0.5164
β_1	200	0.0495	0.0483	0.0466	0.5502	0.5185	0.5182
	800	0.0512	0.0496	0.0514	0.5585	0.5541	0.5641
	50	0.0545	0.0544	0.0625	0.4403	0.5304	0.6015
β_2	200	0.0528	0.0557	0.0525	0.4934	0.5973	0.6484
	800	0.0491	0.0538	0.0545	0.5018	0.5999	0.6710

Table 2.16 RB-TFT with Normal shocks

		Size Distortion (m)			Power (m)		
	n	1	2	4	1	2	4
β_0	50	0.0307	0.0857	0.1335	0.1212	0.4510	0.6115
	200	0.0661	0.0465	0.1128	0.3956	0.2694	0.5642
	800	0.0606	0.0621	0.0702	0.3813	0.3825	0.4167
β_1	50	0.0485	0.0373	0.0537	0.4752	0.3643	0.4670
	200	0.0461	0.0485	0.0464	0.5079	0.5229	0.5127
	800	0.0459	0.0487	0.0479	0.5402	0.5571	0.5486
β_2	50	0.0501	0.0558	0.0604	0.3978	0.5338	0.5979
	200	0.0489	0.0520	0.0547	0.4813	0.5922	0.6472
	800	0.0488	0.0534	0.0525	0.4934	0.6080	0.6700

Table 2.17 AR-sieve with Normal shocks

		Size Distortion (m)			Power (m)		
	n	1	2	4	1	2	4
β_0	50	0.0435	0.0840	0.1314	0.2123	0.4455	0.6169
	200	0.0416	0.0796	0.1029	0.2581	0.4615	0.5449
	800	0.0439	0.0692	0.0784	0.2772	0.4146	0.4689
β_1	50	0.0508	0.0427	0.0571	0.4976	0.4018	0.5001
	200	0.0499	0.0488	0.0474	0.5363	0.5281	0.5276
	800	0.0499	0.0495	0.0518	0.5620	0.5463	0.5604
β_2	50	0.0544	0.0544	0.0611	0.4378	0.5290	0.5883
	200	0.0515	0.0593	0.0544	0.4976	0.6202	0.6578
	800	0.0531	0.0558	0.0528	0.5168	0.6189	0.6637

Table 2.18 Asymptotic test with normal shocks

2.10.3 Fractional Dependence

Fractionally integrated process with integration parameter d

	n	Size Distortion (d)		Power (d)	
		0.1	0.3	0.1	0.3
	50	0.1094	0.2962	0.7619	0.7512
β_0	200	0.1643	0.4057	0.7997	0.7775
	800	0.2013	0.4871	0.7805	0.8220
	50	0.0468	0.0478	0.7144	0.7052
β_1	200	0.0476	0.0510	0.8179	0.7696
	800	0.0501	0.0481	0.8365	0.7540
	50	0.0505	0.0561	0.7775	0.6877
β_2	200	0.0539	0.0533	0.8747	0.6936
	800	0.0499	0.0493	0.8882	0.6764

Table 2.19 AFB with normal shocks

	n	Size Distortion (d)		Power (d)	
		0.1	0.3	0.1	0.3
	50	0.2314	0.4334	0.9113	0.8668
β_0	200	0.3071	0.5617	0.9156	0.8978
	800	0.3564	0.6127	0.8993	0.9119
	50	0.0519	0.0463	0.7400	0.6863
β_1	200	0.0508	0.0526	0.8251	0.7692
	800	0.0510	0.0523	0.8439	0.7670
	50	0.0525	0.0549	0.7817	0.6930
β_2	200	0.0515	0.0541	0.8660	0.7148
	800	0.0513	0.0521	0.8864	0.6744

Table 2.20 RB-TFT with normal shocks

	n	Size Distortion (d)		Power (d)	
		0.1	0.3	0.1	0.3
	50	0.0973	0.2532	0.7175	0.6985
β_0	200	0.1529	0.3712	0.7804	0.7521
	800	0.1893	0.4516	0.7724	0.7999
	50	0.0438	0.0433	0.6900	0.6664
β_1	200	0.0507	0.0494	0.8143	0.7508
	800	0.0481	0.0517	0.8311	0.7605
	50	0.0493	0.0534	0.7663	0.6674
β_2	200	0.0501	0.0523	0.8615	0.6971
	800	0.0489	0.0498	0.8760	0.6677

Table 2.21 AR-sieve with normal shocks

	n	Size Distortion (d)		Power (d)	
		0.1	0.3	0.1	0.3
	50	0.16586	0.16586	0.16586	0.16586
β_0	200	0.18592	0.18592	0.18592	0.18592
	800	0.21870	0.21870	0.21870	0.21870
	50	0.0938	0.0874	0.8932	0.8679
β_1	200	0.0593	0.0604	0.8591	0.8116
	800	0.0546	0.0552	0.8514	0.7864
	50	0.1068	0.1217	0.9278	0.8881
β_2	200	0.0675	0.0802	0.9144	0.8133
	800	0.0562	0.0630	0.9024	0.7390

Table 2.22 Asymptotic test with normal shocks

2.10.4 Phillips-Perron unit root test

The left-hand blocks show rejection rates when the null hypothesis is true (series cumulated in the DGP). The right-hand blocks show rejection rates when the null hypothesis is false (series not cumulated in the DGP).

	Size Distortion (ρ)				Power (ρ)			
n	0	0.3	0.6	0.9	0	0.3	0.6	0.9
50	0.0491	0.0261	0.0381	0.0188	1	0.9997	0.9191	0.0997
200	0.0513	0.0547	0.0450	0.0359	1	1	1	0.8878
800	0.0503	0.0586	0.0474	0.0432	1	1	1	1

Table 2.23 AFB with normal shocks

	Size Distortion (ρ)				Power (ρ)			
n	0	0.3	0.6	0.9	0	0.3	0.6	0.9
50	0.0490	0.0263	0.0343	0.0187	0.9997	0.9977	0.9292	0.0895
200	0.0503	0.0526	0.0439	0.0340	1	1	1	0.8893
800	0.0504	0.0590	0.0475	0.0426	1	1	1	1

Table 2.24 AFB with χ_1^2 shocks

	Size Distortion (ρ)				Power (ρ)			
n	0	0.3	0.6	0.9	0	0.3	0.6	0.9
50	0.0546	0.0382	0.0424	0.0232	0.9995	0.9985	0.9368	0.1067
200	0.0512	0.0575	0.0482	0.0385	1	1	0.9999	0.9063
800	0.0520	0.0594	0.0480	0.0446	1	1	1	1

Table 2.25 AFB with $t(3)$ shocks

	Size Distortion (ρ)				Power (ρ)			
n	0	0.3	0.6	0.9	0	0.3	0.6	0.9
50	0.0433	0.0304	0.0258	0.01036	1	0.9991	0.8601	0.0616
200	0.0432	0.0408	0.0347	0.02064	1	1	1	0.7766
800	0.0435	0.0450	0.0413	0.0262	1	1	1	1

Table 2.26 RB-TFT with Normal shocks

n	Size Distortion (ρ)				Power (ρ)			
	0	0.3	0.6	0.9	0	0.3	0.6	0.9
50	0.0496	0.0361	0.0328	0.0174	1	0.9985	0.73912	0.02368
200	0.0487	0.0477	0.0421	0.0300	1	1	1	0.61282
800	0.0499	0.0491	0.0474	0.0333	1	1	1	1

Table 2.27 MBB with block length $\lceil n^{1/3} \rceil$

n	Size Distortion (ρ)				Power (ρ)			
	0	0.3	0.6	0.9	0	0.3	0.6	0.9
50	0.0515	0.0362	0.0347	0.0183	1	0.9994	0.8359	0.0566
200	0.0515	0.0424	0.0417	0.0330	1	1	1	0.7728
800	0.0486	0.0470	0.0453	0.0364	1	1	1	1

Table 2.28 Stationary blocks bootstrap (SBB) with mean block length $\lceil n^{1/3} \rceil$

n	Size Distortion (ρ)				Power (ρ)			
	0	0.3	0.6	0.9	0	0.3	0.6	0.9
50	0.0676	0.0296	0.0165	0.0058	1	0.9994	0.7471	0.0248
200	0.0571	0.0331	0.0245	0.0167	1	1	1	0.6449
800	0.0540	0.0390	0.0342	0.0214	1	1	1	1

Table 2.29 Asymptotic test with normal shocks

n	Size Distortion (ρ)				Power (ρ)			
	0	0.3	0.6	0.9	0	0.3	0.6	0.9
50	0.0686	0.0330	0.0238	0.0098	0.9998	0.9980	0.8238	0.0348
200	0.0582	0.0357	0.0278	0.0197	1	1	1	0.6785
800	0.0516	0.0407	0.0337	0.0221	1	1	1	1

Table 2.30 Asymptotic test with t3 shocks

2.10.5   Empirical Variances of Series Means

This table is mentioned, as a note, in the conclusion section. The data has been generated as in the model 2.4 with an $AR(1)$ error terms with different ρ (50,000 Monte Carlo replications).

	Sample mean				AFB bootstrap mean			
n	0	0.3	0.6	0.9	0	0.3	0.6	0.9
50	0.979	1.981	5.789	71.94	1.264	2.183	7.328	73.14
200	0.998	2.027	6.184	92.87	1.101	1.993	6.553	85.75
800	0.999	2.043	6.246	98.47	1.106	2.001	6.471	97.22

Table 2.31 Variances of the two sample means

Chapter 3

Frequency-based Bootstrap Methods for DC Pension Plan Strategy Evaluation

3.1 Introduction

In Pension investment literature, it is important to choose the strategy that provides the best real income after retirement, especially seeing that individual investors must manage their funds due to the transition from defined benefit pension (DB) to defined contribution (DC) pension plans (Cannon & Tonks 2008). Despite the abundance of different models for long-term investments, in theory, the choice of the best strategy should be the one that maximises the real reward-to-risk trade off over a specific investment horizon. The computation of the reward-to-risk trade-off of long-term investment strategies requires, among other things, an accumulation model plot and data generation of asset returns which accounts for scenario analysis.

An accumulation model design mainly involves annual contributions, and it may incur lots of other complexity depending on its role to serve the research topic in question. The generation of asset returns requires some assumptions about the possible future situations and takes into consideration previous historical observations. Formerly, practitioners and researchers have applied Monte Carlo and bootstrap simulation-based methods to generate asset-return samples to build different distributions for long-term cash-flow performance (CFP) measures such as terminal accumulated wealth (TAW) (Blake et al. 2001).

In the Monte Carlo simulation of asset returns, practitioners assume that the perceived observations follow a particular data generating process (DGP). To create a TAW distribution, practitioners need to simulate different artificial asset-return paths[1] based on pre-specified

---

[1]Each path represents a hypothetical investor's lifetime return on both equities and bonds.

DGP. Then for each path, only one TAW may be computed; hence, for different artificial samples, various TAW values are observed. Thus, a distribution of TAW can be obtained from many artificial samples resulting in multiple values.

Bootstrap simulations are computer-intensive methods for estimating distributions by resampling the original data. In Pension literature, due to the lack of long-term observations, practitioners often use a bootstrap simulation of asset returns. In this simulation-based approach, the practitioner does not need to assume any particular DGP for the asset returns, and the simulation occurs non-parametrically. Therefore, for each resampled bootstrap sample only one TAW may be computed. When creating many bootstrap samples, a practitioner would be able to produce many artificial observations of TAW, thus creating an empirical distribution of TAW.

Both Monte Carlo and bootstrap simulations of asset returns involve some inaccurate assumptions that have been previously commonplace in finance. For example, the Monte Carlo simulation of asset returns is model-sensitive; hence, if the assets were not truly generated from such a particular distribution the modelling of the TAW distribution, might be misleading. Thus, lots of former researchers have favoured the bootstrap simulation method of asset returns due to its nonparametric aspect (see, among others, Wang 2012, Moore & Nambiar 2012, Dichtl & Drobetz 2011). However, the bootstrap methods were originally designed to build a confidence interval for simple statistics such as the mean (Efron 1979, Cogneau & Zakamouline 2013). Moreover, theoretical econometricians have proven that the original bootstrap is incapable of producing a confidence interval, even for simple statistics, when some characteristics of the DGP are not met. In particular, the standard bootstrap (SB) may not be applied when the data suffer from heavy-tails, even in the independent and identically distributed (i.i.d.) case (Athreya 1987). Moreover, the moving block bootstrap (MBB) method breaks the stationarity of the originally stationary-observed series (Romano & Wolf 1999).

Although these beliefs of bootstrap failure are well-known in econometric literature, in long-term investment literature practitioners still apply the SB and MBB methods, believing that they preserve many of the characteristics of the actual distribution, including the mean, standard deviation, fat tails, and skewness of returns (Lee 2013). As a result, assuming inaccurate assumptions led them to reach the conclusion that there is no significant impact on long-term return analysis between Monte Carlo and bootstrap methods. Based on these false assumptions and other available literature that lay stresses on the fact that there is not much weight on the choices of generating asset returns from Monte Carlo or Bootstrap simulations that affect the outcome of TAW, many researchers use normal distributions to generate asset returns (see, Byrne et al. 2006, p. 219).

Cogneau & Zakamouline (2013) mention the inadequacy of utilizing either the SB or MBB in this setting. In particular, the MBB was designed to deal with constructing a confidence interval for a parameter (also known as a statistic) that is relatively well-understood, like the mean, median, correlation or standard deviation of the unknown primary distribution of the DGP. However, financial practitioners apply the MBB to construct a confidence interval for a statistic that involves successive intervals. For example, they rely on relatively short-term returns, namely annual, to assess long-term CFP measures. Moreover, they have found that the MBB suffers from estimation bias which may only be avoided if specific assumptions are made about the nature of the asset returns DGP.

Inspired by Cogneau & Zakamouline's (2013) findings, some researchers reported results based on a range of block sizes (Forsyth & Vetzal 2016), while Dichtl et al. (2016) used different bootstrap techniques such as the stationary bootstrap of Politis & Romano (1994) and the matched block size method. However, it should be noted that in practice it is often not well-understood how to decide upon which bootstrap method to apply in long-term investment, especially when the DGP of asset returns is unknown in general.

Therefore, choosing between the different available bootstrap methods in the econometric literature to solve practical implementation for assessing long-term decisions requires careful compromises. In particular, for a practitioner to decide which bootstrap simulation-based method must be applied, he or she needs to make some assumptions about, at least, the characteristics of the DGP of asset returns. This may be achieved by forming "Prior knowledge" of asset returns DGP by applying various univariate and multivariate econometric tests on realized observations. Then, based on the "prior" information, one may create models that represent the DGP. It might occur that there is more than one possible econometric model from which the asset returns might be generated, because the DGP of asset returns is never accurately known. Therefore, deciding which bootstrap method needs to be implemented might be a puzzle due to the lack of comparison between different bootstrap methods in the finance literature.

To the best of our knowledge, we believe that the limitation of applying different bootstrap methods to assess a long-term investment outcome has never been studied in the previous literature. Ideally, the bootstrap that needs to be selected in any long-term investment is the bootstrap that generates a distribution of TAW that coincides with the actual TAW distribution, if data is sufficient. Moreover, in the finance literature, frequency-based bootstrap methods and other time-domain bootstrap methods that are gaining growing familiarity in the econometric literature are seldom used. Therefore, our work will address the limitation of applying different bootstrap methods, including the frequency-based ones, to assess a long-term investment outcome. Based on our Monte Carlo experiments, we find that the

Multivariate Fourier bootstrap (MFB) is the most robust bootstrap method that needs to be implemented to generate CFP measures' distribution similar to its original one, if data is available. We will also compare several pension strategies across different countries using previous bootstrap methods such as Standard Bootstrap (SB) and Moving Block Bootstrap (MBB), along with the MFB, to understand the disputes financial researchers have when recommending the optimal strategy.

The remainder of this work is organised as follows: Section 3.2 presents the latest financial background on investment strategies in long-term investment such as pension plans. Section 3.3 proposes the accumulation model which will be used to compare different long-term parameters across different bootstrap methods and countries. Section 3.4 addresses the difficulties in estimating long-term parameters in pension-plan literature using conventional, simulation-based methods used in econometric literature. Section 3.5 reconsiders previous asset returns generating simulation-based methods used in pension literature. Section 3.6 explains the gathered data which will be used for analysis in section 3.7 to give prior information of the DGP of the assets returns. Section 3.8 illustrates our Monte Carlo Simulation exercise for selecting the best bootstrap procedure that mimics the true distribution for each of the various CFP measures. Based on the choice of the chosen bootstrap method, we will rank strategies for a number of countries, based on previous as well as new bootstrap methods in section 3.9.

3.2   An introduction to pension investment strategies

The investigation of the reward-risk trade-off of long-term investment strategies is currently experiencing a boom in the finance literature due to the shift by many governments, public entities, and large corporations from defined benefit (DB) to defined contribution (DC) schemes (Poterba et al. 2006, Cannon & Tonks 2008). In the DC pension plans, the pensioner bears the total risk of the outcome of his or her investment decisions including inadequate contributions, provided asset price, longevity and inflation (see, Blake et al. 2001). Thus, it is quite important for pensioners to recognise the strategy that will generate the highest real return at retirement.

As a resolution, lots of researchers started introducing saving plans, especially for retirement, that reduce risk in the long-run. Specifically, the life-cycle (LC) strategy is the most common approach among governments, institutions, and other entities for investment (Schleef & Eisinger 2007). Under this plan, pensioners invest in stocks in the early years and then gradually switch into less volatile assets like bonds, bills, and cash in later years as the investor ages (Shiller 2006).

In fact, in pension-investment literature, there is little agreement among pension scholars on the asset allocation investment strategies for multi-period long-term investors such as pensioners. For example, Blake et al. (2001) promote high-equity content for long-term investors, whereas Siegel (1998) recommends 100 percent stocks. This proposal is based on the notion of time diversification; equity stocks become less volatile as the time horizon strengthens (Siegel 1998, Campbell & Viceira 2002). Brennan et al. (1997) showing that bonds and stocks' returns are mean-reverting and hence are less risky from the perspective of a long-term investor.

Nonetheless, Basu et al. (2011) show that dynamic switching in asset allocation, considering the retirement-target return, results in a higher terminal value of portfolio retirement compared with the LC plan in the US. In particular, they propose a dynamic lifecycle approach that rearranges asset allocation based upon the annual accumulated wealth compared with the target return. A similar approach was proposed by Wang (2012), which is called dynamic lifecycle with a momentum (DLCM) strategy, and combines LC strategy with market return movements. Furthermore, for comparison reasons, the 50% stocks and 50% bonds strategy is another fixed asset allocation strategy used in the literature (see, Booth 2004, Shiller 2006).

The comparison of most of the above strategies is not uniform. For example, most of these strategies aim to maximise nominal rather than real terminal values (Donnelly et al. 2015). Moreover, pension researchers have shown that their designed plans are superior to other strategies for a particular country like the US, but not globally. Previously, most of the strategies were compared using the TAW distribution rather than other CFP measures like the internal rate of return (IRR) and the geometric mean return (GMR). Three questions are raised in this sense:

1. Would a particular strategy outperform another when real asset returns are compared rather than nominal asset returns?

2. Would an individual strategy outperform another only in particular countries?

3. Would a certain strategy outperform another based on TAW distribution while not based on other CFP measures, like the IRR and the GMR?

We will approach these questions in the following section by implementing a reasonable uniform accumulation model across different countries and selecting a reliable bootstrap method for simulating asset returns that mimics the possible future real scenarios of an individual's terminal wealth based on historical observations.

3.3   The accumulation model

To reduce most of the risks involved in DC investment plan[2] and to compare the outcome of strategies across different countries, our accumulation model considers the homogeneity of pensioners' initial wage; each citizen of a particular country starts with an initial salary, S_0, of 10 units of the country of interest, c. Moreover, our model assumes a homogeneity of the pensioners' contribution rate that is a constant for all countries and set to 10%. Hence, the initial contribution to the pension retirement wealth for each country is equal to one unit of value of the country's currency.

Furthermore, just after the initial salary, the second salary of each hypothetical participant will depend on the annual wage growth $g_{c,t}$, which is heterogeneous and different in each country c at a particular time t. We assumed, in our model, that the real wage growth is constant over time (denoted as g_c). In our simulations, we defined the annual wage growth-rate of country-specific c (g_c) as the average wage growth of that particular country c during the period spanning between 1900 and 2007[3]. Hence, the annual salary of an individual living in the country c at a particular point of time t, $S_{c,t}$, can be given as

$$S_{c,t} = S_0 \left(1 + g_c\right)^t$$

where S_0 represents the initial salary.

A simple retirement accumulation model will be adopted to determine the expected distribution of a retirement portfolio under different strategies in a DC pension plan. We consider the following stream of salaries for individuals living in country c, starting their contribution at year 1, and retiring at period T.

---
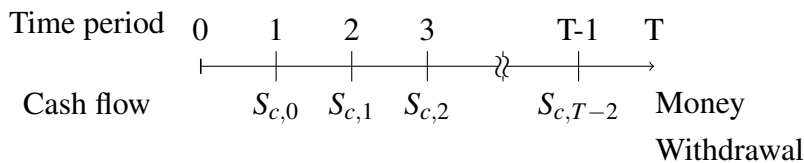
[2]In a DC plan, members suffer from a range of risks involved such as: charges and other costs imposed by the DC plan provider, risk of inadequate contributions, asset price risk, interest-rate risk, and inflation risk (see, Blake et al. 2001, P. 188)

[3]These growth rates are tabulated in Table 3.1

In this case, the TAW value of the pension fund at retirement time T in a particular country c under a given strategy k, $W_{T,c,k}$, is given by

$$W_{T,c,k} = 0.1 \sum_{t=0}^{T-2} S_t \prod_{u=t+2}^{T} \left(1 + r_{c,u,k}\right)$$

where 0.1 stands for the plan contribution rate and $r_{c,u,k}$ represents the rate of return for each investment strategy k, at time u, and in country c.

It is worth noting that the outcome of such simulations would represent, for each country, an index rather than currency quantities in real money (see, Shiller 2006, Basu et al. 2011). Indeed, a 10 unit of fixed starting salary and a 10% fixed contribution rate implies that one unit is invested in the pension fund in the first year. Thus, if the terminal accumulated wealth $W_{T,c,k}$ turns out to be 150, this means that the end-of-period wealth is 150 times the initial contribution.

We find that the value of TAW is dependent on the portion of the investor's annual salary contributed to his or her retirement plan and the rate of return earned during different periods of time until the final year before retirement. It therefore depends on the strategy of investment as well. Considering that the rate of return changes with the investment strategy, to reduce the sources of randomness, we assume that asset-allocation strategies within the DC pension plans are only dependent on shifting between risky equities and riskless bonds. Therefore the portfolio return, $r_{c,t,k}$, is a combination of the return on equities, e_t, and the return on bonds, b_t. Denote by $\alpha_{t,k}$ the proportion invested in equities at time t for each strategy k; therefore, the investment return is defined as

$$r_{t,k} = \alpha_{t,k} e_t + \left(1 - \alpha_{t,k}\right) b_t \tag{3.1}$$

The pension portfolio return, $r_{t,k}$, is a key variable in the model as it indicates investment strategy $k = 1, 2, 3, 4, 5$ that assigns the weights of different assets in the retirement portfolio. For example, in the 50 percent stocks and 50 percent bonds strategy, the variable $\alpha_{t,2}$ is equal to 0.5, which is a constant proportion in equities. Thus, the variable $\alpha_{t,k}$ is determined by the strategy being implemented.

3.3.1 The mathematical representations of different strategy weights in pension portfolio returns

1. In a 100 percent stocks strategy, the investor invests 100 percent in equities and holds them for the entire investment period T, thus $\alpha_{t,1} = 1$, for $t = 2, 3, \ldots, T$.

2. In a 50 percent stocks and 50 percent bonds strategy (50-50 strategy), as discussed earlier, the investor invests 50 percent of his contribution in equity and the rest in bonds, hence, $\alpha_{t,2} = 0.5$, for $t = 2, 3, \ldots, T$.

3. In a simple LC strategy, the investor invests 100 percent in stocks for thirty years and then gradually switches stocks to bonds at the beginning of each year. Thus, the proportion invested in equities, $\alpha_{t,3}$, is dependent on time t, and given by

$$\alpha_{t,3} = \begin{cases} 1 & \text{if } t \leq 30 \\ 1 - 0.1\,(t - 30) & \text{if } t = 31, \ldots, 40 \end{cases}$$

4. In the DLCM strategy, the investor invests 100 percent in a stocks portfolio, as in a simple lifecycle strategy, until the asset-allocation switching period. Rather than automatically switching from stocks to bonds linearly each year, the asset allocation decision is based on the previous year's stock return e_{t-1}. Therefore, the proportion invested in equities, α, depends on time t as it will randomly change depending on the random observed value e_{t-1}. If $t \leq 30$ then $\alpha_{t,4} = 1$. For $t = 31, \ldots, 35$ $\alpha_{t,4}$ is given by:

$$\alpha_{t,4} = \begin{cases} 1 & \text{if } e_{t-1} > 0.1 \\ 0 & \text{if } e_{t-1} < -0.1 \\ 0.8 & \text{if } e_{t-1} \in [-0.1, 0.1] \end{cases}$$

For $t > 35$, α_t is given by:

$$\alpha_{t,4} = \begin{cases} 1 & \text{if } e_{t-1} > 0.1 \\ 0 & \text{if } e_{t-1} < -0.1 \\ 0.6 & \text{if } e_{t-1} \in [-0.1, 0.1] \end{cases}$$

5. In the DLCT accumulation strategy, the investor's asset allocation decision is based on the past performance of the retirement-accumulated wealth. For example, during the switching period, if the investor has accumulated wealth greater than the target accumulation ($W_{t,c,5} > \tilde{W}_{T,c}$) at the beginning of each year, the portfolio is partially switched to bonds; otherwise, it remains as 100 percent in stocks. However, if the switch has begun and the accumulated wealth drops below the target ($W_{t,c,5} < \tilde{W}_{T,c}$), the fund is switched back to stocks. Therefore, again, the proportion invested in equities α is dependent on time t as it will randomly change, depending on the random observed value of the accumulated wealth $W_{t,c,5}$. For $t \leq 30$, $\alpha_{t,5} = 1$. For $t = 31, \ldots, 35$, α_t is

given by:

$$\alpha_{t,5} = \begin{cases} 1 & \text{if } W_{t,c,5} \leq \tilde{W}_{t,c} \\ 0.8 & \text{if } W_{t,c,5} > \tilde{W}_{t,c} \end{cases}$$

For $t > 35$, α_t is given by:

$$\alpha_{t,5} = \begin{cases} 1 & \text{if } W_{t,c,5} \leq \tilde{W}_{t,c} \\ 0.6 & \text{if } W_{t,c,5} > \tilde{W}_{t,c} \end{cases}$$

where $\tilde{W}_{T,c}$ is the target accumulated wealth at period T. Mathematically it is dependent on the growth rate g and target rate of each country c, and is represented by

$$\tilde{W}_{t,c} = \sum_{i=0}^{t-2} (1 + g_c)^i (1 + \tilde{r}_{T,c})^{t-i-1}$$

where $\tilde{r}_{T,c}$ and g_c are the target return represented by the geometric mean of the annual real equity-return and the salary growth-rate of each country c respectively. The investor's target return is different in each country as the stock returns in each country are different. Table 3.1 gives our calculated values of the investors' target return and average wage growth for each country.

Note that the observed accumulated wealth $W_{t,c,5}^*$ for each individual in a given country c is different depending on his or her asset allocation whose the return at a given period t will be denoted by $r_{c,t}^*$. In this case the accumulated wealth in a given country c and time t, $W_{T,c,k}^*$, is computed as

$$W_{t,c,k}^* = \sum_{i=0}^{t-2} (1 + g_c)^i \prod_{j=i+2}^{t} \left(1 + r_{j,c,k}^*\right)$$

Table 3.1 Target return & average real wage growth constant values used across countries

Country	Target return	Average real wage growth
Australia	7.9%	1.0%
Belgium	2.5%	-2.0%
Canada	6.3%	1.0%
Denmark	5.3%	2.0%
France	3.7%	2.0%
Germany	3.4%	1.0%
Ireland	4.6%	1.0%
Italy	2.5%	1.0%
Japan	4.3%	6.0%
Netherlands	5.4%	4.0%
Norway	4.5%	1.0%
South Africa	7.5%	0.0%
Spain	4.0%	2.0%
Sweden	7.8%	2.0%
Switzerland	4.5%	2.0%
UK	5.5%	1.0%
USA	6.5%	1.0%
World	5.07%	1.6%

[1] The Target return $\tilde{r}_{T,c}$ is being computed as the geometric mean across all 108 observations of the annual Real Equity TR (% Total Return) for each country.

[2] The Average real wage growth, g_c, used is the mean and it is dependent on each country c.

The idea of target returns in the target accumulated wealth was introduced by Basu et al. (2011), whereby a dynamic life-cycle strategy with a target return in which the switching of stocks to fixed-income assets depends on whether or not the retirement accumulation meets the investor's target-return at that particular moment in time. In this chapter, the investor's target return is computed as the geometric mean of the annual real stock return for each country.

In theory, maximising the investor's expected utility, or terminal accumulated wealth over a particular holding period, is referred to as the optimal portfolio. In practice, pension-related researchers often lack sufficient data to compute the estimated risk and return. Therefore, some of the practitioners often generate data from estimated parameters or rely on asymptotic theory, both of which provide a poor approximation to the actual distribution of the statistic of interest; hence, the process of doing so has always been associated with uncertainty, as

it may lead to misleading inferences. Thus, finance researchers, to avoid difficulties in the method of generating returns, and to improve their observed finite-sample performance, often refer to powerful simulation techniques such as bootstrap methods.

Before specifying the bootstrap utilised in this chapter, it is of importance to discuss the parameters required for assessing different pension strategies. As discussed earlier, we will be using the TAW, IRR, and GMR distributions for assessing different strategies. In the next section we discuss these parameters with mathematical rigour to show that these statistics depend on a probability distribution of joint returns rather than probability distribution of individual return.

3.4   The concerns of estimating long-term parameters

As discussed by Cogneau & Zakamouline (2013), formerly, bootstrap methods were designed to find estimates of unknown parameters like mean, median, variance, correlation, etc. which involve only one (univariate or multivariate) distribution. However, in finance, investors are not usually interested in estimating such parameters; rather, they opt to study risk and reward over a range of subsequent periods. Hence, temporal structure matters in such parameters. Particularly, in finance, practitioners are interested in determining parameters with temporal joint (Univariate or Multivariate) distributions of successive intervals of returns.

Although our parameters are similar in their construction to those of Cogneau & Zakamouline (2013), our settings are different for two reasons. The first reason is that, in our model, the assessment of strategies in pension planning involves the pension-accumulation model. For example, the Sharpe Ratio will be computed based on simulated observations such as the TAW, IRR, and GMR rather than long-term returns. The second reason is that the performance measures used to evaluate different strategies are not the same as those used in their paper. In particular, our analysis is not limited to the Sharpe Ratio; other performance measures will be employed as well.

Having said this, it is important to mention the mathematical relationship between TAW, IRR, and GMR and the temporal joint distributions of nominal returns. Following the same notation as in previous sections, let $\rho_{T,k} = (r_{1,k}, r_{2,k}, \cdots, r_{T,k})$ be an observed sample of returns which is based on strategy k and a fixed time horizon T that represents intended retirement date. The TAW may be simplified and represented as

$$W_{T,k} = \sum_{i=0}^{T-2} (1+g)^i \prod_{j=i+2}^{T} \left(1+r_{j,k}\right) \tag{3.2}$$

This implies that the TAW may be defined by a function, η, such that

$$W_{T,k} = \eta\left(g, T, \rho_{T,k}\right)$$

where g, T, $\rho_{T,k}$ are as mentioned earlier. Based on the TAW function, the IRR, β_k, may be easily obtained by solving the non-closed form solution of the following equation

$$\left(\frac{1+\beta_k}{\beta_k - g}\right)\left[(1+\beta_k)^{T-1} - (1+g)^{T-1}\right] = \eta\left(g, T, \rho_{T,k}\right) \qquad (3.3)$$

Using Descartes Rule of Sign, there is at most one IRR if the cash flow changes sign once (Turnbull 1952, Norstrøm 1972, p. 99-102). If this happens, then there is a unique solution for equation (3.3), which we denote by

$$\beta_k = \phi\left(g, T, \rho_{T,k}\right)$$

It is also straightforward to observe that the GMR, $\delta_{T,k}$, may be computed by the following function

$$\begin{aligned}
\delta_{T,k} &= \gamma\left(T, \rho_{T,k}\right) \\
&= \left(\prod_{j=2}^{T}\left(1 + r_{j,k}\right)\right)^{1/(T-1)} - 1
\end{aligned} \qquad (3.4)$$

In all of the above CFP measures, the variable $\rho_{T,k}$ is introduced. It represents the temporal joint distribution of the nominal returns $\{r_{j,k}\}_{j=2}^{T}$. Hence, all the represented CFP measures depend on the temporal joint distributions of the nominal returns.

Researchers are usually interested in determining the strategy's risk as well as its performance measure to rank strategies. These measures are estimated based on available data. However, due to the lack of sufficient data, professionals resort to simulation methods. To illustrate the ideas briefly, consider that one has observed a sample of T nominal returns $\rho_{T,k}$. Then one may only compute one value for each of the CFP measures, namely the TAW, IRR, and GMR. The strategy's risk-measures like the VaR or Shortfall Risk (SR) may not be computed based on one sample only. Moreover, even if the observed sample, $\rho_{n,k}$, of nominal returns has a length of size n such that $n > T$, if n is not much greater than T, the intended year of retirement that represents individual lifetime, one may not obtain many non-overlapping blocks to compute these parameters. Therefore, practitioners resort to simulation methods that enable them to create many pseudo-samples that allow them to estimate such long-term parameters.

It should be highlighted, however, that simulation-based approaches, like the bootstrap methods, were not originally designed for computing such parameters, say the Sharpe Ratio which is obtained based on CFP measures (TAW, GMR, IRR) that are themselves dependent on the joint temporal structure of nominal returns $\rho_{T,k}$. Therefore, estimating parameters such as the Sharpe Ratio using a conventional bootstrap approach needs to be justified. We will verify that when applying a particular bootstrap method on datasets that attain some mild assumptions about the DGP, a practitioner may find reasonable estimates of such long-term performance parameters that mimic reality. The next section presents a literature review on simulation-based methods used in pension research to generate pseudo asset-return samples.

3.5   Previous literature on asset return DGP

Previously, scholars and actuaries often favoured the Monte Carlo simulation to evaluate the investment risk for both DB and DC pension plans, (see, Blake et al. 2001, Johnston et al. 2001). However, Monte Carlo simulations have serious consequences. They generate data from a predetermined probability distribution, which of course contradicts reality, as it is never certain how data are generated. Another misconception implemented by a significant number of practitioners, and one that almost all scholars fall into, is the classical idea that the returns are normally distributed and their correlations constant. Therefore, scholars and practitioners have been moving to other simulation-based methods such as the Bootstrap.

Initially, Efron (1979) introduced this method of generating pseudo-samples independently from i.i.d. observations. The practical idea from a financial perspective is that a practitioner may obtain a pseudo-sample from historical observations of returns which contain good and bad events. Therefore, creating new observations based on such historical data provides a hypothetical scenario of reality. Simulating many hypothetical scenarios may allow practitioners to evaluate long-term parameters. Ultimately, this method of simulating returns using Standard Bootstrap methods needs some assumptions on the DGP of stock returns, namely independence, and identical distribution of returns. This approach has been conducted by many authors including Lloyd & Modani (1983), Leibowitz & Langetieg (1989), and Dierkes et al. (2010), among others.

Ever since its introduction, statisticians and econometricians have always tried to weaken those assumptions to produce data from time-dependent observations. One of the most prominent bootstrap methods that are well-known to practitioners is the MBB. Although the MBB may be considered more robust than its counterparts, there is always a loss in the efficiency of such a bootstrap procedure (Liu & Singh 1992). The MBB approach involves partitioning the data into blocks of equal terms or length b, either non-overlappingly (see,

Carlstein 1986) or overlappingly (see, Künsch 1989). In our simulation, we adopted the overlapping case because it may provide a higher estimation efficiency than non-overlapping blocks (see, Hall & Yao 2003). However, there are not much differences between these two methods in the applied world (see, Horowitz 2001). Briefly, the overlapping MBB procedure considers that the vectors in original observations are divided into overlapping blocks of length b. Hence, we would end up with $l = T - b + 1$ blocks. Then, to obtain a bootstrap sample, the algorithm needs to sample randomly from the blocks with replacement from all possible blocks.

Politis & Romano (1994) argue that in the finite dimensional space when the number of observations is limited, the blocks being stationed next to one another will distort the stationarity property of the observed data set. Hence, the pseudo sample becomes non-stationary while the observed sample is stationary. Note that since the MBB re-samples retain the same dependency structure between two dependent data in a given block, such as the annual real returns for stocks and bonds, the MBB preserves the serial correlation and cross-correlation within the elements of the block but not within the blocks themselves.

In finance, the MBB simulation-based approach has been conducted by a number of scholars for estimating long-term parameters: see Hansson & Persson (2000), Mukherji (2011), and Beach (2007), among others. Unfortunately, as correctly noted by Cogneau & Zakamouline (2013), due to the overlapping resampling blocks, the MBB suffers from 'small-sample bias' which leads to downward biasedness in the estimated variance of multi-period returns. Although our settings are different due to the involvement of an accumulation model, such a dilemma continues to exist when computing long-term parameters for assessing different strategies. Unlike the estimation of the long-term performance measures used to evaluate different portfolio investments, the estimation of long-term performance measures used to evaluate different pension strategies, needs to be conducted on CFP measures (TAW, GMR, IRR) rather than multi-period returns.

Having mentioned the obstacles for each of the conventional simulation methods used in finance, it is of interest to study whether other available bootstrap methods in the econometric literature would serve the purpose of the proposed financial application. The bootstrap methods that are suggested are the Multivariate Sieve Bootstrap (MSB), the Multivariate Hybrid Bootstrap (MHB), and the Multivariate Fourier Bootstrap (MFB).

Originally, the Sieve Bootstrap was introduced to compute certain statistics for the univariate causal linear processes time series (Kreiss 1992). Paparoditis (1996) extended the univariate sieve to the multivariate level due to his interest in approximating joint distributions and show the validity in applying them in that framework. Meyer & Kreiss (2015) have investigated further properties of the MSB. In particular, they study whether the MSB

asymptotically works for a specific statistic or not. Briefly, the MSB procedure involves fitting the observed data-set by a predefined order p, computing the residuals, and then drawing with replacement from the set of centered residuals. Afterward, the obtained pseudo sample of residuals is used to generate the bootstrap sample of the observed data-set. This is known as the residual-based bootstrap replicates of the time series.

Due to the exciting features of the transformed time-dependent data in the frequency domain, Jentsch & Kreiss (2010) were the first to introduce the MHB method, which is a bootstrap method that requires two domains, namely the Time and the Frequency domains in the multivariate context. This method is divided into two main steps to generate the bootstrap observations. The first step involves generating bootstrap samples in the time-domain exactly as the MSB, thus capturing most of the dependency structure of the observed multivariate dataset in the time domain via a parametric fit, namely Vector Auto-Regression (VAR) model. In the second step, the procedure uses the bootstrap observations generated in the first step and transforms them to the frequency-domain via the aid of the Discrete Fourier Transform (DFT). In the frequency domain, a nonparametric correction is applied to the transformed pseudo-observations. Then corrected, transformed pseudo-observations are back-transformed via the inverse-DFT to observe the pseudo data-set in the time-domain. Note that the correction implemented in the frequency-domain is nonparametric, and its sole purpose is to widen the class of time-series models that the MSB was not capable of handling. Although the idea of the MHB is simple, its implementation may not be as straightforward and involves lots of computational steps that are outlined in Appendix C (B).

Both the MHB and the MSB are considered to be either parametric or semi-parametric procedures because they entail fitting data to a pre-selected time-series model, namely VAR. However, the MFB may be a new competitive method that overcomes such parametric fitting. In particular, the MFB is completely non-parametric.

The MFB consists of demeaning the data and transforming them into the frequency domain via the aid of a DFT. In the frequency domain, a Rademacher draw is applied to produce different samples in the frequency domain. Each sample created in the frequency domain will be transferred back to the time-domain via an inverse-DFT, hence, generating many pseudo samples in the time domain. However, these pseudo samples are mean-degenerate; thus, for this bootstrap procedure to be fully functional, it will require additional steps to produce a bootstrap pseudo-series in the time domain. In particular, Davidson and Chehab (2017) have augmented the generated time-series data with a normally distributed error that shares the diagonal elements of the heteroskedasticity and autocorrelation consistent (HAC) variance of the original sample data. Unlike the common variance formula, the HAC

variance accounts for autocorrelation in the back-transformed dataset. Therefore, the HAC variance estimator is preferred to the conventional variance formula.

All the bootstrap methods mentioned are capable of generating pseudo time-series data of the original dataset, provided that the dataset in question is generated from a covariance stationary process. The question, however, relies upon which bootstrap methods provide a reliable procedure for producing distributions for CFP measures that are required to compute long-term performance parameters. To address this issue, we need to understand, based on statistical tests, the characteristics of the DGP of the asset returns. Accordingly, we will form a "prior belief" about the DGP of our asset returns. Based on this formulated prior-belief of the DGP, we will use Monte Carlo experiments to evaluate all the mentioned bootstrap methods. In the next section, we will describe the compiled dataset that will be analysed to formulate prior belief about the DGP of asset returns in different countries.

3.6   Data description

3.6.1   Asset-returns data-set

In this chapter, asset class returns were obtained commercially from Ibbotson Associates. The dataset consists of equity and bonds' global returns compiled by Dimson et al. (2002) (DMS). This data set has been further updated, and it currently holds asset-class returns for seventeen countries from 1900 to 2007. The seventeen countries include Australia, Belgium, Canada, Denmark, France, Germany, Ireland, Italy, Japan, the Netherlands, Norway, South Africa, Spain, Sweden, Switzerland, UK and the US. The reason we used this data is the long time-span for different countries and typical characteristics.

DMS have chosen and, when needed, constructed the best index that ensures the widest coverage of the market in each country during each period as the industrial composition has changed dramatically over the covered 108 years. They followed certain criteria to guarantee the accuracy of that index. First, equity indexes should avoid any bias, such as excluding companies that didn't manage to survive long. Second, all asset-returns are total returns that include income, such as dividends, and capital gains or losses. Therefore, equity indexes are constructed from the stock price and dividends while bond indexes use bond-level price and coupons. Most calculations are based on end-of-year index level, but, when not available, time-averaged or intra-year indexes are used. Third, equity indexes represent the market fully. Fourth, long-term return indexes hold constituents weighted according to their market capital. Finally, the use of data over an extended period and across many countries decreases the bias of data, such as using only a prosperous economy or during the post-war period.

Though DMS provide both nominal and real returns, we have chosen real returns, adjusted for inflation, in all our studies. The year-on-year inflation was based on consumer price indexes (CPI). However, since CPI were developed over time, the most reflective index during each period for each country was used. It is understood that it might not be very precise during some periods, but DMS argue that the chosen inflation indexes track long-term currency fluctuations, which gives some reassurance about their inflation measures. Therefore, the returns used are real total returns adjusted for inflation and dividends. It is quite important to be able to compare returns over time and across countries on the same basis.

Hence, the real total net equity index return, now and hereafter, e_t, can be mathematically defined as

$$e_t = (1 + \pi_t)^{-1} \Big(1 + \frac{P_t + D_t}{P_{t-1}} \Big) - 1$$

where P_t, and D_t refers to the price of the index and dividend. π_t refers to the inflation between periods $t - 1$ and t (see, Dimson et al. 2000).

We also include in the study the global index constructed by DMS. The global index was constructed as an international portfolio that invests in all therefore mentioned countries expressed in a common currency, i.e. US dollars. Hence, each country's local currency-return was converted to US dollars. The weightings of the equity index were based on each country's GDP, converted to US dollars between 1900 to 1967, and on equity market capitalization from end of 1967 onward. The weightings of the bond index were based on each country's GDP throughout the whole period covered.

3.6.2   Annual wage-growth data-set

The wage-growth data used was obtained from the Mitchell (1998) labour-earnings indices from 1900 to 1993 and updated from the Hourly Earnings (MEI) data-set from the OECD Main Economic Indicators. All missing years wage-growth data were supplied by the GDP per capital from Maddison-Project (2008), which is probably one of the best estimates of labour earnings growth. These data were previously used by Cannon & Tonks (2013)[4].

---

[4]We are grateful to these authors for providing this data.

3.7   The initial step: Data analysis of asset-returns and their predicted DGP

Following Cogneau & Zakamouline's (2013) remarks, the MBB may be applied when the observations happen to be serially dependent and stationary. In the literature, it was mentioned that practitioners believe stock returns to be mean-reverting. Hence, they apply the MBB without checking whether the data at hand is stationary or not. It is well known that a stationary series is mean reverting but not vice-versa[5]. Therefore, it is essential to test for stationarity and to accurately specify the model of our DGP, based on our available dataset before applying any bootstrap method.

The first step in forming prior knowledge about the DGP entails a detailed analysis of the characteristics of the time series variables observed in our collected dataset. It is important to gain a good understanding of the properties of individual time series by treating each of a country's bond returns and equity returns independently. Then, afterwards, we may study jointly both series, namely the country's bond returns and equity returns to formulate some of the critical properties that are involved in generating the asset returns (bonds and equities) for each country. The next two subsections will cover brief concepts, procedures, and theoretical results that we employ to formulate our prior knowledge on the DGP of asset returns at both levels: univariate and multivariate.

3.7.1   Univariate analysis

The sample autocorrelation function (ACF) and the partial autocorrelation function (PACF) for each asset-return provide a rough indication, for our univariate time series, of whether the DGP has stationarity or not. The corrologram figures, not shown here, indicate that it is not likely for any of our asset-return samples to be generated by a white-noise process, because some of the sample autocorrelations and sample partial autocorrelations cross the dotted lines. However, lots of the coefficients at higher lags are clearly between the dashed lines. Hence, the underlying ACF of most of the countries' asset-return may be in line with a stationary DGP.

Although the ACF and PACF provide an initial test of stationarity, their indication is not evident. Therefore, we applied proper statistical tests for stationarity. One form of testing a stationarity assumption is to test for a unit root. Each of the widely applicable tests of unit roots (a form of non-stationarity) has been applied on each of our asset-return independently, namely the Augmented Dickey-Fuller (ADF) test, the Phillips-Perron (PP) test, and the

---

[5]Therefore, if supposedly a series is mean-reverting, this does not imply that it is stationary.

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. In our series, we observe some degree of disagreement among the results of these tests. For example, with a 5% significance level, we found that France Real Equity return series suffers from a unit root using the ADF test, but this is not the result using the PP test and the KPSS test. However, in general, we may conclude that all our series are weakly stationary[6]. However, we will rely more on the ADF test, because it is more robust than other tests when applied to data-sets that might be correlated or have witnessed some heavy-tail phenomena (see, Wang & Mao 2008).

In the wake of our previous analysis, it seems that our sample of asset returns for all our sample countries is assumed to be stationary. Recall that the data used are described in Section 3.6. These data-sets are annual data; hence, also known as low-frequency data as opposed to the high-frequency data-set. It is valid to argue that since each data point in our data-set is recorded based on annual observation, the dependence structure changes as opposed to data points that are recorded based on a day-by-day basis. Therefore, it would be natural to reject, based on our sample data, that the asset returns originate from a GARCH type time series process.

To conclude, these data-sets may be originating from an Auto-Regressive Moving Average (ARMA) models. Moreover, we are keen to test whether the data at hand suffers from time-variation in the variance. However, before testing for any Auto-Regressive Conditionally Heteroskedastic (ARCH) effect in our asset return series, we would need to ensure that we have taken advantage of all the information we hold to analyse our univariate series. Particularly, we would like to check whether a country's equity-returns play any significant role in predicting the same country's bond returns, or vice-verse.

The sample cross-correlation function (CCF) is a helpful tool for identifying whether the lags of a variable, say a country's equity returns, are useful when predicting the current outcomes of another variable, say a country's bond returns. As a result, we have plotted the CCF for each country's asset-return where we identified that in all our sample countries, lags of one asset show significant ability to predict the outcome of the other asset. However, examining a significant cross-correlation sample does not imply, for example, that the lags of equity returns better explain the current outcome in the bond returns, because we have not accounted for the lags of the bond returns that help to explain the current bond returns. Consequently, the Granger causality test is required to overcome this obstacle.

In the univariate case, we applied the Granger causality test by running two models. The first represents the unrestricted model where one of the asset returns is regressed on its own lag and the lag of the other asset-return while the other model is the asset return is regressed

---

[6]Perron (1989) argue that in case KPSS test results in rejecting its null hypothesis, it might be misleading to infer that the process is not stationary. For example, there might be structural breaks in that data-set that are leading to such inference.

on its lags only. We have observed that in Australia, Ireland, Switzerland, the UK, and the World, both returns of bonds and equities of the aforementioned countries Granger cause one another. Whereas in countries like Denmark, Italy, and South Africa, equity returns Granger causes bond returns. In France, Norway, and Spain bond returns Granger cause equity returns. While the rest of the countries like Belgium, Canada, Germany, Japan, Netherlands, Sweden, and the US neither asset-returns Granger cause one another. Therefore, in the countries where neither asset returns Granger cause the other, accounting for the lags of either asset returns does not help in better explaining the other corresponding current asset-return.

Based on the mentioned analysis, we have tested for ARCH effect on our series accounting for all relevant series that help in predicting it[7]. For example, testing for ARCH effects for the US equity returns involves the following steps: 1) Fit an ARMA model to this series; 2) extract the residuals of the fitted model; 3) regress their squares on its lags; 4) test the joint significance of the newly formed regression. The result reveals that there exist no ARCH effects in the asset returns for each of the countries Belgium, Canada, Germany, Japan, Netherlands, Sweden, and the US.

Although all the discussed univariate tests reveal important details about individual series, our profound understanding of the properties of the DGP have not reached the optimal point because we have not addressed the joint series distributions and their dynamics. The study of joint series is important, especially when we believe that some of the asset returns in some countries happen to Granger cause one another. Therefore, it is important to treat both country-specific asset returns as being a priori endogenous to avoid some miss-inferences due to the structural cross-correlations between both variables (along with their lags). In the following section, we will continue our analysis in a Multivariate context.

3.7.2   Multivariate analysis

For each country, we considered the corresponding bond and equity returns jointly. Based on our univariate analysis, it is clear that our country-specific joint asset returns dataset does not suffer from a deterministic trend, unit roots, or even ARCH effects. Furthermore, in this study, our model assumes that the dependency structure of equities and bonds for all countries is linear and symmetric in time and between series[8]. Therefore, it is reasonable to

---

[7]It is worth noting that in the univariate analysis, when we regress different country-specific asset returns on one another we are treating one of them as being exogenous to the regressand variable.

[8]Recently, a copula-based model for stationary time series with a Markovian structure has been proposed Brechmann & Czado (2015). This model might be a better than any alternative strictly linear stationary process such as the VAR model because it accounts for non-linear correlations when it comes to multiple time series analysis of financial data. However, this is beyond the scope of this chapter as it deviates from the main aims of this paper and requires further independent research.

suggest that a Vector Auto-Regressive (VAR) model may be a potential model to characterise our joint DGP:

$$\begin{bmatrix} e_t \\ b_t \end{bmatrix} = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} \beta_{1i} & \phi_{1i} \\ \beta_{2i} & \phi_{2i} \end{bmatrix} \begin{bmatrix} e_{t-1} \\ b_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{3.5}$$

where e_t and and b_t are the total equity return and the total bond return, respectively, at time t. In our analysis, we have gone through a list of procedures to specify our model and to check it. This includes, but is not limited to, selecting the VAR order and imposing restrictions on the VAR parameters[9]. The list of our Algorithm is described as follows:

Algorithm Procedure 3.7.1.

1. *Specify the VAR model using the Akaike's information criterion (AIC) to determine VAR lag order specification $\hat{p}$.*

2. *Estimate both the restricted and unrestricted levels VAR of order $\hat{p}$ as in the (3.5).*

3. *Check for final model classification of the DGP*

We employed lots of diagnostic tests on the residuals of the fitted VAR model. For example, to check for auto-correlations, we have performed the multivariate Portmanteau test (Ljung-Box statistics), the multivariate Breusch–Godfrey test and the Rank-Based portmanteau test statistics[10]. Moreover, to test for joint non-normality, we have applied Mardia's, Royston's, as well as, Henze-Zirkler's Multivariate Normality tests[11].

Remark: In some countries such as Belgium, we found that it is more appropriate to apply a transformation to the initial dataset before carrying on with our analysis. In particular, for such countries, we employ a logarithmic transformation on the asset returns. Thus, we transformed the real total net index returns to the log of the total real gross returns, hence replacing (e_t and b_t) with ($\log(e_t + 1)$ and $\log(b_t + 1)$) in equation (3.5). In cases where this has been done, we will refer to our model as log-VAR (or log-VARMA) model. Moreover, in some countries such as Australia, although we observe that the residuals may follow the white-noise process, they may not be generated from a Multivariate Normal process.

---

[9]We have also tried fitting VARMA models. Therefore, when a country's data is better estimated with a VARMA or VMA model we would explicitly mention it in our analysis.

[10]Rank-Based portmanteau test performs nicely and is more robust to heavy-tailed distributions

[11]The Henze & Zirkler (1990) test is based on the empirical characteristic function of the sample data at hand. This test is very sensitive and robust when dealing with distributions that possess a normal or light tails but less powerful when it comes to heavy-tails.

Therefore, in this case, we may not classify the multivariate error term $\begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{bmatrix}'$ in equation (3.5) as multivariate normal errors. In fact, it might be that each of the errors ε_{1t} and ε_{2t} is generated from a different DGP. However, since we are studying joint distribution, one may not generate one variable irrespective of the other. For example, a vector whose elements are white noise is not necessarily, in general, a white noise vector. Thus, the only way to classify the joint distribution of the errors $\begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{bmatrix}'$ is to estimate the best bivariate copula model[12] based on the fitted residuals (or maybe the transformed residuals).

The ability to combine (or couple) marginal distributions with a copula model, that is using the converse part of Sklar's theorem, allows us to focus our attention on modelling the univariate distributions for each residual independently without worrying about their serial dependence structure since this has already been accounted for in our VAR model.

We opt to choose the univariate probability distribution of each residual, namely $\hat{\varepsilon}_{1t}$ and $\hat{\varepsilon}_{2t}$, independently of one another. Our choice of residuals' distribution is based on the skewness-kurtosis plot of Cullen & Frey (1999). We compared the values of the unbiased estimation of skewness and Pearson's kurtosis values of the residuals with the skewness and kurtosis values for common distributions (see, Sokal & Rohlf 1999, p. 111-115)[13]. However, once again in some countries, the computed residual from the fitted parametric models may not follow any of the well-known available distributions. Therefore, some transformations on the residuals, independently, need to be carried out to fit such distributions.

Accordingly, some countries in our data sample may be classified as Copula-based multivariate models. In particular, some countries in our sample dataset would be assumed to follow the VAR model with random vector errors generated from non-Guassian joint distribution. In our analysis, we will assume that the variables (bond and equity errors) are continuous, as it is not possible to believe that these variables originate from a discrete DGP. Moreover, the selection of the bi-variate Copula model for the transformed residuals was considered in our analysis and we implemented using the sequential two-step, maximum-likelihood method (TSML) in which the marginals are estimated in the first step and the dependence parameter is estimated in the second step (see, Manner 2007)[14]. In case, we

---

[12]"copulas are multivariate distribution functions whose one-dimensional margins are uniform on the interval (0,1)." (Nelsen 2006, p. 1). For Further mathematical definition see (Definition 1.1, Ibragimov & Prokhorov 2017, p. 5)

[13]Fortunately, there is built-in function called *descdist* in R, the open source programming language and software environment for statistical computing and graphics, that does the following job. Please consult Delignette-Muller & Dutang (2015) for details.

[14]Fortunately, there is a built-in function in R called the *BiCopSelect* which is available by installing either of the R packages *VineCopula* or *CDVine*. This function selects the best fit copula iteratively based on the information criterion of AIC/BIC.

avoided the treatment of the marginal distribution functions, we estimated the copula based on the semi parametric two-step estimator.

In general, a bi-variate Copula, C, is associated with the bivariate joint cumulative distribution function (CDF), F, given as

$$F_{f(\varepsilon_{1t}),g(\varepsilon_{2t})}(y_1,y_2) = C(F_{f(\varepsilon_{1t})}(y_1),F_{g(\varepsilon_{2t})}(y_2);\theta) \qquad (3.6)$$

where f and g are the transformed functions which allow the transformed residuals to fit well-known available distributions. θ is a parameter of the copula C, called the dependence parameter, which measures dependence between the marginals. $F_{f(\varepsilon_{1t})}$ and $F_{f(\varepsilon_{2t})}$ are the CDF associated with the transformed random variables $f(\varepsilon_{1t})$ and $g(\varepsilon_{2t})$, respectively.

Although we believe that the residuals are uncorrelated, the transformations applied on the fitted residuals will impact the joint distribution of the transformed residuals[15]. We will account for this impact by iterating the θ parameter in equation (3.6) when simulating sample-paths for country-specific asset returns.

Based on our analysis, we formulated for each country our "prior knowledge" of its asset-returns DGP. We will present only few of them as follows:

- *Australian Asset Returns DGP*: this model may be classified by a restricted VAR(1) model. We have observed that there is a positive relation between the contemporary total equity return (e_t) and the lag of the total bond return (b_{t-1}). Moreover, it seems that the total bond yield is inversely affected by the lag of the total equity return with statistical significance. We also find that the error terms may be thought of as white noise after applying the Portmanteau test. Moreover, we believe that the errors emerge from normal distribution, especially after applying Henze & Zirkler (1990) Multivariate Normality test on the residuals of the fitted VAR model. Hence, theoretically, one may assume that the Australian asset-returns follow a strictly covariance-stationary process; hence the MBB and the Hybrid bootstrap may be applied to this dataset.

- *Belgian Asset Returns DGP*: this model may be classified as a log-VARMA(6,1) model[16] with normal errors.

- *Canadian Asset Returns DGP*: this model may be classified as a VAR(2) model with normal errors. Although the residuals have suffered from minor heaviness in the tails, the Henze-Zirkler's Multivariate Normality Test failed to reject the normality

---

[15]For arbitrary functions f and g the $corr(\hat{\varepsilon}_{1t},\hat{\varepsilon}_{2t}) \neq corr(f(\hat{\varepsilon}_{1t}),g(\hat{\varepsilon}_{2t}))$

[16]This model has been classified using the Extended Cross-Correlation Matrices of Tiao & Tsay (1983)

null when applied to the residuals. Hence, the Canadian returns follow a strictly covariance-stationary process.

- *Danish Asset-Returns DGP*: this model may be classified as a log-VARMA(1,1) model with errors emerging from a bi-variate function F, with an associate Copula, C, defined as

$$C(F_{f(\varepsilon_{1t})}(y_1), F_{g(\varepsilon_{2t})}(y_2); \theta) \tag{3.7}$$

where $f(x) = \ln(x+1)$, $g(x) = \ln(x+1) + 1$, and $\theta = 0.34$. The marginal CDF are given as $F_{f(\varepsilon_{1t})} \sim \mathcal{N}(0.0026, 0.089)$ and $F_{g(\varepsilon_{2t})} \sim \mathcal{W}eibull(shape = 7.5, scale = 1.05)$. Thus, the Danish asset returns follow a covariance-stationary process that is not a strictly stationary process.

- *Spanish Asset Returns DGP*: it was recognized that one may mimic the Spanish returns by an unrestricted VAR(2) model with Gaussian stationary error terms[17]. Hence, the Spanish returns may be considered a strictly covariance-stationary series.

- *UK Asset-Returns DGP*: this model may be classified as a restricted log-VAR(1) with errors emerging from a bi-variate function F, with an associate Survival Gumbel Copula, C, defined as

$$C(F_{f(\varepsilon_{1t})}(y_1), F_{g(\varepsilon_{2t})}(y_2); \theta = 1.7) \tag{3.8}$$

where $f(x) = x$, $g(x) = x+1$. The marginal CDF are given as $F_{f(\varepsilon_{1t})} \sim \mathcal{L}ogis(location = -0.003, scale = 0.063)$ and $F_{g(\varepsilon_{2t})} \sim \mathcal{W}eibull(shape = 5.9435, scale = 1.0693)$. Hence, the UK asset returns follow a covariance-stationary process that might not be strictly stationary.

- *US Asset Returns DGP*: this model may be classified as a restricted VAR(1) model with normal errors.

3.8   Evaluating bootstrap methods for long-term parameters

Based on our formulated "prior knowledge" about the characteristics of the DGP of asset returns in different countries (see Section 3.7), we opt to use the Warp-Speed method of

---

[17]We have found that the residuals may follow a normal distribution using the Henze-Zirkler's Multivariate Normality Test

Giacomini et al. (2013) for a bootstrap methods evaluation. This approach generates an artificial sample from an assumed DGP, then accordingly draw a single bootstrap sample[18] from the created sample. For each of the 40-year horizon bi-variate (bonds and equities) samples produced, we compute the 40-year univariate returns for different strategies, as mentioned in equation (3.1). Depending on the 40 data points annual returns of strategy returns, we compute the single 40-year point CFP measure, namely the TAW, IRR, and GMR, as described in equations (3.2, 3.3, and 3.4). Using the same steps we produce, for each long-term investment strategy, 10000 Monte Carlo replications. Hence, by the end of the experiments, we have two empirical distributions for each of the CFP measures. We refer to the 'artificial sample statistics' and 'bootstrap-method statistics', respectively.

As mentioned earlier, our primary motivation is based on checking the reliability and robustness of applying different bootstrap methods to several DGP asset returns. To do this, we will use an algorithm which allows us to compare, for various strategies and different DGP asset returns, the artificial with that of the bootstrap sample of CFP measures. Using Cramér–von Mises two samples test, we compare both samples. If the samples match, we accept that the specified bootstrap method is capable of being applied to estimate long-term parameters, such as Sortino Ratio, or VaR when the DGP is assumed to follow the DGP suggested in our Monte Carlo experiment. Otherwise, we tend to reject the notion that the specified bootstrap method gives a reliable estimate to long-term performance parameters.

Algorithm Procedure 3.8.1.

1. *Perform time-series analysis on countries' historical asset-returns to specify a candidate model of the DGP for each country (see, Section 3.7)*

2. *Simulate artificial samples of asset returns based on assumed DGP*

3. *Produce five pseudo bootstrap samples[19] for each artificial sample*

4. *Compute the CFP measures (TAW, GMR, IRR) for each of the six samples along with the five strategies. Hence, a total of 30 combinations of performance measures will be recorded*

5. *Repeat the above steps 10000 times, hence, producing 10000 combinations of CFP measures per strategy*

---

[18]In our case, we will draw five bootstrap samples for each artificial sample since each bootstrap sample represents a different bootstrap method.

[19]There is more than one bootstrap sample because there is more than one bootstrap method where each sample corresponds to a different bootstrap method for the same Artificial Sample

The figure 3.1 illustrate the above description of our algorithm
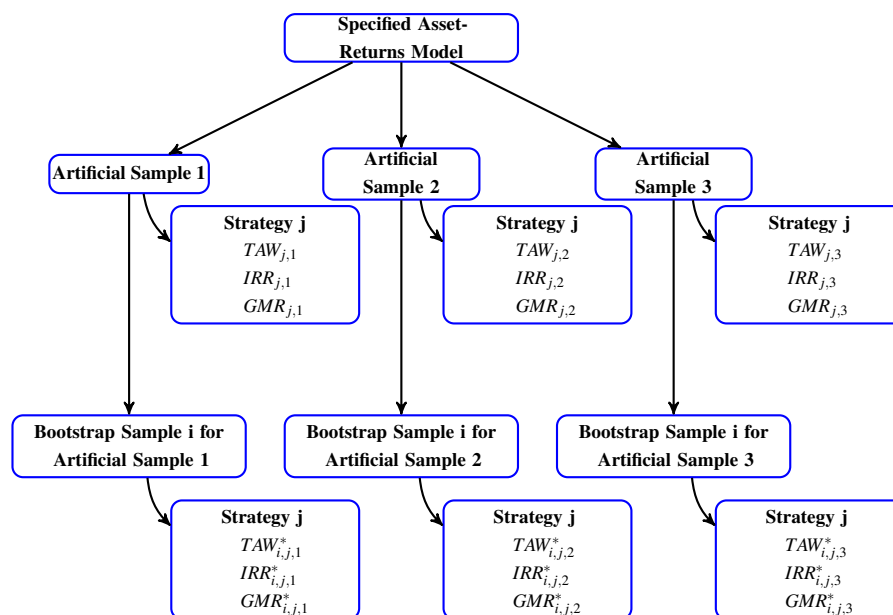
Figure 3.1 Descriptive diagram for modeling cash-flow performance measures

Based on our data analysis in section 3.7, our dataset may be classified to follow one of the following DGP: 1) VAR, log-VAR, or log-VARMA with normal errors models; 2) VAR, log-VAR, or log-VARMA with non-normal errors models. However, in our study, we will also consider additional DGPs, such as 3) Independent Multivariate Normal processes and 4) Copula based-models.

We examined the independent multivariate normal model to observe whether the SB outperformed other mentioned bootstrap methods in the i.i.d. case. The Copula-based multivariate model was chosen, due to its growing popularity in modelling financial time series relative to the traditional correlation-based approach. Modelling dataset with a copula-based model allows a good medium of flexibility for practitioners. In particular, it allows specifying the model's marginal distributions independently from the dependence structure that models the joint distribution of bonds to equity returns. (see, Patton 2012).

Although we considered different parameters for the VAR model in equation (3.5), we found no significant differences by iterating these parameters when comparing different artificial sample statistics to the bootstrap method statistics. Therefore, in our reported Monte Carlo experiments, we will use the estimated parameters of the US asset returns in the observed dataset. For example, using the fitted parameters of the US data, we generate an artificial sample from the following processes:

- **USA Model 1:** A VAR process defined by

$$\begin{bmatrix} e_t \\ b_t \end{bmatrix} = \begin{bmatrix} 0.0761 \\ 0.0257 \end{bmatrix} + \begin{bmatrix} -0.0603 & 0.505 \\ -0.0632 & 0.131 \end{bmatrix} \begin{bmatrix} e_{t-1} \\ b_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{3.9}$$

where the errors are defined as

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} ; \begin{bmatrix} 0.037 & 0.0037 \\ 0.0037 & 0.0094 \end{bmatrix} \right)$$

- **USA Model 2:** A VAR process defined as in equation (3.9) with errors having the same density function as the fitted residuals (kernel density estimator of CDF) with a Gaussian Bi-variate Copula, C with dependence parameter $\theta = 0.21$.

- **USA Model 3:** An Independent Multivariate Normal distribution such that

$$\begin{bmatrix} e_t \\ b_t \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0.08467 \\ 0.0237 \end{bmatrix} ; \begin{bmatrix} 0.0397 & 0 \\ 0 & 0.0095 \end{bmatrix} \right)$$

- **USA Model 4:** A t Copula-based multivariate model[20] with marginal distributions that shares the kernel density estimation of the original observed dataset and dependence parameters ($\rho = 0.2258, \nu = 1.2786 x 10^7$), where ρ and ν represent the correlation parameter and the degrees of freedom parameter.

Figures 3.2 and 3.3 show one simulated sample of each of the above model descriptions for bonds and equities respectively.

---

[20]The Student's t copula was selected because it neither underestimates as the Gaussian Copula nor overestimates as the Gumbel Copula the probability of joint extreme downward movement of portfolios consisting of stocks and bonds (Kole et al. 2007).
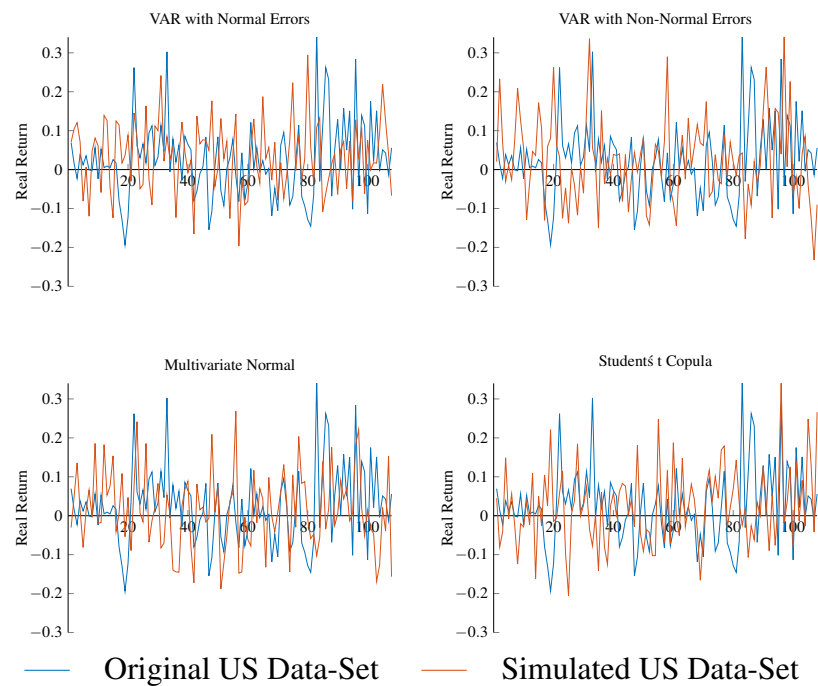
Figure 3.2 The Bond's Real Return of each of the Generating Models for the US Data-Set

As observed graphically, the US equities, as well as bonds, returns may follow any of the suggested DGP models. However, an interesting question is whether any of the earlier suggested bootstrap methods act as a powerful tool in generating distributions of CFP measures for returns that are produced from the above-described models. This is important for ranking different strategies, because the long-term performance parameters are based on the distributions of the cash flows. Therefore, if the bootstrap method is incapable in mimicking the correct distribution of the CFP measure, we might be either underestimating or overestimating a strategy over an alternative; hence, we will have inconsistency in our ranking strategies.
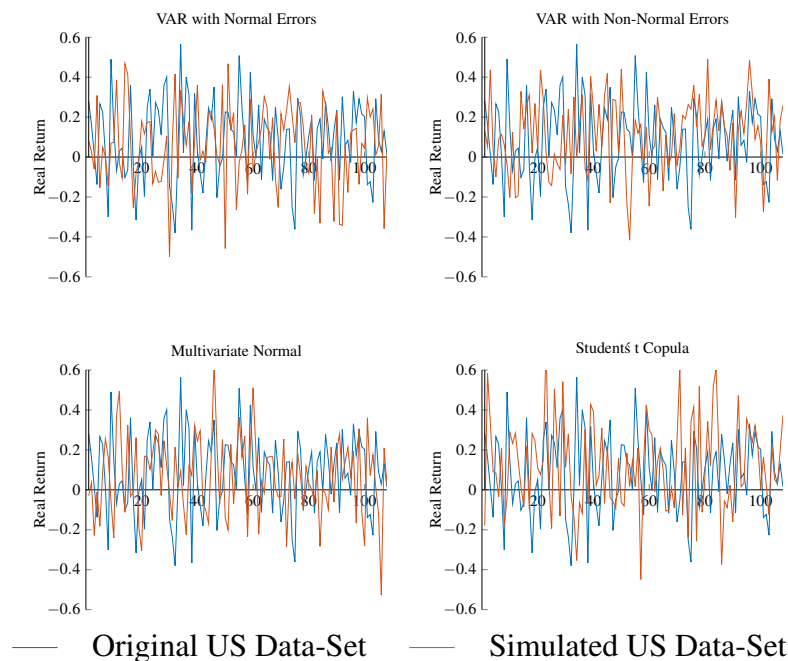
Figure 3.3 The Equity's index Real Return of each of the Generating Models for the US Data-Set

Another issue we need to deal with is whether a strategy alteration would lead to a different favoured bootstrap method. For example, for the same assumed DGP, would it be possible to expect that the SB is better capable of mimicking the TAW distribution than the MBB whereas the MBB outperforms the SB in its capability of mimicking this same distribution for another strategy?

To address these questions, we resorted to Monte Carlo experiments as discussed earlier. For each of the assumed DGP for US data-set, we simulated, for each strategy mentioned in section 3.6, the CFP measures' distribution. Then for each strategy, we compared the empirical distribution of the artificially simulated distribution with the bootstrap distribution via the Two-sample Cramér–von Mises goodness-of-fit hypothesis test (see, Anderson 1962).
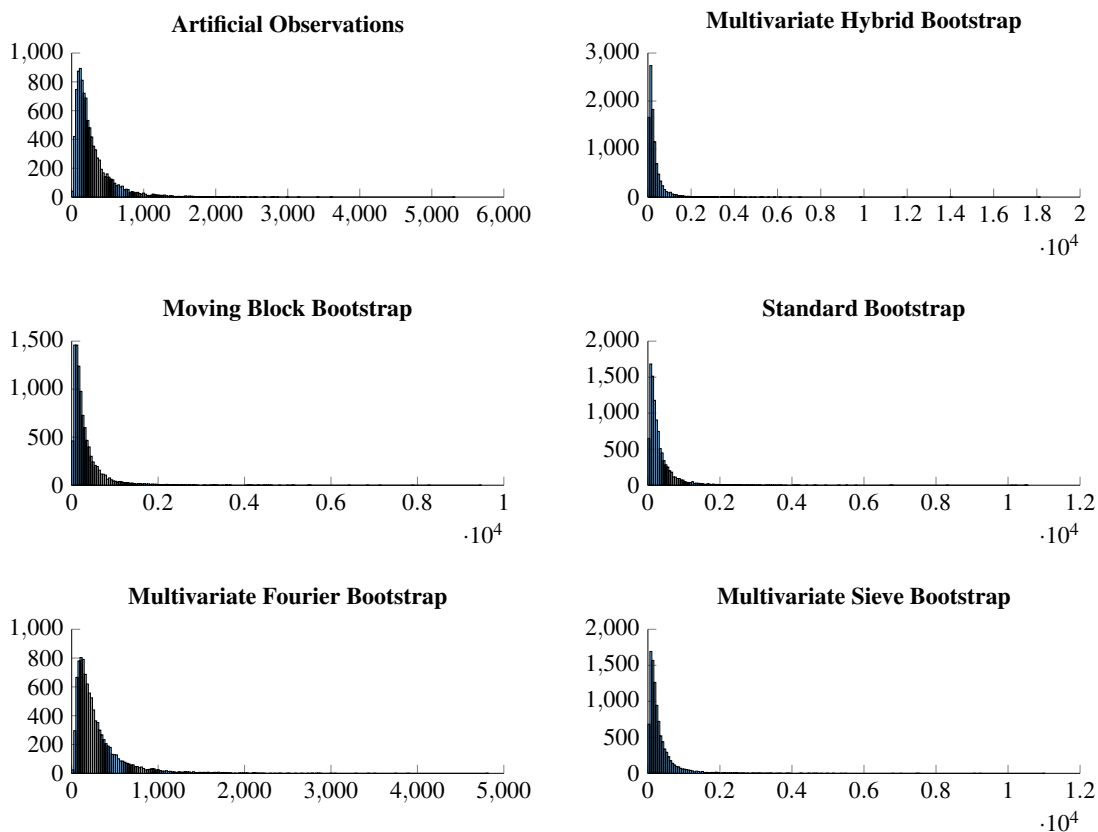
Figure 3.4 The Histograms of the TAW for Strategy One, using the VAR with Normal Error Process - USA model 1

In our simulations, the block length is set at 10 and we randomly choose 4 blocks out of the 100 available blocks. Our choice is motivated by our assumption that a typical participant joins the plan at the age of 25 and retires when (s)he reaches the age of 65. Hence, (s)he has an investment horizon of 40 years. The choice of selecting the value *b* is of great importance but is beyond the scope of this paper. We used the value 10 for comparison reasons only. The selection of block length has been investigated by Cogneau & Zakamouline (2013). For the MHB, we selected a hybrid bandwidth of 0.1; the choice lies within Jentsch & Kreiss's (2010) paper where they used cross-validation and opted to choose the bandwidth depending on the sample size. For the fit of the MSB, we selected the model length based on the AIC criterion. For the MFB, we have utilised the Newey-West bandwidth selection procedure in the HAC estimation.

Figures 3.4, 3.5, and 3.6, represent the histograms for the TAW, GMR, and IRR with different strategies respectively. Figure 3.7 is a zoomed-in view of the TAW distribution. It should be noted that all the histograms have the same number of bins.
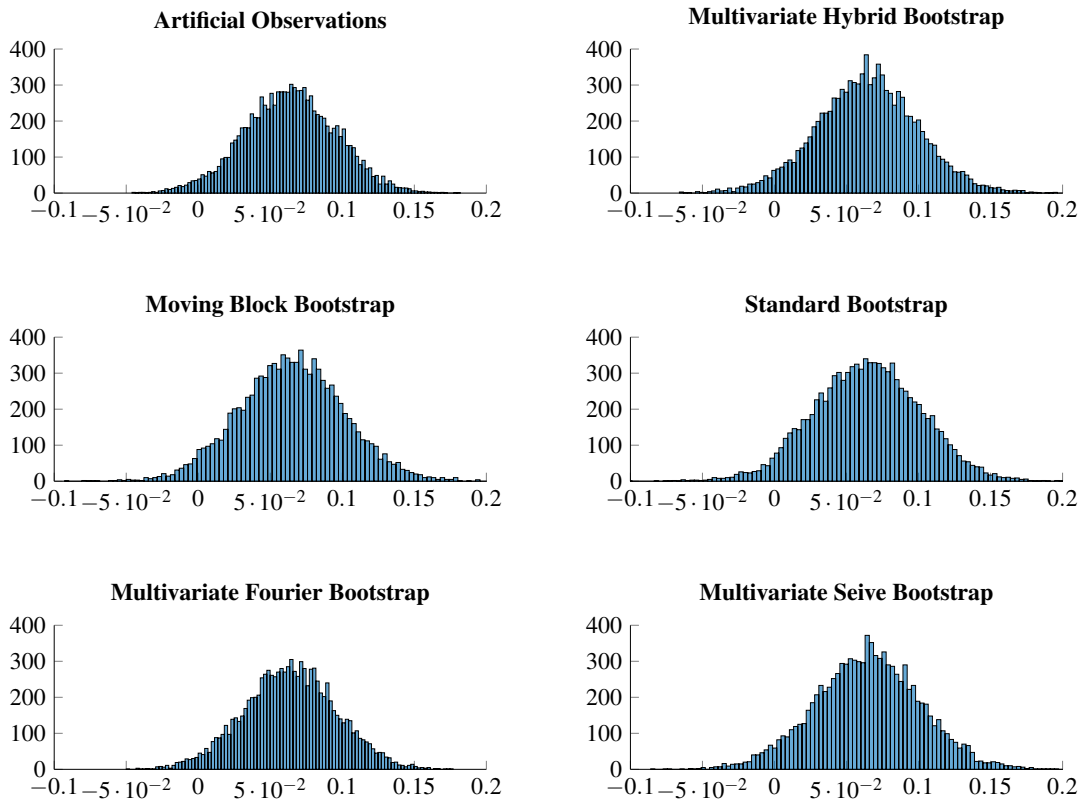
Figure 3.5 The Histograms of the GMR for Strategy One, using the VAR with Normal Error Process - USA model 1

Tables 3.2, 3.3, 3.4, and 3.5 show Cramér von Mises goodness-of-fit statistics of each of the bootstrap methods with the associated assumed US model. The result reveals that all of the assumed DGP (the VAR with normal errors, VAR with non-Normal errors, and Independent Multivariate with Normal errors models as well as Student's t copula with errors generated from the residuals inverse CDF) have favoured the MFB. In fact, it was the only distribution that was not rejected by the Cramér von Mises goodness-of-fit test for different strategies. There were some strategies in these assumed DGP where the MHB was not rejected for the IRR and TAW CFP measures. See, for example, strategies 2 and 3 in table 3.5.

Table 3.2 Two-sample Cramér–von Mises goodness-of-fit statistics for various Bootstraps for US Model 1

USA Model 1					
	MHB	M-MBB	EB	MFB	MSB
Strategy 1					
IRR	0.6937	1.4581	2.9855	0.0656*	2.3078
GMR	1.4209	2.0013	4.1735	0.0730*	3.4749
TAW	0.6937	1.4581	2.9855	0.0656*	2.3078
Strategy 2					
IRR	0.5348	1.3854	1.3951	0.0845*	1.7500
GMR	1.0599	2.6540	2.5229	0.0459*	2.8663
TAW	0.5348	1.3854	1.3951	0.0845*	1.7500
Strategy 3					
IRR	0.5468	1.7654	2.4025	0.0607*	1.8576
GMR	0.9108	1.9520	3.4302	0.0395*	2.5370
TAW	0.5468	1.7654	2.4025	0.0607*	1.8576
Strategy 4					
IRR	0.8030	1.6221	3.4465	0.1031*	2.1262
GMR	1.3591	1.9810	4.1104	0.0706*	3.1615
TAW	0.8030	1.6221	3.4465	0.1031*	2.1262
Strategy 5					
IRR	0.6643	1.6896	3.0927	0.0686*	2.2502
GMR	1.4081	2.1075	4.3852	0.0607*	3.4327
TAW	0.6643	1.6896	3.0927	0.0686*	2.2502

Table 3.3 Two-sample Cramér–von Mises goodness-of-fit statistics for various Bootstraps for US Model 2

US Model 2					
	MHB	M-MBB	EB	MFB	MSB
Strategy 1					
IRR	0.7942	1.2204	4.3788	0.1029*	2.3897
GMR	1.0280	1.7853	4.7543	0.0793*	2.6464
TAW	0.7942	1.2204	4.3788	0.1029*	2.3897
Strategy 2					
IRR	1.2457	1.1238	1.3523	0.1975*	2.4250
GMR	1.5219	1.8726	2.4498	0.2250*	3.1552
TAW	1.2457	1.1238	1.3523	0.1975*	2.4250
Strategy 3					
IRR	0.5119	1.2711	3.5405	0.0822*	1.8489
GMR	0.7737	1.7048	4.3602	0.0569*	2.3179
TAW	0.5119	1.2711	3.5405	0.0822*	1.8489
Strategy 4					
IRR	0.6653	1.2375	3.1061	0.3898*	1.9402
GMR	0.8224	1.6966	3.7642	0.4333*	2.1408
TAW	0.6653	1.2375	3.1061	0.3898*	1.9402
Strategy 5					
IRR	1.2814	1.1098	2.9991	0.2442*	2.2758
GMR	1.4687	1.7935	4.4559	0.1269*	2.9658
TAW	1.2814	1.1098	2.9991	0.2442*	2.2758

Table 3.4 Two-sample Cramér–von Mises goodness-of-fit statistics for various Bootstraps for US Model 3

US Model 3					
	MHB	M-MBB	EB	MFB	MSB
Strategy 1					
IIRR	0.6114	1.5252	2.1921	0.1614*	2.1511
GMR	0.8419	2.0351	3.0992	0.0874*	2.6094
TAW	0.6114	1.5252	2.1921	0.1614*	2.1511
Strategy 2					
IRR	0.8508	1.1233	1.5100	0.2397*	2.4013
GMR	1.4027	1.7612	2.5820	0.1724*	3.3621
TAW	0.8508	1.1233	1.5100	0.2397*	2.4013
Strategy 3					
IRR	0.5586	1.5813	2.2576	0.0716*	2.1421
GMR	0.6629	1.5483	2.3092	0.1675*	2.1096
TAW	0.5586	1.5813	2.2576	0.0716*	2.1421
Strategy 4					
IRR	0.7937	1.4544	2.1626	0.1048*	2.3179
GMR	1.3606	2.1776	3.0236	0.0564*	3.2752
TAW	0.7937	1.4544	2.1626	0.1048*	2.3179
Strategy 5					
IRR	0.8619	1.3601	3.0358	0.1616*	2.8194
GMR	1.3269	1.6077	2.9848	0.0662*	3.3781
TAW	0.8619	1.3601	3.0358	0.1616*	2.8194

Table 3.5 Two-sample Cramér–von Mises goodness-of-fit statistics for various Bootstraps for US Model 4

US Model 4					
	MHB	M-MBB	EB	MFB	MSB
Strategy 1					
IRR	0.7876	1.5204	2.7001	0.0514*	2.7896
GMR	1.4338	2.5696	4.1249	0.0893*	3.6070
TAW	0.7876	1.5204	2.7001	0.0514*	2.7896
Strategy 2					
IRR	0.4173*	1.1884	1.9807	0.0295*	1.6588
GMR	0.8389	1.9504	3.1960	0.0379*	2.3476
TAW	0.4173*	1.1884	1.9807	0.0295*	1.6588
Strategy 3					
IRR	0.3254*	1.2992	1.8426	0.1082*	1.9592
GMR	0.7660	1.7832	2.2319	0.1703*	2.9142
TAW	0.3254*	1.2992	1.8426	0.1082*	1.9592
Strategy 4					
IRR	0.7405	1.4237	1.7486	0.2019*	2.2475
GMR	0.9063	2.1433	2.4066	0.2147*	3.0496
TAW	0.7405	1.4237	1.7486	0.2019*	2.2475
Strategy 5					
IRR	0.7506	1.6261	1.8810	0.2517*	2.5857
GMR	1.0150	2.1123	2.5459	0.2596*	3.3407
TAW	0.7506	1.6261	1.8810	0.2517*	2.5857
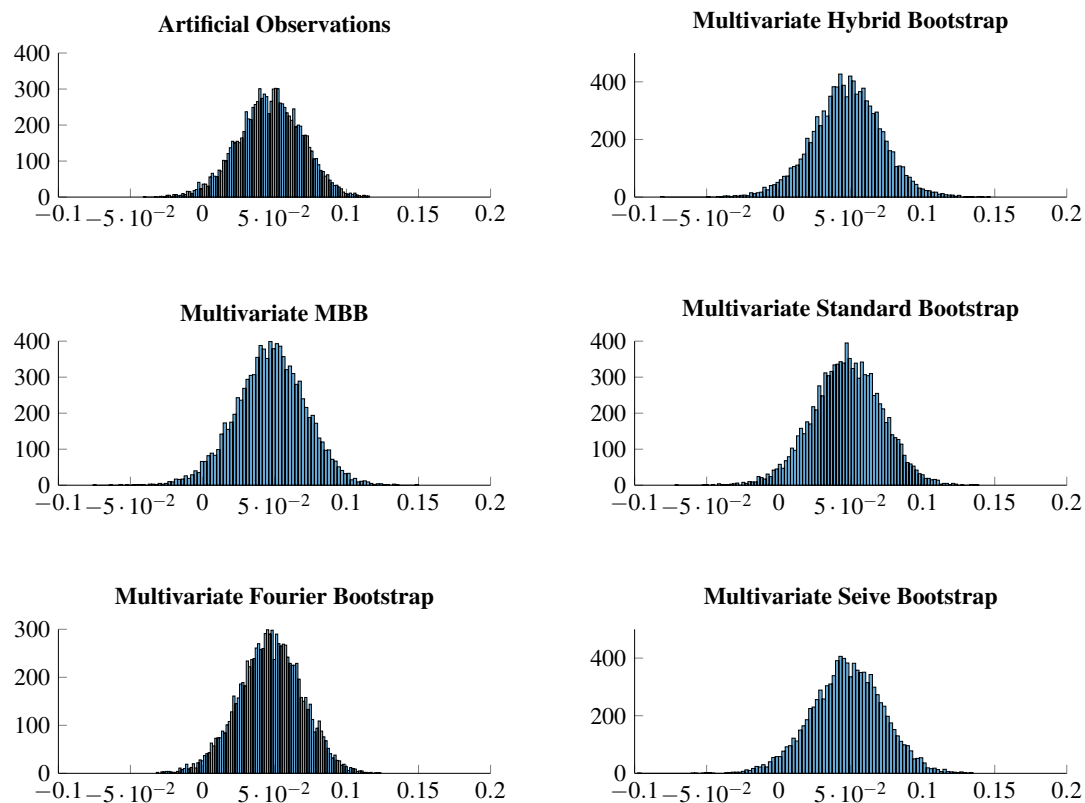
Figure 3.6 The Histograms of the IRR for Strategy Two, using the VAR with Normal Error Process

Furthermore, we simulated artificial VAR models with non-Guassian errors that are more extreme than the Gaussian copula. In particular, we ran simulations for errors emerging from a student's t copula with a UK dataset where the residuals in that country suffered from a more extreme tail than the US residuals distribution. In addition, we also simulated a dataset from a VAR model with error terms being generated from bootstrap residuals. Additionally, we simulated copula-based random variables for Student's t copula without any fitting of Danish asset-returns; however, once again the same conclusion has been reached. Therefore, in all strategies assuming any of the DGP proposed, the MFB was better in modelling the distribution of the CFP measure than the other suggested bootstrap alternatives, provided that before running the bootstrap algorithm, the generated returns are transformed to log of gross returns.

The MFB overcomes other bootstrap methods for such models. In addition, the SB did not mimic the CFP measure's distribution, even for asset returns DGP that originated from an independent multivariate normal distribution, whereas the MFB did.
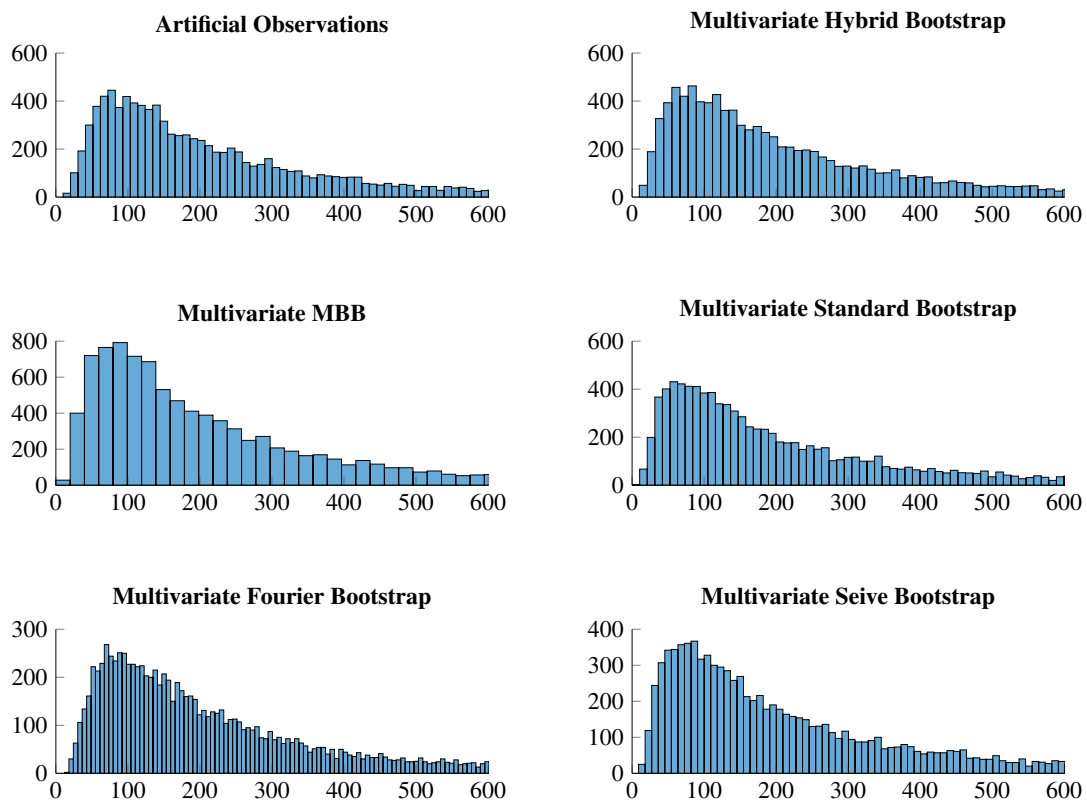
Figure 3.7 The Zoom-in Histograms of the TAW for Strategy four, using the VAR with non-Normal Error Process

Having said this, the next section includes the MFB along with other bootstrap methods utilised in that field, namely the SB as well as the MBB, to rank strategies mentioned in the Pension literature.

3.9   Bootstrap simulations for strategy evaluation

In this section, we consider seventeen different countries and various long-term parameters to rank different strategies. As discussed in previous sections, we perform $10,000$ simulations of CFP measures for the mentioned countries to generate their associated distributions.

To address the arguments, practitioners and researchers had about selecting the best strategy, one needs to acknowledge the fact that they used SB and MBB to generate TAW distribution for each strategy. According to the previous section, when return series are believed to be dependent, such bootstrap methods are incapable of generating the 'right' distribution for long-term return observations. Moreover, they are not superior to their

competitor, the MFB. Therefore, for comparison reasons, we will only consider in this section the SB, the MBB, and the MFB.

As discussed in the previous section, the distribution of the generated future retirement wealth depends on the investment strategy and the bootstrap method utilized. Usually, based on their applied bootstrap algorithm, researchers and practitioners select the strategy that generates TAW distribution with the highest mean and lowest variance (Basu et al. 2011).

In fact, researchers apply various performance and reward-to-risk measures on the simulated TAW observations. For instance, some practitioners have ranked strategies based on VaR (Blake et al. 2001). However, VaR does not address how large the losses can be when extreme events with low probability occur (Bertsimas et al. 2004). Furthermore, as a risk measure, the VaR suffers theoretical property that are necessarily to make it a coherent risk measure, that is, subadditivity and convexity properties (Ibragimov & Prokhorov 2017, p. 26). Therefore, following the recommendation of Tasche (2002), it is important to look at the Conditional VaR (CVaR), also known as the Expected Shortfall (ES)[21] as a better alternative to the VaR. This is due for its theoretical properties: 1) The ES does not require the normality assumption on the data (Bacon 2012); 2) The ES is a subadditive and coherent risk measure (Ibragimov & Prokhorov 2017, p. 114). By definition, the CVaR is "the average of the worst losses of a portfolio, requires existence of first moments of risks to be finite" (Ibragimov & Prokhorov 2017, p. 39); therefore, it ranks strategies based solely on the downside risk level.

Definition 3.9.1. *Let $\alpha \in (0,1)$ be a fixed variable (usually 5%) and Ψ be a real random variable (representing the CFP measure) on a defined probability space such that $\mathbb{E}\big(\max(0,-\Psi)\big) < \infty$. Furthermore, define the VaR of Ψ as*

$$VaR_q(\Psi) := \inf\{z \in \mathbb{R} : \mathbb{P}(\Psi < z) \leq q\}. \tag{3.10}$$

Then, the ES at level α of Ψ is given by[22]:

$$ES_\alpha(\Psi) := \mathbb{E}\big(\Psi \,|\, \Psi \leq VaR_\alpha(\Psi)\big) \tag{3.11}$$

Researchers and practitioners also rank strategies based on reward-to-risk measures. The most common reward-to-risk measure among financial analysts and academic researchers, such as Campbell (2006), is Sharpe ratio, which was introduced by Sharpe (1994) as the

---

[21]Although the CVar and the ES are mathematically defined differently, Corollary 4.3 in (Acerbi & Tasche 2002, p. 1496) show that they are identical.

[22]In general, the probability distribution of the Ψ is unknown, hence the ES can not be computed theoretically as described by equation (3.11). Therefore, the computation of ES will need to based empirically. That is, for each of the n observations of $\{\Psi_i\}_{i=1}^n$, the estimated ES, $\hat{ES}$, is given by $k^{-1}\sum_{i=1}^k \Psi_{(i)}$ where $k := [(1-\alpha)n]$; $\Psi_{(i)}$ the ith smallest observation among the observations $\{\Psi_i\}_{i=1}^n$, and $[]$ is the round function.

reward-to-variability ratio. A similar performance measure that shows the reward to the total risk of the portfolio is the Information Ratio (IR). Assuming the returns are normally distributed, the IR is especially recommended for multi-asset portfolios and for ranking strategies (Kidd 2011). Unlike the Sharpe Ratio, instead of comparing the portfolio to the risk-free asset, it compares the performance of a portfolio to a benchmark (Sharpe 1994). While the Sharpe ratio uses standard deviation, Sortino & Price (1994) argued that only returns that fall below the minimum return to accomplish the goal are associated with risk. Therefore, Sortino & Van Der Meer (1991) introduced the Sortino ratio (SR) which differentiates between positive and negative volatility. This is referred to as downside risk measure. Therefore, the choice of reward-to-risk measure depends on the assumptions behind it and the definition of risk.

We tested for normality on all of the simulated CFP measures (TAW, IRR, GMR). We observed that only very few data-sets of the IRR and GMR are normally distributed. However, we also noticed that this result changes drastically from one country to another and from one bootstrap method to another. Therefore, following Kidd (2011)'s remarks, in our analysis of ranking long-term strategies, we have avoided the usage of Information Ratio (IR). Moreover, since we do not consider the high values of CFP measure to be a risk, we obviate the need to compute the Sharpe ratio. Hence, we will compare reward-to-risk levels of strategies based on the Sortino ratio (SR).

In our simulations, since different strategies represent different investment decisions, we considered the 10000 hypothetical observations of different CFP measures per strategy per bootstrap method, as our returns on these investments that depict that strategy's outcome. Therefore, unlike the conventional application of SR, we will apply it on CFP measures (which are considered equivalent to long-term returns in this case). Moreover, we will consider that the mean of the 50-50 strategy represents the target return in the computation of the SR, as it is the strategy with the lowest mean and variance in all countries and bootstrap methods; hence, it has the spirit of the risk-free return in the conventional SR. Using equation (3.2), (3.3), and (3.4), the SR is defined as follow

$$SR(\Psi_{b,c,j}) = \frac{\mathbb{E}^* \left(\Psi_{b,c,j} \right) - \mathbb{E}^* \left(\Psi_{b,c,2} \right)}{Semi-std^* \left(\Psi_{b,c,j} - \Psi_{b,c,2} \right)},$$

where b represents the bootstrap method SB, MBB, or MFB. c represents the country. j represents the strategy (100% Stock, DLCT, DLCM, LC). $\mathbb{E}^*$ and $Semi-std^*$ are the conditional mean and semi-standard-deviation respectively. Ψ represents the CFP measure of interest (TAW, IRR, or GMR). $\Psi_{b,c,2}$ stands for the CFP measure of the $50-50$ Strategy.

Ranking of strategies varies according to the SR of different CFP measures. In particular, in conventional finance, TAW, unlike the GMR, ranks the investments without factoring out the in-flows in the cash-flows. Therefore, TAW will be used to compare the effect of different bootstrap methods on the same strategy in the same country. As for the GMR, it would be better to examine its SR when comparing different strategies across different countries using the same bootstrap method, because GMR reflects the returns of the strategy in any country of interest. The IRR, on the other hand, is useful to compare different strategies of the same country and same bootstrap method because, as TAW, it takes into account both the inflows and the returns. However, it gives more weight to substantial inflows and returns than moderate ones. Moreover, when comparing strategies, it is important to carefully consider the performance in later years when the investor is close to retirement. In fact, the IRR gives more weight to the performance in later years, as the inflows increase due to the growing wage of the investor.

Note that all rankings represented in the tables below are from highest to least in value. That is, tables that show rankings based on SR, list the strategies with the highest return first while tables for CVaR values list strategies with highest risk first, because they have the highest CVaR value. This allows us to easily notice any association between reward and risk which will be discussed later in this section.

We start by comparing strategies across countries using the CVAR and SR of the GMR, as shown in tables 3.13 and 3.14 respectively. The main thinking behind this is to examine in which country each strategy performs best in terms of return and risk, according to each of the three bootstraps. For example, using the SR of the GMR, we notice that Canada and Australia perform among the top three countries, using all the strategies in all the considered bootstrap methods. This indicates that these countries seem to offer the highest return for all strategies. However, looking at the CVaR, we can also see that these countries have the highets CVaR value. Interestingly, the USA and UK were ranked around the sixth using all CFP measures across the three bootstraps by both the SR and CVaR. In general, though there is no exact consistency across bootstraps and strategies, it can be noted that all countries are ranked around the same level in terms of return and risk level. Countries at the far bottom (top, or middle) are around the same for all strategies and across bootstraps. We can clearly conclude from this that countries at the top offer higher returns on all pension plans relative to countries at the bottom. However, to say whether these countries offer a better environment for pension investments forces us to consider the risk associated with that return, which again confirms our belief that high return comes with high risk.

Furthermore, we are interested to rank strategies in each country using different bootstrap methods and based on different CFP measures. Ranking different strategies based on SR of
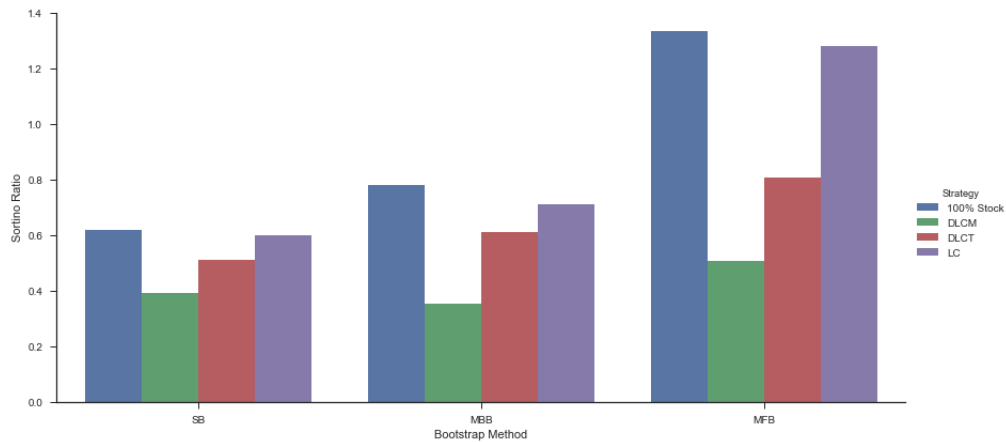
Figure 3.8 UK Sortino Ratio for the IRR using different Bootstrap methods for all strategies.

TAW, IRR, and GMR using the three bootstrap methods (SB, MBB, and MFB) for each of the seventeen countries is represented in tables 3.7, 3.8, and 3.9. We notice that there are differences in ranking across bootstrap methods and across performance measures. Also, while for some countries the ranking is the same across bootstraps and performance measures, the values are not as consistent as we might think.

The differences between the three bootstraps result in the different ranking of strategies and different values of SR of the same performance measure. We observed that SB gave the same ranking across the three measures for all countries, except Canada, France, and Netherlands, using IRR. In particular, SB ranks the strategies as follow: 100% stocks, LC, DLCT, and lastly DLCM. However, it gave different rankings from the MBB and MFB in some countries, such as Canada, using TAW and IRR. This raises concerns around the conclusions arrived by researchers using the SB. Moreover, though the MFB and SB give similar rankings more than other bootstrap pairs across countries and measures, each has a different impact on the values of the SR in the mentioned strategies. For example, in the UK, the values of TAW using SB are the lowest for all strategies compared to other bootstrap methods. Also, comparing the SRs of the IRR in the UK, figure 3.8 shows that the 100% stock strategy overcomes the DLCM and the DLCT using the MFB much more than that observed using the SB and the MBB. Hence, while ranking is important, perhaps studying the values behind the ranking give us a clearer view.

In addition, strategy relative ranking magnitude might vary across the SRs of different performance measures even using the same bootstrap for the same country. For example, in the case of USA (see figures 3.9, 3.10, and 3.11), SB (and MBB) gives same ranking for strategies under the three performance measures, but when we look at the values across the performance measures, it doesn't seem to be as consistent. In particular, under the SB and

MBB, the 100% stocks strategy has much higher SR of TAW than that of DLCM, while the SRs of their IRR (and GMR) are very close. This would lead to confusing conclusions. However, using the MFB, it is notable how the difference in values between strategies across performance measures of the same country is much more consistent. Different bootstrap simulation techniques generates various pseudo returns for the same country's observed data-set. Therefore, the bootstrap method along with the performance measure should be chosen carefully according to the aim of comparison, as explained previously.
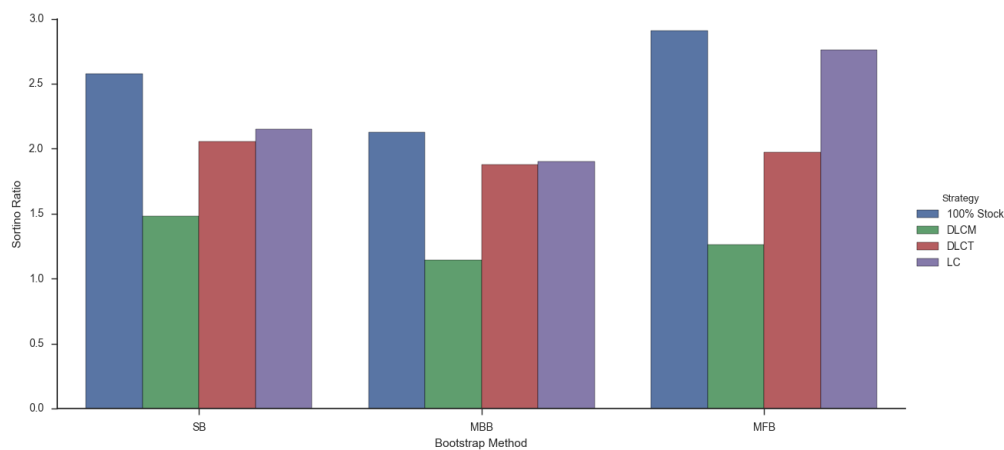
Figure 3.9 USA Sortino Ratio for the TAW using different Bootstrap methods for all strategies.
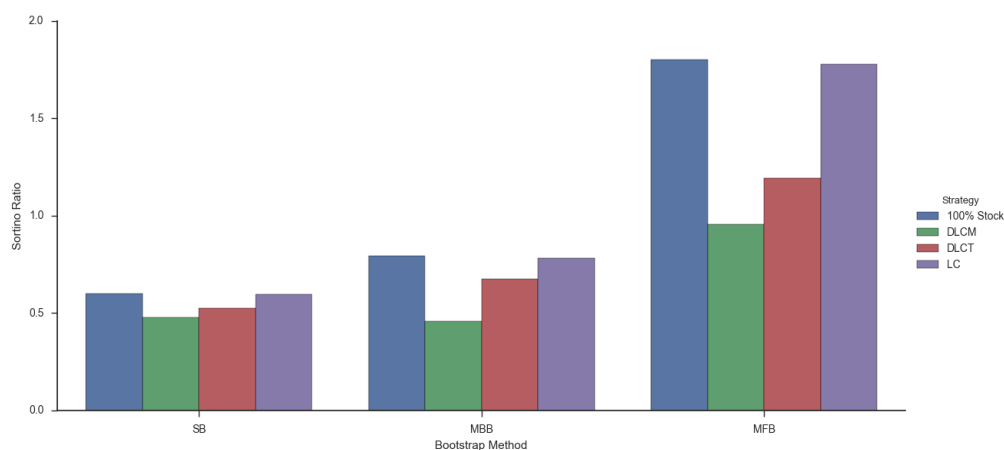
Figure 3.10 USA Sortino Ratio for the IRR using different Bootstrap methods for all strategies.
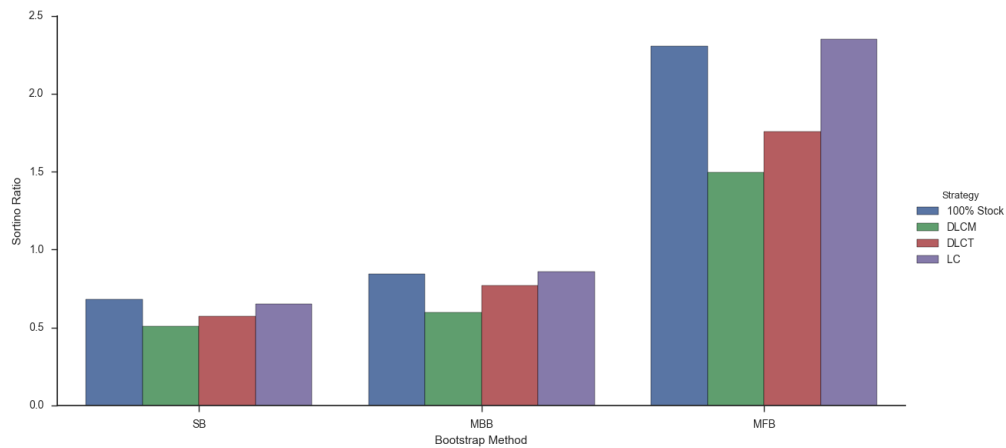
Figure 3.11 USA Sortino Ratio for the GMR using different Bootstrap methods for all strategies.

Using the SR, the choice of both the bootstrap method and performance measure is a critical point that plays a major role in reaching the conclusions. In fact, very few consistencies have been noticed. Only for the UK, Australia, and Denmark the 100% stocks strategy has been ranked first using all three bootstrap methods and all three performance measures. Moreover, all three bootstrap methods give the same ranking using GMR measure only for Australia, Denmark, and the UK. However, using IRR, four countries including Australia, South Africa, UK, and the USA had the same ranking of strategies across bootstraps. Finally, using TAW, five countries - namely Australia, Denmark, Germany, South Africa, and the USA - had the same ranking of strategies across bootstraps. Therefore, to be able to compare the conclusions arrived by different papers, all the tools that were used must be considered.

Furthermore, ranking based on the values of CVaR for each performance measure across bootstrap methods, is based on the risk level of each strategy in each country. On the one hand, when comparing the ranking by SR versus the CVaR, it is observed that it is not the same, which is not surprising. While the Sortino Ratio takes the reward into account and not only the risk, the CVaR ranks the strategy solely based on the risk. For example, the Sortino ratio ranks the 100% stocks first for most of the countries across the bootstraps, and similarly in the cases using CVaR, because there is more dispersion of TAW for the 100% stocks strategy. However, a counter-intuitive example is the case of the USA. Across all bootstrap methods, the 100% stocks strategy has the highest SR of TAW and IRR, while its CVaR of TAW and IRR indicates that is less risky than other strategies, which makes it superior in that case. Ranking strategies based on reward-to-risk differ from ranking them based on safety. In general, we observe that the higher the risk, the higher the reward, despite some deviations.

On the other hand, rankings based on CVAR varied according to the bootstrap method implemented and the performance measure used. For example, Basu et al. (2011) argue that, using bootstrap resampling, the VaR estimate shows that, in the US, the DLCT outperforms the LC in terms of risk. However, it is not the case using CVAR, ceteris paribus. Another interesting example that shows inconsistency across bootstraps is the case of the UK using CVAR of GMR. However, there is noticeable consistency between the CVAR of TAW and IRR for most other countries and bootstrap methods. Only for Canada, Germany, Italy, Netherlands, Switzerland, and the UK, the ranking given by TAW and IRR different only under the SB. Nevertheless, the majority of rankings, according to the CVAR of GMR given by different bootstrap methods, is different from that given by the CVAR of the other two performance measures.

Bootstrap Method	Strategy	Average	Standard Dev.	SR	CVaR
SB	100 % Stocks	316.42	357.18	2.58	39.59
	50-50	148.67	79.79	0	50.44
	DLCM	238.14	220.63	1.48	46.87
	DLCT	280.04	303.39	2.05	41.37
	LC	288.25	274.92	2.15	41.25
MBB	100 % Stocks	270.19	233.92	2.13	53.68
	50-50	150.14	82.46	0	47.78
	DLCM	217.15	173.87	1.14	48.87
	DLCT	262.13	233.41	1.88	50.64
	LC	259.91	202.41	1.9	54.26
MFB	100 % Stocks	227.69	94.39	2.91	94.04
	50-50	141.2	48.41	0	67.96
	DLCM	180.44	63.63	1.26	86.67
	DLCT	208.08	87.24	1.98	83.69
	LC	221.63	81.22	2.76	96.11

Table 3.6 Sortino Ratio and CVaR of the TAW for the US simulations using different bootstrap methods and strategies

3.10   Conclusion

Annual asset returns play a significant role in computing the risk and reward involved in a long-term investment strategy. Moreover, it is understood that different CFP measures grade an investment differently due to the different weighing of the involved inflows, returns, and outflows in a cash-flow diagram. Previously, financial and pension researchers, due to the uncertainty in the nature of the asset return generating process, commonly used conventional bootstrap methods, such as SB and MBB, to produce asset returns to assess reward and risk over a long horizon, using mainly TAW as a CFP measure. However, Cogneau & Zakamouline (2013) demonstrated the potential critical pitfalls that can result in such an improper application of a bootstrap method. The main goal of this paper is to demonstrate the critical importance the bootstrap method and different CFP measures play in determining the optimal strategy.

Specifically, we considered ranking four different strategies (100 % Stocks, DLCM, DLCT, and LC) across several countries using different CFP measures (TAW, IRR, and GMR). First of all, we constructed an accumulation model which reflects pension savings for an ordinary pensioner. Then, for each country, we performed statistical tests on the compiled annual asset returns to explain the key features of the DGP. We found more than one possible DGP that may describe the actual DGP for each country.

We opted to show our work only for the US model for comparison reasons mainly. Then, based on each possible DGP model for the US, we simulated the distribution of the CFP measure for the assumed true DGP model (artificial series) and the several bootstrap pseudo-series. We inferred, through our simulation experiments, that the MFB was the only bootstrap procedure that was able to mimic the distribution of all CFP measures and was model insensitive. Moreover, we also understood that previously implemented bootstrap procedures were incapable of mimicking the distribution of the CFP measures even for the simplest proposed model. In particular, the SB method was incapable of mimicking the distribution of the CFP measures, even when the true distribution originates from an independent DGP.

Building on these conclusions, we report the ranking of various countries across previous bootstrap methods (SB and MBB), along with the new, frequency-based bootstrap method (the MFB). Comparing our results with different published work's procedures prove that the bootstrap method plays a critical role in determining the optimal strategy. Moreover, we observe that different CFP measures rank strategies differently across countries and bootstrap methods. Consequently, this shows that the CFP measure plays a vital role in comparing different strategies based on a relative comparison. For example, if a practitioner would like

to compare a strategy across various countries, it is better to compare the Sortino ratio based on the GMR CFP measure, rather than the IRR or TAW.

In addition, we found that ranking different strategies using different long term parameters is crucial. While the Sortino Ratio shows the best strategy in terms of reward-to-risk, it is worth looking at the CVaR. This is because the CVaR measures the expected losses of a strategy, solely. The best strategy is the one with the highest Sortino Ratio and the lowest CVaR. However, we find that this is very rare and, when found, is not consistent across bootstrap methods and/or CFP measures. Therefore, different participants need to choose their strategy based on their risk-tolerance.

It would be of interest for future research to study whether an accumulation strategy is better than another when further data or data structure is considered. Indeed, it would be interesting to study if the asset returns are being generated from a GARCH or any other model that depicts volatility clustering. It is a standard stylized fact that high-frequency data-set from financial markets exhibits volatility clustering phenomenon. Therefore, it would be interesting to repeat this study using plausible bootstrap procedures with daily data-set returns over the last 30 years to suggest whether a strategy outperforms another strategy. Furthermore, a recent paper suggests the idea of utilizing both low and high-frequency data-set based on their availability (Chambers 2016). Therefore, it would be of future interest to seek if it is possible to take advantage of its finding to devise a bootstrap method that is capable of generating samples that mimic the structure of the original data-generating process which will help in deciding which strategy outperform another using both low frequency and high-frequency data-sets. This will give a more general and robust view on which accumulation strategy outperforms another.

We conclude that bootstrap methods and different CFP measures are of great importance in assessing long-term investment strategies and merit extensive research. One should be very cautious, however, which CFP measure to use. In particular, the CFP measure needs to be unique and a good indicator to use for comparison. For example, if our cash flow had more than one change of sign, the IRR might not be a good measure to use when comparing different strategies across different bootstraps within a country. Moreover, it might be the case, when a different cash-flow is suggested, to repeat our initial steps and compare whether the MFB is able to mimic all the possible DGP before implementing it on the realised dataset.

3.11   Tables

Table 3.7 Ranking Based on Sortino Ratio of TAW

Australia			Belgium			Canada		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	100% Stocks	DLCT	100% Stocks	DLCT	DLCT	100% Stocks
LC	LC	LC	LC	100% Stocks	LC	100% Stocks	100% Stocks	LC
DLCT	DLCT	DLCT	DLCT	LC	DLCT	LC	LC	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM
Denmark			France			Germany		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks
LC	LC	LC	LC	LC	LC	LC	LC	LC
DLCT	DLCT	DLCT	DLCT	DLCT	DLCT	DLCT	DLCM	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCT	DLCM
Ireland			Italy			Japan		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	DLCT	100% Stocks	100% Stocks	100% Stocks	LC	100% Stocks
DLCT	LC	LC	100% Stocks	LC	LC	DLCT	DLCM	LC
LC	DLCT	DLCT	LC	DLCT	DLCT	LC	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	100% Stocks	DLCM
Netherlands			Norway			South Africa		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	DLCT	100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks
DLCT	DLCT	LC	100% Stocks	LC	DLCT	LC	LC	LC
LC	LC	DLCT	LC	DLCT	LC	DLCT	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM
Spain			Sweden			Switzerland		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCT	DLCT	100% Stocks	100% Stocks	100% Stocks	100% Stocks	DLCT	DLCT	100% Stocks
100% Stocks	100% Stocks	LC	DLCT	DLCT	LC	100% Stocks	100% Stocks	LC
LC	LC	DLCT	LC	LC	DLCT	LC	LC	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM
UK			USA			World		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks	DLCT	100% Stocks	100% Stocks
DLCT	LC	LC	LC	LC	LC	100% Stocks	LC	LC
LC	DLCT	DLCT	DLCT	DLCT	DLCT	LC	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM

Table 3.8 Ranking Based on Sortino Ratio of IRR

Australia			Belgium			Canada		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	100% Stocks	DLCT	100% Stocks	DLCT	LC	LC
LC	LC	LC	LC	LC	LC	100% Stocks	DLCT	100% Stocks
DLCT	DLCT	DLCT	DLCT	100% Stocks	DLCT	LC	100% Stocks	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM
Denmark			France			Germany		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks	LC	LC	100% Stocks	100% Stocks
DLCT	LC	LC	LC	LC	100% Stocks	100% Stocks	LC	LC
LC	DLCT	DLCT	DLCT	DLCT	DLCT	DLCM	DLCM	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCT	DLCT	DLCM
Ireland			Italy			Japan		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
LC	100% Stocks	100% Stocks	DLCT	100% Stocks	100% Stocks	LC	LC	100% Stocks
100% Stocks	LC	LC	100% Stocks	LC	LC	100% Stocks	100% Stocks	LC
DLCT	DLCT	DLCT	LC	DLCT	DLCM	DLCT	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCT	DLCM	DLCM	DLCM
Netherlands			Norway			South Africa		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCT	DLCT	100% Stocks	DLCT	100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks
100% Stocks	100% Stocks	LC	LC	LC	DLCT	LC	LC	LC
LC	LC	DLCT	100% Stocks	DLCT	LC	DLCT	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM
Spain			Sweden			Switzerland		
DLCT	DLCT	100% Stocks	DLCT	100% Stocks	100% Stocks	DLCT	DLCT	100% Stocks
100% Stocks	100% Stocks	LC	100% Stocks	DLCT	LC	100% Stocks	LC	LC
LC	LC	DLCT	LC	LC	DLCT	LC	100% Stocks	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM
UK			USA			World		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks	100% Stocks	DLCT	100% Stocks	100% Stocks
LC	LC	LC	LC	LC	LC	100% Stocks	LC	LC
DLCT	DLCT	DLCT	DLCT	DLCT	DLCT	LC	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM

Table 3.9 Ranking Based on Sortino Ratio of GMR

Australia			Belgium			Canada		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	LC	DLCT	100% Stocks	DLCT	DLCT	100% Stocks
LC	LC	LC	100% Stocks	100% Stocks	LC	100% Stocks	LC	LC
DLCT	DLCT	DLCT	DLCT	LC	DLCT	LC	100% Stocks	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM

Denmark			France			Germany		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	100% Stocks	LC	100% Stocks	LC	100% Stocks	100% Stocks
LC	LC	LC	LC	100% Stocks	LC	100% Stocks	LC	LC
DLCT	DLCT	DLCT	DLCT	DLCT	DLCT	DLCM	DLCM	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCT	DLCT	DLCM

Ireland			Italy			Japan		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
LC	LC	100% Stocks	DLCT	100% Stocks	100% Stocks	100% Stocks	LC	100% Stocks
100% Stocks	100% Stocks	LC	100% Stocks	LC	LC	LC	100% Stocks	LC
DLCT	DLCT	DLCT	LC	DLCT	DLCT	DLCT	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM

Netherlands			Norway			South Africa		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	DLCT	100% Stocks	DLCT	100% Stocks	100% Stocks	100% Stocks	LC	100% Stocks
DLCT	100% Stocks	LC	LC	LC	LC	LC	100% Stocks	LC
LC	LC	DLCT	100% Stocks	DLCT	DLCT	DLCT	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM

Spain			Sweden			Switzerland		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCT	DLCT	100% Stocks	DLCT	100% Stocks	100% Stocks	DLCT	DLCT	100% Stocks
LC	LC	LC	100% Stocks	DLCT	LC	100% Stocks	LC	LC
100% Stocks	100% Stocks	DLCT	LC	LC	DLCT	LC	100% Stocks	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM

UK			USA			World		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	LC	LC	100% Stocks	DLCT	100% Stocks	100% Stocks
LC	LC	LC	100% Stocks	100% Stocks	LC	100% Stocks	LC	LC
DLCT	DLCT	DLCT	DLCT	DLCT	DLCT	LC	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM	DLCM

Table 3.10 Ranking Based on CVaR 95 of TAW

| Australia | | | Belgium | | | Canada | | |
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
LC	100 % Stocks	100 % Stocks	LC	LC	DLCM	LC	LC	LC
100 % Stocks	LC	LC	DLCT	DLCT	DLCT	100 % Stocks	DLCT	DLCM
DLCT	DLCT	DLCT	100 % Stocks	100 % Stocks	100 % Stocks	DLCT	100 % Stocks	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	LC	DLCM	DLCM	100 % Stocks

| Denmark | | | France | | | Germany | | |
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCT	LC	DLCM	LC	100 % Stocks	DLCM	LC	100 % Stocks	100 % Stocks
100 % Stocks	100 % Stocks	DLCT	100 % Stocks	LC	LC	100 % Stocks	LC	DLCM
LC	DLCM	LC	DLCM	DLCM	DLCT	DLCM	DLCM	LC
DLCM	DLCT	100 % Stocks	DLCT	DLCT	100 % Stocks	DLCT	DLCT	DLCT

| Ireland | | | Italy | | | Japan | | |
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
LC	LC	DLCM	LC	100 % Stocks	DLCM	LC	LC	DLCM
100 % Stocks	100 % Stocks	100 % Stocks	100 % Stocks	LC	100 % Stocks	100 % Stocks	100 % Stocks	LC
DLCT	DLCM	DLCT	DLCT	DLCM	DLCT	DLCM	DLCT	100 % Stocks
DLCM	DLCT	LC	DLCM	DLCT	LC	DLCT	DLCM	DLCT

| Netherlands | | | Norway | | | South Africa | | |
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCM	DLCM	DLCM	DLCT	DLCM	DLCM	100 % Stocks	LC	100 % Stocks
DLCT	DLCT	DLCT	DLCM	100 % Stocks	DLCT	LC	100 % Stocks	LC
LC	100 % Stocks	100 % Stocks	LC	LC	100 % Stocks	DLCT	DLCT	DLCM
100 % Stocks	LC	LC	100 % Stocks	DLCT	LC	DLCM	DLCM	DLCT

| Spain | | | Sweden | | | Switzerland | | |
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCM	DLCT	DLCM	DLCT	LC	DLCM	DLCT	DLCT	DLCM
DLCT	DLCM	DLCT	100 % Stocks	DLCT	100 % Stocks	DLCM	DLCM	DLCT
100 % Stocks	100 % Stocks	LC	LC	100 % Stocks	LC	LC	LC	100 % Stocks
LC	LC	100 % Stocks	DLCM	DLCM	DLCT	100 % Stocks	100 % Stocks	LC

| UK | | | USA | | | World | | |
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100 % Stocks	LC	LC	LC	LC	DLCM	100 % Stocks	LC	DLCM
LC	100 % Stocks	DLCT	100 % Stocks	100 % Stocks	DLCT	LC	100 % Stocks	LC
DLCT	DLCT	DLCM	DLCT	DLCM	LC	DLCT	DLCT	100 % Stocks
DLCM	DLCM	100 % Stocks	DLCM	DLCT	100 % Stocks	DLCM	DLCM	DLCT

Table 3.11 Ranking Based on CVaR 95 of IRR

Australia			Belgium			Canada		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
LC	100% Stocks	100% Stocks	LC	LC	DLCM	LC	LC	DLCM
100% Stocks	LC	LC	100% Stocks	DLCT	DLCT	100% Stocks	DLCT	LC
DLCT	DLCT	DLCT	DLCT	100% Stocks	100% Stocks	DLCT	100% Stocks	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	LC	DLCM	DLCM	100% Stocks

Denmark			France			Germany		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCT	LC	DLCM	LC	100% Stocks	DLCM	LC	100% Stocks	DLCM
100% Stocks	100% Stocks	DLCT	100% Stocks	LC	LC	100% Stocks	LC	100% Stocks
LC	DLCM	LC	DLCM	DLCM	DLCT	DLCM	DLCM	LC
DLCM	DLCT	100% Stocks	DLCT	DLCT	100% Stocks	DLCT	DLCT	DLCT

Ireland			Italy			Japan		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
LC	LC	DLCM	LC	100% Stocks	DLCM	LC	LC	DLCM
100% Stocks	100% Stocks	100% Stocks	100% Stocks	LC	DLCT	100% Stocks	100% Stocks	LC
DLCT	DLCM	DLCT	DLCT	DLCM	100% Stocks	DLCM	DLCT	100% Stocks
DLCM	DLCT	LC	DLCM	DLCT	LC	DLCT	DLCM	DLCT

Netherlands			Norway			South Africa		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCM	DLCM	DLCM	DLCT	DLCM	DLCM	100% Stocks	LC	100% Stocks
DLCT	DLCT	DLCT	DLCM	100% Stocks	DLCT	LC	100% Stocks	LC
LC	100% Stocks	LC	LC	LC	100% Stocks	DLCT	DLCT	DLCM
100% Stocks	LC	100% Stocks	100% Stocks	DLCT	LC	DLCM	DLCM	DLCT

Spain			Sweden			Switzerland		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCM	DLCT	DLCM	DLCT	LC	DLCM	DLCT	DLCT	DLCM
DLCT	DLCM	DLCT	100% Stocks	DLCT	100% Stocks	DLCM	DLCM	DLCT
100% Stocks	100% Stocks	LC	LC	100% Stocks	LC	LC	LC	LC
LC	LC	100% Stocks	DLCM	DLCM	DLCT	100% Stocks	100% Stocks	100% Stocks

UK			USA			World		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	LC	LC	LC	LC	DLCM	100% Stocks	LC	DLCM
LC	100% Stocks	DLCM	100% Stocks	100% Stocks	DLCT	LC	100% Stocks	LC
DLCT	DLCT	DLCT	DLCT	DLCM	LC	DLCT	DLCT	100% Stocks
DLCM	DLCM	100% Stocks	DLCM	DLCT	100% Stocks	DLCM	DLCM	DLCT

Table 3.12 Ranking Based on CVaR 95 of GMR

Australia			Belgium			Canada		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	100% Stocks	LC	DLCT	DLCM	LC	DLCT	LC
LC	LC	LC	DLCT	100% Stocks	LC	100% Stocks	LC	100% Stocks
DLCT	DLCT	DLCT	100% Stocks	LC	100% Stocks	DLCT	100% Stocks	DLCM
DLCM	DLCM	DLCM	DLCM	DLCM	DLCT	DLCM	DLCM	DLCT
Denmark			France			Germany		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	100% Stocks	DLCM	LC	LC	DLCM	LC	100% Stocks	LC
LC	LC	DLCT	100% Stocks	100% Stocks	LC	100% Stocks	LC	100% Stocks
DLCT	DLCM	100% Stocks	DLCM	DLCM	100% Stocks	DLCM	DLCM	DLCM
DLCM	DLCT	LC	DLCT	DLCT	DLCT	DLCT	DLCT	DLCT
Ireland			Italy			Japan		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
LC	LC	100% Stocks	LC	100% Stocks	LC	100% Stocks	100% Stocks	DLCM
100% Stocks	100% Stocks	DLCM	100% Stocks	LC	DLCT	LC	LC	LC
DLCT	DLCM	DLCT	DLCT	DLCM	100% Stocks	DLCM	DLCT	100% Stocks
DLCM	DLCT	LC	DLCM	DLCT	DLCM	DLCT	DLCM	DLCT
Netherlands			Norway			South Africa		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCT	DLCT	DLCM	DLCT	DLCM	DLCM	100% Stocks	100% Stocks	LC
LC	DLCM	LC	LC	LC	100% Stocks	LC	LC	100% Stocks
100% Stocks	LC	100% Stocks	DLCM	100% Stocks	DLCT	DLCT	DLCT	DLCT
DLCM	100% Stocks	DLCT	100% Stocks	DLCT	LC	DLCM	DLCM	DLCM
Spain			Sweden			Switzerland		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
DLCT	DLCT	DLCM	DLCT	100% Stocks	100% Stocks	DLCT	DLCT	DLCM
DLCM	LC	DLCT	LC	DLCT	LC	DLCM	LC	DLCT
100% Stocks	100% Stocks	100% Stocks	100% Stocks	LC	DLCM	LC	100% Stocks	100% Stocks
LC	DLCM	LC	DLCM	DLCM	DLCT	100% Stocks	DLCM	LC
UK			USA			World		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
100% Stocks	LC	LC	LC	LC	DLCM	100% Stocks	LC	LC
LC	100% Stocks	DLCT	100% Stocks	100% Stocks	LC	DLCT	100% Stocks	100% Stocks
DLCT	DLCT	100% Stocks	DLCT	DLCT	100% Stocks	LC	DLCT	DLCT
DLCM	DLCM	DLCM	DLCM	DLCM	DLCT	DLCM	DLCM	DLCM

Table 3.13 Ranking Countries Based on CVaR 95 of GMR

100% Stocks			DLCM			DLCT			LC		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
Canada	South Africa	Australia	Canada	South Africa	Australia	Canada	South Africa	Australia	Canada	South Africa	Australia
Australia	Australia	Canada	South Africa	Canada	Canada	Sweden	Canada	Canada	Australia	Canada	Canada
South Africa	Canada	Sweden	Sweden	USA	Sweden	Australia	Australia	Sweden	South Africa	Australia	Sweden
Sweden	USA	South Africa	Australia	Australia	South Africa	South Africa	USA	South Africa	Sweden	USA	South Africa
USA	Sweden	World	USA	Sweden	World	Denmark	Sweden	World	USA	Sweden	World
Denmark	World	Denmark	Denmark	World	Denmark	USA	World	Denmark	Denmark	World	Denmark
World	Denmark	USA	World	Denmark	USA	World	Denmark	USA	World	Denmark	USA
UK	UK	Netherlands	UK	UK	Netherlands	UK	UK	Netherlands	UK	UK	Netherlands
Netherlands	Switzerland	UK	Netherlands	Netherlands	UK	Netherlands	Switzerland	UK	Netherlands	Switzerland	UK
Ireland	Netherlands	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Netherlands	Switzerland	Ireland	Netherlands	Switzerland
Switzerland	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland	Switzerland	Ireland	Ireland
Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway
Belgium	Spain	Spain	Belgium	Spain	Spain	Belgium	Spain	Spain	Belgium	Spain	Spain
France	France	France	Spain	France	France	Spain	France	France	France	France	France
Spain	Italy	Belgium	France	Italy	Belgium	France	Italy	Belgium	Spain	Italy	Belgium
Italy	Japan	Italy	Italy	Belgium	Italy	Italy	Belgium	Italy	Italy	Japan	Italy
Germany	Belgium	Japan	Germany	Japan	Japan	Germany	Japan	Japan	Germany	Belgium	Japan
Japan	Germany	Germany	Japan	Germany	Germany	Japan	Germany	Germany	Japan	Germany	Germany

Table 3.14 Ranking Countries Based on Sortino Ratio of GMR

100% Stocks			DLCM			DLCT			LC		
MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB	MBB	MFB	SB
Australia	South Africa	Australia	Australia	Canada	Australia	Canada	Canada	Australia	Australia	South Africa	Australia
South Africa	Australia	South Africa	South Africa	South Africa	South Africa	Australia	South Africa	South Africa	South Africa	Australia	South Africa
Canada	Canada	Canada	Canada	Australia	Canada	South Africa	Australia	Canada	Canada	Canada	Canada
World	USA	World	Sweden	USA	World	Sweden	USA	World	World	USA	World
Sweden	World	Sweden	World	UK	Sweden	World	World	Sweden	Sweden	World	Sweden
UK	UK	UK	USA	World	USA	UK	UK	UK	USA	UK	UK
USA	Sweden	USA	UK	Sweden	UK	USA	Sweden	USA	UK	Sweden	USA
Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Ireland	Netherlands	Netherlands
Ireland	Italy	Ireland	Ireland	Ireland	France	Ireland	Ireland	Ireland	Netherlands	Italy	France
France	Ireland	France	France	France	Ireland	Italy	Denmark	France	France	Ireland	Ireland
Germany	France	Japan	Germany	Italy	Japan	Denmark	Switzerland	Denmark	Germany	France	Japan
Denmark	Japan	Denmark	Denmark	Germany	Denmark	France	Spain	Japan	Denmark	Japan	Denmark
Italy	Germany	Belgium	Italy	Denmark	Italy	Switzerland	France	Belgium	Italy	Germany	Belgium
Belgium	Denmark	Italy	Belgium	Japan	Belgium	Spain	Italy	Germany	Belgium	Denmark	Italy
Japan	Norway	Germany	Japan	Norway	Germany	Belgium	Belgium	Italy	Japan	Switzerland	Germany
Switzerland	Switzerland	Spain	Switzerland	Switzerland	Norway	Germany	Japan	Norway	Switzerland	Spain	Spain
Norway	Spain	Switzerland	Spain	Belgium	Spain	Norway	Norway	Spain	Norway	Norway	Switzerland
Spain	Belgium	Norway	Norway	Spain	Switzerland	Japan	Germany	Switzerland	Spain	Belgium	Norway

Bibliography

Acerbi, C. & Tasche, D. (2002), 'On the coherence of expected shortfall', *Journal of Banking & Finance* **26**(7), 1487 – 1503.

Ahsanullah, M., Nevzorov, V. B. & Shakil, M. (2013), *An introduction to order statistics*, Springer.

Akgiray, V. & Lamoureux, C. G. (1989), 'Estimation of stable-law parameters: A comparative study', *Journal of Business & Economic Statistics* **7**(1), 85–93.

Anderson, T. W. (1962), 'On the distribution of the two-sample cramer-von mises criterion', *The Annals of Mathematical Statistics* pp. 1148–1159.

Andrews, D. W. (2002), 'Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators', *Econometrica* **70**(1), 119–162.

Andrews, D. W. & Monahan, J. C. (1992), 'An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator', *Econometrica: Journal of the Econometric Society* pp. 953–966.

Arcones, M. A. & Giné, E. (1989), 'The bootstrap of the mean with arbitrary bootstrap sample size', *Annales de l'IHP Probabilités et statistiques* **25**(4), 457–481.

Athreya, K. (1987), 'Bootstrap of the mean in the infinite variance case', *Annals of Statistics* **15**, 724–731.

Athreya, K. B. & Lahiri, S. N. (2006), *Measure Theory and Probability Theory*, Springer Science & Business Media.

Bacon, C. R. (2012), *Practical risk-adjusted performance measurement*, John Wiley & Sons.

Baltagi, B. (2011), *Econometrics*, 5 edn, Springer Science & Business Media.

Basawa, I., Mallik, A., McCornick, W., Reeves, J. & Taylor, R. (1991*a*), 'Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes', *Communications in Statistics-Theory and Methods* **20**(3), 1015–1026.

Basawa, I., Mallik, A., McCornick, W., Reeves, J. & Taylor, R. (1991*b*), 'Bootstrapping unstable first-order autoregressive processes', *The Annals of Statistics* **19**(2), 1098–1101.
  URL: *http://www.jstor.org/stable/2242105*

Basu, A. K., Byrne, A. & Drew, M. E. (2011), 'Dynamic lifecycle strategies for target date retirement funds', *The Journal of Portfolio Management* **37**(2), 83–96.

Beach, S. L. (2007), 'Semivariance in asset allocations: Longer horizons can handle riskier holdings', *Journal of Financial Planning* **20**(1), 60–69.

Beran, R. (1988), 'Prepivoting test statistics: A bootstrap view of asymptotic refinements', *Journal of the American Statistical Association* **83**(403), 687–697.
URL: *http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1988.10478649*

Berkes, I., Horváth, L. & Schauer, J. (2011), 'Asymptotics of trimmed cusum statistics', *Bernoulli* **17**(4), 1344–1367.

Berkes, I., Horvath, L. & Schauer, J. (2012), 'Asymptotic behavior of trimmed sums', *Stochastics and Dynamics* **12**(01), 1150002.

Bertail, P., Politis, D. N. & Romano, J. P. (1999), 'On subsampling estimators with unknown rate of convergence', *Journal of the American Statistical Association* **94**(446), 569–579.

Bertsimas, D., Lauprete, G. J. & Samarov, A. (2004), 'Shortfall as a risk measure: properties, optimization and applications', *Journal of Economic Dynamics and control* **28**(7), 1353–1381.

Bickel, P. J. & Ren, J.-J. (2001), 'The bootstrap in hypothesis testing', *Lecture Notes-Monograph Series* pp. 91–112.

Bickel, P. J. & Sakov, A. (2008), 'On the choice of *m* in the *m* out of *n* bootstrap and confidence bounds for extrema', *Statistica Sinica* **18**(3), 967–985.

Blake, D., Cairns, A. J. & Dowd, K. (2001), 'Pensionmetrics: stochastic pension plan design and value-at-risk during the accumulation phase', *Insurance: Mathematics and Economics* **29**(2), 187–215.

Blattberg, R. & Sargent, T. (1971), 'Regression with non-gaussian stable disturbances: Some sampling results', *Econometrica: Journal of the Econometric Society* pp. 501–510.

Blum, M. (1970), 'On the sums of independently distributed pareto variates', *SIAM Journal on Applied Mathematics* **19**(1), 191–198.

Booth, L. (2004), 'Formulating retirement targets and the impact of time horizon on asset allocation', *Financial Services Review* **13**(1), 1.

Brechmann, E. C. & Czado, C. (2015), 'Copar—multivariate time series modeling using the copula autoregressive model', *Applied Stochastic Models in Business and Industry* **31**(4), 495–514.

Brennan, M. J., Schwartz, E. S. & Lagnado, R. (1997), 'Strategic asset allocation', *Journal of Economic Dynamics and Control* **21**(8), 1377–1403.

Brillinger, D. R. (1981), *Time series*, 2 edn, Calif.: Holden-Day Inc.

Brockwell, P. J. & Davis, R. A. (2006), *Time series: theory and methods*, 2 edn, Springer Science & Business Media.

Bühlmann, P. (1997), 'Sieve bootstrap for time series', *Bernoulli* **3**(2), 123–148.

Byrne, A., Dowd, K., Blake, D. P. & Cairns, A. J. (2006), 'Thereś no time like the present: the cost of delaying retirement saving'.

Campbell, J. Y. (2006), 'Household finance', *The journal of finance* **61**(4), 1553–1604.

Campbell, J. Y. & Viceira, L. M. (2002), *Strategic asset allocation: portfolio choice for long-term investors*, Oxford University Press, USA.

Cannon, E. S. & Tonks, I. (2008), *Annuity markets*, Oxford University Press.

Cannon, E. & Tonks, I. (2013), 'The value and risk of defined contribution pension schemes: international evidence', *Journal of Risk and Insurance* **80**(1), 95–119.

Carlstein, E. (1986), 'The use of subseries values for estimating the variance of a general statistic from a stationary sequence', *The Annals of Statistics* pp. 1171–1179.

Cavaliere, G. & Georgiev, I. (2013), 'Exploiting infinite variance through dummy variables in nonstationary autoregressions', *Econometric Theory* **29**(6), 1162–1195.

Cavaliere, G., Georgiev, I. & Robert Taylor, A. (2013), 'Wild bootstrap of the sample mean in the infinite variance case', *Econometric Reviews* **32**(2), 204–219.

Chambers, J. M., Mallows, C. L. & Stuck, B. (1976), 'A method for simulating stable random variables', *Journal of the american statistical association* **71**(354), 340–344.

Chambers, M. J. (2016), 'The estimation of continuous time models with mixed frequency data', *Journal of Econometrics* **193**(2), 390–404.

Chang, Y. & Park, J. Y. (2003), 'A sieve bootstrap for the test of a unit root', *Journal of Time Series Analysis* **24**(4), 379–400.

Cogneau, P. & Zakamouline, V. (2013), 'Block bootstrap methods and the choice of stocks for the long run', *Quantitative Finance* **13**(9), 1443–1457.

Cont, R. (2001), 'Empirical properties of asset returns: stylized facts and statistical issues', *Quantitative Finance* **1**(2), 223–236.

Cornea-Madeira, A. & Davidson, R. (2015), 'A parametric bootstrap for heavy-tailed distributions', *Econometric Theory* **31**(03), 449–470.

Csörgö, S., Haeusler, E. & Mason, D. M. (1988), 'A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables', *Advances in applied mathematics* **9**(3), 259–333.

Csörgő, S., Horváth, L. & Mason, D. M. (1986), 'What portion of the sample makes a partial sum asymptotically stable or normal?', *Probability Theory and Related Fields* **72**(1), 1–16.

Cullen, A. C. & Frey, H. C. (1999), *Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs*, Springer Science & Business Media.

Dahlhaus, R. & Janas, D. (1996), 'A frequency domain bootstrap for ratio statistics in time series analysis', *The Annals of Statistics* **24**(5), 1934–1963.
URL: *http://www.jstor.org/stable/2242635*

Davidson, J. (2000), *Time series analysis*, Blackwell Publishers Ltd, 108 Cowley Road, Oxford, UK.

Davidson, R. (2008), 'Exploring the bootstrap discrepancy', *COMPSTAT 2008* pp. 131–142.

Davidson, R. & MacKinnon, J. G. (2004), *Econometric theory and methods*, Vol. 5, Oxford University Press New York.

Davis, P. J. (2012), *Circulant matrices*, American Mathematical Soc.

Davis, R. A. & Mikosch, T. (1999), 'The maximum of the periodogram of a non-gaussian sequence', *The Annals of Probability* **27**(1), 522–536.

Davison, A. C. & Hinkley, D. V. (1997), *Bootstrap methods and their application*, Vol. 1, Cambridge university press.

del Barrio, E., Janssen, A. & Pauly, M. (2013), 'The m (n) out of k (n) bootstrap for partial sums of st. petersburg type games', *Electronic Communications in Probability* **18**, 1–10.

Delignette-Muller, M. L. & Dutang, C. (2015), 'fitdistrplus: An r package for fitting distributions', *Journal of Statistical Software* **64**(4), 1–34.

Dichtl, H. & Drobetz, W. (2011), 'Dollar-cost averaging and prospect theory investors: an explanation for a popular investment strategy', *Journal of Behavioral Finance* **12**(1), 41–52.

Dichtl, H., Drobetz, W. & Wambach, M. (2016), 'Testing rebalancing strategies for stock-bond portfolios across different asset allocations', *Applied Economics* **48**(9), 772–788.

Dierkes, M., Erner, C. & Zeisberger, S. (2010), 'Investment horizon and the attractiveness of investment strategies: A behavioral approach', *Journal of Banking & Finance* **34**(5), 1032–1046.

Dimson, E., Marsh, P. & Staunton, M. (2000), 'Risk and return in the 20th and 21st centuries', *Business Strategy Review* **11**(2), 1–18.

Dimson, E., Marsh, P. & Staunton, M. (2002), 'Triumph of the optimists', *Princeton University* .

Doeblin, W. (1940), 'Sur l'ensemble de puissances d'une loi de probabilité', *Studia mathematica* **9**(1), 71–96.

Donnelly, C., Guillén, M., Nielsen, J. P. & Pérez-Marín, A. M. (2015), 'On the practical implementation of retirement gains by using an upside and a downside terminal wealth constraint'.

Efron, B. (1979), 'Bootstrap methods: another look at the jackknife', *The annals of Statistics* pp. 1–26.

Elliott, G., Rothenberg, T. J. & Stock, J. H. (1996), 'Efficient tests for an autoregressive unit root', *Econometrica* **64**(4), 813–836.
    URL: *http://www.jstor.org/stable/2171846*

Fama, E. F. (1965), 'The behavior of stock-market prices', *The journal of Business* **38**(1), 34–105.

Fama, E. F. & Roll, R. (1971), 'Parameter estimates for symmetric stable distributions', *Journal of the American Statistical Association* **66**(334), 331–338.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. 2, second edn, Wiley.

Ferretti, N. & Romo, J. (1996), 'Unit root bootstrap tests for ar (1) models', *Biometrika* pp. 849–860.

Fofack, H. & Nolan, J. P. (1999), 'Tail behavior, modes and other characteristics of stable distributions', *Extremes* **2**(1), 39–58.

Forsyth, P. & Vetzal, K. (2016), Robust asset allocation for long-term target-based investing, Technical report, Working Paper, Cheriton School of Computer Science, University of Waterloo.

Franke, J. & Härdle, W. (1992), 'On bootstrapping kernel spectral estimates', *The Annals of Statistics* pp. 121–145.

Gabaix, X. & Ibragimov, R. (2011), 'Rank- 1/2: a simple way to improve the ols estimation of tail exponents', *Journal of Business & Economic Statistics* **29**(1), 24–39.

Giacomini, R., Politis, D. N. & White, H. (2013), 'A warp-speed method for conducting monte carlo experiments involving bootstrap estimators', *Econometric theory* **29**(3), 567.

Gnedenko, B. & Kolmogorov, A. (1968), *Limit distributions for sums of independent random variables*, Addison-Wesley series in statistics, Addison-Wesley.
    URL: *https://books.google.co.uk/books?id=rYsZAQAAIAAJ*

Godfrey, L. (2009), *Bootstrap Tests for Regression Models*, Palgrave Texts in Econometrics, Palgrave MacMillan.
    URL: *https://books.google.co.uk/books?id=XZgdAQAAMAAJ*

Gravetter, F. J. & Wallnau, L. B. (2016), *Statistics for the behavioral sciences*, Cengage Learning.

Gray, R. M. (2006), *Toeplitz and circulant matrices: A review*, now publishers inc.

Griffin, P. S. & Pruitt, W. E. (1987), 'The central limit problem for trimmed sums', *Mathematical Proceedings of the Cambridge Philosophical Society* **102**(02), 329–349.

Gujarati, D. N. (2009), *Basic econometrics*, 5 edn, Tata McGraw-Hill Education.

Gut, A. (2009), *An Intermediate Course in Probabiliy*, Springer Science+Business Media.

Gut, A. (2012), *Probability: a graduate course*, Vol. 75, Springer Science & Business Media.

Hall, P. & Yao, Q. (2003), 'Inference in ARCH and GARCH models with heavy–tailed errors', *Econometrica* **71**(1), 285–317.

Hamilton, J. D. (1994), *Time series analysis*, Princeton University Press, Chichester, West Sussex, UK.

Hannan, E. J. (1970), *Multiple time series*, John Wiley & Sons, Inc.

Hansson, B. & Persson, M. (2000), 'Time diversification and estimation risk', *Financial Analysts Journal* pp. 55–62.

Henze, N. & Zirkler, B. (1990), 'A class of invariant consistent tests for multivariate normality', *Communications in Statistics-Theory and Methods* **19**(10), 3595–3617.

Hidalgo, J. (2003), 'An alternative bootstrap to moving blocks for time series regression models', *Journal of Econometrics* **117**(2), 369–399.

Hill, B. M. (1975), 'A simple general approach to inference about the tail of a distribution', *The annals of statistics* **3**(5), 1163–1174.

Hinkley, D. (1989), 'Bootstrap significance tests', **8**, 65–74.

Hinkley, D. V. (1988), 'Bootstrap methods', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 321–337.

Horowitz, J. L. (2001), 'The bootstrap', *Handbook of econometrics* **5**, 3159–3228.

Hurvich, C. M. & Zeger, S. (1987), *Frequency domain bootstrap methods for time series*, New York University, Graduate School of Business Administration.

Ibragimov, M., Ibragimov, R. & Kattuman, P. (2013), 'Emerging markets and heavy tails', *Journal of Banking & Finance* **37**(7), 2546 – 2559.

Ibragimov, M., Ibragimov, R. & Walden, J. (2015), *Heavy-tailed distributions and robustness in economics and finance*, Vol. 214, Springer.

Ibragimov, R. (2014), 'On the robustness of location estimators in models of firm growth under heavy-tailedness', *Journal of Econometrics* **181**(1), 25 – 33. Heavy Tails and Stable Paretian Distributions.
URL: *http://www.sciencedirect.com/science/article/pii/S0304407614000347*

Ibragimov, R., Jaffee, D. & Walden, J. (2009), 'Nondiversification traps in catastrophe insurance markets', *The Review of Financial Studies* **22**(3), 959–993.
URL: *+ http://dx.doi.org/10.1093/rfs/hhn021*

Ibragimov, R. & Müller, U. K. (2010), 't-statistic based correlation and heterogeneity robust inference', *Journal of Business & Economic Statistics* **28**(4), 453–468.

Janicki, A. & Weron, A. (1994), 'Can one see α-stable variables and processes?', *Statistical Science* pp. 109–126.

Janssen, A. (1989), 'The domain of attraction of stable laws and extreme order statistics', *Probab. Math. Statist* **10**, 205–222.

Janssen, A. (1994), 'Sums of independent triangular arrays and extreme order statistics', *The Annals of Probability* pp. 1766–1793.

Janssen, A. & Pauls, T. (2003), 'How do bootstrap and permutation tests work?', *Annals of statistics* pp. 768–806.

Jentsch, C. & Kreiss, J.-P. (2010), 'The multiple hybrid bootstrap—resampling multivariate linear processes', *Journal of Multivariate Analysis* **101**(10), 2320–2345.

Johnston, K., Forbes, S. & Hatem, J. (2001), 'A comparison of state university defined benefit and defined contribution pension plans: a monte carlo simulation', *Financial Services Review* **10**(1), 37–44.

Kidd, D. (2011), 'The sharpe ratio and the information ratio', *Investment Performance Measurement Feature Articles* **2011**(1), 1–4.

Kinateder, J. (1992), 'An invariance principle applicable to the bootstrap', *Exploring the limits of bootstrap* **270**, 157.

Kirch, C. & Politis, D. N. (2011), 'Tft-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain', *The Annals of Statistics* **39**(3), 1427–1470.

Knight, K. (1989), 'On the bootstrap of the sample mean in the infinite variance case', *The Annals of Statistics* pp. 1168–1175.

Kole, E., Koedijk, K. & Verbeek, M. (2007), 'Selecting copulas for risk management', *Journal of Banking & Finance* **31**(8), 2405–2423.

Koutrouvelis, I. A. (1980), 'Regression-type estimation of the parameters of stable laws', *Journal of the American Statistical Association* **75**(372), 918–928.

Koutrouvelis, I. A. (1981), 'An iterative procedure for the estimation of the parameters of stable laws: An iterative procedure for the estimation', *Communications in Statistics-Simulation and Computation* **10**(1), 17–28.

Kreiss, J.-P. (1988), Asymptotic statistical inference for a class of stochastic processes, PhD thesis, Universität Hamburg.

Kreiss, J.-P. (1992), Bootstrap procedures for $ar(\infty)$—processes, *in* 'Bootstrapping and Related Techniques', Springer, pp. 107–113.

Kreiss, J.-P. & Paparoditis, E. (2003), 'Autoregressive-aided periodogram bootstrap for timeseries', *The Annals of Statistics* **31**(6), 1923–1955.

Kreiss, J.-P. & Paparoditis, E. (2011), 'Bootstrap methods for dependent data: A review', *Journal of the Korean Statistical Society* **40**(4), 357–378.

Kreiss, J.-P., Paparoditis, E. & Politis, D. N. (2011), 'On the range of validity of the autoregressive sieve bootstrap', *The Annals of Statistics* **39**(4), 2103–2130.

Künsch, H. R. (1989), 'The jackknife and the bootstrap for general stationary observations', *The Annals of Statistics* pp. 1217–1241.

Lahiri, S. N. (2003), *Resampling Methods for Dependent Data*, Springer Science & Business Media.

lbragimov, R. & Prokhorov, A. (2017), *Heavy Tails and Copulas*, World Scientific.

Lee, M. I. (2013), 'Stress testing monte carlo assumptions'.

Leibowitz, M. L. & Langetieg, T. C. (1989), 'Shortfall risk and the asset allocation decision: A simulation analysis of stock and bond risk profiles', *The Journal of Portfolio Management* **16**(1), 61–68.

LePage, R., Woodroofe, M. & Zinn, J. (1981), 'Convergence to a stable distribution via order statistics', *The Annals of Probability* pp. 624–632.

Lévy, P. (1925), *Calcul des probabilités*, Vol. 9, Gauthier-Villars Paris.

Lévy, P. & Borel, M. É. (1954), *Théorie de l'addition des variables aléatoires*, Vol. 1, Gauthier-Villars Paris.

Liu, R. Y. (1988), 'Bootstrap procedures under some non-iid models', *The Annals of Statistics* **16**(4), 1696–1708.

Liu, R. Y. & Singh, K. (1992), 'Moving blocks jackknife and bootstrap capture weak dependence', *Exploring the limits of bootstrap* **225**, 248.

Lloyd, W. P. & Modani, N. K. (1983), 'Stocks, bonds, bills, and time diversification', *The Journal of Portfolio Management* **9**(3), 7–11.

Maddison-Project (2008). data retrieved from, http://www.ggdc.net/maddison/maddison-project/home.htm.

Mammen, E. (1992), *When does bootstrap work?: asymptotic results and simulations*, Vol. 77, Springer, New York.

Mandelbrot, B. B. (1997), *The variation of certain speculative prices*, Springer.

Manner, H. (2007), *Estimation and model selection of copulas with an application to exchange rates*, Citeseer.

McCulloch, J. H. (1986), 'Simple consistent estimators of stable distribution parameters', *Communications in Statistics-Simulation and Computation* **15**(4), 1109–1136.

McCulloch, J. H. (1997), 'Measuring tail thickness to estimate the stable index α: A critique', *Journal of Business & Economic Statistics* **15**(1), 74–81.

Meyer, M. & Kreiss, J.-P. (2015), 'On the vector autoregressive sieve bootstrap', *Journal of Time Series Analysis* **36**(3), 377–397.

Mikosch, T. & Starica, C. (1998), *Change of structure in financial time series, long range dependence and the GARCH model*, Chalmers Tekniska Högskola/Göteborgs Universitet. Department of Mathematics.

Mitchell, B. (1998), *International historical statistics: Europe 1750-1993*, Springer.

Moore, J. & Nambiar, V. (2012), 'The journey to ldi and beyond', *The Journal of Investing* **21**(2), 136–151.

Mukherji, S. (2011), 'Are stock returns still mean-reverting?', *Review of Financial Economics* **20**(1), 22–27.

Naidu, P. S. (1995), *Modern spectrum analysis of time Series: Fast algorithms and error control techniques*, CRC Press.

Nelsen, R. (2006), 'An introduction to copulas, 2nd', *New York: SpringerScience Business Media* .

Newey, W. K. & West, K. D. (1994), 'Automatic lag selection in covariance matrix estimation', *The Review of Economic Studies* **61**(4), 631–653.

Newey, W. & West, K. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(1-4), 703–708.

Ng, S. & Perron, P. (2001), 'Lag length selection and the construction of unit root tests with good size and power', *Econometrica* **69**(6), 1519–1554.
URL: *http://www.jstor.org/stable/2692266*

Nolan, J. P. (1997), 'Numerical calculation of stable densities and distribution functions', *Communications in statistics. Stochastic models* **13**(4), 759–774.

Nolan, J. P. (2001), 'Maximum likelihood estimation of stable parameters', *Levy processes: Theory and applications* pp. 379–400.

Nolan, J. P. (2015), *Stable Distributions - Models for Heavy Tailed Data*, Birkhauser, Boston. In progress, Chapter 1 online at academic2.american.edu/~jpnolan.

Norstrøm, C. J. (1972), 'A sufficient condition for a unique nonnegative internal rate of return', *The Journal of Financial and Quantitative Analysis* **7**(3), 1835–1839.
URL: *http://www.jstor.org/stable/2329806*

Paparoditis, E. (1996), 'Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes', *Journal of Multivariate Analysis* **57**(2), 277–296.

Paparoditis, E. & Politis, D. N. (1999), 'The local bootstrap for periodogram statistics', *Journal of Time Series Analysis* **20**(2), 193–222.

Paparoditis, E. & Politis, D. N. (2005), 'Bootstrap hypothesis testing in regression models', *Statistics & probability letters* **74**(4), 356–365.

Park, J. Y. (2002), 'An invariance principle for sieve bootstrap in time series', *Econometric Theory* **18**(02), 469–490.

Park, J. Y. (2003), 'Bootstrap unit root tests', *Econometrica* **71**(6), 1845–1895.

Patton, A. J. (2012), 'A review of copula models for economic time series', *Journal of Multivariate Analysis* **110**, 4–18.

Perron, P. (1989), 'The great crash, the oil price shock, and the unit root hypothesis', *Econometrica* **57**(6), 1361–1401.
URL: *http://www.jstor.org/stable/1913712*

Peters, G. W. & Shevchenko, P. V. (2015), *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*, John Wiley & Sons.

Phillips, P. (1999), Discrete fourier transforms of fractional processes, Technical report, Discussion Paper, Cowles Foundation, Yale University.

Phillips, P. C. (2009), 'Long memory and long run variation', *Journal of Econometrics* **151**(2), 150–158.

Phillips, P. C. B. & Perron, P. (1988), 'Testing for a unit root in time series regression', *Biometrika* **75**(2), 335–346.
URL: *http://www.jstor.org/stable/2336182*

Politis, D. N. & Romano, J. P. (1994), 'The stationary bootstrap', *Journal of the American Statistical association* **89**(428), 1303–1313.

Politis, D. N., Romano, J. P. & Wolf, M. (1999), *Subsampling*, Springer Science & Business Media.
URL: *http://books.google.com/books?id=nGu6rqjE6JoC&pgis=1*

Poterba, J., Rauh, J., Venti, S. & Wise, D. (2006), Lifecycle asset allocation strategies and the distribution of 401 (k) retirement wealth, Technical report, National Bureau of Economic Research.

Romano, J. P. & Wolf, M. (1999), 'Subsampling inference for the mean in the heavy-tailed case', *Metrika* **50**(1), 55–69.

Said, S. E. & Dickey, D. A. (1984), 'Testing for unit roots in autoregressive-moving average models of unknown order', *Biometrika* **71**(3), 599–607.
URL: *http://www.jstor.org/stable/2336570*

Samoradnitsky, G. & Taqqu, M. S. (1994), *Stable non-Gaussian random processes: stochastic models with infinite variance*, Vol. 1, CRC Press.

Sato, K.-i. (2013), *Lévy Processes and Infinitely Divisible Distributions*, Cambridge university press, Cambridge,UK.

Sato, K.-i. & Yamazato, M. (1978), 'On distribution functions of class l', *Probability Theory and Related Fields* **43**(4), 273–308.

Schleef, H. J. & Eisinger, R. M. (2007), 'Hitting or missing the retirement target: Comparing contribution and asset allocation schemes of simulated portfolios', *Financial Services Review* **16**(3), 229.

Shao, J. & Tu, D. (1995), *The jackknife and bootstrap*, Springer Science & Business Media.

Sharpe, W. F. (1994), 'The sharpe ratio', *The journal of portfolio management* **21**(1), 49–58.

Shiller, R. J. (2006), 'Life-cycle personal accounts proposal for social security: An evaluation of president bush's proposal', *Journal of policy modeling* **28**(4), 427–444.

Siegel, J. J. (1998), *Stocks for the Long Run by Jeremy Siegel*, New York: McGraw-Hill.

Simons, G. & Stout, W. (1978), 'A weak invariance principle with applications to domains of attraction', *The Annals of Probability* **6**(2), 294–315.
    URL: *http://www.jstor.org/stable/2243219*

Singh, K. (1981), 'On the asymptotic accuracy of efron's bootstrap', *Ann. Statist.* **9**(6), 1187–1195.
    URL: *http://dx.doi.org/10.1214/aos/1176345636*

Sokal, R. & Rohlf, F. (1999), *Biometry*, W.H. Freeman and Company, USA.

Sortino, F. A. & Price, L. N. (1994), 'Performance measurement in a downside risk framework', *the Journal of Investing* **3**(3), 59–64.

Sortino, F. A. & Van Der Meer, R. (1991), 'Downside risk', *The Journal of Portfolio Management* **17**(4), 27–31.

Tasche, D. (2002), 'Expected shortfall and beyond', *Journal of Banking & Finance* **26**(7), 1519–1533.

Teugels, J. L. (1981), 'Limit theorems on order statistics', *The Annals of Probability* pp. 868–880.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. (1992), 'Testing for nonlinearity in time series: the method of surrogate data', *Physica D: Nonlinear Phenomena* **58**(1-4), 77–94.

Tiao, G. C. & Tsay, R. S. (1983), 'Multiple time series modeling and extended sample cross-correlations', *Journal of Business & Economic Statistics* **1**(1), 43–56.

Turnbull, H. W. (1952), 'Theory of equations'.

Wang, G. & Mao, W.-L. (2008), 'Unit root testing in the presence of heavy-tailed garch errors', *Australian & New Zealand Journal of Statistics* **50**(3), 273–292.

Wang, K. (2012), Long-term Investment and Asset Allocation Strategies in Defined Contribution Pension Plans, PhD thesis, University of Exeter. unpublished thesis.

Weron, A. & Weron, R. (1995), *Computer simulation of Lévy α-stable variables and processes*, Springer.

Wu, C.-F. J. (1986), 'Jackknife, bootstrap and other resampling methods in regression analysis', *the Annals of Statistics* pp. 1261–1295.

Yamazato, M. (1978), 'Unimodality of infinitely divisible distribution functions of class *l*', *The Annals of Probability* **6**(4), 523–531.

Zolotarev, V. M. (1986), *One-dimensional stable distributions*, Vol. 65, American Mathematical Soc.

Appendix A

Mathematical Background For Chapter One

A.1 Order Statistics

Order Statistics are one of the most powerful tools in non-parametric statistics. This power comes from some valuable inferences one may be able to make by ranking observations of a random sample and studying their joint distributions. Unlike Athreya (1987), Knight (1989) has shown that the naive bootstrap fails to mimic the distribution of observations possessing heavy-tails based on the concept of Order Statistics. Knight (1989) has utilized the results of LePage et al. (1981) to prove his theory. LePage et al. (1981) used a valuable property in Order of Statistics to show that the largest few summands determine the limiting distribution. The result LePage et al. (1981) reached is critical for our chapter 1. Therefore, this section gives more insight onto the property LePage et al. (1981) used to reach their conclusion. Moreover, this important property has not been mentioned in the recent literature. This section will therefore be dedicated to provide further explanation of *Order Statistics* and revising previous properties that are critical for our proofs.

The order statistic is the arrangement in a decreasing order, obtained from the original (unordered) sample, through permutation; that is, that:

Definition A.1.1. *For $k = 1, 2, \cdots, n$, let*

$$X_{(k)} = the \ k^{th} \ smallest \ of \ X_1, X_2, \cdots, X_n$$

$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ *is called the order statistic and* $X_{(k)}$ *the* k^{th} *order variable,* $k = 1, 2, \cdots, n.$

The ordered variables $\{X_{(i)}\}_{i=1}^n$ also depend on the number of observations n. Therefore, for complete transparency, it would be better to denote the random variable $X_{(i)}$ as $X_{i:n}$.

Consider $U_1, U_2, \cdots, U_n$ to be independent random variables uniformly distributed on $(0,1)$. The order statistics of the $\{U_i\}$ will be represented by $\{U_{i:n}\}$. Moreover, define a random variable Z_k by

$$Z_k = \sum_{i=1}^k W_i$$

where $k \leq n$, and an i.i.d. sequence $\{W_i\}_{i=1}^n$ of exponential random variables. Then the following proposition holds.

Proposition A.1.2. *The sequence $\{U_{i:n}\}$ has the same joint distribution as $\{Z_k/Z_{n+1}\}$.*

Proof. The proof consists of three steps. The initial step entails developing an equality which will help in developing the joint probability density function of Z_k/Z_{n+1}. The second step expands and reduces the joint probability density of Z_k/Z_{n+1}. The last step explains the joint distribution which will be equal to that of the observed in step two. Denote by the probability measure π defined as

$$\mathbb{P}(Z_1 \in dx_1, Z_2 \in dx_2, \cdots, Z_n \in dx_n)$$

which is defined on the product space of n dimensional $\mathbb{R}$. However, note that: $Z_1 = W_1$, $Z_2 = W_2 + Z_1, \cdots, Z_n = W_n + Z_{n-1}$. Therefore the random variables $\{Z_i\}$ are dependent while the random variables $\{W_i\}$ are independent. Thus it is of interest to transform the set

$$\bigcap_{i=1}^n \{Z_i \in dx_i\} = \bigcap_{i=1}^n \{W_i \in dx_i - x_{i-1}\}$$

where $x_0 = 0$ hence the probability measure π

$$\pi = \mathbb{P}(W_1 \in dx_1, W_2 \in dx_2 - x_1, \cdots, W_n \in dx_n - x_{n-1})$$

observe that $\mathbb{P}(W_1 \in dx_1) \approx f_{W_1}(x_1)dx_1 = e^{-x_1}dx_1$ and that $\mathbb{P}(W_2 \in dx_2 - x_1) \approx f_{W_1}(x_2 - x_1)dx_2$ hence

$$
\begin{aligned}
\pi &\approx f_{W_1}(x_1)f_{W_1}(x_2 - x_1)\cdots f_{W_1}(x_{n+1} - x_n)dx_1\cdots dx_{n+1}\\
&= \prod_{i=1}^{n+1} f_{W_i}(x_i - x_{i-1})dx_i\\
&= \prod_{i=1}^{n+1} e^{-(x_i - x_{i-1})}dx_i\\
&= e^{-x_{n+1}}dx_1\cdots dx_{n+1}
\end{aligned}
$$

Now we need to interpret Z_i as a sum of exponential random variables; hence it would be a Gamma distribution

$$Z_i \sim Gamma(i, 1)$$

hence its pdf is given as

$$f_{Z_i}(x) = \frac{1}{(i-1)!} x^{i-1} e^{-x}$$

Then the following equation shows

$$
\begin{aligned}
&\mathbb{P}(Z_1 \in dx_1, Z_2 \in dx_2, \ldots, Z_n \in dx_n \mid Z_{n+1} = x_{n+1}) \\
&= f_{Z_{n+1}}(x_{n+1})^{-1} \left[\mathbb{P}(\{Z_1 \in dx_1, Z_2 \in dx_2, \ldots, Z_n \in dx_n\} \right. \\
&\qquad \left. \cap \{Z_{n+1} = x_{n+1}\}) \right] \\
&= f_{Z_{n+1}}(x_{n+1})^{-1} \times \pi \\
&= e^{-x_{n+1}} dx_1 \cdots dx_{n+1} / f_{Z_{n+1}}(x_{n+1}) \\
&= n! \, x^{-n} dx_1 \cdots dx_n
\end{aligned}
$$

Then Normalize by dividing all the entries by Z_{n+1}

$$
\begin{aligned}
&\mathbb{P}\left(\frac{Z_1}{Z_{n+1}} \in dy_1, \frac{Z_2}{Z_{n+1}} \in dy_2, \ldots, \frac{Z_n}{Z_{n+1}} \in dy_n \mid Z_{n+1} = x_{n+1} \right) \\
&= n! \, dy_1 \cdots dy_n
\end{aligned}
$$

when $0 \leq y_1 \leq \cdots \leq y_n \leq 1$. On the other hand, consider the i.i.d. sequence of random variables $\{U_i\}$ and its mapping $\{U_{i:n}\}$. This mapping is not one to one as it is a permutation. Observe that there are $n!$ different outcomes that all generate the same order statistic. Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(U_{(1)} \in dy_1, U_{(2)} \in dy_2, \ldots, U_{(n)} \in dy_n \right) \\
&= \sum_{\text{all possible outcomes}} \mathbb{P}\left(U_1 \in dy_1, U_2 \in dy_2, \ldots, U_n \in dy_n \right) \\
&= n! \, dy_1 \cdots dy_n
\end{aligned}
$$

$\square$

One important result in order statistics is the following proposition

Proposition A.1.3. *An n dimensional random vector with i.i.d. elements $\{X_{(i)}\}_{i=1}^{n}$ is equal in distribution to $\{F_X^{-1}(Z_i/Z_{n+1})\}_{i=1}^{n}$.*

Proof. It is easy to observe that if X with continuous CDF F_X, $F_X(X)$ is uniform on $(0,1)$ proof of such result is available in many text books including Ahsanullah et al. (2013) and Gut (2009). Therefore, based on this result, it should be easy to understand that the distribution of the order i.i.d. elements $\{X_{(i)}\}_{i=1}^{n}$ is equal in distribution to $\{U_{(i:n)}\}_{i=1}^{n}$ where $U_{(i:n)}$ is as defined above. Therefore, using Proposition A.1.2, the result follow immediately.   □

Discussion:

Now, and thereafter, let $X \in DA(Z)$ such that $Z \in \mathbb{S}(\alpha)$; and let F_X denote its distribution function. Let S_n be defined as in equation A.25, where the normalizing constants a_n and b_n follow equations A.27 and A.28 respectively. Hence, this statistic, S_n converges in distribution to a stable random variable Z_α. where α is the index of stability.

It may also be shown that based on the views of equation A.22 the following equations hold, as $x \to +\infty$

$$\frac{F_X(-x)}{1 - F_X(x) + F_X(-x)} \to q \qquad \text{and} \qquad \frac{F_X(x)}{1 - F_X(x) + F_X(-x)} \to p \qquad \text{(A.1)}$$

where $p, q \geq 0$ such that $p + q = 1$. Moreover, observe that

$$F_{|X|}(x) = \mathbb{P}(|X| > x) = 1 - F_X(x) + F_X(-x) \qquad \text{(A.2)}$$

Therefore, based on equations A.1, and A.2, it should be clear that as $x \to +\infty$ the following equations holds

$$F_X(-x) \sim qF_{|X|}(|x|) \qquad \text{and} \qquad F_X(x) \sim pF_{|X|}(x)$$

Let $\Gamma_k = (\sum_{i=1}^{k} E_i)$ where $\{E_i\}$ is a sequence of i.i.d. exponential r.v.s with mean one. Then, as a consequence of the Law of Large numbers, it may be shown that:

$$\frac{1}{k}\Gamma_k = \frac{1}{k}\sum_{i=1}^{k} E_i \overset{a.s.}{\to} \mathbb{E}(E_i) = 1$$

Hence, a gamma distributed random variable Γ_k divided by its shape-parameter k converges almost surely to 1 as its shape-parameter approaches infinity, conditional on the fact that the scale is 1.

Let F_X^{-1} be the inverse of F_x defined as:

$$F_X^{-1}(x) = \inf\{y : F_X(y) \le x\}$$

Observe that:

$$
\begin{aligned}
a_n^{-1} F_X^{-1}(\Gamma_k/\Gamma_{n+1}) &= \inf\{a_n^{-1}y : F_X(y) \le \Gamma_k/\Gamma_{n+1}\} \\
&= \inf\left\{z : n\left(\frac{\Gamma_{n+1}}{n+1} \times \frac{n+1}{n}\right) F_X(a_n z) \le \Gamma_k\right\} \\
&\sim
\begin{cases}
\inf\{z : np F_{|X|}(a_n z) \le \Gamma_k\} & \text{for } k \text{ large, hence } z \ge 0 \\
\inf\{z : nq F_{|X|}(-a_n z) \le \Gamma_k\} & \text{for } k \text{ small, hence } z \le 0
\end{cases} \\
&\to
\begin{cases}
\inf\{z : pz^{-\alpha} \le \Gamma_k\} & \text{for } k \text{ large, hence } z \ge 0 \\
\inf\{z : q(-z)^{-\alpha} \le \Gamma_k\} & \text{for } k \text{ small, hence } z \le 0
\end{cases} \\
&=
\begin{cases}
\inf\{z : z \le p^{1/\alpha}\Gamma_k^{-1/\alpha}\} & \text{for } k \text{ large, hence } z \ge 0 \\
\inf\{z : z \le -q^{1/\alpha}\Gamma_k^{-1/\alpha}\} & \text{for } k \text{ small, hence } z \le 0
\end{cases} \\
&=
\begin{cases}
p^{1/\alpha}\Gamma_k^{-1/\alpha} & \text{for } k \text{ large} \\
-q^{1/\alpha}\Gamma_k^{-1/\alpha} & \text{for } k \text{ small}
\end{cases}
\end{aligned}
\tag{A.3}
$$

where z is introduced as a substitution parameter $z := a_n^{-1}y$. The limit $\to$ stand for as $n \to +\infty$. Equation A.27 has been used among the lines.

A.2 Revisiting Various Aspects of Stable Distributions

A.2.1 Stable and Infinitely Divisible Distributions

In general, stability is a condition which describes a class of distribution under which the sum or a linear combination of i.i.d. random variables, that possess this distribution, add up to a random variable that possesses the same distribution. For example, the sum of i.i.d. normal random variables is equal in distribution to a normal random variable. Similarly, this result holds for all stable random variables such as Cauchy and Lévy. There are many equivalent characterizations for the definition of the family of univariate α-stable distribution (see, among others, Samoradnitsky & Taqqu 1994, Nolan 2015, Ibragimov et al. 2015, Peters & Shevchenko 2015). Next, we will define the family of univariate α-stable distribution.

Definition A.2.1. *The probability distribution of a random variable X, F_X, is stable in the broad sense if there exist constants $\alpha \in (0,2]$ and $d_n \in \mathbb{R}$, such that*

$$\Sigma_n \overset{d}{=} n^{1/\alpha} X + d_n \tag{A.4}$$

where Σ_n is the partial sum of the i.i.d. random sequence $\{X_i\}_{i=1}^{n}$ that are identical copies of X. X is strictly stable if and only if $d_n = 0$ for all n.

From the above characterization of the univariate stable distribution, it should be obvious that this class of distributions satisfies the property of closure under convolution. In particular, the tail index α is maintained across the sums.

The property of closure under convolution is closely related to the concept of the domain of attraction. Consider the following definition for the domain of attraction of a non-degenerate distributions.

Definition A.2.2. *An i.i.d. sequence of random variables $\{X_i\}_{i=1}^{n}$ with partial sum, Σ_n, belongs to the domain of attraction of the non-degenerate distribution G if there exists a normalizing sequence $\{a_n > 0, \ n \geq 1\}$ and centering sequence $\{b_n, \ n \geq 1\}$ such that*

$$\frac{\Sigma_n - b_n}{a_n} \overset{d}{\to} Z, \quad as \ n \to \infty \tag{A.5}$$

where Z is a random variable that possess a distribution G. The notation is $F_X \in \mathscr{DA}(G)$, or alternatively, $X \in \mathscr{DA}(Z)$ if $Z \in G$.

Remark: If Z in equation (A.5) is a stable random variable with index of stability α, then the above definition will represent the Generalized central limit theorem (see, Theorem A.2.7) and *the notation is $X \in \mathscr{DA}(Z)$ such that $Z \in \mathbb{S}(\alpha)$ or simply $X \in \mathscr{DA}(\alpha)$.*

Based on the characterization of stable distributions described in definition A.2.1 another equivalent characterization emerges. In particular, if X is an α stable random variable then it belongs to its own domain of attraction. That is, for a suitably normalized sum of a sequence of discrete stable random variables $\{X_i\}_{i=1}^{n}$ there exist constants dependent on n such that the sum converges to the same distribution (see, equation (4.3), Peters & Shevchenko 2015). This characterization poses the question of whether only stable distributions, when suitably normalized, converge in distribution (weak-convergence) to their own law. The answer to this question is not necessarily.

Although Pareto distribution is not among the class of stable distributions, the i.i.d. sum of Pareto random variables converge to a stable law (Blum 1970). Therefore, there

exist non-stable distributions, that belong to the domain of normal attraction of a stable law, that are themselves not stable. Moreover, it is well-known that all stable distributions are infinitely divisible. However, not every infinitely divisible random variable is stably distributed. For example, a sum of two infinitely divisible random variables such as the Gaussian and the Cauchy results in an infinitely divisible random variable, say X, that is not stable but a combination of stable and normal. In fact, the univariate α-stable distribution family represents only a sub-class of the so-called class of $\mathbb{L}$ distributions. The class of $\mathbb{L}$ distributions represents the family of all distributions that are infinitely divisible[1], such that the normed summand are asymptotically constant[2].

Using the properties of closure under convolution and infinite divisibility, Lévy (1925) and Lévy & Borel (1954) specified the distributional membership of the $\mathbb{L}$ distributions family via the characteristic function. However, Khintchine gave a simpler proof on the real line ($\mathbb{R}$). Therefore, this representation has been known as the Lévy-Khintchine representation.

Definition A.2.3 (Lévy-Khintchine representation). *Given a distribution function F_X of an infinitely divisible random variable X, then the logarithm of its characteristic function, Φ_X, admits the Lévy-Khintchine representation given generally by*

$$\Phi_X(t) = \exp(\Psi(t)) \tag{A.6}$$

where the exponent of the characteristic function $\Psi(t)$ is given by

$$\Psi(t) = ia_g t - \frac{1}{2}\sigma^2 t^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{itu} - 1 - itg(u))W(du) \tag{A.7}$$

for some truncation function $g : \mathbb{R} \to \mathbb{R}$ which is measurable and satisfies for all t the condition

$$\int |e^{itu} - 1 - itg(u)|W(du) \tag{A.8}$$

with a drift component $a_g \in \mathbb{R}$ that depends on the choice of the truncation function (g) and a diffusion component σ^2, such that $\sigma > 0$, and the jump component represented by the Lévy-measure[3] W which satisfies $W(0) = 0$ almost surely and

$$\int_{\mathbb{R}\setminus\{0\}} \min\{u^2, 1\}W(du) < \infty \tag{A.9}$$

---

[1] An infinitely divisible random variable is a random variable that may be written in terms of sum of arbitrary number of i.i.d. random variables (see Gut 2012, p. 442).

[2] see (Definition 4.14, Peters & Shevchenko 2015, p. 158)

[3] Is a Borel measure on $\mathbb{R}$, with a certain criteria to be satisfied. It usually describes the distribution of the jumps of the process.

Remarks:

- It should be emphasised that every infinitely divisible probability distribution corresponds in a natural way to a Lévy process (Corollary 11.6, Sato 2013, p. 63). A Lévy process is a stochastic process with independent, stationary increments which have continuous probabilities. In fact, Brownian motion and Poisson process belong to a class called the Levy process with minor additional conditions to be satisfied (see, definition 5.1, Sato 2013, p. 22).

- In general, the jump part of a Levy process is a compensated sum of the Poisson point process with the same characteristic measure as the Levy measure.

- The general Lévy-Khintchine representation may be uniquely defined by the Lévy triplet, also known as the generating triplet, (a_g, σ^2, W) (see, for details, Sato 2013, p. 37-39).

- Using the α-stable Lévy spectral measure which may be defined by

$$W(du) = \frac{P}{u^{1+\alpha}} 1_{(0,+\infty)}(u) du + \frac{Q}{|u|^{1+\alpha}} 1_{(-\infty,0)}(u) du \qquad (A.10)$$

(see, Janicki & Weron 1994) or alternatively by Lévy measure

$$W(u) = -c_1 u^{-\alpha} 1_{(0,+\infty)}(u) + c_2 |u|^{-\alpha} 1_{(-\infty,0)}(u) \qquad (A.11)$$

(see, Theorem C.1, Zolotarev 1986) where P, Q, c_1, c_2 are non-negative numbers[4]. It may be verified that the family of α-stable distributions may be represented by the Lévy-Khintchine representation; hence, it belongs to a sub-class in the class of the $\mathbb{L}$ distributions family[5].

In his paper, Janssen (1994) used the Lévy-Hinčin formula[6] for infinitely divisible distributions and has shown that the convergent normalized sum of infinitely divisible i.i.d. random variables converges to an infinitely divisible random variable, say ζ, that may be represented by a sum of three distinct random variables that is equivalent to it in distribution. In particular, ζ may be represented by a negative Poisson part Δ^-, a positive Poisson part Δ^+, and a Gaussian part N with variance σ^2 that represents the diffusion component in equation

---

[4]It can be shown that $P := \alpha p$ and $Q := \alpha q$. Moreover, it should be noted that $W_n(A) = n\mathbb{P}(a_n^{-1} X \in A)$ converge as $n \to \infty$ to W (see, proof of Lemma 1 and 2, LePage et al. 1981, p. 626)

[5]For details see (Theorem C.2, Zolotarev 1986, p. 9) and (Sato 2013, p. 80)

[6]The Lévy-Hinčin formula is a variant of the Lévy-Khintchine representation with a truncation function $g(u) := u1_{(|u|<\tau)}(u)$ and a drift component $a_g := 0$

(A.7). More rigorously, one may show the following representation of the infinitely divisible random variable.

Theorem A.2.4. *Any infinitely divisible random variable ζ may be represented by the quantile representation*

$$\zeta \stackrel{d}{=} \Delta^- + N + \Delta^+ + K \tag{A.12}$$

where N is a Gaussian random variable with zero mean and variance σ^2 that is identical with the diffusion parameter in equation (A.7) for characteristic equation of Φ_ζ. K is a constant. A negative Poisson part

$$\Delta^- := \sum_{i=1}^{\infty} \left[\psi_1(\Gamma_i) - E\left(\psi_1(\Gamma_i)\mathbf{1}_{(-\tau,0]}(\psi_1(\Gamma_i))\right) \right] \tag{A.13}$$

A positive Poisson part

$$\Delta^+ := \sum_{i=0}^{\infty} \left[\psi_2\left(\tilde{\Gamma}_i\right) - E\left(\psi_2\left(\tilde{\Gamma}_i\right)\mathbf{1}_{(-\tau,0]}\left(\psi_2\left(\tilde{\Gamma}_i\right)\right)\right) \right] \tag{A.14}$$

where Γ_i and $\tilde{\Gamma}_i$ are independent gamma random variables with scale i and shape 1. ψ_1 and ψ_2 are quantile functions of the Lévy measure W in the characteristic equation of Φ_ζ (see, equations (A.10) and (A.11)). They are defined by

$$\psi_1(y) := \min\left\{\inf\left\{t : W(-\infty,t] \ge y\right\}, 0\right\} \tag{A.15}$$

and

$$\psi_2(y) := \max\left\{\sup\left\{t : W[t,\infty) \ge y\right\}, 0\right\} \tag{A.16}$$

Proof. The proofs are identical with Janssen (1994) and Csörgö et al. (1988) □

Remarks:

- The sum of independent random variables, some of which are normal and some of which are stable, converge, when suitably scaled and normalized, to a random variable, say ζ, which consists of a normal part and a Poisson part, as discussed in Theorem A.2.4.

- Each infinitely divisible random variable can uniquely be decomposed into the sum of two independent infinitely divisible random variables. There exist some cases when the Gaussian part of the infinitely divisible random variable, ζ, vanishes when $\sigma^2 = 0$. In this case, the infinitely divisible random variable ζ may be represented by Possion parts only. Stable random variables with $\alpha < 2$ are a case in point.

Janssen (1994) is interested in finding out which parts of the normalized sum Σ_n of an i.i.d. sequence of infinitely divisible random variables $\{X_i\}_{i=1}^n$ are responsible for the normal part and the Poisson part in the limiting random variable. He found that the middle part of the sum converges asymptotically to a normal distribution, while the upper and the lower extremes of the sum converge to the positive and negative spectral Poisson part of the limiting random variable respectively. His findings are for general infinitely divisible distributions that are asymptotically convergent when suitably scaled and normalized. However, in chapter 1, special attention will be devoted to the behaviour of the central part of the sum of random variables that belong to the domain of normal attraction of a stable law, which is a sub-class of the class $\mathbb{L}$.

Remark: Using *Order Statistic* techniques, it has been shown that an n dimensional random vector with i.i.d. elements $\{X_{(i)}\}_{i=1}^n$ is equal in distribution to $\{F_X^{-1}(\Gamma_{i,n}/\Gamma_{n+1,n+1})\}_{i=1}^n$ (see, proposition A.1.3). In fact, Janssen (1994) has used this idea in his proof of Theorem A.2.4 to show its results. However, to show independence between the lower and upper parts of the partial sum of Σ_n in equation (A.25) as $n \to \infty$, Janssen (1989) has written the random variable $\Gamma_{k,n}$ as a sum of two independent series $(\{e_i\}_{i=1}^{\lfloor n/2 \rfloor}, \{\tilde{e}_i\}_{i=1}^{\lfloor n/2 \rfloor})$ with common exponential distribution with mean 1 as

$$\Gamma_{k,n} = \begin{cases} \sum_{i=1}^{\lfloor n/2 \rfloor} e_i & k \leq \lfloor n/2 \rfloor \\ \Gamma_{\lfloor n/2 \rfloor, n} + \sum_{i=1}^{k - \lfloor n/2 \rfloor} \tilde{e}_{n+1-i-\lfloor n/2 \rfloor} & k > \lfloor n/2 \rfloor \end{cases} \tag{A.17}$$

(see, equation (3.5), Janssen 1989, p. 211) where $\lfloor . \rfloor$ is the floor operator.

The most concrete and practical way to describe stable distributions in terms of functionals, is through characteristic functions or Fourier transformation (see Nolan 2015, p. 7). The specification of the characteristic function of α-stable distributions may be derived from the Lévy-Khintchine representation (Theorem C.2., Zolotarev 1986, p. 9) and (Section 4.3.4, Peters & Shevchenko 2015, p. 167). Such a derivation would lead to the A-type parameterization of Zolotarev (1986) of an α-stable characteristic function. Definition A.2.5 gives a variant of the representation of Zolotarev's A type characteristic function.

Definition A.2.5. *A random variable X is stable if and only if there exist $a \neq 0$, $b \in \mathbb{R}$ such that*

$$X \stackrel{d}{=} aZ + b$$

where Z is a random variable with a characteristic function

$$\mathbb{E}(e^{itZ}) = \begin{cases} \exp(-|t|^{\alpha}(1 - i\beta \, \text{sign}(t) \tan(\frac{\pi\alpha}{2})) & \alpha \neq 1 \\ \exp(-|t|(1 + i\frac{2}{\pi}\beta \, \text{sign}(t) \log|t|)) & \alpha = 1 \end{cases}$$

Where $\alpha \in (0, 2]$, and $\beta \in [-1, 1]$.

Based on definition A.2.5, it is obvious that the probability distribution function of a stable distribution, F_X, is characterized by the functionals describing its shape, location, and scale. Indeed, definition A.2.5 has defined those parameters in only one form; however, there are many other forms that may equally define a stable law (see Nolan 2015, p. 7). Originally, Zolotarev (1986) introduced parametric parameterizations of the characteristic exponent of stable laws. The idea of the parameterization of stable law is to select the explanatory parameters precisely, such that it is practical for analysis and gives solid knowledge on the shape of the stable density function.

In their investigation of the tail behavior of non-Gaussian stable distributions, Fofack & Nolan (1999) were engaged in locating the mode of the stable distribution as a function of two parameters that defines the distribution's shape. However, unlike the popular parameterization of Samoradnitsky & Taqqu (1994) (see, definition A.2.6), Fofack & Nolan (1999) used a different parameterization, simply because they may better understand the bounds of the mode employed by the alternative parameterization. In chapter 1, the parameterization of Samoradnitsky & Taqqu (1994) will be adopted. Therefore, it is of great importance to define this parameterization here.

Definition A.2.6. *A random variable X, that possess a stable distribution, has a characteristic function that is mainly classified by four parameters: the index of stability $\alpha \in (0, 2]$, skewness parameter $\beta \in [-1, 1]$, a scale parameter $\sigma > 0$, and a location parameter $\mu \in \mathbb{R}$. It is mathematically defined as*

$$\mathbb{E}(e^{itX}) = \begin{cases} \exp(-\sigma^{\alpha}|u|^{\alpha}[1 - i\beta \, \text{sign}(t) \tan(\frac{\pi\alpha}{2})] + i\mu t) & \alpha \neq 1 \\ \exp(-\sigma|u|[1 + i\frac{2}{\pi}\beta \, \text{sign}(t) \log|t|] + i\mu t) & \alpha = 1 \end{cases} \quad \text{(A.18)}$$

Remarks:

- Equation (A.18) is characterized by four parameters. Therefore, here, and thereafter, $S(\alpha, \beta, \sigma, \mu)$ will be used to denote stable distributions that follow such parameterizations.

- As Weron & Weron (1995) pointed out, there is a natural interpretation for each parameter. The location parameter μ shifts the distribution to the left or right depending on its sign. The scale parameter σ compresses or extends the distribution about μ depending on whether its magnitude is above or below one. The exponent $\alpha \in (0, 2)$ determines the rate at which the tails of the distribution taper off. If $\alpha > 1$, it could be shown that the mean exists and it coincides with the location parameter μ. The parameter β corresponds to the skewness of the distribution. If β is positive then the distribution is skewed to the right, while if negative it is skewed to the left.

The Generalized Central Limit Theorem (GCLT) describes the inter-relationship between the domain of attraction (see Definition A.2.2) and the stable distribution (Nolan 2015, p. 22). In particular, if the identical variable of the sequence of random variables belongs to the domain of attraction of a limiting variable that follows a stable distribution, then this describes the GCLT.

Theorem A.2.7 (Generalized Central Limit Theorem). *A non-degenerate random variable Z is α-stable for some $0 < \alpha \leq 2$ if and only if there is an i.i.d. sequence of random variables $\{X_i\}_{i=1}^n$ and constants $a_n > 0$, $b_n \in \mathbb{R}$ with*

$$a_n^{-1} \sum_{i=1}^n X_i - nb_n \xrightarrow{d} Z \tag{A.19}$$

Theorem A.2.7 has an advantage over the traditional Central Limit Theorem (CLT). For example, the GCLT states that the normalized sum of random variables will converge in distribution to the domain of the normal attraction of a stable law, where the Gaussian law (thus, CLT) is only a particular case of it (when $\alpha = 2$). Consider the normalized sum of binomially distributed observations. It is well known the distribution of such sum converges to the standard normal law. Certainly, this is because the binomial random variable possesses the second moment. However, many other random variables may not enjoy either a first or a second moment such as the Cauchy random variable. The normalized sum of i.i.d. Cauchy random variables offers to epitomize the GCLT. Indeed, the normalized sum of i.i.d. Cauchy random variables converges to a stable law (not Gaussian law). The Cauchy distribution is a special case of the Stable family (see, Nolan 2015).

Having said this, it is worth noting that there is a difference between stating that distribution is stable, and distribution is asymptotically stable. For example, based on the example above, one may state that the sum of binomially distributed observations, when suitably normalized, is asymptotically normal, but one may not say it is an exact normal distribution. Similarly, some distributions are not exactly stably distributed, but in the limit of

their normalized sum are indeed stably distributed. For example, the Pareto distribution does not belong to the stable distribution family, but it belongs to the domain of normal attraction of a stable law (Politis et al. 1999, p. 262).

A.2.2 Tails of Stable Distributions:

The stable distributions, habitually, are connected with heavy tails characteristics. Indeed, Lévy (1925) has shown the asymptotic equivalence between the tail behavior of a Pareto law and non-Gaussian law distributions. Hence, in the literature, the name Paretian tail is always associated with the tail behavior of stable distributions. In his proof of Theorem A.2.7, Feller (1971) used the Lévy measure (defined in A.11) and the uniform convergence of $nF_X(a_n(dx+b_n))$ to the Lévy measure $W(dx)$[7] to show that the tails of the stable probability distribution function, F_Z, such that $Z \in \mathbb{S}(\alpha)$, follow these equations:

$$\mathbb{P}(Z > \varepsilon) = 1 - F_Z(\varepsilon) \sim c_1 \varepsilon^{-\alpha}, \quad \varepsilon \to \infty \tag{A.20}$$

$$\mathbb{P}(Z < \varepsilon) = F_Z(\varepsilon) \sim c_2 |\varepsilon|^{-\alpha}, \quad \varepsilon \to -\infty \tag{A.21}$$

where c_1, and c_2 are some positive real constants such that $c_1 \geq 0, c_2 \geq 0$ such that $c_1 + c_2 > 0$.

The two constants, c_1 and c_2, play an essential role in determining the limiting skewness parameter[8], β, in equation (A.18). Although skewness is usually defined as the measure of the asymmetry of the probability distribution of a random variable about its mean, in our context the skewness refers to the infinite portion of the right-hand tail over the infinite portion of both tails. Therefore, it has nothing to do with the mean, mode, or the median. Note that the mean does not exist when α is smaller than one (see, Lemma, Feller 1971, p. 578).

In general, the mean, which is usually considered to be a measure of central tendency, does not necessarily exist. Therefore, this poses the question of what is the central tendency of a random variable whose distribution is stable. The main goal of central tendency is to find a single value that best represents the entire distribution (Gravetter & Wallnau 2016, p. 75). In symmetric stable distribution with a finite mean, the mode, median, and the mean coincide; hence, all may be considered as the measure of central tendency of such a distribution (see,

---

[7]The uniform convergence is shown indirectly for two reasons. The first is that the spectral measure (in this case Lévy measure) W in A.7 is not being effected by the truncation function g. The second reason is that there exist a unique relationship between the Lévy measure and the canonical measure M in (equation (2.34), Feller 1971, p. 564). Moreover, the constants a_n, b_n are as discussed in equation (A.19)

[8]The skewness parameter β should not be confused with the skewness parameter of a random variable X defined by the third moment of the standardize random variable, which is known as the moment coefficient of skewness.

for example, Blattberg & Sargent 1971). However, when it comes to the general case of stable distributions, to the best of our knowledge, the central tendency of general stable distribution is not clearly defined in the academic literature. Therefore, our intuition suggests that the mode would be considered the best representative of the central tendency of the general stable distribution for two reasons: 1) The mode is unique for stable distributions (Yamazato 1978). 2) Due to the continuity of stable distributions and its well-studied structure, the mode is a representative of the data. In particular, the further an observation is away from the mode, the lower the probability associated with it becomes.

Theorem A.2.8 is critical due to its essential link between the idea of a random variable possessing a domain of normal attraction of a stable law and the asymptotic skewness parameter of the limiting stable variable. This Theorem was initially developed by Gnedenko & Kolmogorov (1968) and Doeblin (1940)

Theorem A.2.8. *A random variable X with distribution function F_X belongs to the domain of attraction of a non-degenerate stable law F_Z ($X \in \mathscr{DA}(Z)$ such that $Z \in \mathbb{S}(\alpha)$) with characteristic exponent $\alpha \in (0,2)$, if and only if, as $x \to \infty$ the following convergence is true*

$$\frac{F_X(-x)}{1 - F_X(x) + F_X(-x)} \to \frac{c_2}{c_1 + c_2} \tag{A.22}$$

for some $c_1, c_2 \geq 0$ and $c_1 + c_2 > 0$, determined by F_Z. In addition, it is required that as $x \to \infty$

$$\frac{1 - F_X(x) + F_X(-x)}{1 - F_X(tx) + F_X(-tx)} \to t^\alpha \tag{A.23}$$

for each constant $t > 0$.

Proof. Available in English at (Gnedenko & Kolmogorov 1968) and in French at (Doeblin 1940). Recently mentioned in (Theorem 4.10, Peters & Shevchenko 2015, p. 155) and (Theorem 3.2, Gut 2009, p. 194). □

Remarks:

- The results in Theorem A.2.8 are similar to the results achieved by Feller (1971) but with a different notation. Consider that $X \in \mathscr{DA}(Z)$ such that $Z \in \mathbb{S}(\alpha)$. Feller (1971) showed that

$$nF_X(a_n dx + b_n/n) \overset{n \to \infty}{\to} F_Z(dx)$$

(Theorem 2, Feller 1971, p. 564). Based on this, it may be shown using equation (A.20) that

$$nx^\alpha[1 - F_X(a_n x + b_n/n)] \overset{n \to \infty}{\to} x^\alpha[1 - F_Z(x)] \overset{x \to \infty}{\to} c_1$$

Hence, one can show that equation (A.22) holds.

- It may be shown that c_1 and c_2 in equation (A.11) can be written as:

$$c_1 = Cp\frac{2-\alpha}{\alpha} \qquad \text{and} \qquad c_2 = Cq\frac{2-\alpha}{\alpha} \tag{A.24}$$

where $C > 0$ and $p, q \in [0,1]$ such that $p + q = 1$ and α is the parameter of stability[9]. Therefore, the convergence constant in equation (A.22) can be rewritten as

$$\frac{c_2}{c_1 + c_2} = q = 1 - p$$

Moreover, the parameter β in equation (A.18) are linked to the constants p, q as follow

$$\beta = p - q = 2p - 1 = 1 - 2q$$

Lemma A.2.9. *Consider a random variable Z such that its distribution F_Z is defined as*

$$F_Z(a) = \frac{1}{2}[F_X(a) + F_X(a+1)]$$

where F_X is cdf of X such that $X \in \mathscr{DA}(E)$ where $E \in \mathbb{S}(\alpha)$. Then Z belongs to the domain of normal attraction of a stable law F_E.

Proof. Theorem A.2.8 needs to be used in the proof. Indeed, if equations (A.22) and (A.23) holds for the distribution function F_X, then Lemma A.2.9 is proved. Observe first that as $a \to \infty$

$$F_X(-a+1) \approx F_X(-a) \qquad \text{and} \qquad F_X(a+1) \approx F_X(a)$$

Hence, it can be shown that as $a \to \infty$

$$\frac{F_Z(a)}{F_X(a)} \to 1 \qquad \text{and} \qquad \frac{F_Z(-a)}{F_X(-a)} \to 1$$

Therefore, using the above, it easily follows that

$$\frac{F_Z(-a)}{1 - F_Z(a) + F_Z(-a)} \approx \frac{F_X(-a)}{1 - F_X(a) + F_X(-a)} \to q$$

---

[9]To verify this one need to use equation (5.12) in (Feller 1971, p. 576) and the fact that the Lévy measure $W(du) = u^{-2}M(du)$ where M is the canonical measure (see, equation (2.34), Feller 1971, p. 564)

Hence Z satisfies equation (A.22). Moreover, it can be shown that

$$
\frac{1 - F_Z(a) + F_Z(-a)}{1 - F_Z(ta) + F_Z(-ta)}
$$
$$
= \frac{(1 - F_X(a) + F_X(-a)) + (1 - F_X(a+1) + F_X(-a+1))}{(1 - F_X(at) + F_X(-at)) + (1 - F_X(at+1) + F_X(-at+1))}
$$
$$
\approx \frac{1 - F_X(a) + F_X(-a)}{1 - F_X(at) + F_X(-at)} \to t^\alpha
$$

Hence Z satisfies equation (A.23). $\qquad\square$

Remarks:

- One should distinguish between infinitely divisible random variables and variables that belong to the domain of the normal attraction of a stable law. Not all distributions that belong to the domain of the normal attraction of a stable law are infinitely divisible. For example, as shown in Lemma A.2.9, the distribution F_Z is not infinitely divisible, but it belongs to the domain of attraction of an infinitely divisible random variable E. In general, a linear combination of distributions that are infinitely divisible is not necessarily infinitely divisible (see, Theorem 4.2, Gut 2012, p. 443). Moreover, note that the distribution F_Z does not belong to the class of $\mathbb{L}$ distributions, since the class of $\mathbb{L}$ distributions will always be infinitely divisible (Remark 4.6 Peters & Shevchenko 2015, p. 159).

- Based on Lemma A.2.9, it may be observed that the non-unimodal random variable Z belongs to the domain of attraction of the unimodal stable random variable E. Therefore, in general, there exist distributions that are not necessarily unimodal but belong to the domain of normal attraction of a stable law.

Lemma A.2.10. *Consider a constant variable $m \in \mathbb{R}$, if $X \in \mathscr{DA}(\alpha)$ then $X - m \in \mathscr{DA}(\alpha)$*

Proof. Observe that
$$
F_{X-m}(-a) = F_X(-a+m)
$$

Therefore, as $a \to \infty$, m becomes negligible, hence, $F_X(-a+m) \sim F_X(-a)$. Thus, both equations (A.22) and (A.23) follow. Hence, the proof is complete. $\qquad\square$

A.2.3   Normalizing Constants of the Convergent Sum

It is of general interest, when addressing distributions that belong to the domain of normal attraction of a stable law, to consider their corresponding normalizing constants. Initially,

the study of the analytical expressions that the normalizing constants need to satisfy was explained in (see, equations (5.23) and (5.28), Feller 1971, p. 579-580).

Let $\{X_i\}_{i=1}^n$ be an i.i.d. sequence of random variables and denote by X a random variable from that sequence such that its distribution function F_x satisfies equations (A.22) and (A.23). Then there exist normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ where $n \in \mathbb{N}$ such that the following statistics

$$S_n = \frac{1}{a_n} \sum_{i=1}^n (X_i - b_n) \tag{A.25}$$

converge in distribution to a stable law with index α. In the case that F_x is the stable distribution $S(\alpha, \beta, \sigma, \mu)$ (see, equation (A.18)) then $a_n = n^{1/\alpha}$ such choice also holds for Pareto, Burr, and t distributions with $\alpha \in (1,2)$, but not for the log-gamma distribution with scale $v > 1$ (Cornea-Madeira & Davidson 2015, p. 451-452).

Later, using probabilistic proofs, Simons & Stout (1978) discuss the relationship between two normalizing sequences a_n and a_n', the first being the normalizing constant of distributions that belong to the domain of attraction of a stable law and the other are stable distributions that belong to the same law, then the division of these normalizing constants is slowly varying. Therefore, in general, the normalizing constant $a_n = n^{1/\alpha} L(n)$ where $L(n)$ is a slow varying function (Samoradnitsky & Taqqu 1994, p. 5) and (Theorem 6, Simons & Stout 1978, p. 309).

For the statistic S_n in equation (A.25), the normalizing constants a_n and b_n must satisfy

$$n \int_{-1}^1 y^2 F_X(a_n dy) \to C \tag{A.26}$$

where C is the same constant defined in equation (A.24) (see, equation (5.23), Feller 1971, p. 579). It may equally be shown that a_n satisfies

$$n(1 - F_X(a_n y) + F_X(-a_n y)) = n\mathbb{P}(|X| > a_n y) \to y^{-\alpha} \tag{A.27}$$

and

$$b_n = \int_{-a_n}^{a_n} y F_X(dy) \tag{A.28}$$

(see, equations (3) and (4), LePage et al. 1981, p. 624).
Remarks:

- In case X is a symmetric random variable around zero, the constant $b_n = 0$ (Feller 1971, p. 574).

- b_n converges as $n \to \infty$ to the location parameter in general. Where the location parameter is the mean of the random variable X if $\alpha > 1$.

A.2.4 On the behaviour of the central part of the Sums:

Understanding properties of partial sums of independent random variables is of critical importance, especially when addressing heavy-tail distribution. Central limit theorem implies weak convergence in probability theory; hence, establishing this property on specified statistics generates useful practical applications, especially when applying hypothesis testing procedures. Therefore, it is of interest to discuss whether a partial sum of random variables that belong to the domain of normal attraction of a stable law, when suitably scaled and normalized, validates the central limit theorem.

Let $\{X_i\}_{i=1}^n$ be an i.i.d. sequence of random variables such that X, a random variable for this sequence, belongs to the domain of normal attraction of a stable law. Then, the statistic S_n defined in equation (A.25) converges in distribution to a stable random variable with index of stability α denoted by Z_α, that is

$$S_n \xrightarrow{d} Z_\alpha \tag{A.29}$$

we are now in a good position to introduce a new perspective on the series $\{X_i\}_{i=1}^n$. Let

$$X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n} \tag{A.30}$$

and

$$|X_n^{(1)}| \ge |X_n^{(2)}| \ge \cdots \ge |X_n^{(n)}| \tag{A.31}$$

denote the *order statistics* and the ordered sample of absolute values, $\{|X_i|\}_{i=1}^n$ the arranged series according to decreasing moduli of $\{X_i\}_{i=1}^n$ respectively.

Initially, LePage et al. (1981) were the first to use properties of order statistics (see, Proposition A.1.3) to give a probabilistic proof of convergence of the normalized partial sums S_n defined in equation (A.25). Their proof enables statisticians to give a representation of the limiting distribution of the normalized sums S_n.

Theorem A.2.11 (Le Page et al. (1981)). *Let $\{X_i\}$ be a sequence of i.i.d. random variables whose common distribution, F_X, satisfies $\lim_{x\to\infty}(\mathbb{P}(X_1 > x)/\mathbb{P}(|X_1| > x)) = p \in [0,1]$ and $\mathbb{P}(|X_1| > x) = x^{-\alpha}L(x)$ where $0 < \alpha < 2$ and L is a slowly varying. $\{|X_n^{(j)}|\}_{j=1}^n$ is defined in equation (A.31). Suppose that the sequence $\{a_n > 0; n > 1\}$ satisfies $a_n = \inf\{a : n\mathbb{P}(|X_1| \ge$*

a) ≤ 1} *as* $n \to \infty^{10}$. *Then the following holds*

$$a_n^{-1}(|X_n^{(1)}|, |X_n^{(2)}|, \cdots, |X_n^{(n)}|, 0, 0, \cdots) \xrightarrow{d} (Z_1, Z_2, \cdots) \tag{A.32}$$

where $Z_k = (\sum_{i=1}^{k} E_i)^{-1/\alpha}$ *and* $\{E_i\}$ *is a sequence of i.i.d. exponential r.v.s with mean one. Moreover, the following also holds*

$$(\delta_{1,n}, \delta_{2,n}, \cdots, \delta_{n,n}, 1, 1, \cdots) \xrightarrow{d} (\delta_1, \delta_2, \cdots) \tag{A.33}$$

where $\delta_{i,n}$ *is the associated value of either* ± 1 *of the random variable* $|X_n|^{(i)}$ *whereas* $\delta_1 \sim Bin(1, -1; p)$ [11]. *In addition, the two limiting sequences of r.v.s are independent.*

As a result of Theorem A.2.11, LePage et al. (1981) were able to conclude that the largest few summands determine the limiting distribution of the statistic S_n. Here, one may ask the interesting question of whether a trimming method of the sum would lead to considerable changes in the limiting behaviour of the statistic S_n.

In fact, as pointed out by Berkes et al. (2012): "Trimming is a standard method to decrease the effect of large sample elements in statistical procedures, used, e.g., for constructing robust estimators. It is also a powerful tool in understanding deeper properties of partial sums of independent random variables." Let the trimmed sum, denoted by $\Sigma_n(r_n, s_n)$, be defined as

$$\Sigma_n(r_n, s_n) = \sum_{i=r_n+1}^{n-s_n} X_{i,n} \tag{A.34}$$

where the sequences $\{r_n, n \in \mathbb{N}\}$ and $\{s_n, n \in \mathbb{N}\}$ are non-negative integer sequences satisfying $r_n + s_n < n$. It is shown in Lemma A.2.12 that there exist some values of r_n and s_n such that the central sum, when suitably scaled and normalized, converge in probability to zero.

Lemma A.2.12 (Janssen (1989)). *Under the assumption that* $\{X_i\}_{i=1}^{n}$ *is an i.i.d. sequence of random variables such that* X_1 *belongs to the domain of normal attraction of a stable law and that it satisfies equations (A.22) and (A.29), if the constants* r_n *and* s_n *satisfies as* $n \to \infty$

$$\frac{q}{r_n} + \frac{p}{s_n} \to 0 \tag{A.35}$$

then there exists a sequence $\{d_n, n \in \mathbb{N}\}$ *of real numbers such that*

$$a_n^{-1}\Sigma_n(r_n, s_n) - d_n \xrightarrow{p} 0 \tag{A.36}$$

---

[10]Hence it also satisfies equation (A.27)

[11]A random variable X defined as $X \sim Bin(1, -1; p)$ means that X is a two-point distribution such that $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = -1) = 1 - p$

Additionally, if the constants r_n and s_n satisfy for all $n \in \mathbb{N}$

$$r_n + 1 > 1/\alpha \qquad and \qquad s_n + 1 > 1/\alpha \qquad (A.37)$$

the d_n in equation (A.36) is equal to the expectation of such a partial sum

$$d_n = a_n^{-1} \mathbb{E}(\Sigma_n(r_n, s_n))$$

Proof. Similar to Janssen (1989); however, discussion in the introduction of this Appendix (A.1) facilitate the proof. □

Remarks:

- Whenever, $q \neq 0$, if $r_n < 1/\alpha$, then the mean of $X_{r_n,n}$ does not exist. A similar argument holds whenever, $p \neq 0$, and $s_n < 1/\alpha$ the mean of $X_{n-s_n+1,n}$ does not exist.

- In case the distribution of X is one-sided, i.e. $\min(p, q) = 0$, one may show that the result of the Lemma also holds if $r_n = 0$ or $s_n = 0$.

- The constants r_n and s_n need to satisfy equation (A.35) as $n \to \infty$. Implicitly this means that as $n \to \infty$ both $r_n, s_n \to \infty$. Moreover, it can be easily shown that $r_n/n, s_n/n \to 0$ as $n \to \infty$.

- The number of terms of the ordered summands $\Sigma_n(r_n, s_n)$ defined in equation (A.34), grows as $n \to \infty$ because as $n \to \infty$

$$n - s_n - r_n = n\left(1 - \frac{s_n}{n} - \frac{r_n}{n}\right) \to \infty$$

Janssen (1989) used the result of Lemma A.2.12 to show a generalization of an important Theorem (see, Theorem 1, LePage et al. 1981, p. 627). In fact, he obtained a limiting representation for stable random variables when the largest k elements are trimmed from the sum S_n defined in equation (A.25).

Theorem A.2.13 (Janssen (1989)). *Under the same assumptions of Lemma A.2.12 it may be shown that*

$$S_n - a_n^{-1} \sum_{i=1}^{k} \delta_{i,n} |X_n^{(i)}| \overset{d}{\to} \sum_{i=1}^{\infty} (\delta_i Z_i - (p - q)\gamma_i) - \sum_{i=1}^{k} \delta_i Z_i + \gamma \qquad (A.38)$$

where S_n is defined in equation (A.25); a_n normalizing constant and the limiting values p and q are as defined in Lemma A.2.12; $\delta_{i,n}$, δ_i, and Z_i are as defined in Theorem A.2.11; γ_k is

defined as

$$
\gamma_k := \begin{cases} 0, & \text{if } k \leq 1/\alpha \\ \mathbb{E}(Z_k), & \text{Otherwise} \end{cases}
$$

the constant, which exists, γ defined as

$$
\gamma := p^{1/\alpha}\Delta_2 - q^{1/\alpha}\Delta_1 - \sum_{i=1}^{\infty}(\delta_i Z_i - (p-q)\gamma_i)
$$

where

$$
\Delta_2 := \sum_{i=1}^{\infty}(Z_i - \gamma_i) \qquad \text{and} \qquad \Delta_1 := \sum_{i=1}^{\infty}(\tilde{Z}_i - \gamma_i)
$$

where $\tilde{Z}_i$ is an independent copy of Z_i

Proof. One could write the constant b_n in equation (A.25) as

$$
nb_n = \sum_{i=1}^{k_n}d_{i,n} + \sum_{i=n+1-r_n}^{n}d_{i,n} + d_n
$$

where d_n is as defined in Lemma A.2.12 and

$$
d_{i,n} := \begin{cases} \mathbb{E}\left(F_X^{-1}\left(\frac{\Gamma_{i,n+1}}{\Gamma_{n+1,n+1}}\right)\right) & \text{if } \min(i, n+1-i) > 1/\alpha \\ 0 & \text{Otherwise} \end{cases}
$$

Then S_n could be split and ordered as

$$
a_n^{-1}\sum_{i=1}^{n}(X_i - b_n) = a_n^{-1}\left(\Sigma(0, n-k_n) - \sum_{i=1}^{k_n}d_{i,n}\right)
$$

$$
+ a_n^{-1}\left(\Sigma(k_n, r_n) - d_n\right)
$$

$$
+ a_n^{-1}\left(\Sigma(n-r_n, 0) - \sum_{i=n+1-r_n}^{n}d_{i,n}\right)
$$

by Lemma A.2.12 the second line of the above equation converges to zero in probability. Hence, the proof of the Theorem is completed by following the same proof in (Theorem 5.2, Janssen 1989, p. 221). □

Remark: Theorem 1 of LePage et al. (1981) is a special case of Theorem A.2.13 when the truncation term k is equal to zero.

Janssen (1989) has shown that the central part, $\Sigma_n(r_n, s_n)$ of the trimmed sum when normalized with the same normalizing constant, a_n of the statistic S_n in equation (A.25), then

$\Sigma_n(r_n, s_n)$ converges in probability to zero under mild assumptions on the factors r_n and s_n. Csörgő et al. (1986) showed that when a certain number of the largest and the smallest order statistics are removed, the partial sum $\Sigma_n(r_n, r_n)$ converges to a normal distribution when suitably scaled, normalized, and when the non-random variable r_n satisfies the *moderately trimmed*[12] sum assumptions. However, many other authors such as Griffin & Pruitt (1987), Teugels (1981), Berkes et al. (2011), and Berkes et al. (2012) were interested in *modulus trimmed sums*. Modulus trimmed sum is associated with the sum when the largest d elements are removed. Following the same notations as in equation (A.31), denote by

$$\widetilde{\Sigma}_n(d_n) = \sum_{i=d_n+1}^{n} X_n^{(i)} \tag{A.39}$$

the partial sum of the sum $\sum_{i=1}^{n} X_i$ where d_n elements with the largest absolute values are removed.

Initially, Teugels (1981) considered the partial sum $\widetilde{\Sigma}_n(d_n)$ with a random norming sequence. He proved that under additional assumptions on the CDF F_X and the constant d_n, the partial sum $\widetilde{\Sigma}_n(d_n)$ is asymptotically normal. Based on the same partial sum, Berkes et al. (2011) showed that under random centring but non-random norming, the central limit theorem holds for all distributions that belong to the domain of attraction of a stable law, provided that in addition to the moderately trimmed sum assumptions, d_n follows $d_n/(log(n))^{\gamma} \to \infty$ for some $\gamma > 7$. Griffin & Pruitt (1987) proved that using non-random centring and norming sequences, say B_n and A_n, the central limit theorem also holds for the partial sum $\widetilde{\Sigma}_n(d_n)$ when one of the i.i.d. summands belong to the domain of attraction of a symmetric stable law. Berkes et al. (2012) generalized the result for i.i.d. summands that belong to the domain of attraction of a stable law. In fact, Berkes et al. (2012) showed that the central limit theorem for the trimmed sum depends not on the symmetry of the CDF F_X, but on the speed of the convergence in relation A.22. Moreover, they showed that increasing the number of trimmed elements does not generally improve the CLT behaviour. This idea looks unnatural since it is believed that if more significant elements are removed from the sum, then the partial sum $\widetilde{\Sigma}_n(d_n)$ should behave "more" normally.

It is of importance in chapter 1 to stress the centring constant B_n of the partial sum. Therefore, Theorem 1.1 of Berkes et al. (2012) is restated here

---

[12]According to standard terminology, the trimming of some elements in the sum S_n in equation (A.25) to get the partial sum $\Sigma_n(r_n, r_n)$ is called light, moderate, or heavy if r_n converges to a constant, infinity with order less than n, or a proportion of n respectively as $n \to \infty$

Theorem A.2.14 (Berkes et al. (2012))**.** *Assume that* $X \in \mathscr{DA}(\alpha)$ *and that*

$$\frac{1 - F_X(x)}{1 - F_X(x) - F_X(-x)} = p, \qquad \textit{for all} \quad x \geq x_0 \tag{A.40}$$

with some $p \in [0,1]$ *and* $x_0 \geq 0$. *If* d_n *follows the moderately trimming criterion then*

$$\frac{1}{A_n}(\widetilde{\Sigma}_n(d_n) - B_n) \xrightarrow{d} N\big(0, \big(\frac{\alpha}{2 - \alpha} + (p - q)^2\big)\big) \tag{A.41}$$

where

$$A_n = d_n^{1/2}(1 - F_{|X|})^{-1}\big(\frac{d_n}{n}\big) \tag{A.42}$$

and

$$B_n = n\Big\{\int_{-x_0}^{x_0} x dF_X(x) + (p - q) \int_{d_n/n}^{1 - F_{|X|}(x_0)} (1 - F_{|X|})^{-1}(t) dt\Big\} \tag{A.43}$$

where $(1 - F_{|X|})^{-1}(t) = \inf\{x : 1 - F_{|X|}(x) \leq t : t \in (0,1)\}$

The Following are important Corollaries of Theorem A.2.14

Corollary A.2.14.1. *Assume that* $X \in \mathscr{DA}(\alpha)$ *and that*

$$\frac{1 - F_X(x)}{1 - F_X(x) - F_X(-x)} = p, \qquad \textit{for all} \quad x \geq x_0 \tag{A.44}$$

with some $p \in [0,1]$ *and* $x_0 \geq 0$. *If* d_n *follows the moderately trimming criterion then*

$$\mathbb{E}\big(X 1(x_0 < |X| \leq v_n)\big) = (p - q) \int_{1 - F_{|X|}(v_n)}^{1 - F_{|X|}(x_0)} (1 - F_{|X|})^{-1}(t) dt$$

where $F_{|X|}(t) = P(|X| \leq t)$, $(1 - F_{|X|})^{-1}(t) = \inf\{x : 1 - F_{|X|}(x) \leq t : t \in (0,1)\}$, *and,* v_n *is some constant such that it approaches infinity as* $n \to \infty$.

Proof. The proof is available implicitly in the proof of Theorem 1.1 of (Berkes et al. 2012, p. 456) where the authors have used Theorem 1 and Lemma 1 of (Csörgő et al. 1986) to show that

$$\mathbb{E}\big(\frac{1}{n}\sum_{i=1}^{n} X_i 1(x_0 < |X_i| \leq \eta_n(d_n))\big) = (p - q) \int_{d_n/n}^{1 - F_{|X|}(x_0)} (1 - F_{|X|})^{-1}(t) dt$$

where $\eta_n(d_n)$ is the d_n^{th} largest element of $\{|X_i| : i \leq n : i \in \mathbb{N}\}$. The argument then easily followed. $\square$

Corollary A.2.14.2. *Under the same conditions and notations as the previous findings, the following results hold as* $n \to \infty$

(a)

$$\frac{\mathbb{E}(|X|1(|X| \leq \eta_n(d)))}{\int_{d/n}^1 (1 - F_{|X|})^{-1}(t)dt} \to 1$$

(b)

$$\frac{\mathbb{E}(X1(|X| \leq \eta_n(d)))}{\int_{d/n}^1 (1 - F_{|X|})^{-1}(t)dt} \to (p - q)$$

(c)

$$\frac{\mathbb{E}(X1(|X| \leq \eta_n(d)))}{\int_{d/n}^1 (1 - F_{|X|})^{-1}(t)dt} \to (p - q)$$

(d)

$$\frac{VAR(|X|1(|X| \leq \eta_n(d)))}{(d/n)[(1 - F_{|X|})^{-1}(d/n)]^2} \to \frac{2}{2 - \alpha}$$

(e)

$$\frac{VAR(X1(|X| \leq \eta_n(d)))}{(d/n)[(1 - F_{|X|})^{-1}(d/n)]^2} \to \frac{2}{2 - \alpha} + (p - q)^2$$

Proof. The above could be easily concluded from equation (2.11) and (2.12) in (Berkes et al. 2012, p. 455-456). $\square$

The first part of B_n in equation (A.43), $n \int_{-x_0}^{x_0} x dF_X(x)$, is very similar to the structure of the location parameter in equation (A.28), whereas the second part of B_n represents some other parameter that centres the distribution around an axis of symmetry in the distribution. Recall that stable distributions belong to the domain of attraction of their own law. Moreover, any distribution that belongs to the domain of attraction, when suitably scaled and normalized, converges to a stable law. Hence, stable distributions satisfy the conditions for Theorem A.2.14 to hold. Hence, the second part of B_n which represents the axis of symmetry which coincides with the mode of stable distribution for many reasons which will be discussed in the subsequent subsection. Therefore, the next subsection is dedicated to give a brief introduction of the mode of stable distributions and its connection with the index of stability and skewness parameter.

A.2.5   Mode of Stable Distributions

Although it is known that all α-stable densities are unimodal (Yamazato 1978), it remains that there is no known formula for the location of the mode; however (Peters & Shevchenko 2015, p. 220). Fofack & Nolan (1999) present numerical analysis of the behaviour of the mode as a function of tail index α and skewness parameter β when the scale is unity. In

fact, they were the first to study the relationship between the mode of the stable distribution and the parameters of the standard stable law, since they acknowledge the importance of the mode in assessing risks in insurance models (Peters & Shevchenko 2015).

It is worth noting that all stable densities are unimodal, bell-shaped around the mode such that the density possesses a positive derivative to the left of the mode and a negative derivative to the right of the mode (Nolan 1997, p. 769). As Sato & Yamazato (1978) put it, a point m is called the unique mode of a distribution function $F_X(x)$ if $f_X(x)$, the density function of the random variable X, is convex on $(-\infty, m)$ and concave on $(m, +\infty)$. In particular, Fofack & Nolan (1999) mention that, given a zero-mean stable random variable $X \sim S(\alpha, \beta, \sigma, 0)$, the mode of it will be given by

$$m_1(\alpha, \beta, \sigma) = \begin{cases} \sigma m(\alpha, \beta) + \beta \sigma \tan(\frac{\alpha\pi}{2}) & \alpha \neq 1 \\ \sigma m(\alpha, \beta) + \frac{2}{\pi} \beta \sigma \log(\sigma) & \alpha = 1 \end{cases} \tag{A.45}$$

where $m(\alpha, \beta)$ is a constant that lies between 1 and -1 depending on α and β provided that σ is equal to one. Moreover, $m(\alpha, 0) = 0$ regardless of α and σ. Fofack & Nolan (1999) have also shown that the rate at which the mode converges to zero as a function of α. This is given by

$$g(\alpha) = \begin{cases} \frac{\tan(\frac{\alpha\pi}{2})\Gamma(1+2/\alpha)}{\Gamma(3/\alpha)} & \alpha \neq 1 \\ \frac{2\gamma - 3}{\pi} & \alpha = 1 \end{cases}$$

where $\gamma \approx 0.57721$ is the *Euler's Constant*

Observe that using the familiar parameterization of Samoradnitsky & Taqqu (1994)[13], the mode, m_1, may exhibit a drawback. In particular, when the index of stability α approaches one the mode m_1 would be infinite. Using another parameterization, the mode m_1 will simply be $m(\alpha, \beta)$ defined in A.45. However, since in chapter 1 we are interested in locating the mode m_1 using the parameterization of Samoradnitsky & Taqqu (1994), we have tabulated the mode m_1 for a variety of α and β with $\sigma = 1$. Observe that $m_1(\alpha, -\beta) = -m_1(\alpha, \beta)$, therefore, we shall only tabulate those values that correspond to the positive values of β.

Although Table A.1 shows that the skewness parameter is positive, the mode switches between positive and negative values depending on the value of the index of stability α. Therefore, one might be interested in understanding whether a relation between the central tendency, or location parameter, and the mode exist. On the one hand the location parameter in equation (A.18) is not influenced by the index of stability α nor by the skewness parameter β, while on the other hand the mode is influenced by both. Therefore, we believe that the

---

[13]Recall that the parameterization of Samoradnitsky & Taqqu (1994) is refer to the structural characteristic function given by equation (A.18) when describing $X \sim S(\alpha, \beta, \sigma, \mu)$

relationship between the mode and the location parameter is linear, in the sense that if the location parameter shifts by a certain number a, the mode will also be shifted by that same number a. Moreover, the mode of the general stable distribution may be computed, based on the mode of the standardized density by a simple scaling factor (Fofack & Nolan 1999, p. 50).

β					α					
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	0	0	0	0	0	0	0	0	0	0
0.1	0	0	0.0005	0.005	0.0205	0.0536	0.1142	0.2317	0.5637	−0.0585
0.2	0	0	0.001	0.0108	0.043	0.11	0.231	0.4659	1.1294	−0.1154
0.3	0	0.0001	0.0018	0.018	0.0683	0.1703	0.3524	0.7043	1.6986	−0.1694
0.4	0.0001	0.0001	0.0028	0.0264	0.0966	0.2351	0.4787	0.9477	2.2724	−0.2197
0.5	0	0.0001	0.0042	0.0364	0.1281	0.3043	0.6104	1.1964	2.8512	−0.2658
0.6	0	0.0001	0.0058	0.0479	0.1627	0.378	0.7473	1.4505	3.435	−0.3074
0.7	0.0001	0.0001	0.0079	0.0611	0.2005	0.4563	0.8893	1.71	4.0238	−0.3444
0.8	0	0.0001	0.0103	0.0758	0.2416	0.5389	1.0364	1.9746	4.6177	−0.377
0.9	0	0.0002	0.0132	0.0924	0.2859	0.6257	1.1882	2.2442	5.2163	−0.4053
1	0.0001	0.0002	0.0165	0.1106	0.3333	0.7169	1.3449	2.5187	5.8196	−0.4293

β	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
0	0	0	0	0	0	0	0	0	0	0
0.1	−0.6806	−0.3482	−0.2287	−0.1629	−0.119	−0.0864	−0.0601	−0.0379	−0.0181	0
0.2	−1.3601	−0.6955	−0.4568	−0.3255	−0.2379	−0.1725	−0.1201	−0.0757	−0.0364	0
0.3	−2.0372	−1.0414	−0.6839	−0.4874	−0.3563	−0.2586	−0.1801	−0.1136	−0.0545	0
0.4	−2.7116	−1.3852	−0.9094	−0.6484	−0.4741	−0.3442	−0.2399	−0.1514	−0.0728	0
0.5	−3.3826	−1.7262	−1.1332	−0.808	−0.5913	−0.4296	−0.2996	−0.1891	−0.0909	0
0.6	−4.0499	−2.0645	−1.355	−0.9663	−0.7075	−0.5143	−0.3589	−0.2268	−0.109	0
0.7	−4.7133	−2.3998	−1.5743	−1.1231	−0.8227	−0.5986	−0.4181	−0.2642	−0.1272	0
0.8	−5.3729	−2.7318	−1.7914	−1.2781	−0.9369	−0.6822	−0.4769	−0.3017	−0.1453	0
0.9	−6.0288	−3.0608	−2.0059	−1.4313	−1.0498	−0.7652	−0.5355	−0.3391	−0.1634	0
1	−6.681	−3.3868	−2.2181	−1.5828	−1.1616	−0.8474	−0.5937	−0.3764	−0.1816	0

Table A.1 The mode, m_1, of a nomalized stable distribution depending on the parameters α and β

Based on the previous discussion, we will introduce one main Lemma of chapter 1 as follows:

Lemma A.2.15. *A random variable whose distribution belongs to the domain of attraction of a stable law but is arbitrary and symmetrically truncated has a mean that consists of two parts: a location parameter and a stabilizing constant that centres the distribution on its axis of symmetry.*

Proof. It is known that

$$1(|X| \leq v_n) = 1(|X| \leq a_n) + 1(a_n \leq |X| \leq v_n)$$

where a_n is defined as the norming constant defined in equation (A.27). Observe that $1 - F_{|X|}(a_n) = 1/n$.

$$
\begin{aligned}
\mathbb{E}(X1(|X| \leq v_n)) &= E(X1(|X| \leq a_n)) + E(X1(a_n \leq |X| \leq v_n)) \\
&= \int_{-a_n}^{a_n} x dF_X(x) + (p-q) \int_{1-F_{|X|}(v_n)}^{1/n} (1-F_{|X|})^{-1}(t) dt
\end{aligned}
\tag{A.46}
$$

by recalling the centering constant b_n defined in equation (A.28). In this case, the first part of $E(X1(|X| \leq v_n))$ represents the location parameter as n approaches infinity. The constant v_n usually represents the d_n^{th} largest element of $\{|X_i| : i \leq n : i \in \mathbb{N}\}$ which coincides with the proof of Theorem 1 in Berkes et al. (2012) and was denoted by the expression the $\eta_n(d_n)$. However, when n goes to infinity, the value $\eta_n(d_n)$ grows, because we can only truncate up to a point d such that it is of lower order than n (by assumptions of moderate trimming). Therefore, simulating such variables, we would witness more tendency to observe more extreme variables; therefore, out of the n observations, the largest d observations are all big enough. In the discrete sense and by definition of the variable $\eta_n(d_n)$, $1 - F_{|X|}(v_n) = P(|X| \geq \eta_n(d_n))$ could be estimated as

$$\frac{1}{n} \sum_{i=1}^{n} 1(|X_i| \geq \eta_n(d_n)) = \frac{d}{n}$$

However, in the continuous sense,

$$1 - F_{|X|}(v_n) \approx d/n \to 0$$

as $n \to \infty$. In this case, observe that

$$v_n = \left(1 - F_{|X|}\right)^{-1}(d/n) \qquad \text{and} \qquad d = n\left(1 - F_{|X|}(v_n)\right) \to \infty, \qquad \text{as} \qquad n \to \infty$$

Hence, $v_n \to \infty$ such that $n\left(1 - F_{|X|}(v_n)\right) \to \infty$ as $n \to \infty$. Following the notations of Theorem A.2.14,

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{X_i 1(|X_i| \leq v_n) - E(X1(|X| \leq v_n))}{v_n(1 - F_{|X|}(v_n))^{1/2}} \xrightarrow{d} N \tag{A.47}$$

Where $N \sim \mathcal{N}(0, \frac{\alpha}{2-\alpha} + (p-q)^2)$. This implies that the second part of $E(X1(|X| \leq v_n))$ plays a central part in centring the distribution. In a sense that $f_{X1(|X| \leq v_n) - E(X1(|X| \leq v_n)) -}(x)$ is convex if $x < 0$ and concave if $x > 0$. Since all stable distributions are unimodal, then, it is believed that such centering would represent the mode of the stable distribution. $\qquad\square$

An important corollary of the above Lemma is the following

Corollary A.2.15.1. *Multiplying the series of a restricted sum of stable i.i.d. random variables by an independent Rademacher distributed random variables, would converge such a distribution to the same limiting normal distribution, using the same norming constant a_n*

Proof. In the proof of Lemma A.2.15, equation (A.47) shows that the random variable

$$\frac{X_i 1(|X_i| \leq v_n) - E(X1(|X| \leq v_n))}{v_n(1 - F_{|X|}(v_n))^{1/2}}$$

converges to a normal random variable $N \sim \mathcal{N}(0, \frac{\alpha}{2-\alpha} + (p-q)^2)$ as $n \to \infty$. Therefore, multiplying this random variable by an independent Rademacher random variable, say ω with mean zero and variance one, will not alter its asymptotic convergence, thanks to the central limit theorem already cited. $\qquad\square$

A.2.6 Estimation of Stable Parameters

Stable distributions are of great interest for applied researchers, not only in financial economics but also in astronomy, physics, biology, and electrical engineering (Akgiray & Lamoureux 1989). Therefore, many authors have studied the different approaches a practitioner may adopt to perform parameter estimation for the α-stable family. In this subsection, we do not aim to provide a comprehensive coverage of this topic, since it is beyond the scope of this thesis. However, several excellent books and articles address these issues: see, for instance Gabaix & Ibragimov (2011), McCulloch (1986), Koutrouvelis (1980, 1981), Akgiray & Lamoureux (1989), Peters & Shevchenko (2015), and Ibragimov et al. (2015), to name a few.

To estimate parameters of the stable law, Akgiray & Lamoureux (1989) examine two relatively accessible and competing estimation techniques: (a) fractile method of Fama &

Roll (1971)[14], which was improved by McCulloch (1986)[15]; (b) iterative regression method of Koutrouvelis (1980, 1981)[16].

In their study, Akgiray & Lamoureux (1989) recommend the use of the regression-type estimator of Koutrouvelis (1980, 1981), when the samples are drawn from a parent distribution of unknown origin. Moreover, Akgiray & Lamoureux (1989) add that both methods were convenient and inexpensive when implemented for this purpose. One of the most common estimates of the index of stability α of stable data is Hill's estimator. However, Hill's estimator, as well as its new competitors, suffer from several problems[17] (see, for details, Gabaix & Ibragimov 2011).

Originally, Hill (1975) suggested ordering the observations of i.i.d. random variables and computing the random estimate of α based on the largest k order statistics $Y_{1,T} > Y_{2,T}, > Y_{3,T} > \cdots > Y_{k,T}$ and it is defined as

$$\hat{\alpha}_T = (\frac{1}{k-1} \sum_{j=1}^{k-1} \log Y_{j,T} - \log Y_{k,T})^{-1}$$

where T is the sample length, $k := k(T) \to \infty$ in an understandable way. It is worth noting that Hill's (1975) proposed method accounts that there exist no parametric form for the entire distribution but only on the asymptotic tail behavior. In particular, Hill (1975) assume that the tails of the stable distribution behave only asymptotically, like a Pareto distribution. However, McCulloch (1997) mentioned that using the Hill estimator when the index of stability is between unity and two may overestimate the parameter α (see, Nolan 2001, p. 5). In fact, Fofack & Nolan (1999) see little reason to use the Hill estimator for a stable distribution and state that the location under which Pareto-like tail behavior starts to occur is highly dependent on the predefined parameterization of the stable law. Additionally, in their analysis of their numerical simulations, Fofack & Nolan (1999) mention that when they selected the parameterization such that the mode is located around or at zero, one must get notably far out on the tail before the power decay is accurate.

Therefore, in the Monte Carlo simulation experiments in section 1.3, we will follow Akgiray & Lamoureux's (1989) recommendation; hence, we will obtain initial estimates of the parameters using McCulloch's (1986) estimation method. Then, we will use the

---

[14]This method has two drawbacks: (1) It is restricted to the case of symmetric stable laws with finite means i.e. $\beta = 0$ and $\alpha > 1$. (2) Its estimators of α and σ are asymptotically bias.

[15]McCulloch (1986) generalises the Fama & Roll (1971) estimation procedure to provide consistent and unbiased estimates of all four parameters such that $\alpha > 0.5$.

[16]"The regression method of Koutrouvelis (1980, 1981) starts with an initial estimate of the parameters and proceeds iteratively until some prespecified convergence criterion is satisfied." (Akgiray & Lamoureux 1989)

[17]For example, the Hill's estimator suffers from sensitivity to dependence and small sample sizes. Furthermore, other estimation method are more complex than the Hill's estimator.

iterative regression-type method of Koutrouvelis (1980, 1981) to estimate consistently all parameters including the skewness parameter β and scale σ of a stable distribution. Note that the parameterization considered in chapter 1 is the one defined by Samoradnitsky & Taqqu (1994).

A.2.7   Generation of Stable Random Variables

There are various ways to create stable random variables due to the existence of more than one form of parameterization. Therefore, depending on the assumptions a practitioner requires in his application, it might be better to consider one form of a characteristic function over another in order to reduce complexities. As stated earlier, Fofack & Nolan (1999) used a different parameterization than that of Samoradnitsky & Taqqu (1994) because they may have better understood the bounds of the mode. However, the S_0 parameterization of Nolan (2015) is superior to that used by Fofack & Nolan (1999) when the continuity of stable parameters is in question (see, Peters & Shevchenko 2015, p. 179-180).

The Parametric Bootstrap method of Cornea-Madeira & Davidson (2015) critically depends on the estimation of the parameters α and β; hence, the continuity of both variables is at the heart of this bootstrap method. Therefore, Cornea-Madeira & Davidson (2015) generate the stable observations based on Chambers et al.'s (1976) algorithm because this algorithm generates the variables from the S_0 parameterization of Nolan (2015). In this situation, Cornea-Madeira & Davidson (2015) avoid the complexity their bootstrap method would have encountered if the parameters were not continuous. In their paper, Cornea-Madeira & Davidson (2015) considered Samoradnitsky & Taqqu (1994) parameterization. Therefore, to generate stable random variables from such parameterization, re-parameterization the algorithm of Chambers et al. (1976) is required (see, Weron & Weron 1995, p. 381).

Our proposed bootstrap method will also require estimation of stable parameters; hence, continuity of all parameters is essential. Therefore, like Cornea-Madeira & Davidson (2015), we will generate stable random variables based on Chambers et al.'s (1976) algorithm, accompanied by a re-parameterization mentioned in Weron & Weron (1995), to generate stable random variables from Samoradnitsky & Taqqu (1994) parameterization.

Appendix B

The Multivariate Hybrid Bootstrap

Consider $\mathscr{Z} = (\mathbf{z}_t : t \in \mathbb{Z})$ to be a VAR model of order p, $VAR(p)$, that is, that

$$\mathbf{z}_t = \phi_0 + \sum_{i=1}^{p} \phi_i \mathbf{z}_{t-i} + \varepsilon_t \tag{B.1}$$

where $\mathbf{z}_t$ is a $(k \times 1)$ multivariate vector time series, $\{\phi_i\}_{i=0}^{p}$ is a sequence of $(k \times k)$ matrices, and ε_t is a $(k \times 1)$ white noise vector such that $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Omega$. Observe that the process $\mathscr{Z}$ may be also represented by

$$\phi(L)\mathbf{z}_t = \phi_0 + \varepsilon_t$$

where $\phi(L) = \mathbf{I}_k - \sum_{i=1}^{p} \phi_i L^i$; and L is the lag operator. The mean of the process may be computed as

$$\mu = \mathbb{E}(\mathbf{z}_t)$$
$$= \phi(1)^{-1}\phi_0$$

Therefore, taking $\mathbf{c}_t = \mathbf{z}_t - \mu$, one may rewrite the $VAR(p)$ process $\mathscr{Z}$, represented by equation B.1, as

$$\phi(L)\mathbf{c}_t = \varepsilon_t$$

Observe that since ε_t is a white noise vector then

$$\phi(L)\Gamma_l = \begin{cases} \Omega & \text{if } l = 0, \\ \mathbf{0} & \text{if } l > 0. \end{cases}$$

where the lth cross-covariance matrix is given by $\Gamma_l = \mathbb{E}(\mathbf{z}_t - \mu)(\mathbf{z}_{t-l} - \mu)'$. The sequence of cross-covariance matrices $\{\Gamma_l\}_{l=0}^{p-1}$ is of special interest. One may rewrite the above equation as a system of matrix equation as

$$
\begin{pmatrix} \Gamma_1 & \cdots & \Gamma_p \end{pmatrix} = \begin{pmatrix} \phi_1 & \cdots & \phi_p \end{pmatrix} \begin{pmatrix} \Gamma_0 & \Gamma_1 & \Gamma_2 & \cdots & \Gamma_{p-1} \\ \Gamma_1' & \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{p-2} \\ \Gamma_2' & \Gamma_1' & \Gamma_0 & \cdots & \Gamma_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Gamma_{p-1}' & \Gamma_{p-2}' & \Gamma_{p-3}' & \cdots & \Gamma_0 \end{pmatrix} \quad \text{(B.2)}
$$

which may be equally represented by:

$$
\Gamma = \Phi \Gamma_0^*
$$

This system of matrix equations is called the *multivariate Yule–Walker equation* for *VAR(p)* models. Observe that the matrix Γ and Φ have an $(pk \times k)$ dimension. The multivariate Yule–Walker equation can be used to obtain the VAR coefficient matrices $\{\phi_j\}_{j=1}^{p}$ from the cross-covariance matrices $\{\Gamma_l\}_{l=0}^{p-1}$. For a stationary *VAR(p)* process, $(\mathbf{z}_t : t \in \mathbb{Z})$, the $(kp \times kp)$ square matrix Γ_0^*, is nonsingular. Therefore, one may equally use the *multivariate Yule–Walker equation* to obtain the cross-covariance matrices. However, before using the multivariate Yule-walker equations, it is important to be able to apply estimation procedure to a given data set.

In practice, non-parametric statisticians, assuming that the underlying generating process of the observed data originates from a VAR model, often apply some initial estimation before determining the order of the VAR process. Here, and throughout the chapter, we assume that we have observations $\{\mathbf{z}_t\}_{t=1}^{T}$ stemming from a strictly stationary *VAR(p)* process $\mathscr{Z}$. The Yule-Walker equation shall be employed on the observations $\{\mathbf{z}_t\}_{t=1}^{T}$ to obtain an initial estimate of the VAR coefficients and the Variance-Covariance matrix of the error terms Ω. To be able to exploit the *Yule-Walker equation* discussed above, a sample estimate of the cross-covariance matrix (Γ_i) for $i = 0, 1, 2, \ldots, p-1$ needs to be computed. The sample estimate of the cross-covariance matrix at lag l, Γ_l, is given by the *sample cross-covariance matrix* at lag l, defined as:

$$
\hat{\Gamma}_l = \frac{1}{T} \sum_{t=1}^{T-l} (\mathbf{z}_t - \bar{\mathbf{z}}_T)(\mathbf{z}_{t+l} - \bar{\mathbf{z}}_T)'
$$

where $\bar{\mathbf{z}}_T$ is the sample mean vector given as:

$$\bar{\mathbf{z}}_T = T^{-1} \sum_{t=1}^{T} \mathbf{z}_t$$

The estimated coefficients may be computed depending on the initial order of the sample, p, as

$$\hat{\Phi} = \hat{\Gamma} \hat{\Gamma}_0^{*-1}$$

where the matrix $\hat{\Gamma}_0^*$ is the estimate of the the $(kp \times kp)$ square matrix Γ_0^* with all the cross-covariance matrices entries replaced by the *sample cross-covariance* matrices, and the matrix $\hat{\Gamma}$ is the estimate of the $(pk \times k)$ matrix Γ defined by

$$\hat{\Gamma} = [\ \hat{\Gamma}_1' \quad \cdots \quad \hat{\Gamma}_{p-1}'\ ]' \tag{B.3}$$

The intercept estimate, $\hat{\phi}_0$, may be computed as

$$\hat{\phi}_0 = (\mathbf{I}_k - \sum_{i=1}^{p} \hat{\phi}_i) \bar{\mathbf{z}}_T$$

The estimated variance-covariance matrix of the error term, Ω, is given by

$$\hat{\Omega}_p = \hat{\Gamma}_0 - \sum_{i=1}^{p} \hat{\phi}_i \hat{\Gamma}_i'$$

All estimated matrices discussed above are based upon a predetermined VAR order p of the stationary $VAR(p)$ process $\mathscr{Z}$. However, when dealing with observations $\{\mathbf{z}_t\}_{t=1}^{T}$, nonparametric statisticians need to estimate the order p.

There have been several proposed methods in the literature regarding *order selection* of a VAR process. Following the hybrid bootstrap algorithm, the framework employed in this paper is the *Akaike information criteria* (AIC). In the time-series literature, *Information criteria* have been shown to be useful in selecting a statistical model, particularly when it comes to the VAR processes. The AIC consists of two components. The first is involved with the goodness-of-fit of the model to the data, while the other penalised more heavily complicated models. The computation of the AIC requires a prior selection of order, starting from one up to a predetermined number. In practice, the predetermined number usually does not exceed 10, due to the sophistication of the other dimensions in the VAR models; hence, in this chapter, ten will be used as the maximum number the observed series may be allowed to each as a VAR order. Then for each order selection, l, the estimated variance-covariance

matrix, $\hat{\Omega}_p$, will be computed, then the AIC will easily follow. In this chapter, the AIC will be computed as:

$$AIC(l) = \ln|\hat{\Omega}_l| + \frac{2}{T}lk^2$$

where l is the a priori selection of order $l \in I = \{0, 1, \cdots, 10\}$, $\hat{\Omega}_l$ is the maximum likelihood (ML) estimate of the matrix Ω, k is the dimension of the vector $\mathbf{z}_t$, and T is the sample size. The estimated VAR order, $\hat{p}$, which will be applied is the argument that minimizes the AIC given by:

$$\hat{p} = \underset{l \in I}{\arg\min}\{AIC(l)\}$$

Selecting the estimated VAR order, $\hat{p}$, to be the VAR order of the observations $\{\mathbf{z}_t\}_{t=1}^{T}$ is the first step in the hybrid bootstrap algorithm. The multivariate hybrid bootstrap procedure consists of the following steps:

1. Select the order of the assumed VAR model using the multivariate Yule Walker equations and the observations $\{\mathbf{z}_t\}_{t=1}^{T}$ stemming from a strictly stationary $VAR(p)$ process, $\mathcal{Z}$. Denote the estimated VAR order as $\hat{p}$.

2. Based on the estimated VAR order $\hat{p}$, estimate the VAR coefficient matrices $\{\hat{\phi}_j\}_{j=1}^{\hat{p}}$ and the covariance matrix $\hat{\Omega}_{\hat{p}}$ using the multivariate Yule Walker equations.

3. Compute the estimated residuals as:

$$\hat{\varepsilon}_t = \mathbf{c}_t - \sum_{i=1}^{\hat{p}} \hat{\phi}_i \mathbf{c}_{t-i}, t = \hat{p}+1, \ldots, T$$

where $\mathbf{c}_t = \mathbf{z}_t - \bar{\mathbf{z}}_T$, and then compute the estimated centered residuals denoted by $\tilde{\varepsilon}_t$ and mathematically represented by:

$$\tilde{\varepsilon}_t = \hat{\mathbf{L}}(\hat{p})^{-1}(\hat{\varepsilon}_t - \frac{1}{T-\hat{p}} \sum_{\tau=\hat{p}+1}^{T} \hat{\varepsilon}_\tau), \qquad \text{where} \qquad t = \hat{p}+1, \ldots, T$$

where

$$\hat{\mathbf{L}}(\hat{p})\hat{\mathbf{L}}(\hat{p})^{'} = \frac{1}{T-\hat{p}} \sum_{t=\hat{p}+1}^{T} (\hat{\varepsilon}_t - \frac{1}{T-\hat{p}} \sum_{\tau=\hat{p}+1}^{T} \hat{\varepsilon}_\tau)(\hat{\varepsilon}_t - \frac{1}{T-\hat{p}} \sum_{\tau=\hat{p}+1}^{T} \hat{\varepsilon}_\tau)^{'}$$

Hence the matrix $\hat{\mathbf{L}}(\hat{p})$ is the lower cholesky decomposition of the covariance matrix of the standardized residuals.

4. Compute the $k \times k$ spectral density of the generate bootstrap time series $\left(\mathbf{z}_t^+ : t \in \mathbb{Z}\right)$, and denote it by $\hat{\mathbf{f}}_{AR}(\omega_\kappa)$, which is given by:

$$
\hat{\mathbf{f}}_{AR}(\omega_\kappa) = \frac{1}{2\pi} \left(I_k - \sum_{k=1}^{\hat{p}} \hat{\phi}_k e^{-ik\omega_\kappa} \right)^{-1} \hat{\Omega}_{\hat{p}} \left(\left(I_k - \sum_{k=1}^{\hat{p}} \hat{\phi}_k e^{-ik\omega_\kappa} \right)^{-1} \right)^H
$$

where $\kappa = 0, \ldots, T-1$.

5. Compute the $k \times k$ Cholesky decomposition of spectral density of the generated bootstrap time series $\left(\mathbf{z}_t^+ : t \in \mathbb{Z}\right)$ in two ways which we shall denote as $\hat{\mathbf{B}}(\omega_\kappa)$ and $\hat{\mathbf{B}}(\omega_h)$, where $\kappa = -N, \ldots, N - c_1$ and $h = 1, \ldots, N$.

6. Compute the $k \times k$ periodogram matrix $\mathbf{I}_T(\omega_k)$ of the mean deviation of the original series $\{\mathbf{z}_t\}$ given by:

$$
\begin{aligned}
\mathbf{I}_T(\omega_k) &= \mathbf{J}(\omega_k)\mathbf{J}^H(\omega_k) \\
&= \frac{1}{2\pi T}\left(\sum_{t=1}^{T} \mathbf{c}_t e^{-it\omega_k}\right)\left(\sum_{t=1}^{T} \mathbf{c}_t e^{-it\omega_k}\right)^H
\end{aligned}
$$

where $k = -[T/2], 1 - [T/2], \ldots, [T/2] - c_1$, $c_1 = 1$ if T is even and zero otherwise, the series $\{\mathbf{c}_t\}$ is the mean deviation of the series of observations $\{\mathbf{z}_t\}$

7. Generate bootstrap observations $\{\mathbf{z}_t^+\}_{t=1}^{T}$ according to the following $VAR(\hat{p})$ model:

$$
\mathbf{c}_t^+ = \sum_{i=1}^{\hat{p}} \hat{\phi}_i \mathbf{c}_{t-i}^+ + \hat{\Omega}_{\hat{p}}^{1/2} \varepsilon_t^+,
$$

where $(\varepsilon_t^+ : t = 1, 2, \ldots T + m)$ is a sequence of independent draws with replacement from the sequence $(\tilde{\varepsilon}_t : t = \hat{p} + 1, \ldots, T)$ and $\hat{\Omega}_{\hat{p}}^{1/2}$ is the lower Cholesky decomposition of the matrix $\hat{\Omega}_{\hat{p}}$. The variable m is chosen to be 50, these are additional drawing so that we may select the last T generation of the series $\{\mathbf{c}_t^+\}$ to remove the start up effects since $(\mathbf{c}_t^+ = 0; t = 1, 2, \ldots, \hat{p})$. We do this step B times where each corresponds to a bootstrap draw.

8. Compute the (multivariate) discrete Fourier transform (mDFT) of the bootstrap observations $\{\mathbf{c}_t^+\}_{t=1}^{T}$, that is, that

$$
\mathbf{J}^+(\omega_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \mathbf{c}_t^+ e^{-it\omega_j}, \, j = 1, \ldots, T
$$

at the Fourier frequencies $\omega_j = 2\pi j/T$, $j = 1, 2, \ldots, [T/2]$.

9. Compute the $k \times k$ spectral density matrix of the time series $(\mathbf{z}_t^+ : t \in \mathbb{Z})$, given by

$$\hat{\mathbf{f}}_{AR}(\omega_\kappa) = \frac{1}{2\pi}(I_k - \sum_{k=1}^{\hat{p}} \hat{\phi}_k e^{-ik\omega_\kappa})^{-1}\hat{\Omega}_{\hat{p}}((I_k - \sum_{k=1}^{\hat{p}} \hat{\phi}_k e^{-ik\omega_\kappa})^{-1})^H$$

where here, and throughout the chapter, the superscript, H, represents the conjugate transpose, and the subscript $\kappa = -[T/2], 1 - [T/2], \ldots, [T/2] - c_1$.

10. Compute the Cholesky decomposition of spectral density $\hat{\mathbf{f}}_{AR}(\omega)$ at each Fourier frequency $(\omega_\kappa : \kappa = -[T/2], 1 - [T/2], \ldots, [T/2] - c_1)$, and denote it by $\hat{\mathbf{B}}(\omega_\kappa)$.

11. Repeat steps 7 and 8 but for Fourier frequency $(\omega_j : j = 1, 2, \ldots, [T/2])$

12. Compute the Cholesky decomposition $\hat{\mathbf{G}}(\omega)$ of the following equation:

$$\hat{\mathbf{G}}(\omega_j)\hat{\mathbf{G}}(\omega_j)^H = \hat{\mathbf{B}}(\omega_j)(\frac{1}{T}\sum_{k=-[T/2]}^{[T/2]-c_1} k_h(\omega_j - \omega_k)\hat{\mathbf{B}}(\omega_k)^{-1}\mathbf{I}_T(\omega_k)[\hat{\mathbf{B}}(\omega_k)^{-1}]^H)\overline{\hat{\mathbf{B}}(\omega_j)^H}$$

where $k_h()$ is known as the Bartlett-Priestley kernel bandwidth-corrected given by:

$$\begin{aligned} k_h(u) &= \frac{1}{h}k(\frac{u}{h}) \\ &= \frac{3}{2h}(1 - (\frac{u}{\pi h})^2)1(|\frac{u}{h}| \leq \pi) \end{aligned}$$

and h is the bandwidth which will be selected as $h \approx T^{-\delta}$; $\delta = 0.317$.

13. Compute the $k \times k$ nonparametric correction matrix $\tilde{\mathbf{Q}}(\omega_j)$, that is, that:

$$\tilde{\mathbf{Q}}(\omega_j) = \hat{\mathbf{G}}(\omega_j)\hat{\mathbf{B}}(\omega_j)^{-1}, j = 1, 2, \ldots, [T/2]$$

14. Compute the corrected (multivariate) discrete Fourier transform (mDFT) of the bootstrap observations $\{\mathbf{c}_t^+\}_{t=1}^T$ using:

$$\mathbf{J}^*(\omega_j) = \tilde{\mathbf{Q}}(\omega_j)\mathbf{J}^+(\omega_j), j = 1, \ldots, N+1$$

15. Complete, correct, and shift so that we observe that:

$$\mathbf{J}^*(\omega_{N+i}) = [\mathbf{J}^*(\omega_{N-i})]^H, i = 1, \ldots, N-1$$

if T is even.

16. Compute the hybrid bootstrap observations $\{\mathbf{c}_t^*\}_{t=1}^T$ which are defined as follows:

$$\mathbf{c}_t^* = \sqrt{\frac{2\pi}{T}} \sum_{j=1}^T \mathbf{J}^*(\omega_j) e^{it\omega_j}, t = 1, \ldots, T.$$