Return Models and Dynamic Asset Allocation Strategies

Wyanet Wen Shi

Doctor of Philosophy

A thesis submitted for the degree of Doctor of Philosophy at
Monash University in 2015 with
the Department of Econometircs and Business Statistics

Copyright Notice

Abstract

This thesis studies the design of optimal investment strategies. A strategy is considered optimal when it minimizes the variance of terminal portfolio wealth for a given level of expected terminal portfolio wealth, or equivalently, maximizes an investor's utility. We study this issue in two particular situations: when asset returns follow a continuous-time path-independent process, and when they follow a discrete-time path-dependent process.

Continuous-time path-independent return models are popular but controversial in the literature. We formulate the criteria for portfolio rules to be considered as optimal in this framework. We construct a portfolio consisting of a risky and a risk-free asset where the return on the risky asset follows a Gaussian diffusion process with non-constant drift and diffusion. Portfolio rules satisfying the specified criteria are shown to be compatible with the objective of utility maximization.

Discrete-time path-dependent return models are more realistic, given the fact that almost all historical return data are measured in discrete time and exhibit serial correlation. Hence, we develop a novel methodology to assess the market efficiency and predictability of Australian index returns on equities, debts and cash, which we show are path-dependent empirically. We propose a one-step Multivariate Semi-parametric Maximum Likelihood Estimation (one-step MSMLE) technique to estimate a Vector-autoregressive Multivariate Generalized Autoregressive Conditional Heteroskedasticity (VAR-MGARCH) model of Australian asset returns. The estimation is done using a "rolling historical window" approach so as to highlight and capture path-dependency in asset returns as well as allow for parameter changes. Serial correlation is found in both the return and the volatility levels of the Australian assets that we consider. Having shown this, we then extend a class of reactive portfolio controls to the case when returns follow a VAR-MGARCH process. The portfolio controls are formulated by solving the Lagrangian which minimizes the variance in next period wealth for a given targeted next period wealth. We quantitatively demonstrate that this class of reactive portfolio controls are efficient, even under the existence of market impacts.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes one original paper published in a peer reviewed journal and two unpublished publications (a conference paper presented in the Australasian Actuarial Education and Research Symposium 2012, and a conference paper accepted by ICSEM 2015, the 17th International Conference on Statistics). The core theme of the thesis is to study the design of optimal investment strategies. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the Doctor of Philosophy under the supervision of Dr. Andrew Leung, Prof. Heather Anderson and A.Prof. Paul Lajbcygier.

In the case of Chapter 2 my contribution to the work involved the following:

Thesis Chapter	Publication Titlel	Publication Status	Nature and Extent of Students Contribution
2	An extension of some results due to Cox and Leland	Publised	60%

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature: 师要

Date: 15/03/2016

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student and the co-author's contributions to this work.

Main Supervisor signature: [redacted]

Date: 15/3/2016

Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. Andrew Leung and Prof. Heather Anderson for providing motivation when required, patience when needed and their expertise. Dr. Andrew Leung provided primary support to the first half of my Ph.D. Prof. Heather Anderson joined later Dr. Andrew Leung resigned, and provided primary support to the second half of my Ph.D. I would also like to thank my associate supervisor A.Prof. Paul Lajbcygier for the continuous inspiration he provided. It can be said with little controversy that my Ph.D would not be what it is without the help of Dr. Leung, Prof. Anderson and A.Prof. Lajbcygier.

Besides my supervisors, I would like to thank the panel members on my final review dissertation, Prof. David Harris, A.Prof. Colin O'Hare and Dr. David Frazier, for their insightful comments and encouragement which inspired me to improve my research from various perspectives. I am also greatly indebted to Prof. Jiti Gao who provided me with profusely detailed and disciplined feedback in his field.

I would like to thank my family and friends for supporting me spiritually throughout writing my thesis and my life in general. There are too many to name here, but I am mostly indebted to my parents. Without them and their precious support it would not be possible to conduct this research.

I know as I finish writing these lines that my fantastic time at Monash will not be an ending but rather a beginning - gate to new experiences. I hope that my thesis will be the first step, however humble, towards a hard-working, yet exciting, career where I will strive to contribute meaningfully to the field and society in general.

Contents

List of Figures

List of Tables

Abbreviations

AB .... **A**symptotic **B**ias

AISB .... **A**symptotic **I**ntegrated **S**quared **B**ias

AIV .... **A**symptotic **I**ntegrated **V**ariance

AISB .... **A**symptotic **M**ean **I**ntegrated **S**quared **E**rror

AR .... **A**uto**r**egression (**A**uto**r**egressive)

BAB .... **B**ank **A**ccepted **B**ill

BEKK .... **B**aba, **E**ngle, **K**raft and **K**roner

CRRA .... **C**onstant **R**elative **R**isk **A**version

GARCH .... **G**eneralized **A**utoregressive **C**onditional **H**eteroskedasticity

GFC .... **G**lobal **F**inancial **C**risis

HARA .... **H**yperbolic **A**bsolute **R**isk **A**version

HJB .... **H**amilton-**J**acobi-**B**ellman

i.i.d .... **i**ndependently and **i**dentically **d**istributed

ISB .... **I**ntegrated **S**quared **B**ias

ISE .... **I**ntegrated **S**quared **E**rror

IV .... **I**ntegrated **V**ariance

MGARCH .... **M**ultivariate-**GARCH**

MLE .... **M**aximum **L**ikelihood **E**stimation (**E**stimator)

MISB .... **M**ean **I**ntegrated **S**quared **E**rror

MSE .... **M**ean **S**quared **E**rror

ODE .... **O**rdinary **D**ifferential **E**quation

OLS .... **O**rdinary **L**east **S**quares

one-step SMLE ... **one-step** **S**emiparametric **M**aximum **L**ikelihood **E**stimation (**E**stimator)

one-step MSMLE . **one-step** **M**ultivariate **S**emiparametric **M**aximum **L**ikelihood **E**stimation (**E**stimator)

PDE .... **P**artial **D**ifferential **E**quation

QMLE .... **Q**uasi - **M**aximum **L**ikelihood **E**stimation (**E**stimator)

RBA .... **R**eserve **B**ank of **A**ustralia

SMLE .... **S**emiparametric **M**aximum **L**ikelihood **E**stimation (**E**stimator)

VAR .... **V**ector-**A**utoregression (**A**utoregressive)

Important Symbols and Notations

t .... Time index

W .... Total portfolio wealth

G .... Amount invisted in a risky asset

H .... Amount invisted in a risk-free asset

K .... Amount withdrawn

s .... Price of a risky asset which has path-independent returns

$B(t)$ .... A Brownian motion

μ .... Drift in the path-independent returns on a risky asset

σ .... Diffusion in the path-independent returns on a risky asset

$r(t)$ .... The riskfree rate

Pf .... Price of the riskfree asset

$u(.)$ .... A utility function

$U(.)$ .... Expected utility of total wealth

$c(.)$ .... Utility from consumption

π .... The density for the physical stock price process

$\bar{\pi}(.)$ .... The density for the risk neutral stock price process

$O(.)$ .... An optimal expected value function

$\hat{f}_n(.)$ .... The density estimate

$f(.)$ .... The true density of multivariate returns

$K(., h)$ .... Kernel function given bandwidth h

$\boldsymbol{\theta}$ .... The parameters in an MGARCH model

Θ .... A compact parameter space

Ω .... The sample space

$F(.)$ .... The information set F is the sigma-algebra generated by Ω

$\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ .... The covariance of the multivariate time series governed by parameters $\boldsymbol{\theta}$

$\mathbf{I}_d$ .... A $d \times d$ identity matrix

$\mathbf{w}_t$ .... ortfolio weights

$\lambda_t^{(EF)}$ .... The positive Lagrangian Multiplier by minimizing variance in the next period wealth for a given expected level of next period wealth

$\lambda_t^{(N)}$ .... The positive Lagrangian Multiplier by minimizing variance in the terminal wealth for a given expected level of terminal wealth

λ_t .... The positive Lagrangian Multiplier by minimizing variance in the terminal wealth for a given expected level of terminal wealth for a given return model

Chapter 1

Introduction

Money has a diminishing time value. A dollar received some time in the future is worth less than a dollar received today. This idea has been universally accepted in finance literature. People prefer to spend money today rather than spending the same amount in the future. Hence, whether to forgo consumption today and invest the money depends on the return one expects to earn from the investment. The higher the expected return, the more likely one will choose to invest instead of spending the money today (Parrino et al. 2014). Therefore, a common objective of an investor is to maximize his investment returns at his accepted level of risk, or equivalently, to minimize risks at his desired level of expected return so as to maximize his utility. Most people invest in portfolios which consist of multiple classes assets. The advantage of creating a portfolio with mixed assets is that it diversifies risks. The impact of a decline in one asset on the entire portfolio is limited at any point in time. An investment portfolio normally comprises at least three asset classes, the principal ones being equity, debt and cash.

Equities represent ownership in a company. Returns on equities can be high but volatile, depending on business performance. Debt instruments generally have lower returns but are less volatile than equities. One distinctive characteristic of debt instruments is their term to maturity. Debt instruments have very different term structures; for instance, they can be in short-term as a ninety-day bank accepted bill, or in long-term as a thirty-year government bond. Cash and cash equivalent securities (easily liquidated assets that can be converted into cash immediately) provide liquidity in a portfolio to purchase new equity or debt. They are transitory assets. Holding cash does not always attract interest, hence funds do not normally reserve a large cash balance. Return on cash refers to the interest earned by deposits in commercial banks. Decision on how much an investor should allocate his fund in each asset requires comprehensive analysis and judgement of each asset class's performance. Generally, there are three main types of asset allocation strategies implemented over time. These are to adopt a fixed asset allocation; to adopt an allocation that evolves according to a fixed and pre-determined schedule; and to adopt a dynamic allocation that is, at any future point of time, not pre-determined, but determined according to the actual investment experience up to that time. We will review some examples comparing these allocations strategies below. Merton (1969) showed that a fixed asset allocation (the first type of strategies mentioned) would be optimal only if asset returns were independently identically distributed (i.i.d.), the utility function was a Constant Relative Risk Aversion (CRRA) power function and there were no transaction costs. His claim was supported and extended by Samuelson (1969). Samuelson (1969), in his landmark paper, investigated the question of how much of an individual's total wealth should be consumed or be invested at any point in time, assuming that no bequest was to be left behind. He utilized a backward dynamic programming recursion to prove that a fixed allocation between a risky and a risk-free asset would be optimal when returns on the risky asset were i.i.d. over time, income was generated only from investments, and consumption was evaluated by using a power utility function over an individual's lifetime. An example of a comparison between a fixed allocation (the first type of strategy mentioned) and a pre-scheduled allocation (the second type of strategy mentioned) was provided in a U.S. patent taken out by Frain and Gallo (2011). They pointed out that many investors failed to change their asset allocation strategies and these strategies become inconsistent with their risk tolerance levels when certain "trigger events" happened. The patent proposed a systematic process to adjust asset allocation rules.Bone and Goddard (2009) provided a quantitative comparison between a fixed proportion allocation (the first type of strategy mentioned) and a contrarian rebalancing allocation (an example of the third type of strategy mentioned). Their contrarian rebalancing rule allocated 75% to equity when the last period's return in equity is higher than the return in debt, and 75% to debt when the opposite occurs. They concluded that, compared to the fixed proportion allocation, the contrarian rebalancing strategy attained a closer terminal wealth to the targeted value. This result may be explained intuitively by the fact that asset returns and volatilities are not constant, and the fixed proportion allocation strategy does not adjust for this flexibility in returns and volatilities.Basu and Drew (2009) demonstrated the risk-reduction advantage of dynamic asset allocations (the third type of strategies mentioned) by constructing a 'contrarian' strategy of weighting towards defensive assets when experience was favorable. They also suggested that most people become more risk-averse as they age, and hence an investor's portfolio should become increasingly conservative. Asset allocation decisions should be made carefully since the size of the portfolio at different stages of the life-cycle substantially influences the final outcome. Each of the three examples listed above (Frain and Gallo (2011), Bone and Goddard (2009) and Basu and Drew (2009)) show that dynamic asset allocations (the third type of strategy mentioned) can perform better than fixed allocations (the first type) or pre-scheduled allocations (the second type), in the sense that the associated risks in the portfolios are smaller while targeted portfolio returns are better maintained. Therefore, research attention has gradually focused on dynamic investment strategies that follow the third approach listed above, which have the potential to reduce risk. This is consistent with an investor's goal to choose the most suitable portfolio wealth distribution given his risk tolerance level. The above discussion is nicely summarized by Trippi and Harriff (1991), who define dynamic asset allocation as 'a class of investment strategies that shifts the content of portfolios between two or more asset classes in response to either changes in the value of the portfolio and/or external economic states, on a more or less continual basis'. The motivation is two-fold: to tailor the distribution of portfolio returns at some future time so that it can be entirely different from that of the market index; and, to exploit predictable regularities including market timing and other tactical allocation strategies. In other words, an investor's goal is to control the distribution of portfolio returns according to his risk tolerance, and to predict and react to changes in the external economy. This thesis focuses on the explicit formulation of dynamic asset allocation under certain situations. Predictability of and reaction to macroeconomic changes are left for future work.

1.1 Considerations in Dynamic Strategies

Explicit formulation of a dynamic investment strategy has three considerations: the definition of the optimization objectives, the expectation of future asset cash flows, and the choice of an optimization approach. We will elaborate on each of these three aspects below. An investment portfolio can drift from its target asset allocation, acquiring risk and return characteristics that may be inconsistent with an investor's goals and preferences. As a portfolio's expected return increases, so does its vulnerability to deviations from the targeted value. The trade-off between higher returns and higher risks, or lower returns and lower risks, can be thought of as a rebalancing frontier. This defines the optimization objectives (the first consideration mentioned). On the other hand, risk measurements such as the standard deviation (or volatility) of portfolio returns, the Sharpe ratio (Sharpe 1994) and Value at Risk (Jorion 2006), are measures describing different aspects of distributions of portfolio returns or wealth. Maximizing the expected utility of an investor is equivalent to choosing a distribution of portfolio wealth from all possibilities with respect to the desired risk measure. This thesis adopts a standard approach which chooses to minimize risks by minimizing the volatility of portfolio return for a desired level of portfolio wealth. The evolution of portfolio wealth depends on the evolution of asset returns (the second consideration mentioned above). Path-independent return models are popular in academic research, for their mathematical convenience. Path-independence refers to a situation in which current or future states of return do not depend on past states; examples include returns being i.i.d. over time, and the distribution of returns being a function of time only. However, it is widely recognized that serial correlation often exists in asset returns. A more practical approach is to apply path-dependent return models, that is, past experience has implications for current and future returns. Although not very realistic, studies based on path-independent return models do provide some insights of the formulation of optimal portfolio allocations. We will provide a detailed review of return models in Section 1.2.1. The main approaches to dynamic asset allocation (the third consideration mentioned above) rely on analytical solutions, stochastic programming and dynamic programming. We now explain and provide examples of each of these approaches in the remainder of this section. Direct application of an analytical solution (the first approach listed) to a class of optimal strategies is not always easy, especially when the functional form of an investor's utility is not explicitly specified. Popular and relevant tools in the maximization of utility involve, but are not limited to, the Hamilton-Jacobi-Bellman (HJB) equation (Bellman 1957), calculus of variations (Gelfand and Fomin 1963) and the Feynmann-Kac formula (Kac 1949). Examples of analytical solutions provided by the HJB equation to the maximization of particular forms of utility functions can be found in the following studies: a Hyperbolic Absolute Risk Aversion (HARA) utility was considered in Merton (1971), an iso-elastic utility was considered in Brennan et al. (1997), and a Constant Relative Risk Aversion (CRRA) power utility was considered in Cvitanic et al. (2008). Examples of applications of the Feynman-Kac formula to maximize given utility functions can be found in two recent papers; Tanaka (2009) and Chiarella et al. (2007). However, the application of the Feynman-Kac formula is subject to certain conditions which impose limitations on the resulting portfolio controls. We will investigate these conditions in Chapter 2. The single-period mean-variance framework proposed by Markowitz (1952) is widely considered to be the starting point of modern research on portfolio optimization, and stochastic programming (the second approach listed) for financial optimization is considered to be a practical multi-period extension to this. Stochastic programming employs mathematical programming in which historical asset returns are viewed as realizations of an assumed distribution of returns, and solutions can either be analytical or empirical. It is an effective tool to deal with practical issues such as transaction costs, market incompleteness, taxes, trading limits and regulatory restrictions. Yu et al. (2003) and Dupačová (1999) provide summaries of examples of stochastic programming in financial optimization which include asset allocations for pension plans and insurance companies, security selections for portfolio managers and currency hedging. Dynamic programming (the third approach listed) considers possible changes in asset returns or portfolio wealths period by period, and solutions are empirical; for example, see Musumeci and Musumeci (1999) and Samuelson (1969).

1.2 Practical Issues

1.2.1 Asset Return Process

Perhaps the most important consideration in the formulation of an investment strategy is our expectation of future returns. We will review some popular models and methods in the analysis of returns below.

Diffusion Process

Merton (1971) was the first to introduce the application of Itô's Lemma and the Fundamental Theorem of Stochastic Dynamic Programming into the systematic construction and analysis of optimal continuous-time dynamic models under uncertainty. He considered two types of stochastic processes, Brownian motion and Poisson processes. Major advantages of his work are its mathematical simplicity as the number of parameters involved is limited, and the relevance of continuous-time models for actual applications. Modelling asset returns based on diffusion processes assumes that investors' decision making is primarily affected by the portfolio's expected return and variance. This idea of focusing on the first two moments of portfolio wealth is widely accepted. However, empirical evidence has shown that asset returns are not normally distributed in practice, and hence the mean and variance do not fully characterize the distribution of portfolio wealth. The development of portfolio allocation theories for non-Gaussian markets has always been challenging, and has generally met limited success. Despite the historical fact that returns (or logarithmic returns) are not normally distributed, Gaussian diffusion processes have been standard for modeling asset returns since Merton's work in 1971. Dynamic asset allocations based on diffusion processes have been prominent. Examples include Black and Scholes (1973), Cox and Ross (1976) and Carr et al. (2002). Hodges and Clarkson (1994) provided a survey of this topic covering the period up to the early 1990s. On the other hand, Merton (1971) used an arbitrary 'bequest function' as a boundary condition in utility maximization, but this idea was criticized by Taksar and Sethi (1988). Taksar and Sethi (1988) pointed out that the boundary behavior around zero terminal wealth might be inconsistent with Merton's 'bequest function'.

Constant Return Characteristics and Path-independence

Cox and Leland (2000) derived criteria for controls that optimize an investor's objectives when returns are path-independent, basing their work on Merton (1969) and Merton (1971). Their study was conducted on a portfolio that consisted of a risky and a risk-free asset, where returns on the risky asset followed a geometric Brownian motion with constant drift and diffusion. Their restriction to a single risky asset involves no significant loss of generality according to the separation theorem of mutual funds (Ross 1978). However, their restriction to the consideration of assets with constant characteristics is not realistic. In addition, their use of the discrete-time binomial model, converging in continuous-time as the time interval shrinks to zero, is cumbersome and detracts from the economics of the issue. Although Cox and Leland (2000) made several limiting assumptions in their work, their results have been extensively cited in the literature, especially in the studies of hedge funds, because the payoff function of hedge funds is naturally path-independent and non-decreasing in the index values (Amin and Kat 2003). Hodges and Clarkson (1994) provided a numerical example comparing different dynamic strategies and claimed that, although the results presented by Cox and Leland (2000) were not well known at that time, path-independence of a strategy is often necessary for a dynamic strategy to be optimal. Vanduffel et al. (2009) extended the relevance of path-independence to the case when prices of risky assets follow an exponential Lévy process. However, path-independent strategies are not always attractive. Kassberger and Liebmann (2012) show that path-dependent strategies are suboptimal for risk-averse investors when the pricing model is a function of the risky asset price at the terminal time, and not surprisingly, path-dependent strategies are preferred if the pricing model of the risky asset is itself path-dependent.

Non-constant Return Characteristics and Path-dependence

Most of the previous studies on investment strategies, like Cox and Leland (2000), are based on the assumption that return processes have constant drift and diffusion. Although this assumption is idealistic, it simplifies the problem and makes the formulation of an optimal portfolio allocation more feasible in reality. Data generating processes of prices of / returns on financial instruments are far more complex. They are often path-dependent, and do not necessarily reflect constant parameters. Models capturing path-dependent asset returns have drawn more and more attention in recent years. One example is the stochastic volatility model. Detailed references for stochastic diffusion processes may be found in Ibe (2009) and Klebaner (2005). One observation that reveals the path-dependency in returns is serial correlation in both asset returns and their volatility. Many models that capture the path-dependency of time series have been developed over the last few decades. Perhaps the most successful and now standard model in the academic literature is the Autoregressive (AR) model for levels (see Yule (1927) and Walker (1931)), and a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model for volatility (see Bollerslev (1986) and Bollerslev (1987)). Examples of asset allocation strategies based on AR return models can be found in Huang and Lee (2010) and Leung (2011). Huang and Lee (2010) adopted an AR model of asset returns when formulating explicit functional forms of the first two moments for accumulated portfolio wealth. They designed an optimal allocation utilizing a numerical approach, and exploited the advantage obtained from approximating an analytical method. Leung (2011) proposed a reactive investment strategy with no restrictions on the path-dependency in the return level. His study provided a method to locate, on the efficient frontier, the corresponding portfolio weights, for a given target portfolio return at each scheduled time. He demonstrated the efficiency of the proposed strategy for a Vector-autoregressive (VAR) return model. In addition to the AR and GARCH effects on the first two moments of asset returns, some researchers suggest that higher moments on total wealth should be taken into account, given the non-normality of returns. Cvitanic et al. (2008) investigated the effect of higher order moments on the optimal investment strategy of a risk-averse investor. They employed a dynamic jump diffusion model, which enabled them to study the third and the fourth moments of the distribution of portfolio wealth, as the effects of higher order moments often arise naturally when there are jumps (that is discontinuities) in returns. They provided tractable, closed-form and inter-temporal portfolio rules for investors with CRRA utility. They showed that observed skewness and kurtosis would lead to lower holdings in risky assets than the standard Merton (1971) model would imply, and a higher chance of overinvestment (attempting to diversify risks by investing in too many projects). Other examples of return models with jumps can be found in Liu et al. (2003) and Das and Uppal (2004).

Continuous Time v. Discrete Time

There is a rich financial literature that employs both continuous-time and discrete-time models. The Gaussian diffusion, stochastic volatility and jump diffusion models mentioned above are examples of continuous-time models. The AR-GARCH model provides an example of discrete time models. Continuous-time models often have elegant mathematical representations and hence they often have compact analytical solutions. Results derived from the discrete time models can be shown to converge to corresponding continuous-time counterparts in most cases, by taking limits as the time interval shrinks to zero. However, the fact that almost all available financial data used by practitioners are measured in discrete time whereas theoretical models are in continuous time raises questions about the compatibility of the theoretical model and empirical application. In addition, analytical solutions to continuous time models often require restrictive assumptions that are not supported in reality. The constant asset characteristics assumption is an example. In contrast, empirical solutions to discrete time models can mitigate the use of restrictive assumptions.

Estimating Asset Returns

AR-GARCH models are often estimated via the Maximum Likelihood Estimation (MLE) approach, but parametric MLE techniques are consistent and efficient only if the true model and distribution of variables are correctly specified. Di and Gangopadhyay (2013) proposed a one-step Semiparametric MLE (one-step SMLE) technique that utilized kernel density estimates. This technique provides a plausible solution to the distribution specification problem, although estimation will no longer be as efficient. On the other hand, parameters in time series are expected to change over a long time horizon. This viewpoint motivates a rolling window estimation approach. Examples of parameter changes in return models can be found in a rolling window study on New York Stock Exchange common stocks conducted by Fama and MacBeth (1973) and in a rolling window study on the UK stock market conducted by Pesaran and Timmermann (2000). The evolution of parameters has also been studied at the volatility level. Examples may be found in Foster and Nelson (1996) and Andreou and Ghysels (2001).

1.2.2 Transaction Costs

Another fundamental question to ask when making investment decisions is how frequently an investor needs to rebalance his portfolio (Seth 2002). Answers to this question depend on the volatility of returns and transaction costs. We will present an extensive quantitative study of the effects of transaction costs in Chapter 5. Transaction costs are the penalty paid for transacting or trading and they are made up of several components. Cheng (2003) provided a summary of the main quantitative costs and these include transaction fees, the bid-ask spread, opportunity cost, liquidity impact and market impact. These costs are detailed below.

Transaction fees

Commissions, fees and taxes (except capital gains tax) are unavoidable costs. They are the charges calculated as the percentage per share paid to the broker and tax office for executing the trade, and hence they can be well anticipated by the market participants. These make the smallest contribution to transactions costs, and they are the easiest to measure.

Bid-ask Spread

The bid-ask spread is the difference between the highest bid and the lowest sell offer for the underlying stock at any given time. It measures the loss from buying one share of a stock and then immediately selling it, and hence it is approximately the cost of trading one share of stock. The bid-ask spread is the main component of transaction costs when trading size is small (Grinold and Kahn 1999).

Opportunity Costs

In economics, opportunity cost is the trade-off between two options. In the context of investments, an investor gives up returns when he chooses a certain portfolio that earns less than the best alternative investment with a similar risk profile (Parrino et al. 2014). One example of an opportunity cost is the loss on any intra-day or day-by-day return when orders are not filled on the same day as they are placed (Grinold and Kahn 1999).

Liquidity Impacts

Liquidity impact arises when an order is larger than the inside market or it requires more immediate liquidity than liquidity providers can provide. The trade becomes visible to the rest of the market because of its size. Due to the immediate demand on liquidity, trading costs will be higher (Cheng 2003).

Market Impacts

Market impact is the effect that is observed when a market participant buys or sells an asset, and the purchase or sale shifts the price of the underlying asset. The price moves upward after a significant purchase and downward after a significant sale. Market impact is positively related to the size of orders and the resulting supply-demand imbalance in assets. The effect of trading on asset returns exhibits a strong positive relationship with contemporaneous returns (Brown, Walsh, and Yuen 1997; Chordia, Roll, and Subrahmanyam 2002; Chordia and Subrahmanyam 2004). This effect can be quantitatively modelled by order imbalance, which exists when the number of buyer-initiated transactions differs from the number of seller-initiated transactions (Kissell and Glantz 2003). Order imbalances cause price pressures. The effects due to positive and negative imbalances are differentiable, because price pressures resulting from large sell orders are greater than those from buy orders (Chan and Fong 2000; Chordia, Roll, and Subrahmanyam 2002). These effects influence asset prices in the aggregate market. Market impact grows as the transaction size increases, and it dominates transaction costs in significant transactions. Although the effects by different market participants may cancel or magnify each other and hence are beyond the traders and the portfolio managers' control, market impacts need to be taken into consideration when making trade decisions. We will provide a numerical example of how market impacts affect the performance of investment strategies in Chapter 5.

1.3 Our Research Framework and Contributions

1.3.1 Research Focus

The research in the thesis studies two different but related questions. First, we consider the case of continuous-time path-independent return models. We investigate the requirements for an allocation rule to be optimal in this situation. The derivation is an extension to Cox and Leland (2000). Second, we proceed to discrete-time path-dependent return models, which are supported by historical data. We generate and assess a particular class of allocation rules introduced by Leung (2011), and implement them within the context of a Vector Autoregressive - Multivariate GARCH (VAR-MGARCH) return framework. We also consider the empirical effect of market impacts on this class of allocation rules.

1.3.2 Our Contributions

Chapter 2 formulates the requirements for portfolio controls to be optimal and compatible with expected utility maximization when asset returns are path-independent. We consider a portfolio that consists of a risky and a risk-free asset. We assume that returns on the risky asset are path-independent with drift and diffusion being functions of the price of the risky asset, portfolio wealth and time. This is an extension to Cox and Leland (2000).

We then assess the predictability and efficiency of the Australian market in Chapter 3. We consider a VAR-MGARCH model which is path-dependent. Inspired by Pesaran and Timmermann (2000) and Di and Gangopadhyay (2013), we present a study of the evolution of parameters in a fitted VAR-MGARCH model for Australian index returns on equities, debt instruments and short-term bank deposits. The study is conducted using a "rolling window" approach which is to capture small parameter changes over different estimation windows. We develop and apply a one-step Multivariate SMLE (one-step MSMLE) method to estimate the model parameters.

We consider a portfolio consisting of three VAR-MGARCH asset classes (equity, debt and cash) and extend the dynamic portfolio rules proposed by Leung (2011) in Chapter 4, after recognizing the path-dependency of asset returns in Chapter 3. We quantitatively demonstrate that the proposed dynamic strategy exhibits excellent efficiency, in the sense that the associated variance is minimized for a given level of expected terminal portfolio wealth.

Chapter 5 investigates the effect of market impacts on the allocation rule proposed in Chapter 4. Our simulated examples show that the proposed strategy maintains its optimality in the presence of market impacts.

Chapter 2

Path-independent Returns and Asset Allocation

Declaration for Thesis Chapter 2

Declaration by candidate

In the case of Chapter 2 the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Equation derivations and text discussions	60%

The following co-author contributed to the work:

Name	Nature of contribution
Dr. Andrew Leung	Equation derivations

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and the co-author's contributions to this work.

Candidate's signature: [redacted]

Date: 15/03/2016

Main Supervisor signature: [redacted]

Date: 15/3/2016

We begin our study of investment strategies by following Merton (1971) to employ geometric Brownian motion models to study asset returns. Merton (1971) provided examples of how to systematically construct and analyze optimal continuous-time dynamic models with geometric Brownian motion and Poisson processes in his landmark paper. The mathematical simplicity of the two continuous-time processes limits the number of parameters in the problem and allows the derivation of analytical solutions.

Early work by Cox and Leland in 1970s, which was published much later in Cox and Leland (2000), derived criteria for portfolio controls to be considered as optimal in the sense that the controls maximize an investor's expected utility. Their study was based on a portfolio consisting of two assets, a risky (a stock) and a risk-free one, where asset returns were assumed to be path-independent and to have constant means and volatilities. Path-independence means that current returns do not depend on past returns. They started with the use of a discrete time binomial model to describe changes in portfolio wealth, and then took limits as the time interval shrank to zero to let the discrete measurements of changes converge to partial differential equations (PDEs) of portfolio wealth.

Ross (1978) addressed the use of two-asset portfolios. A portfolio with a single risky asset can be taken as a mutual fund. Ross (1978) showed that if geometric Brownian motion models are adopted, the separation theorem of mutual funds can be applied: in a portfolio problem of allocating wealth across many risky assets, the problem can be reduced to that of choosing amongst combinations of a few funds formed from these assets. However, asset returns are observed in practice to have non-constant means and volatilities. The use of a discrete-time binomial model, converging in continuous time by taking limits on the time interval, is cumbersome and detracts from the economics of the issue. We generalize some results due to Cox and Leland (2000) in this chapter.

We consider total wealth, $W = G + H$, with G invested in a risky asset (a stock) and H invested in a risk-free asset. Investors are allowed to increase or decrease their investment by an amount K, such that $K > 0$ for cash withdrawals and $K < 0$ for cash injections. W is a function of the price of the risky asset, s, at time, t. G, H and K are functions to be determined.

The price of the risky asset at time is assumed to follow a geometric Brownian motion such that

$$\frac{ds(t)}{s} = \mu dt + \sigma dB(t), \tag{2.1}$$

where $B\left(t\right)$ is a Brownian motion with $dB \sim \mathcal{N}(0, dt)$, μ is the drift and σ is the diffusion. Both μ and σ are assumed to be constants in Cox and Leland (2000). We generalize their results to the case when both μ and σ may be functions of s and t, but the evolution of s is path-independent.

The risk-free rate at time t is denoted by $r(t)$, which is generally independent of the stock price s so that $r_s = \partial r/\partial s = 0$, by virtue of being risk-free. The process of $r(t)$ evolves as

$$\frac{dPf}{Pf} = rdt, \tag{2.2}$$

where Pf is the price of the risk-free asset.

Hence, total wealth W obeys

$$\begin{aligned} dW &= dG + dH \\ &= \frac{ds}{s}G + rH\ dt - Kdt \\ &= [rW + (\mu - r)G - K]\, dt + \sigma G dB. \end{aligned} \tag{2.3}$$

Subscripts in this chapter are reserved solely for derivatives. Most of the following results are published in Leung and Shi (2013).

2.1 Synopsis of Utility Functions

Portfolio controls are considered to be optimal if they maximize an investor's expected utility. Utility functions, $u(.)$, were originally designed as tools for choosing between alternatives that would produce different levels of wellbeing. The function $u(.)$ gives different scores to different arguments (alternatives) and the argument that is both feasible and gives the highest expected score will be chosen. The purpose of setting up an investment strategy is to maximize an investor's expected utility, subject to the evolution of the investor's wealth.

Let $U(W) = \mathbb{E}\left[u(W)\right]$ be the expected utility of total wealth, W. We assume that expected utility is increasing and is also concave in wealth, that is

$$U'(W) \geq 0, \text{ and} \tag{2.4}$$

$$U''(W) \leq 0; \tag{2.5}$$

and these inequalities ensure that it is possible to find an optimal total wealth (that is W^*) that maximizes expected utility.

2.2 Controls Based on the Price of the Risky Asset and Time

Both the risk-free rate, $r(t)$, and the price of the risky asset, $s(t)$, are out of an investor's control. It is interesting to consider the case when an investor sets portfolio controls in response to $s(t)$, that is, when μ, σ and the corresponding controls may depend on both the price of the risky asset and time. This is the case when an investor believes that the portfolio controls and hence wealth are driven by the price of the risky asset.

We note that since $W = G + H$, we only need to control two of W, G and H to identify the entire portfolio.

2.2.1 Zero Cash Withdrawals or Inflow

A portfolio is self-financing when there are no cash withdrawals or injections, that is $K = 0$. The purchase of new assets is funded by the sales of old ones. In this case, the process for total wealth in (2.3) reduces to

$$dW = [rW + (\mu - r)G]\ dt + \sigma G dB. \tag{2.6}$$

By Itô's theorem we have

$$dW = W_t dt + W_s ds + \frac{1}{2} W_{ss}\, (ds)^2 \tag{2.7}$$

$$
\begin{aligned}
&= W_t dt + sW_s \frac{ds}{s} + \frac{1}{2} W_{ss} \left[(\mu s dt)^2 + 2\mu\sigma s^2 dB dt + (\sigma s dB)^2 \right] \\
&= W_t dt + sW_s \left(\mu dt + \sigma dB \right) + \frac{1}{2}\sigma^2 s^2 W_{ss} dt \\
&= \left(W_t + \mu s W_s + \frac{1}{2}\sigma^2 s^2 W_{ss} \right) dt + \sigma s W_s dB
\end{aligned}
$$

where, in the limit as $dt \to 0$, terms of higher orders of dt go to zero faster than first order terms and hence can be ignored.

Knowing the process of W, we can easily deduce the requirements for portfolio controls to be considered optimal in this situation as follows.

Proposition 1. *The optimal controls $W(s,t)$ and $G(s,t)$ satisfy the PDEs:*

$$W_t + rsW_s \ + \frac{1}{2}\sigma^2 s^2 W_{ss} - rW = 0, \tag{2.8}$$

$$G_t + s\left(r + s\sigma\sigma_s\right) G_s + \frac{1}{2}\sigma^2 s^2 G_{ss} - \left(r + s\sigma\sigma_s\right) G = 0, \ \textit{and} \tag{2.9}$$

$$G = sW_s \textit{ from (2.11).}$$

Proof. Equating (2.6) and (2.7), and combining like-terms, we have

$$rW + (\mu - r) G = W_t + \mu s W_s \ + \frac{1}{2}\sigma^2 s^2 W_{ss}, \tag{2.10}$$

and we have

$$G = sW_s. \tag{2.11}$$

Thus

$$rW + (\mu - r) sW_s = W_t + \mu s W_s \ + \frac{1}{2}\sigma^2 s^2 W_{ss},$$

and so

$$W_t + rsW_s + \frac{1}{2}\sigma^2 s^2 W_{ss} - rW = 0. \tag{2.12}$$

These partial differential equations (PDEs) are consistent with the conditions of Proposition 1 in Cox and Leland (2000).

Differentiating (2.12) with respect to s yields:

$$W_{st} + (rsW_{ss} + rW_s) + (\frac{1}{2}\sigma^2 s^2 W_{sss} + \sigma^2 sW_{ss} + \sigma\sigma_s s^2 W_{ss}) - rW_s = 0. \tag{2.13}$$

Multiplying both sides of (2.13) through by s, we obtain

$$sW_{st} + (rs^2 W_{ss} + rsW_s) + \left(\frac{1}{2}\sigma^2 s^3 W_{sss} + \sigma^2 s^2 W_{ss}\right) + \sigma\sigma_s s^3 W_{ss} - rsW_s = 0. \tag{2.14}$$

On the other hand, given the process of G in (2.11), we have the following partial derivatives of G :

$$G_t = sW_{st}, \tag{2.15}$$

$$G_s = W_s + sW_{ss}, \text{ and} \tag{2.16}$$

$$G_{ss} = 2W_{ss} + sW_{sss}. \tag{2.17}$$

Hence (2.14) becomes

$$G_t + rsG_s + \frac{1}{2}\sigma^2 s^2 G_{ss} + \sigma\sigma_s s\left(s^2 W_{ss} + sW_s - sW_s\right) - rG = 0$$
$$G_t + rsG_s + \frac{1}{2}\sigma^2 s^2 G_{ss} + \sigma\sigma_s s\left(sG_s - G\right) - rG = 0.$$

□

Remark 1. Proposition 1 of Cox and Leland (2000) will not apply if $\sigma_s \neq 0$. Note that both (2.8) and (2.9) are parabolic PDEs (Salsa 2008). G has the same functional form as W, but with r replaced by $r + s\sigma\sigma_s$. Therefore, the price of the risky asset in (2.1) has drift equal to $r + s\sigma\sigma_s$, that is

$$ds = (r + s\sigma\sigma_s)sdt + \sigma s dB(t). \tag{2.18}$$

2.2.2 Non-zero Cash Withdrawals or Inflow

We now consider the case when cash withdrawals or inflows are involved, that is, $K(s,t) \neq 0$.

Proposition 2. *Necessary and sufficient conditions for the differentiable functions $G(s,t)$, $W(s,t)$ and $K(s,t)$ to be the optimal controls of an investment strategy are that:*

$$W_t + rsW_s + \frac{1}{2}\sigma^2 s^2 W_{ss} - rW + K = 0, \tag{2.19}$$

$$G_t + s\left(r + s\sigma\sigma_s\right)G_s + \frac{1}{2}\sigma^2 s^2 G_{ss} - \left(r + s\sigma\sigma_s\right)G + sK_s = 0, \text{ and} \tag{2.20}$$

$$G = sW_s. \tag{2.21}$$

for all s and t.

Proof. Equating (2.3) and (2.7), we have

$$\begin{aligned}&[rW + (\mu - r)\,G - K]\,dt + \sigma G dB \\ &= \left[W_t + \mu s W_s \ + \frac{1}{2}\sigma^2 s^2 W_{ss}\right] dt + \sigma s W_s dB.\end{aligned} \tag{2.22}$$

Combining like-terms in (2.22), we obtain

$$sW_s = G,$$

which is the same condition as in (2.11) (when there were no cash withdrawals or inflows), and

$$rW + (\mu - r)\,G - K = W_t + \mu s W_s \ + \frac{1}{2}\sigma^2 s^2 W_{ss}. \tag{2.23}$$

Substituting (2.11) into (2.23), we have

$$rW + (\mu - r)\,sW_s - K = W_t + \mu s W_s \ + \frac{1}{2}\sigma^2 s^2 W_{ss},$$

so

$$W_t + rsW_s \ + \frac{1}{2}\sigma^2 s^2 W_{ss} - rW + K = 0. \tag{2.24}$$

Following the technique from the previous section, on differentiating (2.24) with respect to s and then multiplying through by s, we can easily show that

$$sW_{st} + (rs^2 W_{ss} + rsW_s) + \left(\frac{1}{2}\sigma^2 s^3 W_{sss} + \sigma^2 s^2 W_{ss}\right) + \sigma\sigma_s s^2 W_{ss} - rsW_s + sK_s = 0.$$

Given the partial derivatives of G in (2.15) to (2.17), the above can be rewritten as

$$G_t + rsG_s + \frac{1}{2}\sigma^2 s^2 G_{ss} - \sigma\sigma_s s(G - sG_s) - rG + sK_s = 0,$$

which has an extra term sK_s that accounts for cash withdrawals and inflows, compared to the process in (2.9) when there are no cash withdrawals or inflows. □

Remark 2. Proposition 2 generalizes Proposition 1 of Cox and Leland (2000), since the latter assumes constant diffusion in the price process, that is $\sigma_s = 0$. In addition, the process of W in (2.19) follows a parabolic PDE, and this property will be elaborated upon in Section 2.4.

2.3 Controls Based on Portfolio Wealth and Time

Recall that an investor chooses portfolio controls to maximize the expected utility of his resulting portfolio wealth. It is interesting to explicitly formulate controls as functions of portfolio wealth for this purpose. This case is taking the view that portfolio controls should correspond to the current state of the portfolio wealth.

Let us now consider the case when μ, σ and the corresponding controls may be functions of portfolio wealth and time. Note that μ and σ are characteristics of the risky asset and hence they are also functions of the price of the risky asset. Setting μ and σ as functions of the portfolio wealth reflects the idea that portfolio controls have market impacts on the risky asset.

2.3.1 Zero Cash Withdrawals or Inflows

This section is analogous to Section 2.2.1.

The requirements for portfolio controls to be considered optimal in this situation are as follows.

Proposition 3. *The optimal control $G(W,t) = sW_s$, viewed as a function of total wealth, satisfies:*

$$G_{t|W} + rWG_W + \frac{1}{2}\sigma^2 G^2 G_{WW} - rG = \sigma\sigma_W G^2 (1 - G_W). \tag{2.25}$$

Proof. Total wealth W is a function of (s,t), therefore, we have the following partial derivatives of G and σ:

$$G_t = G_{t|W} + G_W W_t, \tag{2.26}$$

$$G_s = G_W W_s, \tag{2.27}$$

$$G_{ss} = G_W W_{ss} + G_{WW} W_s^2, \text{ and} \tag{2.28}$$

$$\sigma_s = \sigma_W W_s. \tag{2.29}$$

Substituting (2.26) to (2.29) into (2.9), we have

$$\begin{aligned}
& G_t + rsG_s + \frac{1}{2}\sigma^2 s^2 G_{ss} - rG - s\sigma\sigma_s\left(G - sG_s\right) \\
&= \left[\begin{array}{c} \left(G_{t|W} + G_W W_t\right) + \frac{\sigma^2 s^2}{2}\left(G_W W_{ss} + G_{WW} W_s^2\right) \\ + rsG_W W_s - rG - s\sigma\sigma_W W_s \left(G - sG_W W_s\right) \end{array}\right] \\
&= \left[\begin{array}{c} G_{t|W} + G_W \left(W_t + rsW_s + \frac{\sigma^2 s^2}{2} W_{ss}\right) - rG \\ + \frac{\sigma^2 s^2}{2} G_{WW} W_s^2 - \left[\sigma\sigma_W G^2 - s^2\sigma\sigma_W G_W W_s^2\right] \end{array}\right] \\
&= 0.
\end{aligned}$$

Recalling (2.8) and (2.11), the above becomes

$$\begin{aligned}
& G_{t|W} + rWG_W + \frac{1}{2}\sigma^2 G^2 G_{WW} - rG - \sigma\sigma_W G^2\left(1 - G_W\right) \\
&= 0.
\end{aligned}$$

□

Remark 3. Proposition 3 generalizes Proposition 2 of Cox and Leland (2000). Proposition 2 of Cox and Leland (2000) is a special case of (2.25) when $\sigma_W = 0$.

2.3.2 Non-zero Cash Withdrawals or Inflows

This section is in analogy with Section 2.2.2.

Let K be a function of (W, t). We have:

$$K_s = K_W W_s. \tag{2.30}$$

The requirements for portfolio controls to be considered optimal in this situation are as follows.

Proposition 4. *The optimal control $G(W,t) = sW_s$, viewed as a function of total wealth, satisfies:*

$$G_{t|W}+(rW-K)G_W+\frac{1}{2}\sigma^2G^2G_{WW}-(r-K_W)G = \sigma\sigma_W G^2\left(1-G_W\right). \quad (2.31)$$

Proof. Substituting the partial derivatives in (2.26) to (2.29) and (2.30), and condition (2.11) into the process of G in (2.20), we have

$$G_t + s\left(r+s\sigma\sigma_s\right)G_s + \frac{1}{2}\sigma^2 s^2 G_{ss} - \left(r+s\sigma\sigma_s\right)G + sK_s = 0,$$

$$\left[\begin{array}{c} \left(G_{t|W}+G_W W_t\right) + s\left(r+s\sigma\sigma_W W_s\right)G_W W_s \\ +\frac{1}{2}\sigma^2 s^2\left(G_W W_{ss}+G_{WW}W_s^2\right) - \left(r+s\sigma\sigma_W W_s\right)G + sK_W W_s \end{array}\right] = 0,$$

$$\left[\begin{array}{c} G_{t|W}+G_W W_t + \left(r+\sigma\sigma_W G\right)G_W G \\ +\frac{1}{2}\sigma^2 s^2 W_{ss}G_W + \frac{1}{2}\sigma^2 G_{WW}G^2 - \left(r+\sigma\sigma_W G\right)G + K_W G \end{array}\right] = 0,$$

and hence

$$\left[\begin{array}{c} G_{t|W}+G_W\left(W_t+rG+\frac{1}{2}\sigma^2 s^2 W_{ss}\right) \\ +\sigma\sigma_W G_W G^2 + \frac{1}{2}\sigma^2 G_{WW}G^2 - \left(r+\sigma\sigma_W G\right)G + K_W G \end{array}\right] = 0.$$

Using condition (2.24), the above becomes

$$\begin{aligned} & G_{t|W}+\left(rW\ -K\right)G_W+\frac{1}{2}\sigma^2G^2G_{WW}-rG+K_WG-\sigma\sigma_WG^2\left(1-G_W\right) \\ = \ & 0. \end{aligned}$$

□

Remark 4. This result generalizes Proposition 2 of Cox and Leland (2000) when $\sigma_W = 0$. That is, the portfolio controls can have market impacts on the price of the risky asset.

2.4 Controls that are Compatible with Concave Utility

Recall from our discussion of utility maximization that the portfolio controls formulated in Sections 2.2 and 2.3 are optimal if they maximize some expected concave utility $U(W)$. We will show that this concavity requirement is satisfied below.

An investor acquires utility from both the portfolio wealth and consumption from cash withdrawals. Therefore, the investor's problem of choosing the optimal portfolio and consumption rules over N years is formulated as to choose W, G, H and K to maximize

$$U = \mathbb{E}_{\bar{\pi}} \left\{ u(W) + \int_0^N c(K) dt \right\}, \tag{2.32}$$

where $u(W)$ is the utility from portfolio wealth and it is concave in W, $c(K)$ is the utility from consumption and it is concave in K, and the expectation is taken with respect to the physical stock (the risky asset) price process. The density for the physical stock price process is $\bar{\pi}(s_0, 0; S, N)$, where the initial and terminal times are specified at $t = 0$ and $t = N$, $s(0) = s_0$ and

$s(N) = S$. Then the expected utility is evaluated by the integral

$$U = \int_0^\infty \left[u\left[W\left(S\right)\right] + \int\limits_0^N c(K)d\tau \right] \bar{\pi} dS. \tag{2.33}$$

On the other hand, $K(s,t)$ is a control, hence there is no loss of generality by replacing it with $k(s,t)W$, and then (2.24) becomes

$$W_t + rsW_s + \frac{1}{2}\sigma^2 s^2 W_{ss} = (r-k)W, \tag{2.34}$$

which is a parabolic PDE. Therefore, if there exists a solution to $W(s,t)$, the Feynman-Kac (Kac 1949) formula can be applied to find such a solution as

$$\begin{aligned} W(s,t) &= \mathbb{E}_\pi \left[\psi(S)\gamma(t) \left| s(N) = S \right|\right] \\ &= \gamma(t) \int_0^\infty \psi(S)\pi(s_0, 0; S, N) dS, \end{aligned} \tag{2.35}$$

where the expectation is taken with respect to the risk-neutral process $ds = rsdt + \sigma sdB$ with density $\pi(s_0, 0; S, N)$, $\psi(S) = W(S,N)$ is the terminal wealth at time N with stock price S, and

$$\begin{aligned} \gamma(s,t) &= \exp\left\{ \int_t^N \left[k(s,z) - r(z)\right] dz \right\} \\ &= v\left(t\right) \exp\left[\int_t^N k(s,z) dz \right] \end{aligned} \tag{2.36}$$

is the discount factor. Replacing $W(S,N)$ with $\psi(S)$, the expected utility becomes

$$U = \int_0^\infty \left[u\left[\psi(S)\right] + \int\limits_0^N c(K)d\tau \right] \bar{\pi} dS. \tag{2.37}$$

The investment in the risky asset is governed by $G(s,t) = sW_s$ according to the parabolic PDE given in (2.9). Therefore, by similarity, if there exists a solution to the parabolic PDE to G, the solution can be found as

$$G(s,t) = w(t)\mathbb{E}_{\bar{\pi}}\left[S\psi_s(S)\left|s(N) = S\right|\right], \tag{2.38}$$

where the expectation is taken with respect to the physical stock process defined in (2.18) and $w(t) = \exp\left\{-\int_t^N (r + s\sigma\sigma_s)dz\right\}$.

Remark 5. Let us rewrite the PDE of W in (2.34) as:

$$\begin{aligned}
&W_t + rsW_s + \frac{1}{2}\sigma^2 s^2 W_{ss} - (r-k)W \\
&= W_t + pW_s + \frac{1}{2}qW_{ss} - \nu W \\
&= 0.
\end{aligned}$$

The above PDE can be solved utilizing the Feynman-Kac formula as in (2.35) and (2.36) only when the following conditions are satisfied:

C1 p and q are in $C^{0,1}(\mathbb{R}^2)$[1], and are globally bounded above;

C2 q is globally bounded from zero across (s,t);

[1] $C^{k,\alpha}(\Omega)$ is the Hölder space that consists of functions on an open Euclidean subset Ω having continuous derivatives up to order k and such that the kth partial derivatives are Hölder continuous (or satisfy the Hölder condition) with exponent $\alpha \in [0,1]$.

A function f defined on Euclidean space satisfies a Hölder condition when there are nonnegative real constants C and α such that $|f(x) - f(y)| \leq D\,|x-y|^\alpha$, for any x and y in the domain of f. When $\alpha = 1$, f satisfies the Lipschitz condition, and when $\alpha = 0$, f is bounded.

C3 p and q satisfy a global Hölder condition with respect to s;

C4 $p \in C^{2,1}(\mathbb{R}^2)$ and $q \in C^{2,1}(\mathbb{R}^2)$;

C5 p and q and their second derivatives with respect to s are of most polynomial growth;

C6 the growth of p and q is at most linear;

C7 q is continuous in t; and

C8 υ is uniformly bounded and locally Hölder.

Remark 6. Investment in the risky asset is governed by $G(s,t) = sW_s$, and it has a parabolic PDE given in (2.9), when there are no cash withdrawals or inflows. Therefore, by similarity, if there exists a solution to the parabolic PDE to G, the solution can be found as

$$G(s,t) = w(t)\mathbb{E}_{\bar{\pi}}\left[S\psi_s(S)\left|s(N) = S\right|\right], \tag{2.39}$$

where the expectation is taken with respect to the physical stock price process defined in (2.18), and $w(t) = \exp\left\{-\int_t^N (r + s\sigma\sigma_s)d\tau\right\}$.

2.4.1 Zero Cash Withdrawals or Inflows

There is no utility from consumption when cash withdrawals are not allowed, that is, $K(s,t) = 0 = k(s,t))$, and the expected utility for an investor in (2.37) reduces to

$$\begin{aligned} U &= \mathbb{E}_{\bar{\pi}}\left\{u\left[\psi\left(S\right)\right]\right\} \\ &= \int_0^\infty u\left[\psi\left(S\right)\right]\bar{\pi}(s(0),0;S,N)dS. \end{aligned} \tag{2.40}$$

The solution to $W(s,t)$ given by the Feynman-Kac formula in (2.35) reduces to

$$W(s,t) = v(t)\mathbb{E}_{\pi}\left[\psi(S)\left|s(N) = S\right|\right], \tag{2.41}$$

where $v(t) = \exp\left\{-\int_t^N r(z)dz\right\}$ is the discount factor. The initial condition is then

$$\begin{aligned} W(s(0),0) &= v(0)\mathbb{E}_{\pi}\left[\psi(S)\left|s(N) = S\right|\right] \\ &= v(0)\int_0^{\infty} \psi(S)\pi(s(0),0;S,T)dS. \end{aligned} \tag{2.42}$$

The problem of choosing optimal portfolio controls for an investor who operates for a period $[0,N]$ is formulated as to maximize his expected utility of terminal portfolio wealth subject to a given initial wealth. The Lagrangian is constructed to maximize (2.40) subject to (2.42) as

$$\mathcal{L} = \int_0^{\infty} \left[\bar{\pi}u\left[\psi\left(S\right)\right] - \lambda\pi\psi\left(S\right)\right]dS, \tag{2.43}$$

where λ is a positive constant chosen for (2.42) to hold (see Gelfand and Fomin (1963)).

Differentiating (2.43) with respect to terminal wealth $\psi(S)$ and setting the derivative equal to zero:

$$\begin{aligned} \mathcal{L}_W &= \frac{d}{dW}\int_0^{\infty} \left[\bar{\pi}u\left[\psi\left(S\right)\right] - \lambda\pi\psi\left(S\right)\right]dS \\ &= \int_0^{\infty} \left[u_W\bar{\pi} - \lambda\pi\right]dS = 0, \end{aligned}$$

we obtain the first order condition that

$$u_W = \lambda \frac{\pi(s(0), 0; S, N)}{\bar{\pi}(s(0), 0; S, N)} > 0, \tag{2.44}$$

which satisfies (2.4), that the utility function is non-decreasing. The second order condition to maximize (2.43) requires that

$$\begin{aligned} \mathcal{L}_{WW} &= \int u_{WW} \bar{\pi} dS \leq 0, \text{ and hence} \\ u_{WW} &\leq 0, \end{aligned}$$

which satisfies (2.5), that the utility function is concave. Therefore, controls satisfying (2.8), (2.9) and (2.25) are compatible with a concave utility function.

Proposition 5. *A path independent strategy $W(s,t)$ can be found to optimize a given concave utility $u(W)$ if and only if a solution to terminal wealth $W(S,N) = \psi(S)$ can be found to satisfy*

$$u_W = \lambda \frac{\pi(s(0), 0; S, T)}{\bar{\pi}(s(0), 0; S, T)},$$

where λ is a positive constant.

2.4.2 Non-zero Cash Withdrawals or Inflows

When cash withdrawals are allowed, that is then $K(s,t) \neq 0$, we are considering the expected utility from both portfolio returns and consumption for an investor, which is defined in (2.32). Maximizing (2.32) is a continuous stochastic control problem (Dreyfus 1965). We will show that controls satisfying (2.24), (2.20), (2.11) and (2.31) are compatible with the concavity requirements in utility functions in (2.32) using two methods. The first method will solve the fundamental PDE, while the second method will apply calculus of variations.

Fundamental Partial Differential Equation

The expected utility acquired from both investment returns and consumption in (2.32) can be maximized by solving the fundamental PDE. References to this procedure can be found in Dreyfus (1965).

Let us define an optimal expected value function such that

$$\begin{aligned} O(W,t) &= \max_K \mathbb{E}_{\bar{\pi}} \left\{ u\left[\psi\left(S\right)\right] + \int_t^N c(K)dz \right\} \\ &= \max_K \int_0^\infty \left[u\left[\psi(S)\right] + \int\limits_0^N c(K)d\tau \right] \bar{\pi} dS. \end{aligned} \tag{2.45}$$

The physical stock price process outlined in (2.1) terminates at time N. The terminal contribution to the expectation is assessed, and the boundary condition is

$$O(W(S,N),N) = u\left[\psi\left(S\right)\right]. \tag{2.46}$$

Notice that $O(W(S,N),t)$ may be referred to as an indirect utility function.

Recall the process of W in (2.3). It has a drift of $\dot{W} = (\mu G + rH - K)$, and a diffusion of $h^{1/2} = \sigma G$. The fundamental PDE is

$$\begin{aligned} 0 &= \max_K \left[c + O_W \dot{W} + O_t + \frac{1}{2} O_{ss} h \right] \\ &= \max_K \left[c + O_W (\mu G + rH - K) + O_t + \frac{1}{2} O_{ss} \sigma^2 G^2 \right]. \end{aligned} \tag{2.47}$$

The first order condition for a maximum requires

$$\begin{aligned} c_K &= O_W \\ &= \int_0^\infty u_W \bar{\pi} dS \geq 0, \end{aligned} \tag{2.48}$$

given that $u_W > 0$, which satisfies the non-decreasing requirement of a utility function in (2.4). Differentiating (2.48) with respect to K, we have the second order condition for a maximum that

$$c_{KK} = \frac{\int_0^\infty u_{WW} \bar{\pi} dS}{K_W} \leq 0,$$

given that $u_{WW} \leq 0$, which satisfies the concavity requirement of a utility function in (2.5). Therefore, controls satisfying (2.24), (2.20), (2.11) and (2.31) are compatible with concave consumption and utility functions.

Proposition 6. *For a given concave utility function for terminal wealth $u(W)$ and consumption $c(K)$, with terminal wealth $W = \psi(S)$, the optimal cash withdrawal is given by $K(s,t)$ satisfying*

$$c_K = \int_0^\infty u_W \bar{\pi} dS, \ \textit{and} \tag{2.49}$$

$$c_{KK} = \frac{\int_0^\infty u_{WW} \bar{\pi} dS}{K_W}.$$

Notice that Proposition 6 requires that $K_W > 0$, that is, cash withdrawals increase with portfolio wealth. We will elaborate on this relation below in Remark 7.

Calculus of Variations

A simpler approach to utility maximization utilizes the calculus of variation. A good reference can be found in Gelfand and Fomin (1963).

The initial wealth given by (2.35) is

$$\begin{aligned} W(s(0), 0) &= \mathbb{E}_Q \left[\psi(S) \gamma(0) s(N) \left| s(N) = S \right. \right] \qquad (2.50) \\ &= \int_0^\infty \left\{ v(0) \psi(S) \exp \left[\int_0^N k(s, z) dz \right] \pi(s(0), 0; S, N) \right\} dS. \end{aligned}$$

Hence, the Lagrangian that maximizes the expected utility in (2.37) subject to the initial condition in (2.50) is

$$\begin{aligned} \mathcal{L} &= \mathbb{E}_P \left[u \left[\psi \left(S \right) \right] + \int_0^N c(K) d\tau \right] - \omega \mathbb{E}_Q \left[\psi(S) \gamma(0) s(N) \left| s(N) = S \right. \right] \qquad (2.51) \\ &= \int_0^\infty \left[u \left[\psi \left(S \right) \right] + \int_0^N c(K) d\tau \right] \bar{\pi} dS - \omega v(0) \int_0^\infty \psi(S) e^{\int_0^N k(s,z) dz} \pi dS \\ &= \int_0^\infty \left\{ \left[u \left[\psi \left(S \right) \right] + \int_0^N c(K) d\tau \right] \bar{\pi} - \omega v(0) \psi(S) e^{\int_0^N k(s,z) dz} \pi \right\} dS. \end{aligned}$$

where ω is a positive constant chosen to ensure that (2.50) holds.

Given $\psi\left(S\right)$, consider a small variation in k, say Δk, localized at time τ

and in state s so that

$$\Delta k = \eta \text{ in } [\tau, \tau + d\tau]$$

for some constant η, which induces a variation of $\Delta K = \eta W$. This further induces variations in the terms $\int_0^N c(K)d\tau$ and $\gamma(t)$, to the first order in $d\tau$. These are

$$\Delta c(K) = c_K \Delta K + \frac{1}{2} c_{KK} (\Delta K)^2,$$

that is

$$\Delta \int_0^N c(K)d\tau = c_K W \eta d\tau + \frac{1}{2} c_{KK} (\eta W)^2 d\tau,$$

and

$$\Delta\gamma(s,t) = v(s) \exp\left[\int_0^N k(s,\tau)d\tau\right] \cdot \left(e^{\eta d\tau} - 1\right)$$
$$\simeq \gamma(s,t)(\eta d\tau), \text{ as } \eta d\tau \text{ is very small.}$$

Since η is localized at s, $E_Q\left\{\psi(S)\exp\left[\int_0^N k(s,z)dz\right]\right\} = \int_0^\infty \alpha dS$ is a positive constant. The variation in the Lagrangian function (2.51) is

$$\Delta\mathcal{L} = \int_0^\infty \left\{\left[c_K \eta W d\tau + \frac{1}{2} c_{KK} (\eta W)^2 d\tau\right] \bar{\pi} - \omega (\eta d\tau) v(0) \alpha\right\} dS, \quad (2.52)$$

which is nonnegative. We require the variation in the Lagrangian given in (2.52) to be at a maximum, when the Lagrangian is maximized. The first order condition is given by differentiating (2.52) with respect to η to obtain:

$$\int_0^\infty \left\{\left(c_K W d\tau + c_{KK} \eta W^2 d\tau\right) \bar{\pi} - \omega v(0) \alpha d\tau\right\} dS = 0, \text{ and hence} \quad (2.53)$$

$$\left(c_K W + c_{KK}\eta W^2\right)\bar{\pi} - \omega v\left(0\right)\alpha = 0. \tag{2.54}$$

The second order condition is given by differentiating (2.53) with respect to η. This shows that

$$\int_0^\infty c_{KK} W^2 \bar{\pi} dS \leq 0,$$

and hence

$$c_{KK} \leq 0, \tag{2.55}$$

which satisfies (2.5). Substituting (2.55) into (2.54) we have

$$\overline{\pi}\left(s_0, 0; s, \tau\right) c_K W = \kappa \tag{2.56}$$

for some constant $\kappa \geq 0$, which indicates that $c_K \geq 0$. This satisfies the non-decreasing requirement of a feasible utility function in (2.4). Hence we have the following result.

Proposition 7. *Given a concave utility function for terminal wealth $u(W)$ and for consumption $c(K)$, and terminal wealth $\psi\left(S\right)$, the optimal cash withdrawal is given by $K(s,t)$ satisfying:*

$$c_K = \frac{\kappa}{\overline{\pi}\left(s_0, 0; s, t\right) W} \tag{2.57}$$

for some constant $\kappa > 0$. This result is compatible with Proposition 6.

Remark 7. The above proposition indicates that c_K is a decreasing function in W. On the other hand, concavity of utility from consumption requires that c_K is a decreasing function in K, since c_{KK} is non-positive. These two conditions imply that cash withdrawals should increase when greater wealth W is likely to be achieved and decrease when such wealth is less likely to be achieved, that is $K_W > 0$ as in Proposition 6. This corresponds with the conditions contained in Proposition 4 of Cox and Leland (2000).

2.5 Conclusions

This chapter addresses two related issues, building on the work by Cox and Leland (2000). We formed a portfolio that consisted of a risky and a riskfree asset, where the returns are path independent but the expected return and volatility of the risky asset are non-constant, or even stochastic.

We first examined and derived the characteristics of optimal portfolio controls. This was done from two perspectives. One view is that portfolio wealth is driven by the price of the risky asset, and hence we took partial derivatives of the portfolio wealth and controls with respect to the price of the risky asset and time. Another view is that portfolio controls should correspond to the current state of the portfolio wealth, which incorporates market impacts due to rebalancing, and hence we took partial derivatives of the portfolio controls with respect to the portfolio wealth and time. We showed that the optimal control criteria in Cox and Leland (2000) were special cases of ours, in which they assumed that the expected return and volatility of the risky asset were constant. We also showed that the portfolio controls satisfying the PDEs we derived were compatible with utility maximization.

Chapter 3

Asset Returns in the Australian Market

Perhaps, the key to success in investments is the ability to forecast future returns. Finding the appropriate means for modeling uncertainties is critical and drives the results. Investment strategies are useful only insofar as the forecasts of future returns and their variabilities are reliable, which in turn requires the parameters governing the return process to be appropriately estimated.

Kassberger and Liebmann (2012) show that path-dependent strategies are preferred if the pricing model of the underlying risky assets is itself path-dependent. Therefore, if the true data generating process of returns on the risky asset is path-dependent, the criteria for optimal investment strategies we have shown in Chapter 2 will not apply. More realistic asset models are often path-dependent. One continuous time example is when the stock price follows the process

$$\frac{ds}{s} = (\alpha + \beta r)dt + \sigma r^{\gamma} dB, \tag{3.1}$$

where s is the price of the risky asset, r is the risk-free return, σ is the standard deviation of return on the risky asset, B is a Brownian motion, and α, β and γ are coefficients. Equation (3.1) encompasses a number of models used in the literature, as discussed in a survey article by Chan et al. (1992). Examples of discrete time asset pricing models include an Autoregressive (AR) model of returns, in which returns at time t depend on past realizations, and a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model in which current state variance depends on the entire sequence of past returns up to that time.

A question to ask when studying asset returns is to what extent our modeling of returns is going to change over time. Different points of view can be taken on this issue. The most basic one would be to assume that returns are stationary, so that we would not need to update our beliefs about the properties of asset returns as we progress into the future. This is called "fixed estimation", which assumes that parameters are constant over time. Another approach, in contrast, is that as we progress into future planning horizon, we will update the parameters based upon the additional realized returns we observe. This is called "recursive estimation". The "recursive estimate" allows for slow parameter changes over time. Both approaches are based on the assumption that, over a certain period of time, returns can be treated as if they were samples drawn from a distribution characterized by certain parameters and we choose different historical periods to estimate the underlying parameters.

An alternative modeling methodology is to use a "rolling historical window". This methodology divides sequential historical data into windows of equal size. Earlier observations drop out as we "roll into" later windows. The same parametric model is reevaluated for each data window, and if parameters change over time then the rolling window estimates will reflect these changes. An example of the rolling window method can be found in the study of UK stock returns by Pesaran and Timmermann (2000).

Our goal is to study how the prices of Australian securities evolve in this chapter. This is done in a Vector Autoregressive - Multivariate Generalized Autoregressive Conditional Heteroskedasticity (VAR-MGARCH) framework. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are often estimated by Maximum Likelihood Estimation (MLE) methods. When studying asset returns, the underlying likelihood function is often written under the assumption that returns are normally distributed. However, it is widely acknowledged that returns are not normally distributed. Hence the resulting estimator is in fact a pseudo- or quasi-Maximum Likelihood Estimator (QMLE), which is still consistent but not efficient. A non-parametric solution to this problem is to construct the likelihood function utilizing an estimated density of returns. The resulting estimator is consistent and can also be asymptotically efficient if we have a proper estimate of the density.

In this chapter, we develop a one-step multivariate semiparametric maximum likelihood estimation (one-step MSMLE) technique to study MGARCH models, which extends the work of Di and Gangopadhyay (2013). Di and Gangopadhyay (2013) proposed a one-step semiparametric maximum likelihood estimation (one-step SMLE) technique to study univariate GARCH models, utilizing kernel density estimation to investigate the distribution of the data. This technique is one-step in the sense that compared to a two-step SMLE, it does not require a first step to perform QMLE to acquire a fitted residual series. The proposed estimators are consistent and asymptotically unbiased and efficient.

We then estimate an MGARCH model of Australian monthly asset returns using rolling windows of 120 effective months of returns. The estimation is done by our proposed one-step MSMLE. We summarize and discuss the evolution of Australian security prices at the end of this chapter. A key result is that parameters of VAR-MGARCH model are path-dependent and change over time. This supports the adoption of dynamic investment strategies that are studied in Chapter 4.

Subscripts in this chapter and thereafter are reserved for indices on variables.

3.1 One-step MSMLE

Suppose we have an n-period time series $r_t = \sigma_t e_t$, where e_t are independently identically distributed (i.i.d.) innovations with mean 0, variance of 1 and density $f(e_t; \boldsymbol{\theta})$. The conditional variance of r_t, $\sigma_t^2(\boldsymbol{\theta})$ has a GARCH form governed by the parameter $\boldsymbol{\theta}$. We wish to maximize the log-likelihood function

$$l(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^{n} \ln \left[\frac{f(e_t; \boldsymbol{\theta})}{\sigma_t(\boldsymbol{\theta})} \right] \tag{3.2}$$

$$= \frac{1}{n}\sum_{t=1}^{n} \ln\left[\frac{1}{\sigma_t(\boldsymbol{\theta})} f\left(\frac{r_t}{\sigma_t(\boldsymbol{\theta})}\right)\right]$$

to obtain an estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, which belongs to the family of M-estimators Hayashi (2000).

The maximum likelihood estimate (MLE) is consistent and asymptotically efficient, only when the true innovation distribution is correctly specified. However, the true innovation distribution is often unknown. Therefore, in order to perform MLE, we need to replace $f(e_t; \boldsymbol{\theta})$ in (3.2) by a corresponding estimated density. This can be done using two approaches: a QMLE approach or a SMLE approach.

A QMLE proceeds as if the data follows a pre-determined parametric distribution. The resulting estimators are consistent but not efficient.

An SMLE, in contrast, utilizes a non-parametric estimate of the distribution of the data. The resulting estimators are consistent, and they are even asymptotically efficient if the density is estimated properly.

3.1.1 Univariate One-step SMLE

Di and Gangopadhyay (2013) proposed a one-step SMLE technique to study univariate GARCH models by utilizing kernel density estimates. In particular, the estimated density for e is

$$\hat{f}_n\left(e; \boldsymbol{\theta}\right) = \frac{1}{nh_n}\sum_{s=1}^{n} K\left(\frac{e - e_s(\boldsymbol{\theta})}{h_n}\right),$$

where $K(., h_n)$ is a regular kernel function and h_n is the bandwidth.

Substituting the above estimated density in the log-likelihood in (3.2), we have

$$\begin{aligned}\hat{l}_n(\boldsymbol{\theta}|\mathbf{e}) &= \frac{1}{n}\sum_{t=1}^{n}\ln\left[\frac{\hat{f}_n\left(e_t;\boldsymbol{\theta}\right)}{\sigma_t(\boldsymbol{\theta})}\right] \\ &= \frac{1}{n}\sum_{t=1}^{n}\ln\left[\frac{1}{\sigma_t(\boldsymbol{\theta})nh_n}\sum_{s=1}^{n}K\left(\frac{e_t-e_s}{h_n}\right)\right].\end{aligned} \tag{3.3}$$

The one-step semiparametric maximum likelihood estimator (one-step SMLE) of $\boldsymbol{\theta}$, is $\hat{\boldsymbol{\theta}}_{1SMLE}$, which maximizes (3.3).

3.1.2 One-step MSMLE

We extend the above method, proposed by Di and Gangopadhyay (2013), to MGARCH cases as follows.

Suppose that we have a d-variate n-period time series

$$\mathbf{u}_t = \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\mathbf{e}_t, \tag{3.4}$$

where $\mathbf{e}_t$ is an independently identically distributed (i.i.d.) innovation with $\mathbb{E}(\mathbf{e}_t) = \mathbf{0}$ and covariance $\mathbf{I}_d$ (where $\mathbf{I}_d$ is a $d \times d$ identity matrix). The covariance of $\mathbf{u}_t$ is $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$. $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is symmetric and positive-definite, and it has an MGARCH form specified by parameters $\boldsymbol{\theta}$. We write $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ as the lower triangular Cholesky decomposition of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ such that $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\left(\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\right)^T = \boldsymbol{\Sigma}_t(\boldsymbol{\theta})$. The parameter $\boldsymbol{\theta}$ has finite dimension $0 \leq k < \infty$ that $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k$, where Θ is the parameter space.

Suppose further that Ω is the sample space and the information set F is the sigma-algebra generated by Ω.

We also impose some additional regularity conditions as follows.

Assumption 1. $\mathbf{u}_t$ is ergodic stationary and is squared integrable, $\forall\boldsymbol{\theta} \in \Theta$.

Assumption 2. The model in (3.4) is identifiable, that is, if $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_\mathbf{0})$ then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Assumption 3. The parameter space Θ is compact.

Assumption 4. The true parameter $\boldsymbol{\theta}^*$ is in the interior of Θ.

Assumption 5. The density of $\mathbf{e}$ given parameter $\boldsymbol{\theta}$ $f(\mathbf{e};\boldsymbol{\theta}) \in G$ is finite, where

$$G = \begin{array}{c} \{f : \mathbb{R}^d \to \mathbb{R}_{++} \big| \int f(\mathbf{x})d\mathbf{x} = 1, \int \mathbf{x}f(\mathbf{x})d\mathbf{x} = 0, \int \mathbf{x}\mathbf{x}^T f(\mathbf{x})d\mathbf{x} = \mathbf{I}_d, \\ \int \mathbf{x}\mathbf{x}^T f(\mathbf{x})d\mathbf{x} = \mathbf{I}_d, \forall i \sup \left| f^{(i)}(\mathbf{x}) \right| < \infty\} \end{array}.$$

Assumption 6. $\mathbb{E}[|\ln(|\boldsymbol{\Sigma}_\mathbf{t}|)|] < \infty$, $\forall\boldsymbol{\theta} \in \Theta$.

Assumption 7. $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is continuous and measurable on (Ω, F) for all $\boldsymbol{\theta}$ in Θ.

The log-likelihood function of $\boldsymbol{\theta}$ is

$$\begin{aligned} l(\boldsymbol{\theta}|\mathbf{u}) &= \frac{1}{n}\sum\nolimits_{t=1}^{n} \ln f(\mathbf{u}_t;\boldsymbol{\theta}) \\ &= \frac{1}{n}\sum\nolimits_{t=1}^{n} \ln\left[|\Sigma_t(\boldsymbol{\theta})|^{-1/2} f(\mathbf{e}_t;\boldsymbol{\theta})\right] \end{aligned} \tag{3.5}$$

where $f(.)$ is the true density of $\mathbf{e}_t$ given $\boldsymbol{\theta}$

The true distribution of $\mathbf{e}_t$ is unknown. Hence the density $f(\mathbf{e}_t;\boldsymbol{\theta})$ in (3.5) needs to be replaced by its estimate $\hat{f}_n(\mathbf{e}_t;\boldsymbol{\theta})$, and we acquire an semiparametric log-likelihood as

$$l(\boldsymbol{\theta}|\mathbf{u}) = \frac{1}{n}\sum\nolimits_{t=1}^{n} \ln\left[f\left(\mathbf{u}_t;\boldsymbol{\theta}\right)\right] \tag{3.6}$$
$$= \frac{1}{n}\sum\nolimits_{t=1}^{n} \ln\left[|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|^{-1/2} f\left(\mathbf{e}_t;\boldsymbol{\theta}\right)\right].$$

The kernel density estimate of the standardized innovation $\mathbf{e}$ is given by

$$\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right) = n^{-1}|\mathbf{H}|^{-1}\sum_{s=1}^{n} K\left(\mathbf{H}^{-1}\left[\mathbf{e}-\mathbf{e}_s(\boldsymbol{\theta})\right]\right), \tag{3.7}$$

where $\mathbf{H}$ is a symmetric positive-definite bandwidth matrix of dimension $(d \times d)$. The multivariate kernel function $K(\boldsymbol{\omega})$ $(\boldsymbol{\omega} = \mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s))$ satisfies

$$\int_{\mathbb{R}^d} K(\boldsymbol{\omega})d\boldsymbol{\omega} = 1, \tag{3.8}$$

$$\int_{\mathbb{R}^d} \boldsymbol{\omega}K(\boldsymbol{\omega})d\boldsymbol{\omega} = 0, \text{ and} \tag{3.9}$$

$$\int_{\mathbb{R}^d} \boldsymbol{\omega}\boldsymbol{\omega}^T K(\boldsymbol{\omega})d\boldsymbol{\omega} = \mathbf{I}_d \text{ (Scott 1992)}. \tag{3.10}$$

The choice of $K(.)$ is not crucial for the accuracy of density estimates. However, the choice of bandwidth $\mathbf{H}$ is crucial for the accuracy of density estimates. Unfortunately, there is no general closed-form expression for the optimal bandwidth. Popular methods of bandwidth selections up to date include plug-in bandwidths (in particular, rule-of-thumb bandwidths, see for example Scott (1992) and Wand and Jones (1995)) and cross-validation bandwidths (see for example Hall et al. (1992) and Park and Marron (1992)).

The one-step SMLE chooses to maximize the semiparametric likelihood

$$\hat{l}_n(\boldsymbol{\theta}|\mathbf{u}) = \frac{1}{n}\sum\nolimits_{t=1}^{n} \ln\left[|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|^{-1/2}\hat{f}_n\left(\mathbf{e}_t;\boldsymbol{\theta}\right)\right] \tag{3.11}$$

with respect to $\boldsymbol{\theta}$.

Remark 8. The proposed one-step MSMLE method searches for a local maximum of the semiparametric likelihood, and hence the estimated results are very sensitive to the initial values chosen. Different sets of initial values may lead to very different convergence times and estimation results. Although inefficient, the traditional MLEs provide a feasible set of initial values.

3.1.3 Asymptotic Properties of one-step MSMLE

Let us now study the asymptotic properties of the proposed one-step MSMLE. For a diagonal bandwidth matrix $\mathbf{H}$ in (3.7), without loss of generality, we can write

$$\mathbf{H} = |\mathbf{H}|^{1/d}\,\mathbf{A} = h\mathbf{A} \tag{3.12}$$

where $h = |\mathbf{H}|^{1/d}$, and $\mathbf{A}$ is a diagonal matrix with $|\mathbf{A}| = 1$.

Proposition 8. *(Pointwise convergence in probability of $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$) The density estimate $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$ converges pointwise in probability to the true density $f\left(\mathbf{e};\boldsymbol{\theta}\right)$ (that is, $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right) \xrightarrow{p} f\left(\mathbf{e};\boldsymbol{\theta}\right)$ pointwise) when $h \to 0$ and $nh^d = n\left|\mathbf{H}\right| \to \infty$ as $n \to \infty$. (This can easily be proved by using results in Scott (1992).)*

Proposition 9. *(Uniform convergence in probability of $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$) The density estimate $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$ converges uniformly in probability to the true density*

$f(\mathbf{e};\boldsymbol{\theta})$ *(that is,* $\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) \xrightarrow{p} f(\mathbf{e};\boldsymbol{\theta})$ *uniformly), if*

(i) $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ *is uniformly continuous,*

(ii) the Fourier transforms of $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ and $K\left(\mathbf{H}^{-1}\left[\mathbf{e}-\mathbf{e}_s\right]\right)$ exist, and

(iii) $nh^{2d} \rightarrow \infty$ as $n \rightarrow \infty$.(The proof of this proposition is in Appendix A.2.)

Let the semiparametric score be defined as

$$\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta})=\frac{\partial \ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{3.13}$$

and the semiparametric Hessian be defined as

$$\hat{\mathbb{H}}(\mathbf{u},\boldsymbol{\theta}) = \frac{\partial \hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathbf{T}}} = \frac{\partial^2 \ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}. \tag{3.14}$$

Also let

$$\begin{aligned} \mathbb{Q} &= Cov\left[\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta}^*)\right] \\ &= \mathbb{E}\left[\frac{\partial \ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} \frac{\partial \ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}^T}\right], \end{aligned} \tag{3.15}$$

and

$$\begin{aligned} \mathbb{J} &= -\mathbb{E}\left[\hat{\mathbb{H}}(\mathbf{u},\boldsymbol{\theta}^*)\right] \\ &= -\mathbb{E}\left[\frac{\partial^2 \ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right]. \end{aligned} \tag{3.16}$$

Theorem 10. *(Weak Consistency and Strong Consistency of $\hat{\boldsymbol{\theta}}_{1SMLE}$) Suppose that*

(i) $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ *is measurable on* (Ω, F) *for all* $\boldsymbol{\theta}$ *in* Θ*, where* $\mathbf{e} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta})\mathbf{u}$,

(ii) $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ *is continuous in* $\boldsymbol{\theta}$*, and*

(iii) $\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) \xrightarrow{p} f(\mathbf{e};\boldsymbol{\theta})$ *pointwise (conditions for this to be true are outlined in Proposition 8).*

Then, the one-step SMLE $\hat{\boldsymbol{\theta}}_{1SMLE}$ *converges pointwise in probability to the true parameter* $\boldsymbol{\theta}^*$*, that is,*

$$\hat{\boldsymbol{\theta}}_{1SMLE} \xrightarrow{p} \boldsymbol{\theta}^* \text{ pointwise.}$$

The one-step SMLE $\hat{\boldsymbol{\theta}}_{1SMLE}$ *also converges pointwise almost surely to the true parameter* $\boldsymbol{\theta}^*$*, that is,*

$$\hat{\boldsymbol{\theta}}_{1SMLE} \xrightarrow{a.s.} \boldsymbol{\theta}^* \text{ pointwise.}$$

(The proof of this theorem is in Appendix A.3.)

Theorem 11. *(Asymptotic normality of $\hat{\boldsymbol{\theta}}_{1SMLE}$) Suppose that*

(i) $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ *is twice continuously differentiable with respect to* $\boldsymbol{\theta}$*,*

(ii) $\hat{\mathbb{H}}(\mathbf{u},\boldsymbol{\theta})$ *is non-singular, and*

(iii) $\frac{1}{n}\sum_{t=1}^{n} \hat{\mathbb{H}}(\mathbf{u}_t,\tilde{\boldsymbol{\theta}}) \xrightarrow{p} -\mathbb{J}$.

Then, the one-step SMLE $\hat{\boldsymbol{\theta}}_{1SMLE}$ follows that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{1SMLE} - \boldsymbol{\theta}^*) \overset{Dist}{\to} N(\mathbf{0}, \mathbb{J}^{-1}\mathbb{Q}\mathbb{J}^{-1}). \tag{3.17}$$

(The proof of this theorem is in Appendix A.4.)

Theorem 12. *(Asymptotic efficiency of $\hat{\boldsymbol{\theta}}_{1SMLE}$) The one-step SMLE $\hat{\boldsymbol{\theta}}_{1SMLE}$ is asymptotically efficient that*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{1SMLE} - \boldsymbol{\theta}^*) \overset{Dist}{\to} N(\mathbf{0}, \mathbb{I}^{-1}(\boldsymbol{\theta}^*)) \tag{3.18}$$

where $\mathbb{I}(\boldsymbol{\theta}^*)$ is the Fisher Information, when

(i) $\hat{\boldsymbol{\theta}}_{1SMLE}$ *is asymptotically normally distributed (conditions for this to be true are outlined in Theorem 11),*

(ii) $\hat{f}_n(\mathbf{e}; \boldsymbol{\theta}) \overset{p}{\to} f(\mathbf{e}; \boldsymbol{\theta})$ *uniformly (conditions for this to be true are outlined in Proposition 9),*

(iii) $\partial \ln K(\boldsymbol{\omega})/\partial\boldsymbol{\theta}$ *converges uniformly on Θ, and*

(iv) $\partial^2 \ln K(\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ *converges uniformly on Θ.(The proof of this theorem is in Appendix A.5.)*

3.1.4 Small Sample Performance of One-step MSMLE

In this section, we provide numerical studies comparing the performances of 500 replications of QMLE (assuming normality), the proposed one-step MSMLE, two-step MSMLE (following Hafner and Rombouts (2007)) and MLE in a small sample size of 100.

Suppose that $\mathbf{e}_t = \begin{bmatrix} e_{1,t} & e_{2,t} \end{bmatrix}^T$ is i.i.d with mean equals $\mathbf{0}$ and covariance equals $\mathbf{I}_2$, where $e_{1,t}$ is of asymmetric Laplace distribution (location $m = 8/\sqrt{82}$, scale $\lambda = \sqrt{82}/3$, asymmetry $\varkappa = 3$) and $e_{2,t}$ is of Asymmetric Laplace distribution (location $m = -8/\sqrt{82}$, scale $\lambda = \sqrt{82}/3$, asymmetry $\varkappa = 1/3$). Suppose further that $\mathbf{u}_t = \mathbf{\Sigma}_t^{1/2}\mathbf{e}_t$ follows a Baba, Engle, Kraft and Kroner (BEKK) MGARCH model (Engle and Kroner 1995) where the covariance is

$$\mathbf{\Sigma}_t(\boldsymbol{\theta}) = \mathbf{M} + \mathbf{A}_1^T\mathbf{u}_{t-1}\mathbf{u}_{t-1}^T\mathbf{A}_1 + \mathbf{B}_1^T\mathbf{\Sigma}_{t-1}\mathbf{B}_1,$$

and the parameters $\boldsymbol{\theta} = \begin{bmatrix} M_{0,11} & M_{0,12} & M_{0,22} & A_{1,11} & A_{1,12} & A_{1,22} & B_{1,11} & B_{1,12} & B_{1,22} \end{bmatrix}^T$ are chosen to be

$$\mathbf{M} = \begin{bmatrix} M_{0,11} & M_{0,12} \\ M_{0,12} & M_{0,22} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.1 \end{bmatrix},$$

$$\mathbf{A}_1 = \begin{bmatrix} A_{1,11} & A_{1,12} \\ 0 & A_{1,22} \end{bmatrix} = \begin{bmatrix} 0.25 & 0.15 \\ 0 & 0.15 \end{bmatrix}, \text{ and}$$

$$\mathbf{B}_1 = \begin{bmatrix} B_{1,11} & B_{1,12} \\ 0 & B_{1,22} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.3 \\ 0 & 0.1 \end{bmatrix}.$$

A one-step MSMLE utilizes a kernel density estimate. Recall that the choice of kernel is not crucial for the accuracy of density estimates. For elliptically distributed series, a Gaussian kernel is an appropriate candidate.

The Gaussian kernel for a d-variate standardized innovation $\mathbf{e}$ is given by

$$K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right) = (2\pi)^{-d/2}\,|\mathbf{H}|^{-1}\exp\left(-\frac{1}{2}(\mathbf{e}-\mathbf{e}_s)^T\mathbf{H}^{-2}(\mathbf{e}-\mathbf{e}_s)\right). \quad (3.19)$$

Scott's rule of thumb in $\mathbb{R}^d$ (Equation (6.42) in Scott (1992)) suggests that, for a Gaussian kernel, a feasible choice of bandwidth is

$$\tilde{\mathbf{H}} = n^{-1/(d+4)}\tilde{\mathbf{V}}, \quad (3.20)$$

where $\tilde{\mathbf{V}}$ is the prior belief of the covariance of the standardized innovation $\mathbf{e}$. Hence, an appropriate choice of $\mathbf{H}$ could be

$$\tilde{\mathbf{H}} = n^{-1/(d+4)}\mathbf{I}_d. \quad (3.21)$$

Note that, if $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t(\boldsymbol{\theta}))$ and we choose $\mathbf{H} = \mathbf{\Sigma}_t 1/2(\boldsymbol{\theta})$ and $K(.)$ to be a normal kernel, then $|\mathbf{H}|^{-1}K\left(\mathbf{H}^{-1}\mathbf{u}_t\right)$ is the true density of $\mathbf{u}_t$.

Substituting the kernel (3.19) where $d = 2$ and the bandwidth (3.21) into (3.7) we obtain the density estimate of $\mathbf{e}$ as

$$\hat{f}_n(\mathbf{e}, \boldsymbol{\theta}) = \frac{n^{1/3}}{2\pi}\sum_{s=1}^{n}\exp\left(-\frac{n^{1/3}}{2}(\mathbf{e}-\mathbf{e}_s)^T(\mathbf{e}-\mathbf{e}_s)\right). \quad (3.22)$$

Substituting the above density estimate into (3.11) gives the semi-parametric log-likelihood that

$$\hat{l}_n(\boldsymbol{\theta}|\mathbf{u}) = \left\{ \begin{array}{c} \frac{n}{3}\ln n - n\ln(2\pi) \\ +\sum_{t=1}^{n} \left[\begin{array}{c} -\frac{1}{2}\ln|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})| \\ +\ln\sum_{s=1}^{n}\exp\left(-\frac{n^{\frac{1}{3}}}{2}(\mathbf{e}_t-\mathbf{e}_s)^T(\mathbf{e}_t-\mathbf{e}_s)\right) \end{array} \right] \end{array} \right\}. \quad (3.23)$$

The one-step MSMLE $\hat{\boldsymbol{\theta}}_{1MSMLE}$ is obtained by maximizing (3.23).

Comparing the choice of bandwidth in (3.21) with the structure of bandwidth in (3.12), in this case, we can define

$$h = n^{-1/6} \text{ and } \mathbf{A} = \mathbf{I}_d, \quad (3.24)$$

which satisfy

$$\lim_{n\to\infty} h = \lim_{n\to\infty} n^{-1/6} = 0, \quad (3.25)$$

$$\lim_{n\to\infty} nh^d = \lim_{n\to\infty} n^{1/3} = \infty, \text{ and}$$

$$\lim_{n\to\infty} nh^{2d} = \lim_{n\to\infty} n^{2/3} = \infty. \quad (3.26)$$

Hence, according to Proposition 8, 9, Theorem 10 and 11, the kernel estimate is consistent[1] and the resulting one-step MSMLE is consistent and asymptotically normally distributed.

We record in Table 3.1 the averages and the standard deviations based on the 500 replications of the estimates given by a QMLE (assuming nor-

[1]This set of choices of kernel function and bandwidth in (3.19) and (3.21) will not provide strong consistency of the resulting kernel estimate, when dimension is greater than 3.

mality), the proposed one-step MSMLE, a two-step MSMLE and the MLE on a bivariate sample of 100 data points. Our results show that, on average, all four estimation techniques will provide considerably accurate point estimates. Both of the MSMLE techniques outperform QMLE as expected. One-step MSMLE provides more efficient estimates than the two-step technique in this particular example, where the One-step MSMLEs on $\mathbf{M}$, $\mathbf{A}_1$ and $B_{1,11}$ have the smaller standard deviations than the two-step MSMLEs. This is due to the fact that the shape of a bivariate asymmetric Laplace distribution is significantly different from that of a bivariate normal distribution. The poor first step QMLE is a burden on the performance of the two-step SMLE.

We will apply the one-step MSMLE on Australia market returns (the estimation window contains 120 effective observations) in the following section.

3.2 Asset Returns in the Australian Market

An investment portfolio often consists of three asset classes, equity, debt and cash. Understanding of the securities in these three markets is crucial for the success of the portfolio. We will provide a study of the evolution of security prices in Australian markets below.

We obtain a 284-month return series from Australian indexed data dating from December 1989 to August 2013. The S&P/ASX200 index serves as the equity class, which is obtained from Table 8.7 "the Australian Stock Market Indexes, MONTHLY" in Chapter 8 of the report on Australian Economic

Parameter	$\hat{\boldsymbol{\theta}}_{QMLE}$	$\hat{\boldsymbol{\theta}}_{1SMLE}$	$\hat{\boldsymbol{\theta}}_{2SMLE}$	$\hat{\boldsymbol{\theta}}_{MLE}$
$M_{0,11} = 0.2$	0.1430 (0.0761)	0.1710 (0.0244)	0.1831 (0.0463)	0.2001 (0.0162)
$M_{0,12} = 0.1$	0.0701 (0.0560)	0.1138 (0.0124)	0.1234 (0.0667)	0.999 (0.0107)
$M_{0,22} = 0.1$	0.0589 (0.0453)	0.0805 (0.0110)	0.0748 (0.0638)	0.0986 (0.0109)
$A_{1,11} = 0.25$	0.2327 (0.2708)	0.2502 (0.0053)	0.2522 (0.0360)	0.2473 (0.0512)
$A_{1,12} = 0.15$	0.1641 (0.2871)	0.1461 (0.0097)	0.1518 (0.0248)	0.1455 (0.0368)
$A_{1,22} = 0.15$	0.1196 (0.2718)	0.1439 (0.0126)	0.1494 (0.0142)	0.1481 (0.0487)
$B_{1,11} = 0.2$	0.3086 (0.4211)	0.2056 (0.0079)	0.2001 (0.0183)	0.2007 (0.0207)
$B_{1,12} = 0.3$	0.3625 (0.3074)	0.2874 (0.0162)	0.2986 (0.0135)	0.1455 (0.0368)
$B_{1,22} = 0.1$	0.1129 (0.3376)	0.0906 (0.0122)	0.0989 (0.0097)	0.1481 (0.0487)

Table 3.1: Averages of QLEs, one-step MSMLEs, two-step MSMLEs and MLEs for simulated samples of size 100 (the standard deviations of the estimates are outlined in blackets).

Indicators published by the Australian Bureau of Statistics[2].The (logarithmic) return on equity is calculated as the logarithm of the ratio between two consecutive prices. The 90-day Bank Accepted Bill (BAB) index serves as the debt class. This is obtained from Statistical Table F1.1 "Interest Rates and Yields - Money Market - Monthly" published by the Reserve Bank of

[2]Data for period dating from December 1989 to February 2009 is available at `http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1350.0Apr%202009?OpenDocument`. Data for period dating from April 2009 to April 2010 is available at `http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1350.0Jun%202010?OpenDocument`. Data for period dating from May 2010 to May 2012 is available at `http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1350.0Jul%202012?OpenDocument`. Data for period dating from June 2012 to August 2013 is available at `http://www.rba.gov.au/statistics/tables/index.html#share-mkts`.

Australia (RBA)[3]. BAB index yields are quoted as 100 less price. Hence, we can retrieve the price series from the yield, and the return on BABs can then be calculated as the logarithm of the ratio between two consecutive prices. The one-month bank's term deposit index serves as the cash class, which is obtained from Statistical Table F4 "Retail Deposit and Investment Rates" published by the RBA[4]. Term deposit yields are nominal returns quoted in percentage per year, and these can be converted into monthly returns by dividing by 12.

Let $\mathbf{r}_t = \begin{bmatrix} r_{1,t} & r_{2,t} & r_{3,t} \end{bmatrix}^T$ be the vector of asset returns at time t, where r_1 is the return on equity, r_2 is the return on debt and r_3 is the return on cash. The average monthly returns are given by

$$\bar{r}_{i,monthly} = \frac{1}{284} \sum_{t=1}^{284} r_{i,t}, \; i = 1, 2, 3. \tag{3.27}$$

Returns on equity and debt are logarithmic, and hence the average annual returns are given by

$$\bar{r}_{i,annual} = \frac{1}{284} \sum_{t=1}^{284} \left(e^{12 r_{i,t}} - 1 \right), \; i = 1, 2. \tag{3.28}$$

Returns on cash, on the other hand, are quoted on a compound interest basis, and hence the average annual return is given by

$$\bar{r}_{3,annual} = \frac{1}{284} \sum_{t=1}^{284} (1 + r_{3,t})^{12} - 1.$$

[3]Data is acquired from `http://www.rba.gov.au/statistics/tables/index.html#interest-rates`.

[4]Data is acquired from `http://www.rba.gov.au/statistics/tables/index.html#interest-rates`.

The average monthly volatility is given by

$$\bar{\sigma}_{i,monthly} = \sqrt{\frac{1}{284}\sum\nolimits_{t=1}^{284} r_{i,t}^2}, \ i = 1, 2, 3; \tag{3.29}$$

and the average annual volatility is given by

$$\bar{\sigma}_{i,annual} = \sqrt{\frac{12}{284}\sum\nolimits_{t=1}^{284} r_{i,t}^2}, \ i = 1, 2, 3. \tag{3.30}$$

Figure 3.1 plots all three monthly return series, and Table 3.2 provides a summary of both monthly and annual returns.

Summary Statistics	S&P/ASX200	90-day BABs	Cash
Average Monthly Return	0.3998%	0.0601%	0.3006%
Average Monthly Volatility	0.0387%	0.0028%	0.0037%
Average Annual Return	15.61%	0.78%	3.70%
Average Annual Volatility	0.1342%	0.0098%	0.0128%

Table 3.2: Summary statistics of asset returns

3.2.1 A Rolling Window Estimate

We now apply the one-step MSMLE method developed in Section 3.1 to study the multivariate evolution of Australian asset returns. Our study is based on a VAR-MGARCH model of the three asset returns discussed in Section 3.2, and we follow Pesaran and Timmermann (2000), who conducted a similar "rolling window" study of how the parameters in a model of UK stock returns evolve over time.

We include two sets of explanatory variables in our specification of returns:

$$\mathbf{Q}_t = \begin{bmatrix} 1 & \mathbf{X}_t^T \end{bmatrix}^T, ^5$$

where

$$\mathbf{X}_t = \begin{bmatrix} r_{1,t-1} \\ r_{1,t-1} \\ r_{3,t-1} \end{bmatrix}.$$

The explanatory regressor $\mathbf{X}_t$ contains first-order lagged returns, r_{t-1}.

Moreover, we observe heteroskedasticity in asset returns (Figures 3.2 to 3.4) which Pesaran and Timmermann (2000) did not allow for. We capture this effect by utilizing an MGARCH(1,1) model.

The complete model is

$$\underset{(3\times1)}{\mathbf{r}_t} = \underset{(3\times4)}{\mathbf{C}^T}\underset{(4\times1)}{\mathbf{Q}_t} + \mathbf{v}_t \tag{3.31}$$

$$= \begin{bmatrix} C_{10} & C_{11} & C_{12} & 0 \\ C_{20} & C_{21} & C_{22} & 0 \\ C_{30} & 0 & 0 & C_{33} \end{bmatrix} \begin{bmatrix} 1 \\ r_{1,t-1} \\ r_{2,t-1} \\ r_{3,t-1} \end{bmatrix} + \mathbf{v}_t \tag{3.32}$$

$$\mathbf{v}_t = \Sigma_t^{1/2}\mathbf{e}_t \tag{3.33}$$

[5] Pesaran and Timmermann (2000) include a seasonal dummy variable which equals 1 if the return is for January and 0 otherwise. Their model also include a dummy variable D_t. They exclude D_t in the first instance and make a one-step-ahead prediction on returns for the following month. If any of the prediction errors are greater than three standard deviations of the subset of returns included in a regression window, then D_t is set to a vector of ones and the model is reestimated. However, we find that the inclusion of the seasonal dummy and D_t is not necessary. For further details, refer to Pesaran and Timmermann (2000).

$$\underset{(3\times3)}{\mathbf{\Sigma}_t} = \mathbf{A}_0^T\mathbf{A}_0 + \mathbf{A}_1^T\mathbf{v}_{t-1}\mathbf{v}_{t-1}^T\mathbf{A}_1 + \mathbf{B}_1^T\mathbf{\Sigma}_{t-1}\mathbf{B}_1 \tag{3.34}$$

where $\mathbf{e}_t$ is i.i.d. with mean $\mathbf{0}$ and covariance of $\mathbf{I}_3$, and Σ_t is the conditional covariance of asset returns at time t represented by a Baba, Engle, Kraft and Kroner (BEKK) MGARCH model (Engle and Kroner 1995), in which the parameters $\mathbf{A}_0$, $\mathbf{A}_1$ and $\mathbf{B}_1$ are in upper-triangular form. Note that we do not incorporate interactions between equity and cash, or between debt and cash by requiring these coefficients equal 0. This is because, as seen in Figures 3.2 to 3.4 (plots of returns on each asset class), returns on cash exhibit weak or no correlation with returns on equity or debt.

We set up an estimation window containing 120 effective months of observations, and this window is moved forward one observation at a time. The estimated model is updated 164 times as the sample window moves forward.

Let us now construct the semiparametric log-likelihood about parameters $\boldsymbol{\theta} = [\mathbf{A}_0, \mathbf{A}_1, \mathbf{B}_1]$. We again choose a Gaussian kernel and the corresponding Bandwidth by Scott's rule of thumb in $\mathbb{R}^d$ (Equation (6.42) in Scott (1992)) for the standardized innovation $\mathbf{e} = \mathbf{\Sigma}^{-1/2}(\mathbf{r} - \mathbf{QC})$. Thus, in analogy with the example in Section 3.1.4, the semiparametric log-likelihood is

$$\hat{l}_n(\boldsymbol{\theta}|\mathbf{u}) = \left\{ \begin{array}{c} \frac{1}{7}\ln n - \frac{3}{2}\ln(2\pi) \\ +\frac{1}{n}\sum_{t=1}^{n} \left[\begin{array}{c} -\frac{1}{2}\ln|\Sigma_t| \\ +\ln\sum_{s=1}^{n}\exp\left(-\frac{n^{\frac{2}{7}}}{2}(\mathbf{e}_t - \mathbf{e}_s)^T(\mathbf{e}_t - \mathbf{e}_s)\right) \end{array} \right] \end{array} \right\}, \tag{3.35}$$

and the one-step MSMLE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{1MSMLE} = [\hat{\mathbf{A}}_0, \hat{\mathbf{A}}_1, \hat{\mathbf{B}}_1]$, is obtained by maximizing (3.35). Detailed derivations of (3.22) and (3.35) are provided in

Appendix A.7.

Note that, we have $nh^{2d} = n^{\frac{4-d}{d+4}} = n^{\frac{1}{7}} \to \infty$ as $n \to \infty$. Hence, according to Proposition 8, 9, Theorem 10 and 11, the kernel estimate is consistent and the resulting one-step MSMLE is consistent and asymptotically normally distributed.

The log-likelihood in (3.35) is a conditional log-likelihood, conditioned on the initial value $\mathbf{r}_0$. The resulting estimates are M-estimators, which are numerically the same as the equation-by-equation Ordinary Least Square (OLS) estimators (Hayashi 2000). Hence, if the parametric return equation (3.31) is correctly specified, the estimate of the parameter $\mathbf{C}$ is

$$\hat{\mathbf{C}} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{r}, \tag{3.36}$$

where $\mathbf{r} = [\begin{array}{cccc} \mathbf{r}_1 & \mathbf{r}_2 & \dots & \mathbf{r}_{120} \end{array}]^T$ and $\mathbf{Q} = [\begin{array}{cccc} \mathbf{Q}_1 & \mathbf{Q}_2 & \dots & \mathbf{Q}_{120} \end{array}]^T$. The estimate, $\hat{\mathbf{C}}$, is consistent and asymptotically unbiased.

3.3 Estimated Results

The estimated unconditional mean returns are given by

$$\hat{\boldsymbol{\mu}} = (\mathbf{I} - \hat{\mathbf{C}}_1)^{-1}\hat{\mathbf{C}}_0 \tag{3.37}$$

where $\hat{\mathbf{C}}_0$ contains the constant coefficients and $\hat{\mathbf{C}}_1$ contains the coefficients for $\mathbf{X}_t$. The vectorized estimated unconditional covariance matrix is given by

$$vec(\hat{\boldsymbol{\Sigma}}) = [\mathbf{I}_9 - (\hat{\mathbf{A}}_1^T \otimes \hat{\mathbf{A}}_1^T) - (\hat{\mathbf{B}}_1^T \otimes \hat{\mathbf{B}}_1^T)]^{-1} vec(\hat{\mathbf{A}}_0^T \hat{\mathbf{A}}_0), \tag{3.38}$$

where $\mathbf{I}_9$ is a (9×9) identity matrix. Since we are using rolling windows for estimation, each of $\hat{\mathbf{C}}_0$, $\hat{\mathbf{C}}_1$, $\hat{\mathbf{A}}_0$, $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{B}}_1$ are updated as windows are rolled forward.

3.3.1 The Return Equations

Figure 3.5 plots the estimated autocorrelation (AR) coefficients for returns on the S&P/ASX 200 index, Australian 90-day BABs and RBA 1-month bank deposits across time. Figure 3.6 plots the estimated first lag effects of returns on Australian 90-day BABs and returns on S&P ASX 200 on each other. The time-varying AR coefficients support the use of the rolling estimation method, and shows that returns are path dependent.

Returns on the S&P/ASX 200 index exhibit clear mean-reverting pattern from 1989 to early 2008, while returns on Australian 90-day BABs and RBA 1-month bank deposits exhibit mean-aversion (Figures 3.2 to 3.4). Fama and French (1986) suggested that, for long horizons, AR coefficients should be negative for mean-reverting returns and positive for mean-averting returns. Our estimated coefficients are consistent with this claim, in that, the AR coefficients for returns on Australian 90-day BABs and RBA 1-month bank deposits are always positive and the AR coefficients for returns on the S&P/ASX 200 index are mostly negative. Similar conclusions for negative autocorrelation coefficients for returns on Australian equities can be found in research papers published by RBA (see McNelis (1993) and Cecchetti et al. (2005)) and MLC[6] (see Napper (2008)).

On the other hand, the lagged effects due to 90-day BABs returns on

[6]MLC is the wealth management division of the National Australia Bank (NAB)

S&P/ASX200 returns change dramatically while the effects due to equity returns on returns on debt are insignificant over the observation period (Figure 3.6), as the magnitudes of 90-day BABs returns are on average 10 times greater than the magnitudes of 90-day BABs returns.

Figure 3.7 plots the estimated ten-year mean returns on the S&P/ASX 200 index and Australian 90-day BABs across time. Our estimated ten-year mean returns on the S&P/ASX 200 index increase for the 10-year windows ending before November 2000 and then decrease for the 10-year windows ending before late 2003. This is consistent with the historical data. We observe largely negative returns on the S&P/ASX 200 index in 1990 corresponding to the early 1990's recession. Australian investors in equity enjoyed a buoyant period from mid-year 1999 to late 2000, before the recession from mid-year 2001 to early 2003, which corresponded to the collapse in the US equity market due to the growth and collapse in IT (Anderson et al. 2010). The Australian equity market recovered from late 2003 until the outburst of the Global Financial Crisis (GFC) in 2008. More information about the 2008 subprime crisis can be found in Shiller (2008). The dramatic decreases in our estimated ten-year mean returns on the S&P/ASX 200 index for the 10-year windows ending in late 2007 to early 2009 and the maintained lower levels thereafter correspond to this recent crisis.

We also find an interesting pattern when comparing unconditional mean returns on S&P/ASX 200 (equity) and unconditional mean returns on Australian 90-day BABs (debt), in that there is a slightly negative correlation between returns in the equity market and the debt market. Diggle and Brooks (2007) drew a similar conclusion based on their studies of the relationships between target cash rates and investment asset returns in Australia from January 1990 to December 2000. They found that returns on the ASX All Ordinaries (equity) were negatively correlated with RBA target cash rates while yields on Australian Government 13-week Treasury Notes (debt) were positively correlated with RBA target cash rates.

3.3.2 The Covariance Equations

The estimated variance in returns on the S&P/ASX 200 index, Australian 90-day BABs and RBA 1-month bank deposits all increase over time, and this provides additional evidence that return paths are time dependent. The results are comparatively more stable for the ten-year windows before 2007 than for the 10-year windows after 2007. There were a number of financial crises in the first decade of the twenty-first century, for example the early 2000s recession, but by far, the 2008 GFC had the strongest and longest impact. While most countries had not yet recovered from the 2008 GFC, a global recession began in late 2009. The increasing trends that we see in estimated variance are most likely related to this series of crises.

Shiller (2008) suggested that there was a grim feedback loop, in that the failure in the US property price bubble (subprime lending problem) fed into major failures in Europe, which in turn also fed back into the US. The collapse in the property market also contributed to energy and food crises. This loop led to the global recession which began in late 2009, which was the greatest global recession since World War II . There is a great volume of literature addressing the 2008 GFC and the following global recession, for example, see the associated commentary in the International Monetary Fund (2009), International Monetary Fund (2010) and Gore (2010).

3.4 Discussion and Conclusion

We developed a one-step multivariate semiparametric maximum likelihood estimation (one-step MSMLE) technique to study MGARCH models in this chapter. We have shown that the one-step MSMLE provides consistent and asymptotically normally distributed estimates of the MGARCH parameters. We applied this estimation technique to study Australia's security returns. We divided the return data set into rolling windows each containing 120 effective observations and estimated a VAR(1)-MGARCH(1,1) model for each window. We found that parameters in the VAR-MGARCH model change over time, providing strong evidence of path-dependency in returns. Our results also show that returns on equity are mean-reverting, while returns on debts and cash exhibit mean-aversion, while the variances in all asset classes increase over time.

The estimated parameters in this chapter suggest that Australian asset returns are path-dependent. The path-dependency is found in both return and covariance levels. Since the investment rules we developed in chapter 2 relate to path-independent processes, it follows that these rules might not be optimal for Australian market. We will proceed to the development of optimal investment strategies for path-dependent returns in the next chapter.

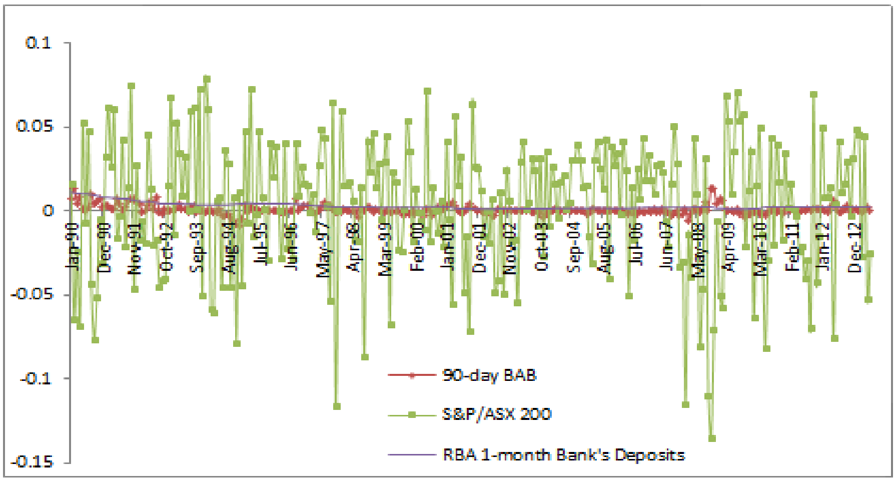

Figure 3.1: Monthly index returns of S&P/ASX200, ASX 90-day Bank Accepted Bills and RBA 1-month Bank's Deposits from January 1990 to August 2013.

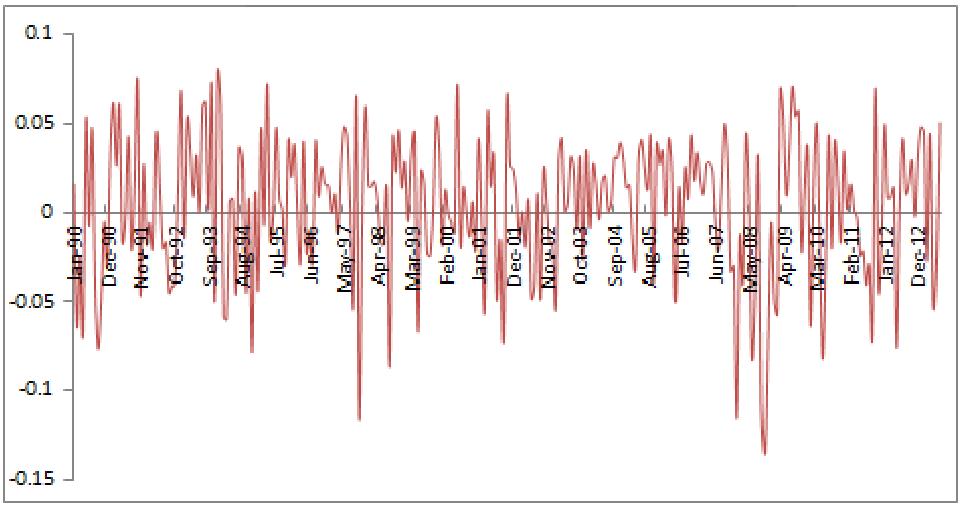

Figure 3.2: Monthly index returns of S&P/ASX200 from January 1990 to August 2013.

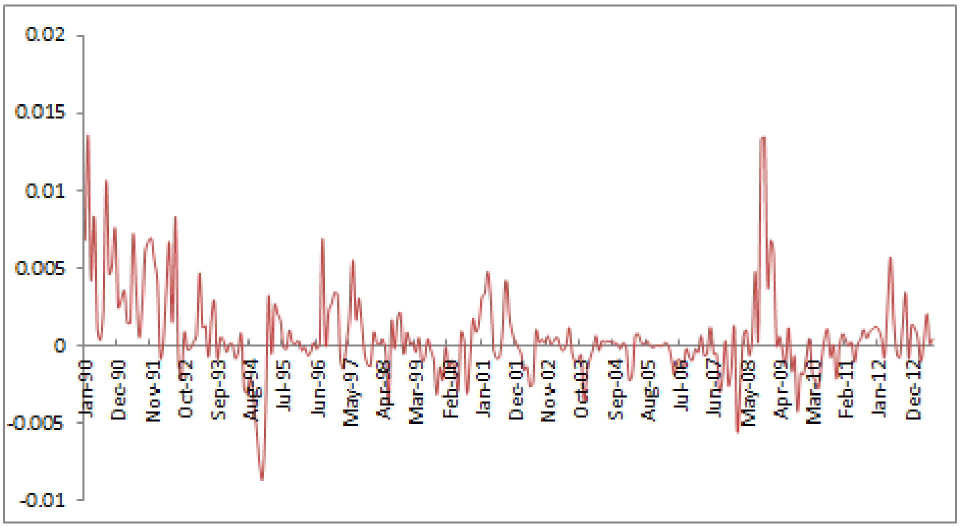

Figure 3.3: Monthly index returns of ASX 90-day Bank Accepted Bills from January 1990 to August 2013.

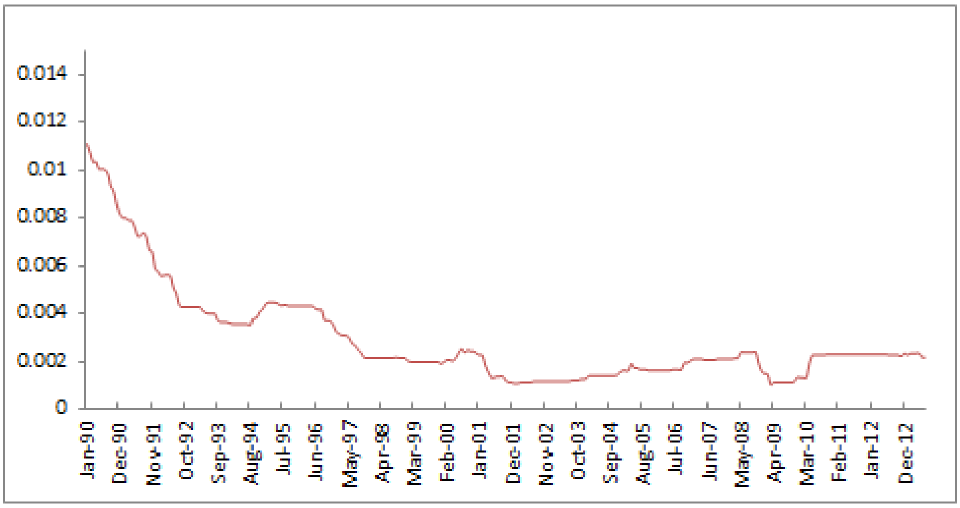

Figure 3.4: Monthly index returns of RBA 1-month Banks' Deposits from January 1990 to August 2013.

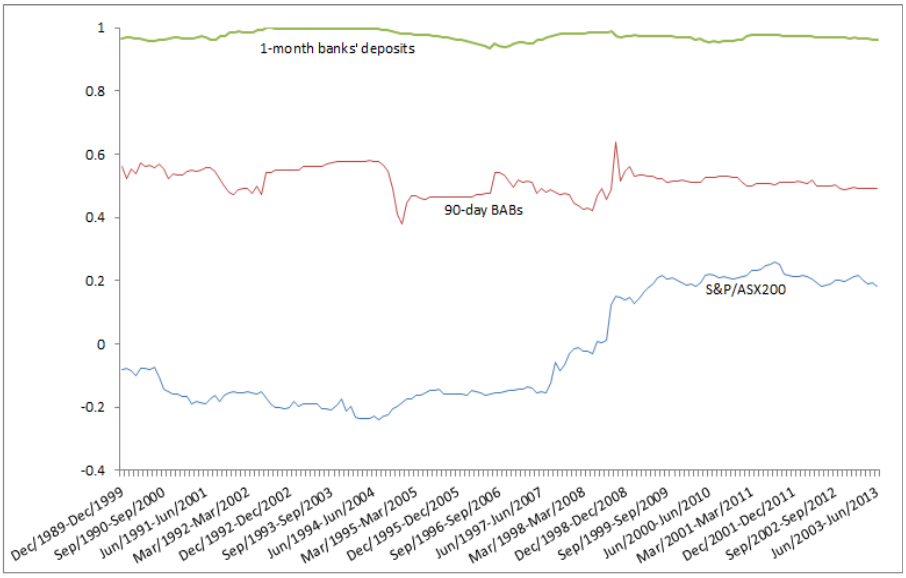

Figure 3.5: First lag autocorrelation coefficients for returns on the S&P/ASX 200 index, Australian 90-day BABs and 1-month bank deposits (X-axis indicates the end of each window)

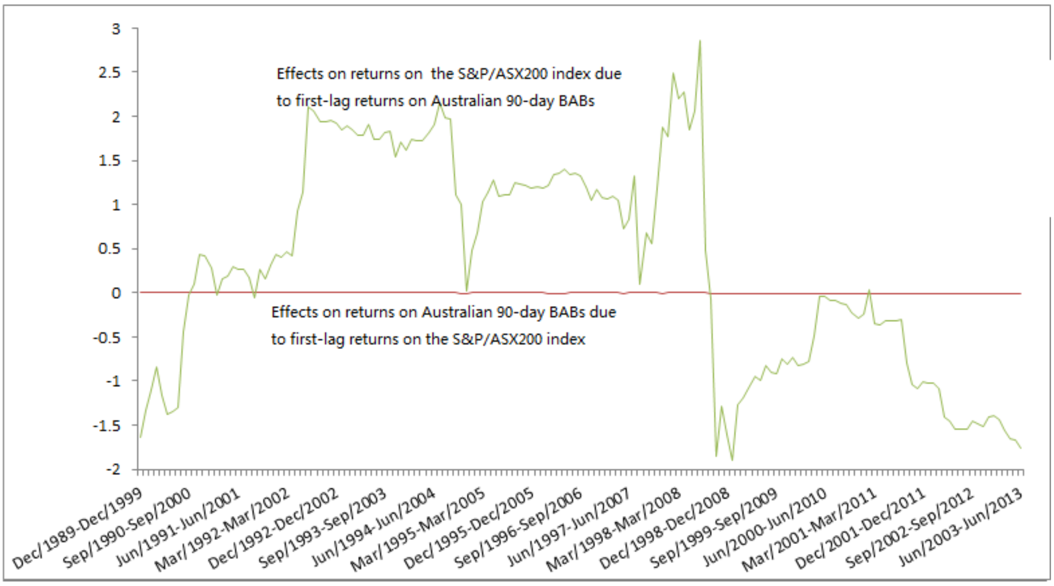

Figure 3.6: First lag cross asset coefficients for returns on the S&P/ASX 200 index and Australian 90-day BABs (X-axis indicates the end of each window)

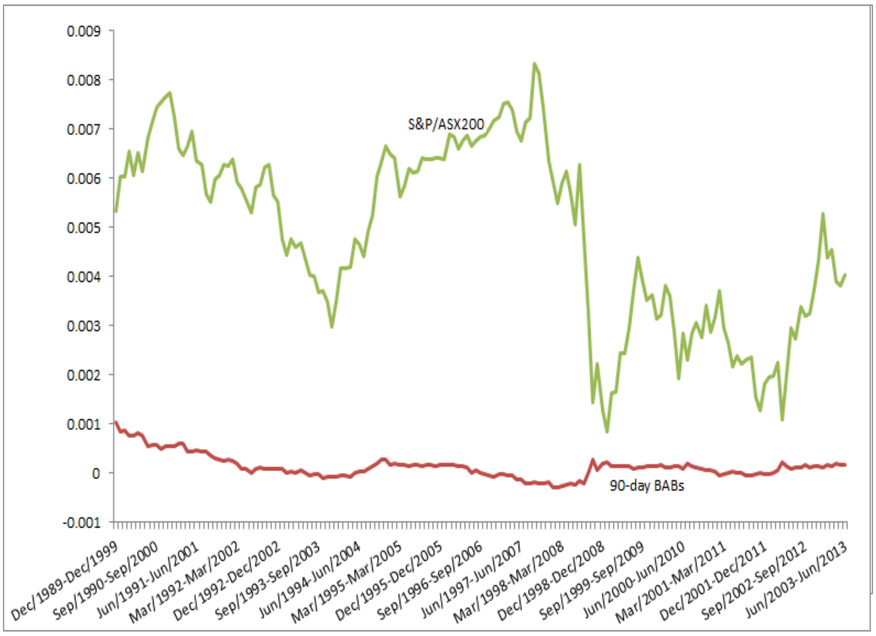

Figure 3.7: Estimated ten-year mean returns on the S&P/ASX200 index and Australian 90-day BABs (X-axis indicates the end of each window)

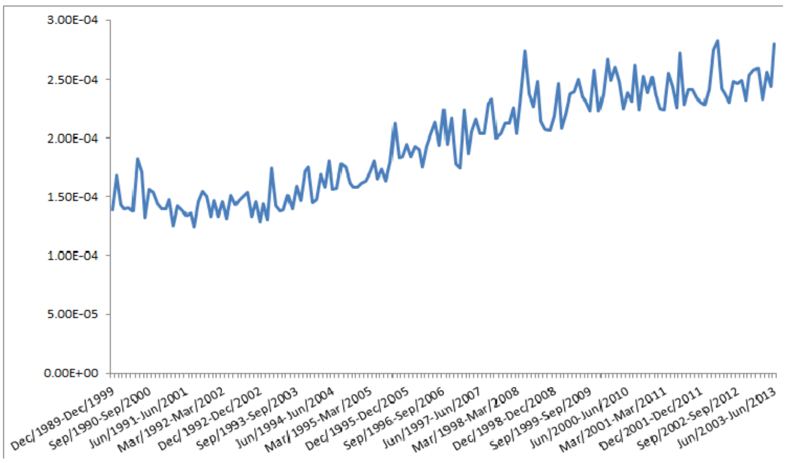

Figure 3.8: Estimated ren-year variances in the S&P/ASX200 index returns (X-axis indicates the end of each window)

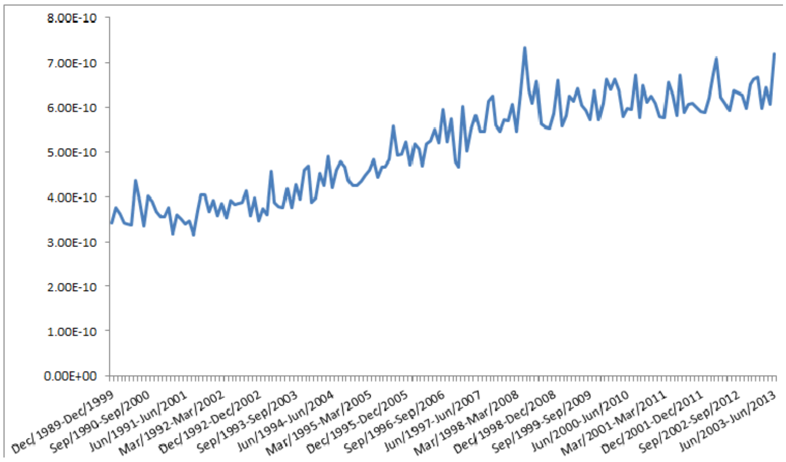

Figure 3.9: Estimated ten-year variances in Australian 90-day BABs returns (X-axis indicates the end of each window)

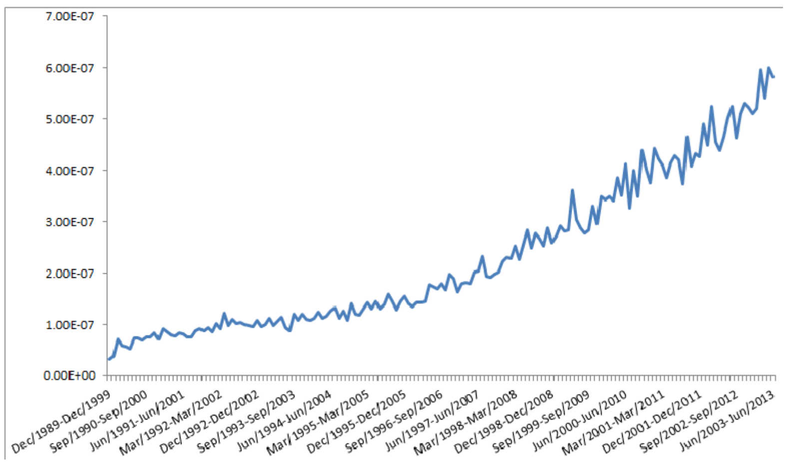

Figure 3.10: Estimated ten-year variances in RBA 1-month banks' deposits returns (X-axis indicates the end of each window)

Chapter 4

VAR-MGARCH Returns and Asset Allocation

We have studied the evolution of Australian security prices in Chapter 3. We found autoregressive (AR) effects in both the return and volatility levels. This is evidence showing that asset returns may be path-dependent; that is, current returns depend on past returns. According to Kassberger and Liebmann (2012), path-dependent strategies are preferred if the pricing model of the risky assets is itself path-dependent, which implies that the portfolio rules we developed in Chapter 2 may not be applicable to Australian markets.

Leung (2011) proposed a class reactive investment strategy which imposes no restrictions on the path-dependency of the mean level of returns. This study provides a method of choosing the targeted portfolio return at each scheduled time according to the difference between current portfolio wealth and the terminal wealth. Portfolios are rebalanced according to the chosen targeted portfolio return. However, Leung (2011) assumed a constant covariance matrix for returns. This condition is not satisfied, given our results from Chapter 3. We will relax this constraint and extend his work to a case when covariances of returns are non-constant. In particular, we will study portfolio wealth when asset returns follow a Vector-Autoregressive (VAR) Multivariate Generalized Autoregressive Conditional Heteroskedasticity (MGARCH) process. Continuing from Chapter 3, we will consider a VAR(1)-MGARCH(1,1) return process as a starting point.

Let W_t denote the portfolio wealth and $\mu(W_t)$ denote the expected portfolio return at time t. For a total of N periods, we take optimization as to minimize the variance $\sigma^2(\mu_t)$ for a given level of μ_t, that is, to locate the portfolio on the efficient frontier for the targeted return. Equivalently, we want to find the minimum $E(W_T^2)$ for a given $E(W_T)$.

The evolution of portfolio wealth is described by a recursive relationship that

$$W_{t+1} = \left(1 + \mathbf{w}_t^T \mathbf{r}_t\right) W_t + K_{t+1}, \tag{4.1}$$

where K_t is the cashflow, $\mathbf{r}_t$ is the vector of asset returns and $\mathbf{w}_t$ contains the portfolio weights at time t.

4.1 Compatibility with Path-independent Controls

We first check if the portfolio rules proposed by Leung (2011) are compatible with the criteria we formulated in Chapter 2, when they are applied to path-independent returns. Asset returns are path-independent when current returns do not depend on past experiences.

We construct a portfolio with two assets, a risky one and a riskfree one for simplicity. We require cash withdrawal and injection to be zeros. Leung (2011) defined cash withdrawals and injections as arbitrary numbers at the end of each scheduled period. Therefore, setting them equal to zero involves no loss of generality in the following proof. Also note that when there are only two assets, any portfolio choice will be on the efficient frontier.

Let s denote the price of the risky asset. Returns on the risky asset are independently normally distributed over time. Let b be the price of the riskfree asset and r_f be the riskfree rate. The processes of the two assets are given by

$$\begin{aligned} \frac{ds}{s} &= \mu dt + \sigma dB, \text{ and} \\ \frac{db}{b} &= r_f dt, \end{aligned}$$

where μ and σ are the mean and standard deviation of returns on the risky asset, and $B(t)$ is Brownian motion as in Chapter 2. Note that σ is required to be a constant over time in Leung (2011), but μ and r_f are not necessarily constant and we only require them to be path-independent.

The above can be expressed in matrix form as

$$\begin{aligned} \mathbf{r}_t &= \begin{bmatrix} \mu \\ r_f \end{bmatrix} + \begin{bmatrix} \sigma e_t \\ 0 \end{bmatrix}, \text{ and} \\ \mathbf{\Sigma} &= \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

where $\mathbf{r}_t$ is the vector of asset returns at time t, and $\mathbf{\Sigma}$ is the covariance matrix in returns.

For given portfolio wealth W_t at time t, according to Leung (2011), portfolio weights should be adjusted to

$$\mathbf{w}_t = \frac{1}{W_t}\left\{\begin{bmatrix}\mu\\ r_f\end{bmatrix}\begin{bmatrix}\mu\\ r_f\end{bmatrix}^T + \mathbf{\Sigma}\right\}^{-1}\left\{[\lambda - W_t]\begin{bmatrix}\mu\\ r_f\end{bmatrix} - \eta\begin{bmatrix}1\\ 1\end{bmatrix}\right\},$$

where λ is the Lagrangian multiplier chosen by the targeted portfolio return, and η is chosen to ensure the weights sum up to 1. Hence, investment in the risky asset G_t and the riskfree asset H_t at time t are given by

$$\begin{aligned}\begin{bmatrix}G_t\\ H_t\end{bmatrix} &= W_t\mathbf{w}_t \qquad (4.2)\\ &= \begin{bmatrix}(\mu^2 + r_f^2 - 2\mu r_f + \sigma^2)^{-1}(r_f^2 - \mu r_f)\eta W_t\\ (\mu^2 + r_f^2 - 2\mu r_f + \sigma^2)^{-1}(\mu r_f - \mu^2)\eta W_t\end{bmatrix}.\end{aligned}$$

On the other hand, according to the evolution in portfolio wealth given in (4.1), when $K = 0$, the change in portfolio value is then:

$$\begin{aligned}\Delta W_t &= W_t\mathbf{w}_t^T\ \Delta\mathbf{r}_t\\ &= \begin{bmatrix}G_t\\ H_t\end{bmatrix}^T \Delta\mathbf{r}(t).\end{aligned}$$

When the time interval shrinks to 0, the above becomes

$$\begin{aligned} dW &= \begin{bmatrix} G \\ H \end{bmatrix}^T d\mathbf{r} \\ &= G\frac{ds}{s} + H\frac{db}{b}, \end{aligned}$$

which is exactly the same as (2.6) in Chapter 2. Therefore, the evolution of portfolio wealth in (2.12) in Proposition 1 in Chapter 2 follows; that is

$$\frac{\partial W}{\partial t} + rs\frac{\partial W}{\partial s} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 W}{\partial s^2} - rW = 0. \tag{4.3}$$

We have shown that G_t and H_t are linear functions of W_t in (4.2). Hence, G_t follows the same process as that for W_t, that is

$$\frac{\partial G}{\partial t} + rs\frac{\partial G}{\partial s} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 G}{\partial s^2} - rG = 0, \tag{4.4}$$

which is consistent with (2.9) in Proposition 1 in Chapter 2 when σ is a constant.

Therefore, we have shown that the class of investment strategies proposed by Leung (2011) satisfies the requirements for being optimal, when asset returns are path-independent and the variance of the return on the risky asset is constant.

4.2 Time-varying Efficient Frontier

Recall that we choose portfolio weights on the efficient frontier of the portfolio. An efficient frontier is the combinations of assets with the minimum portfolio variance for each level of portfolio returns. Put another way, portfolio combinations on the efficient frontier are efficient in the sense that they have the minimum risk compared to other combinations giving the same level of portfolio returns.

Let us study the distribution of portfolio wealth. Note that, for a given level of expected wealth, the variance of portfolio wealth is determined by its second moment. Therefore, we are interested only in the first two moments of portfolio wealth.[1]

Suppose the probability density function of W_t is $f(W_t)$. The targeted portfolio return is μ_t, that is, we wish $\mathbb{E}\left(\mathbf{w}_t^T\mathbf{r}_t\right) = \mu_t$. Also let σ_t^2 be the variance of the portfolio return at time t. The expected next period wealth is then

$$\begin{aligned}\mathbb{E}(W_{t+1}) &= \mathbb{E}\left[\left(1+\mathbf{w}_t^T\mathbf{r}_t\right)W_t\right] + K_{t+1} \\ &= \int \left(1+\mathbf{w}_t^T\mathbf{r}_t\right)W_t f(W_t)dW_t + K_{t+1} \\ &= \mathbb{E}(W_t) + K_{t+1} + \int \mu_t W_t f(W_t)dW_t.\end{aligned} \tag{4.5}$$

The second moment of the next period wealth is

$$\mathbb{E}(W_{t+1}^2) = \mathbb{E}\left[\left(1+\mathbf{w}_t^T\mathbf{r}_t\right)W_t + K_{t+1}\right]^2$$

[1] The first two moments will not fully characterize the return process when returns are not normally distributed. We are considering a VAR-MGARCH framework here and have made the assumption of normality.

$$
\begin{aligned}
&= \mathbb{E}\left[\left(1 + 2\mathbf{w}_t^T\mathbf{r}_t + \left(\mathbf{w}_t^T\mathbf{r}_t\right)^2\right) W_t^2 + K_{t+1}^2 + 2\left(1 + \mathbf{w}_t^T\mathbf{r}_t\right) K_{t+1} W_t\right] \\
&= \mathbb{E}(W_t^2) + \mathbb{E}\left[\begin{array}{c} \left(2\mathbf{w}_t^T\mathbf{r}_t + \left(\mathbf{w}_t^T\mathbf{r}_t\right)^2\right) W_t^2 - K_{t+1}^2 \\ +2K_{t+1}\left(W_t + \mathbf{w}_t^T\mathbf{r}_t W_t + K_{t+1}\right) \end{array}\right] \\
&= \mathbb{E}(W_t^2) + 2K_{t+1}\mathbb{E}(W_{t+1}) - K_{t+1}^2 + \mathbb{E}\left[\left(2\mathbf{w}_t^T\mathbf{r}_t + \left(\mathbf{w}_t^T\mathbf{r}_t\right)^2\right) W_t^2\right].
\end{aligned}
$$

We also have $\mathbb{E}\left(\mathbf{w}_t^T\mathbf{r}_t\right)^2 = Var(\mathbf{w}_t^T\mathbf{r}_t) + \left[\mathbb{E}\left(\mathbf{w}_t^T\mathbf{r}_t\right)\right]^2 = \sigma_t^2 + \mu_t^2$. Hence, the above becomes

$$
\mathbb{E}(W_{t+1}^2) = \left[\begin{array}{c} \mathbb{E}(W_t^2) + 2K_{t+1}\mathbb{E}(W_{t+1}) - K_{t+1}^2 \\ +\int (2\mu_t + \sigma_t^2 + \mu_t^2) W_t^2 f(W_t) dW_t \end{array}\right]. \tag{4.6}
$$

Note that

$$
Var(W_{t+1}) = \mathbb{E}(W_{t+1}^2) - \left[\mathbb{E}(W_{t+1})\right]^2.
$$

Hence, the problem of portfolio control is to minimize $\mathbb{E}(W_{t+1}^2)$ for a given level of $\mathbb{E}(W_{t+1})$. Leaving out the constant components, this reduces to minimizing

$$
\int (2\mu_t + \sigma_t^2 + \mu_t^2) W_t^2 f(W_t) dW_t,
$$

subject to a given level of

$$
\int \mu_t W_t f(W_t) dW_t.
$$

The Lagrangian for the above minimization is

$$
\mathcal{L} = \int (2\mu_t + \sigma_t^2 + \mu_t^2) W_t^2 p(W_t) dW_t - \lambda_t^{(EF)} \int \mu_t W_t f(W_t) dW_t, \tag{4.7}
$$

where $\lambda_t^{(EF)}$ is the positive Lagrangian Multiplier. Note that $\lambda_t^{(EF)}$ has a time subscript which indicates that it is a parameter that may change over time.

The first order variation in $\mathcal{L}$ due to a small variation in $\Delta\mu_t = \varepsilon(W_t)$, such that $\mu_t + \varepsilon(W_t)$ does not exceed the minimum or maximum feasible values of μ_t, is:

$$\Delta\mathcal{L} = \int \varepsilon(W_t)\left[2(1+\sigma_t\frac{\partial\sigma}{\partial\mu}+\mu_t)W_t - \lambda_t^{(EF)}\right] W_t f(W_t) dW_t. \tag{4.8}$$

We require $\Delta\mathcal{L}$ to be insensitive to $\varepsilon(W_t)$ for $\mathcal{L}$ to be at a minimum, that is $\Delta\mathcal{L} = 0$. This is so when

$$f(W_t) = 0, \text{ or}$$

$$\mu_t = \mu_{\min} \text{ or } \mu_t = \mu_{\max}, \text{ or}$$

$$\begin{aligned} 2(1+\sigma_t\frac{d\sigma_t}{d\mu}+\mu_t)W_t &= \lambda_t^{(EF)}, \text{ which implies that} \\ 1+\frac{1}{2}\frac{d\sigma_t^2}{d\mu_t}+\mu_t &= \frac{\lambda_t^{(EF)}}{2W_t}. \end{aligned} \tag{4.9}$$

The Ordinary Differential Equation (ODE) in (4.9) has to be solved for σ_t on the efficient frontier that has a positive-definite quadratic form given by

$$\sigma_t^2 = \varphi_t\mu_t^2 + 2\psi_t\mu_t + \chi_t, \tag{4.10}$$

where φ_t, ψ_t and χ_t are parameters describing the nature of covariances in returns and hence they do not depend on portfolio wealth. Notice that the parameters have time subscripts and may not be constant, meaning that the efficient frontier may shift over time.

Substituting (4.10) back into (4.9), we have the targeted portfolio return at each period being

$$\begin{aligned}\mu_t &= \frac{\lambda_t^{(EF)}}{2W_t(1+\varphi_t)} - \frac{1+\psi_t}{1+\varphi_t} \\ &= \frac{\beta_t}{W_t} + \alpha_t,\end{aligned} \tag{4.11}$$

where

$$\begin{aligned}\beta_t &= \frac{\lambda_t^{(EF)}}{2(1+\varphi_t)}, \text{ and} \\ \alpha_t &= -\frac{1+\psi_t}{1+\varphi_t}\end{aligned} \tag{4.12}$$

4.3 Targeted Portfolio Return

The targeted portfolio return at each rebalancing time is determined by the current expectation of the terminal portfolio wealth. We wish to choose targeted portfolio returns for each period such that for a given level of targeted terminal wealth, the variance of terminal wealth is minimized. We first substitute the functional form of μ_t in (4.11), which has the return characteristics embedded into it, back into the first two moments of next period wealth W_{t+1} in (4.5) and (4.6). We obtain

$$\mathbb{E}(W_{t+1}) = u_t\mathbb{E}(W_t) + b_t, \tag{4.13}$$

and

$$\mathbb{E}(W_{t+1}^2) = v_t \mathbb{E}(W_t^2) + p_t \mathbb{E}(W_t) + q_t, \tag{4.14}$$

where the constants at time t are:

$$\begin{aligned}
u_t &= 1 + \alpha_t, \\
b_t &= \beta_t + K_{t+1}, \\
v_t &= 1 + \chi_t - \frac{(1+\psi_t)^2}{1+\varphi_t}, \\
p_t &= 2(1+\alpha_t)K_{t+1}, \text{ and} \\
q_t &= (\beta_t + K_{t+1})^2 + \varphi_t \beta_t^2.
\end{aligned}$$

See Appendix B.1.1 for the derivation of (4.13) and (4.14).

Let us consider a situation in which the investor has a horizon of N periods. Then the first two moments of terminal wealth W_N evaluated at time t can then be written as:

$$\mathbb{E}_t(W_N) = \sum_{i=0}^{N-t-1} \left(b_{t+i} \prod_{s=i+1}^{N-t-1} u_{t+s} \right) + \mathbb{E}_t(W_t) \prod_{i=0}^{N-t-1} u_{t+i}, \tag{4.15}$$

and

$$\mathbb{E}_t(W_N^2) = \left\{ \begin{array}{c} \sum_{i=0}^{N-t-1} \left[(p_{t+i}\mathbb{E}_t(W_{t+i}) + q_{t+i}) \prod_{s=i+1}^{N-t-1} v_{t+s} \right] \\ +\mathbb{E}_t(W_t^2) \prod_{i=0}^{N-t-1} v_{t+i} \end{array} \right\}. \tag{4.16}$$

The Lagrangian for minimizing $\mathbb{E}_t(W_N^2)$ for a given level of $\mathbb{E}_t(W_N)$ at time t is

$$\mathcal{L}_t = \mathbb{E}_t(W_N^2) - 2\lambda_t^{(N)}\mathbb{E}_t(W_N) \tag{4.17}$$
$$= \left\{ \begin{array}{c} \sum_{i=0}^{N-t-1}\left[(p_{t+i}\mathbb{E}_t(W_{t+i}) + q_{t+i})\prod_{s=i+1}^{N-t-1} v_{t+i}\right] + \mathbb{E}_t(W_t^2)\prod_{i=0}^{N-t-1} v_{t+i} \\ -2\lambda_t^{(N)}\left[\sum_{i=0}^{N-t-1}\left(b_{t+i}\prod_{s=i+1}^{N-t-1} u_{s+i}\right) + \mathbb{E}_t(W_t)\prod_{i=0}^{N-t-1} u_{t+i}\right] \end{array} \right\}.$$

where $\lambda_t^{(N)}$ is the Lagrangian Multiplier.

However, at time t, the best estimate of future returns, and hence φ, ψ, χ, u and v in (4.17) may be the current values. Therefore, we can define the approximated (or estimated) first two moments of terminal wealth W_N evaluated at time t, as

$$\hat{\mathbb{E}}_t(W_N) = \sum_{i=0}^{N-t-1} b_{i+t}u_t^{N-t-1-i} + \mathbb{E}(W_t)u_t^{N-t}, \text{ and} \tag{4.18}$$

$$\hat{\mathbb{E}}_t(W_N^2) = \left\{ \begin{array}{c} \sum_{i=0}^{N-t-1}\left[\left(p_{t+i}\hat{\mathbb{E}}_t(W_{t+i}) + q_{t+i}\right) v_t^{N-t-1-i}\right] \\ +\mathbb{E}_t(W_t^2)v_t^{N-t} \end{array} \right\}. \tag{4.19}$$

Hence, the approximated (or estimated) Lagrangian for minimizing $\hat{\mathbb{E}}_t(W_N^2)$ for a given $\hat{\mathbb{E}}_t(W_N)$ is now

$$\hat{\mathcal{L}}_t = \left\{ \begin{array}{c} \sum_{i=0}^{N-t-1}\left[\left(p_{t+i}\hat{\mathbb{E}}_t(W_{t+i}) + q_{t+i}\right) v_t^{N-t-1-i}\right] + \mathbb{E}_t(W_t^2)v_t^{N-t} \\ -2\lambda_t^{(N)}\left[\sum_{i=0}^{N-t-1} b_{t+i}u_t^{N-t-1-i} + \mathbb{E}_t(W_t)u_t^{N-t}\right] \end{array} \right\}. \tag{4.20}$$

Again, consider a small variation in μ_t, which comes from β_t, $\Delta\beta_t$. We have

$$
\begin{aligned}
\Delta b_t &= \Delta\beta_t, \\
\Delta q_t &= 2\left[(1+\varphi_t)\beta_t + K_{t+1}\right]\Delta\beta_t, \text{ and} \\
\Delta\hat{\mathbb{E}}_t(W_{t+i}) &= \sum_{s=0}^{i-1} \Delta\beta_{t+s} u_t^{i-1-s}.
\end{aligned}
$$

The first order variation in the approximated Lagrangian is then

$$
\Delta\hat{\mathcal{L}}_t = \left\{
\begin{array}{c}
\sum_{i=0}^{N-t-2} \Delta\beta_{t+i} \sum_{s=i+1}^{N-t-1} p_{t+s} u_t^{s-1-i} v_t^{N-t-1-s} \\
+2 \sum_{i=0}^{N-t-1} v_t^{N-t-1-i} \left[(1+\varphi_t)\beta_{t+i} + K_{t+i+1}\right] \Delta\beta_{t+i} \\
-2\lambda_t^{(N)} \sum_{i=0}^{N-t-1} \Delta\beta_{t+i} u_t^{N-t-1-i}
\end{array}
\right\}. \tag{4.21}
$$

For the Lagrangian to be at a minimum, we require it to be insensitive to $\Delta\beta_{t+i}$. Hence the coefficients of $\Delta\beta_{t+i}$ must sum up to 0, that is

$$
\left\{
\begin{array}{c}
\sum_{s=i+1}^{N-t-1} p_{t+s} u_t^{s-1-i} v_t^{N-t-1-s} \\
+2v_t^{N-t-1-i} \left[(1+\varphi_t)\beta_{t+i} + K_{t+i+1}\right] - 2\lambda_t^{(N)} u_t^{N-t-1-i}
\end{array}
\right\} = 0,
$$

which gives

$$
(1+\varphi_t)\beta_{t+i} = \lambda_t^{(N)} \left(\frac{u_t}{v_t}\right)^{N-t-1-i} - \sum_{s=i}^{N-t-1} K_{t+s+1} \left(\frac{u_t}{v_t}\right)^{s-i}, \text{ for } t < T-1. \tag{4.22}
$$

The above functional form of β_{t+i} $(\lambda_t^{(N)})$ should give the targeted value of $\hat{\mathbb{E}}_t(W_N)$ evaluated at time t. Substituting β_{t+i} $(\lambda_t^{(N)})$ into $\hat{\mathbb{E}}_t(W_N)$ in (4.18), we have

$$\hat{\mathbb{E}}_t(W_N) = \left\{ \begin{array}{c} \frac{\lambda_t^{(N)}}{1+\varphi_t}\left[\left(\frac{u_t^2}{v_t}\right)^{N-t} - 1\right](\frac{u_t^2}{v_t} - 1)^{-1} + \mathbb{E}_t(W_t)u_t^{N-t} \\ +u_t^{N-t}\left(1 + \frac{1}{1+\varphi_t}\frac{v_t}{u_t^2 - v_t}\right)\sum_{i=1}^{N-t} K_{t+i}u_t^{-i} - \frac{u_t^{N-t}}{1+\varphi_t}\frac{v_t}{u_t^2 - v_t}\sum_{i=1}^{N-t}\left(\frac{u_t}{v_t}\right)^i K_{t+i} \end{array} \right\}.$$

$\lambda_t^{(N)}$ can be retrieved from $\hat{\mathbb{E}}_t(W_N)$ as

$$\lambda_t^{(N)} = \frac{(1+\varphi_t)(\frac{u_t^2}{v_t} - 1)}{\left(\frac{u_t^2}{v_t}\right)^{N-t} - 1}\left[\begin{array}{c} \hat{\mathbb{E}}_t(W_N) - \mathbb{E}(W_t)u_t^{N-t} \\ -\frac{u_t^{T-t}}{1+\varphi_t}\frac{v_t}{u_t^2 - v_t}\sum_{i=1}^{N-t}\left(\frac{u_t}{v_t}\right)^i K_{t+i} \\ +u_t^{N-t}\left(1 + \frac{1}{1+\varphi_t}\frac{v_{ti}}{u_t^2 - v_t}\right)\sum_{i=1}^{N-t} K_{t+i}u_t^{-i} \end{array} \right]. \quad (4.23)$$

β_t is retrieved by substituting $\lambda_t^{(N)}$ into (4.22).

See Appendix B.1.2 for the detailed derivation of the Lagrangian Multiplier $\lambda_t^{(N)}$ in for minimizing the variance in the (approximated) terminal wealth subject to a given level of this (approximated) terminal wealth.

4.4 Portfolio Weights in a VAR-MGARCH Process

The results of Section 4.2 are now ready to be incorporated into the return process. We will adopt the same return model as in Chapter 3. Suppose that asset returns follow a VAR(1)-MGARCH(1,1) process, where the MGARCH(1,1) effects are governed by a BEKK model (Engle and Kroner 1995). In summary,

$$\begin{aligned}
\mathbf{r}_t &= \boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1} + \mathbf{v}_t, \\
\mathbf{v}_t &= \boldsymbol{\Sigma}_t^{1/2}\mathbf{e}_t, \\
\boldsymbol{\Sigma}_t &= \mathbf{A}_0^T\mathbf{A}_0 + \mathbf{A}_1^T\mathbf{v}_{t-1}\mathbf{v}_{t-1}^T\mathbf{A}_1 + \mathbf{B}_1^T\boldsymbol{\Sigma}_{t-1}\mathbf{B}_1,
\end{aligned} \tag{4.24}$$

where $\mathbf{r}_t$ is a vector of asset returns at time t, $\mathbf{e}_t$ are independently identically distributed (i.i.d.) standard normal innovations and Σ_t is the covariance matrix of returns.

Also suppose that the joint distribution of (r_{t-1}, W_t) is $f(r_{t-1}, W_t)$. Recall that the accumulation of portfolio wealth is specified in (4.1). The first two moments of next period portfolio wealth W_{t+1} are now

$$\mathbb{E}(W_{t+1}) = \mathbb{E}(W_t) + \int\int \mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right)W_t f(\mathbf{r}_{t-1}, W_t)dW_t d\mathbf{r}_{t-1} + K_{t+1}, \tag{4.25}$$

and

$$\mathbb{E}(W_{t+1}^2) = \left\{ \begin{array}{c} \int\int \left[\begin{array}{c} 2\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) \\ +\left[\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right)\right]^2 \\ +\mathbf{w}_t^T\boldsymbol{\Sigma}_t\mathbf{w}_t \end{array} \right] W_t^2 f(\mathbf{r}_{t-1}, W_t)dW_t d\mathbf{r}_{t-1} \\ \mathbb{E}(W_t^2) + 2K_{t+1}\mathbb{E}(W_{t+1}) - K_{t+1}^2 \end{array} \right\}. \tag{4.26}$$

Detailed derivations for (4.25) and (4.26) are outlined in Appendix B.2.

Let us again use the fact that the second moment of portfolio returns equals the sum of the variance and the squared expectation of portfolio returns. The objective of minimizing the variance in next period wealth is equivalent to minimizing $\mathbb{E}(W_{t+1}^2)$ for a given level of $\mathbb{E}(W_{t+1})$. We also require that the portfolio weights sum up to 1, that is $\mathbf{w}_t^T\mathbf{j} = 1$ where $\mathbf{j} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$. The optimization problem can be solved by constructing the Lagrangian function:

$$\begin{aligned}
\mathcal{L} = &\int\int \begin{bmatrix} 2\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) + \mathbf{w}_t^T\boldsymbol{\Sigma}_t\mathbf{w}_t \\ + \left[\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right)\right]^2 \end{bmatrix} W_t^2 f(\mathbf{r}_{t-1}, W_t) dW_t d\mathbf{r}_{t-1} \quad (4.27)\\
&- 2\lambda_t \int\int \mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) W_t f(\mathbf{r}_{t-1}, W_t) dW_t d\mathbf{r}_{t-1} \\
&+ 2\eta_t \int\int \mathbf{w}_t^T\mathbf{j} W_t f(\mathbf{r}_{t-1}, W_t) dW_t d\mathbf{r}_{t-1}.
\end{aligned}$$

where λ_t and η_t are the positive Lagrangian Multipliers.

Let us consider a small variation, $\boldsymbol{\epsilon}(\mathbf{r}_{t-1}, W_t)$, in $\mathbf{w}_t$ such that $\mathbf{j}^T\boldsymbol{\epsilon} = 0$ and $\mu_t + \triangle\mu_t = (\mathbf{w}_t^T + \triangle\mathbf{w}_t^T)r_t \in [\mu_{\min}, \mu_{\max}]$. The resulting variation in $\mathcal{L}$ to the first order is:

$$\begin{aligned}
\triangle\mathcal{L} = &\int\int 2 \begin{bmatrix} \boldsymbol{\epsilon}^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) + \boldsymbol{\epsilon}^T\boldsymbol{\Sigma}_t\mathbf{w}_t \\ +\boldsymbol{\epsilon}^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right)\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) \end{bmatrix} W_t^2 f dW_t d\mathbf{r}_{t-1} \quad (4.28)\\
&- 2\lambda_t \int\int \boldsymbol{\epsilon}^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) W_t f dW_t d\mathbf{r}_{t-1} \\
&+ 2\int\int \boldsymbol{\epsilon}^T\mathbf{j}\eta(\mathbf{r}_{t-1}, W_t) W_t p dW_t d\mathbf{r}_{t-1} + K_{t+1} \\
&= 2\int\int \boldsymbol{\epsilon}^T \left\{ \begin{array}{c} \begin{bmatrix} \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) + \boldsymbol{\Sigma}_t\mathbf{w}_t \\ + \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right)\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) \end{bmatrix} W_t \\ -\lambda_t\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) + \eta_t\mathbf{j} \end{array} \right\} W_t f dW_t d\mathbf{r}_{t-1}
\end{aligned}$$

Again, at a minimum, we require $\triangle\mathcal{L}$ to be insensitive to $\boldsymbol{\epsilon}(\mathbf{r}_{t-1}, W_t)$.

This is so when

$$f(\mathbf{r}_{t-1}, W_t) = 0 \text{ or;}$$

$$\mu_t \text{ is at a minimum or maximum,}$$

or

$$\left[\begin{array}{c} \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) + \boldsymbol{\Sigma}_t \mathbf{w}_t \\ + \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) \mathbf{w}_t^T \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) \end{array} \right] W_t - \lambda_t \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) + \eta_t \mathbf{j} = 0.$$

The last relation gives the portfolio weight at time t as:

$$\mathbf{w}_t = \frac{1}{W_t} \mathbf{G}^{-1} \left[(\lambda_t - W_t) \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) - \eta_t \mathbf{j} \right] \tag{4.29}$$

where

$$\mathbf{G} = \left[\begin{array}{c} \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right)^T \\ + \mathbf{A}_0^T \mathbf{A}_0 + \mathbf{B}_1^T \boldsymbol{\Sigma}_{t-1} \mathbf{B}_1 \\ + \mathbf{A}_1^T \mathbf{v}_{t-1} \mathbf{v}_{t-1}^T \mathbf{A}_1^T \end{array} \right]. \tag{4.30}$$

The Lagrangian Multiplier η_t is set to ensure that the weights to sum up to 1. Thus

$$\begin{aligned} \mathbf{w}_t &= \frac{1}{W_t} \mathbf{G}^{-1} (\lambda_t - W_t) \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right) - \frac{\eta}{W_t} \mathbf{G}^{-1} \mathbf{j} \\ {\mathbf{w}_t}^T \mathbf{j} &= \frac{(\lambda_t - W_t)}{W_t} \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right)^T \left(\mathbf{G}^{-1}\right)^T \mathbf{j} - \frac{\eta}{W_t} \mathbf{j}^T (\mathbf{G}^{-1})^T \mathbf{j} \\ 1 &= \frac{(\lambda_t - W_t)}{W_t} \left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right)^T \left(\mathbf{G}^{-1}\right)^T \mathbf{j} - \frac{\eta}{W_t} \mathbf{j}^T (\mathbf{G}^{-1})^T \mathbf{j}, \end{aligned}$$

and hence

$$\eta_t = \left[(\lambda_t - W_t)\left(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1}\right)^T \left(\mathbf{G}^{-1}\right)^T - W_t\right] \mathbf{j} \left[\mathbf{j}^T (\mathbf{G}^{-1})^T \mathbf{j}\right]^{-1}. \qquad (4.31)$$

If we compare the Lagrangian in (4.27) with the Lagrangian in (4.7) that we set up in Section 4.2, the Lagrangian Multiplier λ_t in (4.27) should be half of the Lagrangian Multiplier $\lambda_t^{(EF)}$ in (4.7), that is

$$\lambda_t = \frac{1}{2}\lambda_t^{(EF)} \qquad (4.32)$$

where $\lambda_t^{(EF)}$ is given by β_t in (4.12).

4.5 The Algorithm to Determine Optimal Portfolio Weights

We have developed a class of optimal dynamic investment strategies, which seek portfolio weights on the efficient frontier at a scheduled rebalancing time. The strategy is summarized below:

1. Assess asset return characteristics φ_t, ψ_t and χ_t, and construct the efficient frontier specified in (4.10).

2. Determine the desired terminal portfolio wealth $\mathbb{E}(W_N)$ and let $\hat{\mathbb{E}}_t(W_N)$ equal this targeted value. Substitute the return characteristics and $\hat{\mathbb{E}}_t(W_N)$ into (4.23) to obtain $\lambda_t^{(N)}$.

3. Substitute $\lambda_t^{(N)}$ into (4.22) to obtain β_t and the desired portfolio return

μ_t.

4. Determine $\lambda_t^{(EF)}$ by substituting β_t into (4.12), and hence determine λ_t in (4.32).

5. Substitute λ_t into (4.29) to obtain the required portfolio weights.

One common constraint in portfolio management is that shortselling is forbidden in many markets, for example, in the Australian market. Let us refer to a strategy as being constrained when the portfolio weights have to lie in the interval $[0, 1]$. The solution to these constrained dynamic portfolio weights $\mathbf{w}_t$ has to be solved numerically. There are two ways of doing this.

The first approach is to numerically minimize the Lagrangian function in (4.27) using a grid search that covers a wide set of combinations of portfolio weights between 0 and 1.

Alternatively, a simpler approach is to apply quadratic programming to minimize $\mathbf{w}_t^T \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{w}_t$ when the forecast portfolio return $\mathbf{w}_t^T(\boldsymbol{\gamma}+\mathbf{C}\mathbf{r}_{t-1})$ equals a chosen level indicated in (4.11). We replace μ_t with the maximum in $(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1})$ when the indicated level μ_t is greater than the maximum in $(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1})$, and similarly, we replace μ_t with the minimum in $(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1})$ if μ_t is less than the minimum in $(\boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1})$.

4.6 Performance of the Dynamic Strategy

Let us study the performance of the proposed strategy by simulating ten thousand sets of twenty-year returns on three asset classes: equity, debt and cash. The return series will be generated under the VAR-MGARCH framework specified in (4.24). Parameters are obtained by estimating (4.24) on 20 years of annual returns on the S&P/ASX 200 index[2], Australian 90-day bank accepted bills (BABs)[3] and Australian one-year bank's term deposit index[4] from 1990 to 2012. This is done by following the one-step multivariate semiparametric maximum likelihood estimation (one-step MSMLE) we developed in Chapter 3.

The estimated VAR(1)-MGARCH(1,1) parameters for the entire sample are

$$\hat{\gamma}= \begin{bmatrix} 0.0724 \\ 0.0044 \\ 0.0081 \end{bmatrix}, \hat{\mathbf{C}} = \begin{bmatrix} -0.3634 & 0.3535 & 0 \\ -0.0333 & 0.2173 & 0 \\ 0 & 0 & 0.6591 \end{bmatrix},$$

$$\hat{\mathbf{A}}_0 = \begin{bmatrix} 0.0243 & 0.0036 & 0.0028 \\ 0 & 1.9795 \times 10^{-7} & 5.2368 \times 10^{-8} \\ 0 & 0 & 1.6966 \times 10^{-11} \end{bmatrix},$$

[2] Annual returns are aggregated monthly returns. Monthly indexed data is acquired from publications of the Australian Bureau of Statistics (1350.0 Australian Economic Indicators, Chapter 8 Table 8.7 Australian stock market indexes, MONTHLY) and the Reserve Bank of Australia (Statistical Table F7 Australian Share Market) Data for the period dating from December 1989 to February 2009 is available at `http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1350.0Apr%202009?OpenDocument`. Data for the period dating from April 2009 to April 2010 is available at `http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1350.0Jun%202010?OpenDocument`. Data for the period dating from May 2010 to May 2012 is available at `http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1350.0Jul%202012?OpenDocument`. Data for the period dating from June 2012 to August 2013 is available at `http://www.rba.gov.au/statistics/tables/index.html#share-mkts`.

[3] Data is acquired from Statistical Table F1.1 Interest Rates and Yields - Money Market - Monthly by the RBA at `http://www.rba.gov.au/statistics/tables/index.html#interest-rates`.

[4] Data is acquired from Statistical Table F4 Retail Deposit and Investment Rates by the RBA at `http://www.rba.gov.au/statistics/tables/index.html#interest-rates`.

$$\hat{\mathbf{A}}_1 = \begin{bmatrix} -4.2493 \times 10^{-4} & -8.6205 \times 10^{-5} & -5.5102 \times 10^{-5} \\ 0 & -1.9429 \times 10^{-4} & -5.1382 \times 10^{-5} \\ 0 & 0 & -3.9935 \times 10^{-7} \end{bmatrix}, \text{ and}$$

$$\hat{\mathbf{B}}_1 = \begin{bmatrix} -0.0653 & -0.0012 & 6.6802 \times 10^{-4} \\ 0 & -0.0568 & 0.0056 \\ 0 & 0 & -0.0783 \end{bmatrix}.$$

Hence, the estimated unconditional mean return is

$$\hat{\boldsymbol{\mu}} = (\mathbf{I} - \hat{\mathbf{C}})^{-1}\hat{\boldsymbol{\gamma}} = \begin{bmatrix} 0.0540 \\ 0.0033 \\ 0.0239 \end{bmatrix}, \tag{4.33a}$$

and the estimated unconditional covariance on returns is

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 5.9378 \times 10^{-4} & 8.6843 \times 10^{-5} & 6.8107 \times 10^{-5} \\ 8.6843 \times 10^{-5} & 1.2701 \times 10^{-5} & 9.9610 \times 10^{-6} \\ 6.8107 \times 10^{-5} & 9.9610 \times 10^{-6} & 7.8119 \times 10^{-6} \end{bmatrix}. \tag{4.34}$$

We set each parameter value in the VAR(1)-MGARCH(1,1) system in (4.24) equal to its estimated value in the work that follow. We also assume that the innovations $\mathbf{e}_t$ in (4.24) follow a standard normal distribution.

Suppose we have an initial wealth of 100 and we are going to invest in three asset classes: equity, debt and cash. The total investment period is twenty years, and we rebalance our portfolio annually. We also assume a constant cash inflow of $K_t = 1$ at the end of each period. We set the targeted annual portfolio return to be $\mu = 0.05$, and hence the targeted terminal wealth is $\mathbb{E}(W_{20}) = 298.3957$ calculated by the recursive relationship in (4.1).

We compare the distribution of portfolio wealth at the end of each rebalancing period, for each of four different investment strategies:

1. an unconstrained dynamic strategy that rebalances portfolio weights to the level indicated by (4.29);

2. a constrained dynamic strategy that rebalances portfolio weights to the constrained numerical minimizer of portfolio variance for the desired portfolio return given in (4.11);

3. an unconstrained static strategy that rebalances portfolio weights to $\mathbf{w}_u = [\ -0.0801 \quad -1.3875 \quad 2.4676\ ]^T$ on equity, debt and cash, found on the longrun efficient frontier[5]; and

4. a constrained static strategy that rebalances portfolio weights to $\mathbf{w}_c = [\ 0.8680 \quad 0 \quad 0.1320\ ]^T$ on equity, debt and cash, found on the constrained longrun efficient frontier[6].

Strategies are constrained in the sense that shortselling is forbidden, that is, portfolio weights lie in the interval $[0, 1]$.

[5] The longrun efficient frontier is constructed with the unconditional mean returns given in (4.33a) and the unconditional covariance in returns given in (4.34).

[6] The longrun efficient frontier is constrained in the sense that it contains portfolios with weights that lie in the interval [0,1].

4.6.1 Simulated Results

Figure 4.1 portrays the distributions of portfolio wealth at the end of the first, fifth, tenth, fifteenth and twentieth years resulting from the four different strategies. Table 4.1 summarizes the mean and standard deviation of portfolio wealth at the end of each year due to the four different strategies.

The average portfolio wealth under the dynamic strategies (both unconstrained and constrained) approach the targeted terminal wealth. The means of terminal portfolio wealth, W_{20}, under the static strategies (both unconstrained and constrained) do not align with the targeted value. Compared to static strategies, the dynamic strategies result in more concentrated distributions of terminal portfolio wealth around the targeted value. In this particular exercise, average W_{20} under an unconstrained dynamic strategy is \$300.73, under a constrained dynamic strategy is \$300.41, under an unconstrained static strategy is \$272.73 and under a constrained static strategy is \$273.24, while the targeted terminal value is \$298.40. The standard deviation in W_{20} under an unconstrained dynamic strategy is 1.15, under a constrained dynamic strategy is 5.70, under an unconstrained static strategy is 24.77 and under a constrained static strategy is 24.85. This pattern is consistent with our goal of minimizing the variance for given level of portfolio wealths.

4.6.2 Constrained Strategies v. Unconstrained Strategies

It is worth pointing out that portfolio wealth under an unconstrained dynamic strategy accumulates fast during early periods and it approaches the targeted terminal value at a decreasing rate, while the portfolio wealth under a constrained dynamic strategy approaches the targeted terminal value at a relatively constant rate. Recall that the desired portfolio return μ_t at time t is defined in (4.11) by the portfolio wealth W_t and the parameter β_t, which is determined by the Lagrangian Multiplier $\lambda_t^{(N)}$ and the target terminal wealth $\mathbb{E}(W_{20})$. β_t are large and W_t are small for early periods, leading to high levels of desired portfolio returns. The resulting portfolio weights for early periods are either much greater than 1 or very negative. An example of portfolio trajectory (β_t, W_t and the corresponding μ_t) is given in Table 4.2. The targeted portfolio returns are 161.23% for the first period, 62.1% for the second period and 21.28% for the third period in this example. The unconstrained dynamic strategy in this case is unrealistic due to huge leverage effects, that is, a great amount needs to be borrowed to invest in equities in order to achieve the desired level of expected portfolio return. Therefore, in practice, one might introduce a regulatory limit on portfolio weights when implementing unconstrained strategies.

4.6.3 Backtesting on Historical Returns

Let us compare the performances of the proposed dynamic strategy with a static strategy on historical returns on equity, debt and cash. Recall that, in Chapter 3, we estimated a VAR(1)-MGARCH(1,1) model for a set of 284-month (from December 1989 to August 2013) returns on the S&P/ASX 200 index, Australian 90-day bank accepted bills and Australia one-year bank term deposit index using a "rolling window" framework. Let us assume that we use the first 120 effective months of data to obtain an initial estimate of the VAR(1)-MGARCH(1,1) model, and we start our investment from January 2000 with an initial input of \$100. Let us also assume that we rebalance the portfolio at the end of each month, and there is a monthly cash injection of \$1. We wish to gain a portfolio return of 0.005 per month, which leads to a targeted terminal wealth of $W_{163} = \$476.38$ by the end of August 2013. We employ our estimation results obtained in Chapter 3 to determine the dynamic portfolio weights proposed in this chapter given by (4.29). The unconstrained static weights of $\mathbf{w}_u = [-0.0532, -2.7166, 3.7698]^T$ and the constrained static weights of $\mathbf{w}_c = [0.8964, 1.4715 \times 10^{-4}, 0.1034]^T$ on equity, debt and cash are found on the unconstrained and the constrained longrun efficient frontier based on the first 120 effective months of returns.

Figure 4.2 plots the evolution of portfolio wealth due to the unconstrained dynamic strategy and the unconstrained static strategy specified above. After 163 months, the unconstrained dynamic strategy leads to a terminal wealth of \$468.76, the constrained dynamic strategy leads to a terminal wealth of \$457.40, the unconstrained static strategy leads to a terminal wealth of \$551.00, and the constrained static strategy leads to a terminal wealth of 557.4536, while the targeted value is \$476.38. These results show that our proposed dynamic strategies achieve terminal values that are closer to the target than the static strategy.

We notice that the trajectory of portfolio wealth resulting from the unconstrained dynamic strategy drops significantly in late 2008. This is due to the Global Financial Crisis bursted in that period. The one-period-ahead forecast according to the previous ten-year experience was not reliable. Hence, the portfolio weights calculated based on this forecast led to a completely different outcome than what we expected. Another informative feature here is the significant differences between the terminal wealth due to the static strategies and the targeted value. We have two explanations. First, static strategies are not risk-minimizing, and they are likely to lead to extreme values. Second, there are changes in the parameters of the return model over time. The static weights are determined based on the first 120 effective months of returns. The parameters in the VAR(1)-MGARCH(1,1) model change over time, and the static allocations given by the initial estimate will not be on the longrun efficient frontiers for later periods.

4.7 Conclusion

This chapter provides a detail derivation of an extension of a type of dynamic portfolio rules, introduced by Leung (2011), to non-constant efficient frontiers of asset returns. We consider the case when returns follow a VAR(1)-MGARCH(1,1) process in particular. The proposed investment strategy determines the desired portfolio returns by assessing the difference between current portfolio wealth and the targeted terminal wealth. It then seeks portfolio weights on the efficient frontier according to the desired portfolio return level at each scheduled rebalancing time.

We quantitatively demonstrate the performance of the proposed dynamic strategy. Our results show that dynamic strategies are more efficient than static strategies, in the sense that the variance of terminal portfolio wealth is reduced.

This analysis may be generalized in several directions. First, when the trading amount is significant, the resulting market impact on asset returns should be included in the formulation of dynamic strategies. This is pursued in the next chapter. Second, other return models may be considered. Third, instead of rebalancing annually at a predetermined time, the rebalance time can be defined as a function of portfolio wealth and returns.

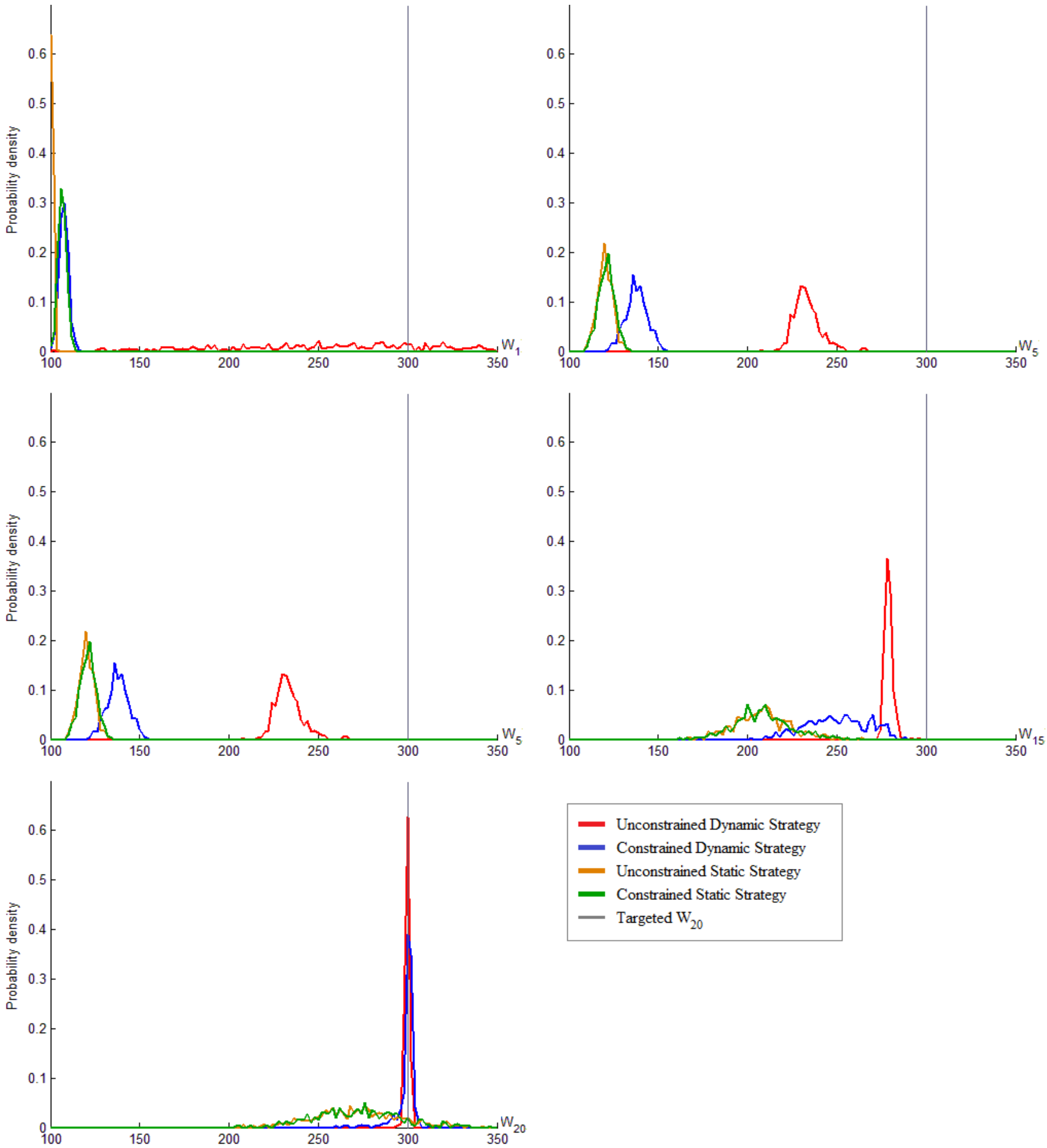

Figure 4.1: Distributions of the portfolio wealth, W_t, at the end of the first, fifth, tenth, fifteenth and twentieth year due to an unconstrained dynamic strategy, a constrained dynamic strategy, an unconstrained static strategy and a constrained static strategy

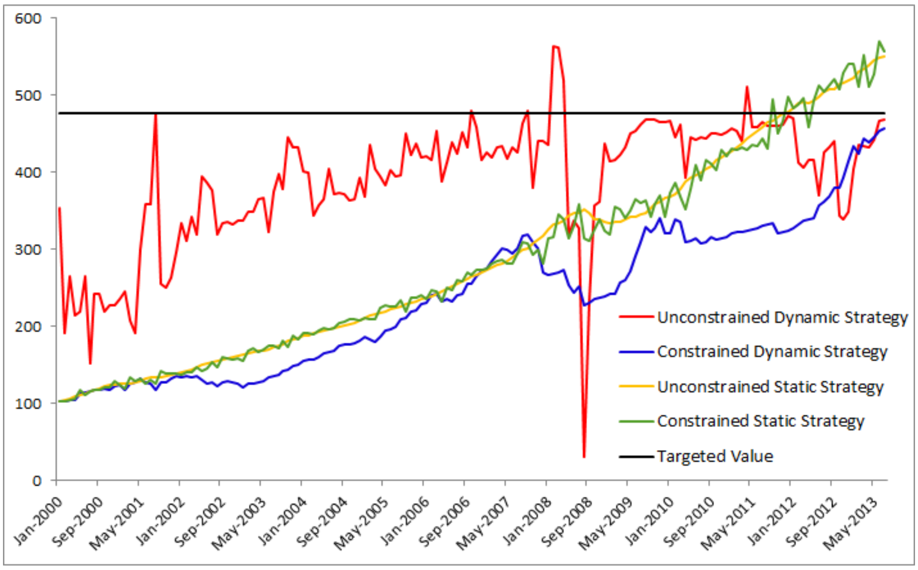

Figure 4.2: Portfolio wealth due to an unconstrained dynamic strategy, a constrained dynamic strategy, an unconstrained static strategy and a constrained static strategy for investments in Australian equity, debt and cash from January 2000 to August 2013

	Average W_t			
t	Unconstrained dynamic	Unconstrained static	Constrained dynamic	Constrained static
1	276.80	101.78	108.56	107.67
2	241.85	105.36	114.67	107.13
3	235.82	109.78	122.61	112.15
4	233.35	115.13	130.57	116.51
5	233.80	121.16	138.77	121.99
6	238.07	127.64	147.77	128.59
7	242.68	134.76	156.97	135.20
8	246.96	142.37	166.37	142.46
9	252.95	150.38	177.06	151.04
10	257.66	158.96	188.00	159.24
11	262.41	168.06	199.42	168.13
12	266.76	177.69	211.24	177.55
13	270.68	187.67	224.03	187.92
14	285.11	198.10	237.03	198.08
15	279.67	208.95	251.13	209.57
16	284.32	220.55	263.98	220.39
17	288.97	232.67	276.38	233.26
18	293.27	245.41	286.29	245.35
19	297.67	258.67	294.54	259.04
20	300.73	272.73	330.41	273.24

	Standard deviations of W_t			
t	Unconstrained dynamic	Unconstrained static	Constrained dynamic	Constrained static
1	75.77	0.67	2.62	2.31
2	25.36	1.43	3.32	2.44
3	11.98	2.27	4.42	3.20
4	12.66	3.25	4.96	3.50
5	12.90	4.22	6.00	4.41
6	6.80	5.21	6.82	5.30
7	5.76	6.18	7.92	6.36
8	11.30	7.19	8.89	7.39
9	4.81	8.33	9.97	8.39
10	14.94	9.49	11.04	9.33
11	4.01	10.61	12.48	10.76
12	3.35	11.81	13.97	12.21
13	6.60	13.07	15.17	13.04
14	6.62	14.39	16.29	14.10
15	6.28	15.84	17.34	15.85
16	4.38	17.45	16.92	17.49
17	1.48	19.20	14.60	19.37
18	1.79	21.01	12.02	21.19
19	0.81	22.93	8.82	22.56
20	1.15	24.77	5.70	24.85

Table 4.1: Statistics of W_t under different strategies

t	β_t	W_t	$\mu_t = \beta_t/W_t + \alpha_t$
0	251.70	100	1.6123
1	163.69	127.13	0.6210
2	162.41	181.20	0.2128
3	213.55	240.74	0.0191
4	153.46	247.11	-0.0241
5	171.79	234.01	-0.0078
6	163.36	230.87	0.0185
7	183.54	236.15	0.0248
8	173.82	242.48	0.0328
9	193.19	254.69	0.0069
10	183.40	256.90	0.0135
11	193.29	261.22	0.0120
12	173.49	264.82	0.0160
13	209.41	269.91	0.0129
14	191.39	273.58	0.0188
15	232.62	280.61	0.0118
16	214.92	285.37	0.0110
17	205.79	287.57	0.0177
18	223.55	294.35	0.0104
19	246.12	296.61	0.0107
20	N.A.	300.71	N.A.

Table 4.2: Example of desired portfolio returns

Chapter 5

Market Impacts and Asset Allocation

One of the most important practical issues in portfolio rebalancing is that every transaction causes market impact. Market impact is shift in the price of an asset when a market participant buys or sells the underlying asset. It is significant when trade size is large. Asset prices move upward after a significant purchase and downward after a significant sale in general. Therefore, the dynamic strategy that we proposed in Chapter 4 has to be adjusted if trade size is significant.

Market impact is positively related to the size of orders and the resulting supply-demand imbalance in assets. Therefore, the effects of trading transactions on asset returns can be modelled by order imbalance, which is the difference between the number of buyer-initiated transactions and the number of seller-initiated transactions (Kissell and Glantz 2003). The effects on asset returns due to excess buy and sell orders are different. Imbalances cause price pressures, and price pressures resulting from large sell orders are greater than those from buy orders (Chan and Fong 2000; Chordia et al. 2002). In addition, order imbalances post temporary impacts on asset prices due to liquidity requirements and permanent impacts due to information leakage (Kissell and Glantz 2003). These effects influence the aggregate market.

This chapter provides a highly simplified quantitative example to demonstrate how the dynamic strategy we proposed in Chapter 4 can be adjusted to incorporate market impacts. We only consider the permanent effects of order imbalances on asset prices in the following analysis. We also assume that order imbalances only shift the mean returns, and any impacts on the covariance level are negligible.

5.1 Order Imbalances and Returns

Let us incorporate market impacts captured by order imbalances into the Vector-Autoregressive Multivariate Generalized Autoregressive Conditional Heteroskedasticity (VAR-MGARCH) model considered in Chapter 4. Let

$$\mathbf{r}_{t^-} = \boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{v}_t; \tag{5.1}$$

$$\mathbf{MI}_t = \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-; \tag{5.2}$$

$$\begin{aligned}\mathbf{r}_t &= \mathbf{r}_{t^-} + \mathbf{MI}_t \\ &= \boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^- + \mathbf{v}_t;\end{aligned} \tag{5.3}$$

$$\mathbf{v}_t = \boldsymbol{\Sigma}_t^{1/2}\mathbf{e}_t; \tag{5.4}$$

$$\boldsymbol{\Sigma}_t=\mathbf{A}_0^T\mathbf{A}_0+\mathbf{A}_1^T\mathbf{v}_{t-1}\mathbf{v}_{t-1}^T\mathbf{A}_1+\mathbf{B}_1^T\boldsymbol{\Sigma}_{t-1}\mathbf{B}_1. \tag{5.5}$$

Trading occurs at time t, $\mathbf{r}_{t^-}$ is the vector of returns at time t^- immediately before the transaction, $\mathbf{r}_t$ is the vector of returns immediately after the transaction, and this latter variable has incorporated market impacts. $\mathbf{MI}_t$ is the market impact caused by the transaction, and it consists of effects due to both excess buy orders, $\mathbf{N}_t^+$, and excess sell orders, $\mathbf{N}_t^-$. $\mathbf{e}_t$ are independently and identically distributed (i.i.d.) innovations, while Σ_t is the covariance matrix which is assumed to be unaffected by market impacts. Order imbalances $\mathbf{N}_t^-$ and $\mathbf{N}_t^+$ are calculated as follows.

Recall from Chapter 4 that the total wealth of a portfolio evolves as

$$W_{t+1} = (1 + \mathbf{w}_t^T \mathbf{r}_t) W_t + K_{t+1}, \tag{5.6}$$

where K_t is the cash injection, and $\mathbf{w}_t$ is the vector of portfolio weights.

Let the weight in asset i at time t^- immediately before the transaction be w_{i,t^-}, and this weight is given by

$$w_{i,t^-} = W_{t-1} w_{i,t-1} (1 + r_{i,t-1}) / W_{t^-},$$

where $W_{t^-} = (1 + \mathbf{w}_{t-1}^T \mathbf{r}_{t-1}) W_{t-1}$ is the instantaneous portfolio value just before the transaction, $w_{i,t-1}$ is the weight on asset i at time $t-1$, and $r_{i,t-1}$ is the return on asset i at time $t-1$.

Suppose we require a weight of $w_{i,t}$ of the total fund to be invested in asset i. Then the transaction should change the weight on asset i by

$$\begin{aligned} \Delta w_{i,t} &= w_{i,t} - w_{i,t^-} \\ &= w_{i,t} - W_{t-1} w_{i,t-1} (1 + r_{i,t-1}) / W_{t^-}. \end{aligned}$$

The accumulated value of asset i over the last period is

$$W_{t^-}\Delta w_{i,t}.$$

Therefore, the number transacted in asset class i at time t, based on the instantaneous asset prices just before the transaction, is

$$\begin{aligned} N_{i,t} &= \frac{W_{t^-}\Delta w_{i,t} + K_t w_{i,t}}{P_{i,t^-}} \qquad (5.7) \\ &= \frac{[(1+\mathbf{w}_{t-1}^T\mathbf{r}_{t-1})W_{t-1} + K_t]w_{i,t} - W_{t-1}w_{i,t-1}(1+r_{i,t-1})}{(1+r_{i,t^-})P_{i,0}\prod_{s=0}^{t-1}(1+r_{i,s})}, \end{aligned}$$

where P_{i,t^-} is the instantaneous price of asset i at time t^- just before the transaction, and $P_{i,0}$ is the initial price of asset i. The vector of number of buy orders is

$$\mathbf{N}_t^+ = \max(\mathbf{0}, \mathbf{N}_t), \qquad (5.8)$$

and the vector of number of sell orders is

$$\mathbf{N}_t^- = \left|\min(\mathbf{0}, \mathbf{N}_t)\right|. \qquad (5.9)$$

Note that $\mathbf{N}_t^-$ is the modulus of the negative order imbalance.

Remark 9. $\mathbf{N}_t^+$ and $\mathbf{N}_t^-$ in our complete system outlined in (5.1) to (5.5) represent the excess buy and sell orders in the whole market. However, the aggregate order imbalance in the market is complicated. It is reasonable for us to assume that transactions by other investors are at equilibrium and the only order imbalance is caused by ours.

5.2 Portfolio Wealth at Time t

This section is analogous to Section 4.4.

Suppose that the joint distribution of asset returns and accumulated values (portfolio wealths) is $f(\mathbf{r}_{t-1}, W_t)$. The first and second moments of the next period's wealth are given by:

$$\begin{aligned}\mathbb{E}(W_{t+1}) &= \mathbb{E}\left[(1+\mathbf{w}_t^T\mathbf{r}_t)W_t + K_{t+1}\right] \\ &= \mathbb{E}(W_t) + \mathbb{E}(\mathbf{w}_t^T\mathbf{r}_tW_t) + K_{t+1} \\ &= \mathbb{E}(W_t) + \int\int \mathbf{w}_t^T \begin{pmatrix} \boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} \\ +\mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^- \end{pmatrix} W_t f dW_t d\mathbf{r}_{t-1} + K_{t+1},\end{aligned} \tag{5.10}$$

and

$$\begin{aligned}\mathbb{E}(W_{t+1}^2) &= \mathbb{E}\begin{bmatrix} W_{t+1}^2 + (\mathbf{w}_t^T\mathbf{r}_tW_t)^2 + 2\mathbf{w}_t^T\mathbf{r}_tW_t^2 \\ +2(1+\mathbf{w}_t^T\mathbf{r}_t)W_tK_{t+1} + K_{t+1}^2 \end{bmatrix} \\ &= \left\{\begin{array}{c} \mathbb{E}(W_t^2) + \mathbb{E}\left\{\left[2\mathbf{w}_t^T\mathbf{r}_t + (\mathbf{w}_t^T\mathbf{r}_t)^2\right]W_t^2\right\} \\ +2K_{t+1}\mathbb{E}[(1+\mathbf{w}_t^T\mathbf{r}_t)W_t] + 2K_{t+1}^2 - K_{t+1}^2 \end{array}\right. \\ &= \left\{\begin{array}{c} \mathbb{E}(W_t^2) + \int\int\left[2\mathbf{w}_t^T\mathbf{r}_t + (\mathbf{w}_t^T\mathbf{r}_t)^2\right]W_t^2 f dW_t d\mathbf{r}_{t-1} \\ +2K_{t+1}\mathbb{E}[(1+\mathbf{w}_t^T\mathbf{r}_t)W_t + K_{t+1}] - K_{t+1}^2 \end{array}\right. \\ &= \left\{\begin{array}{c} \int\int \begin{bmatrix} 2\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right) \\ +\left[\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right)\right]^2 \\ +\mathbf{w}_t^T\boldsymbol{\Sigma}_t\mathbf{w}_t \end{bmatrix} W_t^2 f dW_t d\mathbf{r}_{t-1} \\ +\mathbb{E}(W_t^2) + 2K_{t+1}\mathbb{E}(W_{t+1}) - K_{t+1}^2 \end{array}\right. ,\end{aligned} \tag{5.11}$$

where $\mathbb{E}(W_t)$ and $\mathbb{E}(W_t^2)$ are known at time $t+1$.

Our goal is to minimize the expected variance for the next period's wealth for a given level of the expected next period's wealth. This desired level of the expected next period's wealth is determined by the targeted portfolio return μ_t, which is obtained from the time-varying efficient frontier (see Section 4.3). In summary, the problem is to minimize

$$\int\int \left[\begin{array}{c} 2\mathbf{w}_t^T \left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right) \\ + \left[\mathbf{w}_t^T \left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right)\right]^2 \\ +\mathbf{w}_t^T\boldsymbol{\Sigma}_t\mathbf{w}_t \end{array} \right] W_t^2 f dW_t d\mathbf{r}_{t-1},$$

subject to a given level of

$$\int\int \mathbf{w}_t^T(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-)W_t f dW_t d\mathbf{r}_{t-1}$$

and

$$\mathbf{w}_t^T\mathbf{j} = 1,$$

where $\mathbf{j} = [1\ 1\ 1]^T$ is a vector of ones.

The Lagrangian of the above problem is

$$\begin{aligned} \mathcal{L} = & \int\int \left[\begin{array}{c} 2\mathbf{w}_t^T \left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right) \\ + \left[\mathbf{w}_t^T \left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right)\right]^2 \\ +\mathbf{w}_t^T\boldsymbol{\Sigma}_t\mathbf{w}_t \end{array} \right] W_t^2 f dW_t d\mathbf{r}_{t-1} \\ & - 2\lambda_t \int\int \mathbf{w}_t^T(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-)W_t f dW_t d\mathbf{r}_{t-1} \\ & + 2\eta_t \int\int \mathbf{w}_t^T\mathbf{j}W_t p dW_t d\mathbf{r}_{t-1}, \end{aligned}$$

where the Lagrangian Multipliers λ_t and η_t are positive.

Consider a small variation, say $\boldsymbol{\epsilon}(\mathbf{r}_{t-1}, W_t)$ in $\mathbf{w}_t$, such that $\mathbf{j}^T\boldsymbol{\epsilon} = 0$. The resulting variation in $\mathcal{L}$ to the first order is:

$$\Delta\mathcal{L} = 2\int\int \boldsymbol{\epsilon}_t^T \left\{ \begin{array}{c} \left[\begin{array}{c} \left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right) \\ +\left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right) \\ \times\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right) \\ +\boldsymbol{\epsilon}_t^T\boldsymbol{\Sigma}_t\mathbf{w}_t \end{array} \right] W_t \\ -\lambda_t(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-) + \eta_t\mathbf{j} \end{array} \right\} W_t f dW_t d\mathbf{r}_{t-1}.$$

We require $\Delta\mathcal{L}$ to be insensitive to $\boldsymbol{\epsilon}(\mathbf{r}_{t-1}, W_t)$ when $\mathcal{L}$ is at its minimum. This is so when at least one of the following conditions holds:

$$f(\mathbf{r}_{t-1}, W_t) = 0;$$
$$\mu_t \text{ is at a minimum or maximum;}$$

or

$$\begin{array}{c} \left[\begin{array}{c} \left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right) \\ +\left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right) \\ \times\mathbf{w}_t^T\left(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-\right) \\ +\boldsymbol{\Sigma}_t\mathbf{w}_t \end{array} \right] W_t \\ = \lambda_t(\boldsymbol{\gamma} + \mathbf{C}_0\mathbf{r}_{t-1} + \mathbf{C}_1\mathbf{N}_t^+ + \mathbf{C}_2\mathbf{N}_t^-) - \eta\mathbf{j}. \end{array} \quad (5.12)$$

The Lagrangian Multiplier η_t is chosen to ensure that the weights sum up to 1, and λ_t is determined by the time-varying efficient frontier studied in Chapter 4 (see Section 4.2 to 4.4).

An analytical solution to $\mathbf{w}_t$ in (5.12) can not be obtained easily. The causality between $\mathbf{w}_t$ and $\mathbf{r}_t$ is bidirectional, that is, changes in $\mathbf{w}_t$ affects $\mathbf{r}_t$ and hence the expected portfolio return, while the expected portfolio return determines $\mathbf{w}_t$.

We will attempt to solve this problem numerically. In Chapter 4, we discussed two approaches to obtain numerical solutions to $\mathbf{w}_t$. These approaches can be modified in this case.

The first approach is to numerically solve for $\mathbf{w}_t$ in (5.12) by trying different combinations of portfolio weights. An alternative approach is to numerically minimize the portfolio variance $\mathbf{w}_t^T \boldsymbol{\Sigma}_t \mathbf{w}_t$ for a given targeted level of portfolio return. The intuition is that if we know how the change in portfolio weights will shift asset returns, we can simulate weights which give the desired portfolio return when market impacts exist. The optimal weights will be the set that gives the smallest portfolio variance.

Remark 10. The above algorithms, while they may be intuitively simple, suffer from high computational complexity. The computational requirement increases exponentially with the accuracy level, when constructing and calculating all possible combinations of portfolio weights. Starting from the weight for the first asset being zero and each increment being 10^{-5} until one, in a three-asset case, there will be approximately 10^{15} possible combinations. If we have k periods in each simulation and a total of m repeats, there will be approximately $mk \times 10^{15}$ evaluations of $\mathbf{w}_t^T \Sigma_t \mathbf{w}_t$. We utilized the service from Monash Campus Grid, which allowed us to run the simulations provided in the following section simultaneously on the computer grid. This service significantly shortened the time required.

5.3 Performance of the Dynamic Strategy

We study the performance of the proposed dynamic strategy by considering a portfolio consisting of three asset classes: equity, debt and cash. The total investment period is 20 years. We simulate ten thousand sets of twenty-year returns under the VAR-MGARCH model specified in (5.1) to (5.5). We adopt the same values for parameters $\boldsymbol{\gamma}, \mathbf{C}_0, \mathbf{A}_0, \mathbf{A}_1$ and $\mathbf{B}_1$ as in Section 4.4, for comparison purpose. We also require the innovations $\mathbf{e}_t$ in (5.4) follow a standard normal distribution.

Coefficients for excess buy and excess sell orders, $\mathbf{C}_1$ and $\mathbf{C}_2$, are chosen by following a the study of Chordia et al. (2002). Chordia et al. (2002) investigated the effects of contemporaneous order imbalances on the S&P500 stock market index returns for 1988-1998. They suggested that the coefficients of excess buy orders are positive and the coefficients of excess sell orders are negative on the US market. Brown et al. (1997) drew a similar conclusion on the Australian market, that, on the ASX, a buy order imbalance is weakly associated with higher future prices and a sell order imbalance is weakly associated with lower prices. We adopt the effects of excess buy and excess sell orders on equity from Chordia et al. (2002). The effects of excess buy and excess sell orders on debts (or cash) are chosen as the effects on equity multiplied by the ratio of unconditional variance in debts (or cash) to unconditional variance in equity. These effects are per 1000 shares transacted.

In summary, we have

$$\hat{\gamma} = \begin{bmatrix} 0.0724 \\ 0.0044 \\ 0.0081 \end{bmatrix}, \hat{\mathbf{C}}_0 = \begin{bmatrix} -0.3634 & 0.3535 & 0 \\ -0.0333 & 0.2173 & 0 \\ 0 & 0 & 0.6591 \end{bmatrix},$$

$$\hat{\mathbf{C}}_1 = \begin{bmatrix} 0.000683 & 0 & 0 \\ 0 & 0.000039 & 0 \\ 0 & 0 & 0.000003 \end{bmatrix},$$

$$\hat{\mathbf{C}}_2 = \begin{bmatrix} -0.002244 & 0 & 0 \\ 0 & -0.000129 & 0 \\ 0 & 0 & -0.000010 \end{bmatrix},$$

$$\hat{\mathbf{A}}_0 = \begin{bmatrix} 0.0243 & 0.0036 & 0.0028 \\ 0 & 1.9795 \times 10^{-7} & 5.2368 \times 10^{-8} \\ 0 & 0 & 1.6966 \times 10^{-11} \end{bmatrix},$$

$$\hat{\mathbf{A}}_1 = \begin{bmatrix} -4.2493 \times 10^{-4} & -8.6205 \times 10^{-5} & -5.5102 \times 10^{-5} \\ 0 & -1.9429 \times 10^{-4} & -5.1382 \times 10^{-5} \\ 0 & 0 & -3.9935 \times 10^{-7} \end{bmatrix}, \text{ and}$$

$$\hat{\mathbf{B}}_1 = \begin{bmatrix} -0.0653 & -0.0012 & 6.6802 \times 10^{-4} \\ 0 & -0.0568 & 0.0056 \\ 0 & 0 & -0.0783 \end{bmatrix}.$$

Note that we assume no cross-asset-class market impacts here by setting $\hat{\mathbf{C}}_1$

and $\hat{\mathbf{C}}_2$ to be diagonal. The estimated unconditional mean return is

$$\hat{\boldsymbol{\mu}} = (\mathbf{I} - \hat{\mathbf{C}}_0)^{-1}\hat{\boldsymbol{\gamma}} = \begin{bmatrix} 0.0540 \\ 0.0033 \\ 0.0239 \end{bmatrix}, \tag{5.13}$$

and the estimated unconditional covariance on returns is

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 5.9378 \times 10^{-4} & 8.6843 \times 10^{-5} & 6.8107 \times 10^{-5} \\ 8.6843 \times 10^{-5} & 1.2701 \times 10^{-5} & 9.9610 \times 10^{-6} \\ 6.8107 \times 10^{-5} & 9.9610 \times 10^{-6} & 7.8119 \times 10^{-6} \end{bmatrix}. \tag{5.14}$$

We start with an initial asset price vector of $\mathbf{P}_0 = [\ 1 \quad 1 \quad 1\ ]^T$, and an initial investment of $W_0 = 1000000$. There are constant cash inflows of $K_t = 10000$ at the beginning of each year. Notice that, compared to Chapter 4, we choose a large amount of investments. This is because market impacts are significant only when trade size is large. We set the targeted annual portfolio return to be 0.05 per annum for 20 years, which leads to a targeted terminal wealth of $\mathbb{E}(W_{20}) = 2983957$ calculated by the recursive relationship in (5.6). Portfolios are rebalanced annually under three strategies:

1. an unadjusted constrained dynamic strategy proposed in Chapter 4;
2. an adjusted constrained dynamic strategy proposed in this Chapter; and
3. a constrained static strategy which rebalances portfolio weights to $\mathbf{w}_c = [\ 0.8680 \quad 0 \quad 0.1320\ ]^T$ on equity, debt and cash, found on the con-

strained longrun efficient frontier[1].

A dynamic strategy is unadjusted when it does not consider the effect of market impacts. Strategies are constrained in the sense that shortselling is not allowed, that is, portfolio weights are within $[0, 1]$.

Different choices of portfolio weights will lead to different one-step-ahead returns, when incorporating market impacts into our simulation. Each step contributes to a unique terminal portfolio wealth. Therefore, unlike the simulations in Chapter 4, different strategies are not comparable as they are not based on the same set of asset returns. However, it is still worth plotting the distribution of logarithmic portfolio wealths. The general pattern will give us some insights of how market impact affects the performance of different strategies.

5.3.1 Simulated Results

Figure 5.1 plots the distributions of logarithmic portfolio wealth, $\ln(W_t)$, at the end of the first, the fifth, the tenth, the fifteenth and the twentieth year under an unadjusted constrained dynamic strategy, an adjusted constrained strategy and a constrained static strategy. Table 5.1 presents the means and standard deviations of $\ln(W_t)$ at the end of each period under the three different strategies.

The means of logarithmic portfolio wealth, $\ln(W_t)$, under both unadjusted and adjusted constrained dynamic strategy approach the targeted terminal value of 14.92 ($\ln[\mathbb{E}(W_{20})]$) gradually. The mean of logarithmic terminal

[1] The longrun efficient frontier is constructed with the unconditional mean returns given in (5.13 and the unconditional covariance of returns given in (5.14).

portfolio wealth under the constrained static strategy does not center around the targeted value. The dynamic strategies both provide more concentrated distributions of portfolio wealth, compared to the constrained static strategy, and these distributions become more concentrated over time. Such a result is consistent with our goal of minimizing portfolio variances. In this particular example, the standard deviation of the logarithmic terminal wealth, $\ln(W_{20})$, under an unadjusted constrained dynamic strategy is 0.0173, under an adjusted constrained dynamic strategy is 0.0165, and under a constrained static strategy is 0.0749. This is suggesting that, when trade sizes are significant, considering the effects of potential market impacts by performing an adjusted dynamic strategy can minimize the possibility of obtain extreme (much larger than or much smaller than the targeted value) terminal portfolio wealth.

5.3.2 Larger Coefficients and Larger Fund

The difference between the distributions of terminal wealth under an unadjusted and an adjusted constrained dynamic strategy in Figure 5.1 is small. Let us magnify the coefficients for excess buy and excess sell orders, and the fund size. Let

$$\hat{\mathbf{C}}_1 = \begin{bmatrix} 0.00683 & 0 & 0 \\ 0 & 0.00039 & 0 \\ 0 & 0 & 0.00003 \end{bmatrix} \text{ and}$$

$$\hat{\mathbf{C}}_2 = \begin{bmatrix} -0.02244 & 0 & 0 \\ 0 & -0.00129 & 0 \\ 0 & 0 & -0.00010 \end{bmatrix}$$

per 1000 shares transacted. Also let the initial investment be $W_0 = 1000000$ and the constant cash inflow be $K_t = 100000$ per year. The targeted terminal wealth after 20 years is $\mathbb{E}(W_{20}) = 29839572$ calculated by the recursive relationship in (5.6) for a 5% targeted annual return.

Figure 5.3 plots the distributions of the logarithmic terminal portfolio wealth under an unadjusted dynamic strategy and an adjusted dynamic strategy for the case with smaller coefficients and smaller fund size specified in the previous section, and the case with larger coefficients and larger fund size specified in this section. Our result shows clearly that, with larger coefficients and larger fund size, the adjusted dynamic strategy achieves terminal wealth that is closer to the target than the unadjusted dynamic strategy when market impact is significant.

5.4 Conclusion

In this chapter, we demonstrate an algorithm to modify our dynamic asset allocation rules when market impact is significant. We follow (Kissell and Glantz 2003) and model market impacts by order imbalance. We assume that transactions by other investors are at equilibrium and the majority of order imbalances is caused by us rebalancing our portfolio periodically. The experiment that we conducted is an approximation of reality, and the goal is to determine sensible and robust decisions for the underlying real-life problem.

The constructed model for market impacts considered in this chapter, is a simple template, and it can be generalized to more complicated cases. For example, Chordia and Subrahmanyam (2004) suggested that the expectation of price change is linearly dependent on contemporaneous and lagged order imbalances, and Chan and Fong (2000) suggest that order imbalance and lagged imbalances also have impact on asset volatility. Future work might include lags of order imbalances in both the return and volatility representations.

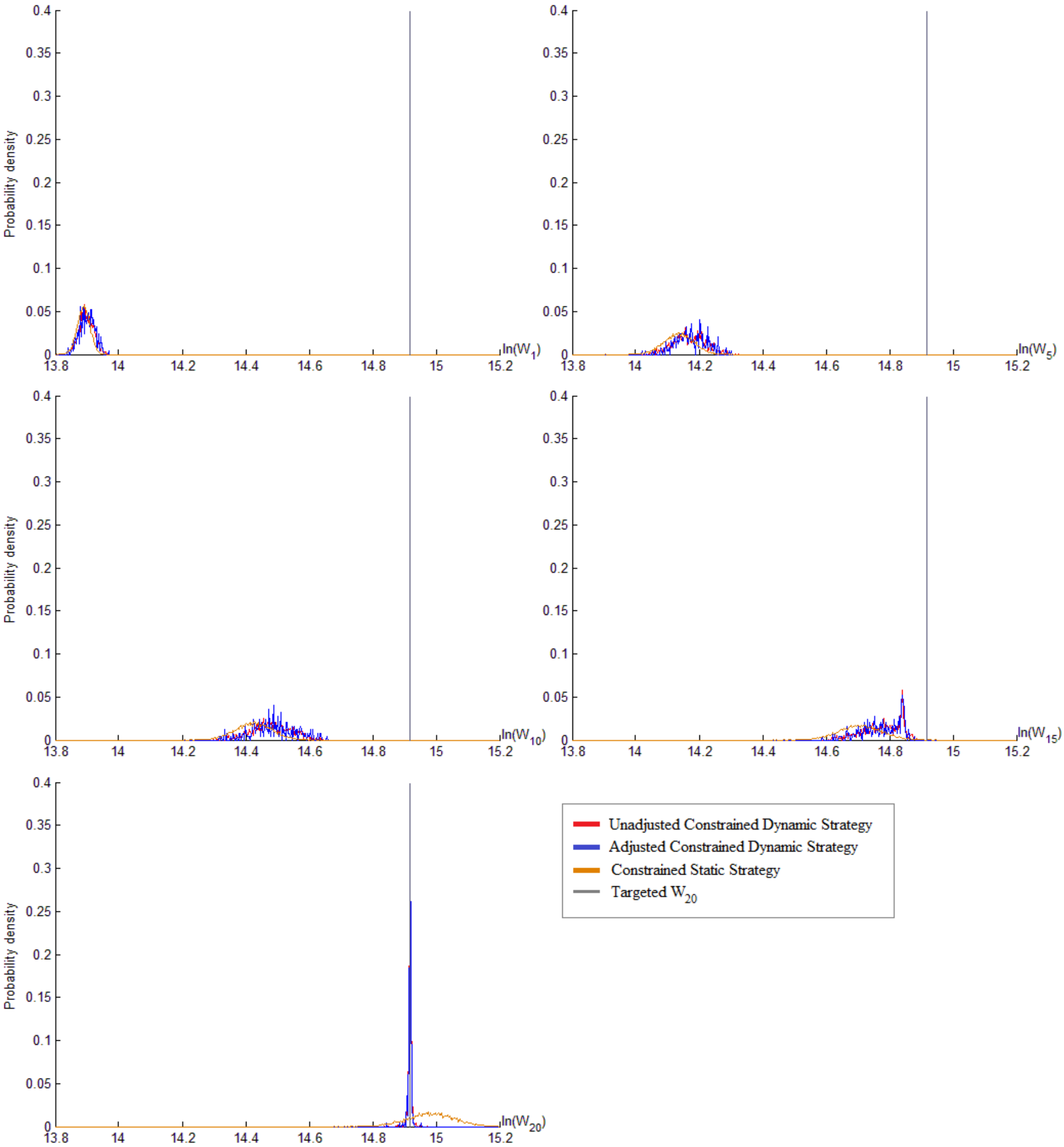

Figure 5.1: Distributions of the logarithmic portfolio wealth, $ln(W_t)$, at the end of the first, fifth, tenth, fifteenth and twentieth year due to an unadjusted constrained dynamic strategy, an adjusted constrained dynamic strategy, and a constrained static strategy.

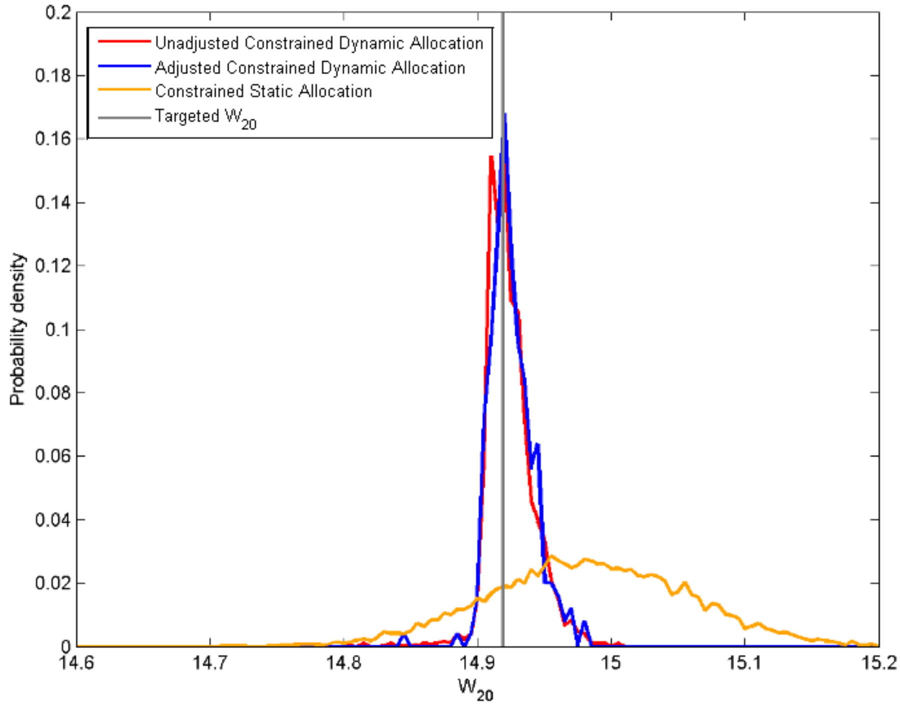

Figure 5.2: Distributions of logarithmic terminal portfolio wealth, $\ln(W_{20})$, under an unadjusted constrained dynamic strategy, an adjusted constrained dynamic strategy and a constrained static strategy.

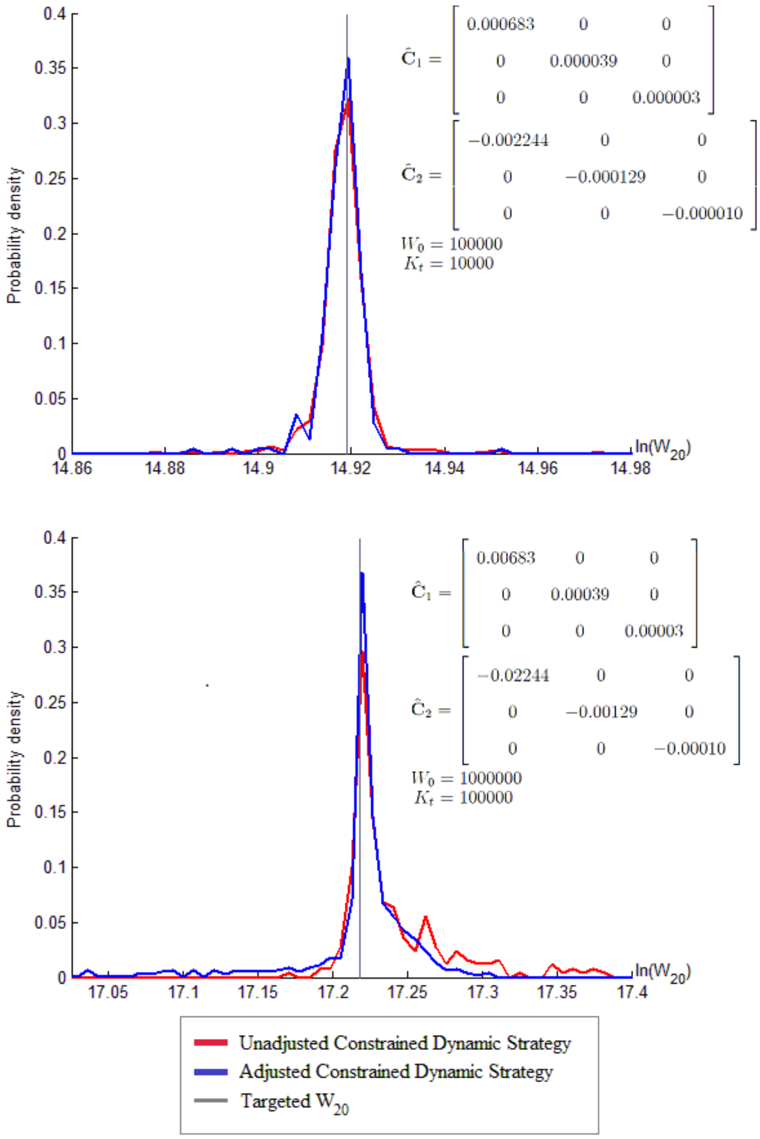

Figure 5.3: Distributions of logarithmic terminal portfolio wealth, $\ln(W_{20})$, under an unadjusted constrained dynamic strategy and an adjusted constrained dynamic strategy for the case with smaller coefficients and smaller fund size, and the case with larger coefficients and larger fund size.

	Average $ln(W_t)$		
t	Unadjusted constrained dynamic	Adjusted constrained dynamic	Constrained static
1	13.9005	13.9025	13.8912
2	13.9853	13.9893	13.9667
3	14.0426	14.0457	14.0198
4	14.1082	14.1110	14.0802
5	14.1715	14.1757	14.1386
6	14.2360	14.2408	14.1975
7	14.2984	14.3021	14.2557
8	14.3611	14.3652	14.3138
9	14.4227	14.4270	14.3718
10	14.4853	14.4912	14.4289
11	14.5480	14.5541	14.4858
12	14.6081	14.6125	14.5427
13	14.6690	14.6741	14.5992
14	14.7263	14.7277	14.6554
15	14.7800	14.7808	14.7115
16	14.8252	14.8249	14.7671
17	14.8610	14.8601	14.8230
18	14.8871	14.8855	14.8783
19	14.9061	14.9053	14.9332
20	14.9193	14.9189	14.9879

	Standard deviations of $ln(W_t)$		
t	Unadjusted constrained dynamic	Adjusted constrained dynamic	Constrained static
1	10.1180	10.1456	9.9716
2	10.85929	10.9217	10.7357
3	10.8972	10.8636	10.7750
4	11.0944	11.0993	10.9355
5	11.1919	11.1685	11.0486
6	11.3135	11.3445	11.1581
7	11.4096	11.4363	11.2639
8	11.5313	11.6098	11.3685
9	11.6316	11.7038	11.4703
10	11.7313	11.8277	11.5668
11	11.8426	11.9233	44.6636
12	11.9305	12.0222	11.7555
13	12.0020	12.0647	11.8379
14	12.0253	12.0291	11.9215
15	11.9668	11.9662	12.0028
16	11.7942	11.7887	12.0853
17	11.4773	11.4295	12.1621
18	11.0766	11.0360	12.2441
19	10.5269	10.4495	12.3223
20	10.0481	9.8956	12.3975

Table 5.1: Statistics of logarithmic portfolio values, $ln(W_t)$, under different strategies

Chapter 6

Conclusion and Possible Further Research

Asset liability management has always been one of the core concerns in finance. Dynamic asset allocation (investment strategy) has been identified as an appropriate approach to manage a portfolio. However, the formulation of an optimal dynamic investment strategy is complicated, and research in this area has had limited success. This thesis has addressed two different but related asset allocation problems and suggests a framework for dynamic asset allocation. A strategy is considered as optimal if it maximizes an investor's expected utility, or equivalently, minimizes the variance of the targeted terminal portfolio wealth.

The first asset allocation problem is when asset return models are path-independent (current or future states of returns do not depend on past states) and in continuous time. Continuous-time path-independent returns models, although not realistic, are popular approximations among academics. Cox and Leland (2000) assessed the criteria for optimal portfolio controls based on a two-asset portfolio, which consists of a risky asset and a risk-free one, and the price of the risky asset follows a geometric Brownian motion with constant drift and diffusion. We extend their work by allowing the drift and diffusion of the process followed by the price of the risky asset to be non-constant, and even path-independent functions of the price of the risky asset, portfolio wealth and time. We study two formulations of portfolio controls: when investments in the assets are functions of the price of the risky asset and time, and when investments in the assets are functions of portfolio wealth and time. The first formulation is adopted by investors who believe that portfolio wealth is purely driven by the price of the risky asset. The second formulation is adopted by investors wishing to manage market impact that results from their trading activities. Our contributions generalize the optimal portfolio control criteria outlined Cox and Leland (2000), and show that the generalized criteria are compatible with utility maximization.

The second part of our research considers asset allocation strategies when return models are path-dependent and in discrete time. In particular, Vector-Autoregressive - Multivariate Generalized Autoregressive Conditional Heteroskedasticity (VAR-MGARCH) models are considered.

We first study securities in the Australian market with a novel estimation technique, namely one-step multivariate semi-parametric maximum likelihood estimation (one-step MSMLE). This technique is an extension to univariate one-step semi-parametric maximum likelihood estimation (one-step SMLE) proposed by Di and Gangopadhyay (2013), and it utilizes kernel density estimation to analyze the joint distribution of asset returns. This type of estimation is one-step in the sense that it does not require a first step that performs a quasi-maximum likelihood estimation (QMLE) to acquire a residual series. We show that our extended estimator produces consistent and asymptotically unbiased and efficient estimates under certain regulatory conditions for kernel estimation. We utilize rolling windows that each consist of 120 effective months of indexed returns on Australian equity, debt and cash to estimate the VAR-MGARCH model. This recursive estimation allows us to investigate slow parameter changes over time. Our result show clear evidence of multivariate serial correlation in both the return and the covariance levels.

We then develop a class of optimal asset allocation strategies based on VAR-MGARCH return models, which extends the work by Leung (2011). The proposed strategy chooses targeted portfolio returns by minimizing the variance of expected terminal wealth at each scheduled rebalance time. It adjusts portfolio weights by attempting to minimize the expected variance for the next period's wealth implied by the targeted portfolio return. The minimization is conducted by looking up the desired portfolio weights on the efficient frontier through a Lagrangian, which is updated by observing the asset returns process and terminal wealth each period. We have also introduced an algorithm to adjust the proposed strategy to account for market impact. We provide a highly simplified but informative numerical example of managing the market impact on portfolio wealth when transaction sizes are significant.

This thesis has mainly focused on the first two moments of portfolio wealth when developing portfolio rules. However, as returns often have asymmetric distributions in practice, bringing in considerations of higher moments of portfolio wealth may improve the performance of investment strategies. Extension of our work may include studying the case when asset prices follow a jump diffusion process (which naturally involves higher moments) or the case when the volatility of asset returns is itself stochastic. On the other hand, we have only considered periodic portfolio rebalances in this thesis. Further work could also consider the situation when the rebalance time is not pre-determined but a function of time and portfolio wealth.

Appendix A

Detailed Working for Chapter 3

A.1 Proof of Proposition 8

We require the following lemma to prove that the density estimate converges pointwise in probability to the true density:

Lemma 1. Convergence in mean square implies convergence in probability. (This can easily be proved using Chebyshev's inequality.)

Let us construct a norm in Lebesgue space with dimension $p = 2$, which is an $\mathcal{L}^2$ norm, which is also known as the pointwise mean squared error (MSE). Then we have

$$\begin{aligned}
\text{MSE} &= \mathbb{E}\left\{\left|\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) - f(\mathbf{e};\boldsymbol{\theta})\right|^2\right\} \qquad \text{(A.1)}\\
&= \left\{\begin{array}{c} \left\{\mathbb{E}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right] - f(\mathbf{e};\boldsymbol{\theta})\right\}^2 \\ +\mathbb{E}\left\{[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})]^2\right\} - \left\{\mathbb{E}[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})]\right\}^2 \end{array}\right. \\
&= \left\{\text{Bias}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right]\right\}^2 + \text{Var}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right],
\end{aligned}$$

where $\text{Bias}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right]$ and $\text{Var}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right]$ are the expected bias and the variance of the density estimate $\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) = |\mathbf{H}|^{-1}K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right)$. The pointwise limit of $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ in $\mathcal{L}^2$ when the MSE goes to 0 will be $f(\mathbf{e};\boldsymbol{\theta})$. This requires that

$$\lim_{n\to\infty}\left\{\text{Bias}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right]\right\}^2 = \lim_{n\to\infty} Var\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right] = 0.$$

The bias and variance of the density estimate $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ are derived following pages 153 to 155 in Scott (1992) below.

Let us change the variables to $\boldsymbol{\omega} = \mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)$ so that $\left|\frac{d\boldsymbol{\omega}}{d\mathbf{e}_s}\right| = |\mathbf{H}|^{-1} = h^{-d}$. The expectation of the density estimate $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ is then

$$\mathbb{E}\left\{\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right\} = \int_{\mathbb{R}^d} K(\boldsymbol{\omega})\, f(\mathbf{e}-\mathbf{H}\boldsymbol{\omega};\boldsymbol{\theta})d\boldsymbol{\varpi}_s \tag{A.2}$$

The second order Taylor expansion of the expectation around $\mathbf{H}\boldsymbol{\omega} = \mathbf{0}$ is

$$\mathbb{E}\left\{\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right\} = \int_{\mathbb{R}^d} K(\boldsymbol{\omega})\begin{bmatrix} f(\mathbf{e};\boldsymbol{\theta}) - \boldsymbol{\omega}^T\mathbf{H}^T\nabla f_n(\mathbf{e};\boldsymbol{\theta}) \\ +\frac{1}{2}\boldsymbol{\varpi}^T\mathbf{H}^T\nabla^2 f(\mathbf{e};\boldsymbol{\theta})\,\mathbf{H}\boldsymbol{\omega} + O(h^4)\end{bmatrix} d\boldsymbol{\omega} \tag{A.3}$$

$$= \left\{\begin{array}{c} f(\mathbf{e};\boldsymbol{\theta})\int_{\mathbb{R}^d} K(\boldsymbol{\omega})\,d\boldsymbol{\omega} - \nabla f(\mathbf{e};\boldsymbol{\theta})\,\mathbf{H}\int_{\mathbb{R}^d}\boldsymbol{\omega}K(\boldsymbol{\omega})\,d\boldsymbol{\omega} \\ +\frac{1}{2}\int_{\mathbb{R}^d} K(\boldsymbol{\omega})\,\boldsymbol{\omega}^T\mathbf{H}^T\nabla^2 f(\mathbf{e};\boldsymbol{\theta})\,\mathbf{H}\boldsymbol{\omega}d\boldsymbol{\omega} + O(h^4)\end{array}\right\}.^1$$

Recalling the properties of $K(\boldsymbol{\omega})$ in (3.8) to (3.10), the above becomes

$$\mathbb{E}\left\{\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right\} = f(\mathbf{e};\boldsymbol{\theta}) - 0 + \frac{1}{2}\int_{\mathbb{R}^d} K(\boldsymbol{\omega})\,\boldsymbol{\omega}^T\mathbf{H}^T\nabla^2 f(\mathbf{e};\boldsymbol{\theta})\,\mathbf{H}\boldsymbol{\omega}d\boldsymbol{\omega} + O(h^4). \tag{A.4}$$

[1] The remainder terms here are $O(h^4)$. This is because kernel functions are symmetric, and their third moments are 0.

The integrand in the third term of (A.4) is a scalar, so its trace equals itself. Hence, using the usual properties of the trace operator, the third term in (A.4) becomes

$$\begin{aligned}
&\frac{1}{2}tr\left[\int_{\mathbb{R}^d} K(\boldsymbol{\omega})\,\boldsymbol{\omega}\boldsymbol{\omega}^T\mathbf{H}^T\nabla^2 f(\mathbf{e};\boldsymbol{\theta})\,\mathbf{H}\mathbf{d}\boldsymbol{\omega}\right] \\
&= \frac{1}{2}tr\left[\left(\int_{\mathbb{R}^d} K(\boldsymbol{\omega})\,\boldsymbol{\omega}\boldsymbol{\omega}^T\mathbf{d}\boldsymbol{\omega}\right)\mathbf{H}^T\nabla^2 f(\mathbf{e};\boldsymbol{\theta})\,\mathbf{H}\right] \\
&= \frac{1}{2}tr\left[\mathbf{H}^T\nabla^2 f(\mathbf{e};\boldsymbol{\theta})\,\mathbf{H}\right] \\
&= \frac{1}{2}h^2 tr(\mathbf{A}\mathbf{A}^T\nabla^2 f(\mathbf{e};\boldsymbol{\theta})\,.
\end{aligned}$$

Hence, the expectation of the density estimate of $\mathbf{e}$ is

$$\mathbb{E}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right] = f(\mathbf{e};\boldsymbol{\theta}) + \frac{1}{2}h^2 tr(\mathbf{A}\mathbf{A}^T\nabla^2 f(\mathbf{e};\boldsymbol{\theta}) + O(h^4), \tag{A.5}$$

and the expected bias in $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ is

$$\text{Bias}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right] = \frac{1}{2}h^2 tr(\mathbf{A}\mathbf{A}^T\nabla^2 f(\mathbf{e};\boldsymbol{\theta}) + O(h^4). \tag{A.6}$$

We deduce that

$$\lim_{n\to\infty}\left\{\text{Bias}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right]\right\}^2 = 0,$$

as $h \to 0$ (from (3.12)).

The variance of $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ may be written as

$$\text{Var}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right] = \text{Var}\left[n^{-1}\left|\mathbf{H}\right|^{-1}\sum_{s=1}^{n} K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right)\right] \tag{A.7}$$

$$= \frac{1}{n}\left[\mathbb{E}\left\{\left[|\mathbf{H}|^{-1}K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right)\right]^2\right\} - \left\{\mathbb{E}\left[|\mathbf{H}|^{-1}K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right)\right]\right\}^2\right].$$

The first term in the square brackets in (A.7) expands as

$$\mathbb{E}\left\{\left[|\mathbf{H}|^{-1}K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right)\right]^2\right\} = \int_{\mathbb{R}^d}\left[|\mathbf{H}|^{-1}K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right)\right]^2 f(\mathbf{e}_s;\boldsymbol{\theta})d\mathbf{e}_s \tag{A.8}$$

$$= h^{-d}\int_{\mathbb{R}^d} K^2\left(\boldsymbol{\omega}\right) f(\mathbf{e};\boldsymbol{\theta})d\boldsymbol{\omega} + O(h^{1-d}),$$

and the second term expands as

$$\left\{\mathbb{E}\left[|\mathbf{H}|^{-1}K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right)\right]\right\}^2 = \left\{f\left(\mathbf{e};\boldsymbol{\theta}\right) + \frac{1}{2}h^2 tr(\mathbf{A}\mathbf{A}^T\nabla^2 f\left(\mathbf{e};\boldsymbol{\theta}\right) + O(h^4)\right\}^2 \tag{A.9}$$

$$= f^2\left(\mathbf{e};\boldsymbol{\theta}\right) + O(h^2).$$

Therefore, for the variance in (A.7) to vanish, we require that $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$.

In summary, $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right) \stackrel{m.s.}{\to} f(\mathbf{e};\boldsymbol{\theta})$ and $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right) \stackrel{p}{\to} f(\mathbf{e};\boldsymbol{\theta})$ when $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$.

A.2 Proof of Proposition 9

Definition 13. *$\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$ uniformly converges in probability to $f(\mathbf{e};\boldsymbol{\theta})$ if for every $\epsilon > 0$,*

$$\mathbb{P}\left[\sup_{\mathbf{e}}\left|\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right) - f(\mathbf{e};\boldsymbol{\theta})\right| < \epsilon\right] = 1. \tag{A.10}$$

We provide the following proof in light of Theorem 3A in Parzen (1962).

To show (A.10) is equivalent to show that

$$\lim_{n\to\infty} \sqrt{\mathbb{E}\left[\sup_{\mathbf{e}} \left|\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right) - f(\mathbf{e};\boldsymbol{\theta})\right|^2\right]} = 0. \tag{A.11}$$

It follows from Proposition 8 that, when $h \to 0$,

$$\lim_{n\to\infty} \sup_{\mathbf{e}} \left|\mathbb{E}\left[\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)\right] - f(\mathbf{e};\boldsymbol{\theta})\right| = 0.$$

Hence, to show (A.11), it suffices to show that

$$\lim_{n\to\infty} \sqrt{\mathbb{E}\left[\sup_{\mathbf{e}} \left|\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right) - \mathbb{E}\left[\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)\right]\right|^2\right]} = 0. \tag{A.12}$$

The Fourier transform $\mathfrak{F}(\boldsymbol{\tau})$ of the density estimate $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$, when both $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$ and $\mathfrak{F}(\boldsymbol{\tau})$ are integrable, is defined as

$$\begin{aligned}\mathfrak{F}(\boldsymbol{\tau}) &= \int_{\mathbb{R}^d} e^{i\boldsymbol{\tau}^T\mathbf{e}} \hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right) d\mathbf{e},\ \boldsymbol{\tau}\in\mathbb{R}^d \\ &= \mathbb{E}_{\hat{f}_n}\left[e^{i\boldsymbol{\tau}^T\mathbf{e}}\right].\end{aligned}$$

Substituting $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$ in (3.7) into its Fourier transform, we have

$$\begin{aligned}\mathfrak{F}(\boldsymbol{\tau}) &= \int_{\mathbb{R}^d} e^{i\boldsymbol{\tau}^T\mathbf{e}} n^{-1}|\mathbf{H}|^{-1} \sum_{s=1}^{n} K\left(\mathbf{H}^{-1}\left[\mathbf{e} - \mathbf{e}_s(\boldsymbol{\theta})\right]\right) d\mathbf{e},\ \boldsymbol{\tau}\in\mathbb{R}^d \\ &= n^{-1}\sum_{s=1}^{n} e^{i\boldsymbol{\tau}^T\mathbf{e}_s} \int_{\mathbb{R}^d} e^{i\mathbf{H}\boldsymbol{\tau}^T\boldsymbol{\varpi}} K\left(\boldsymbol{\omega}\right) d\boldsymbol{\omega} \\ &= n^{-1}\sum_{s=1}^{n} e^{i\boldsymbol{\tau}^T\mathbf{e}_s} \mathfrak{K}(\mathbf{H}\boldsymbol{\tau}),\end{aligned}$$

where $\mathfrak{K}(\mathbf{H}\boldsymbol{\tau})$ is the Fourier transform of $K(\boldsymbol{\omega})$ and $\boldsymbol{\omega}= \mathbf{H}^{-1}[\mathbf{e}-\mathbf{e}_s]$.

$\mathfrak{F}(\boldsymbol{\tau})$ is integrable when $\mathfrak{K}(\mathbf{H}\boldsymbol{\tau})$ is integrable. The integrability of $\mathfrak{F}(\boldsymbol{\tau})$ implies that $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ is uniformly continuous. According to the inversion theorem of Fourier transform, $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ can be written as

$$\begin{aligned}\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) &= (2\pi)^{-1}\int_{\mathbb{R}^d} e^{-i\boldsymbol{\tau}^T\mathbf{e}}\mathfrak{F}(\boldsymbol{\tau})d\boldsymbol{\tau},\ \mathbf{e}\in\mathbb{R}^d \qquad \text{(A.13)}\\ &= (2\pi)^{-1}\int_{\mathbb{R}^d} e^{-i\boldsymbol{\tau}^T\mathbf{e}}n^{-1}\sum_{s=1}^{n} e^{i\boldsymbol{\tau}^T\mathbf{e}_s}\mathfrak{K}(\mathbf{H}\boldsymbol{\tau})d\boldsymbol{\tau}.\end{aligned}$$

Let

$$\phi_n(\boldsymbol{\tau}) = n^{-1}\sum_{s=1}^{n} e^{i\boldsymbol{\tau}^T\mathbf{e}_s}.$$

Taking expectation of (A.13), we have

$$\mathbb{E}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right] = (2\pi)^{-1}\int_{\mathbb{R}^d} e^{-i\boldsymbol{\tau}^T\mathbf{e}}\mathbb{E}\left[\phi_n(\mathbf{t})\right]\mathfrak{K}(\mathbf{H}\boldsymbol{\tau})d\boldsymbol{\tau},$$

and so,

$$\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) - \mathbb{E}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right] = (2\pi)^{-1}\int_{\mathbb{R}^d} e^{-i\boldsymbol{\tau}^T\mathbf{e}}\left\{\phi_n(\boldsymbol{\tau}) - \mathbb{E}\left[\phi_n(\boldsymbol{\tau})\right]\right\}\mathfrak{K}(\mathbf{H}\boldsymbol{\tau})d\boldsymbol{\tau}.$$

The triangle inequality yields that $\left|e^{-i\boldsymbol{\tau}^T\mathbf{e}}\right| \leq 1$, and hence

$$\sup_{\mathbf{e}}\left|\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) - \mathbb{E}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right]\right| = (2\pi)^{-1}\int_{\mathbb{R}^d}\left|\phi_n(\boldsymbol{\tau}) - \mathbb{E}\left[\phi_n(\boldsymbol{\tau})\right]\right|\left|\mathfrak{K}(\mathbf{H}\boldsymbol{\tau})\right|d\boldsymbol{\tau}.$$

The quantity in (A.12) is then

$$\sqrt{\mathbb{E}\left[\sup_{\mathbf{e}}\left|\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) - \mathbb{E}\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta})\right]\right|^2\right]} = (2\pi)^{-1}\int_{\mathbb{R}^d}\sqrt{Var[\phi_n(\boldsymbol{\tau})]}\left|\mathfrak{K}(\mathbf{H}\boldsymbol{\tau})\right|d\boldsymbol{\tau},$$

where $Var[\phi_n(\boldsymbol{\tau})] = \mathbb{E}\left[|\phi_n(\boldsymbol{\tau}) - \mathbb{E}\left[\phi_n(\boldsymbol{\tau})\right]|^2\right]$ due to the complex component involved. Let us rewrite the complex component in $\phi_n(\boldsymbol{\tau})$ as

$$e^{i\boldsymbol{\tau}^T\mathbf{e}_s} = \cos(\boldsymbol{\tau}^T\mathbf{e}_s) + i\sin(\boldsymbol{\tau}^T\mathbf{e}_s).$$

Hence,

$$\begin{aligned} Var(e^{i\boldsymbol{\tau}^T\mathbf{e}_s}) &= Var(\cos(\boldsymbol{\tau}^T\mathbf{e}_s)) + Var(\sin(\boldsymbol{\tau}^T\mathbf{e}_s)) \\ &\leq \mathbb{E}\left[\cos^2(\boldsymbol{\tau}^T\mathbf{e}_s) + \sin^2(\boldsymbol{\tau}^T\mathbf{e}_s)\right] = 1. \end{aligned}$$

Because $\mathbf{e}_s$ are i.i.d., we have

$$Var[\phi_n(\boldsymbol{\tau})] = n^{-2}\sum_{s=1}^{n} Var(e^{i\boldsymbol{\tau}^T\mathbf{e}_s}) \leq \frac{1}{n}.$$

It follows then that

$$\begin{aligned} &\sqrt{\mathbb{E}\left[\sup_{\mathbf{e}}\left|\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right) - \mathbb{E}\left[\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)\right]\right|^2\right]} \\ &\leq (2\pi)^{-1}n^{-1/2}\int_{\mathbb{R}^d}|\mathfrak{K}(\mathbf{H}\boldsymbol{\tau})|\,d\boldsymbol{\tau} \\ &= (2\pi)^{-1}n^{-1/2}\,|\mathbf{H}|^{-1}\int_{\mathbb{R}^d}|\mathfrak{K}(\mathbf{x})|\,d\mathbf{x}, \end{aligned}$$

where $\mathbf{x} = \mathbf{H}\boldsymbol{\tau}$. For this value to vanish, we require that and $nh^{2d} \to \infty$ as $n \to \infty$.

In summary, $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$ converges uniformly in probability to $f(\mathbf{e};\boldsymbol{\theta})$ if

(i) $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$ is uniformly continuous,

(ii) the Fourier transforms of $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ and $K\left(\mathbf{H}^{-1}\left[\mathbf{e}-\mathbf{e}_s\right]\right)$ exist, and

(iii) $nh^{2d} \to \infty$ as $n \to \infty$..

A.3 Proof of Proposition 10

We need the following lemmata to prove the consistency of the one-step MSMLE:

Lemma 2. Let (Ω, F) be a measurable space, where Ω is the sample space and F is the sigma-algebra generated by Ω. Let τ be an arbitrary real-valued function on $\Omega \times \Theta$. If $\tau(x, \boldsymbol{\theta})$ is measurable in x for every $\boldsymbol{\theta}$ in Θ and continuous in $\boldsymbol{\theta}$ for every x in Ω, and Θ is compact, then there exists a measurable function which maps Ω to Θ. (See the proof of Lemma 2 in Jennrich (1969)).

Lemma 3. Consider a probability space $(\Omega, F, \{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\})$, where Ω is the sample space, F is the sigma-algebra generated by Ω, $\Theta \subseteq \mathbb{R}^k$ is a non-empty parameter set and $\{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ is a family of probability measures. Suppose that $g(x, \boldsymbol{\theta})$ is measurable in x for every $\boldsymbol{\theta} \in \Theta$ and continuous in $\boldsymbol{\theta}$ for every $x \in \Omega$, and Θ is compact. Let Γ denote the set of all measurable functions mapping Ω to Θ. If for any $\boldsymbol{\theta} \in \Theta$, there exists a map $\gamma(x) \in \Gamma$ such that $\gamma^{-1}(\boldsymbol{\theta})$ is a measurable set, then

$$\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}[g(x,\boldsymbol{\theta})] = \mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta} g(x,\boldsymbol{\theta})], \tag{A.14}$$

where the expectation is taken over x. (The proof of lemma is in Appendix A.6.)

Let us define

$$\mathbb{D}(\mathbf{u},\boldsymbol{\theta}) = \ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right) - \ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}^*\right). \tag{A.15}$$

The semiparametric log-likelihood $\hat{l}(\boldsymbol{\theta}|\mathbf{u}) = \frac{1}{n}\sum_{t=1}^{n} \ln \hat{f}_n\left(\mathbf{u}_t;\boldsymbol{\theta}\right)$ is continuous in $\boldsymbol{\theta}$, hence $\mathbb{D}(\mathbf{u},\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$.

Assumption 2 (Identification) indicates that, $\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right) \neq \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}^*\right)$ for any $\boldsymbol{\theta}\neq\boldsymbol{\theta}^*$. By Kullback-Leibler information inequality for density functions (Gibbs 1902), We have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}^*}\left[\mathbb{D}(\mathbf{u},\boldsymbol{\theta})\right] = \mathbb{E}_{\boldsymbol{\theta}^*}\left[\ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right) - \ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}^*\right)\right] \\
&= \mathbb{E}_{\boldsymbol{\theta}^*}\left[\ln \frac{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}^*\right)}\right] \\
&\leq \ln \mathbb{E}_{\boldsymbol{\theta}^*}\left[\frac{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}^*\right)}\right] \\
&= \ln\left[\int_{\Omega} \frac{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}^*\right)} \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}^*\right) d\mathbf{u}\right] \\
&= \ln 1 = 0
\end{aligned} \tag{A.16}$$

with equality when $\boldsymbol{\theta}=\boldsymbol{\theta}^*$.

Let

$$U(\boldsymbol{\theta}) := \sup_{\boldsymbol{\theta}\in\Theta} \left|\mathbb{D}(\mathbf{u},\boldsymbol{\theta})\right|,$$

then

$$\mathbb{E}[U(\boldsymbol{\theta})] = \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta} \left|\mathbb{D}(\mathbf{u},\boldsymbol{\theta})\right|\right]$$

$$\leq \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}\left|\ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)\right|\right] + \mathbb{E}\left|\ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}^*\right)\right|.$$

$\mathbb{E}[U(\boldsymbol{\theta})]$ is finite if and only if $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}\left|\ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)\right|\right]$ and $\mathbb{E}\left|\ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}^*\right)\right|$ are finite.

According to Lemma 2 and 3 in this section, the first term above becomes

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}\left|\ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)\right|\right] = \sup_{\boldsymbol{\theta}\in\Theta}\left[\mathbb{E}\left|\ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)\right|\right]. \tag{A.17}$$

$\mathbb{E}\left|\ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)\right|$ expands as follows:

$$\begin{aligned}\mathbb{E}\left|\ln \hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)\right| &= \mathbb{E}\left|-\frac{1}{2}\ln\left|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right| + \ln \hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)\right| \\ &\leq \frac{1}{2}\mathbb{E}\left|\ln\left|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right|\right| + \mathbb{E}\left|\ln \hat{\mathrm{f}}_n\left(\mathbf{e};\boldsymbol{\theta}\right)\right|.\end{aligned} \tag{A.18}$$

We first need to show that $\mathbb{E}\left|\ln\left|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right|\right| < \infty$. $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\,(\boldsymbol{\theta})$ is a positive-definite covariance matrix, hence $\left|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right| > 0$ and $\ln\left|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right| > -\infty$. Let λ_i be the eigenvalues of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, where $i = 1, 2, ..., d$. All λ_i are positive. Then we have

$$\begin{aligned}\ln\left|\Sigma\left(\boldsymbol{\theta}\right)\right| &= \ln\prod_{i=1}^{d}\lambda_i = \sum_{i=1}^{d}\ln\lambda_i \\ &\leq \sum_{i=1}^{d}\lambda_i = tr\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right].\end{aligned}$$

By the squared integrability of $\mathbf{u}$, $\mathbb{E}\left\{vech\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]\right|\} < \infty$ and hence $\ln\left|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right| \leq tr\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right] < \infty$ (Williams 1991).

We now show that $\mathbb{E}\left|\ln \hat{\mathrm{f}}_n\left(\mathbf{e};\boldsymbol{\theta}\right)\right| < \infty$. The density estimate $\hat{f}_n\left(\mathbf{e};\boldsymbol{\theta}\right)$ is always positive. For $0 < \hat{f}_n(\mathbf{e};\boldsymbol{\theta}) \leq 1$, we have $-\infty < \ln \hat{f}_n(\mathbf{e};\boldsymbol{\theta}) \leq 0$ and $0 \leq \mathbb{E}\left|\ln \hat{\mathrm{f}}_n(\mathbf{e};\boldsymbol{\theta})\right| < \infty$. We have shown in Proposition 8 that $\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) \xrightarrow{p} f(\mathbf{e};\boldsymbol{\theta})$ pointwise when $nh \to \infty$ and $h \to 0$. The true density $\mathrm{f}(\mathbf{e},\boldsymbol{\theta})$ is assumed to be finite. Hence when $\hat{\mathrm{f}}_n(\mathbf{e};\boldsymbol{\theta}) > 1$, we have

$$\mathbb{E}\left[\ln \hat{\mathrm{f}}_n(\mathbf{e};\boldsymbol{\theta})\right] \leq \ln \mathbb{E}\left[\hat{\mathrm{f}}_n(\mathbf{e};\boldsymbol{\theta})\right] = \ln \mathrm{f}(\mathbf{e};\boldsymbol{\theta}) < \infty.$$

Recall the assumption that $\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ for every $\mathbf{e}$ in the sample space. In addition, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$. Hence, $\hat{f}_n(\mathbf{u};\boldsymbol{\theta}) = |\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|^{-1/2}\hat{f}_n(\mathbf{e};\boldsymbol{\theta})$ and $\ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$ for every $\mathbf{u}$ in Ω. We now have

$$\begin{aligned}
\mathbb{E}[U(\boldsymbol{\theta})] &\leq \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}\left|\ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta})\right| + \left|\ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta}^*)\right|\right] \\
&= \sup_{\boldsymbol{\theta}\in\Theta}\mathbb{E}\left[\left|\ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta})\right|\right] + \mathbb{E}\left|\ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta}^*)\right| \\
&< \infty,
\end{aligned}$$

and so

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{D}(\mathbf{u}_t,\boldsymbol{\theta}) \xrightarrow{a.s.} \mathbb{E}_{\boldsymbol{\theta}^*}\left[\mathbb{D}(\mathbf{u},\boldsymbol{\theta})\right], \text{ uniformly.} \tag{A.19}$$

To show the weak consistency of $\hat{\boldsymbol{\theta}}_{1SMLE}$, we define a compact set $S := \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \geq \rho, \forall \rho > 0\}$, which is a subset of Θ. By the identification condition in (A.16), we have

$$\sup_{\boldsymbol{\theta}\in S}\frac{1}{n}\sum_{t=1}^{n}\mathbb{D}(\mathbf{u}_t,\boldsymbol{\theta}) < 0 \text{ with probability 1.}$$

Therefore, the $\hat{\boldsymbol{\theta}}_{1SMLE}$, which maximizes $\frac{1}{n}\sum_{t=1}^{n}\mathbb{D}(\mathbf{u}_t,\boldsymbol{\theta})$ and hence maximizes $\frac{1}{n}\sum_{t=1}^{n}\ln\hat{f}_n\left(\mathbf{u}_t;\boldsymbol{\theta}\right)$, is not in S, so that

$$\lim_{n\to\infty}\Pr(\left\|\hat{\boldsymbol{\theta}}_{1SMLE}-\boldsymbol{\theta}^*\right\|\geq\rho)=0,\forall\rho>0,$$

that is

$$\hat{\boldsymbol{\theta}}_{1SMLE}\xrightarrow{p}\boldsymbol{\theta}^*\text{ pointwise.}$$

Let us now define a compact set $S':=\{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\leq\rho,\forall\rho>0\}$, which is also a subset of Θ. Strong consistency of $\hat{\boldsymbol{\theta}}_{1SMLE}$ requires

$$\Pr(\lim_{n\to\infty}\left\|\hat{\boldsymbol{\theta}}_{1SMLE}-\boldsymbol{\theta}^*\right\|<\rho)=1,\forall\rho>0 \tag{A.20}$$

If (A.20) does not hold, there exists a non-null set

$$Z:=\left\{\tilde{\boldsymbol{\theta}}\in\Theta:\lim_{n\to\infty}\sup\left\|\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\|>\rho,\forall\rho>0\right\}$$

. There also exists a non-null subset of Z that,

$$Z'\subset Z=\left\{\tilde{\boldsymbol{\theta}}\in Z:\tilde{\boldsymbol{\theta}}\xrightarrow{a.s.}\boldsymbol{\theta}\in S'\right\}.$$

We then have

$$\begin{aligned}\lim_{n\to\infty}\sup\mathbb{E}_{\boldsymbol{\theta}^*}\left[\mathbb{D}(\mathbf{u},\tilde{\boldsymbol{\theta}})\right]&\leq\mathbb{E}_{\boldsymbol{\theta}^*}\left[\lim_{n\to\infty}\sup_{\tilde{\boldsymbol{\theta}}\in Z'}\mathbb{D}(\mathbf{u},\tilde{\boldsymbol{\theta}})\right]\\&=\mathop{\mathbb{E}_{\boldsymbol{\theta}^*}}_{\boldsymbol{\theta}\in S'}\left[\mathbb{D}(\mathbf{u},\boldsymbol{\theta})\right]\leq 0\\&=\sup_{\boldsymbol{\theta}\in S'}\mathbb{E}_{\boldsymbol{\theta}^*}\left[\mathbb{D}(\mathbf{u},\boldsymbol{\theta})\right].\end{aligned}$$

The above contradicts that $\frac{1}{n}\sum_{t=1}^{n}\mathbb{D}(\mathbf{u}_t,\boldsymbol{\theta})=0$ if and only if $\boldsymbol{\theta}$=$\boldsymbol{\theta}^*$. Therefore, $\hat{\boldsymbol{\theta}}_{1SMLE}$, which maximizes $\frac{1}{n}\sum_{t=1}^{n}\mathbb{D}(\mathbf{u}_t,\boldsymbol{\theta})$ and hence maximizes $\frac{1}{n}\sum_{t=1}^{n}\ln\hat{f}_n\left(\mathbf{u}_t;\boldsymbol{\theta}\right)$, satisfies (A.20).

A.4 Proof of Proposition 11

The following proof is closely in line with pages 470 to 472 in (Hayashi 2000).

Suppose the semiparametric log-likelihood $\hat{l}_n(\boldsymbol{\theta}|\mathbf{u})=\frac{1}{n}\sum_{t=1}^{n}\ln\hat{\mathrm{f}}_n\left(\mathbf{u}_t;\boldsymbol{\theta}\right)$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}_{1SMLE}$ is in the interior of Θ. The first order derivative of $\ln\hat{\mathrm{f}}_n\left(\mathbf{u};\boldsymbol{\theta}\right)$ with respect to $\boldsymbol{\theta}$ is the semiparametric score $\hat{\mathbb{S}}$ defined in (3.13), and the second order derivative of $\ln\hat{\mathrm{f}}_n\left(\mathbf{u};\boldsymbol{\theta}\right)$ with respect to $\boldsymbol{\theta}$ is the semiparametric Hessian $\hat{\mathbb{H}}$ defined in (3.14).

The mean value expansion of $\hat{\mathbb{S}}(\mathbf{u},\hat{\boldsymbol{\theta}}_{1SMLE})$ is

$$\hat{\mathbb{S}}(\mathbf{u},\hat{\boldsymbol{\theta}}_{1SMLE})=\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta}^*)+\hat{\mathbb{H}}(\mathbf{u},\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{1SMLE}-\boldsymbol{\theta}^*), \tag{A.21}$$

where $\tilde{\boldsymbol{\theta}}$ is the mean value lies between $\hat{\boldsymbol{\theta}}_{1SMLE}$ and $\boldsymbol{\theta}^*$. The estimate $\hat{\boldsymbol{\theta}}_{1SMLE}$ is obtained by maximizing $\hat{l}_n(\boldsymbol{\theta}|\mathbf{u})$. Hence,

$$\left.\frac{\partial\hat{l}_n(\boldsymbol{\theta}|\mathbf{u})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{1SMLE}}=\frac{1}{n}\sum_{t=1}^{n}\hat{\mathbb{S}}(\mathbf{u}_t,\hat{\boldsymbol{\theta}}_{1SMLE})=\mathbf{0}. \tag{A.22}$$

If $\frac{1}{n}\sum_{t=1}^{n}\hat{\mathbb{H}}(\mathbf{u}_t,\tilde{\boldsymbol{\theta}})$ is nonsingular, (A.21) and (A.22) together yield

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{1SMLE}-\boldsymbol{\theta}^*)=-\left[\frac{1}{n}\sum_{t=1}^{n}\hat{\mathbb{H}}(\mathbf{u}_t,\tilde{\boldsymbol{\theta}})\right]^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{\mathbf{s}}(\mathbf{u}_t,\boldsymbol{\theta}^*). \tag{A.23}$$

The Central Limit Theorem implies that

$$\sqrt{n}\mathbb{Q}^{-1/2}\left\{\frac{1}{n}\sum_{t=1}^{n}\hat{\mathbb{S}}(\mathbf{u}_t,\boldsymbol{\theta}^*)-\mathbb{E}\left[\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta}^*)\right]\right\}\overset{Dist}{\rightarrow}N(\mathbf{0},\mathbf{I}_d), \tag{A.24}$$

where $\mathbb{Q}^{-1/2}$ is the lower-triangular Cholesky decomposition of the covariance of $\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta}^*)$. The true parameter $\boldsymbol{\theta}^*$ maximizes $\hat{l}_n(\boldsymbol{\theta}|\mathbf{u})$ by definition. Hence, it is reasonable to assume that the expectation of the score evaluated at the true parameter equals $\mathbf{0}$, that is $\mathbb{E}\left[\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta}^*)\right]=\mathbf{0}$. Let us also assume that $\frac{1}{n}\sum_{t=1}^{n}\hat{\mathbb{H}}(\mathbf{u}_t,\tilde{\boldsymbol{\theta}})\overset{p}{\rightarrow}-\mathbb{J}$ ($\mathbb{J}=-\mathbb{E}\left[\hat{\mathbb{H}}(\mathbf{u},\boldsymbol{\theta}^*)\right]$) in the neighborhood of $\boldsymbol{\theta}^*$. Combining (A.24) and (A.23), by Slutsky's theorem we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{1MSMLE}-\boldsymbol{\theta}^*)\overset{Dist}{\rightarrow}N(\mathbf{0},\mathbb{J}^{-1}\mathbb{Q}\mathbb{J}^{-1})$$

as in (3.17).

A.5 Proof of Proposition 12

Continued from Theorem 11, we need the following theorem to prove the efficiency of the one-step SMLE.

Theorem 14. *Suppose that $\{g_n(x)\}$ is a sequence of functions, which are differentiable on $[a,b]$ within the domain of x and such that $\{g_n(x)\}$ converges for some point x_0 on $[a,b]$. If $\{g_n'(x)\}$ converges uniformly on $[a,b]$, then $\{g_n(x)\}$ converges uniformly on $[a,b]$ to a function $g(x)$, and $\{g_n'(x)\}$ converges uniformly on $[a,b]$ to $g'(x)$. (See Theorem 7.17 and the following proof in Rudin (1976))*

The one-step SMLE $\hat{\boldsymbol{\theta}}_{1SMLE}$ is asymptotically efficient, if its asymptotic covariance achieves the inverse of the Fisher Information, which is the lower bound of covariance of estimates (Cramér 1946).

Let us now assume that $f(\mathbf{u};\boldsymbol{\theta})$ is twice continuously differentiable. The first order derivative of $\ln f(\mathbf{u};\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ as the score that

$$\mathbb{S}(\mathbf{u},\boldsymbol{\theta}) = \frac{\partial \ln f(\mathbf{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{A.25}$$

It follows that

$$\begin{aligned}\mathbb{E}\left[\mathbb{S}(\mathbf{u},\boldsymbol{\theta})\right] &= \int_{\mathbb{R}^d} \frac{\partial \ln f(\mathbf{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} f(\mathbf{u};\boldsymbol{\theta})\, d\mathbf{u} \\ &= \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbb{R}^d} f(\mathbf{u};\boldsymbol{\theta})\, d\mathbf{u} = \mathbf{0},\end{aligned}$$

and so $\mathbb{E}\left[\mathbb{S}(\mathbf{u},\boldsymbol{\theta}^*)\right] = \mathbf{0}$ when $\boldsymbol{\theta}=\boldsymbol{\theta}^*$. The second order derivative of $\ln f(\mathbf{u};\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is that

$$\mathbb{H}(\mathbf{u},\boldsymbol{\theta}) = \frac{\partial \mathbf{s}(\mathbf{u},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathbf{T}}} = \frac{\partial^2 \ln f(\mathbf{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}. \tag{A.26}$$

The Fisher Information is given by

$$\mathbb{I}(\boldsymbol{\theta}^*) = \mathbb{E}\left[\mathbb{S}(\mathbf{u},\boldsymbol{\theta}^*)\mathbb{S}(\mathbf{u},\boldsymbol{\theta}^*)^T\right] = -\mathbb{E}\left[\mathbb{H}(\mathbf{u},\boldsymbol{\theta}^*)\right]. \tag{A.27}$$

We first show that $\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta}) \xrightarrow{p} \mathbb{S}(\mathbf{u},\boldsymbol{\theta})$ uniformly. Suppose that the conditions in Proposition 9 are satisfied. We then have

$$\sup \left| \frac{\hat{f}_n(\mathbf{u};\boldsymbol{\theta})}{f(\mathbf{u};\boldsymbol{\theta})} - 1 \right| \xrightarrow{p} 0. \tag{A.28}$$

The Taylor expansion of $\left[f\left(\mathbf{u};\boldsymbol{\theta}\right)/\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)\right]$ around $\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right) = f\left(\mathbf{u};\boldsymbol{\theta}\right)$ is

$$\frac{f\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)} = 1 + \sum\nolimits_{k=1}^{\infty}(-1)^k\frac{1}{(k-1)!}\frac{f\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)}\left[\frac{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right) - f\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)}\right]^k. \quad \text{(A.29)}$$

Multiplying both hand-sides of (A.29) by $\left[\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)/f\left(\mathbf{u};\boldsymbol{\theta}\right)\right]$ and rearranging, we have

$$\frac{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)}{f\left(\mathbf{u};\boldsymbol{\theta}\right)} - 1 = O\left(\frac{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right) - f\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)}\right), \quad \text{(A.30)}$$

if we assume that $\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)$ is close to $f\left(\mathbf{u};\boldsymbol{\theta}\right)$ such that

$$\left|\frac{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right) - f\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)}\right| < 1.$$

Comparing (A.28) and (A.30), it follows that

$$\sup\left|\frac{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right) - f\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)}\right| \xrightarrow{p} 0,$$

that is,

$$\frac{f\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)} \xrightarrow{p} 1 \text{ uniformly.}$$

Similarily,

$$\ln\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right) \xrightarrow{p} \ln f\left(\mathbf{u};\boldsymbol{\theta}\right) \text{ uniformly}$$

when

$$\frac{f\left(\mathbf{u};\boldsymbol{\theta}\right)}{\hat{f}_n\left(\mathbf{u};\boldsymbol{\theta}\right)} \xrightarrow{p} 1 \text{ uniformly.}$$

According to Theorem 14, for

$$\frac{\partial \ln \hat{f}_n(\mathbf{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \xrightarrow{p} \frac{\partial \ln f(\mathbf{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \text{ uniformly,} \tag{A.31}$$

we require an addition assumption that $\partial K(\boldsymbol{\omega})/\partial \boldsymbol{\theta}$ converges uniformly on Θ. The uniform convergence in probability of $\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta})$ to $\mathbb{S}(\mathbf{u},\boldsymbol{\theta})$ is the direct result of (A.31), which also yields that

$$\mathbb{E}\left[\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta})\right] \xrightarrow{p} \mathbb{E}\left[\mathbb{S}(\mathbf{u},\boldsymbol{\theta})\right] = \mathbf{0}, \tag{A.32}$$

and

$$Cov\left[\hat{\mathbb{S}}(\mathbf{u},\boldsymbol{\theta})\right] \xrightarrow{p} Cov\left[\mathbb{S}(\mathbf{u},\boldsymbol{\theta})\right] = \mathbb{I}(\boldsymbol{\theta}^*). \tag{A.33}$$

The convergence of the semiparametric Hessian that

$$\hat{\mathbb{H}}(\mathbf{u},\boldsymbol{\theta}) \xrightarrow{p} \mathbb{H}(\mathbf{u},\boldsymbol{\theta}) \text{ uniformly}$$

also holds if both $\partial \ln K(\boldsymbol{\omega})/\partial \boldsymbol{\theta}$ and $\partial^2 \ln K(\boldsymbol{\omega})/\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T$ converge uniformly on the parameter space Θ. The uniform convergence in probability of $\hat{\mathbb{H}}(\mathbf{u},\boldsymbol{\theta})$ to $\mathbb{H}(\mathbf{u},\boldsymbol{\theta})$ yields that

$$\mathbb{E}\left[\hat{\mathbb{H}}(\mathbf{u},\boldsymbol{\theta}^*)\right] \xrightarrow{p} \mathbb{E}\left[\mathbb{H}(\mathbf{u},\boldsymbol{\theta}^*)\right] = -\mathbb{I}(\boldsymbol{\theta}^*). \tag{A.34}$$

Combining conditions (A.33) and (A.34), by Slutsky's theorem, (3.17) becomes

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{1MSMLE} - \boldsymbol{\theta}^*) \xrightarrow{Dist} N(\mathbf{0}, \mathbb{I}^{-1}(\boldsymbol{\theta}^*))$$

as in (3.18).

A.6 Proof of Lemma 3

Recall the definitions of notations in Lemma 3. According to Lemma 2, (A.14) can be written as

$$\sup_{\gamma(\cdot)\in\Gamma} \mathbb{E}[g(x,\gamma(x))] = \mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta} g(x,\boldsymbol{\theta})]. \tag{A.35}$$

Let

$$\varsigma(x) := \sup_{\theta\in\Theta} g(x,\boldsymbol{\theta}),$$

and $\{\boldsymbol{\theta}_i\}_{i\geq 1}$ be a countable dense set in Θ. Then

$$\varsigma(x) = \sup_{i\geq 1} g(x,\boldsymbol{\theta}_i).$$

Let $\gamma_0 \in \Gamma$ such that $\varsigma(x)$ is bounded below by $g(x,\gamma_0(x))$. It follows that $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta} g(x,\boldsymbol{\theta})]$ is well defined.

It follows that (A.35) always holds if the expectations are $+\infty$. We may assume the expectations on both hand-sides of (A.35) are finite.

Since $g(x,\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ and the set $\{\boldsymbol{\theta}_i\}_{i\geq 1}$ is dense in Θ, for any $\epsilon \in \mathbb{R}^+$ and any $x \in \Omega$, there exists a number $j(x)$ such that

$$g(x,\boldsymbol{\theta}_{j(x)}) > \varsigma(x) - \epsilon,$$

for all $i \geq j$.

We now show the function $x \mapsto \boldsymbol{\theta}_{j(x)}$ is measurable, that is, we show

$$W := \{x \in \Omega : \boldsymbol{\theta}_{j(x)} \in V\}$$

is measurable for any open set $V \subseteq \Theta$.

Let $\{\boldsymbol{\theta}_{j_k}\}_{k\geq 1}$ be a subsequence in $\{\boldsymbol{\theta}_j\}$, then

$$W = \bigcup_{k\geq 1} \{x \in \Omega : \boldsymbol{\theta}_{j_k} = \boldsymbol{\theta}_{j(x)}\},$$

where

$$\begin{aligned}&\{x \in \Omega : \boldsymbol{\theta}_{j_k} = \boldsymbol{\theta}_{j(x)}\} \\ &\quad = \bigcap_{l=1}^{k-1} \{x \in \Omega : g(x, \boldsymbol{\theta}_{j_l}) \leq \varsigma(x) - \epsilon\} \bigcap \{x \in \Omega : g(x, \boldsymbol{\theta}_{j_k}) > \varsigma(x) - \epsilon\}.\end{aligned}$$

This shows that W is a countable union of intersections of measurable sets and hence is measurable.

For an arbitrary $m \in \mathbb{R}^+$, let

$$\Omega_m := \{x \in \Omega : ||\boldsymbol{\theta}_{j(x)}|| < m\}$$

and

$$\gamma_m(x) := \begin{cases} \boldsymbol{\theta}_{j(x)} & : x \in \Omega_m \\ \gamma_0(x) & : x \in \Omega_m^c \end{cases}.$$

Thus, $\gamma_m \in \Gamma$, where Γ denote the set of all measurable functions mapping Ω to Θ.

We now have

$$
\begin{aligned}
\sup_{\gamma\in\Gamma} \mathbb{E}[g(x,\gamma(x))] &\geq \mathbb{E}[g(x,\gamma_m(x))] \\
&= \mathbb{E}\left[g(x,\boldsymbol{\theta}_{j(x)})\mathbb{1}_{\Omega_m}(x)\right] + \mathbb{E}\left[g(x,\gamma_0(u))\mathbb{1}_{\Omega_m^c}(x)\right] \\
&\geq \mathbb{E}[(\varsigma(x)-\epsilon)\mathbb{1}_{\Omega_m}(x)] + \mathbb{E}\left[g(x,\gamma_0(u))\mathbb{1}_{\Omega_m^c}(x)\right].
\end{aligned}
$$

where $\mathbb{1}_{\Omega_m}(x) = 1$ if $x \in \Omega_m$ and 0 otherwise. By the monotone convergence theorem, we have

$$
\mathbb{E}[(\varsigma(x)-\epsilon)\mathbb{1}_{\Omega_m}(x)] + \mathbb{E}[g(x,\gamma_0(x))\mathbb{1}_{\Omega_m^c}(x)] \to \mathbb{E}[(\varsigma(x)] - \epsilon
$$

when $m \to +\infty$. We note that $\epsilon \in \mathbb{R}^+$ is arbitrary, so

$$
\sup_{\gamma\in\Gamma} \mathbb{E}[g(x,\gamma(x))] \geq \mathbb{E}[(\varsigma(x)] = \mathbb{E}[\sup_{\theta\in\Theta} g(x,\boldsymbol{\theta})].
$$

Since $\sup_{\gamma\in\Gamma}\mathbb{E}[g(x,\gamma(x))]$ is the least upper bound of $\mathbb{E}[g(x,\gamma(x))]$ and it is not less than $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta} g(x,\boldsymbol{\theta})]$, the above becomes

$$
\sup_{\gamma\in\Gamma} \mathbb{E}[g(x,\gamma(x))] = \mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta} g(x,\boldsymbol{\theta})],
$$

that is

$$
\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}[g(x,\boldsymbol{\theta})] = \mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta} g(x,\boldsymbol{\theta})].
$$

A.7 Detailed Working for Page 63 to 64

The estimated (or empirical) density of the standardized errors $\mathbf{e}$ is given by

$$\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) = n^{-1}|\mathbf{H}|^{-1}\sum_{s=1}^{n} K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right),$$

Consider a Gaussian kernel in (3.19), given by

$$K\left(\mathbf{H}^{-1}(\mathbf{e}-\mathbf{e}_s)\right) = (2\pi)^{-d/2}\,|\mathbf{H}|^{-1}\exp\left(-\frac{1}{2}(\mathbf{e}-\mathbf{e}_s)^T\mathbf{H}^{-2}(\mathbf{e}-\mathbf{e}_s)\right).$$

We require $\mathbf{e}_t$ to be i.i.d. with a mean of $\mathbf{0}$ and a covariance matrix of $\mathbf{I}_d$. Scott's rule of thumb in $\mathbb{R}^d$ (Equation (6.42) in Scott (1992)) suggests that an appropriate choice of $\mathbf{H}$ for a Gaussian kernel is

$$\tilde{\mathbf{H}} = n^{-1/(d+4)}\mathbf{I}_d.$$

Substituting (3.19) and (3.21) into (3.7), we have the estimated density as

$$\begin{aligned}
\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) &= n^{-1}|\tilde{\mathbf{H}}|^{-1}\sum_{s=1}^{n} K\left(\tilde{\mathbf{H}}^{-1}(\mathbf{e}-\mathbf{e}_s)\right) \\
&= n^{-1}|\tilde{\mathbf{H}}|^{-1}\sum_{s=1}^{n}(2\pi)^{-d/2}\left|\tilde{\mathbf{H}}\right|^{-1}\exp\left[-\frac{1}{2}(\mathbf{e}-\mathbf{e}_s)^T\tilde{\mathbf{H}}^{-2}(\mathbf{e}-\mathbf{e}_s)\right] \\
&= n^{-1}|\tilde{\mathbf{H}}|^{-2}(2\pi)^{-d/2}\sum_{s=1}^{n}\exp\left[-\frac{1}{2}(\mathbf{e}-\mathbf{e}_s)^T\tilde{\mathbf{H}}^{-2}(\mathbf{e}-\mathbf{e}_s)\right].
\end{aligned}$$

where

$$\left|\tilde{\mathbf{H}}\right| = \left|n^{-1/(d+4)}\mathbf{I}_d\right| = n^{-d(d+4)}.$$

Hence, the estimated density becomes

$$\begin{aligned}\hat{f}_n(\mathbf{e};\boldsymbol{\theta}) &= n^{-1}\left(n^{\frac{-d}{d+4}}\right)^{-2}(2\pi)^{-d/2}\sum_{s=1}^{n}\exp\left[-\frac{|\tilde{\mathbf{H}}|^{-\frac{2}{d}}}{2}(\mathbf{e}-\mathbf{e}_s)^T(\mathbf{e}-\mathbf{e}_s)\right]\\ &= n^{-\frac{d-4}{d+4}}(2\pi)^{-d/2}\sum_{s=1}^{n}\exp\left(-\frac{n^{\frac{2}{d+4}}}{2}(\mathbf{e}-\mathbf{e}_s)^T(\mathbf{e}-\mathbf{e}_s)\right),\end{aligned}$$

as in (3.22).

Substituting the above estimated density into (3.6), we have the semi-parametric log-likelihood as becomes

$$\begin{aligned}\hat{l}_n(\boldsymbol{\theta}|\mathbf{u}) &= \frac{1}{n}\sum_{t=1}^{n}\ln\left[|\Sigma_t|^{-1/2}\hat{f}_n(\mathbf{e};\boldsymbol{\theta}^*)\right]\\ &= \frac{1}{n}\sum_{t=1}^{n}\left\{-\frac{1}{2}\ln|\Sigma_t| + \ln\left[\hat{f}_n(\mathbf{e};\boldsymbol{\theta}^*)\right]\right\}\\ &= \frac{1}{n}\sum_{t=1}^{n}\left[-\frac{1}{2}\ln|\Sigma_t|\right] + \frac{1}{n}\sum_{t=1}^{n}\left[\begin{array}{c} -\frac{d-4}{d+4}\ln n - \frac{d}{2}\ln 2\pi + \\ \ln\sum_{s=1}^{n}\exp\left(-\frac{n^{\frac{2}{d+4}}}{2}(\mathbf{e}_t-\mathbf{e}_s)^T(\mathbf{e}_t-\mathbf{e}_s)\right)\end{array}\right]\\ &= \left\{\begin{array}{c} -\frac{4}{d+4}\ln n - \frac{d}{2}\ln(2\pi) \\ +\frac{1}{n}\sum_{t=1}^{n}\left[-\frac{1}{2}\ln|\Sigma_t| + \ln\sum_{s=1}^{n}\exp\left(-\frac{n^{2/(d+4)}}{2}(\mathbf{e}_t-\mathbf{e}_s)^T(\mathbf{e}_t-\mathbf{e}_s)\right)\right]\end{array}\right.\end{aligned}$$

as in (3.35).

Appendix B

Detailed Working for Chapter 4

B.1 Detailed Derivations for Section 4.3

B.1.1 The Next Period Wealth for Given Targeted Portfolio Return

The first two moments of next period wealth W_{t+1} are given by (4.5) and (4.6) as

$$\mathbb{E}(W_{t+1}) = \mathbb{E}(W_t) + K_{t+1} + \int \mu_t W_t p(W_t) dW_t,$$

and

$$\mathbb{E}(W_{t+1}^2) = \left[\begin{array}{c} \mathbb{E}(W_t^2) + 2K_{t+1}\mathbb{E}(W_{t+1}) - K_{t+1}^2 \\ + \int (2\mu_t + \sigma_t^2 + \mu_t^2) W_t^2 p(W_t) dW_t \end{array} \right].$$

Substitute the targeted portfolio return $\mu_t = \beta_t / W_t + \alpha_t$ into the above two moments of W_{t+1} we acquire

$$\mathbb{E}(W_{t+1}) = \mathbb{E}\left(W_t\right) + K_{t+1} + \int \left(\frac{\beta_t}{W_t} + \alpha_t \right) W_t p(W_t) dW_t$$

$$
\begin{aligned}
&= \mathbb{E}\left(W_t\right) + K_{t+1} + \int \beta_t p(W_t) dW_t + \alpha_t \int W_t p(W_t) dW_t \\
&= (1+\alpha_t)\mathbb{E}(W_t) + \beta_t + K_{t+1} \\
&= u_t \mathbb{E}(W_t) + b_t,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}(W_{t+1}^2) &= \left\{ \begin{array}{c} \mathbb{E}(W_t^2) + 2K_{t+1}\mathbb{E}(W_{t+1}) - K_{t+1}^2 \\ +\int (2\mu_t + \sigma_t^2 + \mu_t^2) W_t^2 p(W_t) dW_t \end{array} \right\} \\
&= \left\{ \begin{array}{c} \mathbb{E}(W_t^2) + 2K_{t+1}\left[(1+\alpha_t)\mathbb{E}(W_t) + \beta_t + K_{t+1}\right] - K_{t+1}^2 \\ +\int \left[\begin{array}{c} 2(1+\psi_t)\left(\frac{\beta_t}{W_t} + \alpha_t\right) \\ +(1+\varphi_t)\left(\frac{\beta_t}{W_t} + \alpha_t\right)^2 + \chi_t \end{array} \right] W_t^2 p(W_t) dW_t \end{array} \right\} \\
&= \left\{ \begin{array}{c} \mathbb{E}(W_t^2) + 2(1+\alpha_t)K_{t+1}\mathbb{E}(W_t) + 2\beta_t K_{t+1} + K_{t+1}^2 \\ +2(1+\psi_t)\left[\alpha_t \mathbb{E}(W_t^2) + \beta_t \mathbb{E}(W_t)\right] \\ +(1+\varphi_t)\left[\beta_t^2 + 2\alpha_t\beta_t\mathbb{E}(W_t) + \alpha_t^2\mathbb{E}(W_t^2)\right] + \chi_t \mathbb{E}(W_t^2) \end{array} \right\} \\
&= \left\{ \begin{array}{c} \left[(1+\alpha_t)^2 + 2\psi_t\alpha_t + \varphi_t\alpha_t^2 + \chi_t\right]\mathbb{E}(W_t^2) \\ +\left[2(1+\alpha_t)(K_{t+1}+\beta_t) + 2\beta_t(\psi_t + \alpha_t\varphi_t)\right]\mathbb{E}(W_t) \\ +(\beta_t + K_{t+1})^2 + \varphi_t\beta_t^2 \end{array} \right\} \\
&= \left\{ \begin{array}{c} \left[1 + \chi_t - \frac{(1+\psi_t)^2}{1+\varphi_t}\right]\mathbb{E}(W_t^2) + 2(1+\alpha_t)K_{t+1}\mathbb{E}(W_t) \\ +(\beta_t + K_{t+1})^2 + \varphi_t\beta_t^2 \end{array} \right\} \\
&= v_t\mathbb{E}(W_t^2) + p_t\mathbb{E}(W_t) + q_t,
\end{aligned}
$$

where the constants at time t are:

$$
u_t = 1 + \alpha_t,
$$

$$\begin{aligned}
b_t &= \beta_t + K_{t+1}, \\
v_t &= 1 + \chi_t - \frac{(1+\psi_t)^2}{1+\varphi_t}, \\
p_t &= 2(1+\alpha_t)K_{t+1}, \text{ and} \\
q_t &= (\beta_t + K_{t+1})^2 + \varphi_t \beta_t^2.
\end{aligned}$$

as in (4.13) and (4.14).

B.1.2 The Lagrangian for Minimizing Variance of Terminal Wealth

The approximated (or estimated) Lagrangian, evaluated at time t for minimizing the second of terminal wealth $\hat{\mathbb{E}}_t(W_N^2)$ (given in (4.19)) for a given level of first moment of terminal wealth $\hat{\mathbb{E}}_t(W_N)$ (given in (4.18)) is

$$\hat{\mathcal{L}}_t = \left\{ \begin{array}{c} \sum_{i=0}^{N-t-1} \left[\left(p_{t+i}\hat{\mathbb{E}}_t(W_{t+i}) + q_{t+i} \right) v_t^{N-t-1-i} \right] + \mathbb{E}_t(W_t^2) v_t^{N-t} \\ -2\lambda_t^{(N)} \left[\sum_{i=0}^{N-t-1} b_{t+i} u_t^{N-t-1-i} + \mathbb{E}_t(W_t) u_t^{N-t} \right] \end{array} \right\}$$

as in (4.20).

Consider a small variation in β_t, $\Delta\beta_t$. We have

$$\begin{aligned}
\Delta b_t &= \Delta\beta_t, \\
\Delta q_t &= 2\left[(1+\varphi_t)\beta_t + K_{t+1}\right]\Delta\beta_t, \text{ and} \\
\Delta\hat{\mathbb{E}}_t(W_{t+i}) &= \sum_{s=0}^{i-1} \Delta\beta_{t+s} u_t^{i-1-s}.
\end{aligned}$$

Hence, the first order variation in the approximated Lagrangian due to the small variation $\Delta\beta_t$ is

$$\Delta\hat{\mathcal{L}}_t = \left\{ \begin{array}{c} \sum_{i=0}^{N-t-1} v_t^{N-t-1-i}\left[\left(p_{t+i}\Delta\hat{\mathbb{E}}_t(W_{t+i}) + \Delta q_{t+i}\right)\right] \\ -2\lambda_t^{(N)} \sum\limits_{i=0}^{N-t-1} u_t^{T-t-1-i}\Delta\beta_{t+i} \end{array} \right\}$$

$$= \left\{ \begin{array}{c} \sum\limits_{i=1}^{N-t-1} v_t^{N-t-1-i} p_{t+i}\left(\sum\limits_{s=0}^{i-1} u_t^{i-1-s}\Delta\beta_{t+s}\right) \\ +2\sum\limits_{i=0}^{N-t-1} v_t^{N-t-1-i}\left[(1+\varphi_t)\beta_{t+i} + K_{t+i+1}\right]\Delta\beta_{t+i} \\ -2\lambda_t^{(N)}\sum_{i=0}^{N-t-1}\Delta\beta_{t+i}u_t^{N-t-1-i} \end{array} \right\}$$

$$= \left\{ \begin{array}{c} \sum\limits_{i=0}^{N-t-2}\Delta\beta_{t+i}\sum\limits_{s=i+1}^{N-t-1} p_{t+s}u_t^{s-1-i}v_t^{N-t-1-s} \\ +2\sum\limits_{i=0}^{N-t-1} v_t^{N-t-1-i}\left[(1+\varphi_t)\beta_{t+i} + K_{t+i+1}\right]\Delta\beta_{t+i} \\ -2\lambda_t^{(N)}\sum_{i=0}^{N-t-1}\Delta\beta_{t+i}u_t^{N-t-1-i} \end{array} \right\}$$

as in (4.21).

We require

$$\left\{ \begin{array}{c} \sum\limits_{s=i+1}^{N-t-1} p_{t+s}u_t^{s-1-i}v_t^{N-t-1-s} \\ +2v_t^{N-t-1-i}\left[(1+\varphi_t)\beta_{t+i} + K_{t+i+1}\right] - 2\lambda_t^{(N)}u_t^{N-t-1-i} \end{array} \right\} = 0$$

for the Lagrangian to be at a minimum, that it is insensitive to $\Delta\beta_{t+i}$. Rearrange this condition, we have

$$(1+\varphi_t)\beta_{t+i} = \left[2v_t^{N-t-1-i}\right]^{-1}\left\{ \begin{array}{c} 2\lambda_t u_t^{N-t-1-i} \\ -\sum\limits_{s=i+1}^{N-i-1} p_{i+s}u_i^{s-1-t}v_i^{N-i-1-s} \end{array} \right\} - K_{t+i+1}$$

$$= \lambda_t^{(N)} \left(\frac{u_t}{v_t}\right)^{N-t-1-i} - \sum_{s=i+1}^{N-t-1} u_t K_{t+s+1} u_t^{s-1-i} v_t^{N-t-1-s-(N-t-1-i)} - K_{t+i+1}$$

$$= \lambda_t^{(N)} \left(\frac{u_t}{v_t}\right)^{N-t-1-i} - \sum_{s=i}^{N-t-1} K_{t+s+1} \left(\frac{u_t}{v_t}\right)^{s-i}, \text{ for } t < T-1.$$

as in (4.22).

$\lambda_t^{(N)}$ can be retrieved from $\hat{\mathbb{E}}_t(W_N)$. Substituting $\beta_{t+i}\ (\lambda_t^{(N)})$ into $\hat{\mathbb{E}}_t(W_N)$ in (4.18), we have

$$\begin{aligned}
\hat{\mathbb{E}}_t(W_N) &= \sum_{i=0}^{N-t-1} b_{t+i} u_t^{N-t-1-i} + \mathbb{E}_t(W_t) u_t^{N-t} \\
&= \sum_{i=0}^{N-t-1} \left(\beta_{t+i} u_t^{N-t-1-i}\right) + \sum_{i=0}^{N-t-1} \left(K_{t+i+1} u_t^{N-t-1-i}\right) + \mathbb{E}_t(W_t) u_t^{N-t} \\
&= \left\{ \begin{array}{c} \frac{\lambda_t^{(N)}}{1+\varphi_t} \sum\limits_{i=0}^{N-t-1} \left(\frac{u_t^2}{v_t}\right)^{N-t-1-i} + \sum_{i=0}^{N-t-1} \left(K_{t+i+1} u_t^{N-t-1-i}\right) + \mathbb{E}_t(W_t) u_t^{N-t} \\ -\frac{1}{1+\varphi_t} \sum_{i=0}^{N-t-1} \sum_{s=i}^{N-t-1} K_{t+s+1} \left(\frac{u_t}{v_t}\right)^{s-i} u_t^{N-t-1-i} \end{array} \right\} \\
&= \left\{ \begin{array}{c} \frac{\lambda_t^{(N)}}{1+\varphi_t} \sum\limits_{i=1}^{N-t} \left(\frac{u_{ti}^2}{v_t}\right)^{N-t-i} + \sum_{i=1}^{N-t} \left(K_{t+i} u_t^{N-t-i}\right) + \mathbb{E}_t(W_t) u_t^{N-t} \\ -\frac{1}{1+\varphi_t} \sum_{i=1}^{N-t} u_t^{N-t-i} \sum_{s=i}^{N-t} K_{t+s} \left(\frac{u_t}{v_t}\right)^{s-i} \end{array} \right\} \\
&= \left\{ \begin{array}{c} \frac{\lambda_t^{(N)}}{1+\varphi_t} \left[\left(\frac{u_t^2}{v_{ti}}\right)^{N-t} - 1\right] \left(\frac{u_t^2}{v_t} - 1\right)^{-1} + \sum_{i=1}^{N-t} \left(K_{t+i} u_t^{N-t-i}\right) + \mathbb{E}_t(W_t) u_t^{N-t} \\ -\frac{1}{1+\varphi_t} \sum_{i=1}^{N-t} K_{t+i} \left(\frac{u_t}{v_t}\right)^{i} \sum_{s=1}^{i} u_t^{N-t-s} \left(\frac{u_t}{v_t}\right)^{-s} \end{array} \right\} \\
&= \left\{ \begin{array}{c} \frac{\lambda_t^{(N)}}{1+\varphi_t} \left[\left(\frac{u_t^2}{v_t}\right)^{N-t} - 1\right] \left(\frac{u_t^2}{v_t} - 1\right)^{-1} + \sum_{i=1}^{N-t} \left(K_{t+i} u_t^{N-t-i}\right) + \mathbb{E}_t(W_t) u_t^{N-t} \\ -\frac{1}{1+\varphi_t} \sum_{i=1}^{N-t} K_{t+i} \left(\frac{u_t}{v_t}\right)^{i} \sum_{s=1}^{i} u_t^{N-t} \left(\frac{u_t^2}{v_t}\right)^{-s} \end{array} \right\} \\
&= \left\{ \begin{array}{c} \frac{\lambda_t^{(N)}}{1+\varphi_t} \left[\left(\frac{u_t^2}{v_t}\right)^{N-t} - 1\right] \left(\frac{u_t^2}{v_t} - 1\right)^{-1} + \sum_{i=1}^{N-t} \left(K_{t+i} u_t^{N-t-i}\right) + \mathbb{E}_t(W_t) u_t^{N-t} \\ -\frac{1}{1+\varphi_t} \sum_{i=1}^{N-t} K_{t+i} u_t^{N-t} v_t \left[1 - \left(\frac{u_t^2}{v_t}\right)^{-i}\right] \left(u_t^2 - v_t\right)^{-1} \end{array} \right\}
\end{aligned}$$

$$
= \left\{ \begin{array}{c} \frac{\lambda_t^{(N)}}{1+\varphi_t} \left[\left(\frac{u_t^2}{v_t} \right)^{N-t} - 1 \right] \left(\frac{u_t^2}{v_t} - 1 \right)^{-1} + \sum_{t=1}^{N-i} \left(K_{t+i} u_t^{N-t-i} \right) + \mathbb{E}_t(W_t) u_t^{N-t} \\ - \frac{1}{1+\varphi_t} \sum_{i=1}^{N-t} K_{t+i} u_t^{N-t} v_t \left[\left(\frac{u_t}{v_t} \right)^t - u_t^{-i} \right] (u_t^2 - v_t)^{-1} \end{array} \right\}
$$

$$
= \left\{ \begin{array}{c} \frac{\lambda_t^{(N)}}{1+\varphi_i} \left[\left(\frac{u_t^2}{v_t} \right)^{N-t} - 1 \right] \left(\frac{u_t^2}{v_t} - 1 \right)^{-1} + \mathbb{E}_t(W_t) u_t^{N-t} \\ + u_t^{N-t} \sum_{i=1}^{N-t} K_{t+i} \left(u_t^{-t} - \frac{v_t}{1+\varphi_t} \frac{1}{u_t^2 - v_t} \left[\left(\frac{u_t}{v_t} \right)^i - u_t^{-i} \right] \right) \end{array} \right\}
$$

$$
= \left\{ \begin{array}{c} \frac{\lambda_t^{(N)}}{1+\varphi_t} \left[\left(\frac{u_t^2}{v_t} \right)^{N-t} - 1 \right] \left(\frac{u_t^2}{v_t} - 1 \right)^{-1} + \mathbb{E}_t(W_t) u_t^{N-t} \\ + u_t^{N-t} \left(1 + \frac{1}{1+\varphi_t} \frac{v_t}{u_t^2 - v_t} \right) \sum_{i=1}^{N-t} K_{t+i} u_t^{-i} - \frac{u_t^{N-t}}{1+\varphi_t} \frac{v_t}{u_t^2 - v_t} \sum_{i=1}^{N-t} \left(\frac{u_t}{v_t} \right)^i K_{t+i} \end{array} \right\}.
$$

Rearranging the above equation, we acquire

$$
\lambda_t^{(N)} = \frac{(1+\varphi_t)(\frac{u_t^2}{v_t} - 1)}{\left(\frac{u_t^2}{v_t} \right)^{N-t} - 1} \left[\begin{array}{c} \hat{\mathbb{E}}_t(W_N) - \mathbb{E}(W_t) u_t^{N-t} - \frac{u_t^{T-t}}{1+\varphi_t} \frac{v_t}{u_t^2 - v_t} \sum_{i=1}^{N-t} \left(\frac{u_t}{v_t} \right)^i K_{t+i} \\ + u_t^{N-t} \left(1 + \frac{1}{1+\varphi_t} \frac{v_{ti}}{u_t^2 - v_t} \right) \sum_{i=1}^{N-t} K_{t+i} u_t^{-i} \end{array} \right]
$$

as in (4.23). Note that in the derivation of $\hat{\mathbb{E}}_t(W_N)$ in terms of $\lambda_t^{(N)}$, we have used the fact that

$$
\sum_{s=1}^{t} a^{-s} = \frac{1 - a^{-t}}{a - 1}.
$$

B.2 Some Detailed Derivations for Section 4.4

We adopt a VAR(1)-MGARCH(1,1) return model outlined in (4.24) as

$$
\begin{aligned}
&\mathbf{r}_t = \boldsymbol{\gamma} + \mathbf{C}\mathbf{r}_{t-1} + \mathbf{v}_t, \\
&\mathbf{v}_t = \boldsymbol{\Sigma}_t^{1/2} \mathbf{e}_t, \\
&\boldsymbol{\Sigma}_t = \mathbf{A}_0^T \mathbf{A}_0 + \mathbf{A}_1^T \mathbf{v}_{t-1} \mathbf{v}_{t-1}^T \mathbf{A}_1 + \mathbf{B}_1^T \boldsymbol{\Sigma}_{t-1} \mathbf{B}_1,
\end{aligned}
$$

where $\mathbf{r}_t$ is a vector of asset returns at time t, $\mathbf{e}_t$ are independently identically distributed (i.i.d.) standard normal innovations and Σ_t is the covariance matrix of returns.

Substituting the return process into the recursive relation of portfolio wealth in (4.1),the first two moments of next period wealth are

$$\begin{aligned}
\mathbb{E}(W_{t+1}) &= \mathbb{E}\left[(1+\mathbf{w}_t^T\mathbf{r}_t)W_t\right] + K_{t+1} \\
&= \int\int \left[1+\mathbf{w}_t^T\left(\gamma+\mathbf{C}\mathbf{r}_{t-1}\right)\right] W_t p(\mathbf{r}_{t-1},W_t)dW_t d\mathbf{r}_{t-1} + K_{t+1} \\
&= \mathbb{E}(W_t) + \int\int \mathbf{w}_t^T\left(\gamma+\mathbf{C}\mathbf{r}_{t-1}\right) W_t p(\mathbf{r}_{t-1},W_t)dW_t d\mathbf{r}_{t-1} + K_{t+1},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}(W_{t+1}^2) &= \mathbb{E}\left\{\left[(1+\mathbf{w}_t^T\mathbf{r}_t)W_t + K_{t+1}\right]^2\right\} \\
&= \mathbb{E}\left[(1+\mathbf{w}_t^T\mathbf{r}_t)^2W_t^2 + 2(1+\mathbf{w}_t^T\mathbf{r}_t)K_{t+1}W_t + K_{t+1}^2\right] \\
&= \mathbb{E}\left[(1+2\mathbf{w}_t^T\mathbf{r}_t + \left(\mathbf{w}_t^T\mathbf{r}_t\right)^2)W_t^2\right] + \mathbb{E}\left[2(1+\mathbf{w}_t^T\mathbf{r}_t)K_{t+1}W_t + K_{t+1}^2\right] \\
&= \mathbb{E}\left[(2\mathbf{w}_t^T\mathbf{r}_t + \left(\mathbf{w}_t^T\mathbf{r}_t\right)^2)W_t^2\right] + \mathbb{E}\left(W_t^2\right) + 2K_{t+1}\mathbb{E}(W_{t+1}) - K_{t+1}^2 \\
&= \left\{\begin{array}{c}
\int\int\left[1+\mathbf{E}\left(\mathbf{w}_t^T\left(\gamma+\mathbf{C}\mathbf{r}_{t-1}+\mathbf{v}_t\right)\right)\right]^2 W_t^2 p(\mathbf{r}_{t-1},W_t)dW_t d\mathbf{r}_{t-1} \\
+2K_{t+1}\int\int[1+\mathbf{w}_t^T\left(\gamma+\mathbf{C}\mathbf{r}_{t-1}\right)]W_t p(\mathbf{r}_{t-1},W_t)dW_t d\mathbf{r}_{t-1} + K_{t+1}^2
\end{array}\right\} \\
&= \left\{\begin{array}{c}
\int\int\left[\begin{array}{c} 2\mathbf{w}_t^T\left(\gamma+\mathbf{C}\mathbf{r}_{t-1}\right) \\ +\left[\mathbf{w}_t^T\left(\gamma+\mathbf{C}\mathbf{r}_{t-1}\right)\right]^2 \\ +\mathbf{w}_t^T\mathbf{\Sigma}_t\mathbf{w}_t \end{array}\right] W_t^2 p(\mathbf{r}_{t-1},W_t)dW_t d\mathbf{r}_{t-1} \\
\mathbb{E}(W_t^2) + 2K_{t+1}\mathbb{E}(W_{t+1}) - K_{t+1}^2
\end{array}\right\},
\end{aligned}$$

as in (4.25) and (4.26).

Bibliography

Amin, G. and H. Kat (2003). Hedge fund performance 1990-2000: Do the "Money Machines" really add value? *Journal of Financial and Quantitative Analysis 38*(2), 251–274.

Anderson, K., C. Brooks, and A. Katsaris (2010). Speculative bubbles in the S&P 500: Was the tech bubble confined to the tech sector? *Journal of Empirical Finance 17*, 345–361.

Andreou, E. and E. Ghysels (2001). Rolling-sample volatility estimators: some new theoretical, simulation and empirical results. *Journal of Business and Economic Statistics 20*, 363–376.

Basu, A. and M. Drew (2009). Portfolio size effect in retirement accounts: what does it imply for lifecycle asset allocation funds?

Bellman, R. (1957). *Dynamic Programming.* U.S.A.: Princeton University Press.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy 81*(3), 637–654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics 69*, 542–47.

Bone, S. and A. Goddard (2009, April). Rebalancing revisited. In *the Institute of Actuaries of Australia 2009 Biennial Convention*, Sydney. Institute of Actuaries of Australia.

Brennan, M. J., E. S. Schwartz, and R. Lagnado (1997). Strategic asset allocation. *Journal of Economic Dynamics and Control 21*, 1377–1403.

Brown, P., D. Walsh, and A. Yuen (1997). The interaction between order imbalance and stock price. *Pacific-Basin Finance Journal 5*, 539–557.

Carr, P., H. Geman, and D. Madan (2002). The fine structure of asset returns: An empirical investigation. *Journal of Business 75*(2), 305–332.

Cecchetti, S., A. Flores-Lagunes, and S. Krause (2005, July). Assessing the sources of changes in the volatility of real growth. In *The Changing Nature of the Business Cycle*, Canberra. Reserve Bank of Australia.

Chan, K. and W.-M. Fong (2000). Trade size, order imbalance and the volatility-volume relation. *Journal of Financial Economics 57*, 247–273.

Chan, K., G. Karolyi, F. Longstaff, and A. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance 47*(3), 1209–1227.

Cheng, M. (2003). Pretrade cost analysis and management of implementation shortfall. In *AIMR Conference Proceedings*, Number 7, pp. 26–34.

Chiarella, C., C. Y. Hsaio, and W. Semmler (2007). Intertemporal investment strategies under inflation risk. Technical Report Working Paper, University of Technology Sydney.

Chordia, T., R. Roll, and A. Subrahmanyam (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics 65*, 111–130.

Chordia, T. and A. Subrahmanyam (2004). Order imbalance and individual stock returns: theory and evidence. *Journal of Financial Economics 72*(3), 458–518.

Cox, J. and H. Leland (2000). On dynamic investment strategies. *Journal of Economic Dynamics and Control 24*(11-12), 1859–880.

Cox, J. C. and S. A. Ross (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics 3*(1-2), 145–166.

Cramér, H. (1946). *Mathematical Methods of Statistics.* U.S.A.: Princeton University Press.

Cvitanic, J., V. Polimenis, and F. Zapatero (2008). Optimal portfolio allocation with higher moments. *Annals of Finance 4*, 1–28.

Das, S. and R. Uppal (2004). Systemic risk and international portfolio choice. *The Journal of Finance 59*(6), 2809–2834.

Di, J. and A. Gangopadhyay (2013). One-step semiparametric estimation of the GARCH model. *Journal of Financial Econometrics 12*(2), 382–407.

Diggle, J. and R. Brooks (2007). The target cash rate and its impact on investment asset returns in Australial. *Applied Financial Economics 17*(8), 615–633.

Dreyfus, S. E. (1965). *Dynamic Programming And The Calculus of Variations.* New York: Academic Press INC.

Dupačová, J. (1999). Portfolio optimization via stochastic programming: methods of output analysis. *Mathematical Methods of Operations Research 50*, 245–270.

Engle, R. and K. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 122–150.

Fama, E. and K. French (1986). Common factors in the serial correlation of stock returns. Technical Report CRSP Working Paper No. 200, Anderson Graduate School of Management, UCLA.

Fama, E. and J. MacBeth (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy 81*(3), 607–636.

Foster, D. and D. Nelson (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica 64*, 139–174.

Frain, J. and D. Gallo (2011, 12). System and method for scheduling asset allocation. Patent US 8086512B12.

Gelfand, I. and S. Fomin (1963). *Calculus of Variations.* N.J.: Prentice-Hall, inc.

Gibbs, J. (1902). *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics.* New York: Charles Scribner's Sons.

Gore, C. (2010). The global recession of 2009 in a long-term development perspective. *Journal of International Development 22*, 714–738.

Grinold, R. and R. Kahn (1999). *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk* (2 ed.)., Chapter 16, pp. 445–475. McGraw-Hill.

Hafner, C. and J. Rombouts (2007). Semiparametric multivariate volatility models. *Econometric Theory 23*(02), 251–280.

Hall, P., J. Marron, and B. Park (1992). Smoothed cross-validation. *Theory and Related Fields 92*, 1–20.

Hayashi, F. (2000). *Econometrics.* U.S.A.: Princeton University Press.

Hodges, S. and R. Clarkson (1994). Dynamic Asset Allocation: Insights from Theory. *Philosophical Transactions: Physical Sciences and Engineering 347*(1684), 587–598.

Huang, H. and Y. Lee (2010). Optimal asset allocation for a general portfolio of life insurance policies. *Insurance: Mathematics and Economics 46*(2), 271–280.

Ibe, O. (2009). *Markov processes for stochastic modeling.* Elsevier Academic Press.

International Monetary Fund, R. D. (2009). *April 2009: Crisis and Recovery.* International Monetary Fund, Publication Services.

International Monetary Fund, R. D. (2010). *April 2010: Rebalancing Growth.* International Monetary Fund, Publication Services.

Jennrich, R. (1969). Asymptotic properties of non-linear least-squares estimators. *The Annals of Mathematical Statistics 40*(2), 633–643.

Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk.* U.S.: McGraw-Hill.

Kac, M. (1949). On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society 65*(1), 1–13.

Kassberger, S. and T. Liebmann (2012). When are path-dependent payoffs suboptimal? *Journal of Banking and Finance 36*(6), 1304–1310.

Kissell, R. and M. Glantz (2003). *Optimal Trading Strategies: Quantitative Approaches for Managing Market Impact and Trading Risk.* New York: AMACOM.

Klebaner, F. (2005). *Introduction to Stochastic Calculus with Applications* (2nd ed.). London UK: Imperial College Press.

Leung, A. (2011). Reactive investment strategies. *Insurance: Mathematics and Economics 49*(1), 89–99.

Leung, A. and W. Shi (2013). An extension of some results due to Cox and Leland. *Journal of Mathematical Finance 3*(4), 416–425.

Liu, J., F. Longstaff, and J. Pan (2003). Dynamic asset allocation with event risk. *Journal of Finance 58*(1), 231–259.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 7*, 77–91.

McNelis, P. (1993). The response of Australian stock, foreign exchange and bond markets to foreigh asset returns and volatilities . *RBA Research Discussion Paper* (9301).

Merton, R. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory 3*, 373–413.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. *Review of Economics and Statistics 51*(3), 247–257.

Musumeci, J. and J. Musumeci (1999). A dynamic-programming approach to multiperiod asset allocation. *Journal of Financial Services Research 15*(1), 5–21.

Napper, K. (2008, 12). Refining risk measures: serial correlation. Technical report.

Park, B. and J. Marron (1992). Practical performance of several data driven bandwidth selectors. *Journal of the American Statistical Association 7*, 251–270.

Parrino, R., D. Kidwell, H. Au-Yong, M. Dempsey, N. Morkel-Kingsbury, S. Ekanayake, J. Kofoed, and M. J. (2014). *Fundamentals of Corporate Finance* (2nd ed.). Australia: John Wiley and Sons Australia, Ltd.

Parzen, E. (1962). In estimation of a probability density function and mode. *Annals of Mathematical Statistics 33*, 1065–1076.

Pesaran, M. and A. Timmermann (2000). A recursive modelling approach to predicting UK stock returns. *The Economic Journal 110*, 159–191.

Ross, S. (1978). Mutual fund separation in financial theory-the separating distributions. *Journal of Economic Theory 17*(2), 254–286.

Rudin, W. (1976). *Principles of Mathematical Analysis.* McGraw-Hill.

Salsa, S. (2008). *Partial Differential Equations in Action.* Springer.

Samuelson, P. (1969). Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics 51*, 239–46.

Scott, D. (1992). *Multivariate Density Estimation.* U.S.: John Wiley and Sons, Ltd.

Seth, J. (2002). Rules for rebalancing. *Financial Planning 32*, 89–93.

Sharpe, W. (1994). The Sharpe ratio. *Journal of Portfolio Management 21*, 49–58.

Shiller, R. (2008). *The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It.* Princeton University Press.

Taksar, M. and S. P. Sethi (1988). A note on Merton's "Optimum consumption and portfolio rules in a continuous-time model". *Journal of Economic Theory 46*, 395–401.

Tanaka, K. (2009). Dynamic Asset Allocation under Stochastic Interest Rate and Market Price of Risk. *Working paper*, 1–14.

Trippi, R. R. and R. B. Harriff (1991). Dynamic asset allocation rules: survey and synthesis. *The Journal of Portfolio Management 17*(4), 19–26.

Vanduffel, S., A. Chernih, M. Maj, and W. Schoutens (2009). A Note on the suboptimality of path-dependent pay-offs in Lévy markets. *Applied Mathematical Finance 16*(4), 315–330.

Walker, G. (1931). On periodicity in series of related terms. *Proceedings of the Royal Society of London 131*, 518–532.

Wand, M. and M. Jones (1995). *Kernel Smoothing*, Volume 4 of *Monographs on Statistics and Applied Probability*. London: Chapman & Hall.

Williams, D. (1991). *Probability with Martingales.* U.K.: Cambridge University Press.

Yu, L., X. Ji, and S. Wang (2003). Stochastic programming models in financial optimization: a survey. *Advanced Modeling and Optimization 5*(1), 1–26.

Yule, G. (1927). On a method of investigating periodicities in disturbed series, with special reference to WolferÃŋs sunspot numbers. *Philosophical Transactions of the Royal Society of London 226*, 267–298.