

---

# **SPECIAL TOPICS IN PROBABILISTIC EXCHANGEABILITY AND ITS APPLICATIONS**

Chun-Kai Huang

---

Submitted in fulfillment of the academic  
requirements for the degree of  
Doctor of Philosophy  
in  
Mathematical Statistics  
in the  
School of Mathematics, Statistics and Computer Science  
University of KwaZulu-Natal  
Westville Campus

August 25, 2017

# Disclaimer

This document describes work undertaken for a PhD programme of study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.



School of Mathematics, Statistics and Computer Science  
University of KwaZulu-Natal  
Westville Campus

# Declaration 1 - Supervisors

As the candidate's supervisor/co-supervisor, I agree/do not agree to the submission of this thesis.

---

Prof Delia North (supervisor)

---

Date

---

Prof Temesgen Zewotir (co-supervisor)

---

Date

# Declaration 2 - Plagiarism

I, Chun-Kai Huang, declare that

- (1) The research reported in this thesis, except where otherwise indicated, is my original research.
- (2) This thesis has not been submitted for any degree or examination at any other university.
- (3) This thesis does not contain other persons data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
- (4) This thesis does not contain other persons writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then
  - (a) their words have been re-written but the general information attributed to them has been referenced, or
  - (b) where their exact words have been used, then their writing has been placed in italics and referenced.
- (5) This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the reference sections.

---

Chun-Kai Huang

---

Date

# Declaration 3 - Publications

The following papers have been published from this thesis:

- (1) Huang, C.-K. (2014). On exchangeability and Hausdorff moment problems over  $k$ -dimensional simplexes. *South African Statistical Journal*, 48(1), 73–81.
- (2) Huang, C.-K., North, D., & Zewotir, T. (2016). A conditional extreme value approach to Value-at-Risk estimation with exchangeable innovations. *Proceedings of the 58th Annual Conference of the South African Statistical Association*, 33–40.
- (3) Huang, C.-K., North, D., & Zewotir, T. (2016). Sampling variance of a multi-phase stratified design on partially exchangeable sequences. *Proceedings of the 58th Annual Conference of the South African Statistical Association*, 41–48.
- (4) Huang, C.-K., North, D., & Zewotir, T. (2017). Exchangeability, extreme returns and Value-at-Risk forecasts. *Physica A: Statistical Mechanics and its Applications*, 477, 204–216.

---

Chun-Kai Huang

---

Date

# Acknowledgments

Firstly, I wish to thank my supervisors, Professor Delia North and Professor Temesgen Zewotir, for their continued interest and guidance throughout the process of completing this thesis. This was an especially difficult task, given that I relocated, changed employment and also welcomed two new family members during this time.

I am grateful to both the School of Mathematics, Statistics and Computer Science (University of KwaZulu-Natal), led by Professor Delia North, and the Department of Statistical Sciences (University of Cape Town), led by Professor Francesca Little, for providing space, time and resources needed in undertaking this research. Some computations were also performed using facilities provided by the University of Cape Town's ICTS High Performance Computing Team.

Gratitude also goes to the South African National Research Foundation, the South African Statistical Association, University of KwaZulu-Natal and University of Cape Town for their awards of bursaries, grants and other forms of funding that helped support and stimulate this research.

Finally, and most importantly, I would like to thank my family for their encouragement, patience and understanding throughout this study. Especially to my wife, Shy-Yin, and my children, Mikael, Oskar and Klayton, thank you for allowing me to take away those many hours that should have been spent with you!

# Abstract

This thesis evolves around a probabilistic concept called *exchangeability* and its generalised forms. It is aimed at exploring connections between exchangeability and other sub-areas in mathematical statistics. These connections include theoretical implications, generalisation of existing methodologies and applications to real-world data. There are three topics of particular interest.

The first topic is related to the linkage between de Finetti's representation theorem (for exchangeable sequences) and existence conditions for Hausdorff moment problems over  $k$ -dimensional simplexes. The equivalence of these two results are proved over the most general case in finite spaces. This is a generalisation of existing theory and uses an alternative approach to previous work in the literature. This connection, while theoretically interesting in its own right, may also lead to further cross-field applications, such as distribution re-construction from finite moments or in the approximations to finite exchangeable sequences and finite moment problems.

Secondly, we explore a currently popular topic, namely extreme value theory (EVT), which has been widely applied to areas such as hydrology, earth sciences and finance. Classical results from EVT assume that the data sequence is independent and identically distributed (IID). We generalise this assumption to exchangeable random sequences. This caters for more general approaches to EVT that allows for data dependency. Resampling techniques are utilised for estimating the parameters' prior distributions. We utilise these new methods for Value-at-Risk (VaR) estimation in financial stock returns. This is done for both cases with and without GARCH filters. These new VaR models are also compared to existing models in the literature and shows promising improvements.

For the final topic, exchangeability is applied to two-phase sampling with an auxiliary variable. In particular, our focus is on a two-phase stratified sampling design, under the assumption that readings for the study variable are exchangeable within stratum. This will again provide a generalisation from the usual IID assumption in applications of multiple-phase sampling. It is amalgamated with stationary bootstrapping at various

levels of sampling to estimate within stratum and cross strata covariances. We show that our approach provides a more conservative estimate for the sampling variance of the two-phase estimator for the mean (i.e., the ratio estimator), as compared to the conventional IID method by Rao (1973).



# Abbreviations

- $[g] = \{1, \dots, g\}$
- $\mathbb{N} = \{1, 2, \dots\}$
- $\mathbb{N}_0 = \{0, 1, \dots\}$
- $\mathbb{R}$  = The set of real numbers
- $\mathbb{P}(\dots)$  - The probability of ...
- $E(\dots)$  - The expectation of ...
- $V(\dots)$  - The variance of ...
- $Cov(\dots)$  - The covariance between ...
- $\Pi$  - The set of all finite permutations defined on  $[n]$
- ALSI - FTSE/JSE All-Share Index
- BM - Block maxima
- BS - Bootstrapping
- DPOT - Duration-based POT
- ES - Expected shortfall
- EVT - Extreme value theory
- FTSE100 - Financial Times Stock Exchange 100 Index
- GARCH - Generalised autoregressive conditional heteroskedastic
- GEVD - Generalised extreme value distribution
- GPD - Generalised Pareto distribution
- HMP - Hausdorff moment problem

- HSI - Hang Seng Index
- IID - Independent and identically distributed
- JK - Jackknife
- MLE - Maximum likelihood estimation
- MSCI - MSCI World Index
- PML - Pseudo maximum likelihood
- PORT - Peaks-over-random-threshold
- POT - Peaks-over-threshold
- S&P500 - Standard & Poor 500 Index
- SBS - Stationary bootstrapping
- SRSWOR - Simple random sampling without replacement
- VaR - Value-at-Risk

# Contents

<b>Disclaimer</b>	<b>i</b>
<b>Declaration 1 - Supervisors</b>	<b>ii</b>
<b>Declaration 2 - Plagiarism</b>	<b>iii</b>
<b>Declaration 3 - Publications</b>	<b>iv</b>
<b>Acknowledgments</b>	<b>v</b>
<b>Abstract</b>	<b>vi</b>
<b>Abbreviations</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Literature Review on Exchangeability . . . . .	2
1.2 Objectives and Contributions . . . . .	2
1.3 Chapter Summaries . . . . .	4
1.4 Miscellaneous Tools & Results . . . . .	4
<b>2 Probabilistic Exchangeability</b>	<b>8</b>
2.1 Exchangeable Events . . . . .	8
2.2 Exchangeable Random Variables . . . . .	9
2.3 De Finetti Theorems on Exchangeability . . . . .	11
2.4 De Finetti Theorems on Partial Exchangeability . . . . .	12
2.5 Further Properties of Exchangeability . . . . .	14

<b>3</b>	<b>Exchangeability and Moment Problems</b>	<b>16</b>
3.1	Introduction . . . . .	16
3.2	Hausdorff Moment Problems . . . . .	17
3.3	De Finetti's Representation for Partially Exchangeable Sequences over a Finite Domain . . . . .	19
3.4	Equivalence over a $g$ -tuple of Simplexes . . . . .	20
3.5	Extensions to General Bounded Domains . . . . .	22
3.6	Discussion and Further Research . . . . .	24
<b>4</b>	<b>Exchangeability and Extreme Value Theory</b>	<b>25</b>
4.1	Introduction . . . . .	25
4.2	Extreme Value Models . . . . .	27
4.2.1	GEVD and BM . . . . .	27
4.2.2	GPD and POT . . . . .	28
4.3	EVT for Exchangeable Sequences . . . . .	28
4.4	Parameter Prior Distributions . . . . .	29
4.5	VaR and Backtesting . . . . .	30
4.6	GARCH and VaR . . . . .	32
4.7	Empirical Results - Part I . . . . .	34
4.8	Empirical Results - Part II . . . . .	45
4.9	Discussion and Further Research . . . . .	48
<b>5</b>	<b>Exchangeability and Multi-phase Sampling</b>	<b>51</b>
5.1	Introduction . . . . .	51
5.2	Multi-phase Stratified Sampling . . . . .	52
5.3	Stratified Sampling Design with Partially Exchangeable Sequences . . . . .	53
5.4	A Numerical Example . . . . .	56
5.5	Discussion and Further Research . . . . .	58
<b>6</b>	<b>Conclusions</b>	<b>59</b>

<i>CONTENTS</i>	xii
<b>List of Tables</b>	<b>62</b>
<b>List of Figures</b>	<b>63</b>
<b>Bibliography</b>	<b>64</b>

# Chapter 1

## Introduction

Imagine a coin is tossed 10 times and the outcome of each toss is recorded. Let  $T$  denote the event of obtaining a “tail” and  $H$  is the event of obtaining a “head”. Given that the sequence of outcomes is  $HHTHTHHHTH$ , what is the probability of obtaining a “tail” in the 11<sup>th</sup> toss? If we make the assumption that our inference on this probability solely depends on the frequency of the outcomes (i.e., 7 heads and 3 tails), and not on the order of them, then we are in fact lending ourselves to the judgment of *exchangeability*.

Exchangeable sequences also arise naturally in Pólya’s urn scheme. Consider an urn that initially contains  $w$  white balls and  $b$  black balls. In each trial, a ball is randomly drawn from the urn and is then returned to the urn along with  $c$  new balls of the same colour. If we define  $X_i$  as the colour of the ball drawn in the  $i^{\text{th}}$  trial, then  $\{X_1, X_2, \dots\}$  forms an exchangeable sequence. Simple random sampling without replacement (SRSWOR) forms a special case of the above with  $c = -1$ .

Intuitively, it is easy to see that exchangeability is a generalisation from independent and identically distributed (IID) sequences. The various forms of exchangeability can be characterised by their corresponding representation theorems and they also preserve important properties such as stationarity. These make exchangeability an attractive notion to consider in applications, while the theoretical implications are interesting in their own rights.

In this thesis, we provide a short review of exchangeability and study some special topics related to the theory and applications of exchangeability. In particular, we explore connections between exchangeability and three other areas of interest, namely Hausdorff moment problems (HMPs), extreme value modelling for risk measures and multi-phase sampling with auxiliary variables.

## 1.1 Literature Review on Exchangeability

Haag (1924) was the first to formally discuss the notion of exchangeable events (sometimes also referred to as permutable, interchangeable or symmetric events). He hinted at a representation theorem but did not rigorously state or prove it. Independently, de Finetti defined and characterised exchangeable random variables in the context of personalistic probability specification (see for example de Finetti, 1930, 1974). In the papers, de Finetti also gave his famous representation theorem for the 2-valued case, which states that the joint distribution for any infinite sequence of 2-valued random variables can be represented as a mixture of IID sequences. This was soon extended to real-valued random variables (de Finetti, 1937).

De Finetti then generalised the concept of exchangeability to *partial* exchangeability, which considers a sequence of several types being exchangeable within each type (de Finetti, 1938). The natural refinements to *separately* and *jointly* exchangeable sequences were studied by Hoover (1979) and Aldous (1981). Definite treatments of de Finetti-type results for more general spaces and related notions of symmetry are given in Kallenberg (2005).

It is well known that de Finetti's theorems do not in general hold for finite sequences of exchangeable random variables. However, Diaconis (1977) and Diaconis & Freedman (1980) were able to show that the results are asymptotically true for finite sequences of exchangeable random variables that are *extendable*. They also provided an expression for the variation distance between an extendable finite exchangeable sequence to the closest mixture of IID random variables.

The symmetry (or homogeneity) possessed by exchangeability is central to its usefulness. As such, it is found in various areas of research, for both theory and applications. These include economics, population genetics, psychology, graph theory, random networks, sampling theory, etc. Overview of these developments can be found in Kingman (1978), McCall (1991) and Aldous (2010).

## 1.2 Objectives and Contributions

This study develops around exchangeability, its various generalised forms and the corresponding representation results. In particular, connections with other research fields are drawn. These include theoretical implications, extending existing methodologies and applications to real-world data. The objectives of this thesis can be summarised into the following points:

- To provide a brief review of exchangeability and the corresponding de Finetti-type representation results.
- To generalise and formalise the connection between exchangeability and HMP to the most general sense in finite countable spaces.
- To extend classical extreme value models to cater for exchangeable sequences and utilise the results for financial risk modelling.
- To explore avenues to improve estimation of sampling variance, in a two-phase design, by considering exchangeable sequences.

These bring various contributions to the present literature. In particular, the new results provided by the last three points above may be valuable to the further advancements in their related areas of research. These are summarised into the three topics below:

- The first topic is related to the linkages between de Finetti's representation theorem (for exchangeable sequences) and existence conditions for HMP. The equivalence of these two results, over  $k$ -dimensional simplexes, is proved. This is a generalisation of existing theory and uses an alternative approach to previous work in the literature. This connection, while theoretically interesting in its own right, may also lead to further cross-field applications, such as in distribution re-construction from finite moments or in representations of finite exchangeable sequences.
- Secondly, we study a currently popular topic, namely extreme value theory (EVT), which has been widely applied to areas such as hydrology, earth sciences and finance. Classical results from EVT assumes that the data sequence is IID. We generalise this assumption to exchangeable random sequences. This caters for more general approaches to EVT that allows for data dependency. We utilise these new methods for Value-at-Risk (VaR) estimation in financial stock returns. This is done for both with and without the generalised autoregressive conditional heteroskedastic (GARCH) filters. These VaR models are also compared to existing models in the literature and shows promising improvements.
- For the final topic, exchangeability is applied to multiple-phase sampling. In particular, our focus is on a two-phase stratified sampling design, where observations in the same stratum constitute an exchangeable sequence. This will again provide a generalisation from the usual IID assumption in applications of multiple-phase sampling. We show that our approach provides a more conservative estimate for



the variance of sampling estimator (i.e., the ratio estimator) than the conventional method.

### 1.3 Chapter Summaries

The rest of this thesis proceeds as follows:

Chapter 2 provides a review of the concept of exchangeability. In particular, exchangeable events and random variables are formally defined. This is also extended to partially exchangeable sequences. The corresponding de Finetti-type theorems are given.

The HMP is formulated in Chapter 3. This is stated for the case over a  $g$ -tuple of  $k$ -dimensional simplexes and the corresponding set of existence conditions is derived. Furthermore, the existence conditions are shown to be equivalent to the de Finetti's representation theorem for a  $g$ -fold partially exchangeable sequences taking on values in  $\{0, 1, \dots, k\}$ .

In Chapter 4, classical EVT results are revisited and extended to cater for exchangeable sequences. Resampling techniques are proposed to estimate the empirical prior distributions of the EVT parameters. These are implemented for financial risk modelling. In particular, daily forecasts of VaR for several market indices are calculated over a long period of time and are backtested against real observations. These are also compared to traditional approaches.

In estimating the population mean of a study variable  $y$ , we can often use a ratio-type estimator when a related auxiliary variable  $x$  is available. This is reviewed in Chapter 5 under the situation where  $x$  is qualitative and a two-phase stratified sampling design is utilised. Herein, the IID assumptions within each stratum is relaxed to the judgment of exchangeability. This allows for dependencies within each stratum. A method is proposed for estimating the variance of the ratio estimator under this scenario. An example shows that this method provides a significantly more conservative estimate for the sampling variance, as compared to the standard approach.

Chapter 6 gives an overall conclusion for this thesis. Suggestions for further research is provided along with summary of results achieved in this study.

### 1.4 Miscellaneous Tools & Results

This section reviews some basic results from probability theory, EVT and financial risk modelling that are related to discussions in the later chapters. These results, or tools,

form some of the background foundations of the methodologies used. Readers may skip and return to this section later if desired.

The two theorems below relate to the convergence of probability distributions and are needed to derive the existence conditions of the HMP.

**Theorem 1.1.** (*Helly's theorem*)

Every sequence  $\{F_n : n = 1, 2, 3, \dots\}$  of probability distributions in  $\mathbb{R}^k$  possesses a subsequence  $F_{n_1}, F_{n_2}, \dots$  that converges to a probability distribution  $F$ .

**Proof.** See Feller (1966). □

**Theorem 1.2.** (*Helly-Bray theorem*)

If  $\{F_n\}$  is a sequence of probability distributions over  $\mathbb{R}^k$  with  $F_n \rightarrow F$ , then

$$\int_I f(x) dF_n(x) \rightarrow \int_I f(x) dF(x) \quad \text{as } n \rightarrow \infty, \quad (1.1)$$

where  $I = [a, b]$  is any bounded, closed interval in  $\mathbb{R}^k$  whose boundaries contain no discontinuities of  $F$  and  $f$  is any continuous function over  $I$ .

**Proof.** See Tucker (1967). □

There are two fundamental theorems in EVT that describe the limiting behaviours of normalised block maxima and threshold exceedances. These are vital for the application of EVT and are stated below.

**Theorem 1.3.** (*Fisher-Tippett-Gnedenko theorem*)

Let  $\{X_1, X_2, \dots\}$  be a sequence of IID random variables with an unknown common distribution function  $F$ . Define  $M_n = \max\{X_1, \dots, X_n\}$  to be the (block) maximum of a sample of size  $n$ . If there exist sequences of real constants  $a_n, b_n > 0$  and a non-degenerate distribution function  $H$  such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{M_n - a_n}{b_n} \leq x \right) = H(x), \quad (1.2)$$

i.e.,  $F$  is in the maximum domain of attraction of  $H$  (written as  $F \in MDA(H)$ ), then

$$H(x) = \begin{cases} \exp\{-[1 + \xi x]^{-1/\xi}\} & , \text{ if } \xi \neq 0 \\ \exp[-\exp(-x)] & , \text{ if } \xi = 0 \end{cases}, \quad (1.3)$$

for some  $\xi$ . The three cases  $\xi < 0$ ,  $\xi = 0$  and  $\xi > 0$  are referred to as the Weibull, the Gumbel and the Fréchet distributions, respectively.

**Proof.** See Fisher & Tippett (1928) and Gnedenko (1943).  $\square$

**Theorem 1.4.** (*Pickands-Balkema-de Haan theorem*)

For  $\{X_1, X_2, \dots\}$  as given above and  $x_F$  as the right endpoint of  $F$ , define

$$F_u(x) = \mathbb{P}(X_i - u \leq x | X > u), \quad (1.4)$$

for  $0 \leq x \leq x_F - u$ , as the excess distribution above some threshold level  $u$ . Then,  $F \in \text{MDA}(H)$  if, and only if,

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x \leq x_F - u} |F_u(x) - F_{GPD}(x)| = 0, \quad (1.5)$$

where  $F_{GPD}$  is the generalised Pareto distribution (GPD) defined by

$$F_{GPD}(x|\xi, \beta) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & , \text{ if } \xi \neq 0 \\ 1 - \exp(-x/\beta) & , \text{ if } \xi = 0 \end{cases}, \quad (1.6)$$

as  $u$  tends to  $x_F$ . For this distribution,  $x > 0$  when  $\xi \geq 0$ ,  $0 \leq x \leq -\beta/\xi$  when  $\xi < 0$ , and  $\beta > 0$ .

**Proof.** See Balkema & de Haan (1974) and Pickands (1975).  $\square$

In finance, there are several techniques for evaluating the adequacy of risk models. Two of the popular tests for VaR are the Kupiec likelihood-ratio test (Kupiec, 1995) and the Christoffersen's conditional coverage test (Christoffersen, 1998). These are described in the two remarks below.

*Remark 1.5.* (Kupiec likelihood-ratio unconditional coverage test)

The Kupiec test uses the fact that an adequate VaR model should have its proportion of exceedances close to the pre-specified tail probability level (i.e., unconditional coverage). Let  $x^\alpha$  be the number of times that a sequence of VaR estimates (at level  $\alpha$ ) is exceeded by the corresponding observed values. The null hypothesis of the test is that the expected proportion of exceedances is equal to  $\alpha$  and the resulting test statistic is given by

$$\text{LR}_{\text{UC}} = 2 \ln \left( \left( \frac{x^\alpha}{n} \right)^{x^\alpha} \left( 1 - \frac{x^\alpha}{n} \right)^{n-x^\alpha} \right) - 2 \ln (\alpha^{x^\alpha} (1 - \alpha)^{n-x^\alpha}), \quad (1.7)$$

where  $n$  is the total number of VaR estimates in the sequence. This test statistic is asymptotically distributed as a chi-square distribution with one degree of freedom.

*Remark 1.6.* (Christoffersen's conditional coverage test)

The Christoffersen's test extends the Kupiec test to account for both unconditional coverage and serial independence of VaR exceedances (i.e., clustering of violations). When both of these conditions are satisfied, then the VaR model is said to have the correct conditional coverage. Define  $\alpha_0$  as the probability of observing a VaR exceedance in the currently period, given an exceedance did not occur in the previous period, and  $\alpha_1$  as the probability of observing a VaR exceedance in the currently period, given there was also an exceedance in the previous period. Under the null hypothesis of this test, we have  $\alpha_0 = \alpha_1 = \alpha$ .

Given the observations, we may estimate  $\alpha_0$  and  $\alpha_1$  with  $\hat{\alpha}_0 = x^{01}/(x^{00} + x^{01})$  and  $\hat{\alpha}_1 = x^{11}/(x^{10} + x^{11})$ , respectively, where state 0 denotes a non-exceedance of VaR, state 1 denotes an exceedance of VaR and  $x^{ij}$  denotes the number of periods for which state  $i$  is followed by state  $j$ . Hence, the test statistic can be written as

$$\text{LR}_{\text{CC}} = 2 \ln \left[ \frac{(1 - \hat{\alpha}_0)^{x^{00}} \hat{\alpha}_0^{x^{01}} (1 - \hat{\alpha}_1)^{x^{10}} \hat{\alpha}_1^{x^{11}}}{(1 - \alpha)^{x^{00} + x^{10}} \alpha^{x^{01} + x^{11}}} \right], \quad (1.8)$$

which asymptotically follows a chi-square distribution with two degrees of freedom.

## Chapter 2

# Probabilistic Exchangeability

Exchangeability is a fundamental concept in the subjective approach to probability modelling and replaces the IID concept of the objective theory. In essence, it captures the notion that future samples behave like earlier ones. This is vital in specifying a predictive model for a sequence of observations. It caters for the accumulation of information, where the underlying dependencies are encapsulated through the structure of the joint distribution of the sequence. At the same time, the probabilistic symmetries defined by exchangeability can be characterised by the corresponding de Finetti-type representation theorems. In this chapter, we review the definitions of the various forms of exchangeability and their corresponding representations.

### 2.1 Exchangeable Events

Recall our coin tossing example mentioned at the beginning of the previous chapter. It can be translated into the following formal definition:

**Definition 2.1.** (*Exchangeable events*)

A sequence of events  $\{A_1, A_2, \dots\}$  is said to be exchangeable if the probability that any  $n$  of these events occur depends only on  $n$  and not on the particular events chosen.  $\square$

Now, instead of a single coin, suppose we toss  $g$  coins. Three different cases may arise:

- (1) The coins are perfectly equal. This will generate an exchangeable sequence of events (as in the above definition).

- (2) The coins are completely different. This means each of the  $g$  coins will generate a sequence of exchangeable events, with complete independence between the sequences.
- (3) Some coins are related, i.e. the outcomes of tosses with one coin may influence probabilities with respect to tosses with other coins, but in a less direct manner than in case (1).

In other words, case (3) produces  $g$  exchangeable sequences as in case (2), but with some interdependence between the sequences. This leads to the definition of a partially exchangeable sequence of events.

**Definition 2.2.** (*Partially exchangeable events*)

A sequence of events is said to be  $g$ -fold partially exchangeable if the events split into  $g$  types and events of the same type are exchangeable, i.e. the probability that any  $n_1, n_2, \dots, n_g$  events of types  $1, 2, \dots, g$ , respectively, occur depends only on  $n_1, n_2, \dots, n_g$  and not on the particular events chosen.  $\square$

It may be noted that cases (1) and (2) are just special cases of Definition 2.2. Hence, partial exchangeability is a more general concept than that of exchangeability. It is also easy to deduce from Definition 2.1 that the probability of a singular event occurring, in an exchangeable sequence, is the same for all events in the sequence. This is similarly true for events of the same type in a  $g$ -fold partially exchangeable sequence.

## 2.2 Exchangeable Random Variables

For a sequence of random variables  $\{X_1, X_2, \dots\}$ , the uncertainty relative to some observable sequence of outcomes,  $\{x_1, \dots, x_n\}$  say, in an experiment of size  $n$ , can be determined by making use of the joint distribution function  $F(x_1, \dots, x_n)$ . If we further assume that the sequence  $\{X_1, X_2, \dots\}$  is IID, then  $F(x_1, \dots, x_n) = F(x_1) \cdots F(x_n)$  and it follows immediately that

$$F(x_{m+1}, \dots, x_n | x_1, \dots, x_m) = F(x_{m+1}, \dots, x_n), \quad (2.1)$$

for any  $1 \leq m < n$ . In other words, the predictability of future observations is not abetted by past information. This is clearly inappropriate for specifying a predictive model where we believe that the accumulation of prior observations can provide evidence for future events. Preferably, we would like the structure of the joint density function to encapsulate some form of dependence within the random sequence. One

class of possible subjective judgments is to continue allowing probabilistic symmetry to exist among the random variables. As such we define the following:

**Definition 2.3.** (*Finite exchangeability*)

A finite sequence of random variables  $\{X_1, X_2, \dots, X_n\}$  is said to be exchangeable if their joint distribution function  $F$  satisfies

$$F(x_1, x_2, \dots, x_n) = F(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}), \quad (2.2)$$

for all  $\pi \in \Pi$ , the set of all finite permutations defined on  $[n] = \{1, 2, \dots, n\}$ . In terms of the corresponding density or mass function, the condition reduces to

$$p(x_1, x_2, \dots, x_n) = p(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}), \quad (2.3)$$

where  $p(\cdot)$  is the joint density or mass function.  $\square$

**Definition 2.4.** (*Infinite exchangeability*)

An infinite sequence of random variables  $\{X_1, X_2, \dots\}$  is said to be exchangeable if all its finite subsequences are exchangeable in the sense of Definition 2.3.  $\square$

The case of exchangeability is one where there is a complete symmetry between all the random variables under consideration. However, in practice, one will often find this not to be the case. Thus, exchangeability can only be considered as a limiting case and a more general concept must be introduced, i.e. partial exchangeability.

**Definition 2.5.** (*Finite partially exchangeable random variables*)

A finite sequence of random variables,  $\{X_{ij} : i \in [g], j \in [n_i]\}$ , is said to be *g-fold partially exchangeable* if the joint distribution of the  $n_1, n_2, \dots, n_g$  random variables of types 1, 2,  $\dots$ ,  $g$  respectively, depends only on  $n_1, n_2, \dots, n_g$  and not on the order of the random variables within each type.  $\square$

**Definition 2.6.** (*Infinite partially exchangeable random variables*)

An infinite sequence of random variables,  $\{X_{ij} : i \in [g], j \in \mathbb{N}\}$ , is said to be *g-fold partially exchangeable* if all its finite  $g$ -fold subsequences of random variables, i.e.  $\{X_{ij} : i \in [g], j \in [n_i]\}$  where  $n_i$ 's are finite, are partially exchangeable in the sense of Definition 2.5.  $\square$

We have already mentioned that a sequence of exchangeable random variables can be seen as the result of an experiment in drawing balls from an urn, i.e. an urn model.

This concept can be easily extended to partially exchangeable random variables.

Suppose we have  $g$  urns labeled 1 to  $g$ , where the  $i$ -th urn contains  $n_i$  balls, for all  $i \in [g]$ . Let  $X_{ij}$  denote the result of the  $j$ -th draw from urn  $i$ , then

- (i) an infinite partially exchangeable sequence of random variables is formed by  $\{X_{ij} : i \in [g], j \in \mathbb{N}\}$  when drawing with replacement from the urns.
- (ii) a finite partially exchangeable sequence of random variables is formed by  $\{X_{ij} : i \in [g], j \in [n_i]\}$  when drawing without replacement from the urns.

*Remark 2.7.* There exist other forms of exchangeability, such as joint exchangeability and separate exchangeability (Hoover, 1979; Aldous, 1981), that are not studied in this thesis. Whereas, other classes of probabilistic symmetry (e.g. rotatability, contractability) are detailed in Kallenberg (2005).

### 2.3 De Finetti Theorems on Exchangeability

The various forms of infinite exchangeability are characterised by their corresponding representation theorems. We start by stating the simplest case.

**Theorem 2.8.** (*Representation theorem for  $\{0, 1\}$ -random variables*)

*An infinite sequence of random variables  $\{X_1, X_2, \dots\}$ , taking values in  $\{0, 1\}$ , is exchangeable if, and only if, there exists a distribution function  $F$  such that,  $\forall n \in \mathbb{N}$ ,*

$$p(x_1, x_2, \dots, x_n) = \int_0^1 \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} dF(\theta), \quad (2.4)$$

*where  $\theta$  is the probability of obtaining 1. In other words, the probability that  $k$  out of the  $n$  random variables being equal to 1 is given by*

$$\int_0^1 \binom{n}{k} \theta^k (1 - \theta)^{n-k} dF(\theta). \quad (2.5)$$

**Proof.** See Heath & Sudderth (1976). □

It is important to note that the above theorem implies that, given  $\theta$ , the random variables in an exchangeable sequence are judged to be IID (following the Bernoulli distribution). Or, equivalently, they form a mixture of IID Bernoulli random variables.



On the other hand, it is also easy to see that the set  $\{0, 1\}$  is chosen merely for convenience and the theorem applies to any two-valued space, i.e.,  $\{a, b\}$ . We can in fact extend the result to any finite countable space, in which the integrand would become a multinomial expression (to be discussed in Chapter 3). However, de Finetti-type theorems are less intuitive in the continuous cases. We shall state the result for random variables over the real line. Analogous results can be given to other cases.

**Theorem 2.9.** (*Representation theorem for random variables over  $\mathbb{R}$* )

*An infinite sequence of random variables  $\{X_1, X_2, \dots\}$ , taking values in  $\mathbb{R}$ , is exchangeable if, and only if, there exists a probability measure  $Q$  over  $\tau$  such that,  $\forall n \in \mathbb{N}$ ,*

$$F(x_1, x_2, \dots, x_n) = \int_{\tau} \prod_{i=1}^n G(x_i) dQ(G), \quad (2.6)$$

where  $\tau$  is the set of all distribution functions on  $\mathbb{R}$  and  $Q(G) = \lim_{n \rightarrow \infty} \mathbb{P}(G_n)$ , with  $G_n$  being the empirical distribution function defined by  $\{X_1, \dots, X_n\}$ .

**Proof.** See Chow & Teicher (1988). □

*Remark 2.10.* The above theorem implies that the sequence is conditionally IID (given  $G$ ) and  $Q$  represents some prior belief of what the empirical distribution for  $G$  looks like for large  $n$ . However, the task of describing such a function  $Q$  is not straightforward as  $G$  is in effect an infinite dimensional parameter. In practice, one would often restrict the result to a “nicer” space. Details of such examples can be found in Freedman (1962) and Bernardo & Smith (1994). For our purposes in this thesis, we shall state the result given in Bernardo (1996): *The assumption of exchangeability is characterised by a representation theorem which states that there exists a conditional model  $F_{X|\theta}(x|\theta)$ , where  $\theta \in \Theta$  is the limit of some function of  $x_i$ 's as  $n \rightarrow \infty$ , and a distribution  $F_{\theta}(\theta)$  such that*

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{\Theta} \prod_{i=1}^n F_{X|\theta}(x_i|\theta) dF_{\theta}(\theta), \quad (2.7)$$

where  $F_{\theta}(\theta)$  represents some prior belief for  $\theta$ .

## 2.4 De Finetti Theorems on Partial Exchangeability

We now generalise de Finetti's results to partially exchangeable sequences. Again, let us first consider the  $\{0, 1\}$  case.

**Theorem 2.11.** (*Representation for  $\{0, 1\}$ -partially exchangeable sequences*)

An infinite sequence of random variables  $\{X_{ij} : i \in [g], j \in \mathbb{N}\}$ , taking values in  $\{0, 1\}$ , is  $g$ -fold partially exchangeable if, and only if, there exists a distribution function  $F$  such that, the probability of obtaining  $m_i$  1's from  $n_i$  variables in the  $i$ -th within-type sequence, for all  $i \in [g]$ , is given by

$$\int_{[0,1]^g} \prod_{i=1}^g \binom{n_i}{m_i} \theta_i^{m_i} (1 - \theta_i)^{n_i - m_i} dF(\theta_1, \dots, \theta_g), \quad (2.8)$$

where  $\theta_i$  is the probability of obtaining 1 in the sequence  $\{X_{ij} : j \in \mathbb{N}\}$ .

**Proof.** See Bernardo & Smith (1994). □

The above theorem asserts that each of the  $g$  within-type sequences is conditionally independent and Bernoulli distributed, i.e. for a fixed  $i$ ,  $\{X_{ij} : j \in \mathbb{N}\}$  can be judged to be independent Bernoulli random variables conditioned on some random variable  $\theta_i$ . The dependence structure across the  $g$  within-type sequences is captured by the joint distribution  $F$ . Clearly, if the sequences are mutually independent, then we may write

$$dF(\theta_1, \dots, \theta_g) = dF(\theta_1)dF(\theta_2) \cdots dF(\theta_g),$$

i.e., independent prior distributions.

Again, we extend the above representation to partially exchangeable sequences over the real line. For notational simplicity, we define  $\mathbf{x}_i = (x_{i1}, \dots, x_{in_i})$  for the result below.

**Theorem 2.12.** (*Representation for several sequences in  $\mathbb{R}$* )

An infinite sequence of random variables  $\{X_{ij} : i \in [g], j \in \mathbb{N}\}$ , taking values in  $\mathbb{R}$ , is  $g$ -fold partially exchangeable if, and only if, there exists a probability measure  $Q$  over  $\tau^g$  such that, for all  $n_i$  in  $\mathbb{N}$ ,  $i \in [g]$ ,

$$F(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_g) = \int_{\tau^g} \prod_{i=1}^g \prod_{j=1}^{n_i} G_i(x_{ij}) dQ(G_1, \dots, G_g), \quad (2.9)$$

where  $Q(G_1, \dots, G_g) = \lim_{n_1, \dots, n_g \rightarrow \infty} \mathbb{P}(G_{n_1}, \dots, G_{n_g})$  and  $\tau$  the set of all distribution functions on  $\mathbb{R}$ , with  $G_{n_i}$  being the empirical distribution functions of  $\{X_{ij} : j \in [n_i]\}$  for each  $i = 1, \dots, g$ .

**Proof.** A direct generalisation of Theorem 2.9. □

## 2.5 Further Properties of Exchangeability

The understanding and applicability of exchangeability is further enhanced by its statistical properties and comparisons to other standard probabilistic assumptions. We list some of these in the following remarks.

*Remark 2.13.* It is easy to see that an IID sequence is also exchangeable. However, the converse is not always true. This makes exchangeability a more general concept than IID sequences. In fact, de Finetti's theorem characterises an exchangeable sequence as a mixture of IID sequences. Partial exchangeability is a further generalisation that accounts for exchangeability as the extreme case in which all sequences are identically distributed.

*Remark 2.14.* Random variables in an exchangeable sequence are clearly identically distributed. For example, let us consider a 2-fold partially exchangeable sequence  $\{X_{ij} : i \in [2], j \in \mathbb{N}\}$ . For each  $i \in [2]$ ,  $\{X_{ij} : j \in \mathbb{N}\}$  is an exchangeable sequence, and hence, for any  $a, b \in \mathbb{R}$ ,

$$\begin{aligned} \mathbb{P}(X_{i1} \leq a) &= \lim_{b \rightarrow \infty} \mathbb{P}(X_{i1} \leq a, X_{ij} \leq b) \\ &= \lim_{b \rightarrow \infty} \mathbb{P}(X_{ij} \leq a, X_{i1} \leq b) \\ &= \mathbb{P}(X_{ij} \leq a). \end{aligned}$$

i.e.  $\{X_{ij} : j \in \mathbb{N}\}$  are identically distributed. Similarly, one can easily show that an exchangeable sequence is stationary.

*Remark 2.15.* Random variables in an exchangeable sequence are generally correlated. For an infinite exchangeable sequence  $\{X_1, X_2, \dots\}$ , we have

$$\text{Cov}(X_i, X_j) = V(E(X_i|F_X)) = V(E(X_i|\theta)) \geq 0, \quad (2.10)$$

for all  $i \neq j$ . The conditional expectations above are taken in the usual Bayesian sense, where  $\theta$  is the random parameter of the sampling distribution of  $X$  and  $F_X$  is treated as a function of  $\theta$ . This allows for a more general treatment of many real world data, as compared to the assumption of IID sequences.

*Remark 2.16.* Any subsequence of a finite exchangeable sequence can be shown to be exchangeable. However, extensions to larger exchangeable sequences are not always possible. For example, suppose we have three random variables  $X_1, X_2, X_3$  defined

over  $\{0, 1\}$ , with

$$\begin{aligned}\mathbb{P}(X_1 = 0, X_2 = 1, X_3 = 1) &= \mathbb{P}(X_1 = 1, X_2 = 0, X_3 = 1) \\ &= \mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 0) \\ &= 1/3\end{aligned}$$

and all other combinations of  $X_1, X_2, X_3$  have probability 0. These three random variables are clearly exchangeable. However, it is easy to show that the addition of a further variable  $X_4$ , over the same domain, would make the larger sequence not exchangeable. Gnedin (1995) gave a criterion for extending a finite exchangeable sequence to an infinite one.

*Remark 2.17.* All the de Finetti-type results discussed in this chapter applies to infinite sequences and it is well-known that these representations may fail for finite sequences. However, Diaconis (1977) and Diaconis & Freedman (1980) were able to show that the results are asymptotically true for finite exchangeable sequences that are extendable to larger exchangeable sequences. The bound on the variation distance between an extendable finite sequence and a mixture of IID sequences is dependent on the proportional size between the original sequence and the extended sequence. Related applications on count data may be found in, amongst others, Diaconis & Freedman (1993), Bowman & George (1995), George & Kodell (1996) and Tan et al. (2010).

## Chapter 3

# Exchangeability and Moment Problems

The connection between two classical problems, the HMP and de Finetti's representation theorem for exchangeable random variables, is considered in this chapter. We generalise these problems to a  $g$ -tuple of  $k$ -dimensional simplexes and infinite sequences of  $g$ -fold partially exchangeable random variables in  $\{0, 1, \dots, k\}$ , respectively. The equivalence between them is then formalised and proven. The main results contained in this chapter are published in Huang (2014).

### 3.1 Introduction

Feller (1966), on page 229 of volume II, proved de Finetti's representation theorem for 2-valued exchangeable sequences via the solution to the HMP over the unit interval. He has also suggested that this method can be extended to a sequence that assumes a finite number of values. These serve as a first motivation for the current work.

Various connections between these two theories may also be found in the literature. For example, Dale (1983) gave a probabilistic proof of Hausdorff's theorem for double sequences using de Finetti's result for partially exchangeable events and Peng et al. (2010) showed that partial exchangeability can be characterised by rectangular complete monotonicity. These relations clearly suggest some kind of equivalence between the two theories and Gupta (1999b) established this link between finite HMP and finite symmetric probabilities. Further, Gupta (1999a) has established the equivalence between the HMP over a standard  $k$ -dimensional simplex and an infinite sequence of exchangeable random variables taking values in a discrete finite domain. However, to

the best of our knowledge, equivalence over the infinite case for partially exchangeable sequences has not been formally stated.

The main result of this chapter is to prove that solving the HMP over a  $g$ -tuple of  $k$ -dimensional simplexes is equivalent to the representation theorem for an infinite sequence of  $g$ -fold partially exchangeable random variables that take on values in  $\{0, 1, \dots, k\}$ . We hope that this result serves as a starting point for further generalisations to more complex domains and establishing additional cross-theoretical results.

The remainder of this chapter is structured as follows. In Section 3.2, we solve the HMP over a  $g$ -tuple of  $k$ -dimensional simplexes by generalising the method of Dale (1987). By following Aldous (1985) and Bernardo & Smith (1994), we give de Finetti's representation theorem for an infinite sequence of  $g$ -fold partially exchangeable random variables, that take on values in  $\{0, 1, \dots, k\}$ , in Section 3.3. Section 3.4 provides the main theorem that establishes the equivalence between the above results and Section 3.5 gives some discussion on further extension to general bounded domains. Conclusion and discussion for further work are provided in Section 3.6.

## 3.2 Hausdorff Moment Problems

Suppose  $\{\mu_n : n = 0, 1, \dots\}$  is a sequence of real numbers. The basic HMP is concerned with the existence of a distribution function  $F$  such that

$$\mu_n = \int_0^1 x^n dF(x), \quad n = 0, 1, \dots, \quad (3.1)$$

that is,  $\{\mu_n\}$  is the sequence of moments of some random variable  $X$  with distribution function  $F$ . Hausdorff (1923) showed that a distribution function  $F$  exists, and is unique, with the above property if, and only if, the sequence  $\{\mu_n\}$  is *completely monotonic*, that is

$$\mu_n \geq 0, \quad \mu_0 = 1 \quad \text{and} \quad \Delta^r \mu_n \geq 0, \quad n, r = 0, 1, \dots \quad (3.2)$$

where  $\Delta \mu_n = \mu_n - \mu_{n+1}$ ,  $\Delta^2 \mu_n = \Delta \Delta \mu_n = \Delta \mu_n - \Delta \mu_{n+1}$ , etc. The conditions above are called the *Hausdorff conditions*. In fact, Hausdorff showed that the conditions can be extended to moment problems over any finite interval on the real line, i.e., for intervals  $[a, b]$  such that  $-\infty < a < b < \infty$ .

Various extensions and modifications of this problem have been published since the above result. One paper of particular interest to us is by Dale (1987), where he extended the problem to the standard triangle  $\{(x, y) : x \geq 0, y \geq 0, x + y \leq 1\}$  and solved it by using Bernstein polynomials. We will generalise this method, with the addition of

Helly-Bray theorem, to solve the HMP over a  $g$ -tuple of  $k$ -dimensional simplexes. Let us define, for  $i = 1, 2, \dots$ ,

$$\mathbf{x}_i = (x_{i1}, \dots, x_{ik}), \quad \mathbf{n}_i = (n_{i1}, \dots, n_{ik}), \quad \mathbf{x}_i^{\mathbf{n}_i} = x_{i1}^{n_{i1}} \cdots x_{ik}^{n_{ik}}.$$

and

$$\Omega = \{(\mathbf{x}_1, \dots, \mathbf{x}_g) : x_{ij} \geq 0 \text{ and } \sum_{j=1}^k x_{ij} \leq 1\},$$

which represents a  $g$ -tuple of  $k$ -dimensional simplexes. Subsequently, we have the following result.

**Theorem 3.1.** (*HMP over a  $g$ -tuple of  $k$ -dimensional simplexes*)

For any given sequence  $\{\omega_{\mathbf{n}_1, \dots, \mathbf{n}_g} : \mathbf{n}_i \in [\mathbb{N}_0]^k\}$  of real numbers, there exists a distribution function  $F(\mathbf{x}_1, \dots, \mathbf{x}_g)$  on  $\Omega$  such that

$$\omega_{\mathbf{n}_1, \dots, \mathbf{n}_g} = \int_{\Omega} \mathbf{x}_1^{\mathbf{n}_1} \cdots \mathbf{x}_g^{\mathbf{n}_g} dF(\mathbf{x}_1, \dots, \mathbf{x}_g), \quad n_{ij} = 0, 1, \dots, \quad (3.3)$$

if, and only if,

$$\omega_{\mathbf{0}, \dots, \mathbf{0}} = 1 \quad \text{and} \quad \delta_1^{r_1} \cdots \delta_g^{r_g} \omega_{\mathbf{n}_1, \dots, \mathbf{n}_g} \geq 0 \quad \text{for all } \mathbf{n}_i \in [\mathbb{N}_0]^k, \quad r_i = 0, 1, \dots, \quad (3.4)$$

with  $\delta_i$  defined by

$$\delta_i \omega_{\mathbf{n}_1, \dots, \mathbf{n}_g} = \omega_{\mathbf{n}_1, \dots, \mathbf{n}_g} - \sum_{j=1}^k \omega_{\mathbf{n}_1, \dots, \mathbf{n}_i + \mathbf{1}_j, \dots, \mathbf{n}_g}, \quad (3.5)$$

where  $\mathbf{1}_j$  are  $k$ -dimensional vectors with 1 at the  $j$ -th position, 0 everywhere else, and

$$\delta_i^{r_i} \omega_{\mathbf{n}_1, \dots, \mathbf{n}_g} = \delta_i^{r_i-1} \omega_{\mathbf{n}_1, \dots, \mathbf{n}_g} - \sum_{j=1}^k \delta_i^{r_i-1} \omega_{\mathbf{n}_1, \dots, \mathbf{n}_i + \mathbf{1}_j, \dots, \mathbf{n}_g}. \quad (3.6)$$

**Proof.** (Sketches only)

( $\Rightarrow$ ) Simple.

( $\Leftarrow$ ) First, define the Bernstein polynomial  $B_{m_1, \dots, m_g}^f(\mathbf{x}_1, \dots, \mathbf{x}_g)$  over  $g$  given  $k$ -dimensional simplexes by

$$\sum_{\Gamma} f\left(\frac{\mathbf{u}_1}{m_1}, \dots, \frac{\mathbf{u}_g}{m_g}\right) \prod_{i=1}^g \binom{m_i}{u_{i1} \cdots u_{ik}} x_{i1}^{u_{i1}} \cdots x_{ik}^{u_{ik}} (1 - x_{i1} - \cdots - x_{ik})^{m_i - u_{i1} - \cdots - u_{ik}}$$

where  $\Gamma = \{(\mathbf{u}_1, \dots, \mathbf{u}_g) : u_{ij} \geq 0, \sum_{j=1}^k u_{ij} \leq m_i\}$  and  $f$  is any function defined and bounded on  $\Omega$ . Then we can show (similar to Dale, 1987) that, as  $m_1, \dots, m_g \rightarrow \infty$ ,

$$\mathbb{E}(B_{m_1, \dots, m_g}^f) \rightarrow \mathbb{E}(f). \quad (3.7)$$

Now, let us consider

$$p_{\mathbf{n}_1, \dots, \mathbf{n}_g}^{m_1, \dots, m_g} = \left[ \prod_{i=1}^g \binom{m_i}{n_{i1} \dots n_{ik}} \delta_i^{m_i - n_{i1} - \dots - n_{ik}} \right] \omega_{\mathbf{n}_1, \dots, \mathbf{n}_g}$$

and hence

$$\sum_{\Gamma} \prod_{i=1}^g \prod_{j=1}^k \binom{u_{ij}}{n_{ij}} p_{\mathbf{u}_1, \dots, \mathbf{u}_g}^{m_1, \dots, m_g} = \prod_{i=1}^g \binom{m_i}{n_{i1} \dots n_{ik}} \omega_{\mathbf{n}_1, \dots, \mathbf{n}_g} .$$

By choosing  $n_{ij} = 0$  for all  $i \in [g]$  and all  $j \in [k]$ , we have

$$\sum_{\Gamma} p_{\mathbf{u}_1, \dots, \mathbf{u}_g}^{m_1, \dots, m_g} = 1 .$$

Thus we may define random vectors  $\mathbf{X}_{1, m_1}, \dots, \mathbf{X}_{g, m_g}$ , with  $p_{\mathbf{n}_1, \dots, \mathbf{n}_g}^{m_1, \dots, m_g}$  as the joint atomic distribution, i.e.,

$$\mathbb{P}\left(\mathbf{X}_{1, m_1} = \frac{\mathbf{n}_1}{m_1}, \dots, \mathbf{X}_{g, m_g} = \frac{\mathbf{n}_g}{m_g}\right) = p_{\mathbf{n}_1, \dots, \mathbf{n}_g}^{m_1, \dots, m_g} .$$

By applying result (3.7), Helly's theorem (see Feller, 1966) and Helly-Bray theorem (see Tucker, 1967) we obtain

$$\omega_{\mathbf{n}_1, \dots, \mathbf{n}_g} = \mathbb{E}(\mathbf{X}_1^{n_1} \dots \mathbf{X}_g^{n_g}) ,$$

where  $\mathbf{X}_1, \dots, \mathbf{X}_g$  are random vectors distributed according to the limit distribution  $p_{\mathbf{n}_1, \dots, \mathbf{n}_g}$  over  $\Omega$ , when  $m_1, \dots, m_g$  tend to infinity. This completes the proof of Theorem 3.1.  $\square$

It is easy to see that, for particular choices of  $g$  and  $k$ , the above theorem reduces to HMP over the unit line, the unit square and the standard triangle. In fact, further refinements of the above result are possible, which are discussed in section 3.5.

### 3.3 De Finetti's Representation for Partially Exchangeable Sequences over a Finite Domain

In the theorem below, we extended Theorem 2.11 to an infinite sequence of  $g$ -fold partially exchangeable random variables that take on finitely many values. This extends the binomial expressions in the integrand to multinomials. In particular, we have  $(k + 1)$  possible outcomes instead of just 2. Each sequence has a corresponding set of probability parameters associated with the different possible outcomes.



**Theorem 3.2.** (*Representation for partially exchangeable sequences over a finite domain*)

An infinite sequence of random variables  $\{X_{ij} : i \in [g], j \in \mathbb{N}\}$ , taking values in  $\{0, 1, \dots, k\}$ , is  $g$ -fold partially exchangeable if, and only if, there exists a distribution function  $F$  such that the probability of observing  $n_{i1}$  1's,  $n_{i2}$  2's,  $\dots$  and  $n_{ik}$   $k$ 's out of  $m_i$  variables in the  $i$ -th within-type sequence, for all  $i \in [g]$ , is given by

$$\int_{\Omega} \prod_{i=1}^g \binom{m_i}{n_{i1} \dots n_{ik}} x_{i1}^{n_{i1}} \dots x_{ik}^{n_{ik}} (1 - \sum_j x_{ij})^{m_i - \sum_j n_{ij}} dF(\mathbf{x}_1, \dots, \mathbf{x}_g) \quad (3.8)$$

where  $x_{ij}$  is the probability of observing  $j$  in the  $i$ -th within-type sequence, denoted by  $\{X_{ij} : j \in \mathbb{N}\}$ , and  $\Omega$  is defined as before.

**Proof.** The proof is a straightforward generalisation of the 2-valued result given in Bernardo & Smith (1994), by considering  $g$  urns where each contains  $(k + 1)$  types of items, out of the total of  $m_i$  items in each urn. Subsequently, one can obtain a product of multinomial expressions over the  $g$  urns. See also Aldous (1985).

It is also easy to note that the above theorem is independent of the choice of the set  $\{0, 1, \dots, k\}$ , meaning we could have chosen any arbitrary set that contained  $(k + 1)$  points.

### 3.4 Equivalence over a $g$ -tuple of Simplexes

We now show that solving the HMP over a  $g$ -tuple of  $k$ -dimensional simplexes is equivalent to a  $g$ -fold infinite sequence of random variables, that take on any  $(k + 1)$  values, being partially exchangeable (i.e., the representation theorem for infinite sequences of partially exchangeable random variables). The proof given below is a further generalisation of Gupta (1999a), with the added notion of partial exchangeability, but the route we took is also easier in the sense that we do not introduce infinite product probabilities, nor probability on a class of probabilities. Rather, we explore the simple probabilistic implications of the given distribution.

**Theorem 3.3.** (*Equivalence of HMP and exchangeability over  $k$ -dimensional simplexes*)

An infinite sequences of random variables  $\{X_{ij} : i \in [g], j \in \mathbb{N}\}$ , taking values in  $\{0, 1, \dots, k\}$ , is  $g$ -fold partially exchangeable if, and only if, the corresponding HMP

over a  $g$ -tuple of  $k$ -dimensional simplexes has a solution, i.e., there exists a distribution function  $F$  such that, for  $\Omega$  defined as before and all  $n_{ij} \in \mathbb{N}_0$ , we have

$$\omega_{n_1, \dots, n_g} = \int_{\Omega} \mathbf{x}_1^{n_1} \cdots \mathbf{x}_g^{n_g} dF(\mathbf{x}_1, \dots, \mathbf{x}_g), \quad (3.9)$$

which is equal to the probability of observing 1 for the first  $n_{i1}$  terms, 2 for the next  $n_{i2}$  terms, ... etc., and  $k$  for the last  $n_{ik}$  terms in  $\{X_{ij} : j \in \mathbb{N}\}$ , for each  $i \in [g]$ .

**Proof.** ( $\Rightarrow$ ) Let  $\{X_{ij} : i \in [g], j \in \mathbb{N}\}$  be a set of  $g$ -fold partially exchangeable sequences of random variables taking values in  $\{0, 1, \dots, k\}$  and define  $p_{n_1, \dots, n_g, m}$  as the probability that, for each  $i$  and some  $n_{i1} + \dots + n_{ik} \leq m_i$ , the first  $n_{i1}$   $X_{ij}$ 's equal to 1, followed by  $n_{i2}$  2's, ... etc., and the last  $m_i - n_{i1} - \dots - n_{ik}$   $X_{ij}$ 's equal to 0. Also define

$$\omega_{n_1, \dots, n_g} = p_{n_1, \dots, n_g, \mathbf{n}}$$

for all  $n_{ij} = 0, 1, \dots$ , where  $\mathbf{n} = (\sum_j n_{1j}, \dots, \sum_j n_{gj})$ , and put  $\omega_{\mathbf{0}, \dots, \mathbf{0}} = 1$ . Then we get

$$\begin{aligned} p_{n_1, \dots, n_g, n+1_1} &= p_{n_1, \dots, n_g, \mathbf{n}} - p_{n_1+1_1, n_2, \dots, n_g, n+1_1} - p_{n_1+1_2, n_2, \dots, n_g, n+1_1} \\ &\quad - \cdots - p_{n_1+1_k, n_2, \dots, n_g, n+1_1} \\ &= \omega_{n_1, \dots, n_g} - \omega_{n_1+1_1, \dots, n_g} - \cdots - \omega_{n_1+1_k, \dots, n_g} \\ &= \delta_1 \omega_{n_1, \dots, n_g}, \end{aligned}$$

and similarly we can get  $p_{n_1, \dots, n_g, n+1_j} = \delta_j \omega_{n_1, \dots, n_g}$ , for  $\delta_j$  and  $\mathbf{1}_j$  defined as in Theorem 3.1. Then, by induction, we get

$$p_{n_1, \dots, n_g, n+\mathbf{r}} = \delta_1^{r_1} \delta_2^{r_2} \cdots \delta_g^{r_g} \omega_{n_1, \dots, n_g},$$

which is obviously greater than or equal to zero for any  $\mathbf{r} = (r_1, \dots, r_g) \in [\mathbb{N}_0]^g$ , i.e.  $\{\omega_{n_1, \dots, n_g}\}$  is a moment sequence (by Theorem 3.1) defined by a distribution  $F$  over  $\Omega$ , as defined in the theorem statement, such that

$$\omega_{n_1, \dots, n_g} = \int_{\Omega} \mathbf{x}_1^{n_1} \cdots \mathbf{x}_g^{n_g} dF(\mathbf{x}_1, \dots, \mathbf{x}_g).$$

( $\Leftarrow$ ) Suppose that  $\{\omega_{n_1, \dots, n_g}\}$  is a moment sequence over a  $g$ -tuple of  $k$ -dimensional simplexes, with distribution function  $F$ , then we have

$$\left[ \prod_{i=1}^g \delta_i^{m_i - n_{i1} - \cdots - n_{ik}} \right] \omega_{n_1, \dots, n_g} \geq 0,$$

for all  $\mathbf{m} \in [\mathbb{N}_0]^g$  such that  $n_{i1} + \dots + n_{ik} \leq m_i$ , and we may show

$$\sum_{n_1, \dots, n_g \in Q} \left[ \prod_{i=1}^g \delta_i^{m_i - n_{i1} - \cdots - n_{ik}} \binom{m_i}{n_{i1} \cdots n_{ik}} \right] \omega_{n_1, \dots, n_g}$$

$$\begin{aligned}
&= \sum \int_{\Omega} \left[ \prod_{i=1}^g \binom{m_i}{n_{i1} \dots n_{ik}} x_{i1}^{n_{i1}} \dots x_{ik}^{n_{ik}} (1 - \sum_j x_{ij})^{m_i - \sum_j n_{ij}} \right] dF \\
&= 1
\end{aligned}$$

for  $Q = \{(\mathbf{n}_1, \dots, \mathbf{n}_g) : n_{ij} \geq 0, \sum_j n_{ij} \leq m_i\}$ . We may then define a set of sequences  $\{X_{ij} : i \in [g], j \in \mathbb{N}\}$ , taking values in  $\{0, 1, \dots, k\}$ , such that

$$\begin{aligned}
&\mathbb{P} \left( \begin{array}{l} \text{for each } i = 1, \dots, g, \text{ a particular choice of } n_{ij} \text{ of } X_{ij}\text{'s} \\ \text{equal to } j, \text{ for all } j, \text{ from the total of } m_i \text{ } X_{ij}\text{'s} \end{array} \right) \\
&= \left[ \prod_{i=1}^g \delta_i^{m_i - n_{i1} - \dots - n_{ik}} \right] \omega_{\mathbf{n}_1, \dots, \mathbf{n}_g}
\end{aligned}$$

and thus

$$\begin{aligned}
&\mathbb{P} \left( \begin{array}{l} \text{for each } i = 1, \dots, g, \text{ any choice of } n_{ij} \text{ of } X_{ij}\text{'s} \\ \text{equal to } j, \text{ for all } j, \text{ from the total of } m_i \text{ } X_{ij}\text{'s} \end{array} \right) \\
&= \left[ \prod_{i=1}^g \delta_i^{m_i - n_{i1} - \dots - n_{ik}} \binom{m_i}{n_{i1} \dots n_{ik}} \right] \omega_{\mathbf{n}_1, \dots, \mathbf{n}_g}.
\end{aligned}$$

Then, clearly,  $\{X_{ij} : i \in [g], j \in \mathbb{N}\}$  is  $g$ -fold partially exchangeable sequence. This completes the proof for Theorem 3.3.  $\square$

The above result clearly covers all particular cases of infinite exchangeable, and partially exchangeable, sequences of random variables that take on finitely many values. Figure 3.1 provides a summary linking the different cases of HMP to the corresponding equivalent cases in exchangeability.

### 3.5 Extensions to General Bounded Domains

Stockbridge (2003) provided the solutions to HMP over polytopes

$$B = \{(\theta_1, \dots, \theta_k) : a_{h1}\theta_1 + \dots + a_{hk}\theta_k \leq a_{h0}, h = 1, \dots, g\}$$

and more general bounded regions defined by

$$C = \{(\theta_1, \dots, \theta_k) : a_{h1}\theta_1^{b_{h1}} + \dots + a_{hk}\theta_k^{b_{hk}} \leq a_{h0}, h = 1, \dots, g\}.$$

We can deal with these cases through defining the increments and the  $p$  terms differently, then we may obtain the solutions in similar ways as in Theorem 3.3. For example, in the basic case of the HMP over  $[0, b]$ , we define the increments by

$$\Delta\mu_n = b\mu_n - \mu_{n+1}$$

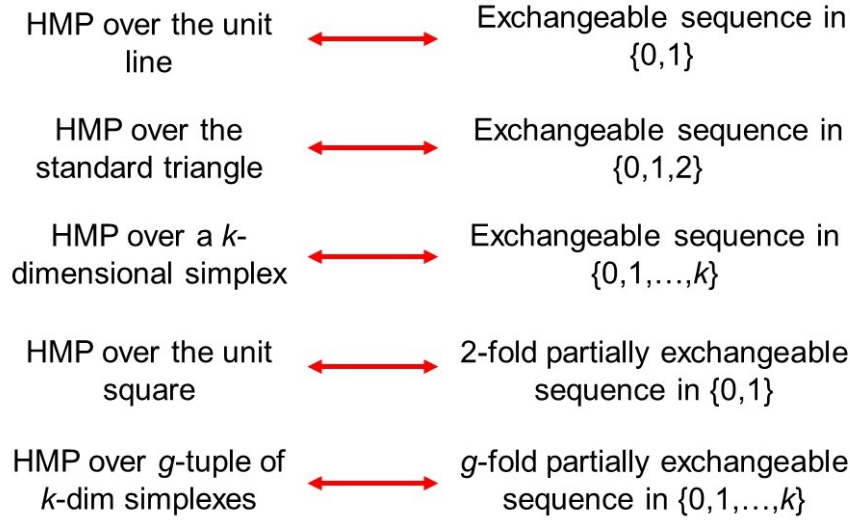


Figure 3.1: Corresponding equivalent sub-cases between HMP and exchangeable sequences.

which will eventually lead to

$$\sum_{k=0}^n \binom{n}{k} \left(\frac{1}{b}\right)^n \Delta^{n-k} \mu_k = 1 .$$

So we can define an exchangeable sequence  $X_1, X_2, \dots$  in  $\{0, 1\}$  by

$$\mathbb{P}(k \text{ out of } n \text{ } X_i\text{'s are equal to } 1) = \binom{n}{k} \left(\frac{1}{b}\right)^n \Delta^{n-k} \mu_k .$$

Conversely, define

$$p_{k,n} = b^n \mathbb{P}(X_1 = 1, \dots, X_k = 1, X_{k+1} = 0, \dots, X_n = 0)$$

and

$$\mu_n = p_{n,n} = \mathbb{P}(X_1 = 1, \dots, X_n = 1)$$

then we get

$$p_{n-1,n} = b p_{n-1,n-1} - p_{n,n} = \Delta \mu_{n-1}$$

and hence

$$p_{n-2,n} = b p_{n-2,n-1} - p_{n-1,n} = \Delta^2 \mu_{n-2} .$$

Therefore, by induction, we have for all  $k < n$

$$p_{k,n} = b p_{k,n-1} - p_{k+1,n} = \Delta^{n-k} \mu_k .$$

So, this means that we obtain the same result for HMP over  $[0, b]$  as for HMP over  $[0, 1]$ . A similar approach can be applied to HMP over polytopes and more general bounded regions defined above.

### 3.6 Discussion and Further Research

Finite HMPs refer to cases where only a finite number of moments are available. Likewise, one may consider a sequence of only finitely many exchangeable random variables taking values in a finite discrete domain. Equivalence under these cases was dealt with by Gupta (1999b). In particular, he established the links between finite Hausdorff moment problems and finite symmetric probabilities. However, a possible interest for future research may be to look at translating the various approximation procedures for finite moment problems (such as Talenti, 1987; Inglese, 1995) to finite exchangeable sequences, and vice-versa.

Extensions to exchangeable sequences that take on values in a continuous domain are much more complicated. This is due to the fact that the corresponding representation theorem result in an integral defined over the set of all distribution functions in the domain of interest and a probability measure over the set of all probability measures in the same domain (see, for example, Aldous, 1985). Insights into the extensions for such cases would be of interest for further exploration.

It is also somewhat obvious that we have been dealing with HMP over bounded regions because we are, in essence, looking for distribution functions and probability measures which are clearly bounded. In theory, one may consider general measures that need not be bounded and this could relate to more general types of moment problems over unbounded domains, for example, the Stieltjes moment problem over  $[0, \infty)$  and the Hamburger moment problem over  $(-\infty, \infty)$ . It would be interesting to see how the two concepts of exchangeability and moment problems may connect or disconnect under these cases. These are left, perhaps, for future research.

## Chapter 4

# Exchangeability and Extreme Value Theory

In this chapter, we propose new approaches to extreme value modelling for the forecasting of VaR. In particular, the block maxima (BM) and the peaks-over-threshold (POT) methods are generalised to cater for exchangeable random sequences. This caters for the dependencies, such as serial autocorrelation, of financial returns observed empirically. In addition, these approaches allow for parameter variations within each VaR estimation window. Empirical prior distributions of the extreme value parameters are attained by using resampling procedures. We compare the results of our VaR forecasts to that of the unconditional EVT approach and the conditional GARCH-EVT model for robust conclusions. As a further exploration, we also extend the GARCH-EVT model under the assumption of exchangeable innovations. The content of this chapter is included in the papers Huang et al. (2016a) and Huang et al. (2017).

### 4.1 Introduction

VaR is a commonly used benchmark for quantifying financial risk and is intended to measure the maximum possible loss of a portfolio over a specified time horizon. VaR has also prevailed as an important and widely used risk measure since the occurrence of numerous noteworthy risk management failures in the early 1990s. Most computations of VaR critically depend on an underlying distributional assumption and focus mainly on the tail behaviours (Jorion, 2006). Consequently, the selection of an appropriate distribution, or a related methodology, to accurately reflect the behaviour of financial returns has become a vital topic of research over the past two decades.

It has been well documented that empirical distribution of financial returns contradicts the classical Gaussian assumptions. For instance, Fama (1965) showed that extreme movements in financial returns emerge more frequently than estimated by Gaussian models (i.e., they exhibit heavy-tails). Aas & Haff (2006) also showed that asset returns data often exhibit skewness in distribution, with dissimilar tail behaviours. Further stylised facts, such as volatility clustering and long range dependency, are also discussed by Tsay (2010). Hence, the conjecture of a potential distribution for financial returns must be able to capture such properties in order to obtain accurate VaR estimates.

EVT has emerged as a suitable candidate for modelling VaR as it can account for both heavy-tails and skewness. Consequently, it has been extensively applied to model tail probabilities in financial returns. Koedijk et al. (1990) was among the first to apply EVT to the financial framework, by using the methods to study fat-tail behaviours in foreign exchange rate returns. Shortly thereafter, Jansen & de Vries (1991) used EVT to generate robust probabilities for large returns on share prices. Ho et al. (2000) and Longin (2000) estimated VaR using extreme value distributions, while McNeil & Frey (2000) further combined EVT innovations with the GARCH models for evaluating VaR and expected shortfall (ES). Recent work on financial applications of EVT to emerging markets include, amongst others, those by Gencay & Selcuk (2004), Huang et al. (2014) and Chinhamu et al. (2015).

The generalised extreme value distribution (GEVD) and the GPD arise as limiting distributions of BM and threshold exceedances in a sequence of IID random variables, respectively (Coles, 2001). However, the IID assumption for financial variables is largely debatable, as may be traced back to earlier studies, such as King (1966). EVT have subsequently been amalgamated with GARCH modelling in an attempt to overcome such a shortfall. In particular, GARCH models can be utilised to filter out some volatility dependence in the financial data and, subsequently, EVT can be applied to the (near) IID residuals (see, for instance, McNeil & Frey, 2000; Byström, 2004; Zhao et al., 2011). However, the appropriateness in assuming IID residuals is also dependent on how well the data series is depicted by the GARCH model. Further interesting approaches are proposed by Chavez-Demoulin et al. (2014) and de Haan et al. (2016). The former proposed a nonparametric Bayesian smoothing approach for allowing time-varying POT parameters while the latter studied the extreme value index under  $\beta$ -mixing conditions.

As a different way to circumvent the assumption of IID, we propose to generalise the EVT methods to exchangeable sequences. Some related theoretical aspects of EVT on exchangeable sequences were dealt with in Galambos (1987). Hill (1994)

also discussed the forecasting of extreme values in an exchangeable sequence using a Bayesian approach, with a modified Hills estimator. However, to the best of our knowledge, very limited applications of exchangeability exist for financial risk modelling via EVT.

## 4.2 Extreme Value Models

The BM and the POT methods are two fundamental techniques for identifying extremes in EVT. The former focuses on the distribution of block maxima, which can be modelled by GEVD. The latter identifies realised exceedances, over a predefined high threshold, which can be described by GPD (Coles, 2001). These results are reviewed below, with discussions on how they are implemented in practice.

### 4.2.1 GEVD and BM

Let  $\{X_1, X_2, \dots\}$  be a sequence of IID random variables with common (but unknown) distribution function  $F$  and let  $M_n$  denote the maximum of  $\{X_1, \dots, X_n\}$ , for a sample of size  $n$ . Intermediately, we can write down the distribution of  $M_n$ , in terms of  $F$ ,

$$\begin{aligned} F_{M_n}(x) &= \mathbb{P}(M_n \leq x) \\ &= \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \\ &= [F(x)]^n. \end{aligned} \tag{4.1}$$

Naturally, this expression still depends on the unknown distribution function  $F$ . However, Fisher & Tippett (1928) and Gnedenko (1943) have shown that, regardless of the form of  $F$  (as long as  $F \in \text{MDA}(H)$ ), the asymptotic distribution of properly normalised  $M_n$ , as  $n$  tends to infinity, is given by (see Theorem 1.3)

$$H(x) = \begin{cases} \exp\{-[1 + \xi x]^{-1/\xi}\} & , \text{ if } \xi \neq 0 \\ \exp[-\exp(-x)] & , \text{ if } \xi = 0 \end{cases} . \tag{4.2}$$

In practice, one cannot identify the normalising constants, since  $F$  is unknown. Hence, one can alternatively (see, for example, Longin, 1996; McNeil, 1998; Bali, 2003; Byström, 2004; Gilli & Këllezzi, 2006; Faranda et al., 2011) fit the sequence of maxima to the 3-parameter GEVD, given by

$$F_{GEVD}(x|\xi, \sigma, \mu) = \begin{cases} \exp\{-[1 + \xi(\frac{x-\mu}{\sigma})]^{-1/\xi}\} & , \text{ if } \xi \neq 0 \\ \exp[-\exp(-\frac{x-\mu}{\sigma})] & , \text{ if } \xi = 0 \end{cases} , \tag{4.3}$$

with  $\sigma > 0$  and  $1 + \xi(x - \mu)/\sigma > 0$ , where  $\mu$  is the location parameter,  $\sigma$  is the scale parameter and  $\xi$  is the shape parameter. This is a unified representation of the heavy-tailed Frechet distributions ( $\xi > 0$ ), the short-tailed Weibull class of distributions ( $\xi <$



0) and the light-tailed Gumbel class of distributions ( $\xi = 0$ ). In practical applications, we would divide the data into non-overlapping blocks (of some pre-specified size) and identify the maximum in each block. Subsequently, the series of BM is utilised in maximum likelihood estimation (MLE) to find parameter estimates for  $\mu$ ,  $\sigma$  and  $\xi$ . This is referred to as the BM method and enables one to approximate extreme statistics in the sequence (Coles, 2001).

### 4.2.2 GPD and POT

For the POT method, we assume  $\{X_1, X_2, \dots\}$  as above, i.e., an IID sequence of random variables with common distribution function  $F$ . Suppose  $u$  is some predetermined high threshold value for  $F$ . Conditional on  $X_i$  being observed in excess of  $u$ , we can express the probability of the magnitude of exceedance above  $u$  as

$$\begin{aligned} F_u(x) &= \mathbb{P}(X_i - u \leq x | X_i > u) \\ &= \frac{F(x + u) - F(u)}{1 - F(u)}. \end{aligned} \tag{4.4}$$

Balkema & de Haan (1974) and Pickands (1975) identifies  $F_u(x)$  asymptotically (see Theorem 1.4) with GPD, i.e.,

$$F_{GPD}(x|\xi, \beta) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & , \text{ if } \xi \neq 0 \\ 1 - \exp(-x/\beta) & , \text{ if } \xi = 0 \end{cases}, \tag{4.5}$$

as  $u$  tends to the right end point of  $X_i$ . For this distribution,  $x > 0$  when  $\xi \geq 0$ ,  $0 \leq x \leq -\beta/\xi$  when  $\xi < 0$ , and  $\beta > 0$ . To estimate the parameters, we first choose a threshold level  $u$ , then identify those values that lie above  $u$  and calculate  $x - u$ , the exceedances. Subsequently, MLE is implemented using these exceedances and estimates for  $\beta$ , and  $\xi$ , are obtained (Coles, 2001).

## 4.3 EVT for Exchangeable Sequences

Now suppose we are interested in the maximum of a subset of an exchangeable sequence and define  $M_n$  as before. According to Remark 2.10 and Remark 2.17, we can write

$$\begin{aligned} F_{M_n}(x) &= \mathbb{P}(M_n \leq x) \\ &= F(x, \dots, x) \\ &= \int_{\Theta} [F_{X|\theta}(x|\theta)]^n dF_{\theta}(\theta). \end{aligned} \tag{4.6}$$

Given the fact that  $\{X_1|\theta, X_2|\theta, \dots\}$  are IID random variables (conditional on  $\theta$ ), it is easily seen, using Lebesgue's dominated convergence theorem and standard EVT results, that

$$\begin{aligned} F_{M_n}(x) &\approx \lim_{n \rightarrow \infty} \int_{\Theta} [F_{X|\theta}(x|\theta)]^n dF_{\theta}(\theta) \\ &\approx \int_{\xi, \sigma, \mu} F_{GEVD}(x|\xi, \sigma, \mu) dF_{\xi, \sigma, \mu}(\xi, \sigma, \mu), \end{aligned} \quad (4.7)$$

for large  $n$  and for some joint prior distribution  $F_{\xi, \sigma, \mu}(\xi, \sigma, \mu)$  of  $\xi$ ,  $\sigma$  and  $\mu$  (as parameters of GEVD). Similarly, given the facts that exchangeability is preserved under location shifts and any subsequence of an exchangeable sequence is also exchangeable, we may express the distribution of exceedances above a large threshold  $u$  as

$$\begin{aligned} F_u(x) &= \int_{\Theta} \mathbb{P}(X - u \leq x | X > u, \theta) dF_{\theta}(\theta) \\ &\approx \int_{\xi, \beta} F_{GPD}(x|\xi, \beta) dF_{\xi, \beta}(\xi, \beta). \end{aligned} \quad (4.8)$$

where the POT method is applied to the sequence of exceedances  $X_i - u$ , which is IID given  $\theta$ .

These results imply that the distribution of BM and threshold exceedances, for an exchangeable sequence, can be approximated by uncountable mixtures (or, compound distributions) of the GEVD and GPD family, respectively. They also provide a motivation for a conditional approach in estimating the distribution of BM and threshold exceedances, where variations in the parameters are accounted for. The only hurdle now is to specify our beliefs of the prior distributions  $F_{\xi, \sigma, \mu}(\xi, \sigma, \mu)$  and  $F_{\xi, \beta}(\xi, \beta)$ .

## 4.4 Parameter Prior Distributions

To obtain the empirical prior distributions of GEVD and GPD parameters, we utilise three different resampling procedures. In particular, we make use of the standard IID bootstrapping (BS), stationary bootstrapping (SBS) and a jackknife (JK) procedure adapted for extreme values.

The SBS, introduced by Politis & Romano (1994), is a generalisation of the standard BS procedure. The method is as follows. A value  $p \in (0, 1]$  is predefined, which is optimally taken to be  $c^{-1}n^{-1/3}$  for a data set of size  $n^1$ . When deciding whether

---

<sup>1</sup>The SBS procedure described here is implemented by the R function `tsbootstrap`, which by default defines  $c = 3.15$ . This value is obtained through Monte Carlo simulation. An automatic block-length selection procedure has been suggested by Politis & White (2004) and Patton et al. (2009), which can be implemented by the function `b.star` in the R package `np`. This has been executed over the

an observation is to be included in a block, a number  $u$  is randomly drawn from the  $UNIF(0, 1)$  distribution. If  $u$  is less than  $1 - p$ , we include the observation into the current block. If  $u$  is greater than  $1 - p$ , then a new block is started. This algorithm is continued until all the observations have been selected into blocks. Hence, the block length is a random variable following a geometric distribution with parameter  $p$  (i.e., mean block length is  $1/p = cn^{1/3}$ ). It also consequently renders the number of blocks as a random variable. The blocks are then resampled with replacement to form new samples. This procedure caters for dependency in the data set while also preserving stationarity.

The classical JK procedure systematically leaves out an observation from the data set, each time generating a resample of size one less than the original data set. However, BM and POT methods are only concerned with data points that constitute an extreme value. Hence, we propose and implement an adapted JK as follows. In the case of BM, remove blocks, of the pre-specified size, one at a time. This effectively removes one of the BM at a time and creates a series of resamples of size one-block less than the original data. Similarly, for the POT method, we first identify observations that are above the pre-specified threshold value, then remove these exceedances from the data one at a time.

In each resampling process mentioned above, the EVT parameters are re-estimated for each resample and an empirical distribution is subsequently constructed to approximate the integrals in Equations 4.7 and 4.8. These calculations are then implemented in a rolling window procedure, explained in the next two sections.

Dependencies between the parameters can be examined by computing a generalised form of Hoeffdings D statistics, described in Hollander & Wolfe (1973), for every pair of parameters. The value of this statistic ranges between -0.5 and 1. A statistic value close to 1 indicates a strong dependency in the pair of variables, for a wide range of alternatives to independence, such as non-monotonic relationships.

## 4.5 VaR and Backtesting

The magnitude of market risk capital, reserved by financial institutes as per the Basel Accord, is directly related to the level of portfolio risk. VaR is a popular measure used rolling periods of our data sets. For example, ALSI returns gave block lengths with a mean of 2.95 and a standard deviation of 2.72 (similar values were obtained for other data sets). These result in similar values for  $c$  as set by default in `tsbootstrap`. In addition, considering the data sets are weakly dependent, the efficiency on computing power and the fact that SBS is less sensitive to block size selection (relative to other block bootstrap), we have implemented the above default value for  $c$ .

to quantify this risk by computing the maximum possible loss for a portfolio over a specified time period. Its calculations focus on the tails of a distribution.

Suppose  $X$  is a random variable with distribution function  $F$ . VaR over a specified time period, for a given probability  $\alpha$ , can be defined as the  $\alpha$ -th upper quantile of  $F$ , i.e.,

$$\text{VaR}_\alpha = F^{-1}(1 - \alpha) \tag{4.9}$$

where  $F^{-1}$  is the corresponding quantile function. Under the unconditional EVT approach, we can estimate VaR using quantiles of the fitted model.

Let  $\alpha_{\text{ext}}$  be the probability that a block maximum, observed over a period of  $n$  time units, exceeds  $\text{VaR}_\alpha$ , i.e.,  $\mathbb{P}(M_n > \text{VaR}_\alpha)$ . We then deduce the following from expression (4.1),

$$\begin{aligned} \alpha_{\text{ext}} &= \mathbb{P}(M_n > \text{VaR}_\alpha) \\ &= 1 - \mathbb{P}(X_1 < \text{VaR}_\alpha, \dots, X_n < \text{VaR}_\alpha) \\ &= 1 - (1 - \alpha)^n, \end{aligned} \tag{4.10}$$

given that the underlying sequence is IID. Subsequently, if we want to estimate VaR at level  $\alpha$  using the unconditional BM method, we can simply compute the corresponding quantile from GEVD at  $\alpha_{\text{ext}}$ . However, the above argument is not attainable for an exchangeable sequence, since an exchangeable sequence is not necessarily IID. Alternatively, by noting that an exchangeable sequence is strictly stationary, we can obtain a generalised approximate form of the above, i.e.,

$$\begin{aligned} \alpha_{\text{ext}} &= \mathbb{P}(M_n > \text{VaR}_\alpha) \\ &= 1 - F_{M_n}(x) \\ &\approx 1 - (1 - \alpha)^{n^\lambda}, \end{aligned} \tag{4.11}$$

where  $\lambda$  is a constant known as the extremal index (Smith & Weissman, 1994). For the current work, we estimate  $\lambda$  by using the *blocks method* described in Embrechts et al. (1997).

As for the POT method, we rearrange expression (4.4), and replace  $x$  with  $x - u$ , to obtain

$$F(x) = (1 - F(u))F_u(x - u) + F(u). \tag{4.12}$$

For a sufficiently high value of  $u$ , we can estimate  $F(u)$  by  $(1 - N_u/N)$ , where  $N$  is the total number of observations in the data and  $N_u$  is the number of observations exceeding  $u$ . Further,  $F_u(x - u)$  can be estimated by a GPD, for the unconditional IID case, or by expression (4.8), in the case of an exchangeable sequence. These methods thus allow us to estimate the inverse probability  $F^{-1}(1 - \alpha)$ .

The steps to estimate VaR for the next time point (say  $t + 1$ ) in an exchangeable sequence are summarised below:

- (1) Consider a historic sequence of observations (up to the current time point  $t$ ) as an exchangeable sequence. Hence, resample (using BS, SBS or JK) from this sequence for a large number of times and, each time, apply the BM or POT method to the resample (for a pre-specified block size  $n$  or threshold level  $u$ , respectively). This will form a range of values for the EVT parameters.
- (2) Construct an empirical distribution for the EVT parameters, using the result from above. For the BM method, estimate  $\lambda$  and search for an estimate for  $F_{M_n}^{-1}(\alpha)$  in expression (4.7). Similarly, use a search algorithm to find an estimate for  $F_u^{-1}(\alpha)$  in expression (4.8) for the POT approach.

The above steps are implemented in a rolling window procedure to produce consecutive estimates of VaR over a long period of time. We then backtest these VaR estimates, against the realised values, with the widely used Kupiec likelihood-ratio test (see Remark 1.5) and the Christoffersen's conditional coverage test (see Remark 1.6).

In this work, we also compare our new methods against the classical unconditional EVT approaches. Furthermore, we benchmark our results against the GARCH-EVT framework introduced by McNeil & Frey (2000). This framework is a conditional approach for VaR modelling, where POT is used for estimating the tail of the innovation distribution of the GARCH model. This is described in the next section.

## 4.6 GARCH and VaR

Instead of modelling the distribution of  $X$  directly, let us now consider a conditional model as follows. For a stationary sequence  $\{X_1, X_2, \dots\}$ , we assume

$$X_t = \mu_t + \sigma_t Z_t \tag{4.13}$$

where  $Z_t$  are the innovations with marginal distribution  $F_Z(z)$ , while  $\mu_t$  and  $\sigma_t$  are assumed to be measurable with respect to  $\Phi_{t-1}$ , the information about the process available up to time  $t - 1$ . Also,  $F_X(x)$  denotes the marginal distribution of  $X_t$ . Hence, we may write

$$\begin{aligned} F_{X_{t+1}|\Phi_t}(x) &= P(\mu_{t+1} + \sigma_{t+1}Z_{t+1} \leq x|\Phi_t) \\ &= F_Z\left(\frac{x - \mu_{t+1}}{\sigma_{t+1}}\right) \end{aligned} \tag{4.14}$$

and define VaR for day  $t + 1$  simply as

$$\text{VaR}_\alpha(t + 1) = \mu_{t+1} + \sigma_{t+1}z_\alpha \tag{4.15}$$

where  $z_\alpha$  denotes the upper  $\alpha$ -th quantile of  $Z_t$ . To estimate this, we need to specify a model for the dynamics of the conditional mean and volatility. This is typically done using the GARCH(1,1) process for the volatility and the AR(1) model for the conditional mean, i.e.,

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1\epsilon_t^2 + \beta\sigma_t^2 \quad \text{and} \quad \mu_{t+1} = \phi X_t, \tag{4.16}$$

where  $\epsilon_t = \sigma_t Z_t$ ,  $\alpha_0 > 0$ ,  $\alpha_1 > 0$ ,  $\beta > 0$ , and  $\phi$  is the AR(1) coefficient. It is also commonly assumed that the GARCH model is fitted using MLE, where  $Z_t$  follows a standard normal (i.e.,  $z_\alpha$  is simply a standard normal quantile), and  $\mu_{t+1}$  and  $\sigma_{t+1}$  are estimated using standard 1-day forecasts McNeil & Frey (2000).

McNeil & Frey (2000) also proposed amalgamating the GARCH model with the POT method, to produce a conditional GARCH-EVT approach for estimating VaR. They used a pseudo maximum likelihood (PML) procedure, to minimise assumptions about the model innovations, when estimating the GARCH parameters. This procedure still assumes normality for the likelihood construction, but uses robust standard errors for inference and yields consistent estimators. They further estimated the innovation quantile using POT. This is then combined with the forecasts for mean and volatility to obtain an estimate for VaR. One possible drawback of the GARCH-EVT approach is that the procedure assumes the model innovations are IID (since the classical POT method is based on an IID sequence). We again propose to relax the IID assumption by considering exchangeable sequences. We will use SBS for estimating the prior distribution for the POT parameters.

We extend the steps described in Section 4.5 as below:

- (1) Fit a GARCH model to a historic window of returns of size  $w$ , i.e.,  $\{x_{t-w+1}, \dots, x_t\}$ , using the PML approach. Forecast  $\mu_{t+1}$  and  $\sigma_{t+1}$ , using the fitted model and extract the corresponding residuals (i.e., innovations).
- (2) Consider the residuals as an exchangeable sequence. Hence, resample from this sequence of residuals, a large number of times, using SBS and, each time, apply the POT method to the bootstrapped sample (for a pre-specified threshold level  $u$ ). This will form a range of values for  $\hat{\xi}$  and  $\hat{\beta}$ .
- (3) Construct an empirical distribution  $\hat{F}_{\hat{\xi}, \hat{\beta}}$  for  $\xi$  and  $\beta$ , using the result from above. Use a search algorithm to find an estimate for  $z_\alpha = F^{-1}(1 - \alpha)$  and, hence, calculate  $\text{VaR}_\alpha(t + 1)$ .

This procedure is again implemented in a rolling window to produce consecutive estimates for VaR over a long period.

## 4.7 Empirical Results - Part I

To illustrate our methodology introduced from Section 4.5, and to compare model performances across different markets, we have chosen data sets extracted from the following indices: Standard & Poor 500 Index (S&P500), Financial Times Stock Exchange 100 Index (FTSE100), MSCI World Index (MSCI), Hang Seng Index (HSI) and FTSE/JSE All-Share Index (ALSI). All data series are comprised of daily closing prices, obtained from McGregor BFA. For daily stock prices  $\{P_1, P_2, \dots\}$ , the log-returns (or, simply returns) are calculated as follows

$$X_i = \ln(P_i) - \ln(P_{i-1}) . \quad (4.17)$$

The daily index prices for S&P500, FTSE100, MSCI and HSI were recorded from 21 November 1994 to 21 November 2014 (i.e., total of 5219 daily returns) and the prices of ALSI were recorded from 30 June 1995 to 21 November 2014 (i.e., total of 4847 daily returns).

Figures 4.1 to 4.5 present the time series plot of daily index prices and the time series plot of daily returns for each return series, respectively. The data sets spread include various important events in the financial history and are evidenced in the figures. For example, extreme price movements (i.e., extreme observations in the return series) are observable for the 1997 Asian Financial Crisis, Russian defaults in 1998, the economic recession in the early 2000s and the 2007-2008 Global Financial Crisis. Overall, the price of all indices significantly increased over the chosen time periods. The figures further indicate presence of heteroscedasticity and volatility clustering for the return series.

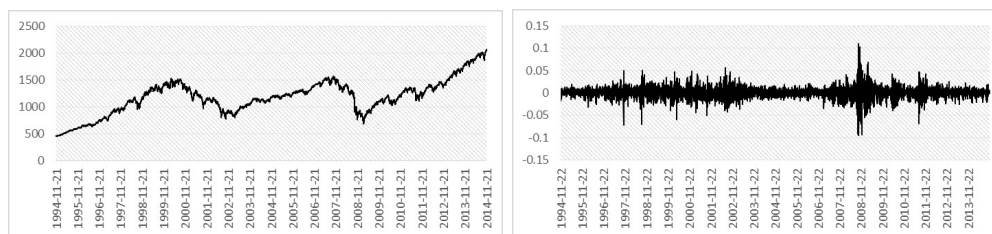


Figure 4.1: Time series plots for daily closing values (left) and returns (right) for S&P500.

Table 4.1 provides the descriptive summaries for returns of the five indices. All return series produced a slightly positive mean of returns. They are all characterised by

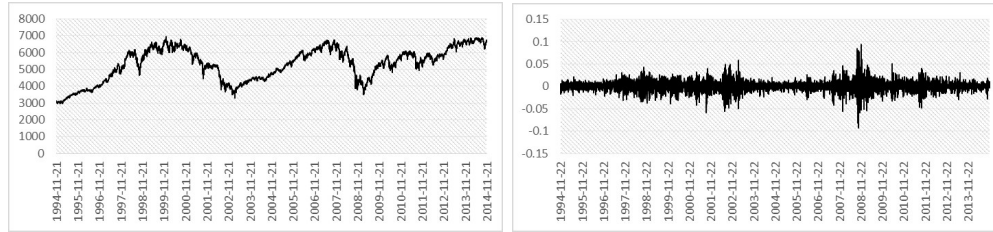


Figure 4.2: Time series plots for daily closing values (left) and returns (right) for FTSE100.

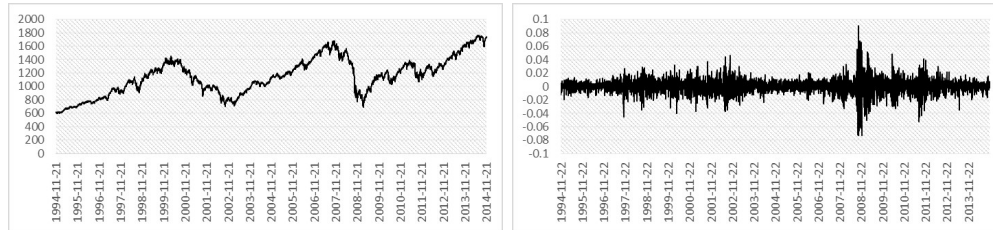


Figure 4.3: Time series plots for daily closing values (left) and returns (right) for MSCI.

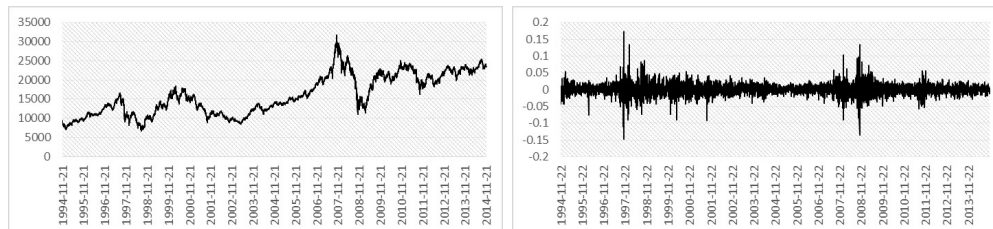


Figure 4.4: Time series plots for daily closing values (left) and returns (right) for HSI.

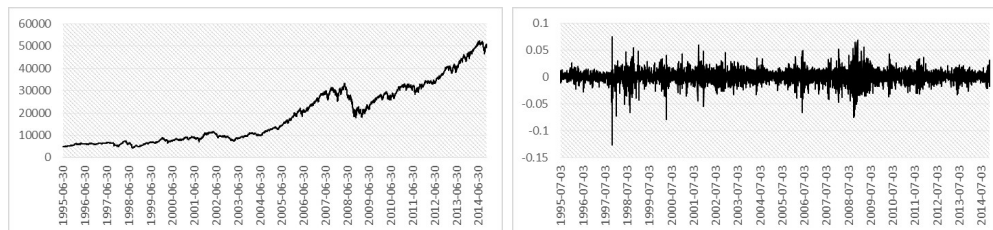


Figure 4.5: Time series plots for daily closing values (left) and returns (right) for ALSI.

substantial skewness and high excess kurtosis. These are results of the leptokurtic and asymmetric behaviours in the return series, as commonly suggested in the literature. These non-Gaussian characteristics are also confirmed by rejections in the Anderson-Darling normality test.

For the purpose of VaR modelling, our analyses are focused on extreme negative returns in the data. Hence, for convenience and for ease of computation, we utilise the relation  $\min\{X_1, \dots, X_n\} = -\max\{-X_1, \dots, -X_n\}$  and multiply each return series by -1. In accordance with McNeil & Frey (2000), we also use a rolling window of 1000 days of historical observations to predict VaR for the next day.

In the analyses to follow, block sizes of 5, 10 and 21 are selected for the BM method



Table 4.1: Descriptive statistics for daily returns of the five indices.

Index	Min	Max	Mean	Std. dev.	Skewness	Excess Kurtosis	Anderson-Darling test statistic ( $p$ -value)
S&P500	-0.0947	0.1096	0.0003	0.0120	-0.2503	8.5299	1966.15 (<0.0001)
FTSE100	-0.0927	0.0938	0.0001	0.0116	-0.1624	6.3388	1966.80 (<0.0001)
MSCI	-0.0733	0.0910	0.0002	0.0098	-0.3734	7.9793	1974.75 (<0.0001)
HSI	-0.1473	0.1725	0.0002	0.0163	0.0894	10.6296	1948.76 (<0.0001)
ALSI	-0.1269	0.0742	0.0005	0.0124	-0.4741	6.2700	1822.66 (<0.0001)

(for both IID and exchangeable sequences), resulting in estimates for weekly, fortnightly and monthly maxima series, respectively. On the other hand, threshold values at 80%, 90% and 95% sample quantiles (re-calculated for each window) are used for the POT method (for both IID and exchangeable sequences). These choices of threshold level are supported by examining the mean excess plot for each data set, where these quantiles lie on a positive straight line (Coles, 2001). As for the GARCH-EVT approach, we follow McNeil and Frey in selecting 90% quantile as the threshold level for the innovations and the GARCH model is fitted using PML estimation.

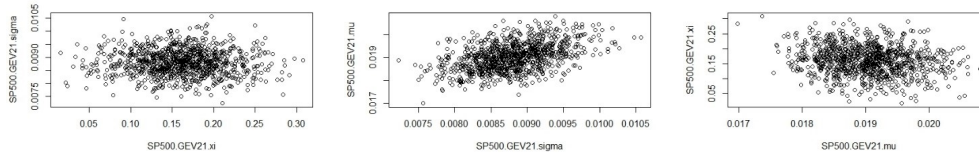


Figure 4.6: Scatter plots of GEVD parameters for S&amp;P500, with BS resampling and block size 21.

Table 4.2: Hoeffding's D Statistics for parameters estimated from fitting GEVD to S&amp;P500.

Distribution	Jackknife			Bootstrap		
	$\xi$ vs $\sigma$	$\sigma$ vs $\mu$	$\mu$ vs $\xi$	$\xi$ vs $\sigma$	$\sigma$ vs $\mu$	$\mu$ vs $\xi$
GEVD5	0.21	0.52	0.33	0.03	0.09	0.03
GEVD10	0.25	0.50	0.35	0.00	0.06	0.02
GEVD21	0.26	0.56	0.39	0.00	0.06	0.01

For each window (when using the exchangeable EVT approach), 1000 resamples are generated using BS, SBS and the adapted JK as described earlier. Parameter estimation is then performed for each resample, which would generate empirical prior distributions for the EVT parameters in each window. Dependencies between parameters

Table 4.3: Hoeffding’s D Statistics for parameters estimated from fitting GPD to S&P500.

	<b>Jackknife</b>	<b>Bootstrap</b>
<b>Distribution</b>	$\xi$ vs $\beta$	$\xi$ vs $\beta$
GPD80	0.71	0.11
GPD90	0.65	0.12
GPD95	0.62	0.10

were examined using pairwise scatterplots and the Hoeffdings D statistic. A sample of these results for S&P500 are presented in Figure 4.6 and Tables 4.2 to 4.3. The results provide very little evidence of strong dependencies between the parameters (although some weak dependencies were observed). Similar results were obtained for all other models and indices. This led us to treating the parameters as mutually independent. The extremal index for the BM method on exchangeable sequences is also re-estimated in every window. The search algorithm we shall use, for estimating  $F_u^{-1}(1 - \alpha)$  and  $F_{M_n}^{-1}(1 - \alpha)$ , is a combination of golden search algorithm and successive parabolic interpolation (Brent, 1973), as built in the R function `optimize()`.

VaR estimation and backtesting are performed at two confidence levels, namely at 1% and 2.5%, using the Kupiec likelihood ratio test and the Christoffersen conditional coverage test. The results for various models and indices are recorded in Tables 4.6 to 4.10, given at the end of this section. To better compare the overall performance of our new models with existing approaches, we further summarise the results in Tables 4.4 and 4.5.

Table 4.4 compares the new EVT approach for exchangeable sequences to the classical unconditional EVT approach for IID sequences. This comparison is examined for both BM and POT methods and at three different block sizes and threshold levels, respectively. In 47 out of 60 cases for the BM method, one can improve the VaR adequacy by switching to the BM method for exchangeable sequences. Whereas, the assumption of exchangeability can improve only 35 out of 60 cases of VaR estimates when utilising the POT method. The exchangeability assumption seems to be have a stronger effect on the BM method. On a related note, prior research (see, for example, Gilli & K ellezi, 2006; Huang et al., 2014; Chinghamu et al., 2015) has often shown that GPD performs better than GEVD in evaluating financial risk when using the classical unconditional EVT approach. However, under the assumption of exchangeability, this is not always the case. In particular, by comparing values across Tables 4.6 to 4.10, we may observe numerous cases where GEVD, using our resampling approach, produced higher Kupiec

Table 4.4: Indication of whether the exchangeable EVT approach do improve on the classical unconditional EVT models, in terms of one-day-ahead VaR forecasting.

Data	Test	Level	BM			POT		
			$n = 5$	$n = 10$	$n = 21$	$u = 0.8$	$u = 0.9$	$u = 0.95$
S&P500	Kupiec	1%	Y	Y	Y	Y	Y	Y
		2.5%	Y	Y	Y	Y	Y	Y
	Christoffersen	1%	Y	Y	Y	N	Y	Y
		2.5%	Y	Y	Y	Y	Y	Y
FTSE100	Kupiec	1%	Y	Y	Y	Y	Y	Y
		2.5%	N	Y	Y	N	Y	Y
	Christoffersen	1%	Y	Y	Y	N	Y	Y
		2.5%	Y	Y	Y	N	N	N
MSCI	Kupiec	1%	Y	Y	Y	Y	Y	Y
		2.5%	Y	Y	Y	Y	Y	Y
	Christoffersen	1%	N	N	N	N	N	N
		2.5%	N	N	N	Y	N	N
HSI	Kupiec	1%	N	Y	Y	Y	Y	N
		2.5%	N	Y	Y	N	Y	Y
	Christoffersen	1%	N	Y	Y	N	Y	Y
		2.5%	N	N	N	N	N	Y
ALSI	Kupiec	1%	Y	Y	Y	N	Y	N
		2.5%	Y	Y	Y	N	Y	N
	Christoffersen	1%	Y	Y	Y	N	N	N
		2.5%	Y	Y	Y	N	Y	N

Notes: This table provides a partial summary of results from Tables 4.6 to 4.10, in terms of whether the exchangeable EVT approaches can improve on the classical unconditional EVT methods. In particular, this is examined for each block size level of the BM method and for each threshold level of the POT method. Y = the classical unconditional EVT method is outperformed by a corresponding exchangeable EVT method (highlighted in grey); N = no improvement is observed by using the exchangeable EVT approaches (not highlighted).

or Christoffersen  $p$ -values than the corresponding GPD estimates. This is an indication that, by accommodating for some prior belief in the parameter variations (in our case, the empirical prior distributions), we have made the relative performances between BM and POT methods more comparable with each other. When comparing across indices, we find that MSCI and HIS have the least cases of VaR improvement when switched to the exchangeability assumption (i.e. 11 out of 24 cases). Interestingly, MSCI deviates the most from normality and HSI have the highest excess kurtosis (see Table 1). On the other hand, ALSI is the closest to normality (relatively) with the smallest excess kurtosis. This may be attributed to the fact that JSE is a relatively smaller market with prudent fiscal and monetary policies, making it less affected by global events, relative to its international counterparts. Consequently, the risk in ALSI seems better captured by the various models, with its best performing models producing higher  $p$ -values, as compared to other indices (especially for the Christoffersen conditional coverage test).

Table 4.5: Model(s) with the most adequate VaR forecasts for various indices, at different VaR levels.

Data	Kupiec		Christoffersen	
	1%	2.5%	1%	2.5%
S&P500	GPD90-SBS	GPD90-SBS	GPD80-IID	GPD80-SBS
FTSE100	GEVD21-BS	GPD95-SBS	GPD95-SBS	GEVD10-BS
MSCI	GEVD10-BS	GPD90-SBS	-	GPD80-SBS
HSI	GPD80-SBS/GPD90-JK/ GARCH-EVT	GEVD21-JK	GEVD10-JK	GARCH-EVT
ALSI	GEVD5-SBS	GPD90-SBS	GEVD21-JK	GARCH-EVT

Notes: This table records the model(s) that produced the most adequate VaR forecasts for each index, at each of two different VaR confidence levels, by comparing results obtained in Tables 4.6 to 4.10. GEVD5-SBS denotes the BM method for exchangeable sequences with block size 5 and using SBS resampling, while GPD80-IID denotes the classical unconditional POT method with threshold level at 80% sample quantile, etc. GARCH-EVT represents the conditional approach proposed by McNeil and Frey (2000). No adequate model was observed for MSCI at 1% VaR level when using the Christoffersen conditional coverage test.

Table 4.5 records the best performing model(s) for each index and for the two VaR adequacy tests at 1% and 2.5% levels. It is observed that the best model involves using SBS in 10 out of 20 cases. This is to some extent expected as SBS caters for dependencies in the data set when resampling from each window. We also note that the GARCH-EVT model only produced the best result in 3 out of the 20 cases, while the best VaR estimates can be achieved by our new approaches in 16 out of 20 cases. This is a strong evidence that our proposed new EVT approaches for exchangeable sequences can produce superior forecasting performances as compared to the GARCH-EVT model.

Table 4.6: One-day-ahead out-of-sample test for VaR estimates of S&amp;P500.

VaR levels	No. of exceedances (expected)		Kupiec $p$ -value		Christoffersen $p$ -value	
	1% (42)	2.5% (105)	1%	2.5%	1%	2.5%
GEVD5-IID	62	141	0.0042	0.0009	0.0001	<0.0001
GEVD5-JK	78	150	<0.0001	<0.0001	<0.0001	<0.0001
GEVD5-BS	53	111	0.1077	0.5890	0.0048	0.0029
GEVD5-SBS	57	126	0.0296	0.0494	0.0004	0.0010
GEVD10-IID	68	165	0.0002	<0.0001	<0.0001	<0.0001
GEVD10-JK	75	130	<0.0001	0.0195	<0.0001	<0.0001
GEVD10-BS	47	99	0.4648	0.5189	0.0005	0.0004
GEVD10-SBS	57	124	0.0296	0.0754	0.0004	0.0011
GEVD21-IID	89	203	<0.0001	<0.0001	<0.0001	<0.0001
GEVD21-JK	58	126	0.0206	0.0494	<0.0001	<0.0001
GEVD21-BS	50	94	0.2403	0.2491	0.0057	0.0028
GEVD21-SBS	56	122	0.0419	0.1117	0.0005	0.0003
GPD80-IID	55	115	0.0583	0.3545	0.0073	0.0073
GPD80-JK	54	115	0.0799	0.3545	0.0043	0.0037
GPD80-BS	53	113	0.1077	0.4631	0.0048	0.0033
GPD80-SBS	53	116	0.1077	0.3069	0.0048	0.0377
GPD90-IID	57	121	0.0296	0.1345	0.0004	0.0003
GPD90-JK	55	119	0.0583	0.1911	0.0005	0.0003
GPD90-BS	51	112	0.1869	0.5240	0.0006	0.0008
GPD90-SBS	46	107	0.5612	0.8807	0.0052	0.0018
GPD95-IID	61	124	0.0064	0.0754	0.0002	0.0003
GPD95-JK	58	121	0.0206	0.1345	0.0003	0.0011
GPD95-BS	56	110	0.0419	0.6578	0.0004	0.0006
GPD95-SBS	52	101	0.1430	0.6568	0.0006	0.0007
GARCH-EVT	56	120	0.0419	0.1609	<0.0001	<0.0001

Notes: This table shows results for the testing of VaR estimates derived from the various methods discussed in this section. The tests are performed at 1% and 2.5% VaR confidence levels for losses using the Kupiec likelihood ratio test and the Christoffersen conditional coverage test. GEVD5-IID and GPD80-IID denotes the classical unconditional EVT methods with block size 5 and threshold level at 80% sample quantile, respectively. GEVD5-JK, GEVD5-BS and GEVD5-SBS denotes the BM method for exchangeable sequences, with block size 5, using JK, BS and SBS resampling, respectively. Similar notations analogously apply to other models and GARCH-EVT refers to the approach proposed by McNeil and Frey (2000).

Table 4.7: One-day-ahead out-of-sample test for VaR estimates of FTSE100.

VaR levels	No. of exceedances (expected)		Kupiec $p$ -value		Christoffersen $p$ -value	
	1% (42)	2.5% (105)	1%	2.5%	1%	2.5%
GEVD5-IID	73	143	<0.0001	0.0004	<0.0001	<0.0001
GEVD5-JK	88	154	<0.0001	<0.0001	<0.0001	<0.0001
GEVD5-BS	59	117	0.0141	0.2639	0.0018	<0.0001
GEVD5-SBS	65	131	0.0011	0.0152	<0.0001	<0.0001
GEVD10-IID	79	161	<0.0001	<.0001	<0.0001	<0.0001
GEVD10-JK	75	142	<0.0001	0.0006	<0.0001	<0.0001
GEVD10-BS	53	101	0.1077	0.6568	0.0048	0.0007
GEVD10-SBS	68	130	0.0002	0.0195	<0.0001	<0.0001
GEVD21-IID	93	202	<0.0001	<0.0001	<0.0001	<0.0001
GEVD21-JK	72	145	<0.0001	0.0002	<0.0001	<0.0001
GEVD21-BS	44	89	0.7809	0.0952	0.0043	0.0001
GEVD21-SBS	67	128	0.0004	0.0315	<0.0001	<0.0001
GPD80-IID	66	127	0.0007	0.0396	<0.0001	<0.0001
GPD80-JK	66	127	0.0007	0.0396	<0.0001	<0.0001
GPD80-BS	66	127	0.0007	0.0396	<0.0001	<0.0001
GPD80-SBS	65	127	0.0011	0.0396	<0.0001	<0.0001
GPD90-IID	67	123	0.0004	0.0921	<0.0001	<0.0001
GPD90-JK	67	123	0.0004	0.0921	<0.0001	<0.0001
GPD90-BS	61	117	0.0064	0.2639	0.0011	<0.0001
GPD90-SBS	64	115	0.0017	0.3545	<0.0001	<0.0001
GPD95-IID	60	121	0.0096	0.1345	0.0014	<0.0001
GPD95-JK	60	121	0.0096	0.1345	0.0014	<0.0001
GPD95-BS	60	115	0.0096	0.3545	0.0014	<0.0001
GPD95-SBS	55	103	0.0583	0.8065	0.0202	<0.0001
GARCH-EVT	67	122	0.0004	0.1117	<0.0001	<0.0001

Notes: This table shows results for the testing of VaR estimates derived from the various methods discussed in this section. The tests are performed at 1% and 2.5% VaR confidence levels for losses using the Kupiec likelihood ratio test and the Christoffersen conditional coverage test. GEVD5-IID and GPD80-IID denotes the classical unconditional EVT methods with block size 5 and threshold level at 80% sample quantile, respectively. GEVD5-JK, GEVD5-BS and GEVD5-SBS denotes the BM method for exchangeable sequences, with block size 5, using JK, BS and SBS resampling, respectively. Similar notations analogously apply to other models and GARCH-EVT refers to the approach proposed by McNeil and Frey (2000).

Table 4.8: One-day-ahead out-of-sample test for VaR estimates of MSCI.

VaR levels	No. of exceedances (expected)		Kupiec $p$ -value		Christoffersen $p$ -value	
	1% (42)	2.5% (105)	1%	2.5%	1%	2.5%
GEVD5-IID	73	151	<0.0001	<0.0001	<0.0001	<0.0001
GEVD5-JK	115	175	<0.0001	<0.0001	<0.0001	<0.0001
GEVD5-BS	51	103	0.1869	0.8065	<0.0001	<0.0001
GEVD5-SBS	61	124	0.0064	0.0754	<0.0001	<0.0001
GEVD10-IID	81	178	<0.0001	<0.0001	<0.0001	<0.0001
GEVD10-JK	60	130	0.0096	0.0195	<0.0001	<0.0001
GEVD10-BS	44	94	0.7809	0.2491	<0.0001	<0.0001
GEVD10-SBS	58	127	0.0206	0.0396	<0.0001	<0.0001
GEVD21-IID	90	209	<0.0001	<0.0001	<0.0001	<0.0001
GEVD21-JK	92	166	<0.0001	<0.0001	<0.0001	<0.0001
GEVD21-BS	40	85	0.7325	0.0368	<0.0001	<0.0001
GEVD21-SBS	56	123	0.0419	0.0921	<0.0001	<0.0001
GPD80-IID	58	123	0.0206	0.0921	<0.0001	<0.0001
GPD80-JK	57	122	0.0296	0.1117	<0.0001	<0.0001
GPD80-BS	51	108	0.1869	0.8041	<0.0001	<0.0001
GPD80-SBS	52	100	0.1430	0.5861	<0.0001	0.0102
GPD90-IID	58	119	0.0206	0.1911	<0.0001	<0.0001
GPD90-JK	53	114	0.1077	0.4066	<0.0001	<0.0001
GPD90-BS	55	113	0.0583	0.4631	<0.0001	<0.0001
GPD90-SBS	52	107	0.1430	0.8807	<0.0001	<0.0001
GPD95-IID	57	118	0.0296	0.2254	<0.0001	<0.0001
GPD95-JK	52	109	0.1430	0.7295	<0.0001	<0.0001
GPD95-BS	48	86	0.3790	0.0474	<0.0001	<0.0001
GPD95-SBS	49	86	0.3042	0.0474	<0.0001	<0.0001
GARCH-EVT	59	122	0.0141	0.1117	<0.0001	0.0003

Notes: This table shows results for the testing of VaR estimates derived from the various methods discussed in this section. The tests are performed at 1% and 2.5% VaR confidence levels for losses using the Kupiec likelihood ratio test and the Christoffersen conditional coverage test. GEVD5-IID and GPD80-IID denotes the classical unconditional EVT methods with block size 5 and threshold level at 80% sample quantile, respectively. GEVD5-JK, GEVD5-BS and GEVD5-SBS denotes the BM method for exchangeable sequences, with block size 5, using JK, BS and SBS resampling, respectively. Similar notations analogously apply to other models and GARCH-EVT refers to the approach proposed by McNeil and Frey (2000).

Table 4.9: One-day-ahead out-of-sample test for VaR estimates of HSI.

VaR levels	No. of exceedances (expected)		Kupiec $p$ -value		Christoffersen $p$ -value	
	1% (42)	2.5% (105)	1%	2.5%	1%	2.5%
GEVD5-IID	40	111	0.7325	0.5890	0.0020	<0.0001
GEVD5-JK	32	85	0.0996	0.0368	0.0017	<0.0001
GEVD5-BS	34	79	0.1897	0.0063	0.0003	<0.0001
GEVD5-SBS	34	97	0.1897	0.3970	0.0003	<0.0001
GEVD10-IID	61	125	0.0064	0.0613	<0.0001	<0.0001
GEVD10-JK	43	83	0.9006	0.0214	0.0342	<0.0001
GEVD10-BS	37	70	0.4121	0.0002	0.0008	<0.0001
GEVD10-SBS	41	96	0.8532	0.3429	0.0026	<0.0001
GEVD21-IID	70	162	<0.0001	<0.0001	<0.0001	<0.0001
GEVD21-JK	52	101	0.1430	0.6568	<0.0001	<0.0001
GEVD21-BS	35	61	0.2517	<0.0001	0.0004	<0.0001
GEVD21-SBS	43	93	0.9006	0.2095	0.0003	<0.0001
GPD80-IID	41	88	0.8532	0.0762	0.0002	<0.0001
GPD80-JK	41	88	0.8532	0.0762	0.0002	<0.0001
GPD80-BS	41	87	0.8532	0.0604	0.0002	<0.0001
GPD80-SBS	42	88	0.9765	0.0762	0.0002	<0.0001
GPD90-IID	43	88	0.9006	0.0762	0.0003	<0.0001
GPD90-JK	42	88	0.9765	0.0762	0.0002	<0.0001
GPD90-BS	44	88	0.7809	0.0762	0.0004	<0.0001
GPD90-SBS	44	94	0.7809	0.2491	0.0004	<0.0001
GPD95-IID	43	91	0.9006	0.1440	0.0003	0.0002
GPD95-JK	43	92	0.9006	0.1745	0.0003	0.0003
GPD95-BS	41	87	0.8532	0.0604	0.0002	<0.0001
GPD95-SBS	44	91	0.7809	0.1440	0.0004	0.0002
GARCH-EVT	42	89	0.9765	0.0952	0.0002	0.0007

Notes: This table shows results for the testing of VaR estimates derived from the various methods discussed in this section. The tests are performed at 1% and 2.5% VaR confidence levels for losses using the Kupiec likelihood ratio test and the Christoffersen conditional coverage test. GEVD5-IID and GPD80-IID denotes the classical unconditional EVT methods with block size 5 and threshold level at 80% sample quantile, respectively. GEVD5-JK, GEVD5-BS and GEVD5-SBS denotes the BM method for exchangeable sequences, with block size 5, using JK, BS and SBS resampling, respectively. Similar notations analogously apply to other models and GARCH-EVT refers to the approach proposed by McNeil and Frey (2000).



Table 4.10: One-day-ahead out-of-sample test for VaR estimates of ALSI.

VaR levels	No. of exceedances (expected)		Kupiec $p$ -value		Christoffersen $p$ -value	
	1% (42)	2.5% (105)	1%	2.5%	1%	2.5%
GEVD5-IID	41	102	0.6850	0.5513	0.5921	0.0015
GEVD5-JK	36	92	0.6858	0.6641	0.6557	0.0015
GEVD5-BS	33	78	0.3636	0.0524	0.4974	0.0033
GEVD5-SBS	37	94	0.8105	0.8216	<0.0001	0.0094
GEVD10-IID	46	128	0.2365	0.0017	0.2844	<0.0001
GEVD10-JK	23	75	0.0067	0.0230	0.0222	0.0047
GEVD10-BS	30	70	0.1534	0.0045	0.2851	0.0073
GEVD10-SBS	35	93	0.5680	0.7417	<0.0001	0.0300
GEVD21-IID	54	147	0.0177	<0.0001	0.0290	<0.0001
GEVD21-JK	41	104	0.6850	0.4250	0.7013	<0.0001
GEVD21-BS	27	65	0.0496	0.0006	0.1202	0.0021
GEVD21-SBS	34	91	0.4600	0.5898	<0.0001	0.0055
GPD80-IID	36	94	0.6858	0.8216	<0.0001	0.0094
GPD80-JK	36	94	0.6858	0.8216	<0.0001	0.0094
GPD80-BS	34	90	0.4600	0.5192	<0.0001	0.0045
GPD80-SBS	34	90	0.4600	0.5192	<0.0001	0.0045
GPD90-IID	35	94	0.5680	0.8216	<0.0001	0.0094
GPD90-JK	35	94	0.5680	0.8216	<0.0001	0.0094
GPD90-BS	35	94	0.5680	0.8216	<0.0001	0.0094
GPD90-SBS	36	96	0.6858	0.9858	<0.0001	0.0124
GPD95-IID	35	92	0.5680	0.6641	<0.0001	0.0067
GPD95-JK	35	92	0.5680	0.6641	<0.0001	0.0067
GPD95-BS	28	75	0.0746	0.0230	<0.0001	0.0002
GPD95-SBS	32	81	0.2803	0.1074	<0.0001	0.0004
GARCH-EVT	35	90	0.5680	0.5192	<0.0001	0.3980

Notes: This table shows results for the testing of VaR estimates derived from the various methods discussed in this section. The tests are performed at 1% and 2.5% VaR confidence levels for losses using the Kupiec likelihood ratio test and the Christoffersen conditional coverage test. GEVD5-IID and GPD80-IID denotes the classical unconditional EVT methods with block size 5 and threshold level at 80% sample quantile, respectively. GEVD5-JK, GEVD5-BS and GEVD5-SBS denotes the BM method for exchangeable sequences, with block size 5, using JK, BS and SBS resampling, respectively. Similar notations analogously apply to other models and GARCH-EVT refers to the approach proposed by McNeil and Frey (2000).

## 4.8 Empirical Results - Part II

We now extend our empirical analysis to include a GARCH filter for the negative daily returns, as discussed in Section 4.6. However, we shall focus only on two indices: FTSE100 and ALSI. As in McNeil & Frey (2000), we use a GARCH(1,1) filter (which can be easily generalised to other GARCH-type models) and have the POT method implemented with a threshold level at the 90% sample quantile. Only the SBS procedure is considered for estimating the empirical distribution of the POT parameters. All other variables are set as in the previous section.

For a preliminary investigation, we fit the GARCH model on rolling windows of 1000 daily returns and construct various autocorrelation function (ACF) and partial ACF (PACF) plots of the corresponding innovations. As an illustration, the PACF plot of FTSE100 residuals and the PACF plot of absolute values of ALSI residuals are shown in Figure 4.7, for the window from observation 3001 to 4000 (which includes the 2008/2009 financial crisis period). Both plots show a slow (if any) decay of autocorrelation, indicating possible long range dependency in the GARCH innovations (as one would also get for the original return series). Similar results were obtained for a large number of windows.

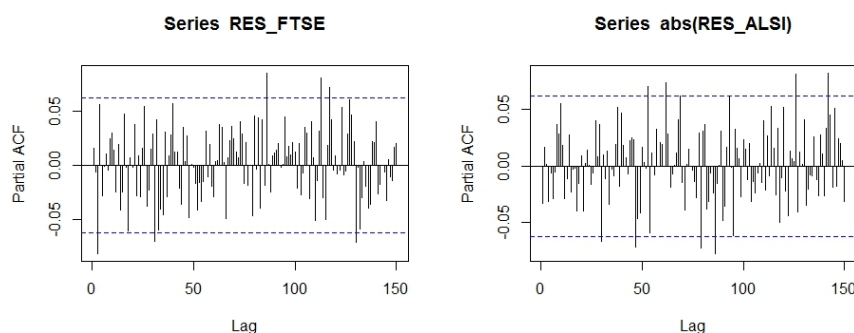


Figure 4.7: Partial ACF plots of GARCH residuals for FTSE100 and the absolute value of GARCH residuals for ALSI.

Figures 4.8 and 4.9 indicate records of kurtosis and skewness of all rolling windows, for both FTSE100 and ALSI. The black solid lines represent the kurtosis and skewness of the original negative returns series, while the red dotted lines represent the corresponding values for the model innovations. We note that the rolling residual kurtosis values are significantly lower than the kurtosis of the original return series. This is expected as the GARCH filters correct for some volatility clustering inherent in the returns. However, the residual kurtosis, for each data set, still remains above 3, and often peaks much higher than 3. This is consistent with a stylised fact that says financial returns have conditional heavy tails (Cont, 2001). The presence of high residual kurtosis and

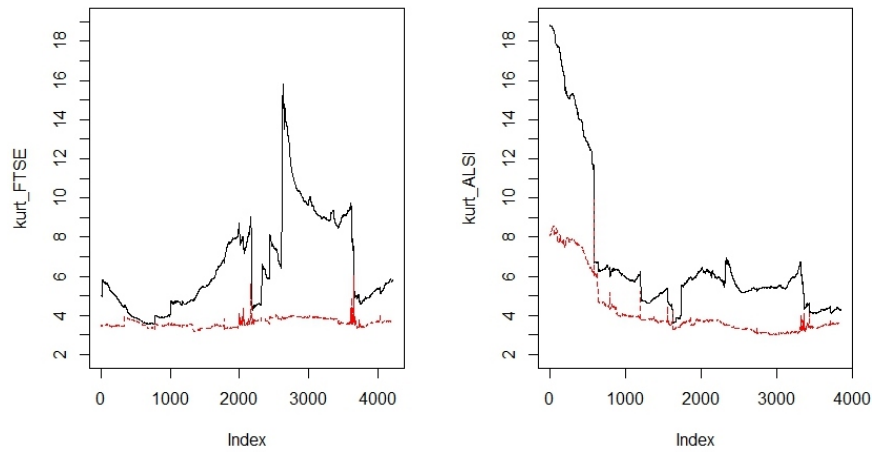


Figure 4.8: Kurtosis of GARCH residuals for rolling windows of 1000 daily returns.

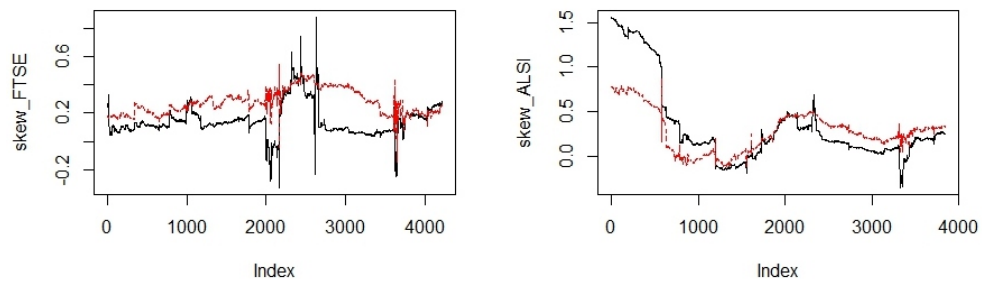


Figure 4.9: Skewness of GARCH residuals for rolling windows of 1000 daily returns.

skewness recorded at various windows is also supportive of the choice of using an EVT method for modelling the tails of the innovations. Although, it is interesting to note how both ALSI kurtosis series significantly dropped down and seemed to stabilise as the rolling window moved forward in time.

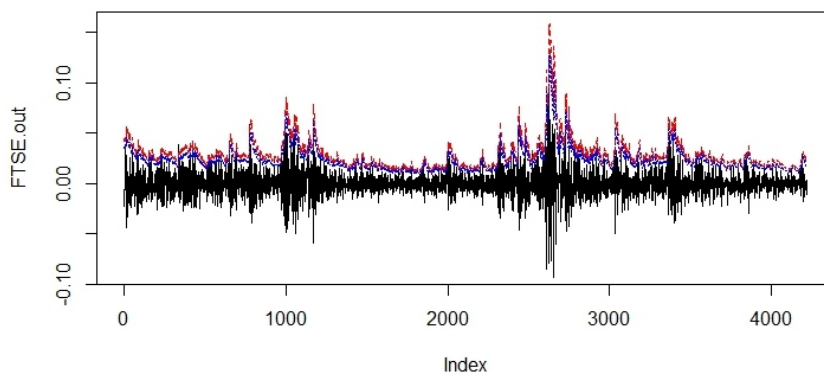


Figure 4.10: VaR estimates for negative FTSE100 returns using GARCH-EVT with exchangeable innovations (black line = daily returns; red line = 1% VaR estimates; blue line = 2.5% VaR estimates).

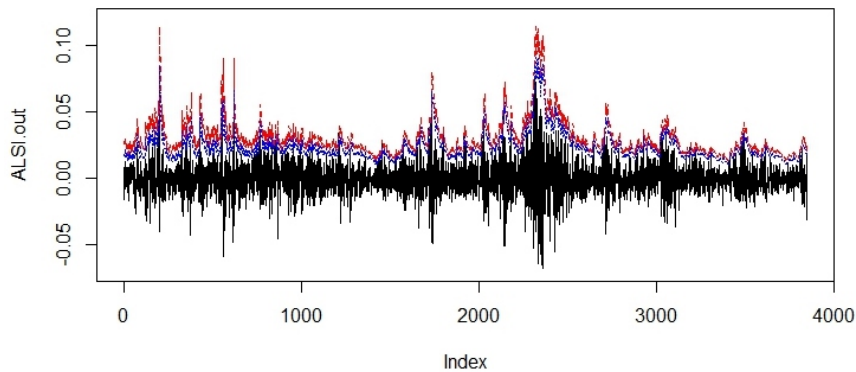


Figure 4.11: VaR estimates for negative ALSI returns using GARCH-EVT with exchangeable innovations (black line = daily returns; red line = 1% VaR estimates; blue line = 2.5% VaR estimates).

VaR estimates are calculated at both 2.5% and 1% VaR levels. Figures 4.10 and 4.11 present the negative daily returns of both data sets and the corresponding daily VaR estimates calculated using our new approach. The graphs show that the VaR model reacts consistently to market changes, especially during financial turmoil. For example, the graphs depict the highest peak during the 2008/2009 financial crisis, with the VaR estimates adjusting accordingly.

Table 4.11 shows the results of backtesting the VaR estimates against the corresponding realised negative return values. We contrast the performance of our new approach (denoted as GARCH-EVT-Exch) with the unconditional EVT approach (denoted as uncond. EVT), the GARCH-norm approach (where GARCH is fitted using ML and the innovation distribution is assumed to be standard normal) and the GARCH-EVT approach by McNeil & Frey (2000). The unconditional EVT approach is the application of the POT method on the original negative daily returns, without using a GARCH filter. The number of violations of VaR, the corresponding Kupiec likelihood-ratio test  $p$ -value and Christoffersen conditional coverage test  $p$ -value are recorded for the various models, at the two different VaR levels.

For FTSE100, it is quite clear that our new approach produced better VaR estimates than the other three models. As is well-known, the GARCH-norm model tends to underestimate the conditional heavy tails, resulting in excess amount of VaR violations. The unconditional EVT approach cannot respond quickly to changing volatility and tends to record consecutive violations during a stress period. The GARCH-EVT approach brings a slight improvement. However, its assumption of IID innovations depended on how well the data set is depicted by the GARCH filter. In this scenario, our model, which caters for residual dependencies, produced a significantly better VaR

Table 4.11: Backtesting of VaR estimates for negative daily returns.

Data	Method	No. of violations		Kupiec test		Christoffersen test	
		2.5%	1%	2.5%	1%	2.5%	1%
FTSE100	Expected	105	42	-	-	-	-
	Uncond. EVT	123	67	0.0921	0.0004	<0.0001	<0.0001
	GARCH-norm	154	74	<0.0001	<0.0001	<0.0001	<0.0001
	GARCH-EVT	122	67	0.1117	0.0004	<0.0001	<0.0001
	GARCH-EVT-Exch	107	44	0.8807	0.7809	0.7462	<0.0001
ALSI	Expected	96	38	-	-	-	-
	Uncond. EVT	94	35	0.8216	0.5680	0.0094	<0.0001
	GARCH-norm	111	63	0.1349	0.0003	0.3249	<0.0001
	GARCH-EVT	90	35	0.5192	0.5680	0.3980	<0.0001
	GARCH-EVT-Exch	89	40	0.4531	0.8054	0.7540	<0.0001

model, evidenced by the superior results from both tests.

A more interesting set of results is revealed for ALSI. The GARCH filter did not bring improvement to the unconditional EVT, in terms of the Kupiec test. We suggest this may be due to the effect of the stabilising kurtosis, as we observed earlier. However, GARCH-EVT produced a better result for the Christoffersen's test at the 2.5% level. Our new approach further improves on the result of this test at 2.5%. At the same time, a higher Kupiec  $p$ -value was also recorded for 1% VaR. Given that the Christoffersen's test is a much stricter test (i.e., tests for both correct number of exceedances and independence), we believe our new approach still produced a very competitive (if not better) model for ALSI.

## 4.9 Discussion and Further Research

In this chapter, we considered the notion of exchangeability<sup>2</sup> for EVT. We derived corresponding expressions for the distribution of BM and threshold exceedances, of an exchangeable sequence, and utilised them for financial risk valuation. Our prior belief in distributions of the parameters were based on resampling using BS, SBS and an adapted JK method for extreme observations. As such, we have shown that generalisations to

<sup>2</sup>Note that here exchangeability is viewed simply as a modelling assumption. In practice, an informed judgment is required to establish this subjective view (approximately), with the use of graphical analysis and other statistical tests related to properties of exchangeability. Like the assumption of IID, there is no definite test of exchangeability and it is most likely that one can never be definitely certain of this assumption. Nevertheless, exchangeability is a more general assumption than IID and is, in theory, expected to produce more robust results.

exchangeable sequences have the potential to improve EVT-based VaR estimation. In particular, it was also observed that, by accounting for parameter variations through exchangeability, the performance of BM and POT methods in estimating VaR have improved in general. This is supported by the empirical results of our out-of-sample VaR forecasts in five different indices. It is also worthwhile noting that the relative performances between the BM and POT methods are more comparable under the assumption of exchangeability. This is in contrast to the unconditional IID approach where the POT method often outperforms the BM in financial risk modelling. We also provided evidence that our new approaches can give superior forecasting performances than the GARCH-EVT model proposed by McNeil & Frey (2000).

We further proposed an extension of the GARCH-EVT approach to VaR estimation by catering for exchangeable innovations. This in turn allows for dependencies between innovations. We tested our new GARCH-EVT-Exch approach against the unconditional EVT, the GARCH-norm and the GARCH-EVT procedures using negative daily returns in FTSE100 and ALSI. GARCH-EVT-Exch produced convincingly the most robust model for FTSE100. However, the results are more mixed for ALSI, likely due to its generally low (apart from the early periods in the data set) and stabilising kurtosis.

The results obtained in this chapter act as an initiation for further development of the methodology proposed. For example, one could also use other subjective priors for the GEVD and GPD parameters to specify different dependency structures of the innovations. Further considerations of various partial exchangeability and probabilistic symmetry assumptions for the model innovations are possible. On the other hand, a fully Bayesian GARCH-EVT model may be considered, though estimation is difficult due to large number of parameters. Generalisations to other GARCH-type families requires further investigation, while comparison with the CaViaR method of Engel & Manganelli (2004) and the aforementioned methods by Chavez-Demoulin et al. (2014) and de Haan et al. (2016) may also be of interest.

Another potentially interesting matter is the choices of window size, block size and threshold level for VaR estimation. We designated a window size of 1000 days, as per the analysis of McNeil & Frey (2000). Block sizes were chosen as per practical implications of weekly, fortnightly and monthly returns. Simulations in McNeil & Frey (2000), Chavez-Demoulin et al. (2014) and de Haan et al. (2016) have suggested that the 90% sample quantile may be a suitable threshold choice. Nevertheless, it is not immediately obvious whether our new approach is sensitive to these choices. This exploration will require advanced computing power and is also left for future work. Finally, other prospective research could include: examining the exchangeability assumption for peaks-over-random-threshold (PORT) and duration-based POT models

(DPOT) (Santos et al., 2013); and utilising the new approaches for estimating ES (another popular risk measure).

Computations in this Chapter were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing Team.

## Chapter 5

# Exchangeability and Multi-phase Sampling

This chapter investigates the use of exchangeability in multi-phase sampling. Consider the problem of estimating the population mean of a study variable  $y$ , which is difficult to measure, but a related auxiliary variable  $x$ , with improved accessibility, is available. We can often apply a ratio-type estimator in this scenario. In cases where  $x$  is qualitative, or may be categorised, and a double sampling plan is used, we may consider a two-phase stratified sampling design. Traditionally, it is assumed that the  $N$  variables representing the readings on  $y$  are IID within and across strata. In this chapter, we relax this assumption to a judgment of exchangeable sequences within each stratum, while still maintaining the assumption of independence across strata. This caters for the existence of dependence structures for within-stratum readings. We propose a methodology for estimating the variance of the ratio estimator under this scenario. Through an example, we show that this method provides a significantly more conservative estimate for the sampling variance, as compared to the standard approach. The results obtained are published in Huang et al. (2016b).

### 5.1 Introduction

When considering the task of estimating the population mean of a study variable  $y$ , it is often the case that information on an auxiliary variable  $x$  is readily available for all units in the population. In such situations, it is common to utilise a ratio- or regression- type estimator to improve the efficiency in estimation (Cochran, 1977). However, when  $x$  is not known over the whole population, but still easier to obtain than  $y$ , we may implement a two-phase, or double, sampling design. The value of  $x$  is



observed for a large sample in phase 1 and  $y$  is subsequently recorded for a subsample in phase 2. This can be generalised to cater for multiple auxiliary variables with varying levels of accessibility and correlation, where several chain-type estimators are proposed (Mukerjee et al., 1987; Singh et al., 1994; Ahmed, 1998; Bhushan et al., 2008; Hamad et al., 2013).

As a way to measure how good a sampling estimator is, the estimator variance, or mean square error in the case of biased estimators, needs to be estimated. These are usually approximated by their corresponding asymptotic expressions, which commonly assumes IID observations. A way to relax the IID condition is to take on the Bayesian approach to finite population sampling, which assumes that the observations are *exchangeable* (Ericson, 1969; Treder & Sedransk, 1996). However, this approach also requires formalisation of prior information and known sampling distributions (or at least estimates of them).

In this chapter, we consider the case where  $x$  is a stratification variable, which is more easily accessible, and observations for  $y$  are obtained through phase 2 sampling from each stratum. We further assume the judgment of exchangeability within each stratum, while strata are mutually independent. This corresponds to finite population sampling without replacement. We propose a way to approximate the estimator variance under this scenario, using SBS at different levels of the sampling process. An example is considered which shows the standard procedure estimate underestimating the estimator variance, while our method provides an improvement.

## 5.2 Multi-phase Stratified Sampling

Let  $U = \{1, 2, \dots, N\}$  be the index set of a finite population of size  $N$  and  $y$  be the primary variable of interest. Suppose  $x$  is an auxiliary variable related to  $y$ , which is less expensive or is easier to measure. In this situation, it is common to consider a two-phase sampling design. In the first phase a large sample  $S' \subset U$  of size  $n'$  is drawn using SRSWOR and the auxiliary variable  $x$  is observed. Subsequently, a subsample  $S \subset S'$  of size  $n$  is drawn, using SRSWOR, to observe  $y$ . One way of incorporating the auxiliary information into the estimation of the population mean  $\bar{y}_U$ , is to use a ratio estimator

$$\bar{t}_{rat} = \frac{\bar{y}_n}{\bar{x}_n} \bar{x}_{n'} , \quad (5.1)$$

where  $\bar{y}_n = n^{-1} \sum_{i \in S} y_i$ ,  $\bar{x}_n = n^{-1} \sum_{i \in S} x_i$  and  $\bar{x}_{n'} = (n')^{-1} \sum_{i \in S'} x_i$ .

Often members of  $U$  can be cross-classified into groups based on the auxiliary variable; either the variable is qualitative in nature (e.g. gender), or may be categorised (e.g.

age). This scenario is classically associated with stratified sampling design, with unknown population stratum sizes. Suppose that the stratification variable  $x \in \{1, \dots, H\}$  is only observed after phase 1 and samples  $S_h$  (of sizes  $m_h$ ) are subsequently drawn from each stratum using SRSWOR. This results in an estimator for  $\bar{y}_U$  as

$$\bar{t}_{str} = \frac{1}{n'} \sum_{h=1}^H n_h \bar{y}_h, \quad (5.2)$$

where  $n_h$  is the number of units in  $S$  with  $x = h$  and  $\bar{y}_h = m_h^{-1} \sum_{i \in S_h} y_i$ . The variance for this estimator is given by

$$V(\bar{t}_{str}) = \left(1 - \frac{n'}{N}\right) \frac{S_y^2}{n'} + E \left[ \sum_{h=1}^H \left(\frac{n_h}{n'}\right)^2 \left(1 - \frac{m_h}{n_h}\right) \frac{s_h^2}{m_h} \right], \quad (5.3)$$

where  $S_y^2$  is the population variance of  $y$  and  $s_h^2$  is the sample variance of  $y$  in stratum  $h$ , from phase 1 if we observe them all. This can be estimated by

$$\begin{aligned} \hat{V}(\bar{t}_{str}) &= \frac{N-1}{N} \sum_{h=1}^H \left(\frac{n_h-1}{n'-1} - \frac{m_h-1}{N-1}\right) \frac{n_h}{n'} \frac{s_h^2}{m_h} \\ &\quad + \frac{1}{n'-1} \left(1 - \frac{n'}{N}\right) \sum_{h=1}^H \frac{n_h}{n'} (\bar{y}_h - \bar{t}_{str})^2, \end{aligned} \quad (5.4)$$

where  $s_h^2$  is the sample variance of  $y$  in stratum  $h$  from phase 2 (Rao, 1973).

### 5.3 Stratified Sampling Design with Partially Exchangeable Sequences

Under the model-based approach to sampling, the variance and estimated variance of  $\bar{t}_{str}$  are derived based on the underlying assumption of IID of the random sequence  $\{Y_1, \dots, Y_N\}$  (for which  $\{y_1, \dots, y_N\}$  is a particular realisation) within stratum and between strata. We aim to explore situations where such assumptions may prove to be too restrictive. Although, it may still often be the case that the order in which units are chosen is not important. This leads to a natural generalisation to exchangeable sequences.

We consider observations within the same stratum to be exchangeable and rewrite Theorem 2.12 as follow: Suppose that we can categorise a sequence  $\{Y_1, Y_2, \dots\}$  into  $H$  disjoint exchangeable subsequences and let  $\mathbf{Y}_h$  denote a finite subset of those  $Y_i$ 's that are in subsequence  $h$  (with the index subset denoted by  $S_h$ ). Then, if  $\mathbf{y}_h$  is a realisation of  $\mathbf{Y}_h$ , we have the following representation

$$f_{\mathbf{Y}_1, \dots, \mathbf{Y}_H}(\mathbf{y}_1, \dots, \mathbf{y}_H) = \int_{\Theta} \prod_{h=1}^H \prod_{i \in S_h} f_{Y|\theta_h}(y_i|\theta_h) f_{\theta_1, \dots, \theta_H}(\theta_1, \dots, \theta_H) d\theta_1 \dots d\theta_H, \quad (5.5)$$

where  $\theta_h$  is the set of underlying parameters associated with sequence  $h$ . If we further set  $|S_h| = m_h$ , then we have a scenario analogous to the two-phase stratified sampling in Section 5.2. Here, we consider the sequence of observations in individual strata to be exchangeable and dependencies across strata are characterised by the joint distribution  $f_{\theta_1, \dots, \theta_H}(\theta_1, \dots, \theta_H)$  of the underlying parameter sets.

Under the above assumptions, it remains mathematically feasible to use the estimator  $\bar{t}_{str}$  for  $\bar{y}_U$ . However, the calculation and estimation of  $V(\bar{t}_{str})$  may become more cumbersome. Let  $Z_i$  be the indicator variable on unit  $i$  being selected for the first phase sample and  $\mathbf{Z} = (Z_1, \dots, Z_N)$ . Consequently,

$$\begin{aligned} V(\bar{t}_{str}) &= V(E[\bar{t}_{str}|\mathbf{Z}]) + E(V[\bar{t}_{str}|\mathbf{Z}]) \\ &= V(\bar{t}^{(1)}) + E\left(V\left[\frac{1}{n'} \sum_{h=1}^H n_h \bar{y}_h | \mathbf{Z}\right]\right) \\ &= V(\bar{t}^{(1)}) + E\left(\sum_{h=1}^H \left(\frac{n_h}{n'}\right)^2 V[\bar{y}_h | \mathbf{Z}] + 2 \sum_{a < b} \frac{n_a n_b}{(n')^2} Cov(\bar{y}_a, \bar{y}_b | \mathbf{Z})\right), \end{aligned} \quad (5.6)$$

where  $\bar{t}^{(1)}$  is the sample mean from phase 1, assuming we know  $y_i$  for all  $i \in S'$ . The first term is the variance resulted from phase 1 sampling and the second term is the additional variance resulted from the subsampling in phase 2.

Now, assuming  $Y_1, \dots, Y_N$  are still identically distributed with mean  $\mu$  and variance  $\sigma^2$ , we can write the first term in expression (5.6) as

$$\begin{aligned} V(\bar{t}^{(1)}) &= E\left[(\bar{t}^{(1)} - \bar{y}_U)^2\right] \\ &= E\left[\left(\frac{1}{n'} \sum_{i \in S'} Y_i - \frac{1}{N} \sum_{i \in U} Y_i\right)^2\right] \\ &= E\left[\left(\left(\frac{1}{n'} - \frac{1}{N}\right) \sum_{i \in S'} Y_i - \frac{1}{N} \sum_{i \notin S'} Y_i\right)^2\right] \\ &= E\left[\left(\left(\frac{1}{n'} - \frac{1}{N}\right) \sum_{i \in S'} Y_i - \frac{1}{N} \sum_{i \notin S'} Y_i - \left(\frac{1}{n'} - \frac{1}{N}\right)n'\mu + \frac{1}{N}(N - n')\mu\right)^2\right] \\ &= E\left[\left(\frac{1}{n'} - \frac{1}{N}\right)^2 \left(\sum_{i \in S'} Y_i - n'\mu\right)^2 + \left(\frac{1}{N}\right)^2 \left(\sum_{i \notin S'} Y_i - (N - n')\mu\right)^2\right. \\ &\quad \left. - 2\left(\frac{1}{n'} - \frac{1}{N}\right)\left(\frac{1}{N}\right) \left(\sum_{i \in S'} Y_i - n'\mu\right) \left(\sum_{i \notin S'} Y_i - (N - n')\mu\right)\right] \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1}{n'} - \frac{1}{N}\right)^2 \left[ n' \sigma^2 + 2 \sum_{i,j \in S', i < j} \text{Cov}(Y_i, Y_j) \right] \\
&\quad + \left(\frac{1}{N}\right)^2 \left[ (N - n') \sigma^2 + 2 \sum_{i,j \notin S', i < j} \text{Cov}(Y_i, Y_j) \right] \\
&\quad - 2 \left(\frac{1}{n'} - \frac{1}{N}\right) \left(\frac{1}{N}\right) \sum_{i \in S', j \notin S'} \text{Cov}(Y_i, Y_j). \tag{5.7}
\end{aligned}$$

The subsequent problem is in estimating the covariance terms

$$\sum_{i,j \in S', i < j} \text{Cov}(Y_i, Y_j), \quad \sum_{i,j \notin S', i < j} \text{Cov}(Y_i, Y_j) \quad \text{and} \quad \sum_{i \in S', j \notin S'} \text{Cov}(Y_i, Y_j), \tag{5.8}$$

which incorporates covariances between  $Y_i$ 's from the same stratum and across stratum. Now, for  $i$  and  $j$  in the same stratum, i.e.,  $Y_i$  and  $Y_j$  are exchangeable, we may write (see Remark 2.15)

$$\rho_h := \text{Cov}(Y_i, Y_j) \approx V(E(Y_i | \theta)) = V(E(Y_i | F_{\mathbf{Y}_h})), \tag{5.9}$$

where  $F_{\mathbf{Y}_h}$  is the limiting empirical distribution of  $Y_i$ 's in stratum  $h$ , if  $n_h$  is large and  $m_h/n_h$  is relatively small. We suggest estimating these within-stratum covariance terms using SBS (Politis & Romano, 1994) in each stratum. This is a generalisation to the standard BS, in which data are divided into blocks of random sizes (block sizes following a geometric distribution) and the blocks are resampled to form new samples. For simplicity, we also assume independence across strata (this can also be motivated practically when one agrees that changes in one stratum does effect others, or when such effects are considered minimal). Hence,  $\text{Cov}(Y_i, Y_j) = 0$  for any pair  $i$  and  $j$ , from different strata. This will result in

$$\sum_{i,j \in S', i < j} \text{Cov}(Y_i, Y_j) \approx \sum_{h=1}^H \binom{n_h}{2} \rho_h \tag{5.10}$$

$$\sum_{i,j \notin S', i < j} \text{Cov}(Y_i, Y_j) \approx \sum_{h=1}^H \binom{\lceil n_h(N/n - 1) \rceil}{2} \rho_h \tag{5.11}$$

$$\sum_{i \in S', j \notin S'} \text{Cov}(Y_i, Y_j) \approx \sum_{h=1}^H n_h (\lceil n_h(N/n - 1) \rceil) \rho_h \tag{5.12}$$

where  $\lceil n_h(N/n - 1) \rceil$  is used to approximate  $N_h - n_h$  and given that individuals in an exchangeable sequence behave similarly to each other (allowing us to approximate out-of-sample covariances with in-sample ones). We will also estimate  $\sigma^2$  using the sample variance of all observed  $y$ .

The second term in (5.6), given independence across strata, is equal to

$$\tau := E \left( \sum_{h=1}^H \left(\frac{n_h}{n'}\right)^2 V[\bar{y}_h | \mathbf{Z}] \right). \tag{5.13}$$

This expectation is taken over all values of  $\mathbf{Z}$  and cannot be evaluated given only one sample. Consequently, we propose estimating this expression again by using SBS. Although, the resampling here is taken over the union of  $S_h$ , i.e.,  $m_h$  may change from resample to resample, and within each stratum of the resample (allowing the estimations of  $V[\bar{y}_h|\mathbf{Z}]$ ). Within each resample,  $n_h$  is also estimated by  $m_h n' / \sum m_h$ .

## 5.4 A Numerical Example

To implement our proposed methodology, we consider a practical example using the Australian AIDS survival data set<sup>1</sup>. In all steps where SBS is required, we set the BS parameter optimally to  $p = c^{-1}(n^*)^{-1/3}$  (Politis & Romano, 1994), where  $n^*$  is the size of the sample we are resampling from and  $c$  is set to be 3.15 (as in the previous chapter). The number of bootstrap samples is set to 1000.

The variable of interest  $y$  is the age (years) of patients at diagnosis. This is recorded for 2843 patients across Australia. An auxiliary variable  $x$  is readily available, which indicates the state of origin of each patient:

NSW = New South Wales,

QLD = Queensland,

VIC = Victoria,

Other = all other states.

For our purpose here, let us assume this is our population and we aim to estimate  $\bar{y}$ , the average age of those in the study of interest. However, we do not know the population stratum sizes  $N_h$ . Meanwhile, we undertake the judgment that  $\{Y_1, Y_2, \dots\}$  are independent across strata (states) and are exchangeable within stratum (which may not at all be an unreasonable judgment!)<sup>2</sup>

We draw a sample  $S'$  using SRSWOR in phase 1 (and observe readings on  $x$ ) and subsamples  $S_h$  are drawn from each strata in phase 2 using SRSWOR (and observe readings on  $y$ ). A summary of the sample information is given in Table 5.1. The value of the corresponding two-phase stratified design estimator is given as  $\bar{t}_{str} = 37.78714$ ,

<sup>1</sup>Data by Australian National Centre in HIV Epidemiology and Clinical Research. Available in R package “MASS”.

<sup>2</sup>Intuitively, this means we believe that there exist some kind of dependence amongst members in the same stratum. However, the order in which the readings are observed are not meaningful, i.e., individual identifications are not important. This is certainly a more general assumption than IID and is expected to give more conservative results by incorporating covariances amongst members in the same stratum.

which can be compared to the true population mean  $\bar{y} = 37.40907$ . Sample variances seem to significantly vary across strata.

Table 5.1: Sample information for two-phase sampling on Australian AIDS survival data.

$N$	$n'$	$h$	$n_h$	$m_h$	$\bar{y}_h$	$s_h^2$	$\bar{t}_{str}$
2843	500	NSW	331	200	38.13	118.2142	37.78714
		QLD	40	27	37.7037	218.755	
		VIC	101	68	37.29412	90.30026	
		Other	28	19	35.63158	89.80117	

Table 5.2 records the estimated values for  $\rho_h$  and  $\tau$ . The estimates for  $\rho_h$  are obtained through resampling within each stratum. The value for  $\tau$  is obtained by both resampling the union of  $S_h$  and resampling within the resultant strata.

Table 5.2: Estimated variance and covariance using stationary bootstrapping.

$h$	$\rho_h$	$\tau$
NSW	0.892956	0.3534294
QLD	3.464333	
VIC	1.074241	
Other	1.821655	

Table 5.3: Comparing sampling variance for  $\bar{t}_{str}$ .

Method/Assumption	Variance	Std. Dev.
Rao	0.0006403	0.02530365
Simulated	0.9817275	0.9908216
Exchangeable	0.5336707	0.7391561

The value of  $\hat{V}(\bar{t}_{str})$  (and the corresponding standard deviation), under three different approaches, are presented in Table 5.3. The first estimate is obtained using the formula by Rao (1973), as given in Section 5.2. The second value is obtained from the population data, by re-calculating  $\bar{t}_{str}$  repeatedly using random samples of size 500 and randomised phase 2 sampling ratio (all samples obtained using SRSWOR). This calculation is done for 10000 iterations and the sample variance of  $\bar{t}_{str}$  across iterations is obtained. The formula by Rao (1973) seems to underestimate the variance of  $\bar{t}_{str}$ , potentially due to the assumption of IID observations. Meanwhile, our proposed approach, which caters for within stratum dependencies, produced an improved estimate for the variance (closer

to the simulated variance from the population data).

## 5.5 Discussion and Further Research

In this chapter, we considered a scenario of two-phase stratification sampling design, where observations within strata are assumed to be exchangeable and strata are assumed to be mutually independent. A method is proposed for estimating the variance of the ratio estimator using SBS at various levels of the sampling procedure. An example considered here demonstrates that the standard variance estimate significantly overestimates the performance of the ratio estimator, while our method provided a more conservative approximation.

There are several limitations to our approach that may be generalised or improved. Firstly, we have implemented a very Bayesian-unlike approach, in the sense that we did not specify a prior distribution for  $\theta$ , nor a sampling distribution. More precisely, our method tries to capture the varying effect of  $\theta$  through the bootstrapped samples. This is of course allowing the data to overtake any form of subjective prior information we may have for  $y$ , apart from the observed  $x$  values. Secondly, we have assumed independence across strata. Consequently, all covariances across strata were assumed to be zero. Relaxing this would again relate to specifying or estimating the joint behaviour between  $\theta_h$  in expression (5.5). In addition, the example in Section 5.4 is based on a singular sample we have taken<sup>3</sup> and further simulation is required to observe the overall performance of our method. Further work should be done to compare our method to other general approaches to estimating variance in complex designs (Lohr, 2010).

---

<sup>3</sup>Although we did test the method on a few other samples and have observed similar results.

## Chapter 6

# Conclusions

This thesis has striven to explore several topics related to the concept of exchangeability. These topics involved both theoretical derivations and practical implementations. These ventures are essentially articulated in three main chapters of this thesis. They are precluded by a short chapter providing a review of exchangeability. This encompassed formal definitions, remarks on related characteristics and de Finetti's representation theorems for the various forms of exchangeability. Many of the concepts introduced were used in later chapters. Then, the remainder of the thesis focused on three related areas of study.

Firstly, connections between HMP and exchangeability were revisited. The existence conditions for HMP were generalised to the case over a domain coupling a  $g$  tuple of  $k$ -dimensional simplexes. This was aided by the use of a generalised Bernstein polynomial, Helly's theorem and Helly-Bray theorem. On the other hand, de Finetti's theorem was also extended to cater for a  $g$ -fold partially exchangeable sequence in  $\{0, 1, \dots, k\}$ . The consideration of an urn scheme with  $g$  urns, having  $(k + 1)$  types of items in each urn, was suggested for deriving this representation. The pinnacle of this chapter was to formally state and prove the equivalence between the above two theories. This provided an addition to existing literature, while the derivations used are also different to previous work. This is an exciting mathematical outcome, but also provokes further investigations into topics such as: cross-application between approximation methods for finite moment problems and finite exchangeable sequences; and, relations between unbounded moment problems and general symmetric measures, etc.

EVT has proven to be popular in various areas of application, such as hydrology, earth sciences and structural engineering. This is due to recognitions of the several advantages in implementing an EVT approach. These are comprised of its solitary focus on extremes of the data set (hence minimising the bias caused by rest of the data),



separate investigation of the two tails of the data set (hence catering for asymmetry) and accommodation for heavy-tails. These also make extreme value models pragmatic candidates for estimating financial risk measures. However, a potential drawback lies in the fundamental assumption of IID sequences in the EVT theorems. In Chapter 4, we provided an alternative by extending the existing results (both the BM and the POT methods) to cater for exchangeable sequences. We also constructed the empirical prior distributions of the EVT parameters using resampling techniques (namely, BS, SBS and JK procedures).

Our new approaches were implemented in a rolling window procedure to forecast daily VaR figures. These were compared to existing models in the literature through back-testing against actual observed data in S&P500, FTSE100, MSCI, HSI and ALSI negative daily returns. It was evidenced that the performances of both BM and POT methods generally improved after catering for exchangeable sequences (although the improvements were more pronounced for BM). Correspondingly, the relative performances between the two methods became more comparable (as compared to classical approaches).

In addition, we also amalgamated our new POT approach with the GARCH model. This formed a direct extension to the popular GARCH-EVT model proposed by McNeil & Frey (2000). This again showed significant improvements as compared to other competing models. Suggestions for further work included exploring different subjective priors for the EVT parameters (instead of an empirical one), combining partial exchangeability with change-point analysis to cater for the regime switching nature of the financial series and comparing with, or extending, further advanced EVT models (such as the CAViAR, DPOT and PORT models).

The final topic of interest was on the multiple-phased sampling design with auxiliary variables. In particular, we considered a two-phase sampling design, where the auxiliary variable is categorical. This is identified with a stratified sampling design, with unknown population stratum sizes. It is assumed that the auxiliary (or stratification) variable is observed from the phase 1 sample and readings for the study variable is only drawn at phase 2. This resulted in an estimator for the population mean of the study variable in the form of a generalised ratio estimator. This estimator is essentially a weighted sum of ratio estimators from each stratum. Traditionally, the sampling variance of this estimator is estimated by considering its corresponding asymptotic expression that assumes IID observations.

We extended the above scenario by assuming exchangeability among observations from a common stratum, while retaining independence across strata. The sampling variance is then expressed in terms of the various covariance terms (e.g., in-sample against

out-of-sample, within stratum against crossing strata, etc.). SBS (which preserves the stationarity property of an exchangeable sequence) was used at various levels of sampling to estimate these covariance terms. An example utilising data from the Australian National Centre in HIV Epidemiology and Clinical Research was considered. The empirical results showed that the standard variance estimate significantly overestimate the performance of the ratio estimator, while our new method provided a more conservative approximation.

In conclusion, exchangeability is a powerful and elegant concept of probability that permeates and unifies many real world processes. Its symmetric property, while simplifying, provides a more general view-point than the over-elaborated IID concept. In particular, the underlying framework allows for a more flexible modelling of complex procedures by catering for prior information. Conceptually, the corresponding de Finetti-type results also provide simple passages between the subjectivist and objectivist perceptions. These comments are key factors in the use of exchangeability, for both theoretical studies and applications. We hope this thesis has embroidered the concept and applicability of exchangeability. This was done with the aim to inspire new insights into topics involving exchangeability and provide stimulus for further explorations in future research.

# List of Tables

4.1	Descriptive statistics for daily returns of the five indices. . . . .	36
4.2	Hoeffding's D Statistics for parameters estimated from fitting GEVD to S&P500. . . . .	36
4.3	Hoeffding's D Statistics for parameters estimated from fitting GPD to S&P500. . . . .	37
4.4	Indication of whether the exchangeable EVT approach do improve on the classical unconditional EVT models, in terms of one-day-ahead VaR forecasting. . . . .	38
4.5	Model(s) with the most adequate VaR forecasts for various indices, at different VaR levels. . . . .	39
4.6	One-day-ahead out-of-sample test for VaR estimates of S&P500. . . . .	40
4.7	One-day-ahead out-of-sample test for VaR estimates of FTSE100. . . . .	41
4.8	One-day-ahead out-of-sample test for VaR estimates of MSCI. . . . .	42
4.9	One-day-ahead out-of-sample test for VaR estimates of HSI. . . . .	43
4.10	One-day-ahead out-of-sample test for VaR estimates of ALSI. . . . .	44
4.11	Backtesting of VaR estimates for negative daily returns. . . . .	48
5.1	Sample information for two-phase sampling on Australian AIDS survival data. . . . .	57
5.2	Estimated variance and covariance using stationary bootstrapping. . . . .	57
5.3	Comparing sampling variance for $\bar{t}_{str}$ . . . . .	57

# List of Figures

3.1	Corresponding equivalent sub-cases between HMP and exchangeable sequences. . . . .	23
4.1	Time series plots for daily closing values (left) and returns (right) for S&P500. . . . .	34
4.2	Time series plots for daily closing values (left) and returns (right) for FTSE100. . . . .	35
4.3	Time series plots for daily closing values (left) and returns (right) for MSCI. . . . .	35
4.4	Time series plots for daily closing values (left) and returns (right) for HSI. . . . .	35
4.5	Time series plots for daily closing values (left) and returns (right) for ALSI. . . . .	35
4.6	Scatter plots of GEVD parameters for S&P500, with BS resampling and block size 21. . . . .	36
4.7	Partial ACF plots of GARCH residuals for FTSE100 and the absolute value of GARCH residuals for ALSI. . . . .	45
4.8	Kurtosis of GARCH residuals for rolling windows of 1000 daily returns. . . . .	46
4.9	Skewness of GARCH residuals for rolling windows of 1000 daily returns. . . . .	46
4.10	VaR estimates for negative FTSE100 returns using GARCH-EVT with exchangeable innovations (black line = daily returns; red line = 1% VaR estimates; blue line = 2.5% VaR estimates). . . . .	46
4.11	VaR estimates for negative ALSI returns using GARCH-EVT with exchangeable innovations (black line = daily returns; red line = 1% VaR estimates; blue line = 2.5% VaR estimates). . . . .	47

# Bibliography

- Aas, K. & Haff, D. H. (2006). The generalized hyperbolic skew Student's  $t$ -distribution. *Journal of Financial Econometrics*, 4(2), 275–309.
- Ahmed, M. S. (1998). A note on regression-type estimators using multiple auxiliary information. *Australian & New Zealand Journal of Statistics*, 40(3), 373–376.
- Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4), 581–598.
- Aldous, D. J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII–1983, Lecture Notes in Mathematics*, volume 1117 (pp. 1–198). Berlin: Springer-Verlag.
- Aldous, D. J. (2010). More uses of exchangeability: Representations of complex random structures. In N. H. Bingham & C. M. Goldie (Eds.), *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman* (pp. 35–63). Cambridge: Cambridge University Press.
- Bali, T. G. (2003). An extreme value approach to estimating volatility and value at risk. *The Journal of Business*, 76(1), 83–108.
- Balkema, A. & de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2(5), 792–804.
- Bernardo, J. M. (1996). The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences*, Spec. II, 111–121.
- Bernardo, J. M. & Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.
- Bhushan, S., Pandey, A., & Katara, S. (2008). A class of estimators in double sampling using two auxiliary variables. *Journal of Reliability and Statistical Studies*, 1(1), 67–73.

- Bowman, D. & George, E. O. (1995). A saturated model for analyzing exchangeable binary data: Applications to clinical and developmental toxicity studies. *Journal of the American Statistical Association*, 90(431), 871–879.
- Brent, R. (1973). *Algorithms for Minimization without Derivatives*. New Jersey: Prentice-Hall.
- Byström, H. N. E. (2004). Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis*, 13(2), 133–152.
- Chavez-Demoulin, V., Embrechts, P., & Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 181(1), 44–52.
- Chinhamu, K., Huang, C.-K., Huang, C.-S., & Hammujuddy, M. J. (2015). Empirical analyses of extreme value models for the South African Mining Index. *South African Journal of Economics*, 83(1), 41–55.
- Chow, Y. S. & Teicher, H. (1988). *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer-Verlag.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862.
- Cochran, W. G. (1977). *Sampling Techniques*. 3rd edition. New York: Wiley.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- Dale, A. I. (1983). A probabilistic proof of Hausdorff's theorem for double sequences. *Sankhyā: The Indian Journal of Statistics, Series A*, 45(3), 391–394.
- Dale, A. I. (1987). Two-dimensional moment problems. *The Mathematical Scientist*, 12(1), 21–29.
- de Finetti, B. (1930). Funzione caratteristica di un fenomeno aleatorio. *Atti Reale Accademia Nazionale dei Lincei*, Mem. 4, 86–133.
- de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7(1), 1–68.
- de Finetti, B. (1938). Sur la condition d'équivalence partielle. *Actualités Scientifiques et Industrielles*, 739, 5–18.

- de Finetti, B. (1974). *Theory of Probability*, volume II. New York: Wiley.
- de Haan, L., Mercadier, C., & Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance Stochastics*, 20(2), 321–354.
- Diaconis, P. (1977). Finite forms of de Finetti’s theorem on exchangeability. *Synthese*, 36(2), 271–281.
- Diaconis, P. & Freedman, D. (1980). Finite exchangeable sequences. *The Annals of Probability*, 8(4), 745–764.
- Diaconis, P. & Freedman, D. (1993). Nonparametric binary regression: A Bayesian approach. *The Annals of Statistics*, 21(4), 2108–2137.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance*. Berlin: Springer-Verlag.
- Engel, R. F. & Manganelli, S. (2004). CAViAR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, 2(4), 367–381.
- Ericson, W. A. (1969). Subjective Bayesian models in sampling finite population. *Journal of the Royal Statistical Society, Series B*, 31(2), 195–233.
- Fama, E. (1965). The behavior of stock market prices. *Journal of Business*, 38(1), 34–105.
- Faranda, D., Lucarini, V., Turchetti, G., & Vaienti, S. (2011). Numerical convergence of the block-maxima approach to the generalized extreme value distribution. *Journal of Statistical Physics*, 145(5), 1156–1180.
- Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, volume II. New York: Wiley.
- Fisher, R. A. & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24 (pp. 180–190).
- Freedman, D. A. (1962). Invariants under mixing which generalize de Finetti’s theorem. *The Annals of Mathematical Statistics*, 33(3), 916–923.
- Galambos, J. (1987). *The asymptotic theory of extreme order statistics*. 2nd edition. Melbourne, Florida: Krieger.

- Gencay, R. & Selcuk, F. (2004). Extreme value theory and value-at-risk: relative performance in emerging markets. *International Journal of Forecasting*, 20(2), 287–303.
- George, E. O. & Kodell, R. L. (1996). Tests of independence, treatment heterogeneity dose-related trend with exchangeable binary data. *Journal of the American Statistical Association*, 91(436), 1602–1610.
- Gilli, M. & K ellezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(2), 207–228.
- Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une s erie al eatoire. *Annals of Mathematics*, 44(3), 423–453.
- Gnedin, A. V. (1995). On a class of exchangeable sequences. *Statistics & Probability Letters*, 25(4), 351–355.
- Gupta, J. C. (1999a). The moment problem for the standard  $k$ -dimensional simplex. *Sankhy : The Indian Journal of Statistics, Series A*, 61(2), 286–291.
- Gupta, J. C. (1999b). Partial Hausdorff sequences and symmetric probabilities on finite products of  $\{0,1\}$ . *Sankhy : The Indian Journal of Statistics, Series A*, 61(3), 347–357.
- Haag, J. (1924). Sur un probl eme g en eral de probabilit es et ses diverses applications. In *Proceedings of the International Congress of Mathematicians* (pp. 659–674).
- Hamad, N., Hanif, M., & Haider, N. (2013). A regression type estimator with two auxiliary variables for two-phase sampling. *Open Journal of Statistics*, 3, 74–78.
- Hausdorff, F. (1923). Momentprobleme f ur ein endliches intervall. *Mathematische Zeitschrift*, 16(1), 220–248.
- Heath, D. & Sudderth, W. (1976). De Finetti's theorem on exchangeable random variables. *The American Statistician*, 30(4), 188–189.
- Hill, B. M. (1994). Bayesian forecasting of extreme values in an exchangeable sequence. *Journal of Research of the National Institute of Standards and Technology*, 99(4), 521–538.
- Ho, L., Burridge, P., Cadle, J., & Theobald, M. (2000). Value-at-risk: applying the extreme value approach to Asian markets in the recent financial turmoil. *Pacific-Basin Finance Journal*, 8(2), 249–275.



- Hollander, M. & Wolfe, D. A. (1973). *Nonparametric statistical methods*. New York: Wiley.
- Hoover, D. (1979). Relations on probability spaces and arrays of random variables. Preprint, Institute for Advanced Study, Princeton, NJ.
- Huang, C.-K. (2014). On exchangeability and Hausdorff moment problems over  $k$ -dimensional simplexes. *South African Statistical Journal*, 48(1), 73–81.
- Huang, C.-K., North, D., & Zewotir, T. (2016a). A conditional extreme value approach to value-at-risk estimation with exchangeable innovations. In *Proceedings of the 58th Annual Conference of the South African Statistical Association* (pp. 33–40).
- Huang, C.-K., North, D., & Zewotir, T. (2016b). Sampling variance of a multi-phase stratified design on partially exchangeable sequences. In *Proceedings of the 58th Annual Conference of the South African Statistical Association* (pp. 41–48).
- Huang, C.-K., North, D., & Zewotir, T. (2017). Exchangeability, extreme returns and value-at-risk forecasts. *Physica A: Statistical Mechanics and its Applications*, 477, 204–216.
- Huang, C.-S., Huang, C.-K., & Chinhamu, K. (2014). Assessing the relative performance of heavy-tailed distributions: empirical evidence from the Johannesburg Stock Exchange. *Journal of Applied Business Research*, 30(4), 1263–1286.
- Inglese, G. (1995). Christoffel functions and finite moment problems. *Inverse Problems*, 11(4), 949–960.
- Jansen, D. W. & de Vries, C. G. (1991). On the frequency of large stock returns: putting booms and busts into perspective. *The Review of Economics and Statistics*, 73(1), 18–24.
- Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*. New York: McGraw-Hill.
- Kallenberg, O. (2005). *Probabilistic Symmetries and Invariance Principles*. New York: Springer.
- King, B. F. (1966). Market and industry factors in stock price behavior. *Journal of Business*, 39(1), 139–190.
- Kingman, J. F. C. (1978). Uses of exchangeability. *The Annals of Probability*, 6(2), 183–197.

- Koedijk, K. G., Schafgans, M. M. A., & de Vries, C. G. (1990). The tail index of exchange rate returns. *Journal of International Economics*, 29(1-2), 93–108.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 2, 173–184.
- Lohr, S. L. (2010). *Sampling: Design and Analysis*. 2nd edition. Boston: Brooks/Cole.
- Longin, F. M. (1996). The asymptotic distribution of extreme stock market returns. *The Journal of Business*, 69(3), 383–408.
- Longin, F. M. (2000). From value at risk to stress testing: the extreme value approach. *Journal of Banking & Finance*, 24(7), 1097–1130.
- McCall, J. J. (1991). Exchangeability and its economic applications. *Journal of Economic Dynamics and Control*, 15(3), 549–568.
- McNeil, A. (1998). Calculating quantile risk measures for financial return series using extreme value theory. Preprint, ETH Zürich.
- McNeil, A. & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4), 271–300.
- Mukerjee, R., Rao, T. J., & Vijayan, K. (1987). Regression type estimators using multiple auxiliary information. *Australian Journal of Statistics*, 29(3), 244–254.
- Patton, A., Politis, D. N., & White, H. (2009). Correction to “Automatic block-length selection for the dependent bootstrap” by D. Politis and H. White. *Econometric Reviews*, 28(4), 372–375.
- Peng, H., Dang, X., & Wang, X. (2010). The distribution of partially exchangeable random variables. *Statistics & Probability Letters*, 80(11-12), 932–938.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1), 119–131.
- Politis, D. N. & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428), 1303–1313.
- Politis, D. N. & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1), 53–70.
- Rao, J. N. K. (1973). On double sampling for stratification and analytical surveys. *Biometrika*, 60(1), 125–133.

- Santos, P. A., Alves, I. F., & Hammoudeh, S. (2013). High quantiles estimation with Quasi-PORT and DPOT: An application to value-at-risk for financial variables. *The North American Journal of Economics and Finance*, 26, 487–496.
- Singh, V. K., Singh, G. N., & Shukla, D. (1994). A class of chain ratio type estimators with two auxiliary variables under double sampling scheme. *Sankhyā: The Indian Journal of Statistics, Series B*, 56(2), 209–221.
- Smith, R. L. & Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society - Series B*, 56(3), 515–528.
- Stockbridge, R. H. (2003). The problem of moments on polytopes and other bounded regions. *Journal of Mathematical Analysis and Applications*, 285(2), 356–375.
- Talenti, G. (1987). Recovering a function from a finite number of moments. *Inverse Problems*, 3(3), 501–517.
- Tan, F., Rayner, G. J., Wang, X., & Peng, H. (2010). A full likelihood procedure of exchangeable negative binomials for modelling correlated and overdispersed count data. *Journal of Statistical Planning and Inference*, 140(10), 2849–2859.
- Treder, R. P. & Sedransk, J. (1996). Bayesian sequential two-phase sampling. *Journal of the American Statistical Association*, 91(434), 782–790.
- Tsay, R. S. (2010). *Analysis of Financial Time Series*. 3rd edition. New Jersey: Wiley & Sons.
- Tucker, H. G. (1967). *A Graduate Course in Probability*. New York: Academic Press.
- Zhao, X., Scarrott, C. J., Oxley, L., & Reale, M. (2011). Garch dependence in extreme value models with Bayesian inference. *Mathematics and Computers in Simulation*, 81(7), 1430–1440.