Topics on Forward Investment Theory

Pedro Miguel Almeida Serra Costa Vitória
Magdalen College
University of Oxford

a thesis submitted for the degree of
doctor of philosophy at the university of oxford
Hilary term 2015

Dedicated to my parents, Carlos and Filomena.

Acknowledgements

I would like to thank my supervisor Thaleia Zariphopoulou for all her advice, dedication and support. I am also thankful to Terry Lyons for agreeing to co–supervise my research and being most helpful and supportive. I would also like to thank Marek Musiela for numerous valuable discussions. Furthermore, I am grateful to René Carmona, Ben Hambly, Jan Obłój and Xunyu Zhou for providing helpful feedback on preliminary versions of this thesis.

I have had a great time during my stay in Oxford and I owe most of it to the people I met (certainly not to the weather). I thank my friends and colleagues at the Mathematical Institute and Oxford–Man Institute for the many stimulating discussions and support. Many thanks in particular to Bahman Angoshtari, Horatio Boedihardjo, Philippe Charmoy, Yoel Furman, Sigrid Källblad, Abby Levenberg, Wande McCunn, Wei Pan, Bary Pradelski, Johannes Ruf, Galen Sher and Rasmus Wissmann. I am also thankful to my friends at Magdalen who have contributed to such a great, fun and relaxing atmosphere in college. Most of all, I am grateful to my girlfriend Maria Mestre for always being here for me.

I am extremely indebted to my parents Carlos and Filomena, and to my sisters Joana and Carolina, for their unconditional love and support. I will never be able to thank them enough; I just hope to make them proud. I am also grateful to my uncles José Serra, who always fostered my interest in maths, and Vitor Serra, who was extremely supportive during my undergraduate studies.

Financial support from the Fundação para a Ciência e Tecnologia — doctoral grant SFRH / BD / 68331 / 2010 — is gratefully acknowledged. I am also thankful to the Oxford–Man Institute for financial support and for providing excellent working conditions, and to the University of Texas at Austin for hosting me in a number of occasions.

Topics on Forward Investment Theory

A thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2015

Pedro M. Almeida Serra Costa Vitória

Magdalen College

Abstract

In this thesis, we study three topics in optimal portfolio selection that are relevant to the theory of forward investment performance processes.

In Chapter 1, we develop a connection between the classical mean–variance optimisation and time–monotone forward performance processes for infinitesimal trading times. Namely, we consider consecutive mean–variance problems and we show that, for an appropriate choice of the corresponding mean–variance trade-off coefficients, the wealth process that is generated converges (as the trading interval goes to zero) to the optimal wealth process generated by a time–monotone forward performance process. The choice of the trade-off coefficients is made in accordance to the evolution of the risk tolerance process of the forward performance process. This result allows us to provide a fresh view on the issue of time–consistency of mean–variance analysis, for we propose a method to update mean–variance risk preferences forward in time. As a by-product, our convergence theorem generalises a result by Gyöngy (1998) on the convergence of the Euler scheme for SDEs. We also provide novel results on the Lipschitz regularity of the local risk tolerance function of forward investment performance processes. The material in this chapter is joint work with Marek Musiela and Thaleia Zariphopoulou.

Chapter 2 combines forward investment theory and partial information. Specifically, we construct forward investment performance processes in models where the drift is a random variable distributed according to a known distribution. The forward performance processes we consider are of the type

$$U(t,x) = u(t,x,R_t),$$

where $R.$ denotes the process of cumulative excess returns, and $u(t,x,z) : [0,\infty) \times \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}$ is such that $u(t,\cdot,z)$ is a utility function satisfying Inada's conditions. We derive the Hamilton–Jacobi–Bellman (HJB) equation for $u(\cdot)$. The HJB equation is linearised into the ill-posed heat equation; then, using the multidimensional version of Widder's theorem, we fully characterise the solutions to this equation in terms of a collection of positive measures; the result is an integral

representation of the convex conjugate function of $u(t, \cdot, z)$. We construct several examples, and we show how these can be combined, in the dual domain, to generate mixtures of forward investment performance processes. We also show that the volatility of these processes is intrinsic, in that it is not generated by changes of numéraire/measure.

In Chapter 3, we provide an extension of the Black–Litterman model to the continuous time setting. Our extension is different from, and complements that of, Frey, Gabih, and Wunderlich (2012) and Davis and Lleo (2013). Specifically, we develop a novel robust estimator of instantaneous expected returns which is continuously shrunk towards the predictions of an asset pricing theory, such as the CAPM. We derive this estimator fairly explicitly and study some of its properties. As in the Black–Litterman model, such an estimator can be used to make optimal asset allocation problems in continuous time more robust with respect to estimation errors. We provide explicit solutions to the problem of maximising expected power utility of terminal wealth, when our estimator is used to estimate the drift. As an example, we illustrate our results explicitly in the case of a multifactor model, where Arbitrage Pricing Theory predicts that alphas should be approximately zero.

Contents

Introduction

This thesis is a contribution to the theory of forward investment performance processes. Such processes are used in optimal asset allocation problems. In this introductory chapter, we introduce the most widely studied frameworks used in optimal asset allocation — mean–variance analysis and expected utility maximisation — before reviewing briefly some of the main results on forward investment performance processes. We then describe the structure of this thesis.

Mean–variance analysis

Harry Markowitz is considered the father of modern portfolio theory due to his pioneering work on mean–variance analysis — see Markowitz (1952, 1959). Therein, he proposes that investors should hold an efficient portfolio, a portfolio that minimises the risk while maximising the expected return. In other words, the objective of a mean–variance investor is

$$\min_{\pi} \left(\mathrm{Var}(X_T^{\pi}) - \eta \mathbb{E}[X_T^{\pi}] \right), \tag{1}$$

where T is the trading horizon, X_T^{π} is the terminal wealth generated by the trading strategy π, and η is a coefficient which measures the rate at which the investor is willing to trade variance for expected return.

Since its inception, mean–variance analysis has been the topic of many academic papers and is widely used in the investment management industry. In particular, it has been extended to multi–period and continuous time trading — see Li and Ng (2000) and Zhou and Li (2000) for the case of deterministic market parameters, and Lim and Zhou (2002) and Lim (2004) for the case of random market parameters. However, one of mean–variance main drawbacks is that, in a multi–period context, it yields time–inconsistent policies.

Expected utility maximisation

An alternative framework for optimal asset allocation problems is expected utility maximisation. Its axiomatic foundation goes back to von Neumann and Morgenstern (1947) and Savage (1954); in the context of portfolio selection it was first studied in Samuelson (1969) and Merton (1969), in the discrete and continuous time settings, respectively. In this framework, the objective is to maximise the expected utility of terminal wealth, that is,

$$\max_{\pi} \mathbb{E}\left[U\left(X_T^{\pi}\right)\right], \tag{2}$$

where $U(\cdot)$ is a utility function. There are essentially two approaches to solving (2). One is the dynamic programming method, pioneered by Bellman (1957) and used in the original works of Merton and Samuelson; the other is the martingale method — see, for instance, Karatzas, Lehoczky, and Shreve (1987).

One advantage of expected utility maximisation is that it is a time–consistent criterion, in the sense that Bellman's principle of optimality holds. More precisely, define the indirect value function,

$$u(t,x) = \max_{\pi} \mathbb{E}[U(X_T^{\pi})|X_t^{\pi} = x];$$

then, for any s and t such that $t \leq s \leq T$,

$$u(t,x) = \max_{\pi} \mathbb{E}[u(s,X_s^{\pi})|X_t^{\pi} = x]. \tag{3}$$

This time–consistency property leads to the interpretability of the optimal policy and it provides a method to solve (2) — the dynamic programming method[1].

Drawbacks of classic frameworks for asset allocation problems

One drawback of both mean–variance analysis and expected utility maximisation is the presence of a fixed investment horizon. In some cases, the investment horizon may not be fixed or known. Similarly, some derivative contracts have payoffs at random times, *e.g.* American options; to price such contracts via utility indifference, one needs to specify a utility function for each time and it is not obvious how this should be done. For example, in Henderson and Hobson (2007), the authors introduce the concept of horizon–unbiased utilities to price real options.

Recently, in Musiela and Zariphopoulou (2008, 2009), the authors introduce a new family of stochastic utilities, the so-called forward investment performance processes. These are defined for all times and, therefore, can be used in problems where the investment horizon is not set. For example, they are successfully used in Leung, Sircar, and Zariphopoulou (2012) to price American options. Furthermore, this new class of utilities is sufficiently rich to incorporate dynamic, random, risk preferences.

Another drawback of classic frameworks for asset allocation problems, such as mean–variance analysis and expected utility maximisation, is the ambiguity faced when choosing a market model. There are two ways of dealing with this ambiguity. One possibility is to consider a family of models (measures) together with a penalisation function and maximise the worst case penalised pay-off, where each model is penalised according to their perceived likeliness. This minimax approach is often referred to as Knightian uncertainty — a reference to the original work of Knight (1921) — or, simply, as model uncertainty. There has been a growing number of works on this topic; we refer to Maccheroni et al. (2006), Schied (2007) and Schied, Föllmer, and Weber (2009) and the references therein

---

[1] There are some measurability subtleties associated with the dynamic programming principle, particularly with equation (3) — see, for instance, Bouchard and Touzi (2009).

for an overview. More recently, in Källblad, Obłój, and Zariphopoulou (2013), the authors combine Knightian uncertainty with forward investment performance processes.

The other possibility is to consider a probability distribution supported on a class of market models — the so-called prior distribution — and maximise the average pay-off. This approach is often referred to as partial information and is of a more practical nature than model uncertainty. For example, one can consider the class of Black–Scholes models with a fixed volatility and different drifts which are distributed according to some prior — this particular type of partial information is called Bayesian uncertainty. Expected utility maximisation problems under partial information have been studied extensively in the literature — see, for instance, Lakner (1995, 1998). Likewise, the mean–variance problem under partial information is solved in Xiong and Zhou (2007). In Chapter 2, we combine partial information with forward investment performance processes.

Forward investment performance processes

Before proceeding, we briefly review some important results on forward investment performance processes. We work on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a Brownian motion $W_\cdot$. We denote by $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$ the filtration generated by $W_\cdot$.

Given a progressively measurable process $\pi_\cdot$, we define $X^\pi_\cdot$ as a solution to

$$dX^\pi_t = X^\pi_t \pi'_t \sigma_t (\lambda_t dt + dW_t),$$

where $\lambda_\cdot$ and $\sigma_\cdot$ are progressively measurable processes. We denote by $\mathcal{A}$ the set of admissible trading strategies, defined as

$$\mathcal{A} = \left\{ \pi_\cdot \text{ progressively measurable} : X^\pi_\cdot > 0,\ \mathbb{E}\left[\sup_{s \in [0,\cdot]} (X^\pi_s)^2 \right] < \infty \right\}. \tag{4}$$

The definition of forward investment performance processes is as follows.

Definition 1. *Let $\mathcal{A}$ denote the set of admissible portfolios. A $\mathbb{F}$–adapted process $U(t,x) : [0,\infty) \times \mathbb{R}^+ \times \Omega \to \mathbb{R}$ is a forward investment performance process if*

(i) for each $t \geq 0$, the mapping $U(t,\cdot)$ is increasing and strictly concave,

(ii) for each $\pi \in \mathcal{A}$, $t \mapsto U(t, X_t^\pi)$ is a supermartingale, and

(iii) for each $x_0 > 0$, there exists a $\pi^*_\cdot \in \mathcal{A}$ such that $X_0^{\pi^*} = x_0$ and the process $t \mapsto U(t, X_t^{\pi^*})$ is a martingale.

Given a forward investment performance process $U(t,x)$, conditions (ii) and (iii) in Definition 1 imply that the following time–consistency condition[2] is satisfied,

$$U(t,x) = \sup_{\pi \in \mathcal{A}} \mathbb{E}[U(T, X_T^\pi) | \mathcal{F}_t,\, X_t^\pi = x], \tag{5}$$

for any t and T such that $t \leq T$.

We observe that the existence and characterisation of forward investment performance processes does not follow immediately from their definition. Indeed, this is an open area of research. There are two main results obtained in this direction. In Musiela and Zariphopoulou (2010b), the authors show that all, sufficiently smooth, forward investment performance processes satisfy the following stochastic partial differential equation (SPDE),

$$dU(t,x) = \frac{1}{2} \frac{1}{U_{xx}(t,x)} \|U_x(t,x)\lambda_t + a_x(t,x)\|^2 dt + a(t,x)' dW_t. \tag{6}$$

This SPDE, also known as the forward stochastic Hamilton–Jacobi–Bellman (HJB) equation, lacks general existence and uniqueness results. Some progress in solving (6) has been made in El Karoui and Mrad (2013, 2010).

---

[2] Equation (5) is also called the self–generating condition in Žitković (2009) and the horizon–unbiased condition in Henderson and Hobson (2007).

In Musiela and Zariphopoulou (2010a) and Berrier, Rogers, and Tehranchi (2009), the authors characterise all the forward investment performance processes $U(t,x)$ which are monotonically decreasing, i.e. such that $U(\cdot,x)$ is decreasing. It follows easily from (6) that any such $U(\cdot)$ in this class of processes is of the form

$$U(t,x) = u(A_t,x),$$

where $u(\cdot)$ is a solution of the partial differential equation (PDE),

$$u_t - \frac{1}{2}\frac{u_x^2}{u_{xx}} = 0,$$

and the process $A_\cdot$ is defined as

$$A_t = \int_0^t |\lambda_s|^2 ds. \tag{7}$$

The optimal portfolio for a monotonically decreasing forward investment performance process can be written in feedback form as

$$\pi^*(t,x) = \frac{r(A_t,x)}{x}\Sigma_t^{-1}\sigma_t\lambda_t, \tag{8}$$

where $\Sigma_t = \sigma_t\sigma_t'$ and $r(\cdot)$ is the local risk tolerance function,

$$r(t,x) = -\frac{u_x(t,x)}{u_{xx}(t,x)}. \tag{9}$$

Furthermore, $r(\cdot)$ satisfies the fast diffusion equation

$$r_t + \frac{1}{2}r^2 r_{xx} = 0. \tag{10}$$

As shown in Musiela and Zariphopoulou (2010a), monotonically decreasing forward investment performance processes are characterised by a positive Borel measure $\nu \in \mathcal{B}^+(\mathbb{R})$, where

$$\mathcal{B}^+(\mathbb{R}) = \left\{\nu \in \mathcal{B}(\mathbb{R}) : \nu \geq 0 \text{ and } \int_{\mathbb{R}} e^{xy}\nu(dx) < \infty,\ y \in \mathbb{R}\right\}. \tag{11}$$

More precisely, it is shown that

$$h(t,y) = \int_{\mathbb{R}} \frac{e^{xy-\frac{1}{2}x^2t} - 1}{x}\nu(dx) + h(0,0), \tag{12}$$

where $h(\cdot)$ is defined implicitly by

$$u_x(t,h(t,y)) = e^{-y+\frac{t}{2}}.$$

In particular, the characterisation of monotonically decreasing forward investment performance processes shows that Definition 1 is not vacuous. In Chapter 2, we establish the characterisation of a class of forward investment performance processes which are not monotonically decreasing.

Remark 1. *In Berrier, Rogers, and Tehranchi (2009), the authors use duality methods to obtain an integral representation for the convex conjugate function of* $u(t,\cdot)$*. More precisely, define* $v(t,y)$ *by*

$$v(t,y)=\sup_{x>0}\big(u(t,x)-xy\big).$$

Then, $v(\cdot)$ *admits the integral representation*

$$v(t,y)=\int_{\mathbb{R}^+}\frac{1}{1-z}\left(1-y^{1-z}e^{\frac{z(1-z)}{2}t}\right)\nu(dz)+v(0,1),$$

for some measure $\nu\in\mathcal{B}^+(\mathbb{R})$*.*

Structure of the thesis

Chapter 1 develops a connection between the classical mean–variance optimisation and time–monotone forward performance processes for infinitesimal trading times. Namely, we consider consecutive mean–variance problems and we show that, for an appropriate choice of the corresponding mean–variance trade-off coefficients, the wealth process that is generated converges (as the trading interval goes to zero) to the optimal wealth process generated by a time–monotone forward performance process. The choice of the trade-off coefficients — the η in (1) — is made in accordance to the evolution of the risk tolerance process of the forward performance process — the $r(A_t,x)$ in (8).

This result allows us to provide a fresh view on the issue of time–consistency of mean–variance analysis, for we propose a method to update mean–variance risk preferences forward in time. Specifically, we propose that the mean–variance trade-off coefficient should be updated forward in time according to the fast diffusion equation (10).

To prove our convergence result, we generalise a result by Gyöngy (1998) on the convergence of the Euler scheme for SDEs. This result, provided in Section 1.5, is stated in very general terms and, therefore, may have an interest on its own. In Section 1.6, we also study the regularity, in space and time, of the local risk tolerance function of monotonically decreasing forward investment performance processes; this novel result may also have an interest on its own.

In Chapter 2, we provide a characterisation of a class of forward investment performance processes in models with partial information. More precisely, we consider models where the drift is a random variable distributed according to a known distribution. Then, we characterise forward investment performance processes of the type $U(t,x) = u(t,x,R_t)$, where $R_.$ is the process of cumulative excess returns. Our characterisation is in terms of an integral representation for the convex conjugate function of $u(t,\cdot,z)$.

Our result relies on a standard separation theorem, proved in Section 2.2. Such a separation theorem allows us to show, in Section 2.3, that the function $u(\cdot)$ satisfies an ill-posed HJB equation. The HJB equation can be linearised and solved using the multidimensional version of Widder's theorem, as shown in Section 2.4.

We see that the class of processes studied in this chapter is quite rich and, in Section 2.5, we construct several such examples. We show that some of these forward investment performance processes, which we call mixtures of power utilities, have an intrinsic volatility, in the sense that it is not generated by changes of numéraire/measure — see Definition 5, for the precise definition. To the best of our knowledge, these constitute the first examples of such processes.

Chapter 3 provides a generalisation of the Black–Litterman model to the continuous time setting. This generalisation is different from, and complements that of, Frey, Gabih, and Wunderlich (2012) and Davis and Lleo (2013).

In the Black–Litterman model, an asset pricing theory (such as the CAPM) is used in conjunction with investor's views to form an estimate of one–period expected returns. The objective of such an approach is to make asset allocation problems more robust with respect to estimation errors. In the recent papers Frey, Gabih, and Wunderlich (2012) and Davis and Lleo (2013), the authors combine investor's views with past returns to form an estimate of instantaneous expected returns.

We develop a framework in which it is possible to combine an asset pricing theory with past returns in order to estimate the instantaneous expected returns. The idea is very simple: instead of taking instantaneous expected returns to be constant, we model them as an Ornstein–Uhlenbeck process which mean–reverts around their theoretical value — the value predicted by an asset pricing theory. In other words, we take the theoretical value as a shrinkage target to which estimates of instantaneous expected returns should be continuously shrunk.

We derive our estimator using standard filtering tools in Section 3.2. Then, in Section 3.3, we explicitly solve the problem of maximising expected power utility of terminal wealth when the drift is estimated using such an estimator. We make these results very concrete by considering the specific example of a multifactor model and illustrating them using data on the American equities market.

This last chapter places less emphasis on the asset allocation problem and more on the problem of estimating the market parameters. We observe that, despite the shift in focus, this chapter follows naturally from Chapter 2, for both chapters address the issue of uncertainty associated with model specification.

Chapter 1

Infinitesimal Mean–Variance and Forward Investment

Mean–variance analysis is one of the oldest and most widely studied frameworks for asset allocation problems. However, it has one main drawback, in that it yields time–inconsistent policies in a multi–period context. In the literature, there are two ways of dealing with time–inconsistency in control problems. One way is to consider the precommitted strategy, that is the allocation which is optimal at time zero only; this allocation is derived in Li and Ng (2000), in the discrete time case, and in Zhou and Li (2000), in the continuous time case. The other way is to frame it within a game–theoretic context. In this case, one views the mean–variance problem as a game where the players are different incarnations of the same investor at different times (with one player for each time); the optimal strategy is then a Nash equilibrium point. The study of mean–variance analysis using a game–theoretic approach first appeared in Basak and Chabakauri (2010). A more general and more economically sensible result can be found in Björk, Murgoci, and Zhou (2014). There, the authors make use of a general theory for time–inconsistent problems developed in Björk and Murgoci (2010).

Another drawback of mean–variance analysis is having a fixed investment horizon. In particular, this criterion does not provide the flexibility to update risk preferences forward in time. We observe that the paradigm of updating preferences forward is advocated, for instance, in Black (1968).

More recently, in Musiela and Zariphopoulou (2008, 2009), the authors introduce a new framework for asset allocation, using the so-called forward investment performance processes. These processes are updated forward in time and do not have a fixed investment horizon. Furthermore, they yield time–consistent allocations. Therefore, they do not suffer from the aforementioned drawbacks.

In this chapter, we develop a connection between the classical mean–variance optimisation and time–monotone forward performance processes for infinitesimal trading times. Namely, we consider consecutive mean–variance problems and we show that, for an appropriate choice of the corresponding mean–variance trade-off coefficients, the wealth process that is generated converges (as the trading interval goes to zero) to the optimal wealth process generated by a time–monotone forward performance process. The choice of the trade-off coefficients is made in accordance to the evolution of the risk tolerance process of the forward performance process.

This convergence result allows us to provide a fresh view on the issues of time–inconsistency and horizon inflexibility of mean–variance analysis, for we propose a method to update mean–variance risk preferences forward in time.

Our approach has two drawbacks. One is that our conclusions are based on a limiting result and, therefore, only hold in the infinitesimal time-scale. The other drawback of this approach is that it is not purely based on mean–variance analysis. Instead, we focus on how forward investment performance processes may be used to produce a forward version of the mean–variance criterion.

The remainder of this chapter is organised as follows. In Section 1.1, we briefly review the well known mean–variance problem in a single period. Then, in Section 1.2, we consider two consecutive mean–variance problems. In Section 1.3, we consider several consecutive mean–variance problems, each with an investment horizon of length $\delta > 0$. Then, we show that, under certain regularity assumptions, the optimal wealth process generated by these consecutive problems converges, as $\delta \to 0$, to a limiting process. Using the results of Section 1.3, in Section 1.4, we establish a connection between the mean–variance problem and monotonically decreasing forward investment performance processes. Section 1.5 is devoted to proving an extension of a result by Gyöngy on the convergence of the Euler scheme — a technical result which is used in Section 1.3. In Section 1.6, we study the regularity of the local risk tolerance function of time–monotone forward performance processes; this auxiliary result is required in Section 1.4.

1.1 Single–period mean–variance

Consider a market with N risky assets and one riskless asset which is used as a numéraire. In a single–period mean–variance model, one considers two dates: the initial date $t = 0$, at which the investment decision is made, and the investment horizon $t = T$, at which returns are realised. We denote the excess return of the i^{th} risky asset over the investment period $[0, T]$ by R^i_T. It is assumed that R^i_T has finite variance. We define the vector of expected excess returns, $\mu_0 = \left(\mu^1_0, \cdots, \mu^N_0\right)$, and the covariance matrix of excess returns, $\Sigma_0 = \left(\Sigma^{ij}_0\right)_{ij}$, where

$$\mu^i_0 = \mathbb{E}\left[R^i_T\right] \text{ and } \Sigma^{ij}_0 = \mathrm{Cov}\left(R^i_T, R^j_T\right).$$

For simplicity, we let $R_T = \left(R^1_T, \cdots, R^N_T\right)$ denote the vector of excess returns.

The proportion of wealth invested in the i^{th} risky asset at time 0 is denoted by π^i_0. We say that $\pi_0 = \left(\pi^1_0, \cdots, \pi^N_0\right)$ forms a portfolio. The excess return of π_0 realised over the investment period $[0, T]$ is denoted by $R^{\pi_0}_T$ and satisfies

$$R^{\pi_0}_T = \sum_{i=1}^{N} \pi^i_0 R^i_T.$$

At time T, the discounted wealth associated with a portfolio π_0 is given by

$$X^{\pi_0}_T = x_0 \left(1 + R^{\pi_0}_T\right),$$

where x_0 is the initial wealth.

In single–period mean–variance analysis, the objective of an investor at time 0 is the minimisation,

$$\inf_{\pi_0 \in \mathbb{R}^N} \left(\operatorname{Var}\left(X_T^{\pi_0}\right) - \eta_0 \mathbb{E}\left[X_T^{\pi_0}\right] \right), \tag{1.1}$$

where the scalar η_0 is an exogenous quantity, specified by the investor at time 0. We observe that η_0 characterises the rate at which the investor is willing to trade variance for expected return; we refer to this parameter as the mean–variance trade-off coefficient.

A solution to the unconstrained minimization problem (1.1) yields the optimal investment decision for a single–period mean–variance investor, as in the following theorem due to Markowitz.

Theorem 1 (Markowitz (1959))**.** *The optimal portfolio for a mean–variance investor with objective given by* (1.1) *at time* 0 *is given by*

$$\pi_0^* = \frac{\eta_0}{x_0} \Sigma_0^{-1} \mu_0. \tag{1.2}$$

To ease the notation, we denote the terminal value of the optimal discounted wealth process as X_T^*,

$$X_T^* = x_0 + \eta_0 \mu_0' \Sigma_0^{-1} R_T. \tag{1.3}$$

In Figure 1.1, we represent schematically the single–period mean–variance problem.

Figure 1.1: Diagram representing single–period mean–variance analysis.

1.2 Beyond a single period

In practice, at the end of an investment period, investors are likely to face subsequent investment decisions. Suppose that, indeed, this is the case and, at time

$t = T$, one has a new investment horizon, $2T$. We denote by $R_{2T} = \left(R_{2T}^1, \cdots, R_{2T}^N\right)$ the vector of cumulative excess returns at time $2T$. In other words, $R_{2T}^i - R_T^i$ is the excess return of the i^{th} risky asset over the period $[T, 2T]$. As in the previous section, we denote

$$\mu_T^i = \mathbb{E}_T\left[R_{2T}^i - R_T^i\right] \text{ and } \Sigma_T^{ij} = \mathrm{Cov}_T\left(R_{2T}^i - R_T^i, R_{2T}^j - R_T^j\right),$$

where $\mathbb{E}_T$ and Cov_T are, respectively, the expectation and covariance operators conditioned on the realisation of R_T.

The investment decision at time T is represented by a portfolio, denoted by $\pi_T = \left(\pi_T^1, \cdots, \pi_T^N\right)$. Each quantity π_T^i represents the proportion of discounted wealth invested in the i^{th} risky asset. Note that, since π_T is chosen at time T, it can depend on the realisation of returns R_T. For simplicity, we let

$$\pi = (\pi_0, \pi_T)$$

denote the sequence of portfolios at times 0 and T.

At time $2T$, the cumulative excess return of the sequence of investment decisions π is given by

$$R_{2T}^\pi = R_T^\pi + \sum_{i=1}^N \pi_T^i\left(R_{2T}^i - R_T^i\right).$$

Similarly, at time $2T$, the discounted wealth associated with the above π is

$$X_{2T}^\pi = X_T^\pi\left(1 + R_{2T}^\pi - R_T^\pi\right).$$

The mean–variance problem for the second period is similar to (1.1); the main difference is that, at time T, the problem is to be solved conditioned on the realisation of returns in the first period. More precisely, given the initial portfolio π_0, at time T, the investor faces the minimisation

$$\inf_{\pi_T \in \mathbb{R}^N}\left(\mathrm{Var}_T\left(X_{2T}^\pi\right) - \eta_T \mathbb{E}_T\left[X_{2T}^\pi\right]\right), \tag{1.4}$$

where $\pi = (\pi_0, \pi_T)$.

As in the previous section, the mean–variance trade-off coefficient η_T is exogenously specified by the investor and should reflect his risk preferences at time T.

However, η_T *differs* from η_0 in that it may depend on the realisation of returns R_T. More precisely, we write $\eta_T \in \sigma(R_T)$, where $\sigma(R_T)$ denotes the σ–algebra generated by R_T. In particular, we may have $\eta_T \neq \eta_0$.

It is easy to deduce that the optimal portfolio, π_T^*, for the objective (1.4) is given by

$$\pi_T^* = \frac{\eta_T}{X_T^*}\Sigma_T^{-1}\mu_T. \tag{1.5}$$

Respectively, the optimal discounted wealth process at time $2T$, denoted by X_{2T}^*, is given by

$$X_{2T}^* = X_T^* + \eta_T \mu_T' \Sigma_T^{-1}(R_{2T} - R_T).$$

A question which arises naturally at this point is whether π_T^* is time–consistent, in some sense, with π_0^*. *However, it is not clear what is meant by time–consistency in this case, since the second investment period is unaccounted for at the time of the first investment decision.* Such time–consistency considerations may impose restrictions on the evolution of the investor's risk preferences, i.e. the specification of the mean–variance trade-off coefficient η_T. In the remainder of this chapter, we analyse this question of time–consistency with the lenses of forward investment performance processes. Specifically, we propose a method to update the mean–variance trade-off coefficient forward in time; this updating method is justified in the limit as $T \to 0$.

Remark 2. *The two–step mean–variance procedure described in this section is not equivalent to what is known in the literature as multi–period mean–variance analysis. In the latter, at the beginning of the first period the investor already accounts for trading in all subsequent periods. In other words, in multi–period mean–variance analysis with two periods, the objective of an investor is*

$$\inf_{\pi=(\pi_0,\pi_T)} \Big(\mathrm{Var}\left(X_{2T}^\pi\right) - \eta_0 \mathbb{E}\left[X_{2T}^\pi\right] \Big). \tag{1.6}$$

The two procedures are represented schematically in Figures 1.2 and 1.3. Note

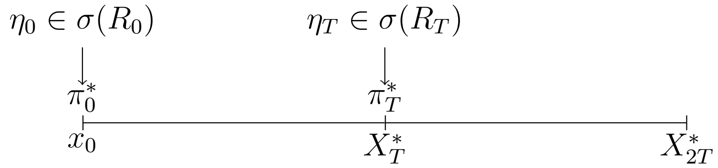

Figure 1.2: Diagram representing mean–variance analysis in two consecutive periods.

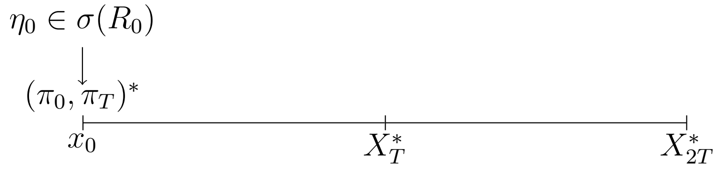

Figure 1.3: Diagram representing multi–period mean–variance analysis with two periods.

that in multi–period mean–variance analysis the investor sets his preferences only once, at the beginning of the first period, and does not have the flexibility to revise them later. The difference between the two procedures is analogous to the difference between (backwards) expected utility maximisation and forward investment performance processes.

Nevertheless, even in multi–period mean–variance analysis one eventually faces the questions raised in this section. Indeed, in multi–period mean–variance analysis, when the final period ends and new investment decisions have to be made — decisions unaccounted for at time 0 *— the issue of how to update risk preferences forward in a time–consistent fashion arises.*

1.3 Infinitesimal mean–variance

In this section, we study the application of mean–variance analysis in consecutive periods, as in Section 1.2. We focus on the wealth process generated by such a procedure; our objective is to study the structure of this process, in the limit as the length of each period shrinks to zero.

We work in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a standard $\mathbb{R}^N$–valued Brownian motion $W_\cdot$. We denote by $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$ the completion of the filtration generated by $W_\cdot$.

As in the previous sections, we consider N risky assets and one riskless asset which is used as a numéraire. Excess returns are measured discretely in intervals of length δ. Therefore, the cumulative excess returns of the risky assets are defined on the grid of points

$$\Theta^\delta = \left(t_i^\delta\right)_{i \in \mathbb{N}}, \tag{1.7}$$

where $t_i^\delta = i\,\delta$. For a fixed time $t \in \Theta^\delta$, we denote the vector of cumulative excess returns by $R_t = \left(R_t^1, \cdots, R_t^N\right)$.

We assume that excess returns are conditionally normally distributed. More precisely, we assume that

$$R^\delta_{t^\delta_{i+1}} = R^\delta_{t^\delta_i} + \mu_{t^\delta_i}\left(t^\delta_{i+1} - t^\delta_i\right) + \sigma_{t^\delta_i}\left(W_{t^\delta_{i+1}} - W_{t^\delta_i}\right), \tag{1.8}$$

where $\mu_t : [0, \infty) \times \Omega \to \mathbb{R}^N$ and $\sigma_t : [0, \infty) \times \Omega \to \mathbb{R}^{N \times N}$ are $\mathbb{F}$–progressively measurable processes. Furthermore, we assume that σ_t is invertible, and define $\lambda_t = \sigma_t^{-1}\mu_t$.

A portfolio $\pi_\cdot$ is a discrete process, $\pi_\cdot = (\pi_t)_{t \in \Theta^\delta}$, such that, for any $t \in \Theta^\delta$, $\pi_t \in \mathcal{F}_t$. The process of discounted wealth associated with $\pi_\cdot$ is denoted by $X_\cdot^{\pi,\delta}$ and is defined recursively, for all $t \in \Theta^\delta$, by

$$X^{\pi,\delta}_{t^\delta_{i+1}} = X^{\pi,\delta}_{t^\delta_i}\left(1 + \pi'_{t^\delta_i}\left(R^\delta_{t^\delta_{i+1}} - R^\delta_{t^\delta_i}\right)\right),$$

with initial condition $X_0^{\pi,\delta} = x_0$, where x_0 is the initial wealth.

An admissible portfolio, $\pi_\cdot$, is such that $\pi_t = 0$ whenever $X_t^{\pi,\delta} \leq 0$. More precisely, given a portfolio $\pi_\cdot$, we define its bankruptcy time as

$$\tau^{\pi,\delta} = \inf \left\{ t \in \Theta^\delta : X_t^{\pi,\delta} \leq 0 \right\}. \tag{1.9}$$

Then, the set of admissible portfolios is

$$\mathcal{A}^\delta = \left\{ \pi_\cdot : X_\cdot^{\pi,\delta} = X_{\cdot \wedge \tau^{\pi,\delta}}^{\pi,\delta} \right\}. \tag{1.10}$$

We observe that, by making this choice of admissibility, we are implicitly taking the wealth domain to be[1] $\mathbb{R}^+$.

Next, we consider portfolios in a feedback form. Specifically, let $\pi(t,x) : \Theta^\delta \times \mathbb{R} \times \Omega \to \mathbb{R}^N$ be a function such that, for all $x \in \mathbb{R}$ and $t \in \Theta^\delta$, $\pi(t,x) \in \mathcal{F}_t$; as usual in the literature, we omit the dependence of $\pi(t,x)$ on $\omega \in \Omega$. We define recursively the associated wealth process $X_\cdot^{\pi,\delta}$ by

$$X_{t_{i+1}^\delta}^{\pi,\delta} = X_{t_i^\delta}^{\pi,\delta} \left(1 + \pi\left(t_i^\delta, X_{t_i^\delta}^{\pi,\delta} \right)' \left(R_{t_{i+1}^\delta}^\delta - R_{t_i^\delta}^\delta \right) \right).$$

Then, the portfolio associated with $\pi(\cdot)$ is denoted, with a slight abuse of notation, by $\pi_\cdot$ and defined for all $t \in \Theta^\delta$ as

$$\pi_t = \pi(t, X_t^{\pi,\delta}).$$

Hereafter, we use $\pi(\cdot)$ interchangeably to refer to both the function $\pi(t,x) : \Theta^\delta \times \mathbb{R} \times \Omega \to \mathbb{R}^N$ and the associated portfolio $\pi_t : \Theta^\delta \times \Omega \to \mathbb{R}^N$. We observe that, a feedback portfolio, $\pi(\cdot)$, is admissible if and only if $\pi(\cdot, x) = 0$ whenever $x \leq 0$.

Next, we consider for each period $\left[t_i^\delta, t_{i+1}^\delta\right]$ an associated mean–variance problem as follows. At any time t_i^δ, the optimal mean–variance allocation is given by the minimisation

---

[1] This is for the sake of concreteness. It is possible to consider other wealth domains, such as, for example, $\mathbb{R}$. Indeed, the technical results in Section 1.5, which constitute the backbone of the chapter, are stated for a general wealth domain D.

$$\inf_{\pi\in\mathbb{R}^N}\left(\mathrm{Var}_{t_i^\delta}\left(X^{\pi}_{t^\delta_{i+1}}\right)-\eta_{t_i^\delta}\mathbb{E}_{t_i^\delta}\left[X^{\pi}_{t^\delta_{i+1}}\right]\right), \tag{1.11}$$

where $\mathbb{E}_t$ and Var_t are, respectively, the expectation and variance operators conditioned on $\mathcal{F}_t$.

The mean–variance trade-off coefficient $\eta_{t_i^\delta}$ is exogenously specified by the investor and is taken to satisfy $\eta_{t_i^\delta}\in\mathcal{F}_{t_i^\delta}$.

The optimal portfolio for objective (1.11) is then defined for all $t\in\Theta^\delta$, in a feedback form, by

$$\pi^{\eta,\delta}(t,x)=\frac{\eta_t}{x}\Sigma_t^{-1}\mu_t, \tag{1.12}$$

where $\Sigma_t=\sigma_t\sigma_t'$. The associated wealth process is denoted, for simplicity, as $X^{\eta,\delta}_{\cdot}$ and is given recursively by

$$X^{\eta,\delta}_{t^\delta_{i+1}}=X^{\eta,\delta}_{t^\delta_i}+\eta_{t_i^\delta}\mu'_{t_i^\delta}\Sigma^{-1}_{t_i^\delta}\left(R^\delta_{t^\delta_{i+1}}-R^\delta_{t^\delta_i}\right),$$

with initial condition $X_0^{\eta,\delta}=x_0$. Figure 1.4 provides a schematic representation of the consecutive mean–variance problems.

Analogously to feedback portfolios, it is possible to specify the sequence of trade-off coefficients, $(\eta_t)_{t\in\Theta^\delta}$, in a feedback form. More precisely, let $\eta(t,x)$: $\Theta^\delta\times\mathbb{R}\times\Omega\to\mathbb{R}$ be a function such that, for all $x\in\mathbb{R}$ and $t\in\Theta^\delta$, $\eta(t,x)\in\mathcal{F}_t$.

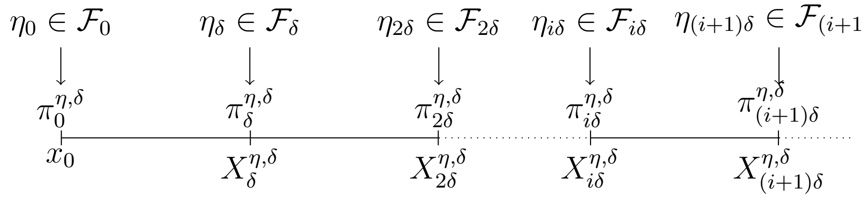

Figure 1.4: Diagram representing consecutive mean–variance problems in periods of length δ.

We define the associated wealth process $X^{\eta,\delta}_{\cdot}$ recursively, for all $t \in \Theta^{\delta}$, by

$$X^{\eta,\delta}_{t^{\delta}_{i+1}} = X^{\eta,\delta}_{t^{\delta}_{i}} + \eta\left(t^{\delta}_{i}, X^{\eta,\delta}_{t^{\delta}_{i}}\right)\mu'_{t^{\delta}_{i}}\Sigma^{-1}_{t^{\delta}_{i}}\left(R^{\delta}_{t^{\delta}_{i+1}} - R^{\delta}_{t^{\delta}_{i}}\right). \tag{1.13}$$

Then, the risk tolerance process associated with $\eta(\cdot)$ is denoted, with a slight abuse of notation, by $\eta_{\cdot}$ and defined for all $t \in \Theta^{\delta}$ as

$$\eta_t = \eta\left(t, X^{\eta,\delta}_{t}\right).$$

We observe that the portfolio associated with $\eta(\cdot)$, $\pi^{\eta,\delta}_{\cdot}$ defined as in (1.12), is admissible if and only if $\eta(\cdot, x) = 0$ whenever $x \leq 0$. The bankruptcy time of $\pi^{\eta,\delta}_{\cdot}$ is denoted, for simplicity, as $\tau^{\eta,\delta}$; with this notation, (1.9) is rewritten as

$$\tau^{\eta,\delta} = \inf\left\{t \in \Theta^{\delta} : X^{\eta,\delta}_{t} \leq 0\right\}. \tag{1.14}$$

In the sequel, we consider only admissible portfolios. Therefore, feedback portfolios and trade-off functions are taken to be defined on the wealth domain, $\mathbb{R}^+$; we assume that outside this domain these functions take the value 0.

Convergence as $\delta \to 0$

We provide auxiliary results on the structure of the process $X^{\eta,\delta}_{\cdot}$, as $\delta \to 0$. To this end, we introduce the following regularity assumptions.

Assumption 1 (Regularity of the market coefficients)**.** *The processes $\mu_{\cdot}$, $\sigma_{\cdot}$, and $\Sigma^{-1}_{\cdot}$ in* (1.8) *are almost surely bounded and $\alpha-$Hölder continuous.*

Assumption 2 (Regularity of the mean–variance trade-off coefficient)**.** *Let $\eta(t, x) : [0, \infty) \times \mathbb{R}^+ \times \Omega \to \mathbb{R}$.*

For each compact set $E \subset \mathbb{R}^+$, and each $T \geq 0$, there exists $K > 0$, such that, *almost surely,*

$$
\begin{aligned}
|\eta(t,x)-\eta(t,y)| &\leq K|x-y|, \\
|\eta(t,x)| &\leq K(1+|x|), \\
|\eta(t,x)-\eta(s,x)| &\leq K(1+|x|)|t-s|^{\alpha},
\end{aligned} \tag{1.15}
$$

for all $x, y \in E$, *and* $t, s \leq T$.

We say that the constant K *is global if it does not depend on the set* E, *i.e. if* (1.15) *holds for all* $x, y > 0$.

The following theorem shows that, under the assumptions above, the sequence of discounted wealth processes $\left(X^{\eta,\delta}_{\cdot}\right)_{\delta>0}$ converges, as $\delta \to 0$, to a process, denoted by $X^{\eta}_{\cdot}$. Furthermore, this process is itself a discounted wealth process generated by a continuous time trading strategy. In other words, Theorem 2 shows that performing mean–variance analysis in infinitesimally small periods of time is equivalent to following a continuous time trading strategy.

Theorem 2. *Suppose that Assumption 1 holds and* $\eta(t,x) : [0,\infty) \times \mathbb{R}^{+} \times \Omega \to \mathbb{R}$ *satisfies Assumption 2.*

Let $X^{\eta,\delta}_{\cdot}$ *be the unique strong solution to* (1.13). *Similarly, let* $X^{\eta}_{\cdot}$ *be the unique strong solution to*

$$
dX^{\eta}_t = \theta^{\eta}\left(t, X^{\eta}_t\right)' \left(\mu_t dt + \sigma_t dW_t\right),
$$

where $\theta^{\eta}(t,x) = \eta(t,x)\Sigma_t^{-1}\mu_t$, *and denote by* τ^{η} *its explosion time, that is,*

$$
\tau^{\eta} = \inf\left\{t : X^{\eta}_t \notin \mathbb{R}^{+}\right\}. \tag{1.16}
$$

Then, for any bounded stopping time τ *such that* $\tau < \tau^{\eta}$, *the following hold:*

(i) if $(\delta_n)_{n\in\mathbb{N}}$ *is such that* $\sum_n \delta_n^{\beta} < \infty$, *for some* $\beta > 0$, *then*

$$
\sup_{t\in\Theta^{\delta_n}:t<\tau} \left|X^{\eta,\delta_n}_t - X^{\eta}_t\right| \to 0, \ \text{a.s.}, \tag{1.17}
$$

as $n \to \infty$;

(ii) if the constant K in Assumption 2 is global, then

$$\mathbb{E}\left[\sup_{t\in\Theta^\delta : t<\tau}\left|X_t^{\eta,\delta} - X_t^\eta\right|^p\right] \to 0, \tag{1.18}$$

as $\delta \to 0$, for any $p \geq 2$.

Proof. To prove this result it is convenient to interpolate the discrete time processes $X_\cdot^{\eta,\delta}$. Specifically, we define the auxiliary process $\bar{X}_\cdot^{\eta,\delta}$ as a solution to

$$d\bar{X}_t^{\eta,\delta} = \theta\left(\kappa_t^\delta, \bar{X}_{\kappa_t^\delta}^{\eta,\delta}\right)'\left(\mu_{\kappa_t^\delta}dt + \sigma_{\kappa_t^\delta}dW_t\right),$$

where $\kappa_t^\delta = \delta\left\lfloor\frac{t}{\delta}\right\rfloor$. Additionally, we define

$$\bar{\tau}^{\eta,\delta} = \inf\left\{t : \bar{X}_{\kappa_t^\delta}^{\eta,\delta} \leq 0\right\}.$$

We recall that $\eta(\cdot, x) = 0$ for all $x \leq 0$. Therefore, $\bar{X}_t^{\eta,\delta}$ is defined for all $t \geq 0$, although it remains constant for $t \geq \bar{\tau}^{\eta,\delta}$.

Then,

$$\bar{X}_t^{\eta,\delta} = X_t^{\eta,\delta},$$

for all $t \in \Theta^\delta$. Therefore, we can prove (1.17) and (1.18) with $\bar{X}_\cdot^{\eta,\delta}$ instead of $X_\cdot^{\eta,\delta}$.

Furthermore, by Theorem 7 proved in Section 1.5, the following properties hold:

- as $\delta \to 0$, we have

$$\mathbb{P}\left(\bar{\tau}^{\eta,\delta} < \tau\right) \to 0; \tag{1.19}$$

- under hypothesis (i), we have, almost surely,

$$\liminf_n \bar{\tau}^{\eta,\delta_n} \geq \tau^\eta, \text{ and} \tag{1.20}$$

$$\sup_{t<\bar{\tau}^{\eta,\delta}\wedge\tau}\left|\bar{X}_t^{\eta,\delta_n} - X_t^\eta\right| \to 0, \tag{1.21}$$

as $n \to \infty$;

- under hypothesis (ii),

$$\mathbb{E}\left[\sup_{t<\bar{\tau}^{\eta,\delta}\wedge\tau}\left|\bar{X}_t^{\eta,\delta_n} - X_t^{\eta}\right|^p\right] \to 0, \tag{1.22}$$

as $\delta \to 0$, for any $p \geq 2$.

Let $(\delta_n)_{n\in\mathbb{N}}$ be as in hypothesis (i) and let $\omega \in \Omega$ be such that (1.20) and (1.21) hold. Then, for such an ω and n large enough,

$$\begin{aligned}\sup_{t\in\Theta^{\delta_n}:t<\tau}\left|X_t^{\eta,\delta_n} - X_t^{\eta}\right| &= \sup_{t\in\Theta^{\delta_n}:t<\bar{\tau}^{\eta,\delta_n}\wedge\tau}\left|\bar{X}_t^{\eta,\delta_n} - X_t^{\eta}\right| \\ &\leq \sup_{t<\bar{\tau}^{\eta,\delta_n}\wedge\tau}\left|\bar{X}_t^{\eta,\delta_n} - X_t^{\eta}\right|.\end{aligned}$$

Therefore,

$$\sup_{t\in\Theta^{\delta_n}:t<\tau}\left|X_t^{\eta,\delta_n} - X_t^{\eta}\right| \to 0,$$

as $n \to \infty$. Since (1.20) and (1.21) hold almost surely, then so does (1.17).

Now, suppose that hypothesis (ii) holds. We observe that

$$\begin{aligned}\sup_{t\in\Theta^{\delta}:t<\tau}\left|X_t^{\eta,\delta} - X_t^{\eta}\right|^p &= \sup_{t\in\Theta^{\delta}:t<\bar{\tau}^{\eta,\delta}\wedge\tau}\left|\bar{X}_t^{\eta,\delta} - X_t^{\eta}\right|^p \mathbb{1}_{\{\bar{\tau}^{\eta,\delta}\geq\tau\}} \\ &\quad + \sup_{t\in\Theta^{\delta}:t<\tau}\left|\bar{X}_t^{\eta,\delta} - X_t^{\eta}\right|^p \mathbb{1}_{\{\bar{\tau}^{\eta,\delta}<\tau\}} \\ &\leq \sup_{t<\bar{\tau}^{\eta,\delta}\wedge\tau}\left|\bar{X}_t^{\eta,\delta} - X_t^{\eta}\right|^p + \sup_{t<\tau}\left|\bar{X}_t^{\eta,\delta} - X_t^{\eta}\right|^p \mathbb{1}_{\{\bar{\tau}^{\eta,\delta}<\tau\}}\end{aligned}$$

Therefore, since (1.22) holds, to prove (1.18), it suffices to show that

$$\mathbb{E}\left[\sup_{t<\tau}\left|\bar{X}_t^{\eta,\delta} - X_t^{\eta}\right|^p \mathbb{1}_{\{\bar{\tau}^{\eta,\delta}<\tau\}}\right] \to 0, \tag{1.23}$$

as $\delta \to 0$. The convergence result above follows from the Cauchy–Schwarz inequality. Indeed, we observe that, by Theorem 6 proved in Section 1.5, there exists a constant C, independent of δ, such that

$$\mathbb{E}\left[\sup_{t<\tau\wedge\bar{\tau}^{\eta,\delta}}\left|\bar{X}_t^{\eta,\delta}\right|^{2p}\right] < C \text{ and } \mathbb{E}\left[\sup_{t<\tau}|X_t^{\eta}|^{2p}\right] < C. \tag{1.24}$$

Furthermore, since $t \mapsto \bar{X}_t^{\eta,\delta}$ is continuous and remains constant for $t \geq \tau^{\eta,\delta}$, we deduce that

$$\sup_{t<\tau\wedge\bar{\tau}^{\eta,\delta}}\left|\bar{X}_t^{\eta,\delta}\right| = \sup_{t<\tau}\left|\bar{X}_t^{\eta,\delta}\right|.$$

Then, (1.19) and (1.24) together with the Cauchy–Schwarz inequality yield (1.23). □

Theorem 2 is concerned with the convergence of the wealth processes generated by consecutive mean–variance problems, $\left(X_\cdot^{\eta,\delta}\right)_{\delta>0}$. It states that the limit of these processes is itself a wealth process, $X_\cdot^{\eta}$, which is generated by the continuous time strategy $\pi_\cdot^{\eta}$, where

$$\pi_t^{\eta} = \frac{\eta\left(t, X_t^{\eta}\right)}{X_t^{\eta}}\Sigma_t^{-1}\mu_t. \tag{1.25}$$

A related question concerns the convergence of the sequence of optimal mean–variance allocations, $\left(\pi_\cdot^{\eta,\delta}\right)_{\delta>0}$. As expected, this sequence of portfolios converges to $\pi_\cdot^{\eta}$, as shown in the following proposition.

Proposition 1. *Suppose that Assumption 1 holds and* $\eta(t,x) : [0,\infty)\times\mathbb{R}^{+}\times\Omega \to \mathbb{R}$ *satisfies Assumption 2.*

Let $X_\cdot^{\eta,\delta}$, $X_\cdot^{\eta}$ *and* τ^{η} *be as in the hypothesis of Theorem 2.*

Define $\pi_\cdot^{\eta}$ *as in* (1.25), *and* $\pi_\cdot^{\eta,\delta}$ *as in*

$$\pi_t^{\eta,\delta} = \frac{\eta\left(t, X_t^{\eta,\delta}\right)}{X_t^{\eta,\delta}}\Sigma_t^{-1}\mu_t.$$

Let τ *be a bounded stopping time such that* $\tau < \tau^{\eta}$. *Then, the following hold:*

(i) if $(\delta_n)_{n\in\mathbb{N}}$ *is such that* $\sum_n \delta_n^{\beta} < \infty$, *for some* $\beta > 0$, *then*

$$\sup_{t\in\Theta^{\delta_n}:t<\tau}\left|\pi_t^{\eta,\delta_n} - \pi_t^{\eta}\right| \to 0, \ a.s.; \tag{1.26}$$

(ii) if $\frac{\eta(t,x)}{x}$ is almost surely bounded, then

$$\mathbb{E}\left[\sup_{t\in\Theta^{\delta}:t<\tau}\left|\pi_t^{\eta,\delta}-\pi_t^{\eta}\right|^p\right]\to 0, \tag{1.27}$$

as $\delta \to 0$.

Proof. To prove (i), we observe that, since the function $x \mapsto \frac{\eta(t,x)}{x}$ is locally Lipschitz and

$$\min_{t\leq\tau} X_t^{\eta} > 0,$$

then, for each ω, there exists K such that

$$\left|\frac{\eta\left(t,X_t^{\eta,\delta_n}\right)}{X_t^{\eta,\delta_n}}-\frac{\eta\left(t,X_t^{\eta}\right)}{X_t^{\eta}}\right|\leq K\left|X_t^{\eta,\delta_n}-X_t^{\eta}\right|,$$

for all $t\in\Theta^{\delta}$ such that $t<\tau$, and for all n large enough. Then, (i) follows from (1.17).

The result in (ii) is a consequence of (i). Indeed, let $(\delta_n)_{n\in\mathbb{N}}$ be an arbitrary sequence such that $\delta_n \to 0$. Then, there is a subsequence $(\delta_{u_n})_{n\in\mathbb{N}}$, such that $\sum_n \delta_{u_n} < \infty$. By (1.26),

$$\sup_{t\in\Theta^{\delta_{u_n}}:t<\tau}\left|\pi_t^{\eta,\delta_{u_n}}-\pi_t^{\eta}\right|\to 0, \ a.s.,$$

and, therefore, by the dominated convergence theorem,

$$\mathbb{E}\left[\sup_{t\in\Theta^{\delta_{u_n}}:t<\tau}\left|\pi_t^{\eta,\delta_{u_n}}-\pi_t^{\eta}\right|^p\right]\to 0,$$

as $n\to\infty$. Since the sequence $(\delta_n)_{n\in\mathbb{N}}$ is arbitrary, we conclude that (1.27) holds. □

Remark 3. *In Skiadas (2008), a continuous time allocation of the form*

$$\pi_t=\eta_t\Sigma_t^{-1}\mu_t,$$

for some process $\eta_\cdot$, is said to be instantaneously mean–variance efficient. We observe that the limiting allocation $\pi^\eta_\cdot$, defined in (1.25)*, has this form. Indeed, with Theorem 2 and Proposition 1, we provide a theoretical foundation to support the notion of instantaneous mean–variance efficiency.*

Mathematically, Theorem 2 does not encompass any surprising result. Indeed, it is essentially a consequence of the convergence of the Euler approximating scheme for SDEs, or more precisely of an interpolation of the Euler scheme. The convergence of such schemes has been widely studied in the literature. Among others, in Gyöngy and Krylov (1996) and Gyöngy (1998), the authors prove the convergence in probability and almost surely, respectively, of a similar scheme, under somewhat different conditions. In Section 1.5, we establish the convergence of a larger class of schemes, a result on which the proof of Theorem 2 is based.

We end this section with a numerical illustration of the convergence result in Theorem 2. We consider the risk tolerance process $\eta(t,x) = x^+$ and a constant $|\lambda_\cdot| = 1$. In Figure 1.5, we compare the distributions of X^η_T and $X^{\eta,\delta}_T$, where $T = 1$ (in years) and $\delta \in \{1/2, 1/12\}$. More precisely, with a solid line, we plot the density of $\frac{X^\eta_T - x_0}{x_0}$; using histograms, we plot the approximate density of $\frac{X^{\eta,\delta}_T - x_0}{x_0}$ for $\delta \in \{1/2, 1/12\}$ (in red, we consider semi–annual periods, $\delta = 1/2$, and, in green, monthly periods, $\delta = 1/12$).

1.4 Mean–variance and forward investment

We recall some of the results on monotonically decreasing forward investment performance processes; these, and other, results are briefly reviewed in the Introduction.

We consider a continuously traded market with N risky assets. The process of cumulative excess returns is denoted by $R_\cdot$ and taken to satisfy

$$dR_t = \mu_t dt + \sigma_t dW_t,$$

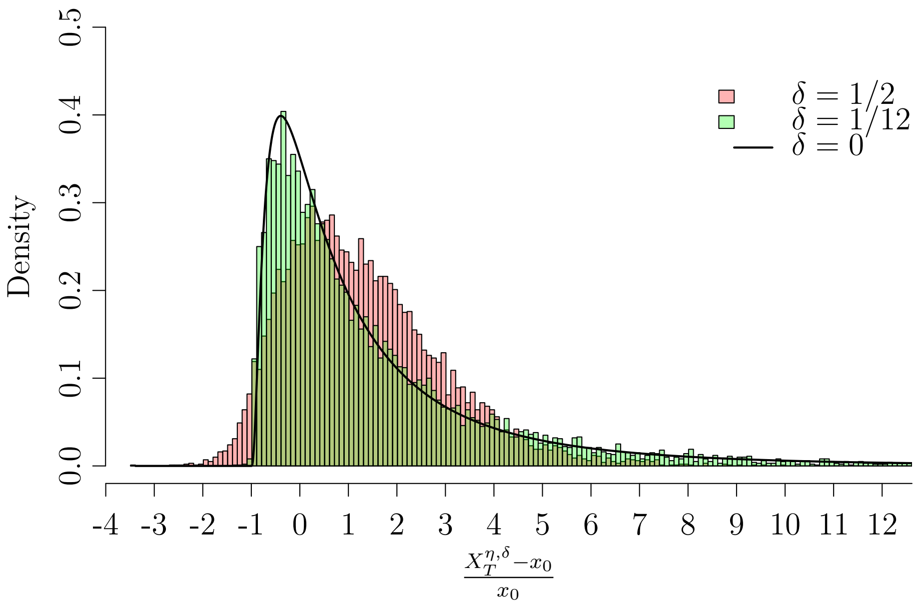

Figure 1.5: Consider $\eta(t,x) = x^+$, $T = 1$ (in years), and a constant $|\lambda_\cdot| = 1$. With a solid line, we plot the density of $\frac{X_T^{\eta} - x_0}{x_0}$. The histograms represent 10000 samples of the distribution of $\frac{X_T^{\eta,\delta} - x_0}{x_0}$ for $\delta \in \{1/2, 1/12\}$.

where the processes $\mu_\cdot$ and $\sigma_\cdot$ and the Brownian motion $W_\cdot$ coincide with the ones defined in the previous section.

Given a progressively measurable process $\pi_\cdot$, we define the associated discounted wealth process, $X_\cdot^\pi$, as the solution to

$$dX_t^\pi = X_t^\pi \pi_t' dR_t.$$

We denote by $\mathcal{A}$ the set of admissible trading strategies, defined as

$$\mathcal{A} = \left\{ \pi_\cdot \text{ progressively measurable} : X_\cdot^\pi > 0,\ \mathbb{E}\left[\sup_{s \in [0,\cdot]} (X_s^\pi)^2 \right] < \infty \right\}. \tag{1.28}$$

Let $U(t,x) = u(A_t, x)$ be a time–monotone forward performance process, where

$$A_t = \int_0^t |\lambda_s|^2 ds. \tag{1.29}$$

Then, the optimal portfolio for $U(t,x)$ is given in feedback form by

$$\pi^*(t,x) = \frac{r(A_t,x)}{x}\Sigma_t^{-1}\mu_t, \tag{1.30}$$

where $r(\cdot)$ is the local risk tolerance function, defined as

$$r(t,x) = -\frac{u_x(t,x)}{u_{xx}(t,x)}. \tag{1.31}$$

Furthermore, $r(\cdot)$ satisfies the fast diffusion equation

$$r_t + \frac{1}{2}r^2 r_{xx} = 0. \tag{1.32}$$

A well known characterisation of time–monotone forward performance processes establishes that there exists $\nu \in \mathcal{B}^+(\mathbb{R})$ such that

$$h(t,y) = \int_{\mathbb{R}} \frac{e^{xy-\frac{1}{2}x^2t} - 1}{x}\nu(dx) + h(0,0), \tag{1.33}$$

where the function $h(\cdot)$ is defined implicitly by

$$u_x(t,h(t,y)) = e^{-y+\frac{t}{2}}, \tag{1.34}$$

and $\mathcal{B}^+(\mathbb{R})$ is defined as in (11). In Section 1.6, we study the regularity of the local risk tolerance function $r(\cdot)$ in terms of the characterising measure ν.

Comparing the portfolio which is optimal for consecutive mean–variance problems, (1.12), and the one which is optimal for a time–monotone forward performance process, (1.30), we deduce the following result.

Theorem 3. *Suppose that Assumption 1 holds.*

Let $U(t,x) = u(A_t,x)$ *be a monotonically decreasing forward investment performance process; define its* $C^{1,2}$ *local risk tolerance function,* $r(\cdot)$*, as in* (1.31)*, and its associated measure,* $\nu \in \mathcal{B}^+(\mathbb{R})$*, as in* (1.33)*; denote by* $X^*_\cdot$ *its optimal* *wealth process.*

Define $\eta(t,x) : [0,\infty) \times \mathbb{R}^+ \times \Omega \to \mathbb{R}$ *by*

$$\eta(t,x) = r(A_t, x), \tag{1.35}$$

and let $X^{\eta,\delta}_\cdot$ *be the unique strong solution to* (1.13).

Then, for any bounded stopping time τ*, the following hold:*

(i) if $(\delta_n)_{n\in\mathbb{N}}$ *is such that* $\sum_n \delta_n^\beta < \infty$*, for some* $\beta > 0$*, then*

$$\sup_{t\in\Theta^{\delta_n}:t<\tau} \left| X_t^{\eta,\delta_n} - X_t^* \right| \to 0, \ a.s.,$$

as $n \to \infty$*;*

(ii) if the measure ν *is compactly supported, then*

$$\mathbb{E}\left[\sup_{t\in\Theta^{\delta}:t<\tau} \left| X_t^{\eta,\delta} - X_t^* \right|^p \right] \to 0,$$

as $\delta \to 0$*, for any* $p \geq 2$*.*

Proof. This result is a consequence of Theorem 2. We make a few remarks:

- the process $X^\eta_\cdot$ in Theorem 2 coincides with $X^*_\cdot$ and solves

$$dX_t^* = \eta(t, X_t^*)\mu_t' \Sigma_t^{-1}(\mu_t dt + \sigma_t dW_t);$$

- since, by definition, forward investment performance processes are defined for all times, we have $\tau^\eta = \infty$, where τ^η is the explosion time of $X^\eta_\cdot$, defined in (1.16).

- Assumption 2 holds since $r(\cdot)$ is $C^{1,2}$ and, by Assumption 1, $\lambda_\cdot$ is a bounded process;

- it follows from Theorems 8 and 9 in the appendix that, if the measure ν is compactly supported, then $\eta(\cdot)$ satisfies Assumption 2 with a global Lipschitz constant.

□

Theorem 3 shows that, for an appropriate choice of trade-off coefficients, mean–variance analysis yields, in the limit, the same wealth process as a time–monotone forward performance process. Furthermore, by Proposition 1, the optimal portfolios of the two procedures also coincide in the limit.

It is well known that forward investment performance processes yield time–consistent policies. Therefore, equation (1.35) provides a way to update risk tolerance in a time–consistent way. More precisely, Theorem 3 suggests that a mean–variance trade-off coefficient $\eta(\cdot)$ is (infinitesimally) time–consistent if it is given by $\eta(t, x) = r(A_t, x)$, where $r(\cdot)$ satisfies the fast diffusion PDE (1.32). With this choice of trade-off coefficient, we are able to apply mean–variance analysis forward in time without the restriction of a fixed investment horizon.

Remark 4. *There is a well known connection between expected utility maximisation and mean–variance analysis with a fixed horizon. More precisely, it is known that mean–variance analysis with a given trade–off coefficient yields the same optimal allocation as a specific quadratic utility. The result in Theorem 3 also establishes such a connection, but in the context where there is no fixed horizon and risk preferences evolve forward in time (as opposed to backwards). However, the connection in Theorem 3 is not as clear-cut as the one in the backwards case because it holds only infinitesimally.*

Remark 5. *In this chapter, we restrict ourselves to monotonically decreasing forward investment performance processes. It should be possible to relate mean–variance analysis to other (non–monotone) forward utilities. To this end, the results in El Karoui and Mrad (2010) on the construction of forward utilities with* *a given optimal allocation may be very useful. We leave this research direction for future work.*

Remark 6. *The result in Theorem 3 is strongest when the measure ν associated with the forward investment performance process has a compact support. This is not a very stringent assumption. Indeed, the best known monotonically decreasing forward investment performance processes have an associated measure comprised of a single Dirac mass:*

- *Exponential utility,* $U(t,x) = 1 - \gamma e^{-\frac{x}{\gamma}+\frac{A_t}{2}}$*, has associated measure* $\nu = \gamma\delta_0$*;*
- *Logarithmic utility,* $U(t,x) = \log(x) - \frac{A_t}{2}$*, has associated measure* $\nu = \delta_1$*;*
- *Power utility,* $U(t,x) = \frac{\gamma^{\frac{\gamma-1}{\gamma}}}{\gamma-1} x^{\frac{\gamma-1}{\gamma}} e^{-\frac{\gamma-1}{2}A_t}$*, has associated measure* $\nu = \delta_\gamma$*, where* $\gamma \in (1,+\infty)$*.*

Example 1. *In the case of logarithmic utility, we have*

$$u(t,x) = \log(x) - \frac{t}{2}.$$

In this case, the local risk tolerance function, $r(\cdot)$*, is constant across time,*

$$r(t,x) = x.$$

Therefore, an investor who performs mean–variance analysis with a trade-off coefficient given by $\eta(t,x) = x^+$ *is (infinitesimally) time–consistent. For this choice of utility, an illustration of the convergence result in Theorem 3 is given in Figure 1.5.*

1.5 A result on Euler's approximation scheme

In this section, we prove the convergence of an approximation scheme, similar to Euler's scheme. The result is stated in very general terms and, therefore, we believe it has an interest on its own. It is inspired by a similar result in Gyöngy (1998); in fact, Gyöngy (1998) is a particular case of the result in this section, when $\mu(\tilde{x},\tilde{t},x,t,\delta)=\mu(\tilde{x},t)$ and $\sigma(\tilde{x},\tilde{t},x,t,\delta)=\sigma(\tilde{x},t)$, below.

We work in a filtered probability space $(\Omega,\mathcal{F},\mathbb{P},(\mathcal{F}_t)_{t\geq 0})$. We assume that, in this space, there exists a standard $d-$dimensional Brownian motion, which we denote by $W_\cdot$.

The sequence of approximations we consider is denoted by $\left(X^\delta_\cdot\right)_{\delta>0}$. For each $\delta>0$, the process $X^\delta_\cdot$ takes values in $\mathbb{R}^N$ and is a solution of

$$\begin{cases} dX^\delta_t &= \mu\left(X^\delta_{\kappa^\delta_t},\kappa^\delta_t,X^\delta_t,t,\delta\right)dt+\sigma\left(X^\delta_{\kappa^\delta_t},\kappa^\delta_t,X^\delta_t,t,\delta\right)dW_t \\ X^\delta_0 &= x_0 \end{cases}, \tag{1.36}$$

where

$$\kappa^\delta_t=\delta\left\lfloor\frac{t}{\delta}\right\rfloor,$$

and $\mu(\tilde{x},\tilde{t},x,\cdot,\delta)$ and $\sigma(\tilde{x},\tilde{t},x,\cdot,\delta)$ are progressively measurable processes taking values in $\mathbb{R}^N$ and $\mathbb{R}^{N\times d}$, respectively. We assume that the functions $\mu(\cdot)$ and $\sigma(\cdot)$ are defined on $\tilde{D}\times[0,\infty)\times D\times[0,\infty)\times[0,\infty)\times\Omega$, where $\tilde{D},D\subset\mathbb{R}^N$ are open sets such that $x_0\in\tilde{D}\cap D$. The variables $(x,\tilde{x})$ and $(t,\tilde{t})$ are called spatial and time variables, respectively, and $\delta>0$ is the time step of the scheme. We refer to $\tilde{D}$ and D as the spatial domains for the equation (1.36).

Before continuing, we simplify the notation by defining

$$\mu^\delta_t=\mu\left(X^\delta_{\kappa^\delta_t},\kappa^\delta_t,X^\delta_t,t,\delta\right),$$

and adopt an analogous notation for σ^δ_t.

Remark 7. *We observe that, if $\mu(\tilde{x},\tilde{t},x,t,\delta)=\mu(\tilde{x},\tilde{t})$ and $\sigma(\tilde{x},\tilde{t},x,t,\delta)=\sigma(\tilde{x},\tilde{t})$, then* (1.36) *reduces to an interpolation of the classical Euler scheme.*

Also, if $\mu(\tilde{x},\tilde{t},x,t,\delta)=\mu(x,t)$ and $\sigma(\tilde{x},\tilde{t},x,t,\delta)=\sigma(x,t)$, then (1.36) *reduces* *to a stochastic differential equation (SDE).*

We show that, under certain assumptions on the regularity of the functions $\mu(\cdot)$ and $\sigma(\cdot)$, the sequence of processes $\left(X_\cdot^\delta\right)_{\delta>0}$ converges, as $\delta \to 0$, to a process, which we denote by $X_\cdot$. The limiting process $X_\cdot$ solves itself an SDE. We proceed as follows:

1. we prove an auxiliary result on the regularity of solutions to (1.36);
2. we obtain a first convergence result, assuming strong regularity of the functions $\mu(\cdot)$ and $\sigma(\cdot)$. In particular, we assume that these functions are globally defined, that is $\tilde{D} = D = \mathbb{R}^N$, and globally Lipschitz;
3. we relax the assumptions of the previous result to their local counterparts and prove convergence using a localisation argument.

Convergence under global assumptions

The main assumption required to prove the strong convergence of the sequence of processes $\left(X_\cdot^\delta\right)_{\delta>0}$ is the following.

Assumption 3 (Global Lipschitz)**.** *The functions $\mu(\cdot)$ and $\sigma(\cdot)$ are globally defined, that is $\tilde{D} = D = \mathbb{R}^N$.*

Furthermore, for each $p \geq 1$, there is $K > 0$ such that, almost surely,

$$\begin{aligned} |\mu(\tilde{x}, \tilde{t}, x, t, \delta) - \mu(\tilde{y}, \tilde{t}, y, t, \delta)|^p &\leq K(|x-y|^p + |\tilde{x} - \tilde{y}|^p), \\ |\mu(\tilde{x}, \tilde{t}, x, t, \delta)|^p &\leq K(1 + |x|^p + |\tilde{x}|^p), \end{aligned} \tag{1.37}$$

and analogous inequalities hold for $\sigma(\cdot)$.

We say that the constant K above is the Lipschitz constant of $\mu(\cdot)$ and $\sigma(\cdot)$.

Under Assumption 3, we deduce, from the classical theory of stochastic differential equations, that there exists a unique strong solution to (1.36). Furthermore, such solution has bounded moments of all orders. We do not prove the existence result here, but do prove, in the next lemma, the boundedness of its moments and an auxiliary result which is required in the sequel.

Lemma 1. *Suppose that Assumption 3 holds and let K be the Lipschitz constant of $\mu(\cdot)$ and $\sigma(\cdot)$.*

For each $\delta > 0$, let $X^\delta_\cdot$ be the unique strong solution to (1.36).

Then, for each $p \geq 2$ and $T \geq 0$, there exists a constant C, depending only on x_0, p, T and K, such that,

$$\mathbb{E}\left[\sup_{t\leq T} |X^\delta_t|^p\right] \leq C, \tag{1.38}$$

$$\mathbb{E}\left[\sup_{t\leq T} \left|X^\delta_t - X^\delta_{\kappa^\delta_t}\right|^p\right] \leq C\delta^{\frac{p}{2}}, \tag{1.39}$$

for all $\delta \leq 1$.

Proof. We start by establishing (1.38). By Jensen's inequality, and the Burkholder–Davis–Gundy inequality,

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\leq T} |X^\delta_t|^p\right] &= \mathbb{E}\left[\sup_{t\leq T} \left|x_0 + \int_0^t \mu^\delta_s ds + \int_0^t \sigma^\delta_s dW_s\right|^p\right] \\
&\leq 3^{p-1}\left(|x_0|^p + T^{p-1}\int_0^T \mathbb{E}\left[|\mu^\delta_s|^p\right] ds + C_p T^{\frac{p}{2}-1}\int_0^T \mathbb{E}\left[|\sigma^\delta_s|^p\right] ds\right) \\
&\leq C\left(1 + \int_0^T \mathbb{E}\left[|\mu^\delta_s|^p + |\sigma^\delta_s|^p\right] ds\right) \\
&\leq C\left(1 + \int_0^T \mathbb{E}\left[\left|X^\delta_{\kappa^\delta_s}\right|^p + |X^\delta_s|^p\right] ds\right) \\
&\leq C\left(1 + \int_0^T \mathbb{E}\left[\sup_{t\leq s} |X^\delta_t|^p\right] ds\right),
\end{aligned}$$

where the constant C above may change from line to line. Therefore, using Gronwall's inequality, we deduce that

$$\mathbb{E}\left[\sup_{t\leq T} |X^\delta_t|^p\right] \leq C.$$

Next, we prove that (1.39) holds. Suppose that $\delta \le 1$. Then, by Jensen's inequality and the Burkholder–Davis–Gundy inequality,

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\le T}\left|X_t^\delta - X_{\kappa_t^\delta}^\delta\right|^p\right] &\le 2^{p-1}\mathbb{E}\left[\sup_{t\le T}\left(\delta^p\left|\mu_t^\delta\right|^p + C_p\delta^{\frac{p}{2}}\left|\sigma_t^\delta\right|^p\right)\right] \\
&\le 2^{p-1}C\delta^{\frac{p}{2}}\mathbb{E}\left[\sup_{t\le T}\left(1+\left|X_t^\delta\right|^p\right)\right] \\
&\le C\delta^{\frac{p}{2}}.
\end{aligned}$$

□

To prove the convergence of $\left(X_\cdot^\delta\right)_{\delta>0}$, we need to account for two factors: the time discretisation in (1.36), and the presence of the time step δ as a variable in the functions $\mu(\cdot)$ and $\sigma(\cdot)$. For this reason, we introduce the next assumption on the regularity of $\mu(\cdot)$ and $\sigma(\cdot)$ with respect to the variables $\tilde{t}$ and δ.

Assumption 4 (Global Hölder)**.** *For each $p \ge 1$, there exist constants $K > 0$ and $\alpha > 0$, such that, almost surely,*

$$\begin{aligned}
|\mu(\tilde{x},\tilde{t},x,t,\delta) - \mu(\tilde{x},\tilde{s},x,t,\delta)|^p &\le K(1+|x|^p+|\tilde{x}|^p)(\tilde{t}-\tilde{s})^{\alpha p}, \\
|\mu(\tilde{x},\tilde{t},x,t,\delta) - \mu(\tilde{x},\tilde{t},x,t,0)|^p &\le K(1+|x|^p+|\tilde{x}|^p)\delta^{\alpha p},
\end{aligned} \tag{1.40}$$

and analogous inequalities hold for $\sigma(\cdot)$. Furthermore, the Hölder exponent, α, is independent of p.

We say that the constant K above is the Hölder constant of $\mu(\cdot)$ and $\sigma(\cdot)$.

Now, we define the limiting process of $\left(X_\cdot^\delta\right)_{\delta>0}$. With a slight abuse of notation, we denote

$$\mu(x,t) = \mu(x,t,x,t,0)$$

and

$$\mu_t = \mu(X_t,t),$$

and use an analogous notation for $\sigma(x,t)$ and σ_t. Let $X.$ be the solution of the SDE

$$\begin{cases} dX_t &= \mu(X_t,t)dt + \sigma(X_t,t)dW_t \\ X_0 &= x_0 \end{cases}. \tag{1.41}$$

We observe that, if Assumption 3 holds, then $X.$ is a well defined process with bounded moments of all orders.

Theorem 4. *Suppose that Assumptions 3 and 4 hold. For each $\delta > 0$, let $X^\delta_.$ and $X.$ denote the unique solutions to* (1.36) *and* (1.41)*, respectively.*

Then, for any $T > 0$, the following hold:

(i) for each $p \geq 2$ and $\delta \leq 1$,

$$\mathbb{E}\left[\sup_{t\leq T} \left|X^\delta_t - X_t\right|^p\right] \leq C\delta^{\left(\alpha\wedge\frac{1}{2}\right)p}, \tag{1.42}$$

where C is a constant depending only on x_0, p, T and K;

(ii) if $(\delta_n)_{n\in\mathbb{N}}$ is such that $\sum_n \delta_n^\beta < \infty$, for some $\beta > 0$, then

$$\sup_{t\leq T} \left|X^{\delta_n}_t - X_t\right| \to 0, \text{ a.s.}, \tag{1.43}$$

as $n \to \infty$.

Proof. We start by establishing (i). By Jensen's inequality and the Burkholder–Davis–Gundy inequality,

$$\begin{aligned}\mathbb{E}\left[\sup_{t\leq T} \left|X^\delta_t - X_t\right|^p\right] &= \mathbb{E}\left[\sup_{t\leq T} \left|\int_0^t \mu^\delta_s - \mu_s ds + \int_0^t \sigma^\delta_s - \sigma_s dW_s\right|^p\right] \\ &\leq 2^{p-1}\Bigg(T^{p-1}\int_0^T \mathbb{E}\left[\left|\mu^\delta_s - \mu_s\right|^p\right] ds \\ &\qquad\qquad + C_p T^{p/2-1}\int_0^T \mathbb{E}\left[\left|\sigma^\delta_s - \sigma_s\right|^p\right] ds\Bigg) \\ &\leq C\int_0^T \mathbb{E}\left[\left|\mu^\delta_s - \mu_s\right|^p\right] + \mathbb{E}\left[\left|\sigma^\delta_s - \sigma_s\right|^p\right] ds.\end{aligned}$$

In order to bound the right-hand side of the inequality above we proceed as follows. By Jensen's inequality, Lemma 1 and Assumptions 3 and 4, we have

$$\begin{aligned}
\mathbb{E}\left[\left|\mu_s^\delta - \mu_s\right|^p\right] \leq 4^{p-1}\mathbb{E}\Big[& \left|\mu\left(X^\delta_{\kappa^\delta_s}, \kappa^\delta_s, X^\delta_s, s, \delta\right) - \mu\left(X^\delta_{\kappa^\delta_s}, \kappa^\delta_s, X^\delta_s, s, 0\right)\right|^p \\
& + \left|\mu\left(X^\delta_{\kappa^\delta_s}, \kappa^\delta_s, X^\delta_s, s, 0\right) - \mu\left(X^\delta_{\kappa^\delta_s}, s, X^\delta_s, s, 0\right)\right|^p \\
& + \left|\mu\left(X^\delta_{\kappa^\delta_s}, s, X^\delta_s, s0\right) - \mu\left(X^\delta_s, s, X^\delta_s, s, 0\right)\right|^p \\
& + \left|\mu\left(X^\delta_s, s, X^\delta_s, s, 0\right) - \mu(X_s, s, X_s, s, 0)\right|^p \Big] \\
\leq C & \left(\delta^{\left(\alpha\wedge\frac{1}{2}\right)p} + \mathbb{E}\left[\sup_{t\leq s}\left|X^\delta_t - X_t\right|^p\right]\right).
\end{aligned}$$

The term $\mathbb{E}\left[\left|\sigma_s^\delta - \sigma_s\right|^p\right]$ can be bounded in a similar way.

Therefore,

$$\mathbb{E}\left[\sup_{t\leq T}\left|X^\delta_t - X_t\right|^p\right] \leq C\left(\delta^{\left(\alpha\wedge\frac{1}{2}\right)p} + \int_0^T \mathbb{E}\left[\sup_{t\leq s}\left|X^\delta_t - X_t\right|^p\right] ds\right),$$

and, by Gronwall's inequality, we conclude that

$$\mathbb{E}\left[\sup_{t\leq T}\left|X^\delta_t - X_t\right|^p\right] \leq C\delta^{\left(\alpha\wedge\frac{1}{2}\right)p}. \tag{1.44}$$

To establish *(ii)*, we make use of (1.44) for p large enough. Indeed, let $(\delta_n)_{n\in\mathbb{N}}$ and $\beta > 0$ be as in the hypothesis of the theorem. For a given $\gamma \in \left(0, \alpha\wedge\frac{1}{2}\right)$, let p be such that $p\left(\alpha\wedge\frac{1}{2} - \gamma\right) \geq \beta$. Then, by Markov's inequality and (1.44),

$$\begin{aligned}
\sum_n \mathbb{P}\left(\sup_{t\leq T}\left|X^{\delta_n}_t - X_t\right| \geq \delta_n^\gamma\right) &= \sum_n \mathbb{P}\left(\sup_{t\leq T}\left|X^{\delta_n}_t - X_t\right|^p \geq \delta_n^{\gamma p}\right) \\
&\leq C\sum_n \delta_n^{p\left(\alpha\wedge\frac{1}{2}-\gamma\right)} \\
&\leq C\sum_n \delta_n^\beta \\
&< \infty.
\end{aligned}$$

Therefore, $\xi < \infty$, almost surely, where

$$\xi = \sup_n \frac{\sup_{t \leq T} \left| X_t^{\delta_n} - X_t \right|}{\delta_n^\gamma}.$$

Finally, by the definition of ξ,

$$\sup_{t \leq T} \left| X_t^{\delta_n} - X_t \right| \leq \xi \delta_n^\gamma, \ a.s., \tag{1.45}$$

for all n, and we conclude that (1.43) holds. $\square$

Remark 8. *We observe that, from* (1.45)*, the rate of convergence in* (1.43) *is* γ*, for any* $\gamma \in \left(0, \alpha \wedge \frac{1}{2}\right)$*.*

Convergence under local assumptions

The Lipschitz and Hölder properties, (1.37) and (1.40), are global, in the sense that they must hold for all $\tilde{x}, x, \tilde{y}, y \in \mathbb{R}^N$. These can be made local, as in the following assumptions.

Assumption 5 (Local Lipschitz)**.** *We recall that* $\tilde{D}$ *and* D *are the spatial domains of the functions* $\mu(\cdot)$ *and* $\sigma(\cdot)$*.*

For each $p \geq 1$ *and compact sets* $\tilde{E} \subset \tilde{D}$ *and* $E \subset D$*, there exists* $K > 0$ *such that, for all* $(\tilde{x}, x, \tilde{y}, y) \in \tilde{E} \times E \times \tilde{E} \times E$*, almost surely,*

$$\begin{aligned} |\mu(\tilde{x}, \tilde{t}, x, t, \delta) - \mu(\tilde{y}, \tilde{t}, y, t, \delta)|^p &\leq K(|x-y|^p + |\tilde{x} - \tilde{y}|^p), \\ |\mu(\tilde{x}, \tilde{t}, x, t, \delta)|^p &\leq K(1 + |x|^p + |\tilde{x}|^p), \end{aligned} \tag{1.46}$$

and analogous inequalities hold for $\sigma(\cdot)$*.*

We say that the Lipschitz constant K *is global if it does not depend on* $\tilde{E}$ *and* E*, i.e. if* (1.46) *holds for all* $(\tilde{x}, x, \tilde{y}, y) \in \tilde{D} \times D \times \tilde{D} \times D$*.*

Assumption 6 (Local Hölder)**.** *We recall that* $\tilde{D}$ *and* D *are the spatial domains of the functions* $\mu(\cdot)$ *and* $\sigma(\cdot)$*.*

For each $p \geq 1$ and compact sets $\tilde{E} \subset \tilde{D}$ and $E \subset D$, there exists $K > 0$ and $\alpha > 0$ such that, for all $(\tilde{x}, x) \in \tilde{E} \times E$, almost surely,

$$\begin{aligned} |\mu(\tilde{x}, \tilde{t}, x, t, \delta) - \mu(\tilde{x}, \tilde{s}, x, t, \delta)|^p &\leq K(1 + |x|^p + |\tilde{x}|^p)(\tilde{t} - \tilde{s})^{\alpha p}, \\ |\mu(\tilde{x}, \tilde{t}, x, t, \delta) - \mu(\tilde{x}, \tilde{t}, x, t, 0)|^p &\leq K(1 + |x|^p + |\tilde{x}|^p)\delta^{\alpha p}, \end{aligned} \tag{1.47}$$

and analogous inequalities hold for $\sigma(\cdot)$. Furthermore, the Hölder exponent, α, does not depend on p, $\tilde{E}$ and E.

We say that the Hölder constant K is global if it does not depend on $\tilde{E}$ and E, i.e. if (1.47) holds for all $(\tilde{x}, x) \in \tilde{D} \times D$.

The McShane–Whitney extension theorem shows that any Lipschitz function $f(x) : E \to \mathbb{R}$ can be extended to a Lipschitz function $F(x) : \mathbb{R}^N \to \mathbb{R}$ with the same Lipschitz constant — see Whitney (1934) and McShane (1934). Next, we adapt the construction in McShane (1934) to show that functions satisfying the regularity conditions (1.46) and (1.47) can also be extended while preserving the Lipschitz and Hölder constants (up to multiplication by a fixed constant).

Theorem 5 (McShane–Whitney extension theorem)**.** *Let $E \subset \mathbb{R}^N$ be a star domain with vantage point e_0, and $f(x,t) : E \times [0, \infty) \to \mathbb{R}$. Suppose that there exists $K > 0$ and $\alpha > 0$ such that*

$$\begin{aligned} |f(x,t) - f(y,t)| &\leq K|x - y| \\ |f(x,t)| &\leq K(1 + |x|) \\ |f(x,t) - f(x,s)| &\leq K(1 + |x|)|t - s|^\alpha \end{aligned} \tag{1.48}$$

for any $x, y \in E$ and $t, s \geq 0$.

Then, there exists $F(x,t) : \mathbb{R}^N \times [0, \infty) \to \mathbb{R}$ such that

- *(i) $F(x,t) = f(x,t)$, for any $x \in E$, and*
- *(ii) $F(\cdot)$ satisfies (1.48) with K replaced by $3K + 4|e_0|$.*

Proof. For any $e \in E$, $x \in \mathbb{R}^N$ and $t \geq 0$, we define

$$f_e(x,t) = f(e,t) + K|x-e|,$$

and

$$F(x,t) = \inf_{e \in E \colon |e-e_0| \leq |x-e_0|} f_e(x,t).$$

Suppose that $x \in E$. On the one hand,

$$F(x,t) \leq f_x(x,t) = f(x,t).$$

On the other hand, for any $e \in E$,

$$f_e(x,t) = f(e,t) + K|x-e| \geq f(e,t) + |f(x,t) - f(e,t)| \geq f(x,t).$$

Therefore, we deduce that $F(x,t) = f(x,t)$.

Let $x \in \mathbb{R}^N$ be arbitrary. Then, for any $e, d \in E$,

$$\begin{aligned} f_e(x,t) &= f(e,t) + K|d-e| + K|x-e| - K|d-e| \\ &\geq f(d,t) + K(|x-e| - |d-e|) \\ &\geq f(d,t) - K(|x| + |d|). \end{aligned}$$

Taking the infimum with respect to e in the inequality above, we deduce that $F(x,t) > -\infty$. In other words, $F(x,t)$ is well defined for all $x \in \mathbb{R}^N$.

Now, we show that (ii) holds. We observe that, for any $e, d \in E$,

$$\begin{aligned} f_e(x,t) &= f_d(y,t) + f_e(x,t) - f_d(y,t) \\ &= f_d(y,t) + f(e,t) - f(d,t) + K(|x-e| - |y-d|) \\ &\leq f_d(y,t) + K|e-d| + K(|x-e| - |y-d|) \\ &\leq f_d(y,t) + K(|x-y| + 2|e-d|). \end{aligned}$$

Taking the infimum in the inequality above with respect to (e,d) in the set

$$\Lambda = \left\{(e,d) \in E \times E : |e-e_0| \leq |x-e_0| \text{ and } |e-d| \leq |x-y|\right\},$$

we deduce that

$$F(x,t) \leq \inf_{(e,d)\in\Lambda} f_d(y,t) + 3K|x-y|.$$

Furthermore, since E is a star domain, we have

$$\inf_{(e,d)\in\Lambda} f_d(y,t) \leq F(y,t).$$

Therefore, we conclude that

$$F(x,t) - F(y,t) \leq 3K|x-y|.$$

Swapping the roles of x and y in the inequality above, we deduce that the first inequality in (1.48) holds.

Next, we observe that, for any $x \in \mathbb{R}^N$ and $e \in E$ such that $|e - e_0| \leq |x - e_0|$,

$$\begin{aligned} f_e(x,t) &= f(e,t) + K|x - e| \\ &\leq K(1 + 2|e| + |x|) \\ &\leq K(1 + 2|e_0| + 2|e - e_0| + |x|) \\ &\leq K(1 + 2|e_0| + 2|x - e_0| + |x|) \\ &\leq K(1 + 4|e_0| + 3|x|) \\ &\leq (3K + 4|e_0|)(1 + |x|). \end{aligned}$$

Therefore, taking the infimum with respect to e in the inequality above, we deduce that the second inequality in (1.48) holds.

Finally, let $x \in \mathbb{R}^N$ and $e \in E$ be such that $|e - e_0| \leq |x - e_0|$. Then,

$$\begin{aligned} f_e(x,t) &= f_e(x,s) + f_e(x,t) - f_e(x,s) \\ &\leq f_e(x,s) + K(1 + |e|)|t - s|^\alpha \\ &\leq f_e(x,s) + K(1 + 2|e_0| + |x|)|t - s|^\alpha. \end{aligned}$$

Since e is arbitrary, we deduce that

$$F(x,t) - F(x,s) \leq (K + 2|e_0|)(1 + |x|)|t - s|^\alpha.$$

Swapping t and s in the inequality above we conclude that the third inequality in (1.48) holds. $\square$

Remark 9. *The original theorem of McShane and Whitney does not assume that the domain of $f(\cdot)$ is a star domain. In our case, we use such an assumption to prove that the third inequality in* (1.48) *can be extended along with the first two.*

If the coefficients of an SDE are locally Lipschitz or have a restricted spatial domain, it is well known that a unique strong solution exists but is only defined up to explosion times. In the case of equation (1.36), an analogous result holds, as shown in Theorem 6 below. This result is proved using standard localisation arguments. In particular it makes use of a sequence of functions $(\mu_k(\cdot))_{k\in\mathbb{N}}$ and $(\sigma_k(\cdot))_{k\in\mathbb{N}}$ which are defined globally and approximate $\mu(\cdot)$ and $\sigma(\cdot)$ as in the following lemma.

Lemma 2. *Suppose that Assumptions 5 and 6 hold. Let $(\tilde{D}_k)_{k\in\mathbb{N}}$ and $(D_k)_{k\in\mathbb{N}}$ be sequences of open sets such that $\tilde{D}_k \subset \tilde{D}$, $D_k \subset D$ and*

(a) the closure of $\tilde{D}_k$ (respectively, D_k), is a compact set contained in $\tilde{D}_{k+1}$ (respectively, D_{k+1});

(b) $\tilde{D}_k \nearrow \tilde{D}$ and $D_k \nearrow D$, that is

$$\bigcup_k \tilde{D}_k = \tilde{D} \text{ and } \bigcup_k D_k = D.$$

Then, for each k, there exist functions $\mu_k(\cdot)$ and $\sigma_k(\cdot)$ defined on $\mathbb{R}^N \times [0,\infty) \times \mathbb{R}^N \times [0,\infty) \times [0,\infty) \times \Omega \to \mathbb{R}^N$ such that

(i) for all $(\tilde{x}, x) \in \tilde{D}_k \times D_k$,

$$\mu_k(\tilde{x}, \tilde{t}, x, t, \delta) = \mu(\tilde{x}, \tilde{t}, x, t, \delta),$$
$$\sigma_k(\tilde{x}, \tilde{t}, x, t, \delta) = \sigma(\tilde{x}, \tilde{t}, x, t, \delta);$$

(ii) $\mu_k(\cdot)$ and $\sigma_k(\cdot)$ satisfy Assumptions 3 and 4.

Furthermore, if $\tilde{D}$ and D are star domains, and the Lipschitz and Hölder constants in Assumptions 5 and 6 are global, then the functions $\mu_k(\cdot)$ and $\sigma_k(\cdot)$ can be taken to be independent of k.

Proof. Functions $\mu_k(\cdot)$ and $\sigma_k(\cdot)$, such that (i) and (ii) hold, can be constructed using cutoff functions. To show that (*ii*) holds, we use the inequality

$$|f(x)g(x) - f(y)g(y)| \leq |f(x)||g(x) - g(y)| + |g(y)||f(x) - f(y)|,$$

valid for any two functions $f(\cdot) : \mathbb{R}^N \to \mathbb{R}^M$ and $g(\cdot) : \mathbb{R}^N \to \mathbb{R}$.

Finally, when $\tilde{D}$ and D are star domains and the Lipschitz and Hölder constants in Assumptions 5 and 6 are global, we use Theorem 5 to extend $\mu(\cdot)$ and $\sigma(\cdot)$ to functions $\mu_0(\tilde{x}, \tilde{t}, x, t, \delta)$ and $\sigma_0(\tilde{x}, \tilde{t}, x, t, \delta)$ defined for all $\tilde{x}, x \in \mathbb{R}^N$. □

Theorem 6. *Let $\tilde{D}, D \subset \mathbb{R}^N$ be open sets. For each $(\tilde{x}, x) \in \tilde{D} \times D$, $\tilde{t} \geq 0$ and $\delta > 0$, let $\mu(\tilde{x}, \tilde{t}, x, \cdot, \delta)$ and $\sigma(\tilde{x}, \tilde{t}, x, \cdot, \delta)$ be progressively measurable processes. Suppose that $\mu(\cdot)$ and $\sigma(\cdot)$ satisfy Assumption 5.*

Let $(\tilde{D}_k)_{k\in\mathbb{N}}$ and $(D_k)_{k\in\mathbb{N}}$ be such that $\tilde{D}_k \nearrow \tilde{D}$ and $D_k \nearrow D$. For each k, let $\mu_k(\cdot)$ and $\sigma_k(\cdot)$ be functions with domain $\mathbb{R}^N \times [0, \infty) \times \mathbb{R}^N \times [0, \infty) \times [0, \infty) \times \Omega$ such that they satisfy Assumption 3 and

$$\mu_k(\tilde{x}, \tilde{t}, x, t, \delta) = \mu(\tilde{x}, \tilde{t}, x, t, \delta),$$
$$\sigma_k(\tilde{x}, \tilde{t}, x, t, \delta) = \sigma(\tilde{x}, \tilde{t}, x, t, \delta),$$

for all $(\tilde{x}, x) \in \tilde{D}_k \times D_k$.

For a fixed $\delta > 0$, and for each $k \in \mathbb{N}$, let the process $X^k_\cdot$ be the unique strong solution to

$$\begin{cases} dX_t &= \mu\left(X_{\kappa^\delta_t}, \kappa^\delta_t, X_t, t, \delta\right) dt + \sigma\left(X_{\kappa^\delta_t}, \kappa^\delta_t, X_t, t, \delta\right) dW_t \\ X_0 &= x_0 \end{cases}, \tag{1.49}$$

when $\mu(\cdot)$ and $\sigma(\cdot)$ are replaced by $\mu_k(\cdot)$ and $\sigma_k(\cdot)$, respectively.

Let $(\tau_k)_{k\in\mathbb{N}}$ be a sequence of exit times, defined as

$$\tau^k = \inf\ \left\{t : X^k_t \notin D_k \ \text{ or } \ X^k_{\kappa^\delta_t} \notin \tilde{D}_k\right\}.$$

Then, $\tau^k \nearrow \tau := \lim \tau^k$, *and*

$$X_t = \lim_k X_t^k, \;\; t \in [0, \tau),$$

is the unique solution of (1.49) *in* $[0, \tau)$. *Furthermore, the following hold:*

(i) the stopping time τ *is maximal, in the sense that the solution* $X_.$ *cannot be extended to any later stopping time;*

(ii) for each stopping time $\hat{\tau}$, *such that* $\hat{\tau} < \tau$, $X_.^k$ *eventually coincides with* $X_.$ *in* $[0, \hat{\tau}]$, *that is*

$$\mathbb{P}\left(\bigcup_{m=1}^{\infty}\bigcap_{k=m}^{\infty}\left\{\sup_{t\leq\hat{\tau}}\left|X_t^k - X_t\right| = 0\right\}\right) = 1; \tag{1.50}$$

(iii) if the Lipschitz constant of $\mu_k(\cdot)$ *and* $\sigma_k(\cdot)$ *does not depend on* k, *then, for each* $p \geq 2$ *and* $T > 0$, *there exists a constant* C, *independent of* δ, *such that*

$$\mathbb{E}\left[\sup_{t<\tau\wedge T} |X_t|^p\right] < C, \tag{1.51}$$

and, for each bounded stopping time $\hat{\tau}$ *such that* $\hat{\tau} < \tau$,

$$\lim_{k\to\infty} \mathbb{E}\left[\sup_{t<\hat{\tau}} \left|X_t^k - X_t\right|^p\right] = 0. \tag{1.52}$$

Proof. Let $X_.^k$ and τ^k be as in the hypothesis of the theorem. Then, since $(\tilde{D}_k)_{k\in\mathbb{N}}$ and $(D_k)_{k\in\mathbb{N}}$ are increasing sequences of sets, we have $X_t^k = X_t^{k+1}$ for $t \in [0, \tau^k]$. Thus, $\tau^k \nearrow \tau$, and

$$X_t = \lim_k X_t^k$$

is well defined for $t \in [0, \tau)$. Note that, by construction, $X_.$ solves (1.49) in $[0, \tau)$.

To establish (i), we observe that one of the following three assertions must hold:

- $\tau = \infty$;
- $\tau < \infty$ and $\lim_{t \to \tau} X_t$ does not exist;
- $\tau < \infty$ and $\lim_{t \to \tau} X_t$ exists but is not in D.

Result (1.50) follows immediately from the properties of $X^k_\cdot$. Indeed, let $\hat{\tau}$ be a stopping time such that $\hat{\tau} < \tau$. The graph of X_t, for $t \in [0, \hat{\tau}]$, is a compact subset of D, and, therefore, for each ω, there is k such that $X_t(\omega) \in D_k$, for all $t \in [0, \hat{\tau}(\omega)]$. Thus, we have

$$\sup_{t \leq \hat{\tau}} \left| X^k_t - X_t \right| = 0.$$

almost surely.

Now, we show that *(iii)* holds. Suppose that the Lipschitz constant of $\mu_k(\cdot)$ and $\sigma_k(\cdot)$ does not depend on k. Then, by (1.38), for any $p \geq 2$, there exists a constant C, independent of k and δ, such that

$$\mathbb{E}\left[\sup_{t \leq T} \left|X^k_t\right|^p\right] \leq C. \tag{1.53}$$

Therefore, from the monotone convergence theorem, we deduce that

$$\begin{aligned} \mathbb{E}\left[\sup_{t < \tau \wedge T} \left|X_t\right|^p\right] &= \mathbb{E}\left[\lim_k \sup_{t < \tau^k \wedge T} \left|X_t\right|^p\right] \\ &= \lim_k \mathbb{E}\left[\sup_{t < \tau^k \wedge T} \left|X^k_t\right|^p\right] \\ &\leq C. \end{aligned}$$

Finally, to deduce (1.52), we can use (1.50), (1.51), and (1.53), together with the dominated convergence theorem. □

Next, we prove a local version of Theorem 4.

Theorem 7. *Suppose that Assumptions 5 and 6 hold. Let $X_\cdot$ denote the unique solution to (1.41) and τ its explosion time, as given by Theorem 6. Similarly, for each δ, let $X_\cdot^\delta$ denote the unique solution to (1.36), and τ^δ its explosion time.*

Then,

$$\mathbb{P}\left(\tau^\delta < \tau\right) \to 0, \tag{1.54}$$

as $\delta \to 0$.

Furthermore, for any bounded stopping time $\hat{\tau}$ such that $\hat{\tau} < \tau$, the following hold:

(i) if $(\delta_n)_{n\in\mathbb{N}}$ is such that $\sum_n \delta_n^\beta < \infty$, for some $\beta > 0$, then

$$\liminf_n \tau^{\delta_n} \geq \tau, \ a.s., \tag{1.55}$$

and

$$\sup_{t<\hat{\tau}\wedge\tau^{\delta_n}} \left|X_t^{\delta_n} - X_t\right| \to 0, \ a.s., \tag{1.56}$$

as $n \to \infty$;

(ii) if the spatial domains $\tilde{D}$ and D are star domains, and the Lipschitz and Hölder constants in Assumptions 5 and 6 are global, then

$$\mathbb{E}\left[\sup_{t<\hat{\tau}\wedge\tau^\delta} \left|X_t^\delta - X_t\right|^p\right] \to 0, \tag{1.57}$$

as $\delta \to 0$, for any $p \geq 2$.

Proof. We recall that by construction of $X_\cdot$ and $X_\cdot^\delta$, see Theorem 6 and Lemma 2, there exist increasing sequences of sets $(\tilde{D}_k)_{k\in\mathbb{N}}$ and $(D_k)_{k\in\mathbb{N}}$, and sequences of functions $(\mu_k(\cdot))_{k\in\mathbb{N}}$ and $(\sigma_k(\cdot))_{k\in\mathbb{N}}$ satisfying Assumptions 3 and 4, such that:

- $X_\cdot$ and $X_\cdot^\delta$ are defined as

$$X_\cdot = \lim_k X_\cdot^k \text{ and } X_\cdot^\delta = \lim_k X_\cdot^{\delta,k},$$

where $X^{\delta,k}_{\cdot}$ and $X^k_{\cdot}$ are the solutions of (1.36) and (1.41), respectively, when $\mu(\cdot)$ and $\sigma(\cdot)$ are replaced by $\mu^k(\cdot)$ and $\sigma^k(\cdot)$;

- if we define

$$\tau^k = \inf\left\{t : X^k_t \notin D_k \text{ or } X^k_{\kappa^\delta_t} \notin \tilde{D}_k\right\},$$
$$\tau^{\delta,k} = \inf\left\{t : X^{\delta,k}_t \notin D_k \text{ or } X^{\delta,k}_{\kappa^\delta_t} \notin \tilde{D}_k\right\},$$

  then $\tau^k \nearrow \tau$, $\tau^{\delta,k} \nearrow \tau^\delta$, and

$$X_{\cdot\wedge\tau^k} = X^k_{\cdot\wedge\tau^k} \text{ and } X^\delta_{\cdot\wedge\tau^{\delta,k}} = X^{\delta,k}_{\cdot\wedge\tau^{\delta,k}}.$$

We first show that (1.55) holds. By Theorem 4, we have that, for each $k \in \mathbb{N}$ and $T \geq 0$,

$$\sup_{t\leq T}\left|X^{\delta_n,k}_t - X^k_t\right| \to_n 0. \tag{1.58}$$

Thus, for each k, T, and ω, there is N such that

$$\tau^{k-1} \wedge T \leq \tau^{\delta_n,k} \wedge T \leq \tau^{\delta_n} \wedge T,$$

for all $n \geq N$. Therefore, taking the limit, first in n and then in k, we deduce that

$$\tau \wedge T \leq \liminf_n \tau^{\delta_n} \wedge T.$$

Finally, taking the limit as $T \to \infty$, we conclude that

$$\tau \leq \liminf_n \tau^{\delta_n}.$$

Next, we show that (1.56) holds. Let $\hat{\tau} < \tau$ be a bounded stopping time. For each ω, since $(X_t)_{t\in[0,\hat{\tau}]}$ is a compact set, there is k, such that $X_t \in D_{k-1}$, for all $t \in [0, \hat{\tau}]$. Therefore, by (1.58), there is N such that, for all $n \geq N$, $X^{\delta_n}_t \in D_k$, for all $t \in [0, \hat{\tau}]$. Thus, we conclude that

$$\sup_{t < \hat{\tau} \wedge \tau^{\delta_n}} \left| X_t^{\delta_n} - X_t \right| = \sup_{t < \hat{\tau} \wedge \tau^{\delta_n}} \left| X_t^{\delta_n, k} - X_t^k \right| \to 0,$$

as $n \to \infty$.

Now, suppose that the Lipschitz and Hölder constants in Assumptions 5 and 6 are global. Then, by (1.42), we have that, for each $k \in \mathbb{N}$,

$$\mathbb{E}\left[\sup_{t \le \hat{\tau}} \left| X_t^{\delta, k} - X_t^k \right|^p \right] \le C \delta^{\left(\alpha \wedge \frac{1}{2}\right) p},$$

where the constant C does not depend on k. Therefore, by the monotone convergence theorem,

$$\begin{aligned} \mathbb{E}\left[\sup_{t < \hat{\tau} \wedge \tau^{\delta}} \left| X_t^{\delta} - X_t \right|^p \right] &= \mathbb{E}\left[\lim_k \sup_{t < \hat{\tau} \wedge \tau^{\delta, k} \wedge \tau^k} \left| X_t^{\delta} - X_t \right|^p \right] \\ &= \lim_k \mathbb{E}\left[\sup_{t < \hat{\tau} \wedge \tau^{\delta, k} \wedge \tau^k} \left| X_t^{\delta, k} - X_t^k \right|^p \right] \\ &\le C \delta^{\left(\alpha \wedge \frac{1}{2}\right) p}, \end{aligned}$$

and (1.57) follows immediately.

Finally, we show that (1.54) holds. Let $(\delta_n)_{n \in \mathbb{N}}$ be an arbitrary sequence such that $\delta_n \to 0$. Then, there exists a subsequence $(\delta_{u_n})_{n \in \mathbb{N}}$ such that $\sum_n \delta_{u_n} < \infty$. Therefore, from (1.55), we deduce that

$$\liminf_n \tau^{\delta_{u_n}} \ge \tau.$$

In particular, it follows that

$$\mathbb{1}_{\left\{\tau^{\delta_{u_n}} < \tau\right\}} \to 0, \ \ a.s.,$$

and, by the dominated convergence theorem,

$$\mathbb{P}\left(\tau^{\delta_{u_n}} < \tau\right) \to 0.$$

Since the sequence $(\delta_n)_{n\in\mathbb{N}}$ is arbitrary, we conclude that (1.54) holds. $\square$

1.6 On the regularity of the local risk tolerance function

Recall from Section 1.4 that the local risk tolerance function associated with a time–monotone forward performance process, $r(t,x) : [0,\infty) \times \mathbb{R}^+ \to \mathbb{R}$, satisfies the fast diffusion PDE (1.32). Furthermore, it is known that

$$r(t, h(t,y)) = h_y(t,y), \tag{1.59}$$

where the function $h(t,y) : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ admits the integral representation (1.33).

Remark 10. *Any function $h(\cdot)$ that satisfies* (1.33) *must be C^∞. This observation is a consequence of the Leibniz integral rule. Indeed, we observe that the $(n+1)^{th}$ derivative of the integrand in* (1.33) *satisfies the inequality*

$$|x|^n e^{xy-\frac{1}{2}x^2t} \leq e^{xy} + e^{x(y+n)} + e^{x(y-n)}.$$

Furthermore, since $\nu \in \mathcal{B}^+(\mathbb{R})$,

$$\int_{\mathbb{R}} e^{xy}\nu(dx) < \infty,$$

for any $y \in \mathbb{R}$.

Remark 11. *The definition of $h(\cdot)$ in* (1.34) *implies that its range coincides with* *the domain of* $r(t,\cdot)$*, that is* $\mathbb{R}^+$*.*

In this section, we study the regularity of the function $r(\cdot)$. We start by obtaining necessary and sufficient conditions under which the function $r(t,\cdot)$ is globally Lipschitz. We observe that, since the function $r(t,\cdot)$ is continuously differentiable, it is locally Lipschitz.

Theorem 8. *Let* $r(t,x) : [0,\infty)\times\mathbb{R}^+ \to \mathbb{R}$ *be a function satisfying* (1.59)*, and let* $\nu \in \mathcal{B}^+(\mathbb{R})$ *be the measure associated with* $r(\cdot)$ *through* (1.33)*. Then, the function* $r(t,\cdot)$ *is globally Lipschitz, with Lipschitz constant* K*, if and only if the measure* ν *is concentrated in* $[-K,K]$*.*

Proof. First, we observe that the function $r(t,\cdot)$ has Lipschitz constant K if and only if, for each $t \geq 0$ and $x \in \mathbb{R}^+$,

$$|r_x(t,x)| \leq K.$$

By (1.59), we have that

$$r_x(t,h(t,y)) = \frac{h_{yy}(t,y)}{h_y(t,y)}.$$

Therefore, since the range of $h(t,\cdot)$ is $\mathbb{R}^+$, $r(t,\cdot)$ has Lipschitz constant K if and only if, for all $y \in \mathbb{R}$,

$$\left|\frac{h_{yy}(t,y)}{h_y(t,y)}\right| \leq K. \tag{1.60}$$

Next, we assume that the measure ν is concentrated in $[-K,K]$. Then,

$$\begin{aligned} |h_{yy}(t,y)| &= \left|\int_{\mathbb{R}} x e^{xy-\frac{1}{2}x^2t}\nu(dx)\right| \\ &\leq K\int_{\mathbb{R}} e^{xy-\frac{1}{2}x^2t}\nu(dx) \\ &= Kh_y(t,y), \end{aligned}$$

and we easily conclude that $r(t,\cdot)$ has Lipschitz constant K.

To show the opposite direction, we argue by contradiction. To this end, suppose that $\nu((K,+\infty))>0$, and let ε be such that $\nu([K+\varepsilon,+\infty))>0$. We show that, in this case, $r(0,\cdot)$ does not have Lipschitz constant K. To ease the notation we introduce the function

$$f(y)=h_y(0,y)=\int_{\mathbb{R}} e^{xy}\nu(dx).$$

Then, for each $y\geq 1$, we have

$$\begin{aligned}
f'(y) &= \int_{\mathbb{R}} xe^{xy}\nu(dx)\\
&= \int_0^{+\infty} xe^{xy}\nu(dx) - \int_{-\infty}^0 |x|e^{-|x|y}\nu(dx)\\
&\geq Ke^{Ky}\nu([K,+\infty)) - \int_{-\infty}^0 |x|e^{-|x|}\nu(dx).
\end{aligned}$$

Therefore, we conclude that

$$\lim_{y\to+\infty} f'(y)=+\infty,$$

and, in particular, $f'(y)>0$, for all y's large enough.

Then, we easily deduce that, for y large enough,

$$\left|\frac{f'(y)}{f(y)}\right|\leq K \Leftrightarrow f'(y)\leq Kf(y)$$

if and only if

$$\int_{\mathbb{R}} (x-K)e^{xy}\nu(dx)\leq 0. \tag{1.61}$$

Note, however, that

$$\begin{aligned}\int_{\mathbb{R}}(x-K)e^{xy}\nu(dx) = &\int_{K}^{+\infty}(x-K)e^{xy}\nu(dx)-\int_{-\infty}^{K}(K-x)e^{xy}\nu(dx)\\ \geq &\,\varepsilon e^{(K+\varepsilon)y}\nu([K+\varepsilon,+\infty))+\\ &-\sup_{x\in(-\infty,K]}\left((K-x)e^{xy}\right)\nu((-\infty,K])\\ = &\,\varepsilon e^{(K+\varepsilon)y}\nu([K+\varepsilon,+\infty))-\frac{e^{Ky-1}}{y}\nu((-\infty,K]),\end{aligned}$$

and, therefore,

$$\lim_{y\to\infty}\int_{\mathbb{R}}(x-K)e^{xy}\nu(dx)=+\infty,$$

which contradicts (1.61). We conclude that $r(0,\cdot)$ does not have Lipschitz constant K.

Finally, we suppose that $\nu((-\infty,K))>0$. Let $\tilde{\nu}$ denote the push forward of the measure ν by the map $x\mapsto -x$, that is, $\tilde{\nu}(B)=\nu(-B)$, for any $B\in\mathcal{B}(\mathbb{R})$. Then,

$$\begin{aligned}f(y)&=\int_{\mathbb{R}}e^{xy}\nu(dx)\\ &=\int_{\mathbb{R}}e^{-xy}\tilde{\nu}(dx)=\tilde{f}(-y),\end{aligned}$$

where the function $\tilde{f}(y):\mathbb{R}\to D$ is defined as $\tilde{f}(y)=\int_{\mathbb{R}}e^{xy}\tilde{\nu}(dx)$. Furthermore,

$$\tilde{\nu}((K,+\infty))=\nu((-\infty,-K))>0.$$

Therefore, we can repeat the argument used in the case where $\nu((K,+\infty))>0$ to deduce that

$$\left|\frac{f'(-y)}{f(-y)}\right|=\left|\frac{\tilde{f}'(y)}{\tilde{f}(y)}\right|>K,$$

for y large enough. We conclude that $r(0,\cdot)$ does not have Lipschitz constant K.

This ends the proof. $\square$

Next, we study the regularity of the function $r(\cdot, x)$.

Theorem 9. *Let* $r(t,x) : [0,\infty) \times \mathbb{R}^+ \to \mathbb{R}$ *be a function satisfying* (1.59)*, and let* $\nu \in \mathcal{B}^+(\mathbb{R})$ *be the measure associated with* $r(\cdot)$ *through* (1.33)*.*

Suppose that the measure ν *is concentrated in* $[-K, K]$*. Then, for each* $T \geq 0$*, there is* $C \geq 0$ *such that*

$$|r(t,x) - r(s,x)| \leq C(1 + |x|)|t - s|,$$

for all $t, s \in [0, T]$*.*

Proof. Let $t \in [0,T]$, for T fixed but arbitrary, and let K be as in the hypothesis of the theorem. By (1.59), we have that

$$\begin{aligned} r_t(t, h(t,y)) &= h_{ty}(t,y) - r_x(t, h(t,y))h_t(t,y) \\ &= h_{yyy}(t,y) - r_x(t, h(t,y))h_{yy}(t,y). \end{aligned}$$

Next, we estimate $|h_{yy}|$ in terms of $|h|$. To this end, let y be such that $|y| \geq \frac{KT}{2}$. Then, for each $x \in [-K, K]$, we have that

$$\operatorname{sgn}\left(\frac{e^{xy - \frac{1}{2}x^2 t} - 1}{x}\right) = \operatorname{sgn}(y).$$

Thus,

$$\int_{\mathbb{R}} \left|\frac{e^{xy - \frac{1}{2}x^2 t} - 1}{x}\right| \nu(dx) = \left|\int_{\mathbb{R}} \frac{e^{xy - \frac{1}{2}x^2 t} - 1}{x} \nu(dx)\right| \leq |h(t,y)| + |h(0,0)|.$$

Using the inequality above, we conclude that

$$\begin{aligned}|h_{yy}(t,y)| &= \left|\int_{\mathbb{R}} x e^{xy-\frac{1}{2}x^2 t}\nu(dx)\right| \\ &\leq \left|\int_{\mathbb{R}} x^2 \frac{e^{xy-\frac{1}{2}x^2 t}-1}{x}\nu(dx)\right| + \left|\int_{\mathbb{R}} x\nu(dx)\right| \\ &\leq K^2(|h(t,y)|+|h(0,0)|)+K\nu(\mathbb{R}),\end{aligned}$$

for any y such that $|y| \geq \frac{KT}{2}$. Therefore, for each $y \in \mathbb{R}$,

$$|h_{yy}(t,y)| \leq K^2(|h(t,y)|+|h(0,0)|)+K\nu(\mathbb{R})+M, \tag{1.62}$$

where the constant M is defined as

$$M = \max_{(t,y):t\leq T,\,|y|\leq\frac{KT}{2}} |h_{yy}(t,y)|.$$

We work similarly for the estimates on $|h_{yyy}(\cdot)|$.

Since the measure ν is supported in $[-K,K]$, we have, by Theorem 8, that $|r_x(\cdot)|$ is bounded by K. Combining this bound with (1.62), we deduce that, for all $y \in \mathbb{R}$,

$$|r_t(t,h(t,y))| \leq C(1+|h(t,y)|).$$

Therefore, since the domain of $r(\cdot)$ is the range of $h(\cdot)$,

$$|r_t(t,x)| \leq C(1+|x|),$$

for all $x \in \mathbb{R}^+$ and $t \leq T$. This concludes the proof of the theorem. □

Chapter 2

Forward Investment and Partial Information

In classical frameworks for asset allocation problems, such as mean–variance analysis and expected utility maximisation, one typically assumes a model and then solves an optimisation problem. For example, in a one–period model, if we assume that excess returns have mean μ and covariance Σ, then mean–variance analysis yields the optimal allocation

$$\pi^{MV} = \frac{\eta}{x_0}\Sigma^{-1}\mu, \tag{2.1}$$

where η is an exogenous parameter representing the investor's willingness to trade variance for expected return, and x_0 is the initial wealth.

This classical approach to asset allocation has one major drawback: it does not take into account that the true model is never known exactly and, in fact, may be very hard to estimate. For example, it is well known that, even if one assumes that daily returns are normally distributed and i.i.d., it is notoriously difficult to estimate their mean — see, for instance, Merton (1980) and Black (1993).

Furthermore, as with most optimisation procedures, classical asset allocation problems are highly sensitive with respect to their input, of which the market model is a part of. For example, the mean–variance optimal allocation (2.1) depends crucially on μ. Indeed, in practice, one often works with a large number of correlated assets which have an almost singular covariance matrix Σ; in this case, Σ^{-1} is ill–conditioned and amplifies any errors that the estimate of μ may have.

One approach which can be used to mitigate the ambiguity faced when choosing a model is to consider a prior distribution supported on a class of market models. In other words, the model itself is a random variable. Then, the observation of returns can be used to perform Bayesian inference and derive the posterior distribution of the model. This approach is often referred to as partial information. For example, one can consider the class of Black–Scholes models with a fixed volatility and different drifts which are distributed according to some prior — this particular type of partial information is called Bayesian uncertainty.

Asset allocation problems with partial information have been extensively studied in the literature. Typically, the unobservable drift is either modelled as a Gaussian process or a hidden Markov model (HMM). In one of the earliest works, Browne and Whitt (1996) solve the problem of maximising expected log utility — the well known Kelly criterion — under Bayesian uncertainty. In Lakner (1995), the author uses martingale methods to study the problem of expected utility maximisation under partial information for general diffusions; these results are specialised in Lakner (1998) to the case where the drift is an unobserved Gaussian process. In Brendle (2006), the drift is an Ornstein–Uhlenbeck process and the problem of maximising expected power utility of terminal wealth is explicitly solved. Danilova et al. (2010) is concerned with the case where there is additional insider information. In Karatzas and Zhao (2001), the authors study in detail the problem of expected utility maximisation under Bayesian uncertainty and provide two solutions: using duality methods and partial differential equation (PDE) techniques. In Sass and Haussmann (2004) and Rieder and Bäuerle (2005), the authors model the drift using a HMM. Also using a HMM, the authors in Frey, Gabih, and Wunderlich (2012) solve the problem when expert opinions are used to bias the estimate of the drift. More recently, in Björk, Davis, and Landén (2010), these results are generalised. In Xiong and Zhou (2007), the authors study mean–variance analysis under partial information.

Forward investment performance processes were introduced in Musiela and Zariphopoulou (2008, 2009); similar definitions appear in Henderson and Hobson (2007) and Žitković (2009). In Musiela and Zariphopoulou (2010a), Berrier et al. (2009) and El Karoui and Mrad (2013), the authors provide an integral representation for the class of monotonically decreasing forward investment performance processes. In general, there is no explicit characterisation of nonmonotone processes. The most general result is proved in Musiela and Zariphopoulou (2010b), where it is shown that a general forward investment performance process solves a stochastic partial differential equation (SPDE); in El Karoui and Mrad (2013, 2010), the authors establish a connection between this SPDE and two stochastic differential equations (SDEs). Recently, in Nadtochiy and Tehranchi (2012), the authors prove an integral representation for a class of forward investment performance processes arising in factor models. In Nadtochiy and Zariphopoulou (2014), the authors provide examples of forward investment performance processes with nonzero volatility.

The aforementioned literature on partial information focuses exclusively on expected utility maximisation and mean–variance analysis. In particular, to the best of our knowledge, the problem of constructing forward investment performance processes in models with partial information has not yet been studied.

In this chapter, we study forward investment performance processes in models with partial information. More precisely, we work on models with Bayesian uncertainty on the drift, that is, models where the drift is a random variable with a known distribution. We construct forward investment performance processes that are of the type $U(t,x) = u(t,x,R_t)$ where $R_.$ is the process of cumulative excess returns. As in the case of full information, we construct these processes by providing an integral representation for the convex conjugate function of $u(t,\cdot,z)$.

This representation is obtained by solving an ill-posed Hamilton–Jacobi–Bellman (HJB) equation. As expected, the forward investment performance processes we construct have a nonzero volatility; in fact, we see that this volatility is intrinsic, in the sense that it is not generated by changes of numéraire/measure — we refer to Definition 5 in Section 2.5 for the precise definition.

This chapter is structured as follows. In Section 2.1, we introduce the model and define forward investment performance processes in models with partial information. In Section 2.2, we use standard filtering techniques to prove a separation theorem which allows us to separate the construction of forward investment performance processes from the problem of partial information. In Section 2.3, we derive the ill-posed HJB equation associated with the forward investment problem and prove a verification theorem. Section 2.4 is concerned with solving the HJB equation. More precisely, we provide an integral representation of forward investment performance measurement processes; this constitutes the main result of the chapter. Finally, in Section 2.5, we provide several examples of forward investment performance processes, and show that some of these processes have a volatility which is not generated by changes of numéraire/measure.

2.1 Market model and investment criterion

We work on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with

(i) a $\mathbb{R}^N$–valued Brownian motion $W_\cdot$, as well as

(ii) a random variable $\lambda : \Omega \to \mathbb{R}^N$, independent of the Brownian motion $W_\cdot$ under the probability measure $\mathbb{P}$.

The distribution of λ is denoted by $\mu_0(A) = \mathbb{P}(\lambda \in A)$. The random variable λ models the uncertainty on the risk premium. The case of full information can be recovered from our model by letting λ be a Dirac point mass. The following assumption plays an important role in the sequel.

Assumption 7. *The distribution of λ has a finite Laplace transform, that is*

$$\int e^{z'y}\mu_0(dy) < \infty, \quad \textit{for all } z \in \mathbb{R}^N.$$

We denote by $\mathbb{F}^W = \left\{\mathcal{F}_t^W : t \geq 0\right\}$ the augmentation of the filtration generated by the process $W_\cdot$. Additionally, we let

$$\mathbb{F}^{\lambda,W} = \left\{\mathcal{F}_t^W \vee \sigma(\lambda) : t \geq 0\right\},$$

where $\sigma(\lambda)$ denotes the σ–algebra generated by λ.

We consider a complete financial market consisting of N risky securities whose process of cumulative excess returns is denoted by $R_\cdot$ and satisfies

$$R_t = \sigma(\lambda t + W_t), \tag{2.2}$$

where $\sigma \in \mathbb{R}^{N\times N}$ is an invertible matrix. We denote by $\mathbb{F}^R = \left\{\mathcal{F}_t^R : t \geq 0\right\}$ the augmentation of the filtration generated by the process $R_\cdot$.

The different filtrations correspond to different levels of information. An investor with access to the filtration $\mathbb{F}^{\lambda,W}$ is an investor with full information, that is, his strategies can depend on the value taken by λ, as well as past prices. An investor who only observes the price time-series has access to the filtration $\mathbb{F}^R$ — in particular, the value of λ is unobservable to such an investor. We observe that

$$\frac{R_t}{t} \to \lambda,$$

and, therefore, an investor who only observes prices can, eventually, estimate the value of λ with any desired level of accuracy. In this chapter, we only consider investors with partial information. This constrains the set of admissible allocations.

Given a $\mathbb{F}^R$–progressively measurable process $\pi.$ such that

$$\int_0^t |\pi_s|^2 ds < \infty, \text{ a.s. for any } t \geq 0,$$

we define the process of discounted wealth process as the unique strong solution to the following SDE

$$dX_t^\pi = X_t^\pi \pi_t' dR_t.$$

With the definition above, the process $\pi.$ corresponds to the proportion of discounted wealth invested in the risky securities.

We consider only wealth processes that remain strictly positive and satisfy the additional technical condition

$$\mathbb{E}\left[\sup_{s\in[0,t]} |X_s^\pi|^2\right] < \infty, \text{ for all } t \geq 0. \tag{2.3}$$

The formal definition of the set of admissible allocations is as follows.

Definition 2. *The set of all admissible allocations, denoted by $\mathcal{A}$, is defined as*

$$\mathcal{A} = \left\{\pi.\ \mathbb{F}^R\textit{–progressively measurable}: X_\cdot^\pi > 0,\ X_\cdot^\pi \textit{ satisfies } (2.3)\right\}. \tag{2.4}$$

Due to the randomness in λ it is not entirely obvious what strategies, if any, are in $\mathcal{A}$. In the following proposition we observe that $\mathcal{A}$ is not empty and that, in particular, it contains all constant allocations.

Proposition 2. *Suppose that Assumption 7 holds and let $\pi. = \bar{\pi}$ be a constant process. Then, $\pi. \in \mathcal{A}$.*

Proof. Let $\pi.$ be as in the hypothesis of the proposition. Then,

$$X_t^\pi = x_0 \exp\left(\bar{\pi}'\sigma\lambda t + \bar{\pi}'\sigma W_t - \frac{1}{2}\bar{\pi}'\Sigma\bar{\pi}t\right),$$

where $\Sigma = \sigma\sigma'$. In particular,

$$|X_t^\pi|^2 = x_0^2 \exp(2\bar{\pi}'\sigma\lambda t)\exp(\bar{\pi}'\Sigma\bar{\pi}t)M_t,$$

where $M_\cdot$ is a martingale defined by

$$M_t = \exp(2\bar{\pi}'\sigma W_t - 2\bar{\pi}'\Sigma\bar{\pi}t).$$

Since, $M_\cdot$ and λ are independent, we deduce that

$$\mathbb{E}\left[\sup_{s\in[0,t]}|X_s^\pi|^2\right] \le x_0^2\exp(\bar{\pi}'\Sigma\bar{\pi}t)(1+\mathbb{E}[\exp(2\bar{\pi}'\sigma\lambda t)])\mathbb{E}\left[\sup_{s\in[0,t]}|M_s|^2\right].$$

We end the proof by using the Burkholder–Davis–Gundy inequality and Assumption 7. □

Remark 12. *Proposition 2 highlights the importance of Assumption 7. Indeed, if Assumption 7 does not hold, then, for each γ, there exists at least one constant policy $\pi_\cdot = \bar{\pi}$ such that*

$$\mathbb{E}\left[(X_t^\pi)^\gamma\right] = \infty, \ \ \textit{for all } t > 0.$$

As mentioned earlier in the chapter, we are interested in constructing forward investment performance processes, the definition of which is as follows.

Definition 3. *Let $\mathcal{A}$ denote the set of admissible portfolios. A $\mathbb{F}^R$–adapted process $U(t,x) : [0,\infty)\times\mathbb{R}^+\times\Omega\to\mathbb{R}$ is a forward investment performance process if*

- *(i) for each $t \ge 0$, the mapping $U(t,\cdot)$ is increasing and strictly concave,*
- *(ii) for each $\pi \in \mathcal{A}$, the process $t \mapsto U(t, X_t^\pi)$ is a $\mathbb{F}^R$–supermartingale, and*
- *(iii) for each $(t_0, x_0) \in [0,\infty)\times\mathbb{R}^+$, there exists a $\pi_\cdot^* \in \mathcal{A}$ such that $X_{t_0}^{\pi^*} = x_0$ and the process $t \mapsto U\left(t, X_t^{\pi^*}\right)$ is a $\mathbb{F}^R$–martingale in $[t_0,\infty)$.*

The above definition of forward investment performance process differs from Definition 1 in two aspects. First, to account for partial information, the supermartingality and martingality properties need to be taken with respect to the filtration $\mathbb{F}^R$. Second, we require the existence of an optimal wealth process for all possible initial conditions $(t_0, x_0) \in [0, \infty) \times \mathbb{R}^+$. This requirement is necessary in the proof of Proposition 4, below.

We observe that the existence of forward investment performance processes does not follow from their definition. Nevertheless, by providing an integral representation in Section 2.4 and constructing some examples in Section 2.5, we show that Definition 3 is not vacuous.

2.2 Separation theorem

Asset allocation problems in models with partial information are often amenable due to the existence of a so-called separation theorem. Such a theorem allows one to disentangle the optimal control problem from that of optimal filtering. In other words, it reduces the original problem to a standard asset allocation problem under full information. In our case, we also have a separation theorem as shown in the next theorem. It is a consequence of well known results in linear filtering, see for instance Liptser and Shiryaev (2001, Theorem 10.3, p. 396). We provide the proof here, as its details are relevant for the sequel.

Theorem 10. *Let $R_t = \sigma(\lambda t + W_t)$, where $W.$ is a Brownian motion, σ an invertible matrix, and λ a random variable with law μ_0. Suppose that λ has finite mean and is independent of $W.$.*

Let $\mathbb{F}^R = \left\{\mathcal{F}_t^R : t \geq 0\right\}$ be the augmentation of the filtration generated by $R.$ and $\hat{\lambda}_t = \mathbb{E}\left[\lambda | \mathcal{F}_t^R\right]$. Then,

$$R_t = \sigma\left(\int_0^t \hat{\lambda}_s ds + B_t\right),$$

where $B.$ is a $\mathbb{F}^R-$Brownian motion. Furthermore, we have $\hat{\lambda}_t = G\left(t, \sigma^{-1}R_t\right)$, where

$$G(t,z) = \frac{\nabla F(t,z)}{F(t,z)} \quad \textit{and} \quad F(t,z) = \int e^{z'y - \frac{1}{2}\|y\|^2 t} \mu_0(dy), \tag{2.5}$$

for any $t > 0$.

Proof. Let

$$B_t = \sigma^{-1}R_t - \int_0^t \hat{\lambda}_s ds = W_t + \lambda t - \int_0^t \hat{\lambda}_s ds.$$

Then, $B.$ is adapted to the filtration $\mathbb{F}^R$ and, for any t and T such that $t \leq T$,

$$\mathbb{E}\left[B_T|\mathcal{F}_t^R\right] = W_t + \lambda t + \mathbb{E}\left[W_T + \lambda T - W_t - \lambda t|\mathcal{F}_t^R\right] - \int_0^T \mathbb{E}\left[\hat{\lambda}_s|\mathcal{F}_t^R\right] ds = B_t.$$

In other words, $B.$ is a $\mathbb{F}^R$–martingale. Furthermore, $\langle B \rangle_t = t$. Therefore, by Lévy's characterisation theorem, $B.$ is a $\mathbb{F}^R$–Brownian motion.

Next, we compute $\hat{\lambda}_T$ explicitly, for any $T > 0$. Define

$$Z_t = \exp\left(-\lambda' W_t - \frac{1}{2}\lambda^2 t\right).$$

Then, $Z.$ is a $\left(\mathbb{F}^{\lambda,W}, \mathbb{P}\right)$–martingale. Therefore, Z_T defines a probability measure $\mathbb{Q}$, equivalent to $\mathbb{P}$ on $\mathcal{F}_T^{\lambda,W}$, such that

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = Z_T.$$

By Girsanov's theorem, $\tilde{W}_t = W_t + \lambda t = \sigma^{-1}R_t$ is a $\left(\mathbb{F}^{\lambda,W}, \mathbb{Q}\right)$–Brownian motion for $t \in [0,T]$. Furthermore, since $\mathcal{F}_0^{\lambda,W} = \sigma(\lambda)$, $\tilde{W}.$ is independent of λ, under $\mathbb{Q}$.

Similarly,

$$\frac{1}{Z_t} = \exp\left(\lambda' W_t + \frac{1}{2}\lambda^2 t\right) = \exp\left(\lambda' \tilde{W}_t - \frac{1}{2}\lambda^2 t\right)$$

is a $\left(\mathbb{F}^{\lambda,W}, \mathbb{Q}\right)$–martingale for $t \in [0,T]$ and we have $\frac{d\mathbb{P}}{d\mathbb{Q}} = \frac{1}{Z_T}$.

Since $\tilde{W}_T$ and λ are independent under $\mathbb{Q}$, then

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{F}_T^R\right] = F(T, \tilde{W}_T), \tag{2.6}$$

where

$$F(t, z) = \int e^{z'y - \frac{1}{2}\|y\|^2 t}\mu_0(dy).$$

Finally, we obtain

$$\hat{\lambda}_T = \mathbb{E}\left[\lambda|\mathcal{F}_T^R\right] = \frac{\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}}\lambda|\mathcal{F}_T^R\right]}{\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{F}_T^R\right]} = \frac{\nabla F(T, \tilde{W}_T)}{F(T, \tilde{W}_T)},$$

which conclude the proof of the theorem. □

Remark 13. *Note that the result of Theorem 10 does not depend on Assumption 7. If Assumption 7 does not hold, then the functions* $F(t, z)$ *and* $G(t, z)$ *given by* (2.5) *are not defined at* $t = 0$*. This is not a problem since* (2.5) *is said to hold for* $t > 0$*.*

Remark 14. *The measure* $\mathbb{Q}$ *introduced in the proof of Theorem 10 is the so-called equivalent martingale measure (EMM). This is because, under* $\mathbb{Q}$*, any admissible wealth process is a martingale. Indeed, recall from the proof of Theorem 10, that* $\tilde{W}_\cdot = \sigma^{-1}R_\cdot$ *is a Brownian motion under* $\mathbb{Q}$*. Therefore, any admissible wealth process* $X_\cdot^\pi$ *is a local martingale under* $\mathbb{Q}$*,*

$$dX_t^\pi = X_t^\pi \pi_t' \sigma d\tilde{W}_t.$$

Furthermore, since, under $\mathbb{Q}$*,* $\mathcal{F}_T^R$ *is independent of* λ*,*

$$\mathbb{E}^{\mathbb{Q}}\left[\sup_{t\in[0,T]}|X_t^\pi|\right] = \mathbb{E}^{\mathbb{Q}}\left[\sup_{t\in[0,T]}|X_t^\pi|\middle|\sigma(\lambda)\right] = \mathbb{E}^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\sup_{t\in[0,T]}|X_t^\pi|\middle|\sigma(\lambda)\right] < \infty.$$

The last inequality follows from the fact that $\sup_{t\in[0,T]}|X_t^\pi|$ *and* $\frac{d\mathbb{Q}}{d\mathbb{P}}$ *are square* *integrable,* conditioned *on the value of* λ*. Thus,* $X^{\pi}_{\cdot}$ *is a martingale under* $\mathbb{Q}$*.*

The separation theorem above effectively disentangles the problem of model uncertainty from that of portfolio choice. It allows us to work with the so-called *a posteriori* dynamics for $R_{\cdot}$,

$$dR_t = \sigma\left(G\left(t, \sigma^{-1}R_t\right)dt + dB_t\right), \tag{2.7}$$

where $G(\cdot)$ is defined in (2.5) and $B_{\cdot}$ is a $\mathbb{F}^R$–Brownian motion. In the sequel, we shall work on this *a posteriori* model. In this model, we can abstract from the problem of partial information and work as if there were full information. We observe that the dynamics in (2.7) are those of a complete market with a perfectly correlated stochastic risk premium. This risk premium, although no longer constant, has some structure which we exploit in order to solve the problem of constructing forward performance processes.

We observe that, as a consequence of Theorem 10, it is now trivial to construct time–monotone forward performance processes. Indeed, any monotonically decreasing forward investment performance process is of the type

$$U(t,x) = u(A_t, x),$$

where

$$A_t = \int_0^t \left\|G\left(s, \sigma^{-1}R_s\right)\right\|^2 ds,$$

and $u(\cdot)$ solves $u_t - \frac{1}{2}\frac{u_x^2}{u_{xx}} = 0$. In the sequel, we exploit the structure of the function $G(\cdot)$ and characterise other, more complex, types of forward investment processes. More precisely, we consider forward investment performance processes of the form

$$U(t,x) = u\left(t, x, \sigma^{-1}R_t\right), \tag{2.8}$$

where $u(\cdot)$ is a deterministic function.

There are several reasons why we are interested in considering forward investment performance processes of the type (2.8). On the one hand, it is reasonable to assume that the measure of performance depends on $R.$ since the estimate of the risk premium itself depends on the process of cumulative returns. On the other hand, as we shall see, processes of the type (2.8) are still tractable and it is possible to characterise them in terms of a family of measures.

We end this section with a trivial application of Theorem 10, for two examples of measures μ_0.

Example 2. *We consider two simple laws for μ_0, a Dirac measure and a Gaussian distribution:*

- *In the case of full information we have $\mu_0 = \delta_b$. Then,*

$$F(t,z) = e^{z'b - \frac{1}{2}\|b\|^2 t} \quad \textit{and, in turn} \quad G(t,z) = b.$$

- *If $\mu_0 \sim \mathcal{N}(b, B)$, then*

$$\begin{aligned} F(t,z) &= \frac{\sqrt{|B_t|}}{\sqrt{|B|}} e^{-\frac{1}{2}\left(b'B^{-1}b - (b+z)'B_t(b+z)\right)}, \\ G(t,z) &= B_t(b+z), \end{aligned}$$

where $B_t = (B^{-1} + t)^{-1}$.

2.3 The Hamilton–Jacobi–Bellman equation

In this section, we derive the Hamilton–Jacobi–Bellman (HJB) equation associated with the construction of forward investment performance processes. Note that, although we work with forward investment performance processes, the arguments used in the proofs are entirely analogous to those used in classical expected utility maximisation — see, for instance, Touzi (2010, Sections 6.3 and 8.1).

Before proceeding, we establish some notation. Let $\phi(t,x,z) : [0,\infty) \times \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}$ be a $C^{1,2,2}$ function. We denote by $\nabla\phi$ and $\triangle\phi$ the gradient and Laplacian of $\phi(t,x,z)$ with respect to z, respectively. Furthermore, for each $\pi \in \mathbb{R}^N$, we define the linear operators

$$\mathcal{L}^{\pi}\phi = \phi_t + x\phi_x\pi'\sigma G + \frac{1}{2}x^2\phi_{xx}\pi'\Sigma\pi + x\pi'\sigma\nabla\phi_x + \nabla\phi' G + \frac{1}{2}\triangle\phi,$$
$$\mathcal{I}^{\pi}\phi = x\phi_x\pi'\sigma + \nabla\phi',$$

where $G(t,z)$ is defined by (2.5).

We observe that, by Itô's formula,

$$\phi\left(T, X_T^{\pi}, \sigma^{-1}R_T\right) = \phi\left(t, X_t^{\pi}, \sigma^{-1}R_t\right) + \int_t^T \mathcal{L}^{\pi_s}\phi\left(s, X_s^{\pi}, \sigma^{-1}R_s\right) ds + \int_t^T \mathcal{I}^{\pi_s}\phi\left(s, X_s^{\pi}, \sigma^{-1}R_s\right) dB_s,$$

for any admissible $\pi_\cdot \in \mathcal{A}$. Furthermore, if $\phi(t,\cdot,z)$ is strictly concave, then

$$\begin{aligned}\sup_{\pi\in\mathbb{R}^N} \mathcal{L}^{\pi}\phi &= \phi_t + \nabla\phi' G + \frac{1}{2}\triangle\phi - \frac{1}{2}\frac{1}{\phi_{xx}}\|\phi_x G + \nabla\phi_x\|^2. \\ &= \phi_t + H(\phi_x, \nabla\phi, \phi_{xx}, \triangle\phi, \nabla\phi_x),\end{aligned}$$

where $H(\cdot)$, the so-called Hamiltonian, is a continuous function.

Next, we prove that any forward investment performance process of the form (2.8) solves the HJB equation. This is the content of the following theorem.

Theorem 11. *Suppose that Assumption 7 holds and* $U(t,x) = u\left(t,x,\sigma^{-1}R_t\right)$ *is a forward investment performance process in the sense of Definition 3, where* $u(t,x,z) : [0,\infty) \times \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}$ *is a* $C^{1,2,2}$ *function. Then,* $u(\cdot)$ *satisfies the HJB equation,*

$$\sup_{\pi\in\mathbb{R}^N} \mathcal{L}^{\pi}u = 0. \tag{2.9}$$

Proof. The proof is split in the following two propositions. □

Proposition 3. *Suppose that Assumption 7 holds and let* $u(\cdot)$ *be as in the hypothesis of Theorem 11. Then,*

$$\sup_{\pi\in\mathbb{R}^N} \mathcal{L}^{\pi} u \leq 0. \tag{2.10}$$

Proof. Let $t \geq 0$, $x > 0$ and $\bar{\pi} \in \mathbb{R}^N$ be arbitrary, and define the process $\pi_{\cdot} = \mathbf{1}_{\cdot\geq t}\bar{\pi}$. Then, $\pi_{\cdot} \in \mathcal{A}$ and $X_t^{\pi} = x$. For fixed $h > 0$ and $\varepsilon > 0$, we define the stopping time

$$\theta^h = \inf\left\{s : |X_s^{\pi} - X_t^{\pi}| > \varepsilon \text{ or } |R_s - R_t| > \varepsilon \text{ or } s - t > h\right\}.$$

We observe that, by continuity of $X_{\cdot}^{\pi}$ and $R_{\cdot}$,

$$\lim_{h\to 0} \frac{\theta_h}{h} = 1, \ \ a.s..$$

Since $\pi_{\cdot} \in \mathcal{A}$ and $U(t,x) = u\left(t, x, \sigma^{-1}R_t\right)$ is a forward investment performance process, we must have

$$\mathbb{E}\left[u\left(\theta_h, X_{\theta_h}^{\pi}, \sigma^{-1}R_{\theta_h}\right) | \mathcal{F}_t^R\right] \leq u\left(t, x, \sigma^{-1}R_t\right).$$

By Itô's formula,

$$\begin{aligned} u\left(\theta_h, X_{\theta_h}^{\pi}, \sigma^{-1}R_{\theta_h}\right) =& u\left(t, x, \sigma^{-1}R_t\right) + \int_t^{\theta_h} \mathcal{L}^{\bar{\pi}} u\left(s, X_s^{\pi}, \sigma^{-1}R_s\right) ds + \\ & \int_t^{\theta_h} \mathcal{I}^{\bar{\pi}} u\left(s, X_s^{\pi}, \sigma^{-1}R_s\right) dB_s. \end{aligned}$$

Therefore,

$$\mathbb{E}\left[\int_t^{\theta_h} \mathcal{L}^{\bar{\pi}} u\left(s, X_s^{\pi}, \sigma^{-1}R_s\right) ds \,\middle|\, \mathcal{F}_t^R\right] \leq 0.$$

By the dominated convergence theorem,

$$\lim_{h\to 0} \frac{1}{h}\mathbb{E}\left[\int_t^{\theta_h} \mathcal{L}^{\bar{\pi}} u\left(s, X_s^{\pi}, \sigma^{-1}R_s\right) ds \,\middle|\, \mathcal{F}_t^R\right] = \mathcal{L}^{\bar{\pi}} u\left(t, x, \sigma^{-1}R_t\right).$$

Therefore, we conclude that

$$\mathcal{L}^{\bar{\pi}} u\left(t, x, \sigma^{-1} R_t\right) \leq 0,$$

and, since t, x and $\bar{\pi}$ are arbitrary and R_t takes any value in $\mathbb{R}^N$, the proof is complete. $\square$

Proposition 4. *Let $u(\cdot)$ be as in the hypothesis of Theorem 11. Then,*

$$\sup_{\pi \in \mathbb{R}^N} \mathcal{L}^{\pi} u \geq 0. \tag{2.11}$$

Proof. Suppose, by contradiction, that there exist t_0, x_0, z_0 and δ such that

$$\sup_{\pi \in \mathbb{R}^N} \mathcal{L}^{\pi} u(t_0, x_0, z_0) = \left(\phi_t + H(\phi_x, \nabla\phi, \phi_{xx}, \triangle\phi, \nabla\phi_x)\right)(t_0, x_0, z_0) < -2\delta.$$

By continuity of the function $H(\cdot)$, there exists $\varepsilon > 0$ such that

$$\sup_{\pi \in \mathbb{R}^N} \mathcal{L}^{\pi} u(t, x, z) < -\delta,$$

for all t, x and z such that $|(t, x, z) - (t_0, x_0, z_0)| < \varepsilon$.

Let $X^{\pi}_{\cdot}$ be the optimal wealth process for the forward performance process $U(t, x)$ such that $X^{\pi}_{t_0} = x_0$. We define the stopping time θ by

$$\theta = \inf\left\{t : |X^{\pi}_t - x_0| > \varepsilon \text{ or } \left|\sigma^{-1} R_t - z_0\right| > \varepsilon \text{ or } t - t_0 > \varepsilon\right\}.$$

Then,

$$\mathbb{E}\left[u\left(\theta, X^{\pi}_{\theta}, \sigma^{-1} R_{\theta}\right) | \mathcal{F}^R_{t_0}\right] = u\left(t_0, x_0, \sigma^{-1} R_{t_0}\right).$$

By Itô's formula,

$$\begin{aligned} u\left(\theta, X^{\pi}_{\theta}, \sigma^{-1} R_{\theta}\right) =& u\left(t_0, x_0, \sigma^{-1} R_{t_0}\right) + \int_{t_0}^{\theta} \mathcal{L}^{\pi_t} u\left(t, X^{\pi}_t, \sigma^{-1} R_t\right) dt + \\ & \int_{t_0}^{\theta} \mathcal{I}^{\pi_t} u\left(t, X^{\pi}_t, \sigma^{-1} R_t\right) dB_t. \end{aligned}$$

Therefore,

$$\mathbb{E}\left[\int_{t_0}^{\theta} \mathcal{L}^{\pi_t} u\left(t, X_t^{\pi}, \sigma^{-1}R_t\right) dt \middle| \mathcal{F}_{t_0}^{R}\right] = 0.$$

In particular,

$$\begin{aligned} 0 &= \mathbb{E}\left[\int_{t_0}^{\theta} \mathcal{L}^{\pi_t} u\left(t, X_t^{\pi}, \sigma^{-1}R_t\right) dt \middle| \sigma^{-1}R_{t_0} = z_0\right] \\ &\leq -\delta\mathbb{E}\left[\theta - t_0 | \sigma^{-1}R_{t_0} = z_0\right] \\ &< 0, \end{aligned}$$

which gives the desired contradiction. □

Verification theorem

The converse to Theorem 11 is known as a verification theorem. It yields sufficient conditions for a process of the type (2.8) to be a forward investment performance process.

Theorem 12. *Suppose Assumption 7 holds. Let* $u(t,x,z) : [0,\infty) \times \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}$ *be a solution to the HJB equation,* (2.9)*, satisfying the growth constraint*

$$u(t,x,z) \leq C_t \left(1 + |x|^2 + e^{|z|}\right), \tag{2.12}$$

for some $C_\cdot$ *such that* $\sup_{t\in[0,\cdot]} C_t < \infty$*. Suppose that* $u(t,\cdot,z)$ *is increasing and strictly concave, for all* $(t,z) \in [0,\infty) \times \mathbb{R}^N$*.*

Define $\bar{\pi}(t,x,z)$ *by*

$$\bar{\pi}(t,x,z) = -\left(\frac{1}{x u_{xx}(\cdot)} \left(\sigma^{-1}\right)' \left(u_x(\cdot) G(t,z) + \nabla u_x(\cdot)\right)\right)(t,x,z). \tag{2.13}$$

Suppose the SDE

$$dX_t^* = X_t^* \bar{\pi}\left(t, X_t^*, \sigma^{-1}R_t\right)' dR_t,$$

has a unique strong solution for each initial condition $X_t^* = x$. *Furthermore, assume that* $\pi_\cdot^* \in \mathcal{A}$, *where* $\pi_t^* = \bar{\pi}\left(t, X_t^*, \sigma^{-1} R_t\right)$.

Then $U(t,x) = u\left(t, x, \sigma^{-1} R_t\right)$ *is a forward investment performance process and its optimal wealth process is* $X_\cdot^*$.

Proof. Let $\pi_\cdot \in \mathcal{A}$ be an arbitrary admissible portfolio. For fixed t and T such that $t \leq T$, and for each $n \in \mathbb{N}$, we define the stopping time

$$\theta_n = \inf\left\{s \geq t : s > T \text{ or } |X_s^\pi - X_t^\pi| > n \text{ or } |R_s - R_t| > n\right\}.$$

Then,

$$\lim_{n \to \infty} \theta_n = T, \quad a.s..$$

By Itô's formula,

$$u\left(\theta_n, X_{\theta_n}^\pi, \sigma^{-1} R_{\theta_n}\right) = u\left(t, X_t^\pi, \sigma^{-1} R_t\right) + \int_t^{\theta_n} \mathcal{L}^{\pi_s} u\left(s, X_s^\pi, \sigma^{-1} R_s\right) ds + \\ \int_t^{\theta_n} \mathcal{I}^{\pi_s} u\left(t, X_s^\pi, \sigma^{-1} R_s\right) dB_s.$$

Therefore, since $u(\cdot)$ solves (2.9),

$$\begin{aligned} \mathbb{E}\left[u\left(\theta_n, X_{\theta_n}^\pi, \sigma^{-1} R_{\theta_n}\right) | \mathcal{F}_t^R\right] = u\left(t, X_t^\pi, \sigma^{-1} R_t\right) + \\ \mathbb{E}\left[\left.\int_t^{\theta_n} \mathcal{L}^{\pi_s} u\left(s, X_s^\pi, \sigma^{-1} R_s\right) ds \right| \mathcal{F}_t^R\right] \\ \leq u\left(t, X_t^\pi, \sigma^{-1} R_t\right). \end{aligned} \tag{2.14}$$

Now, we define the random variable

$$Z = \sup_{s \in [t,T]} u\left(s, X_s^\pi, \sigma^{-1} R_s\right).$$

Then, by (2.12), there is a constant C such that

$$Z \leq C\left(1 + \sup_{s \in [t,T]} |X_s^\pi|^2 + \sup_{s \in [t,T]} \exp\left(\left|\sigma^{-1} R_s\right|\right)\right).$$

We recall that $\sigma^{-1}R_s = \lambda s + W_s$, and that λ is independent of $W_\cdot$. Therefore, since $\pi_\cdot \in \mathcal{A}$ and Assumption 7 holds, we deduce that $\mathbb{E}[|Z|] < \infty$.

Finally, we observe that, for all n,

$$u\left(\theta_n, X^\pi_{\theta_n}, \sigma^{-1}R_{\theta_n}\right) \leq Z.$$

Therefore, we can apply the dominated convergence theorem to conclude that

$$\mathbb{E}\left[u\left(T, X^\pi_T, \sigma^{-1}R_T\right)|\mathcal{F}^R_t\right] \leq u\left(t, X^\pi_t, \sigma^{-1}R_t\right),$$

that is, $t \mapsto U(t, X^\pi_t)$ is a supermartingale for all $\pi_\cdot \in \mathcal{A}$.

To conclude the proof, we need to show that $t \mapsto U(t, X^*_t)$ is a martingale. To this end we just need to note that, by construction, $X^*_\cdot$ satisfies (2.14) with equality, since

$$\mathcal{L}^{\bar{\pi}(t, X^*_t, \sigma^{-1}R_t)}u\left(t, X^*_t, \sigma^{-1}R_t\right) = 0.$$

□

2.4 Characterisation of forward investment performance processes

The results in this section provide necessary and sufficient conditions for a process of the form

$$U(t,x) = u\left(t, x, \sigma^{-1}R_t\right)$$

to be a forward investment performance process. To this end, we solve the HJB equation, (2.9), proceeding in two steps:

1. linearise the HJB equation, transforming it in the forward heat equation;

2. use the multidimensional version of Widder's theorem on the positive solu-

tions to the heat equation.

After establishing the solutions to the ill–posed HJB equation, we conclude by using the Theorems 11 and 12 of the previous section to provide a characterisation of forward investment performance processes.

Linearising the HJB equation

In order to linearise the HJB equation we assume that its solution satisfies the so-called Inada's conditions.

Definition 4. *Let $u(x) : \mathbb{R}^+ \to \mathbb{R}$ be a smooth, increasing and strictly concave function. We say that $u(\cdot)$ satisfies Inada's conditions if*

$$\lim_{x \to 0^+} u'(x) = \infty \text{ and } \lim_{x \to \infty} u'(x) = 0.$$

Theorem 13. *Suppose that $u(t,x,z) : [0,\infty) \times \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}$ is a $C^{1,3,3}$ function such that $u(t,\cdot,z)$ is increasing, strictly concave and satisfies Inada's conditions. For $t > 0$, let*

$$\tilde{u}(t,x,z) = u(t,x,z)F(t,z),$$

where $F(t,z)$ is defined by (2.5)*. Additionally, let $\tilde{v}(t,\cdot,z)$ be the convex conjugate function of $\tilde{u}(t,\cdot,z)$,*

$$\begin{aligned} \tilde{v}(t,y,z) &= \sup_{x>0} \{\tilde{u}(t,x,z) - xy\} \\ &= \tilde{u}\left(t, \tilde{I}(t,y,z), z\right) - y\tilde{I}(t,y,z), \end{aligned} \tag{2.15}$$

where $\tilde{I}(\cdot)$ is such that

$$\tilde{u}_x\left(t, \tilde{I}(t,y,z), z\right) = y,\ y \in \mathbb{R}^+. \tag{2.16}$$

Then, $u(\cdot)$ *solves* (2.9) *if and only if* $\tilde{v}(\cdot)$ *solves*

$$\tilde{v}_t + \frac{1}{2}\triangle \tilde{v} = 0, \tag{2.17}$$

where $\triangle \tilde{v}$ *denotes the Laplacian of* $\tilde{v}(t, y, z)$ *with respect to* z.

Proof. The proof is separated in Lemmas 3 and 4, below. □

Before stating and proving Lemmas 3 and 4, and since their proof is purely technical, we outline the intuition behind these results. We recall from Remark 14 that $F(\cdot)$ is such that

$$F\left(t, \sigma^{-1}R_t\right) = \mathbb{E}^{\mathbb{Q}}\left[\left.\frac{d\mathbb{P}}{d\mathbb{Q}}\right| \mathcal{F}_t^R\right],$$

where $\mathbb{Q}$ is the EMM.

We then have

$$\mathbb{E}\left[U(t,X_t)|\mathcal{F}_s^R\right] = \frac{\mathbb{E}^{\mathbb{Q}}\left[\left.\frac{d\mathbb{P}}{d\mathbb{Q}}U(t,X_t)\right| \mathcal{F}_s^R\right]}{\mathbb{E}^{\mathbb{Q}}\left[\left.\frac{d\mathbb{P}}{d\mathbb{Q}}\right| \mathcal{F}_s^R\right]} = \frac{\mathbb{E}^{\mathbb{Q}}\left[F\left(t,\sigma^{-1}R_t\right)U(t,X_t)|\mathcal{F}_s^R\right]}{F\left(s,\sigma^{-1}R_s\right)}.$$

Therefore, $U(t,x) = u\left(t,x,\sigma^{-1}R_t\right)$ is a forward investment performance process under the measure $\mathbb{P}$ if and only if $\tilde{U}(t,x) = F\left(t,\sigma^{-1}R_t\right)u\left(t,x,\sigma^{-1}R_t\right)$ is a forward investment performance process under the measure $\mathbb{Q}$. Since $R_.$ has no drift under $\mathbb{Q}$, we conclude that $\tilde{u}(t,x,z) = F(t,z)u(t,x,z)$ must satisfy the HJB equation, (2.9), with $G(\cdot)$ replaced by zero, that is (2.18), below. This result is proved in Lemma 3.

Since $\tilde{u}\left(t,x,\sigma^{-1}R_t\right)$ defines a forward investment performance process under the equivalent martingale measure $\mathbb{Q}$, we have

$$\tilde{u}(t,x,z) = \sup_{\pi\in\mathcal{A}} \mathbb{E}^{\mathbb{Q}}\left[\tilde{u}\left(T, X_T^{\pi}, \sigma^{-1}R_T\right)|X_t^{\pi} = x,\ R_t = \sigma z\right].$$

Let $\tilde{v}(t,\cdot,z)$ be the convex conjugate function of $\tilde{u}(t,\cdot,z)$ defined as in (2.15).

Then, the following properties hold:

(i) $\tilde{u}(t,x,z) \leq \tilde{v}(t,y,z) + xy$, for any x and y, with equality if and only if $x = \tilde{I}(t,y,z)$.

(ii) $\tilde{u}(t,\cdot,z)$ is the concave conjugate function of $\tilde{v}(t,\cdot,z)$, that is

$$\tilde{u}(t,x,z) = \inf_{y>0} \{\tilde{v}(t,y,z) + xy\}.$$

From (i), and since $\mathbb{Q}$ is the EMM, we have

$$\tilde{u}(t,x,z) \leq \mathbb{E}^{\mathbb{Q}} \left[\tilde{v}\left(T,y,\sigma^{-1}R_T\right) | R_t = \sigma z\right] + xy,$$

with equality if and only if y is such that $x = \mathbb{E}^{\mathbb{Q}} \left[\tilde{I}\left(T,y,\sigma^{-1}R_T\right) \middle| R_t = \sigma z \right]$. In other words,

$$\tilde{u}(t,x,z) = \inf_{y>0} \left\{\mathbb{E}^{\mathbb{Q}} \left[\tilde{v}\left(T,y,\sigma^{-1}R_T\right) | R_t = \sigma z\right] + xy\right\}.$$

Therefore, by (ii), $\tilde{v}(\cdot)$ satisfies

$$\tilde{v}(t,y,z) = \mathbb{E}^{\mathbb{Q}} \left[\tilde{v}\left(T,y,\sigma^{-1}R_T\right) | R_t = \sigma z\right],$$

and, using Feynman–Kac's formula, we conclude that $\tilde{v}(\cdot)$ solves the heat equation

$$\tilde{v}_t + \frac{1}{2} \triangle \tilde{v} = 0,$$

where $\triangle \tilde{v}$ denotes the Laplacian of $\tilde{v}(t,y,z)$ with respect to z. This result is established formally in Lemma 4.

In terms of notation, a tilde is used to denote functions which arise naturally in the risk–neutral world, where the probability measure is $\mathbb{Q}$. The variable x is reserved for variables in the wealth domain, while y denotes variables in the dual domain.

Lemma 3. *Let $F(t,z)$ be as in* (2.5)*. The function $u(t,x,z) : [0,\infty)\times\mathbb{R}^+\times\mathbb{R}^N \to \mathbb{R}$ is a solution to* (2.9) *if and only if $\tilde{u}(\cdot)$ solves*

$$\tilde{u}_t + \frac{1}{2}\triangle\tilde{u} - \frac{1}{2}\frac{\|\nabla\tilde{u}_x\|^2}{\tilde{u}_{xx}} = 0, \tag{2.18}$$

where

$$\tilde{u}(t,x,z) = u(t,x,z)F(t,z). \tag{2.19}$$

Proof. Recall the definitions of $F(\cdot)$ and $G(\cdot)$, given in (2.5). From the definition of $\tilde{u}(t,x,z)$, (2.19), it follows easily that

$$u_t = \frac{\tilde{u}_t}{F} - \frac{F_t}{F^2}\tilde{u}, \quad \nabla u = \frac{\nabla\tilde{u}}{F} - \frac{\tilde{u}}{F}G,$$

and

$$\nabla^2 u = \frac{1}{F}\nabla^2\tilde{u} - 2\nabla\tilde{u}\frac{G'}{F} - \tilde{u}\frac{\nabla^2 F}{F^2} + 2\frac{\tilde{u}}{F}GG'.$$

Then, noting that $F(\cdot)$ solves the heat equation $F_t + \frac{1}{2}\triangle F = 0$, we conclude from the three equalities above that

$$u_t + \nabla u'G + \frac{1}{2}\triangle u - \frac{1}{2}\frac{1}{u_{xx}}\|u_x G + \nabla u_x\|^2 = \frac{1}{F}\left(\tilde{u}_t + \frac{1}{2}\triangle\tilde{u} - \frac{1}{2}\frac{\|\nabla\tilde{u}_x\|^2}{\tilde{u}_{xx}}\right).$$

Therefore, $u(\cdot)$ is a solution to (2.9) if and only if $\tilde{u}(\cdot)$ solves (2.18). $\square$

Lemma 4. *Suppose that $\tilde{u}(t,x,z) : [0,\infty)\times\mathbb{R}^+\times\mathbb{R}^N \to \mathbb{R}$ is a $C^{1,3,3}$ function such that $\tilde{u}(t,\cdot,z)$ is increasing, strictly concave, and satisfies Inada's conditions. Let $\tilde{v}(t,\cdot,z)$ be the convex conjugate function of $\tilde{u}(t,\cdot,z)$, defined as in* (2.15)*. Then, $\tilde{u}(\cdot)$ solves* (2.18) *if and only if $\tilde{v}(\cdot)$ solves*

$$\tilde{v}_t + \frac{1}{2}\triangle\tilde{v} = 0.$$

Proof. Recall the definitions of $\tilde{v}(\cdot)$ and $\tilde{I}(\cdot)$,

$$\tilde{v}(t,y,z) = \tilde{u}\left(t, \tilde{I}(t,y,z), z\right) - y\tilde{I}(t,y,z), \tag{2.20}$$

and

$$\tilde{u}_x\left(t, \tilde{I}(t,y,z), z\right) = y, \ \ y \in \mathbb{R}^+. \tag{2.21}$$

Differentiating (2.21), we deduce that

$$\tilde{u}_{xx} = \frac{1}{\tilde{I}_y}, \ \ \tilde{u}_{xxx} = \frac{-\tilde{I}_{yy}}{\tilde{I}_y^3}, \ \ \tilde{u}_{xt} = \frac{-\tilde{I}_t}{\tilde{I}_y}, \ \ \nabla\tilde{u}_x = \frac{-\nabla\tilde{I}}{\tilde{I}_y},$$

$$\nabla\tilde{u}_{xx} = \frac{1}{\tilde{I}_y}\left(\frac{-\nabla\tilde{I}_y}{\tilde{I}_y} + \frac{\tilde{I}_{yy}\nabla\tilde{I}}{\tilde{I}_y^2}\right),$$

$$\nabla^2\tilde{u}_x = \frac{2}{\tilde{I}_y^2}\nabla\tilde{I}_y(\nabla\tilde{I})' - \frac{\tilde{I}_{yy}}{\tilde{I}_y^3}\nabla\tilde{I}(\nabla\tilde{I})' - \frac{\nabla^2\tilde{I}}{\tilde{I}_y}.$$

Therefore, from (2.20), we deduce that

$$\begin{aligned} \tilde{v}_t(t,y,z) &= \tilde{u}_t\left(t, \tilde{I}(t,y,z), z\right), \\ \triangle\tilde{v}(t,y,z) &= \triangle\tilde{u}\left(t, \tilde{I}(t,y,z), z\right) - \frac{\left\|\nabla\tilde{u}_x\left(t, \tilde{I}(t,y,z), z\right)\right\|^2}{\tilde{u}_{xx}\left(t, \tilde{I}(t,y,z), z\right)}, \end{aligned}$$

which proves the result. □

The functions $\tilde{I}(\cdot)$ and $\tilde{v}(\cdot)$ defined in Theorem 13 satisfy the identity

$$\tilde{I}(t,y,z) = -\tilde{v}_y(t,y,z).$$

Therefore, if $\tilde{v}(\cdot)$ is a solution to the heat equation (2.17), then $\tilde{I}(\cdot)$ is also a solution,

$$\tilde{I}_t + \frac{1}{2}\triangle\tilde{I} = 0. \tag{2.22}$$

Furthermore, $\tilde{I}(\cdot)$ is strictly positive. For these two reasons, $\tilde{I}(\cdot)$ is instrumental in the sequel.

We observe that it is possible to write $\tilde{I}(\cdot)$ in terms of $u(\cdot)$ directly, bypassing the definitions of $\tilde{v}(\cdot)$ and $\tilde{u}(\cdot)$, as

$$\tilde{I}(t,y,z) = u_x^{(-1)}\left(t, \frac{y}{F(t,z)}, z\right), \tag{2.23}$$

where $u_x^{(-1)}(t,y,z)$ denotes the inverse of $u_x(t,x,z)$ with respect to x, and $F(t,z)$ is defined as in (2.5). The local risk tolerance of $u(\cdot)$ is given in terms of $\tilde{I}(\cdot)$ by

$$-\frac{u_x}{u_{xx}}\left(t, \tilde{I}(t,y,z), z\right) = -y\tilde{I}_y(t,y,z). \tag{2.24}$$

Similarly, the optimal portfolio function, (2.13), can also be written in terms of $\tilde{I}(\cdot)$ as

$$\bar{\pi}\left(t, \tilde{I}(t,y,z), z\right) = \left(\sigma^{-1}\right)' \frac{\nabla \tilde{I}}{\tilde{I}}(t,y,z).$$

Widder's theorem

It is well known that equation (2.22) does not have a unique solution. In fact, it may not have a solution at all depending on the initial condition. This equation is known as the ill–posed backward heat equation. In dimension one, a classical theorem of Widder, provides an integral representation to the positive solutions of the backward heat equation. For higher dimensions, an analogous result holds, as the following theorem shows.

In analogy with (11), we define $\mathcal{B}^+\left(\mathbb{R}^N\right)$ as

$$\mathcal{B}^+\left(\mathbb{R}^N\right) = \left\{\nu \in \mathcal{B}\left(\mathbb{R}^N\right) : \nu \geq 0 \text{ and } \int_{\mathbb{R}^N} e^{x'z}\nu(dz) < \infty,\ z \in \mathbb{R}^N\right\}. \tag{2.25}$$

Theorem 14 (Widder (1975)). *Let $f(t,x) : [0,\infty) \times \mathbb{R}^N \to \mathbb{R}$ be a positive $C^{1,2}$ function. The following are equivalent:*

(*i*) *$f(\cdot)$ solves the heat equation, $f_t + \frac{1}{2}\triangle f = 0$;*

(*ii*) *there exists a measure $\nu \in \mathcal{B}^+\left(\mathbb{R}^N\right)$ such that*

$$f(t,x) = \int e^{z'x - \frac{1}{2}\|x\|^2 t}\nu(dz).$$

Proof. This result is well known in dimension $N = 1$, see Widder (1975, p. 235). For higher dimensions, in Robbins and Siegmund (1973), the authors explain how to proceed with the proof. It is also a particular case of a result in Nadtochiy and Tehranchi (2012). For completeness, we prove this result here.

The proof of the implication $(ii) \Rightarrow (i)$ is trivial. Here, we prove the implication $(i) \Rightarrow (ii)$.

Let $f(t,x) : [0,\infty)\times\mathbb{R}^N \to \mathbb{R}$ be a positive $C^{1,2}$ function such that $f_t+\frac{1}{2}\triangle f = 0$. Then, for each $\tau > 0$ and $t \in [0,\tau)$, we have

$$f(t,x) = \int_{\mathbb{R}^N} \Gamma(t,x;\tau,y) f(\tau,y)dy, \tag{2.26}$$

where

$$\Gamma(t,x;\tau,y) = \frac{1}{\left(\sqrt{2\pi}\right)^N}(\tau-t)^{-\frac{N}{2}} \exp\left(-\frac{|y-x|^2}{2(\tau-t)}\right),$$

see Corollary 2 in Friedman (1964, p. 48).

Doing the transformation $y = \tau z$ in (2.26), we get

$$f(t,x) = \exp\left(\frac{-|x|^2}{2(\tau-t)}\right)\left(\sqrt{\frac{\tau}{\tau-t}}\right)^N \int_{\mathbb{R}^N} \exp\left(\frac{\tau}{\tau-t}\left(z'x - \frac{|z|^2 t}{2}\right)\right)\nu_\tau(dz), \tag{2.27}$$

where

$$\nu_\tau(dz) = \left(\sqrt{\frac{\tau}{2\pi}}\right)^N \exp\left(\frac{|z|^2\tau}{2}\right) f(\tau z,\tau)dz.$$

We observe that

$$\int_{\mathbb{R}^N} \nu_\tau(du) = f(0,0) < \infty.$$

Therefore, by Helly's selection theorem — see, for instance, Billingsley (1968, p. 227) — there exists a measure μ and a subsequence ν_{τ_k}, such that $\nu_{\tau_k} \to \mu$ weakly.

Next, we show that $\nu \in \mathcal{B}^+ \left(\mathbb{R}^N\right)$. It follows from (2.27), that

$$f(0,x) = e^{\frac{-|x|^2}{2\tau}} \int_{\mathbb{R}^N} e^{z'x} \nu_\tau(dz),$$

for any $\tau > 0$. Therefore,

$$\int_{B_R(0)} e^{z'x} \nu_\tau(dz) \leq f(0,x) e^{\frac{|x|^2}{2\tau}}, \tag{2.28}$$

for any $R > 0$, where $B_R(0) = \left\{z \in \mathbb{R}^N : |z| < R\right\}$ is the ball of radius R centered at 0. Taking the limit as $\tau \to \infty$ in (2.28), we deduce that

$$\int_{B_R(0)} e^{z'x} \nu(dz) \leq f(0,x),$$

and, since $R > 0$ is arbitrary, we conclude that

$$\int_{\mathbb{R}^N} e^{z'x} \nu(dz) < \infty.$$

To conclude the proof, we take the limit as $\tau \to \infty$ in (2.27). Then, it follows that

$$f(t,x) = \int_{\mathbb{R}^N} \exp\left(z'x - \frac{|z|^2 t}{2}\right) \nu(dz).$$

$\square$

Characterising forward investment performance processes

The following two theorems form the core of the chapter. Their proofs follow easily from the previous results.

Theorem 15. *Suppose that Assumption 7 holds. Let* $u(t,x,z) : [0,\infty) \times \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}$ *be a* $C^{1,3,3}$ *function such that* $u(t,\cdot,z)$ *satisfies Inada's conditions. Sup-*

pose that $U(t,x) = u\left(t,x,\sigma^{-1}R_t\right)$ *is a forward investment performance process. Define* $\tilde{I}(\cdot)$ *as in* (2.23).

Then,

$$\tilde{I}(t,y,z) = \int e^{z'x - \frac{1}{2}\|x\|^2 t}\nu_y(dx), \tag{2.29}$$

for some collection of measures $\{\nu_y\}_{y\in\mathbb{R}^+}$ *such that* $\nu_y \in \mathcal{B}^+\left(\mathbb{R}^N\right)$, *for all* $y \in \mathbb{R}^+$.

Proof. By Theorem 11, $u(\cdot)$ solves the HJB equation, (2.9). Define $\tilde{u}(t,x,z) = u(t,x,z)F(t,z)$, where $F(t,z)$ is as in (2.5). Let $\tilde{v}(t,\cdot,z)$ be the convex conjugate function of $\tilde{u}(t,\cdot,z)$, defined as in (2.15). Then, by Theorem 13, $\tilde{v}(\cdot)$ solves the heat equation,

$$\tilde{v}_t + \frac{1}{2}\triangle\tilde{v} = 0.$$

Since $\tilde{I}(t,y,z) = -\tilde{v}_y(t,y,z)$, we deduce that $\tilde{I}(\cdot)$ solves the same equation.

Finally, since $\tilde{I}(\cdot)$ is a positive function, we deduce from Theorem 14 that, for each y, there exists a measure $\nu_y \in \mathcal{B}^+\left(\mathbb{R}^N\right)$ such that

$$\tilde{I}(t,y,z) = \int e^{z'x - \frac{1}{2}\|x\|^2 t}\nu_y(dx).$$

□

The next theorem is the converse to Theorem 15. In order to state it, we need a procedure that takes a function $\tilde{I}(\cdot)$, such as the one in (2.29), and constructs a candidate to forward investment performance process; this is done in the following remark.

Remark 15. *Given a function* $\tilde{I}(t,y,z) : [0,\infty)\times\mathbb{R}^+\times\mathbb{R}^N \to \mathbb{R}$ *such that* $\tilde{I}_y(\cdot) < 0$, *we construct a candidate to forward investment performance process,* $U(t,x)$, *as follows.*

Define

$$\tilde{v}(t,y,z) = -\int_{y_0}^{y}\tilde{I}(t,x,z)dx + H(t,z),$$

where $H(\cdot)$ *is an arbitrary solution to the heat equation,* $H_t + \frac{1}{2}\triangle H = 0$ *— in particular, we may have* $H(\cdot) = 0$*.*

Let $\tilde{u}(t,\cdot,z)$ *denote the concave conjugate function to* $\tilde{v}(t,\cdot,z)$*,*

$$\begin{aligned}\tilde{u}(t,x,z) &= \inf_{y>0}\left\{\tilde{v}(t,y,z) + xy\right\}\\ &= \tilde{v}\left(t, \tilde{I}^{(-1)}(t,x,z), z\right) + x\tilde{I}^{(-1)}(t,x,z),\end{aligned}$$

where $\tilde{I}^{(-1)}(t,x,z)$ *denotes the inverse of* $\tilde{I}(t,y,z)$ *with respect to* y*.*

Finally, define

$$u(t,x,z) = \frac{\tilde{u}(t,x,z)}{F(t,z)},$$

where $F(t,z)$ *is given by* (2.5)*.*

Then, the candidate to forward investment performance process is

$$U(t,x) = u\left(t, x, \sigma^{-1}R_t\right).$$

Theorem 16. *Suppose that Assumption 7 holds. Let* $\{\nu_y\}_{y\in\mathbb{R}^+}$ *be a collection of measures such that* $\nu_y \in \mathcal{B}^+\left(\mathbb{R}^N\right)$ *and* $\tilde{I}_y(t,y,z) < 0$*, where*

$$\tilde{I}(t,y,z) = \int e^{z'x - \frac{1}{2}\|x\|^2 t}\nu_y(dx).$$

With $\tilde{I}(\cdot)$*, construct* $U(t,x) = u(t,x,\sigma^{-1}R_t)$ *as in Remark 15 and suppose that* $u(\cdot)$ *satisfies the growth constraint* (2.12)*.*

Define $\bar{\pi}(t,x,z)$ *by*

$$\bar{\pi}(t,x,z) = \frac{1}{x}\left(\sigma^{-1}\right)' \nabla\tilde{I}\left(t, \tilde{I}^{(-1)}(t,x,z), z\right), \tag{2.30}$$

where $\tilde{I}^{(-1)}(t,x,z)$ *denotes the inverse of* $\tilde{I}(t,y,z)$ *with respect to* y*. Suppose that the SDE*

$$dX_t^* = X_t^*\bar{\pi}\left(t, X_t^*, \sigma^{-1}R_t\right)dR_t$$

has a unique strong solution for each initial condition $X_t^* = x$. *Furthermore, assume that* $\pi_\cdot^* \in \mathcal{A}$, *where* $\pi_t^* = \bar{\pi}\left(t, X_t^*, \sigma^{-1} R_t\right)$.

Then, $U(t,x)$ *is a forward investment performance process.*

Proof. Given $\tilde{I}(t,y,z)$ as in the hypothesis of the theorem, we define $\tilde{v}(t,y,z)$ and $u(t,x,z)$ as in Remark 15. Then, by construction, $\tilde{v}(\cdot)$ solves (2.17). Therefore, by Theorem 13, $u(\cdot)$ solves the HJB equation, (2.9). Then, the result follows from the verification theorem, Theorem 11. □

Remark 16. *We observe that Theorem 16 and, in particular, the examples constructed in the following section establish the existence of forward investment performance processes.*

As with time–monotone forward performance processes, under partial information, forward investment performance processes of the form (2.8) are also characterised in terms of measures. However, in this case we need to specify a continuum of measures, as opposed to just one, all of which are measures in $\mathcal{B}^+\left(\mathbb{R}^N\right)$, as opposed to $\mathcal{B}^+(\mathbb{R})$. In fact, because we consider processes which may have a nonzero volatility, the characterisation provided in this chapter is more general than that of monotone utilities. In the following section, we recover the representation of monotonically decreasing forward investment performance processes under full information.

2.5 Examples

In this section, we construct some examples of forward investment performance processes using the results of the previous section. These examples are presented in order of increasing complexity, starting with the well known monotonically decreasing forward investment performance processes.

Recovering monotonically decreasing forward investment performance processes

Recall that a forward investment performance process, $U(t,x)$, is said to be monotonically decreasing if $U(\cdot,x)$ is almost surely decreasing. In the case of full information, there is a well known characterisation of these processes, as described in the Introduction. We now recover this result.

Firstly, we observe that in the case of full information, $\mu_0 = \delta_b$, we have from (2.5) that

$$F(t,z) = e^{z'b - \frac{1}{2}\|b\|^2 t} \text{ and } G(t,z) = b.$$

Secondly, we note that, in the case of monotonically decreasing forward investment performance processes there is no dependence on R_t. In particular, we must have

$$\nabla u = \nabla u_x = 0. \tag{2.31}$$

In terms of the function $\tilde{I}(\cdot)$ defined as in (2.23), the second equality in (2.31) can be rewritten as

$$\nabla \tilde{I}(t,y,z) = -y\tilde{I}_y(t,y,z)G(t,z). \tag{2.32}$$

Theorem 15 provides an integral representation for $\tilde{I}(\cdot)$ in terms of a collection of measures $\{\nu_y\}_{y\in\mathbb{R}^+}$,

$$\tilde{I}(t,y,z) = \int e^{z'x - \frac{1}{2}\|x\|^2 t} \nu_y(dx). \tag{2.33}$$

If all the measures ν_y are equivalent to each other, then, there exists a measure, ν, and some change of measure, $f(y,x)$, such that

$$\nu_y(dx) = f(y,x)\nu(dx),$$

for each $y \in \mathbb{R}^+$.

Now, suppose that the function $f(\cdot, x)$ is differentiable. Then, by (2.32), we deduce that

$$\int xe^{z'x-\frac{1}{2}\|x\|^2t}f(y,x)\nu(dx) = -yb\int e^{z'x-\frac{1}{2}\|x\|^2t}f_y(y,x)\nu(dx),$$

that is,

$$xf(y,x)\nu(dx) = -ybf_y(y,x)\nu(dx).$$

Therefore, ν must be supported on the set $\{\alpha b : \alpha \in \mathbb{R}\} \subset \mathbb{R}^N$, and

$$\alpha f(y, \alpha b) = -yf_y(y, \alpha b), \tag{2.34}$$

for all $\alpha \in \mathbb{R}$.

Integrating (2.34) with respect to y, we deduce that

$$f(y, \alpha b) = C(\alpha)y^{-\alpha},$$

for some function $C(\cdot)$. Plugging the expression for $f(y, x)$ in the integral representation of $\tilde{I}(\cdot)$, (2.33), we deduce that

$$\tilde{I}(t, y, z) = \int_{\mathbb{R}} y^{-\alpha}e^{\alpha z'b-\frac{1}{2}\alpha^2\|b\|^2t}\tilde{\nu}(d\alpha),$$

for some measure $\tilde{\nu}$.

Finally, if we define $I(t, y) = U_x^{(-1)}(t, y)$, we deduce that

$$I(t, y) = \tilde{I}(t, yF(t, z), z) = \int_{\mathbb{R}} y^{-\alpha}e^{-\frac{1}{2}\alpha(\alpha-1)\|b\|^2t}\tilde{\nu}(d\alpha).$$

Since $U(t, \cdot)$ is assumed to satisfy Inada's conditions, we deduce that the measure $\tilde{\nu}$ must be supported in $\mathbb{R}^+$. This integral representation coincides with the one found in Berrier, Rogers, and Tehranchi (2009) — see Remark 1 in the Introduction.

Myopic–like forward investment performance processes

The optimal portfolio for a time–monotone logarithmic forward performance process is

$$\pi_t^{myopic} = \left(\sigma^{-1}\right)' G\left(t, \sigma^{-1} R_t\right) = \left(\sigma^{-1}\right)' \frac{\nabla F\left(t, \sigma^{-1} R_t\right)}{F\left(t, \sigma^{-1} R_t\right)}, \tag{2.35}$$

which is known as the myopic portfolio. In this section, we construct forward performance processes that have an optimal allocation with the same functional form as $\pi_\cdot^{myopic}$. Although the functional form is the same, in general, they do not coincide exactly with $\pi_\cdot^{myopic}$ and therefore cannot be called myopic.

In the integral representation found in Theorems 15 and 16, we consider a family of measures $\{\nu_y\}_{y \in \mathbb{R}^+}$. One of the simplest examples one can consider is the one where $\nu_y = I_0(y)\nu$, for some smooth, strictly decreasing function $I_0(\cdot)$, and some measure $\nu \in \mathcal{B}^+\left(\mathbb{R}^N\right)$. In this case, we have

$$\tilde{I}(t, y, z) = I_0(y) F_\nu(t, z),$$

where we define $F_\nu(t, z)$, for any measure $\nu \in \mathcal{B}^+\left(\mathbb{R}^N\right)$, as

$$F_\nu(t, z) = \int e^{z'x - \frac{1}{2}\|x\|^2 t} \nu(dx). \tag{2.36}$$

Then, the forward investment performance process associated with this family of measures is defined as $U(t, x) = u\left(t, x, \sigma^{-1} R_t\right)$, where the function $u(\cdot)$ is given by

$$u(t, x, z) = \frac{F_\nu(t, z)}{F_{\mu_0}(t, z)} u_0\left(\frac{x}{F_\nu(t, z)}\right).$$

We observe that this family of forward performance processes is indexed by the initial condition $u_0(\cdot)$ and the measure ν. However, it is interesting to notice that the associated optimal portfolio does not depend on $u_0(\cdot)$,

$$\pi_t^* = \left(\sigma^{-1}\right)' \frac{\nabla F_\nu\left(t, \sigma^{-1} R_t\right)}{F_\nu\left(t, \sigma^{-1} R_t\right)}.$$

Note that, with the same notation as in (2.36), the myopic portfolio can be rewritten as

$$\pi_t^{myopic} = \left(\sigma^{-1}\right)' \frac{\nabla F_{\mu_0}\left(t, \sigma^{-1} R_t\right)}{F_{\mu_0}\left(t, \sigma^{-1} R_t\right)}.$$

In particular, $\pi^*_\cdot$ is the myopic portfolio whenever $\nu = \mu_0$.

Power forward investment performance processes

We define a power forward investment performance process as a process $U(t,x)$ such that

$$U(t,x) = x^{1-\frac{1}{\gamma}} \phi_t,$$

for some $\gamma \in (1, \infty)$ and some process $\phi_\cdot$.

If $U(t,x) = u\left(t, x, \sigma^{-1} R_t\right)$ is a power forward investment performance process and we define $\tilde{I}(t,y,z)$ as in (2.23), then

$$\tilde{I}(t,y,z) = (yC(t,z))^{-\gamma},$$

for some function $C(\cdot)$. These processes constitute a particular example of the ones introduced in the previous subsection and can be generated by considering the family of measures $\{\nu_y\}_{y \in \mathbb{R}^+}$ where $\nu_y = y^{-\gamma}\nu$, for some $\nu \in \mathcal{B}^+\left(\mathbb{R}^N\right)$.

In this case, the optimal portfolio is given in terms of $C(\cdot)$ by

$$\bar{\pi}(t,x,z) = -\gamma \left(\sigma^{-1}\right)' \frac{\nabla C(t,z)}{C(t,z)}, \tag{2.37}$$

and the local risk tolerance function is

$$-\frac{u_x}{u_{xx}}(t,x,z) = \gamma x. \tag{2.38}$$

We observe that (2.38) coincides with the risk tolerance function of classic power utilities.

Mixtures of power utilities

In this subsection, we use Theorem 16 to construct forward investment performance processes which can be thought of as mixtures of power utilities. Our objective is twofold. On the one hand, we illustrate how to combine, in the dual domain, two or more forward investment performance processes into a more complex process. On the other hand, in the next subsection, we show that these forward investment performance processes have a complex volatility process, in the sense that it is not generated by changes of measure and/or numéraire.

Given two forward investment performance processes, $U_1(t,x) = u_1(t,x,\sigma^{-1}R_t)$ and $U_2(t,x) = u_2(t,x,\sigma^{-1}R_t)$, the average of the two is not, in general, a forward investment performance process. This a consequence of the nonlinearity of the HJB equation, (2.9). However, in the dual domain, it is easy to combine the two processes into a more complex process.

Indeed, using the transformation in (2.23), we can define $\tilde{I}_1(\cdot)$ and $\tilde{I}_2(\cdot)$ which, by Theorem 15, satisfy the heat equation. Then, given $w \in [0,1]$, the function $\tilde{I}(\cdot)$ also satisfies the heat equation, where

$$\tilde{I}(t,y,z) = w\tilde{I}_1(t,y,z) + (1-w)\tilde{I}_2(t,y,z).$$

Furthermore, $\tilde{I}_y(t,y,z) < 0$. Therefore, as in Remark 15, $\tilde{I}(\cdot)$ can be used to define a forward investment performance process $U(t,x) = u\left(t,x,\sigma^{-1}R_t\right)$. We observe that the optimal portfolio for the process $U(t,x)$ is a weighted average of the optimal portfolios for $U_1(t,x)$ and $U_2(t,x)$. Also, from (2.24), we deduce that the risk tolerance of $U(t,x)$ is a weighted average of the risk tolerances of $U_1(t,x)$ and $U_2(t,x)$. In this case, we say that the forward investment performance process $U(t,x)$ is a mixture of $U_1(t,x)$ and $U_2(t,x)$.

The mixing procedure introduced in the previous paragraph can be applied to power forward investment performance processes. The advantage of working with such processes is that, due to their tractability, the computations can be made fairly explicitly. To this end, we define

$$\tilde{I}(t,y,z) = \int (yC(t,z;\gamma))^{-\gamma}\nu(d\gamma)$$

and

$$\tilde{v}(t,y,z) = -\int \frac{y^{1-\gamma}}{1-\gamma}C(t,z;\gamma)^{-\gamma}\nu(d\gamma),$$

for some collection of positive functions $\{C(t,z;\gamma)\}_{\gamma\in(1,\infty)}$, and for some measure ν concentrated on $(1,\infty)$. In line with Theorem 16, we assume that, for each γ, $C(\cdot;\gamma)^{-\gamma}$ solves the heat equation.

We observe that $\tilde{v}_y(t,y,z) = -\tilde{I}(t,y,z) < 0$. Therefore, the concave conjugate function of $\tilde{v}(t,\cdot,z)$ is $\tilde{u}(t,\cdot,z)$, given explicitly by

$$\begin{aligned}\tilde{u}(t,x,z) &= \tilde{v}\left(t,\tilde{I}^{(-1)}(t,y,z),z\right) + x\tilde{I}^{(-1)}(t,x,z)\\ &= -\int \frac{\gamma}{1-\gamma}\tilde{I}^{(-1)}(t,x,z)^{1-\gamma}C(t,z;\gamma)^{-\gamma}\nu(d\gamma).\end{aligned}$$

Under the hypothesis of Theorem 16, we then have that $U(t,x) = u\left(t,x,\sigma^{-1}R_t\right)$, where

$$u(t,x,z) = \frac{\tilde{u}(t,x,z)}{F(t,z)},$$

and $F(t,z)$ is defined in (2.5), is a forward investment performance process.

We say that $U(t,x)$ is a mixture of power forward investment performance processes. The optimal portfolio for $U(t,x)$ is given by

$$\bar{\pi}(t,x,z) = -\left(\sigma^{-1}\right)' \int \gamma\frac{\nabla C(t,z;\gamma)}{C(t,z;\gamma)}\nu_{t,x,z}(d\gamma), \tag{2.39}$$

where $\nu_{t,x,z}$ is a probability measure, defined as

$$\nu_{t,x,z}(d\gamma) = \frac{1}{x}\left(\tilde{I}^{(-1)}(t,x,z)C(t,z;\gamma)\right)^{-\gamma}\nu(d\gamma).$$

Similarly, the local risk tolerance function is given by

$$-\frac{u_x}{u_{xx}}(t,x,z) = x\int \gamma \nu_{t,x,z}(d\gamma). \tag{2.40}$$

Note the similarities between (2.39), (2.40) and (2.37), (2.38), respectively. These justify why we say that $U(t,x)$ is a mixture of power forward investment performance processes; in this case, the mixing measure is ν. In the following two examples, we make this construction more explicit by considering two specific mixing measures.

Example 3. *If $\nu = \delta_\gamma$, for some $\gamma > 1$, then*

$$\tilde{I}^{(-1)}(t,x,z) = \frac{x^{-\frac{1}{\gamma}}}{C(t,z;\gamma)},$$

and

$$\tilde{u}(t,x,z) = \frac{\gamma}{\gamma-1}\frac{x^{1-\frac{1}{\gamma}}}{C(t,z;\gamma)}.$$

Therefore, and as expected, the resulting forward investment performance process is the one associated with a power utility. In this example, there is no mixing.

Example 4. *In this example, we consider the measure*

$$\nu = w_1\delta_\gamma + w_2\delta_{2\gamma},$$

where $\gamma \in (1,\infty)$ and $w_1 + w_2 = 1$. Then,

$$\tilde{I}^{(-1)}(t,x,z) = f\left(x, w_2C(t,z;2\gamma)^{-2\gamma}, w_1C(t,z;\gamma)^{-\gamma}\right)^{-\frac{1}{\gamma}},$$

where

$$f(x,a,b) = \frac{-b+\sqrt{b^2+4ax}}{2a}.$$

The expressions for the forward investment performance process and the optimal

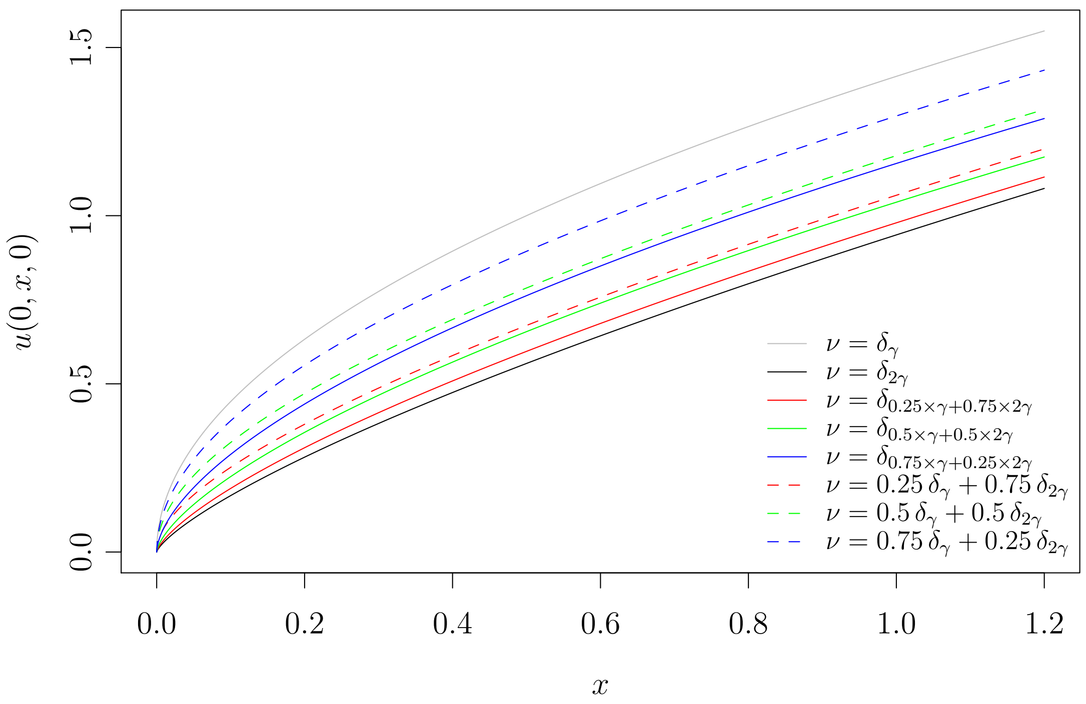

Figure 2.1: Plots of the initial utility $u(0,\cdot,0)$ for different mixing measures ν, where we set $\gamma = 2$.

portfolio can also be worked out explicitly. In this case, the resulting forward investment performance is a mixture of two power utilities with relative risk tolerances γ *and* 2γ*.*

In Figure 2.1, we plot the initial utility $u(0,\cdot,0)$ *for different mixing measures* ν*. We observe that a utility which results from mixing two power utilities with relative risk tolerances* γ *and* 2γ *does not coincide with the power utility which has relative risk tolerance equal to the average of* γ *and* 2γ*. Nevertheless, these two different utilities have the same risk tolerance.*

Processes with intrinsic volatility

In the remainder of this chapter, we show that the mixtures of forward investment performance processes introduced in the previous section have an intrinsic volatility, in the sense that it is not generated by changes of measure and/or numéraire

— see Definition 5, below, for the precise definition.

The volatility of a forward investment performance process, $U(t,x)$, is defined as the process, $a(t,x)$, such that

$$dU(t,x) = b(t,x)dt + a(t,x)dB_t.$$

Time–monotone forward performance processes must have zero volatility. In Musiela and Zariphopoulou (2009), the authors show how to construct processes with nonzero volatility by combining monotonically decreasing processes with changes of measure and changes of numéraire. Processes constructed in such a way have a volatility, $a(t,x)$, which satisfies

$$a(t,x) = \phi_t U(t,x) + \psi_t x U_x(t,x), \tag{2.41}$$

for some processes $\phi_.$ and $\psi_.$.

Now, suppose that $U(t,x)$ is a process with a volatility of the type (2.41). We define the process $V(t,x)$ by

$$V(t,x) = \frac{U\left(t, \frac{x}{\mathcal{E}(\psi_.)_t}\right)}{\mathcal{E}(\phi_.)_t},$$

where $\mathcal{E}(\psi_.)$ denotes the exponential martingale associated with $\psi_.$, defined as the solution to the SDE

$$d\mathcal{E}(\psi_.)_t = \psi_t \mathcal{E}(\psi_.)_t dB_t.$$

Then, it is easy to see that $V(t,x)$ has a zero volatility. Therefore, despite being nonzero, volatilities satisfying (2.41) are rather contrived, in that they are generated by changes of numéraire and/or measure[1]. This observation leads to the following definition.

---

[1] Recently, the authors in Nadtochiy and Zariphopoulou (2014) construct homothetic forward investment performance processes with a nonzero volatility. The volatility of such processes is also of the type (2.41).

Definition 5. *Given a forward investment performance process, $U(t, x)$, its volatility is defined as the process $a(t, x)$ such that*

$$dU(t, x) = b(t, x)dt + a(t, x)dB_t.$$

The volatility process, $a(t, x)$, is said to be generated by changes of numéraire/measure if it satisfies (2.41)*. In other words, $U(t, x)$ has a volatility generated by changes of numéraire/measure if there exist strictly positive processes $Z_\cdot$ and $Y_\cdot$, such that the process $V(t, x)$, defined as*

$$V(t, x) = \frac{U(t, xY_t)}{Z_t},$$

has zero volatility — in this case, $Y_\cdot$ and $Z_\cdot$ correspond to the change of numéraire and change of measure, respectively.

We recall that the forward investment performance processes constructed in this chapter are of the type $U(t, x) = u\,(t, x, \sigma^{-1}R_t)$. The volatility of such a process is

$$a(t, x) = \nabla u\left(t, x, \sigma^{-1}R_t\right).$$

A necessary condition for such a volatility to be generated by changes of numéraire/measure is given in the following lemma.

Lemma 5. *Let $U(t, x) = u\,(t, x, \sigma^{-1}R_t)$ be a forward investment performance process. Define $\tilde{I}(\cdot)$ as in* (2.23)*. If the volatility of $U(t, x)$ is generated by changes of numéraire/measure, then there exist functions $\phi(\cdot)$ and $\psi(\cdot)$ such that*

$$\nabla\tilde{I}(t, y, z) = \phi(t, z)y\tilde{I}_y(t, y, z) + \psi(t, z)\tilde{I}(t, y, z). \tag{2.42}$$

Proof. Let $a(t, x)$ be the volatility of $U(t, x)$ and suppose that it satisfies (2.41).

Then, there exist functions $\phi(\cdot)$ and $\psi(\cdot)$ such that

$$\nabla u_x(t,x,z) = \phi(t,z)u_x(t,x,z) + \psi(t,z)xu_{xx}(t,x,z). \tag{2.43}$$

We observe that, on the one hand, $\nabla u_x(t,x,z)$ satisfies

$$\nabla u_x\left(t,\tilde{I}(t,y,z),z\right) = -\frac{1}{F(t,z)}\left(\frac{\nabla\tilde{I}}{\tilde{I}_y}(t,y,z) + yG(t,z)\right), \tag{2.44}$$

where $F(t,z)$ and $G(t,z)$ are defined as in (2.5). On the other hand, replacing x by $\tilde{I}(t,y,z)$ in (2.43), leads to

$$\nabla u_x\left(t,\tilde{I}(t,y,z),z\right) = \frac{1}{F(t,z)}\left(\phi(t,z)y + \psi(t,z)\frac{\tilde{I}(t,y,z)}{\tilde{I}_y(t,y,z)}\right). \tag{2.45}$$

We conclude the proof by comparing (2.44) and (2.45). □

In the next proposition, we show that the volatility of a mixture of power forward performance processes becomes more complex as the number of mixed processes grows. In particular, when one mixes more than two processes, the volatility of the resulting forward performance process is not generated by changes of numéraire/measure.

Proposition 5. *Given a collection of distinct $\gamma_i \in (1,\infty)$, let*

$$\tilde{I}(t,y,z) = \sum_{i=1}^{n} w_i y^{-\gamma_i} e^{z'b_i - \frac{1}{2}\|b_i\|^2 t}, \tag{2.46}$$

where $\sum_{i=1}^n w_i = 1$, and $b_i \in \mathbb{R}^N$ are such that $b_i \neq 0$, for $i = 1,\cdots,n$.

For each $m \le n$, define the Vandermonde matrix

$$\Gamma_m = \begin{pmatrix} 1 & -\gamma_1 & \cdots & (-\gamma_1)^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & -\gamma_n & \cdots & (-\gamma_n)^m \end{pmatrix}$$

and the matrix $B = [b_1, b_2, \cdots, b_n]$*. Suppose that at least one of the rows of* B *is not in the column space of* Γ_{n-2}*.*

With $\tilde{I}(\cdot)$*, define* $U(t,x)$ *as in Theorem 16. Then, if* $n > 2$*, the volatility of the forward investment performance process* $U(t,x)$ *is not generated by changes of numéraire/measure.*

Proof. By Lemma 5, it suffices to show that there do not exist functions $\phi(\cdot)$ and $\psi(\cdot)$ such that

$$\nabla\tilde{I}(t,y,z) = \phi(t,z)y\tilde{I}_y(t,y,z) + \psi(t,z)\tilde{I}(t,y,z).$$

We prove that, if there exist functions $\phi_0(\cdot), \cdots, \phi_m(\cdot)$ such that

$$\nabla\tilde{I}(t,y,z) = \sum_{j=0}^{m} \phi_j(t,z) \left(\frac{\partial}{\partial \log(y)}\right)^j \tilde{I}(t,y,z), \tag{2.47}$$

then $m \geq n-1$.

Indeed, suppose that (2.47) holds. Then, from the definition of $\tilde{I}(t,y,z)$ in (2.46), we deduce that

$$\sum_{i=1}^{n} y^{-\gamma_i} w_i b_i e^{z'b_i - \frac{1}{2}\|b_i\|^2 t} = \sum_{i=1}^{n}\sum_{j=0}^{m} \phi_j(t,z)(-\gamma_i)^j y^{-\gamma_i} w_i e^{z'b_i - \frac{1}{2}\|b_i\|^2 t}.$$

Since $\gamma_i \neq \gamma_j$ for $i \neq j$, we conclude that

$$b_i = \sum_{j=0}^{m} \phi_j(t,z)(-\gamma_i)^j, \quad \text{for all } i = 1, \cdots, n. \tag{2.48}$$

We observe that the equation above can be written in matrix notation as

$$B = \Phi(t,z)\Gamma_m',$$

where $\Phi(t,z) = [\phi_1(t,z), \cdots, \phi_m(t,z)]$. Since Γ_{n-1} is invertible, we deduce that it is possible to find functions $\phi_j(\cdot)$ that satisfy (2.48), whenever $m \geq n-1$. However, for $m < n-1$ the same is not true. The reason is that, if (2.48) holds for $m < n-1$, then all the rows of the matrix B are in the column space of Γ_{n-2}, contradicting the hypothesis of the proposition. In particular, if $m = 1$ and $n \geq 3$, (2.48) does not hold which concludes the proof. $\square$

Chapter 3

Partial Information and the Black–Litterman Model

In Chapter 2, we develop the theory of forward investment under partial information in order to account for model ambiguity and, in particular, the uncertainty of expected returns. In this chapter, we develop further the topic of uncertainty of expected returns, but with a focus on how these can be better estimated. Such a study is of practical importance not only to forward investment theory, but to asset allocation problems in general. In fact, we observe that expected returns play an important role in most asset allocation problems — including those in forward investment. However, expected returns are notoriously difficult to estimate, see Merton (1969) and Black (1993).

In the papers by Black and Litterman (1991, 1992), the authors introduce the original Black–Litterman model for a single–period asset allocation problem. In this model, the prediction of an asset pricing theory, such as the Capital Assset Pricing Model (CAPM), is taken as the shrinkage point, or prior belief, in the estimation of expected returns. This prior belief is then combined with the so-called investor's views, which reflect the extra information the investor may have on the distribution of returns.

There have been recent developments on translating the Black–Litterman methodology to continuous time models, notably in Frey, Gabih, and Wunderlich (2012) and Davis and Lleo (2013). In these works, the authors consider the problem of estimating latent factors when expert/analyst opinions are used to bias these estimates. More precisely, in Frey, Gabih, and Wunderlich (2012), the latent factors are modelled as a Markov chain and expert opinions form a marked point process. In Davis and Lleo (2013), both the latent factors and analyst recommendations follow stochastic differential equations. In both cases, filtering tools are used to estimate the latent factors. Then, a separation theorem allows the authors to use the filtered factors to solve an expected utility maximisation problem.

In Frey, Gabih, and Wunderlich (2012) and Davis and Lleo (2013), the focus is clearly on investor's views and how these can be used to bias the estimate of expected returns. However, in the original Black–Litterman model, it is the prior belief — the predictions of an equilibrium theory such as the CAPM — which is used as a shrinkage point in the estimation of expected returns. In this chapter, we present an alternative approach to developing a continuous time version of the Black–Litterman model, one where equilibrium is the main component[1].

In our model, the estimator of instantaneous expected returns is continuously shrunk towards its theoretical equilibrium value. To achieve this, we model expected returns not as a constant random variable but as an unobserved stochastic process which mean–reverts around its equilibrium value. Then, a simple application of the Kalman–Bucy filter yields the estimator. This estimator is the main novelty of the chapter.

To better motivate our choice of model and shrinking procedure, we introduce an example which we shall resume later in Section 3.4. Consider a factor model with m factor portfolios and n risky assets. The process of cumulative excess returns of the factor portfolios is denoted by $R^F_\cdot = \left(R^{F,1}_\cdot, \cdots, R^{F,m}_\cdot\right)$. The process of cumulative excess returns of the remaining risky assets is denoted by $R_\cdot =$

---

[1] In the sequel, we do not consider investor's views. This is for simplicity, as these can be easily added as extra state variables.

$(R^1_\cdot, \cdots, R^n_\cdot)$. The dynamics of $\left(R^F_\cdot, R_\cdot\right)$ is given by

$$\begin{cases} dR^F_t = \mu dt + \sigma_F dW^F_t \\ dR_t = \alpha dt + \beta dR^F_t + \sigma_a dW^a_t, \end{cases} \tag{3.1}$$

where σ_F is a $m \times m$ matrix, σ_a is a diagonal $n \times n$ matrix, β is the $n \times m$ matrix of factor loadings, and $\left(W^F, W^a\right)$ is a $(m+n)$–dimensional Brownian motion. Note that $W^a_\cdot$ drives the idiosyncratic shocks, while $W^F_\cdot$ drives the common shocks. We assume that σ_F, σ_a and β are known, but μ and α have to be estimated. In this example, we focus on the estimate for α.

According to Arbitrage Pricing Theory (APT) — see Ross (1976) — for $n \gg m$, we should have $\alpha \approx 0$ in (3.1). In other words, investors should not be compensated for idiosyncratic risk because, for n large enough, this can be hedged away. If α is indeed zero, it is pointless to estimate it. However, in practice, α is not always zero, an observation which fuels the active management industry and the search for alpha. A deviation of α from zero can be interpreted as a mispricing — when $\alpha_i > 0$ (respectively, $\alpha_i < 0$) the i^{th} security is overpriced (respectively, underpriced). Such mispricings may be due to small shocks which generate temporary misalignments between supply and demand.

The existence of mispricings creates arbitrage opportunities which, due to the presence of arbitrageurs, should fade away. In other words, the presence of arbitrageurs creates a force which keeps α close to its equilibrium value of zero. One way to make the APT consistent with the existence of temporary mispricings is to assume that α is not constant. Furthermore, due to the nature of the arbitrage clearing mechanism, it is reasonable to assume that α mean–reverts around zero. More precisely,

$$d\alpha_t = -\kappa_\alpha \alpha_t dt + \sigma_\alpha dW^\alpha_t, \tag{3.2}$$

where σ_α and κ_α are $n \times n$ positive definite diagonal matrices, and $W^\alpha_\cdot$ is a n–

dimensional Brownian motion.

Using standard linear filtering theory, it is possible to derive the optimal estimator, $\hat{\alpha}_{\cdot}$, for the process $\alpha_{\cdot}$ in (3.2). It is easy to see that, by construction, $\hat{\alpha}_{\cdot}$ is continuously shrunk towards its equilibrium value of 0. In other words, in this simple example, the APT prediction provides a 'center of gravity' to which the statistical estimator of α is continuously attracted.

The remainder of this chapter is structured as follows. In Section 3.1, we briefly review the classic Black–Litterman model for a single–period mean–variance allocation problem. In Section 3.2, we derive our novel estimator for instantaneous expected returns and study some of its properties. Section 3.3 is concerned with the problem of optimal asset allocation. There, we make direct use of the results in Kim and Omberg (1996) and Brendle (2006) to solve explicitly the problem of maximising expected power utility of terminal wealth. Finally, in Section 3.4, we make our results more specific by resuming the study of the multifactor model introduced earlier; these results are illustrated using data on the American equities market.

3.1 The Black–Litterman model

In this section, we review the classic Black–Litterman model for single–period asset allocation, following closely the exposition in He and Litterman (1999).

We consider a one–period model with n risky assets. We denote the excess returns of these assets by $R = (R_1, \cdots, R_n)$ and we assume that they are normally distributed,

$$R \sim \mathcal{N}(\mu, \Sigma),$$

for some vector $\mu \in \mathbb{R}^n$ and positive–definite matrix $\Sigma \in \mathbb{R}^{n\times n}$.

Given an allocation $\pi \in \mathbb{R}^n$, the discounted wealth generated by π at the end of the period is given by

$$X^{\pi} = x_0(1 + \pi' R),$$

where x_0 is the initial wealth.

Then, the optimal asset allocation for the mean–variance criterion,

$$\min_{\pi \in \mathbb{R}^n} \Big(\mathrm{Var}(X^\pi) - \eta \mathbb{E}[X^\pi] \Big),$$

is given by

$$\pi^{MV} = \frac{\eta}{x_0} \Sigma^{-1} \mu, \tag{3.3}$$

where the parameter η measures the willingness of the investor to trade variance for expected return.

The optimal mean–variance allocation in (3.3) is known to have some problems. Specifically, it is very unstable with respect to the inputs of the model, μ and Σ. Indeed, more often than not, Σ is a large matrix with some almost null eigenvalues — a consequence of the correlation between different assets. Therefore, Σ^{-1} is ill-conditioned and amplifies any errors arising in the estimation of μ. Given that μ is difficult to estimate — see, for instance, Merton (1980) and Black (1993) — this is an undesirable feature for any asset allocation.

The CAPM predicts that the investors as a whole hold the market portfolio

$$\pi^{\text{market}} = \eta^{\text{market}} \Sigma^{-1} \mu,$$

where $\eta^{\text{market}} = \left(\mathbf{1}' \Sigma^{-1} \mu\right)^{-1}$ is a proxy for the average mean–variance trade-off coefficient of the market as a whole. Therefore, the CAPM predicts that, in equilibrium, expected returns are

$$\mu^{\text{eq}} = \frac{1}{\eta^{\text{market}}} \Sigma \pi^{\text{market}}. \tag{3.4}$$

Since π^{market} in (3.4) is directly observable, whereas μ is not, the theoretical relation in (3.4) can be used to estimate μ. Another, completely orthogonal, approach to estimating μ is to analyse historical data using some statistical method.

As we see in this section, the Black–Litterman model can be used to benefit from both approaches — this is the main observation behind the approach developed in the subsequent sections.

The Black–Litterman model uses a Bayesian approach to estimate the vector of expected returns, μ. Therefore, μ is itself a random variable, initially distributed according to some prior belief; this prior belief is given by the CAPM. More precisely, we assume that

$$\mu \sim \mathcal{N}(\mu^{\mathrm{eq}}, \tau\Sigma), \tag{3.5}$$

where μ^{eq} is the CAPM prediction, given by (3.4), and the scalar τ represents the confidence in the CAPM prior[2].

We define the error of the CAPM prior as $\varepsilon^{\mathrm{eq}} = \mu - \mu^{\mathrm{eq}}$. We observe that $\varepsilon^{\mathrm{eq}}$ is also normally distributed,

$$\varepsilon^{\mathrm{eq}} \sim \mathcal{N}(0, \tau\Sigma).$$

Furthermore, $\varepsilon^{\mathrm{eq}}$ can be used to rewrite (3.5) as

$$\mu = \mu^{\mathrm{eq}} + \varepsilon^{\mathrm{eq}}.$$

In addition to the CAPM, it is assumed that the investor has his own views on market returns. These views are formulated as a probabilistic statement on the returns of the so-called view portfolios. Loosely speaking, one has

$$P\mu \sim \mathcal{N}(v, \Omega),$$

where P is the matrix used to construct the view portfolios, v is the vector of expected returns of these portfolios according to the investor's views and Ω represents the uncertainty of the investor in his views. More precisely, we assume

---

[2] The covariance matrix of μ does not need to be a multiple of Σ but this is usually the case in the literature.

that

$$P\mu = v + \varepsilon^{\text{views}},$$

where $\varepsilon^{\text{views}}$ is a random variable corresponding to the error in the investor's views. Furthermore, it is assumed that

$$\varepsilon^{\text{views}} \sim \mathcal{N}(0, \Omega),$$

and that $(\varepsilon^{\text{eq}}, \varepsilon^{\text{views}})$ are jointly Gaussian and independent.

It is easy to deduce that the posterior distribution for μ is again normally distributed, $\mathcal{N}(\hat{\mu}, \hat{\Sigma})$, with parameters

$$\begin{aligned}
\hat{\mu} &= \hat{\Sigma}\left((\tau\Sigma)^{-1}\mu^{\text{eq}} + P'\Omega^{-1}v\right), \\
\hat{\Sigma} &= \left((\tau\Sigma)^{-1} + P'\Omega^{-1}P\right)^{-1}.
\end{aligned}$$

As a consequence, the vector of excess returns is distributed according to a Gaussian distribution, $R \sim \mathcal{N}\left(\hat{\mu}, \Sigma + \hat{\Sigma}\right)$, and the optimal allocation is

$$\pi^{BL} = \frac{\eta}{x_0}\left(\Sigma + \hat{\Sigma}\right)^{-1}\hat{\mu},$$

where, as before, η is the mean–variance trade-off coefficient of the investor and x_0 is his initial wealth.

We observe that subjective views may be due to analyst's recommendations, or they may result from an application of statistical methods to historical data, as, for example, in Beach and Orlov (2007). As a simple illustration, suppose that the views result from using the sample mean to estimate expected returns on historical data spanning T years. In this case, we have $P = \text{Id}$ and $v = \bar{\mu}$ where $\bar{\mu}$ is the sample mean. Furthermore, given the properties of the sample mean estimator, it makes sense to consider

$$\Omega = \frac{1}{T}\Sigma,$$

which corresponds to the variance of the estimate $\bar{\mu}$. Then, we can rewrite $\hat{\mu}, \hat{\Sigma}$ and π^{BL} as

$$\begin{aligned} \hat{\Sigma} &= \frac{1}{T+\tau^{-1}}\Sigma, \\ \hat{\mu} &= \frac{\tau^{-1}\mu^{\text{eq}} + T\bar{\mu}}{\tau^{-1}+T}, \\ \pi^{BL} &= \frac{\eta}{x_0}\frac{\tau^{-1}\Sigma^{-1}\mu^{\text{eq}} + T\Sigma^{-1}\bar{\mu}}{1+\tau^{-1}+T}. \end{aligned}$$

Note that $\hat{\mu}$ is a weighted average of the CAPM equilibrium, μ^{eq}, and the sample mean estimator, $\bar{\mu}$. In other words, $\hat{\mu}$ is the result of shrinking the sample mean estimator to the CAPM equilibrium.

In the sequel, we follow a similar approach, in that we shrink the estimator of expected returns to its equilibrium value. The difference is that we do so in continuous time. As in this section, we work in a Bayesian framework.

3.2 The drift estimator

In this section, we derive a novel estimator for the drift of a diffusion. This estimator has the property of being continuously shrunk towards a target process. We obtain such an estimator by making the drift a mean–reverting process and applying standard linear filtering techniques.

Estimating a constant drift in a diffusion

It is well known that shrinkage can help reduce estimation error — see, for instance, Stein (1956). The objective of this subsection is to motivate the use of shrinkage techniques in the specific context of drift estimation in a diffusion. For simplicity, we restrict ourselves to the one dimensional case.

Suppose a given process, denoted by $R_\cdot$, is known to follow the one dimensional stochastic differential equation (SDE)

$$dR_t = \mu dt + \sigma dW_t,$$

where σ is known but μ is unknown. The problem we are faced with is that of estimating μ.

If this process is sampled up to time t, then the maximum likelihood estimate of μ is simply

$$\hat{\mu}_t^{MLE} = \frac{R_t}{t}.$$

This estimator is unbiased and consistent, but has the disadvantage of having large variance for small t. Specifically, the variance of $\hat{\mu}_t^{MLE}$ is given by

$$\mathrm{Var}\left(\hat{\mu}_t^{MLE}\right) = \frac{\sigma^2}{t}.$$

Furthermore, given typical values for μ and σ, $\mathrm{Var}\left(\hat{\mu}_t^{MLE}\right)$ decreases too slowly with t. For example, if $\mu = 0.1$ and $\sigma = 0.2$, then it would take approximately 16 years of historical data to reject the hypothesis $\mu \neq 0$ at the 95% confidence level.

An alternative approach to maximum likelihood estimation is Bayesian estimation. In this case, one needs to specify a prior for μ. It is convenient to consider a Gaussian prior, denoted by

$$\mu_0 \sim \mathcal{N}(\hat{\mu}_0, \hat{\sigma}_0^2). \tag{3.6}$$

Given a sample of $R_\cdot$ up to time t, and for the choice of prior above, it is easy to show that the posterior distribution is also Gaussian, $\mu_t \sim \mathcal{N}\left(\hat{\mu}_t^{\text{Bayes}}, \hat{\sigma}_t^2\right)$, where

$$\hat{\mu}_t^{\text{Bayes}} = \frac{\hat{\sigma}_0^{-2}\hat{\mu}_0 + \sigma^{-2}R_t}{\hat{\sigma}_0^{-2} + \sigma^{-2}t} \text{ and } \hat{\sigma}_t^2 = \frac{1}{\hat{\sigma}_0^{-2} + \sigma^{-2}t}. \tag{3.7}$$

Given the posterior distribution, μ_t, natural estimators of μ are, for example, the mean, mode or median of μ_t. Since the posterior distribution is Gaussian, these estimators all coincide with $\hat{\mu}_t^{\text{Bayes}}$. This estimator is consistent and asymptotically unbiased. However, it may be biased — in fact, it is unbiased if and only if $\hat{\mu}_0 = \mu$. The advantage of $\hat{\mu}^{\text{Bayes}}$ over the maximum likelihood estimator is that it has a lower variance and, in particular, for small t, its variance does not explode.

The choice of prior in the Bayesian estimation of μ should reflect what is known as the bias–variance trade-off. In a nutshell, by decreasing the variance of the prior, $\hat{\sigma}_0^2$ in (3.6), one obtains an estimator with less variance but more bias. Since both the variance and bias of an estimator contribute to its mean squared error (MSE), it is possible to reduce the MSE by trading bias for variance. The MSE of the maximum likelihood and Bayesian estimators are, respectively,

$$\mathrm{MSE}_t^{MLE} = \frac{\sigma^2}{t} \quad \text{and} \quad \mathrm{MSE}_t^{\text{Bayes}}(\hat{\mu}_0, \hat{\sigma}_0) = \frac{\sigma^{-2}t + \hat{\sigma}_0^{-4}(\hat{\mu}_0 - \mu)^2}{(\hat{\sigma}_0^{-2} + \sigma^{-2}t)^2}. \tag{3.8}$$

We observe that, given the prior mean $\hat{\mu}_0$, $\mathrm{MSE}_t^{\text{Bayes}}(\hat{\mu}_0, \cdot)$ is minimised at the point

$$\sigma_0^* = |\hat{\mu}_0 - \mu|.$$

Also, it is interesting to note that

$$\begin{aligned}\mathrm{MSE}_t^{\text{Bayes}}(\hat{\mu}_0, \hat{\sigma}_0) &= \frac{1}{\hat{\sigma}_0^{-2} + \sigma^{-2}t} + \frac{\hat{\sigma}_0^{-4}(\hat{\mu}_0 - \mu)^2 - \hat{\sigma}_0^{-2}}{(\hat{\sigma}_0^{-2} + \sigma^{-2}t)^2} \\ &< \mathrm{MSE}_t^{MLE} + \frac{\hat{\sigma}_0^{-4}(\hat{\mu}_0 - \mu)^2 - \hat{\sigma}_0^{-2}}{(\hat{\sigma}_0^{-2} + \sigma^{-2}t)^2}.\end{aligned}$$

Therefore, if $\hat{\sigma}_0 \geq \sigma_0^*$, then

$$\mathrm{MSE}_t^{\text{Bayes}}(\hat{\mu}_0, \hat{\sigma}_0) < \mathrm{MSE}_t^{MLE}.$$

In particular, if $|\mu|$ is known to be strictly smaller than some constant c, then

$$\mathrm{MSE}_t^{\mathrm{Bayes}}(0, c) < \mathrm{MSE}_t^{MLE}.$$

In other words, shrinking to zero is beneficial whenever μ is sufficiently small.

The objective of the rest of this section is to develop an alternative Bayesian estimator for the drift of a diffusion. We follow a Bayesian approach similar to the one introduced here. However, in addition to the prior belief, we consider a shrinkage target which continuously attracts the estimator.

Preliminaries

Throughout the rest of the chapter, we work on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with

(i) two $\mathbb{R}^n$–valued Brownian motions, $W_\cdot^\mu$ and $W_\cdot^R$, such that

$$d\left\langle W^{\mu,i}, W^{R,j}\right\rangle_t = \rho_{i,j}^{\mu,R} dt;$$

(ii) a random variable $\mu_0 : \Omega \to \mathbb{R}^n$, independent of $\left(W_\cdot^\mu, W_\cdot^R\right)$ and normally distributed,

$$\mu_0 \sim \mathcal{N}\left(\hat{\mu}_0, \hat{\Sigma}_0\right). \tag{3.9}$$

We note that the correlated Brownian motions $W_\cdot^\mu$ and $W_\cdot^R$ can be easily constructed from a standard $\mathbb{R}^{2n}$–valued Brownian motion. We denote by $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$ the augmentation of the filtration generated by $\left(W_\cdot^\mu, W_\cdot^R\right)$.

We consider a market with n risky assets and denote their cumulative excess returns up to time t by R_t. We assume that $R_\cdot$ satisfies

$$\begin{cases} dR_t = \mu_t dt + \sigma dW_t^R \\ d\mu_t = \kappa(\mu_t^{\mathrm{eq}} - \mu_t)dt + \sigma_\mu dW_t^\mu, \end{cases} \tag{3.10}$$

where σ and σ_μ are $n \times n$ invertible matrices, κ is a $n \times n$ positive–definite matrix and $\mu^{\mathrm{eq}}_\cdot$ is a process.

We denote by $\mathbb{F}^R = \left\{\mathcal{F}^R_t : t \geq 0\right\}$ the augmentation of the filtration generated by the process $R_\cdot$.

The process $\mu_\cdot$ is unobservable. We follow the Bayesian approach and assume that there is a prior belief on the initial value of $\mu_\cdot$, i.e. μ_0 is a random variable. Furthermore, we assume this prior is independent of $W^\mu_\cdot$ and $W^R_\cdot$ and has a Gaussian distribution, (3.9).

The process $\mu^{\mathrm{eq}}_\cdot$ is the target of the shrinkage. Note that, by construction in (3.10), $\mu^{\mathrm{eq}}_\cdot$ acts as a center of gravity around which $\mu_\cdot$ should mean–revert. As a consequence, the estimator of μ_t we obtain for model (3.10) is biased towards μ^{eq}_t. The superscript in $\mu^{\mathrm{eq}}_\cdot$ stands for equilibrium in an allusion to the Black–Litterman setting where there is a theoretical equilibrium value for expected returns. We observe that $\mu^{\mathrm{eq}}_\cdot$ should only incorporate the information accessible to the investor. For this reason, we assume that $\mu^{\mathrm{eq}}_t \in \mathcal{F}^R_t$.

The estimator, $\hat{\mu}_t$

We now obtain the optimal estimator of μ_t, given the model in (3.10) and the choice of prior (3.9). By optimal estimator we mean the one that minimises the MSE. As expected, the optimal estimator is simply a conditional expectation, as shown in the following proposition.

Proposition 6. *The best estimator of μ_t in the mean squared error sense, is the conditional expectation $\hat{\mu}_t = \mathbb{E}\left[\mu_t | \mathcal{F}^R_t\right]$. More precisely,*

$$\min_{\hat{\mu} \in \mathcal{F}^R_t} \mathbb{E}\left[(\hat{\mu} - \mu_t)^2\right] = \mathbb{E}\left[(\hat{\mu}_t - \mu_t)^2\right].$$

Proof. The proof follows immediately from the following equality

$$\mathbb{E}\left[(\hat{\mu}-\mu_t)^2\right]=\mathbb{E}\left[(\hat{\mu}-\hat{\mu}_t)^2+(\hat{\mu}_t-\mu_t)^2\right],$$

which holds for any $\hat{\mu}\in\mathcal{F}_t^R$. □

It follows from the choice of model, (3.10), that the distribution of μ_t conditioned on $\mathcal{F}_t^R$ is Gaussian and, therefore, it is characterised by its first two moments,

$$\hat{\mu}_t=\mathbb{E}\left[\mu_t|\mathcal{F}_t^R\right], \tag{3.11}$$

$$\hat{\Sigma}_t=\mathbb{E}\left[(\mu_t-\hat{\mu}_t)(\mu_t-\hat{\mu}_t)'|\mathcal{F}_t^R\right]. \tag{3.12}$$

The dynamics of $\hat{\mu}_\cdot$ and $\hat{\Sigma}_\cdot$ can be computed explicitly using the well known Kalman–Bucy filter. This is the content of the following result.

Theorem 17. *Let* $(R_\cdot,\mu_\cdot)$ *satisfy* (3.10) *with initial condition* (3.9). *Assume that* μ_0 *is independent of* $W_\cdot^R$ *and* $W_\cdot^\mu$, *and that* $\mu_t^{eq}\in\mathcal{F}_t^R$. *Define* $\hat{\mu}_t$ *and* $\hat{\Sigma}_t$ *as in* (3.11) *and* (3.12), *respectively.*

Then, $\hat{\Sigma}_t=\mathbb{E}\left[\hat{\Sigma}_t\right]$ *is deterministic and*

$$\begin{cases} d\hat{\mu}_t=\kappa(\mu_t^{eq}-\hat{\mu}_t)dt+\left(\sigma_\mu\rho^{\mu,R}\sigma'+\hat{\Sigma}_t\right)\Sigma^{-1}\left(dR_t-\hat{\mu}_t dt\right), \\ d\hat{\Sigma}_t=\left(-\kappa\hat{\Sigma}_t-\hat{\Sigma}_t\kappa'-\left(\sigma_\mu\rho^{\mu,R}\sigma'+\hat{\Sigma}_t\right)\Sigma^{-1}\left(\sigma_\mu\rho^{\mu,R}\sigma'+\hat{\Sigma}_t\right)'+\Sigma_\mu\right)dt, \end{cases} \tag{3.13}$$

where $\Sigma=\sigma\sigma'$ *and* $\Sigma_\mu=\sigma_\mu\sigma_\mu'$. *Furthermore, we have*

$$dR_t=\hat{\mu}_t dt+\sigma dW_t, \tag{3.14}$$

where $W_\cdot$ *is a* $\mathbb{F}^R$*–Brownian motion.*

Proof. The proof follows from a simple application of the standard Kalman–Bucy filter — see, for instance, Liptser and Shiryaev (2001, Theorem 10.3, p. 396). However, because μ_t^{eq} may be stochastic, we need to perform a change of variables before applying the Kalman–Bucy filter. More precisely, we define

$$\begin{aligned}\tilde{\mu}_t &= \mu_t - e^{-\kappa t}\int_0^t e^{\kappa s}\kappa\mu_s^{\mathrm{eq}}ds,\\ \tilde{R}_t &= R_t - \int_0^t \left(\mu_s - \tilde{\mu}_s\right)ds.\end{aligned}$$

Then,

$$\begin{cases} d\tilde{R}_t = \tilde{\mu}_t dt + \sigma dW_t^R \\ d\tilde{\mu}_t = -\kappa\tilde{\mu}_t dt + \sigma_\mu dW_t^\mu. \end{cases} \tag{3.15}$$

Furthermore, we observe that $\tilde{R}_t \in \mathcal{F}_t^R$.

Applying the Kalman–Bucy filter to estimate $\tilde{\mu}_t$ in (3.15), we deduce that

$$\begin{cases} d\hat{\tilde{\mu}}_t = -\kappa\hat{\tilde{\mu}}dt + \left(\sigma_\mu\rho^{\mu,R}\sigma' + \hat{\Sigma}_t\right)\Sigma^{-1}\left(d\tilde{R}_t - \hat{\tilde{\mu}}dt\right), \\ d\hat{\Sigma}_t = \left(-\kappa\hat{\Sigma}_t - \hat{\Sigma}_t\kappa' - \left(\sigma_\mu\rho^{\mu,R}\sigma' + \hat{\Sigma}_t\right)\Sigma^{-1}\left(\sigma_\mu\rho^{\mu,R}\sigma' + \hat{\Sigma}_t\right)' + \Sigma_\mu\right)dt, \end{cases}$$

where $\hat{\tilde{\mu}}_t = \mathbb{E}\left[\tilde{\mu}_t|\mathcal{F}_t^R\right]$ and $\hat{\Sigma}_t$ is as in (3.12). Furthermore,

$$d\tilde{R}_t = \hat{\tilde{\mu}}_t dt + \sigma dW_t,$$

where $W_.$ is a $\mathbb{F}^R$–Brownian motion.

Finally, we observe that

$$\begin{aligned} d\hat{\tilde{\mu}}_t &= d\hat{\mu}_t + \kappa\left(\hat{\mu}_t - \hat{\tilde{\mu}}_t - \mu_t^{eq}\right)dt,\\ d\tilde{R}_t &= dR_t - (\mu_t - \tilde{\mu}_t)dt = dR_t - \left(\hat{\mu}_t - \hat{\tilde{\mu}}_t\right)dt, \end{aligned}$$

which concludes the proof. □

The equation for $\hat{\Sigma}_.$ in (3.13) is a multidimensional Riccati equation. In general, if the dimension is higher than one, it does not admit an explicit solution[3].

---

[3] In general, the Riccati equation may not even have a solution. However, since equation (3.13) can be written as $d\hat{\Sigma}_t = \left(-\hat{\Sigma}_t BB'\hat{\Sigma}_t - \hat{\Sigma}_t A - A'\hat{\Sigma}_t + C\right)dt$, for some A, B and C with

In the following corollary, we make simplifying assumptions which allow us to solve the Riccati equation and provide an explicit formula for the estimator $\hat{\mu}_t$. We observe that these assumptions are automatically satisfied in the unidimensional case. However, in dimensions greater than one, such assumptions are quite stringent[4].

Corollary 1. *In Theorem 17, suppose that* $\sigma_\mu = \tau\sigma$, $\hat{\Sigma}_0 = \tau_0\Sigma$, $\rho^{\mu,R} = \rho\, Id$ *and* $\kappa = k\, Id$, *where* Id *is the* $n \times n$ *identity matrix and* τ_0, τ, k *and* ρ *are scalars. Let* $\alpha = \sqrt{(k+\rho\tau)^2 + \tau^2(1-\rho^2)}$ *and assume that* $\alpha \neq 0$. *Then,* $\hat{\Sigma}_t = \tau_t\Sigma$, *where*

$$\tau_t = -k - \rho\tau + \alpha\frac{\phi^+(t) + \phi^-(t)}{\phi^+(t) - \phi^-(t)}, \tag{3.16}$$

$$\phi^\pm(t) = \frac{(\tau_0 + k + \rho\tau \pm \alpha)}{2\alpha}e^{\pm\alpha t}. \tag{3.17}$$

Furthermore, the optimal filter, $\hat{\mu}_t$, *can be written as*

$$\hat{\mu}_t = \frac{\hat{\mu}_0 + \int_0^t \phi_s^{eq}\mu_s^{eq}ds + \int_0^t \phi_s^R dR_s}{1 + \int_0^t \phi_s^{eq}ds + \int_0^t \phi_s^R ds}, \tag{3.18}$$

where

$$\begin{cases} \phi_t^{eq} = k(\phi^+(t) - \phi^-(t)) \\ \phi_t^R = \alpha(\phi^+(t) + \phi^-(t)) - \phi_t^{eq}. \end{cases} \tag{3.19}$$

The denominator in (3.18) *is given by*

$$1 + \int_0^t \phi_s^{eq}ds + \int_0^t \phi_s^R ds = \phi^+(t) - \phi^-(t).$$

Proof. This result is an immediate consequence of Theorem 17. The solution to the Riccati equation is provided in Appendix 3.A. □

$C \geq 0$ such a solution exists and is unique. Furthermore, since $\hat{\Sigma}_0 \geq 0$, then $\hat{\Sigma}_t \geq 0$, for all $t \geq 0$ — see Theorems 8 and 9 in Kučera (1973).

[4] Although stringent, we observe that the assumption $\hat{\Sigma}_0 = \tau_0\Sigma$ is a common one in the Black–Litterman literature — the assumption $\sigma_\mu = \tau\sigma$ is obviously very similar in nature. The assumption $\rho^{\mu,R} = \rho\,\mathrm{Id}$ is also a very strong one but it includes the important case $\rho^{\mu,R} = 0$.

Under the hypothesis of Corollary 1, the dynamics for the filter $\hat{\mu}_\cdot$ can be simplified to

$$d\hat{\mu}_t = (-k - \rho\tau - \tau_t)\hat{\mu}_t dt + k\mu_t^{\text{eq}} dt + (\tau_t + \rho\tau) dR_t.$$

Discretising this equation using a time step δ, we obtain the approximate formula to update $\hat{\mu}_\cdot$:

$$\hat{\mu}_{t+\delta} \approx (1 - \delta\tau_t - \delta k - \delta\rho\tau)\hat{\mu}_t + \delta k\mu_t^{\text{eq}} + \delta(\tau_t + \rho\tau)\frac{R_{t+\delta} - R_t}{\delta}. \tag{3.20}$$

The updating formula (3.20) reveals the intuition behind $\hat{\mu}_\cdot$. Essentially, $\hat{\mu}_\cdot$ is an exponential moving average. At time t, the weight of a new observation of μ_t^{eq} is δk, while the weight given to a new observation of returns is $\delta(\tau_t + \rho\tau)$. We observe that, given the approximate updating formula (3.20), one can estimate the drift in real time very efficiently.

The variance of the estimator $\hat{\mu}_t$ is determined by the function $\tau_\cdot$ defined in (3.16). In practice, we want τ_t to be small, so that the estimator $\hat{\mu}_t$ does not to fluctuate too much. We observe that $\tau_\cdot$ is monotone and satisfies

$$\tau_\infty := \lim_{t\to\infty} \tau_t = \alpha - k - \rho\tau = \sqrt{(k + \rho\tau)^2 + \tau^2(1 - \rho^2)} - k - \rho\tau.$$

In Figure 3.1, we plot the function $\tau_\cdot$ for different values of τ and k, when $\tau_0 = 10^{-1}$ and $\rho = 0$.

If the objective is to predict future returns, then the quantity of interest is $\mathbb{E}\left[R_T | \mathcal{F}_t^R\right]$ for some $T > t$. This can be computed explicitly from the results in Theorem 17. Indeed, from

$$\mathbb{E}\left[\hat{\mu}_T | \mathcal{F}_t^R\right] = e^{-\kappa(T-t)}\hat{\mu}_t + \int_t^T \kappa e^{-\kappa(T-s)} \mathbb{E}\left[\mu_s^{\text{eq}} | \mathcal{F}_t^R\right] ds,$$

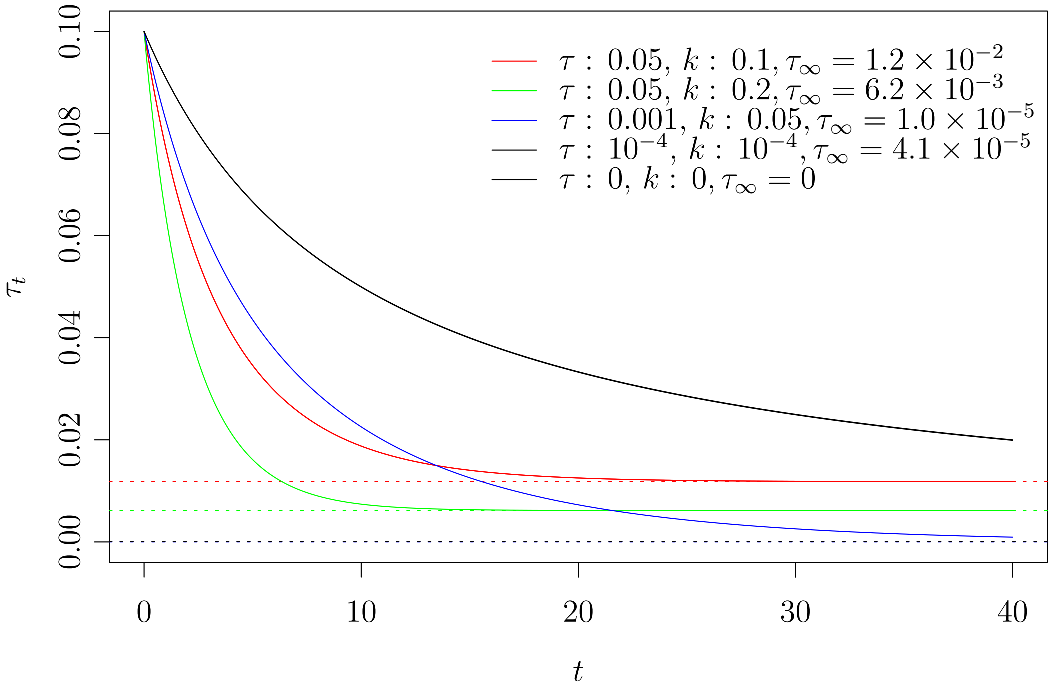

Figure 3.1: Plot of $\tau_\cdot$ for different values of τ and k, with $\tau_0 = 10^{-1}$ and $\rho = 0$.

we deduce that the optimal estimator for future cumulative excess returns is given by

$$\mathbb{E}\left[R_T | \mathcal{F}_t^R\right] = R_t + \hat{\mu}_t(T-t) + \int_t^T \left(\mathrm{Id} - e^{-\kappa(T-s)}\right)\left(\mathbb{E}\left[\mu_s^{\mathrm{eq}} | \mathcal{F}_t^R\right] - \hat{\mu}_t\right) ds. \quad (3.21)$$

For $T \approx t$, one can approximate (3.21) by

$$\mathbb{E}\left[R_T | \mathcal{F}_t^R\right] \approx R_t + \hat{\mu}_t(T-t).$$

Before proceeding, we consider three limiting cases of the result in Corollary 1:

- If $\tau = 0$, then $\phi_t^R = \tau_0 e^{-kt}$ and

  $$\tau_\infty = 0.$$

  In this case, μ_t approaches μ_t^{eq} exponentially fast. Therefore, returns contribute very little to estimate μ_t and the variance of the estimator goes to zero.

- If $k \to \infty$, then, for any $t > 0$,

  $$\phi_t^R \to 0 \text{ and } \tau_t \to 0.$$

  In this case, $k \to \infty$ forces μ_t to coincide with μ_t^{eq}. Therefore, returns are not used to estimate μ_t and there is no uncertainty at all on the estimation.

- If $k = 0$, then

  $$\phi_t^{\text{eq}} = 0 \text{ and } \tau_\infty = (1-\rho)\tau.$$

  In this case, μ_t^{eq} does not contain any information about μ_t. However, unless $\rho = 1$ or $\tau = 0$, the variance of the estimator does not go to zero because of the random shocks perturbing $\mu_.$.

- If $k^2 + \tau^2 \to 0$, then, for any $t > 0$,

  $$\phi_t^{\text{eq}} \to 0, \;\; \phi_t^R \to \tau_0, \;\; \tau_t \to \frac{\tau_0}{1+\tau_0 t} \text{ and } \tau_\infty \to 0.$$

  This corresponds to the case where the vector of expected excess returns is unknown but constant — the so-called Bayesian uncertainty.

  In this case, $\hat{\mu}_t$ reduces to

  $$\frac{\hat{\mu}_0 + \tau_0 R_t}{1 + \tau_0 t} = \hat{\mu}_t^{\text{Bayes}},$$

  where $\hat{\mu}_t^{\text{Bayes}}$ is defined in (3.7).

Properties of the estimator $\hat{\mu}_t$

Now, we study some of the properties of the estimator $\hat{\mu}_t$ given by (3.18). We assume that there is an unknown parameter μ such that

$$R_t = \mu t + \sigma W_t^R.$$

The estimator $\hat{\mu}_t$ is a biased one. However, as the following proposition shows, it may be asymptotically unbiased if the target process $\mu^{\text{eq}}_{\cdot}$ converges, on average, to μ.

Proposition 7. *Suppose that*

$$\mathbb{E}\left[\mu_t^{eq}\right] \to \mu,$$

as $t \to \infty$*. Then,* $\hat{\mu}_t$ *is asymptotically unbiased, that is,*

$$\mathbb{E}[\hat{\mu}_t] \to \mu,$$

as $t \to \infty$*.*

Proof. We note that

$$\mathbb{E}[\hat{\mu}_t] - \mu = \frac{\hat{\mu}_0 - \mu + \int_0^t \phi_s^{\text{eq}} \left(\mathbb{E}[\mu_s^{\text{eq}}] - \mu\right) ds}{1 + \int_0^t \phi_s^{\text{eq}} ds + \int_0^t \phi_s^R ds}.$$

Therefore, it suffices to show that

$$\frac{\int_0^t \phi_s^{\text{eq}} \left|\mathbb{E}[\mu_s^{\text{eq}}] - \mu\right| ds}{\int_0^t \phi_s^{\text{eq}} ds} \to 0, \tag{3.22}$$

as $t \to \infty$.

We observe that

$$\frac{\int_{t_n}^{t_{n+1}} \phi_s^{\text{eq}} \left|\mathbb{E}[\mu_s^{\text{eq}}] - \mu\right| ds}{\int_{t_n}^{t_{n+1}} \phi_s^{\text{eq}} ds} = \xi_n,$$

where ξ_n is such that

$$\xi_n \in \left[\inf_{s\in[t_n,t_{n+1}]} \left|\mathbb{E}[\mu_s^{\text{eq}}] - \mu\right|, \sup_{s\in[t_n,t_{n+1}]} \left|\mathbb{E}[\mu_s^{\text{eq}}] - \mu\right|\right].$$

Therefore, since $\xi_n \to 0$, as $n \to \infty$, the limit in (3.22) follows easily from the Stolz–Cesàro theorem — see, for instance, Theorem 1.22 in Mureşan (2009, p. 85). □

The variance and mean squared error of the estimator depend on the properties of $\mu^{\text{eq}}_{\cdot}$. For this reason, in the subsequent analysis, we consider only the particular case of constant $\mu^{\text{eq}}_{\cdot}$. More precisely, we assume that $\mu^{\text{eq}}_{\cdot} = \hat{\mu}_0$. In other words, we continuously shrink the estimate towards the initial target $\hat{\mu}_0$. In this case, we have

$$\hat{\mu}_t = \frac{\hat{\mu}_0 \left(1 + \int_0^t \phi^{\text{eq}}_s ds\right) + \int_0^t \phi^R_s dR_s}{1 + \int_0^t \phi^{\text{eq}}_s ds + \int_0^t \phi^R_s ds}.$$

Alternatively, it is possible to write $\hat{\mu}_t$ as

$$\hat{\mu}_t = \frac{\hat{\mu}_0 + \int_0^t \varphi_s dR_s}{1 + \int_0^t \varphi_s ds}, \tag{3.23}$$

where $\varphi_{\cdot}$ is a deterministic function, the definition of which we omit as it is not relevant in the sequel. Note the similarities between $\hat{\mu}_t$, given by (3.23), and $\hat{\mu}_t^{\text{Bayes}}$, given by (3.7). Indeed, $\hat{\mu}_t^{\text{Bayes}}$ is a particular case of $\hat{\mu}_t$.

The mean squared error of $\hat{\mu}_t$ can be easily computed,

$$\text{MSE}_t(\hat{\mu}_0, \tau_0, k, \tau) = \frac{\text{Tr}(\Sigma) \int_0^t (\phi^R_s)^2 ds + |\hat{\mu}_0 - \mu|^2 \left(1 + \int_0^t \phi^{\text{eq}}_s ds\right)^2}{\left(1 + \int_0^t \phi^R_s ds + \int_0^t \phi^{\text{eq}}_s ds\right)^2},$$

where the dependence on τ_0, k and τ is hidden in the definition of the functions $\phi^R_{\cdot}$ and $\phi^{\text{eq}}_{\cdot}$. Recall the definition of $\text{MSE}_t^{\text{Bayes}}(\hat{\mu}_0, \hat{\sigma}_0)$, given in (3.8). Then, since $\hat{\mu}_t^{\text{Bayes}}$ is a particular case of $\hat{\mu}_t$, we must have

$$\lim_{k,\tau \to 0} \text{MSE}_t(\hat{\mu}_0, \tau_0, k, \tau) = \text{MSE}_t^{\text{Bayes}}\left(\hat{\mu}_0, \sqrt{\tau_0}\sigma\right).$$

It is interesting to check whether there exist parameters k and τ such that

$$\text{MSE}_t(\hat{\mu}_0, \tau_0, k, \tau) < \text{MSE}_t^{\text{Bayes}}(\hat{\mu}_0, \sqrt{\tau_0}\sigma). \tag{3.24}$$

If the inequality above holds, then the estimator $\hat{\mu}_t$ performs better than $\hat{\mu}_t^{\text{Bayes}}$, in estimating μ. It is easy to find examples of parameters τ_0, k and τ for which (3.24) does indeed hold. However, as the next result shows, we have

$$\min_{k,\tau,\tau_0} \mathrm{MSE}_t(\hat{\mu}_0, \tau_0, k, \tau) = \min_{\tau_0} \mathrm{MSE}_t^{\mathrm{Bayes}}(\hat{\mu}_0, \sqrt{\tau_0}\sigma).$$

Proposition 8. *Let $\hat{\mu}_t$ be an estimator of μ given by* (3.23)*, for some function $\varphi_\cdot$. Then, its mean squared error, $MSE_t = \mathbb{E}\left[|\hat{\mu}_t - \mu|^2\right]$, satisfies the inequality*

$$MSE_t \geq \frac{|\hat{\mu}_0 - \mu|^2}{1 + \frac{|\hat{\mu}_0 - \mu|^2}{\mathrm{Tr}(\Sigma)} t}.$$

Furthermore, this lower bound is attained when $\varphi_\cdot$ is constant and

$$\varphi_t = \frac{|\hat{\mu}_0 - \mu|^2}{\mathrm{Tr}(\Sigma)}.$$

Proof. Let $\hat{\mu}_t$ and $\varphi_\cdot$ be as in the hypothesis of the proposition. Then, the mean squared error of $\hat{\mu}_t$ is given by

$$\mathrm{MSE}_t = \frac{\mathrm{Tr}(\Sigma) \int_0^t \varphi_s^2 ds + |\hat{\mu}_0 - \mu|^2}{\left(1 + \int_0^t \varphi_s ds\right)^2}.$$

Therefore, by Jensen's inequality,

$$\mathrm{MSE}_t \geq \frac{\frac{\mathrm{Tr}(\Sigma)}{t} \left(\int_0^t \varphi_s ds\right)^2 + |\hat{\mu}_0 - \mu|^2}{\left(1 + \int_0^t \varphi_s ds\right)^2}. \tag{3.25}$$

It is now an easy exercise in differential calculus to show that the right-hand side of (3.25) is minimised when

$$\int_0^t \varphi_s ds = t \frac{|\hat{\mu}_0 - \mu|^2}{\mathrm{Tr}(\Sigma)}.$$

This concludes the proof. $\square$

A consequence of the previous proposition is that

$$\mathrm{MSE}_t(\hat{\mu}_0, \tau_0^*, k, \tau) \geq \mathrm{MSE}_t^{\mathrm{Bayes}}\left(\hat{\mu}_0, \sqrt{\tau_0^*}\sigma\right),$$

where $\tau_0^* = \frac{|\hat{\mu}_0 - \mu|^2}{\mathrm{Tr}(\Sigma)}$. In general, it can be shown that, if $\tau_0 \leq \tau_0^*$, then

$$\mathrm{MSE}_t(\hat{\mu}_0, \tau_0, k, \tau) \geq \mathrm{MSE}_t^{\mathrm{Bayes}}(\hat{\mu}_0, \sqrt{\tau_0}\sigma),$$

for any k and τ. In other words, if the initial shrinkage is strong enough (τ_0 is small enough), then any further shrinkage (by means of k and τ) is not beneficial. Therefore, we reach the negative conclusion that, if the drift is fixed but unknown, the estimator $\hat{\mu}_t$ given by (3.18) adds little, if any, value. This is not surprising since the Bayes estimator, $\hat{\mu}_t^{\mathrm{Bayes}}$, is designed specifically for the case of a constant drift. However, in the context where there is an equilibrium state from which the drift might occasionally deviate — for example, when there may be temporary mispricings — such an estimator may prove valuable.

3.3 The optimal allocation

In this section, we study the problem of optimal asset allocation for the model (3.10).

We recall that, as a consequence of Theorem 17, cumulative excess returns satisfy the so-called *a posteriori* dynamics,

$$dR_t = \hat{\mu}_t dt + \sigma dW_t, \tag{3.26}$$

where $\hat{\mu}_.$ satisfies the system of equations (3.13) and $W_.$ is a $\mathbb{F}^R$–Brownian motion.

Given a $\mathbb{F}^R$–progressively measurable allocation $\pi_.$, we denote its associated discounted wealth process by $X^\pi_\cdot$ and define it as the solution to the SDE

$$dX^\pi_t = X^\pi_t \pi'_t dR_t,$$

with initial condition $X^\pi_0 = x_0$. For a given horizon T, the set of admissible allocations is defined as

$$\mathcal{A}_T = \left\{ \pi_\cdot \ \mathbb{F}^R\text{–progressively measurable} : X^\pi_\cdot > 0, \, \mathbb{E}\left[\sup_{s\in[0,T]} |X^\pi_s|^2 \right] < \infty \right\}.$$

Forward investment

In this case, the definition of a forward investment performance process coincides with that of Chapter 2, i.e. Definition 3 which, for convenience, we rewrite below.

Definition 6. *Let $\mathcal{A} = \cup_{T>0}\mathcal{A}_T$ denote the set of admissible portfolios. A $\mathbb{F}^R$–adapted process $U(t,x) : [0,\infty)\times\mathbb{R}^+\times\Omega \to \mathbb{R}$ is a forward investment performance process if*

(i) for each $t \geq 0$, the mapping $U(t,\cdot)$ is increasing and strictly concave,

(ii) for each $\pi \in \mathcal{A}$, the process $t \mapsto U(t, X^\pi_t)$ is a $\mathbb{F}^R$–supermartingale, and

(iii) for each $(t_0, x_0) \in [0,\infty)\times\mathbb{R}^+$, there exists a $\pi^*_\cdot \in \mathcal{A}$ such that $X^{\pi^*}_{t_0} = x_0$ and the process $t \mapsto U\left(t, X^{\pi^*}_t\right)$ is a $\mathbb{F}^R$–martingale in $[t_0,\infty)$.

Given a, sufficiently regular, forward investment performance process $U(t,x)$, we know from Musiela and Zariphopoulou (2010b) that the following forward SPDE holds,

$$dU(t,x) = \frac{1}{2}\frac{\|U_x(t,x)\sigma^{-1}\hat{\mu}_t + a_x(t,x)\|^2}{U_{xx}(t,x)}dt + a(t,x)'dW_t, \tag{3.27}$$

with initial condition $U(0,x) = U(x)$.

In the sequel we consider another, closely related, asset allocation problem — expected utility maximisation. In that case, we make use of the results in Brendle (2006) to derive the optimal allocation for some, widely used, terminal utilities.

Expected utility maximisation

In an expected utility maximisation problem, the optimal allocation is a solution to the problem

$$\sup_{\pi \in \mathcal{A}_T} \mathbb{E}[U(X_T^\pi) | X_0^\pi = x_0], \tag{3.28}$$

where $U(\cdot)$ is a given utility function and x_0 is the initial wealth. In order to find the optimal allocation to (3.28), it is typical to introduce the (indirect) value function, $U(t, x) : [0, T] \times \mathbb{R}^+ \to \mathbb{R}$, defined as

$$U(t, x) = \sup_{\pi \in \mathcal{A}_T} \mathbb{E}\left[U(X_T^\pi) | \mathcal{F}_t^R, \, X_t^\pi = x\right],$$

and derive the associated Hamilton–Jacobi–Bellman (HJB) equation.

Due to the presence of the process $\mu_\cdot^{\mathrm{eq}}$ in the dynamics of $\hat{\mu}_\cdot$, the system of equations comprised of (3.26) and (3.13) may not be Markovian. For this reason, it may not be possible to write the stochastic process $U(t, x)$ as a deterministic function of the state variables R_t and $\hat{\mu}_t$. In this general case, the following stochastic HJB equation holds,

$$dU(t, x) = \frac{1}{2} \frac{\|U_x(t, x)\sigma^{-1}\hat{\mu}_t + a_x(t, x)\|^2}{U_{xx}(t, x)} dt + a(t, x)' dW_t, \tag{3.29}$$

with terminal condition $U(T, \cdot) = U(\cdot)$, see Peng (1992). We observe that in (3.29), the process $a(t, x)$ is also a part of the solution and is required in order for the process $U(t, x)$ to be adapted — as opposed to (3.27) where $a(t, x)$ is given. This is a common feature of backward stochastic differential equations (BSDEs), see, for instance, El Karoui, Peng, and Quenez (1997).

If $\mu_t^{\mathrm{eq}} = f(t, R_t)$, for some function $f(t,z) : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$, then the system comprised of (3.26) and (3.13) becomes a Markovian system. As a consequence, $U(t,x)$ is a deterministic function of R_t and $\hat{\mu}_t$, that is

$$U(t,x) = u(t, x, R_t, \hat{\mu}_t),$$

where $u(t,x,z,m) : [0,\infty) \times \mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined as

$$u(t,x,z,m) = \sup_{\pi \in \mathcal{A}_T} \mathbb{E}\left[U(X_T^\pi) | X_t^\pi = x, R_t = z, \hat{\mu}_t = m\right].$$

In this case, the SPDE (3.29) leads to the following PDE for $u(\cdot)$,

$$\begin{aligned} 0 = & u_t + ku_m'(f(t,z) - m) + u_z' m + \frac{1}{2}\mathrm{Tr}(u_{zz}\Sigma) + \mathrm{Tr}\left(u_{mz}\left(\sigma_\mu \rho^{\mu,R}\sigma' + \hat{\Sigma}_t\right)\right) + \\ & \frac{1}{2}\mathrm{Tr}\left(u_{mm}\left(\sigma_\mu \rho^{\mu,R}\sigma' + \hat{\Sigma}_t\right)' \Sigma^{-1}\left(\sigma_\mu \rho^{\mu,R}\sigma' + \hat{\Sigma}_t\right)\right) + \\ & - \frac{1}{2}\frac{\left\| u_x \sigma^{-1} m + \sigma^{-1}\left(\sigma_\mu \rho^{\mu,R}\sigma' + \hat{\Sigma}_t\right)' u_{xm} + \sigma' u_{xz}\right\|^2}{u_{xx}}. \end{aligned} \tag{3.30}$$

The process $a(t,x)$ in (3.29) is written in terms of $u(\cdot)$ as

$$a(t,x) = \sigma' u_z(t,x,R_t,\hat{\mu}_t) + \sigma^{-1}\left(\sigma_\mu \rho^{\mu,R}\sigma' + \hat{\Sigma}_t\right)' u_m(t,x,R_t,\hat{\mu}_t).$$

Furthermore, the optimal portfolio for the expected utility maximisation problem (3.28) is given, in a feedback form, by

$$\begin{aligned} \pi^*(t,x,z,m) = & - \frac{u_x}{x u_{xx}}(t,x,z,m)\Sigma^{-1} m + \\ & - \frac{u_x}{x u_{xx}}\left(\left(\sigma_\mu \rho^{\mu,R}\sigma' + \hat{\Sigma}_t\right)' \frac{u_{xm}}{u_x} + \Sigma \frac{u_{xz}}{u_x}\right)(t,x,z,m). \end{aligned} \tag{3.31}$$

For a general utility, $U(\cdot)$, it is not possible to explicitly solve (3.30). However, for logarithmic and power utilities explicit solutions are available, as we show in the next subsections.

Log utility

Suppose that the utility function, $U(\cdot)$, is logarithmic,

$$U(x) = \log(x).$$

Then, it is easy to see that the solution to the HJB equation (3.30) can be decomposed as

$$u(t, x, z, m) = \log(x) + G(t, z, m),$$

where $G(t, z, m) = u(t, 1, z, m)$. In this case, the optimal allocation for the expected utility maximisation problem is the so-called myopic allocation, $\pi^{\text{myopic}}_{\cdot}$, defined as

$$\pi^{\text{myopic}}_t = \Sigma^{-1} \hat{\mu}_t. \tag{3.32}$$

If, furthermore, $\mu^{\text{eq}}_{\cdot} = \bar{\mu}$ is a constant, then

$$G(t, z, m) = m' A(t) m + m' B(t) + C(t),$$

for some functions $A(\cdot), B(\cdot)$ and $C(\cdot)$ satisfying the terminal condition $A(T) = B(T) = C(T) = 0$.

We observe that, because $\hat{\mu}_{\cdot}$ is correlated with the process of cumulative excess returns, the myopic allocation (3.32) has some of the features of a trend–following allocation. Indeed, $d\pi^{\text{myopic}}_t$ has an instantaneous perfect correlation with dR_t, in the sense that

$$\frac{d\langle \pi^{\text{myopic}}, R' \rangle_t}{dt} = \sqrt{\frac{d\langle \pi^{\text{myopic}}, \pi^{\text{myopic}\prime} \rangle_t}{dt} \frac{d\langle R, R' \rangle_t}{dt}}.$$

In other words, if we observe a positive (respectively, negative) return, then the myopic allocation increases (respectively, decreases).

Power utility

Suppose that $\mu_t^{\mathrm{eq}} = \bar{\mu}$ is a constant and

$$U(x) = \frac{x^{1-\gamma}}{1-\gamma},$$

for some $\gamma > 1$. In this case, it is easy to see that the solution to the HJB equation, $u(\cdot)$, can be factored as

$$u(t, x, z, m) = U(x) \exp\left(G(t, m)\right),$$

where $G(\cdot)$ is given below.

Furthermore, from the results in Brendle (2006), we deduce that

$$G(t, m) = m'\tilde{A}(t)m + m'\tilde{B}(t) + \tilde{C}(t),$$

where $\tilde{C}(t) = \log|(1-\gamma)u(t, 1, z, 0)|$,

$$\begin{aligned} \tilde{A}(t) &= \left(\mathrm{Id} - 2A(t)\hat{\Sigma}_t\right)^{-1} A(t), \\ \tilde{B}(t) &= \left(\mathrm{Id} - 2A(t)\hat{\Sigma}_t\right)^{-1} B(t), \end{aligned}$$

and the functions $A(\cdot)$ and $B(\cdot)$ satisfy the system of equations

$$\begin{cases} \dot{A} = -\frac{1}{2}\frac{1-\gamma}{\gamma}\left(\mathrm{Id} + 2\sigma\rho^{\mu,R'}\sigma_\mu' A\right)' \Sigma^{-1} \left(\mathrm{Id} + 2\sigma\rho^{\mu,R'}\sigma_\mu' A\right) \\ \quad + \kappa' A + A\kappa - 2A\Sigma_\mu A \\ \dot{B} = -2A\left(\Sigma_\mu + \frac{1-\gamma}{\gamma}\sigma_\mu\rho^{\mu,R}\rho^{\mu,R'}\sigma_\mu'\right) B - \frac{1-\gamma}{\gamma}\Sigma^{-1}\sigma\rho^{\mu,R'}\sigma_\mu' B \\ \quad + \kappa B - 2A\kappa\bar{\mu}, \end{cases} \tag{3.33}$$

with terminal conditions $A(T) = B(T) = 0$. We observe that, if $\bar{\mu} = 0$, then $B(\cdot) = 0$. Furthermore, from well known results on the matrix Riccati equation, it follows that $A(\cdot) \leq 0$, see, for instance, Kučera (1973).

In this case, the optimal allocation is given in terms of $A(\cdot)$ and $B(\cdot)$ by

$$\pi_t^* = \frac{1}{\gamma}\Sigma^{-1}\left(\hat{\mu}_t + \left(\sigma_\mu \rho^{\mu,R}\sigma + \hat{\Sigma}_t\right)'\left(\mathrm{Id} - 2A(t)\hat{\Sigma}_t\right)^{-1}(2A(t)\hat{\mu}_t + B(t))\right). \quad (3.34)$$

In general, equation (3.33) does not admit an explicit solution in dimensions higher than one. We recall that $\hat{\Sigma}_\cdot$ solves the multidimensional Riccati equation, (3.13), which also lacks explicit solutions in dimensions higher than one.

However, as in Corollary 1, under simplifying assumptions, it is possible to reduce (3.33) to a system of unidimensional ordinary differential equations (ODEs).

Proposition 9. *Suppose that* $\hat{\Sigma}_0 = \tau_0 \Sigma$ *in* (3.9), *and that* $\sigma_\mu = \tau\sigma$, $\kappa = k\,Id$, *and* $\rho^{\mu,R} = \rho\,Id$ *in* (3.10), *where* τ_0, τ, k *and* ρ *are scalars.*

Let $\left(\hat{\mu}_\cdot, \hat{\Sigma}_\cdot\right)$ *be the solution to* (3.13), *and let* $A(\cdot)$ *and* $B(\cdot)$ *be the solutions to* (3.33).

Then,

$$\hat{\Sigma}_t = \tau_t\Sigma, \;\; A(t) = a(t)\Sigma^{-1}, \;\text{ and }\; B(t) = b(t)\Sigma^{-1}\bar{\mu},$$

where

$$\dot{\tau}_t = -2k\tau_t - (\tau\rho + \tau_t)^2 + \tau^2$$

and

$$\begin{cases} \dot{a} = -\frac{1}{2}\frac{1-\gamma}{\gamma}(1 + 2\tau\rho a)^2 + 2ka - 2\tau^2 a^2 \\ \dot{b} = -2ab\left(\tau^2 + \frac{1-\gamma}{\gamma}\rho^2\tau^2\right) - \frac{1-\gamma}{\gamma}\rho\tau b + kb - 2ka. \end{cases} \quad (3.35)$$

Furthermore, the optimal allocation, $\pi_\cdot^*$, *defined as in* (3.34), *is given by*

$$\pi_t^* = \frac{1}{\gamma(1 - 2a(t)\tau_t)}\Sigma^{-1}\left((1 + 2\rho\tau a(t))\hat{\mu}_t + (\rho\tau + \tau_t)b(t)\bar{\mu}\right).$$

Proof. The proof follows by a direct manipulation of (3.33). □

We observe that the system of equations in (3.35) has an explicit solution, as shown in Appendix 3.A.

3.4 Example: a multifactor model

In this section, we apply the results of the previous sections to the multifactor model introduced earlier in the chapter. These results are illustrated using data on the American equities market.

The model

First, we recall the model. The market consists of n risky assets whose returns are explained by the returns on m factor portfolios. Specifically, we denote the process of cumulative excess returns of the risky assets by $R_\cdot = (R^1_\cdot, \cdots, R^n_\cdot)$ and we denote the process of cumulative excess returns of the factor portfolio by $R^F_\cdot = \left(R^{F,1}_\cdot, \cdots, R^{F,m}_\cdot\right)$. Then, we assume that $\left(R^F_\cdot, R_\cdot\right)$ satisfy

$$\begin{cases} dR^F_t = \mu_t dt + \sigma_F dW^F_t, \\ dR_t = \alpha_t dt + \beta dR^F_t + \sigma_a dW^a_t, \end{cases} \tag{3.36}$$

where σ_F is an $m \times m$ matrix, σ_a is a diagonal $n \times n$ matrix, β is the $n \times m$ matrix of factor loadings, and $\left(W^F_\cdot, W^a_\cdot\right)$ is a standard $(m+n)$–dimensional Brownian motion. The dynamics of the processes $\mu_\cdot$ and $\alpha_\cdot$ are specified below.

We suppose that $n \gg m$. Then, the APT predicts that $\alpha_t \approx 0$; we use this theoretical prediction as a shrinkage target for the estimator of α_t. More precisely, we specify

$$d\alpha_t = -\kappa_\alpha \alpha_t dt + \sigma_\alpha dW^\alpha_t, \tag{3.37}$$

where σ_α is a diagonal $n \times n$ matrix, κ_α is a diagonal positive–definite $n \times n$

matrix and $W_\cdot^\alpha$ is a n–dimensional Brownian motion. For simplicity, we assume that $W_\cdot^\alpha$ is independent of $\left(W_\cdot^F, W_\cdot^a\right)$. The initial value of $\alpha_\cdot$ in (3.37) is a normally distributed random variable,

$$\alpha_0 \sim \mathcal{N}\left(0, \hat{\Sigma}_0^\alpha\right), \tag{3.38}$$

where $\hat{\Sigma}_0^\alpha$ is a diagonal positive–definite $n \times n$ matrix.

Remark 17. *The parameters σ_α and κ_α, in* (3.37)*, control the strength of the shrinkage towards zero of the estimator $\hat{\alpha}_t$, introduced below. Therefore, they should reflect the investor's belief in the APT prediction. For example, letting $\kappa_\alpha \to \infty$ and $\sigma_\alpha \to 0$, forces $\hat{\alpha}_t \to 0$, meaning absolute confidence in the APT prediction; similarly, letting $\kappa_\alpha \to 0$ and $\sigma_\alpha \to 0$, leads to an estimator $\hat{\alpha}_t$ which is not biased by the APT prediction.*

We assume that the risk factors have a theoretical price, denoted by $\bar{\mu}$. More precisely, we assume that the infinitesimal expected returns of the factor portfolios have a theoretical value of $\bar{\mu}\, dt$. As with the vector of alphas, we shrink our estimate of μ_t towards this theoretical prediction. More precisely, we specify

$$d\mu_t = -\kappa_\mu(\mu_t - \bar{\mu})dt + \sigma_\mu dW_t^\mu, \tag{3.39}$$

where σ_μ is a $m \times m$ matrix, κ_μ is a positive–definite $m \times m$ matrix and $W_\cdot^\mu$ is a m–dimensional Brownian motion. For simplicity, we assume that $W_\cdot^\mu$ is independent of $\left(W_\cdot^F, W_\cdot^a, W_\cdot^\alpha\right)$. The initial value of $\mu_\cdot$ in (3.39) is a normally distributed random variable,

$$\mu_0 \sim \mathcal{N}\left(\bar{\mu}, \hat{\Sigma}_0^\mu\right), \tag{3.40}$$

where $\hat{\Sigma}_0^\mu$ is a positive–definite $n \times n$ matrix.

Remark 18. *If a theoretical price for the risk factors, $\bar{\mu}$, is not available, then we can let $\kappa_\mu \to 0$ and $\sigma_\mu \to 0$ in* (3.39)*.*

Equations (3.36)–(3.39) constitute the multifactor model which we study in the sequel. We observe that such a model has a strong structure, particularly due to the independence assumptions; this structure is exploited leading to simplified versions of the results obtained in Sections 3.2 and 3.3.

The estimators, $\hat{\mu}_t$ and $\hat{\alpha}_t$

To exploit the independence structure of our model, it is convenient to introduce the process of abnormal returns, denoted as $R^a_\cdot$ and defined by

$$R^a_t = R_t - \beta R^F_t. \tag{3.41}$$

We observe that $R^F_\cdot$ and $R^a_\cdot$ are independent. Therefore, the problems of estimating the processes $\mu_\cdot$ and $\alpha_\cdot$ can be decoupled. Moreover, since the components of the vector R^a_t are also independent, we can estimate the different components of α_t separately. More precisely, it follows from Theorem 17 that the filtering equation for $\mu_\cdot$ is

$$\begin{cases} d\hat{\mu}_t = -\kappa_\mu(\hat{\mu}_t - \bar{\mu})dt + \hat{\Sigma}^\mu_t \Sigma_F^{-1}\left(dR^F_t - \hat{\mu}_t dt\right), \\ d\hat{\Sigma}^\mu_t = \left(-\kappa_\mu \hat{\Sigma}^\mu_t - \hat{\Sigma}^\mu_t \kappa'_\mu - \hat{\Sigma}^\mu_t \Sigma_F^{-1} \hat{\Sigma}^\mu_t + \Sigma_\mu\right) dt, \end{cases} \tag{3.42}$$

where $\Sigma_F = \sigma_F \sigma'_F$ and $\Sigma_\mu = \sigma_\mu \sigma'_\mu$.

Similarly, the estimator of α_t, $\hat{\alpha}_t = (\hat{\alpha}^1_t, \cdots, \hat{\alpha}^n_t)$, solves

$$\begin{cases} d\hat{\alpha}^i_t = -\kappa^i_\alpha \hat{\alpha}^i_t dt + \left(\sigma^i_a\right)^{-2} \hat{\Sigma}^{\alpha,i}_t \left(dR^{a,i}_t - \hat{\alpha}^i_t dt\right), \\ d\hat{\Sigma}^{\alpha,i}_t = \left(-2\kappa^i_\alpha \hat{\Sigma}^{\alpha,i}_t - \left(\sigma^i_a\right)^{-2}\left(\hat{\Sigma}^{\alpha,i}_t\right)^2 + \left(\sigma^i_\alpha\right)^2\right) dt, \end{cases} \tag{3.43}$$

where σ^i_a, σ^i_α and κ^i_α denote the i^{th} diagonal component of the matrices σ_a, σ_α and κ_α, respectively. The initial conditions for (3.42) and (3.43) are determined from (3.40) and (3.38), respectively.

It is possible to solve (3.43) explicitly without any extra assumptions. To solve

(3.42), we need assumptions on σ_F, σ_μ and κ_μ similar to the ones in Corollary 1.

The optimal allocation

Each allocation, $\pi_\cdot$, has two components,

$$\pi_t = \left(\tilde{\pi}_t^F, \pi_t^a\right),$$

where $\tilde{\pi}_t^F \in \mathbb{R}^m$ and $\pi_t^a \in \mathbb{R}^n$ represent the proportion of discounted wealth invested in the factor portfolios and in the risky assets, respectively.

We observe that, because the returns on the risky assets are themselves exposed to the returns on the factor portfolios, the component $\tilde{\pi}_\cdot^F$ does not represent the proportion of discounted wealth exposed to the m risk factors. In fact, the proportion of discounted wealth exposed to the risk factors is $\pi_\cdot^F$, where

$$\pi_t^F = \tilde{\pi}_t^F + \beta' \pi_t^a.$$

It is convenient to represent allocations in terms of the processes $\pi_\cdot^F$ and $\pi_\cdot^a$, instead of $\tilde{\pi}_\cdot^F$ and $\pi_\cdot^a$. Indeed, with such a representation, the dynamics of a wealth process, $X_\cdot^\pi$, can be decomposed in two independent components, that is

$$dX_t^\pi = X_t^\pi \left(\pi_t^{F\prime} dR_t^F + \pi_t^{a\prime} dR_t^a\right).$$

In other words, the instantaneous excess return on a portfolio $\pi_\cdot$ has two independent contributions: the return from the risk factors, $\pi_t^{F\prime} dR_t^F$, and the return from the idiosyncratic risks, $\pi_t^{a\prime} dR_t^a$.

In order to obtain an explicit optimal allocation, we consider the problem of expected utility maximisation with a power utility function $U(\cdot)$, where

$$U(x) = \frac{x^{1-\gamma}}{1-\gamma},$$

for some $\gamma > 1$. In this case, the the indirect value function takes the form

$$u(t,x,m,a) = \sup_{\pi \in \mathcal{A}_T} \ \mathbb{E}\left[U(X_T^\pi) | X_t^\pi = x, \ \hat{\mu}_t = m, \ \hat{\alpha}_t = a\right].$$

It follows from the structure of the model and the results in the previous section that

$$u(t,x,m,a) = U(x) \exp\left(G_F(t,m) + \sum_{i=1}^n G_i(t,a_i)\right), \tag{3.44}$$

where $a = (a_1, \cdots, a_n)$ and

$$\begin{aligned} G_F(t,m) &= m' \tilde{A}_F(t) m + m' \tilde{B}_F(t) + \tilde{C}_F(t), \\ G_i(t,a_i) &= a_i^2 \tilde{A}_i(t) + \tilde{C}_i(t). \end{aligned}$$

Furthermore, we have

$$\begin{aligned} \tilde{A}_F(t) &= \left(\mathrm{Id} - 2A_F(t)\hat{\Sigma}_t^\mu\right)^{-1} A_F(t), \\ \tilde{B}_F(t) &= \left(\mathrm{Id} - 2A_F(t)\hat{\Sigma}_t^\mu\right)^{-1} B_F(t), \\ \tilde{A}_i(t) &= \left(1 - 2A_i(t)\hat{\Sigma}_t^{\alpha,i}\right)^{-1} A_i(t), \end{aligned}$$

where $A_F(\cdot)$ and $B_F(\cdot)$ satisfy the system of equations

$$\begin{cases} \dot{A_F} = -\frac{1}{2}\frac{1-\gamma}{\gamma}\Sigma_F^{-1} + \kappa_\mu' A_F + A_F \kappa_\mu - 2A_F \Sigma_\mu A_F \\ \dot{B_F} = -2A_F \Sigma_\mu B_F + \kappa_\mu B_F - 2A_F \kappa_\mu \bar{\mu}. \end{cases} \tag{3.45}$$

Analogously, each of the the functions $A_i(\cdot)$ satisfies a one–dimensional Riccati equation which can be solved explicitly, (3.47) presented below.

Proposition 10. *The optimal allocation in the i^{th} risky asset at time t is given by*

$$\pi_t^{a,i,*} = \frac{1}{1 - 2A_i(t)\hat{\Sigma}_t^{\alpha,i}} \frac{\hat{\alpha}_t^i}{\gamma(\sigma_a^i)^2}, \tag{3.46}$$

where $\left(\hat{\alpha}_\cdot^i, \hat{\Sigma}_\cdot^{\alpha,i}\right)$ satisfy the system of equations (3.43)*, and $A_i(\cdot)$ is a solution to* *the Riccati equation*

$$\dot{A}_i = -\frac{1}{2}\frac{1-\gamma}{\gamma}\frac{\left(1+2\sigma_\alpha^i \rho_i^{\alpha,a}\sigma_a^i A_i\right)^2}{\left(\sigma_a^i\right)^2} + 2A_i\kappa_\alpha^i - \left(\sigma_\alpha^i\right)^2 A_i^2, \tag{3.47}$$

with terminal condition $A_i(T) = 0$.

Proof. This result is a simple consequence of Proposition 9. □

Under more stringent assumptions, it is also possible to derive explicitly the optimal factor exposure, $\pi_\cdot^{F,*}$, as shown in the following proposition.

Proposition 11. *Suppose that* $\sigma_\mu = \tau\sigma_F$*,* $\kappa_\mu = k\, Id$*, and* $\hat{\Sigma}_0^\mu = \tau_0\Sigma$*, where* τ_0, τ *and* k *are scalars. Let* $\left(\hat{\mu}_\cdot, \hat{\Sigma}_\cdot^\mu\right)$ *be the solution to the system of equations* (3.42)*. Then,*

$$\hat{\Sigma}_t^\mu = \tau_t\Sigma_F, \;\; A_F(t) = a_F(t)\Sigma_F^{-1} \text{ and } B_F(t) = b_F(t)\Sigma_F^{-1}\bar{\mu},$$

where

$$\dot{\tau_t} = -2k\tau_t - \left(\tau\rho + \tau_t\right)^2 + \tau^2 \tag{3.48}$$

and

$$\begin{cases} \dot{a_F} = -\frac{1}{2}\frac{1-\gamma}{\gamma}\left(1+2\tau\rho a_F\right)^2 + 2ka_F - 2\tau^2 a_F^2 \\ \dot{b_F} = -2a_F b_F\left(\tau^2 + \frac{1-\gamma}{\gamma}\rho^2\tau^2\right) - \frac{1-\gamma}{\gamma}\rho\tau b_F + kb_F - 2ka_F. \end{cases} \tag{3.49}$$

Therefore, the optimal exposure to risk factors is given by

$$\pi_t^{F,*} = \frac{1}{\gamma(1-2a_F(t)\tau_t)}\Sigma_F^{-1}\left(\left(1+2\rho\tau a_F(t)\right)\hat{\mu}_t + \left(\rho\tau + \tau_t\right)b_F(t)\bar{\mu}\right).$$

Proof. This result is essentially a restatement of Proposition 9, in the context of the multifactor model of this section. □

It is possible to solve explicitly equation (3.47) as well as the system of equations (3.49) — see Appendix 3.A. Therefore, under the assumptions of Proposition 11, the problem of optimal asset allocation with a power utility for the multifactor model (3.36) can be explicitly solved. We stress that, even if the conditions in Proposition 11 are not met, we can nevertheless compute explicitly the optimal exposure to the idiosyncratic risks.

Empirical work

We end this section with an application of our model to data on the American equities market. The objective of this simple exercise is to illustrate how the estimator can be applied and how it performs relatively to a baseline. As baselines we take the APT prediction itself (no filtering) and the simple Kalman filter estimator (no shrinkage towards the APT). The data consist of daily total return time–series for 246 constituents of the S&P500 index[5], from 1990 until 2014.

The factor portfolios are the Fama–French factors: $R_m - R_f$ (excess return on the market), SMB (Small Minus Big) and HML (High Minus Low). For more details on the Fama–French factors we refer to Fama and French (1993).

Recall the definition of abnormal returns, given in (3.41). Using the Fama–French factors, the cumulative excess return of stock i at time k is given by

$$R_{ik} = \beta_{i1}[R_m - R_f]_k + \beta_{i2}SMB_k + \beta_{i3}HML_k + R^a_{ik},$$

where R^a_{ik} denotes the cumulative abnormal return of stock i at time k. Therefore, we identify the abnormal returns with the residuals of the regression of excess returns on the Fama–French factors. In Figure 3.2, we compare the correlation matrix of the excess returns and that of the abnormal returns. We observe that the correlation coefficients of abnormal returns are much closer to zero than those of excess returns — we recall that the model (3.36) assumes that the abnormal returns are independent.

---

[5] The only criterion for the choice of stocks is the availability of data since 1990.

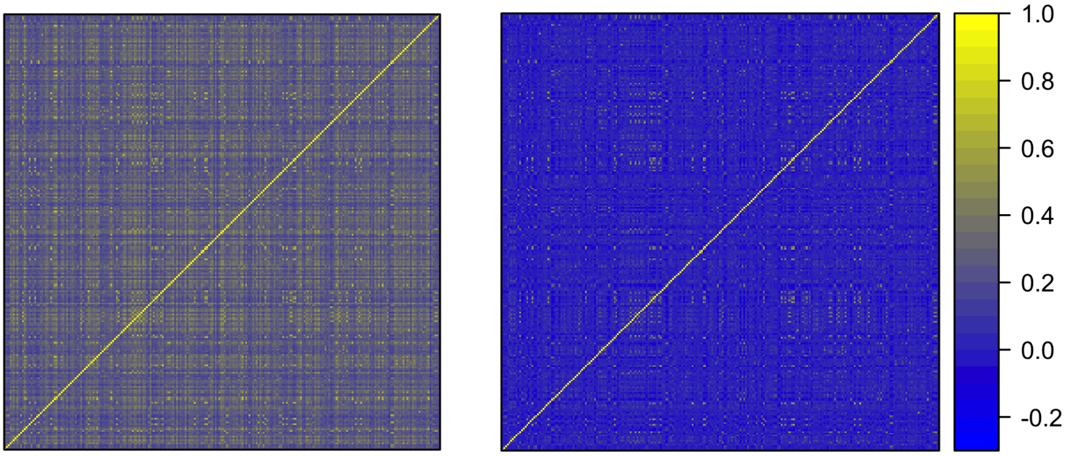

Figure 3.2: On the left, we plot the correlation matrix of the excess returns for the 246 stocks during the period $1990 - 2014$. On the right, we plot the correlation matrix of the abnormal returns for the same stocks during the same period. The abnormal returns are the residuals of the regression of the excess returns on the Fama–French factors.

In the sequel, we focus on the problem of estimating alpha from the process of cumulative abnormal returns. We simplify the model (3.36), and consider instead

$$\begin{cases} dR_t^{a,i} = \alpha_t^i dt + \sigma_i dW_t^{a,i} \\ d\alpha_t^i = -k\alpha_t^i dt + \tau\sigma_i dW_t^{\alpha,i}, \end{cases} \tag{3.50}$$

where the parameters k and τ are the same for all the stocks. The initial condition for $\alpha_\cdot$ in (3.50) is

$$\alpha_0^i \sim \mathcal{N}(0, \tau_0\sigma_i^2), \tag{3.51}$$

where τ_0 reflects the initial shrinkage. Hereafter, unless otherwise stated, we let $\tau_0 = 10^{-1}$.

The parameters k and τ determine the strength of continuous time shrinkage. In the sequel, we consider different (arbitrary) combinations of values for these parameters representing different levels of shrinkage. We use the following two benchmarks as reference:

- take the estimate of alpha to be zero, $\hat{\alpha}_\cdot^i = 0$ — this coincides with the APT prediction and corresponds to the case in which $k \to \infty$ and $\tau \to 0$;

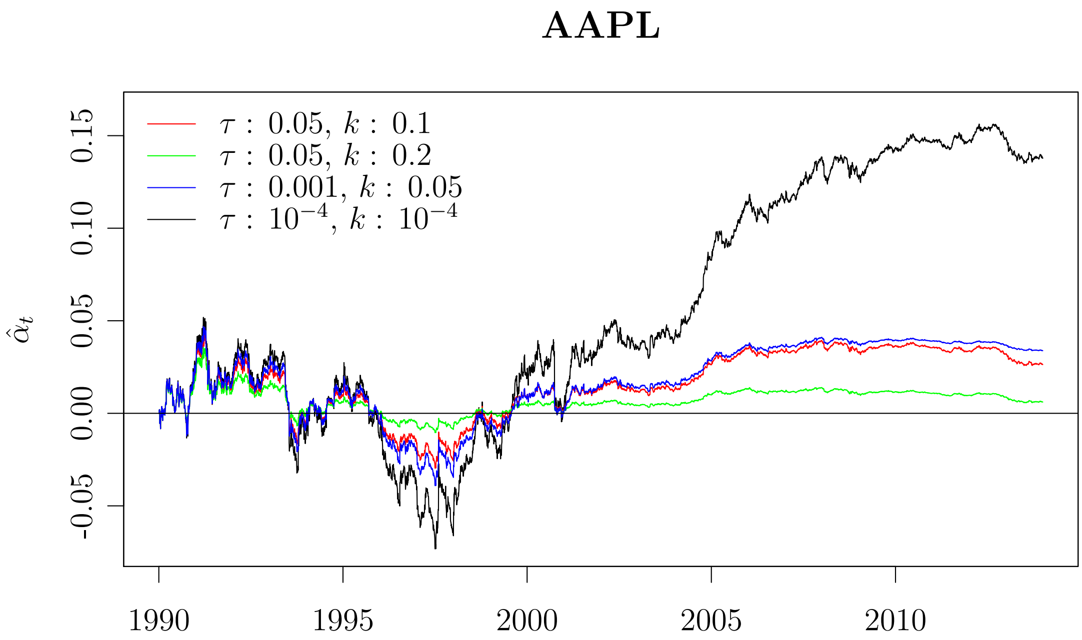

Figure 3.3: Plot of the estimate of alpha, $\hat{\alpha}_t$, for the Apple stock for different values of k, τ.

- no continuous time shrinkage, that is[6] $k = \tau = 0$ — in this case, the estimator coincides with the standard Kalman filter.

In Figures 3.3 and 3.4, we plot the estimate of alpha, $\hat{\alpha}_\cdot$, for two different stocks. It can be seen that the effects of continuous time shrinkage are two–fold: the estimator $\hat{\alpha}_t$ remains closer to zero, as expected, and it is less volatile than its non–shrinked counterpart.

We compare the performance of the estimators for different values of k and τ in a prediction exercise. Specifically, we use the estimate of alpha at time t, $\hat{\alpha}_t$, to predict the cumulative excess abnormal return at time $t + \Delta$,

$$\hat{R}^a_{t+\Delta} = R^a_t + \hat{\alpha}_t \Delta.$$

The prediction horizon is set at 6 months, that is $\Delta = 0.5$. We compare the prediction error of the different estimators at times $t_i = 0.5\, i$. All the results are reported relative to the benchmark of predicting zero ($\hat{\alpha}_\cdot = 0$). We denote the relative performance of an estimator versus this benchmark by Gain and define it

---

[6] In fact, we consider $k = \tau = 10^{-4}$. It can be seen that there is no substantial difference between this choice and $k = \tau = 10^{-5}$, for example.

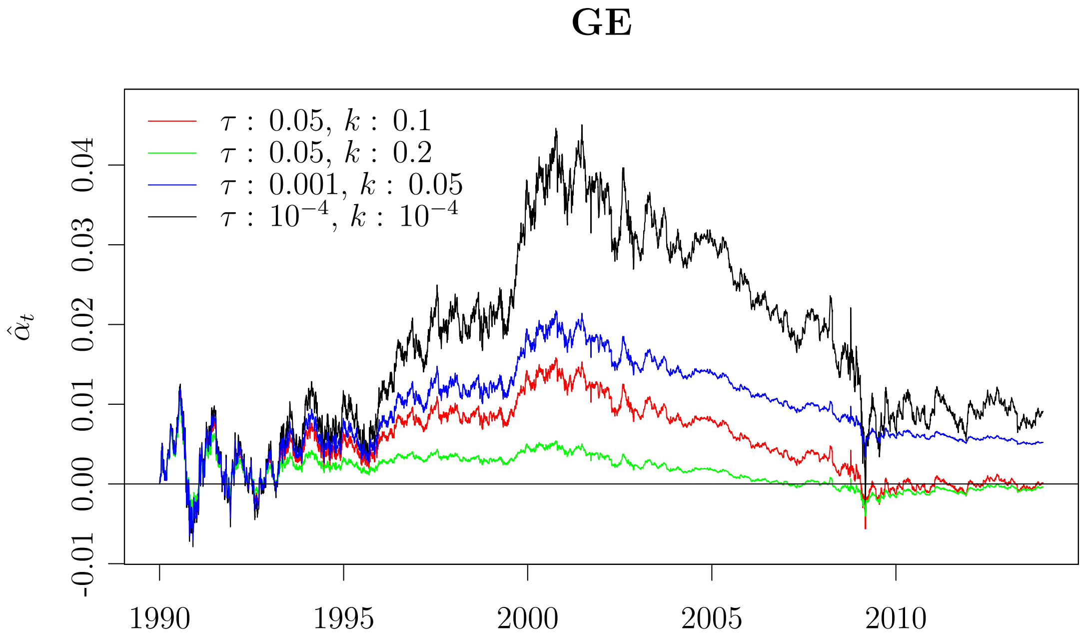

Figure 3.4: Plot of the estimate of alpha, $\hat{\alpha}_t$, for the General Electric stock for different values of k, τ.

simply as

$$\text{Gain} = \frac{\text{MSPE}^{benchmark} - \text{MSPE}}{\text{MSPE}^{benchmark}},$$

where MSPE and $\text{MSPE}^{benchmark}$ denote the mean squared prediction errors of the estimator and of the benchmark, respectively. The mean can be computed across time, cross–sectionally or across time and section; we perform the three analyses.

In Figure 3.5, we plot the Gain of different estimators across time — the mean squared prediction errors are computed cross–sectionally. We can see that, if no shrinkage is performed, the Gain of the resulting estimator swings widely, sometimes underperforming by as much as 15%.

In Figure 3.6, we plot the Gain across the different stocks — the mean squared prediction errors are computed across time. The stocks are ordered according to their signal–to–noise ratio (SNR), i.e. their Sharpe ratio. As expected, in stocks with low SNR, none of the estimators achieves a positive Gain. In fact, Gain seems to be a linear function of the square of the SNR — varying the amount of shrinkage changes mostly the slope of this linear relationship.

In Table 3.1, we compute the total Gain — the mean squared prediction errors are computed across time and cross–sectionally — for different values of τ, k and

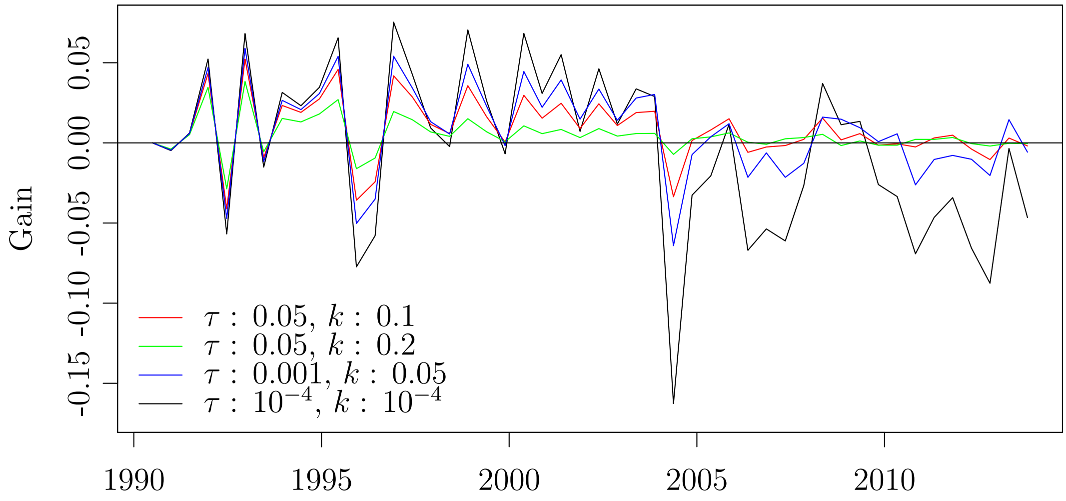

Figure 3.5: Plot of the Gain of the different estimators across time — the mean squared prediction errors are computed cross–sectionally.

τ_0	$\tau : 0.05,\ k : 0.1$	$\tau : 0.05,\ k : 0.2$	$\tau : 10^{-3},\ k : 0.05$	$\tau : 10^{-4},\ k : 10^{-4}$
1.00	0.35	0.40	-0.37	-2.64
0.50	0.94	0.69	0.54	-1.35
0.20	1.10	0.63	1.12	-0.02
0.10	0.90	0.46	1.09	0.59
0.05	0.65	0.31	0.82	0.79

Table 3.1: Gain (in %) of the different estimators relative to the benchmark, for different values of τ, k and τ_0. The mean squared prediction errors are computed across time and cross–sectionally.

also τ_0. As expected, as τ_0 increases (less initial shrinkage), the performance of the non–shrinked estimator deteriorates.

We recall that the optimal allocation on the i^{th} idiosyncratic risk for a power utility is given by

$$\pi_t^i = \frac{1}{1 - 2A_i(t)\hat{\Sigma}_t^{\alpha,i}} \frac{\hat{\alpha}_t^i}{\gamma\sigma_i^2}, \tag{3.52}$$

where γ is the risk aversion and the functions $A_i(\cdot)$ and $\hat{\Sigma}_{\cdot}^{\alpha,i}$ solve different Riccati equations. In a model with full information on α, the optimal portfolio for the

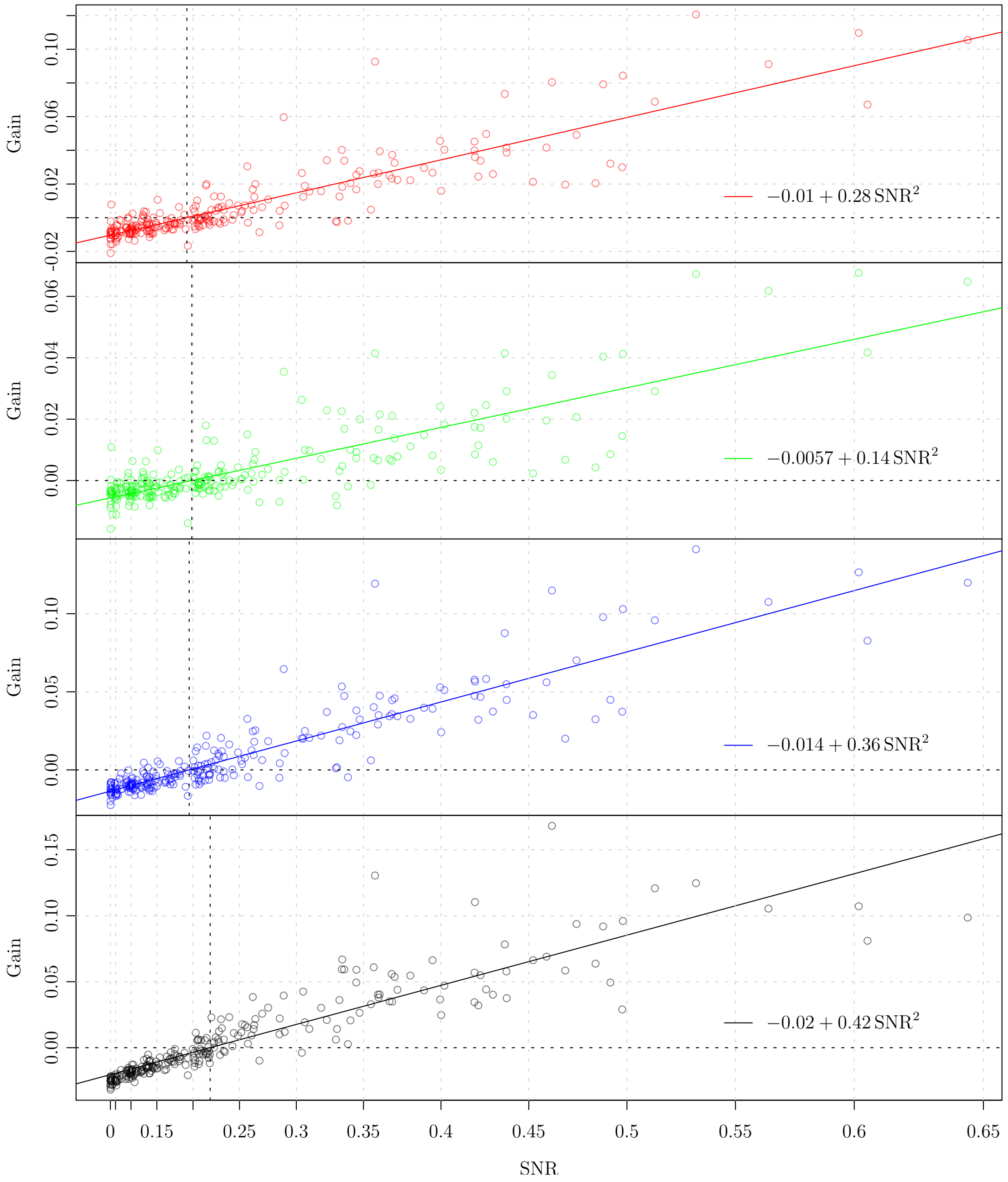

Figure 3.6: Plots of the cross-sectional Gain of the different estimators — the mean squared prediction errors are computed across time. From top to bottom, the parameters of the estimators used to produce the plots are, as before, ($\tau = 0.05, k = 0.1$), ($\tau = 0.05, k = 0.2$), ($\tau = 0.001, k = 0.05$) and ($\tau = 10^{-4}, k = 10^{-4}$).

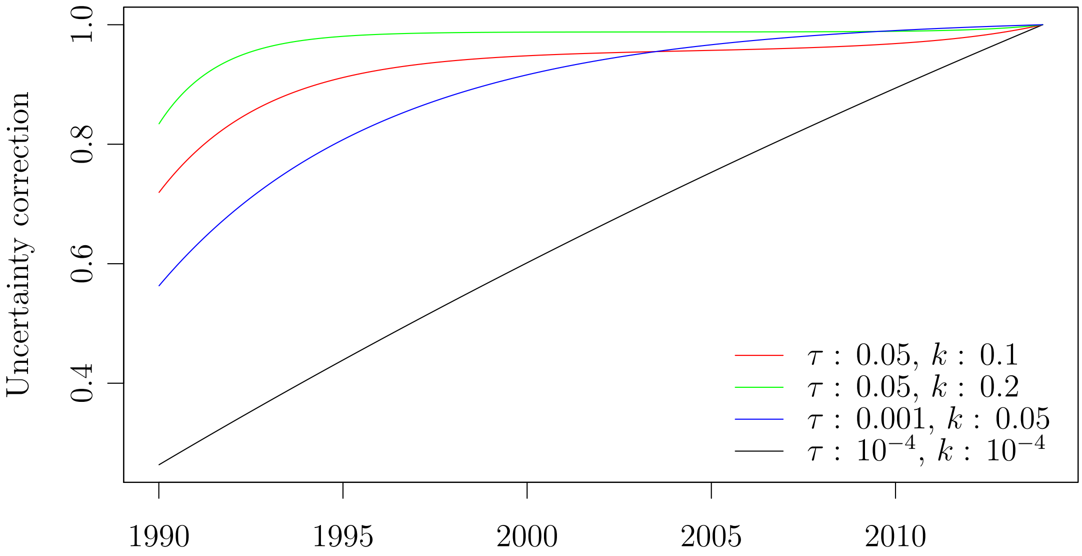

Figure 3.7: Plot of the uncertainty correction for different values of τ and k. The risk aversion is $\gamma = 5$

same utility would be $\frac{\alpha^i}{\gamma\sigma^2}$. Therefore, the term

$$\frac{1}{1 - 2A_i(t)\hat{\Sigma}_t^{\alpha,i}}$$

provides an uncertainty correction to the optimal portfolio — it scales allocations down to account for the uncertainty on α. Such a correction depends on the parameters τ and k, as well as on the risk aversion parameter γ. In Figure 3.7, we plot this uncertainty correction for the different values of τ and k considered above, and for a risk aversion $\gamma = 5$. We observe that, as expected, when the shrinkage is strongest, the uncertainty correction is closest to 1.

The optimal allocation in (3.52) depends on the volatility of the abnormal returns, σ_i. This is assumed to be a constant in the model (3.50). However, in the data, one can observe that volatility is not constant. Therefore, to account for changing volatility in the allocation, we replace σ_i in (3.52) by a (non forward–looking) rolling estimate of the volatility[7], $\hat{\sigma}^i_\cdot$ — in this case, a simple exponential moving average of past squared returns. In Figures 3.8 and 3.9, we plot the rolling

---

[7] This is a standard practice in industry; it helps risk management by scaling down positions in times of higher volatility.

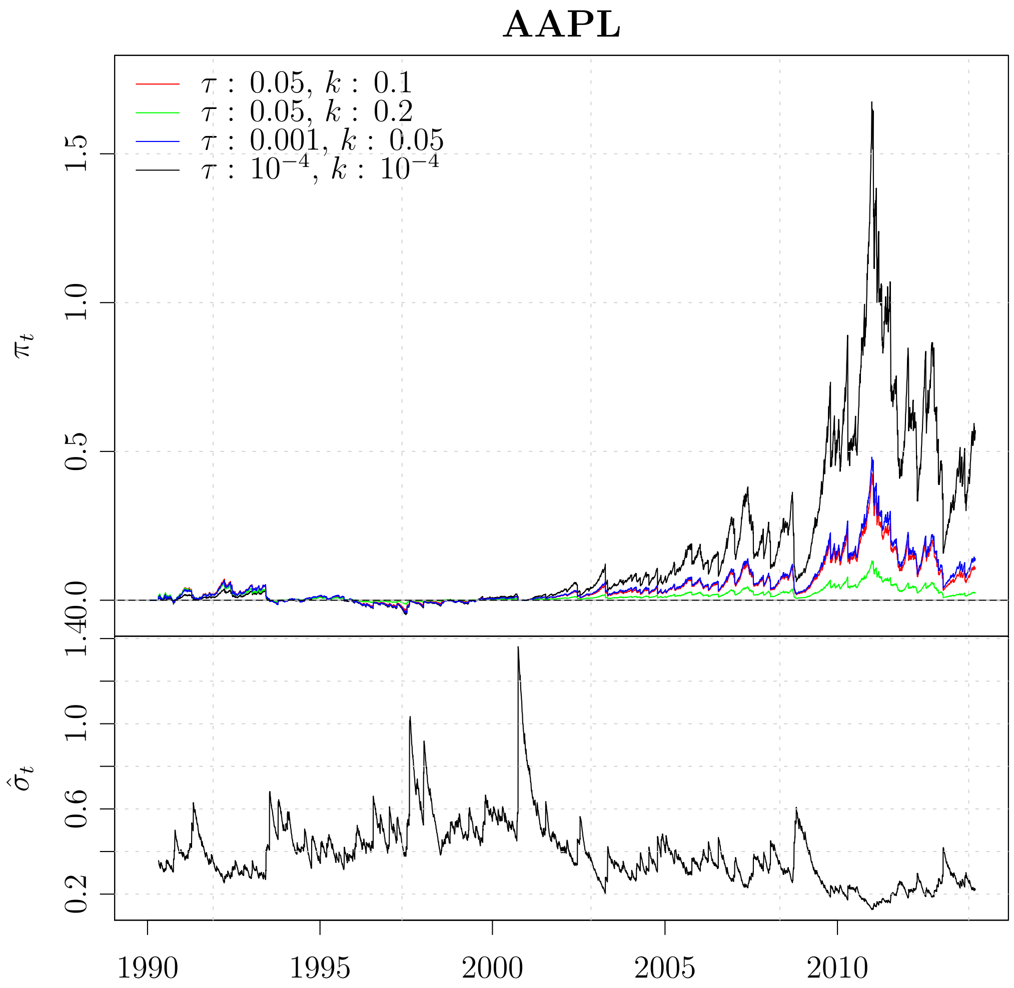

Figure 3.8: Optimal allocation and rolling estimate of volatility for the Apple stock. The rolling estimate of volatility uses a window of 80 days — the corresponding exponential smoothing weight is 2/81. The risk aversion is $\gamma = 5$.

estimate of volatility and the associated optimal allocation for two stocks when the risk aversion is $\gamma = 5$.

We compare the performance of the different allocations by comparing the respective wealth processes. More precisely, for each allocation π. we compute the associated wealth process, $X^{\pi}_{\cdot}$,

$$dX^{\pi}_t = X^{\pi}_t \pi'_t dR^a_t.$$

In Figure 3.10, we plot the wealth process associated with the optimal allocations for different parameters τ and k, and for a risk aversion $\gamma = 5$; note that the scale is logarithmic.

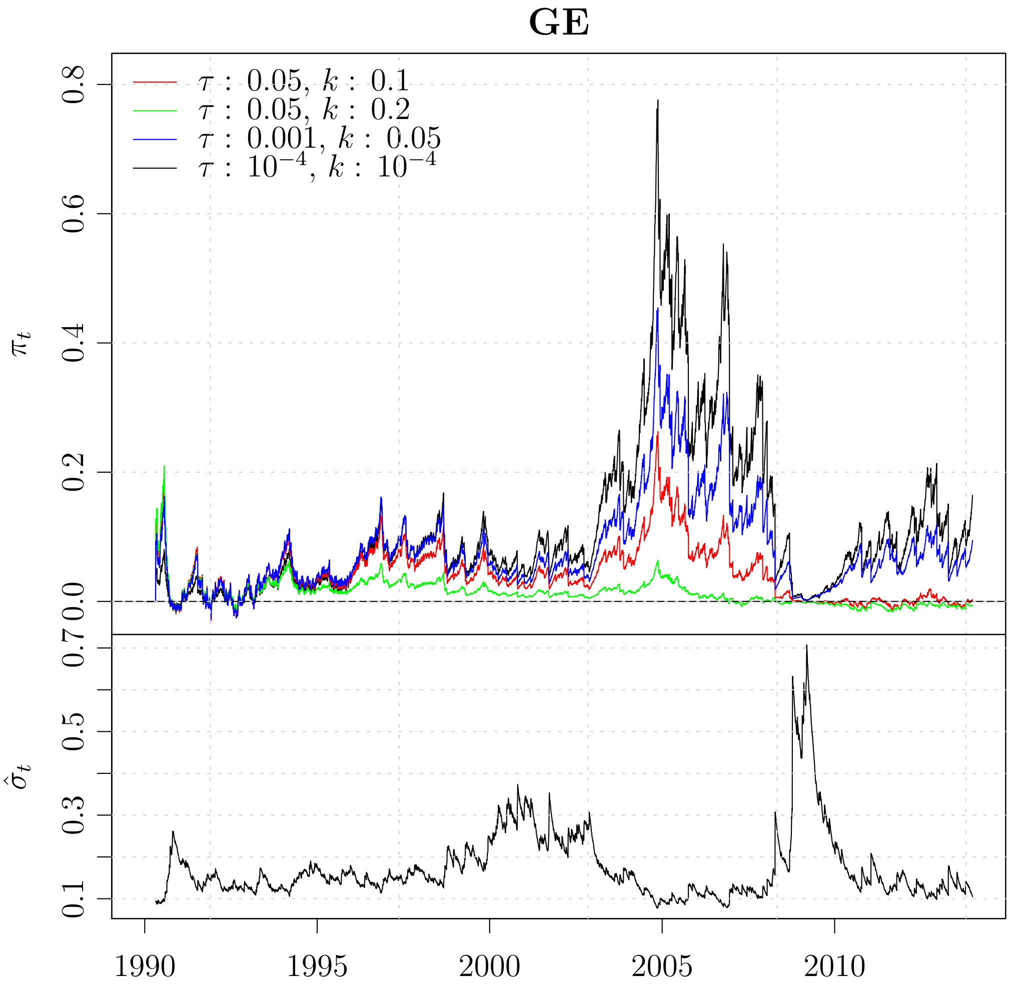

Figure 3.9: Optimal allocation and rolling estimate of volatility for the General Electric stock. The rolling estimate of volatility uses a window of 80 days — the corresponding exponential smoothing weight is 2/81. The risk aversion is $\gamma = 5$.

For different values of γ, the results are summarised in Table 3.2. There, we compare the Sharpe ratios of the wealth processes associated with the optimal allocations for different values of τ, k and γ. Sharpe ratios are computed using daily log returns. From the table, one can see that the strategy using the non–shrinked estimator always underperforms. The difference in performance is more obvious for low risk aversions, as expected. Nevertheless, even for a risk aversion as high as 20, performing shrinkage improves the Sharpe ratio by as much as 40%.

We end this section with a note on how to choose the mean–reversion parameters τ and k. This is an important issue which we overlook in the previous sections. It is tantamount to the choice of prior in any Bayesian method, or to the amount of penalisation in any shrinkage method.

One possibility is to infer τ and k from the data, via maximum likelihood estimation, for example. In theory, this is possible and relatively straightforward (from a computational point of view), since $R^a_\cdot$ is a Gaussian process with a covariance function which can be computed explicitly. In practice, however, this is an unrealistic approach. Indeed, we recall that it is notoriously difficult to estimate a *constant* drift, see Merton (1980) and Black (1993) — in the case of a mean–reverting drift, the problem is obviously harder.

The other possibility is to choose the parameters τ and k that maximise the out–of–sample performance, or minimise the out–of–sample prediction error. For example, for several values of τ and k on a grid, one can build a table such as Table 3.2 and choose the parameters that maximise the Sharpe ratio (or any other measure of performance). This seems like a more *ad–hoc* approach. However, we observe that it coincides with how penalisation parameters are typically chosen in popular penalised regression methods such as Ridge regression or LASSO. In the statistics community, this approach is known as cross–validation.

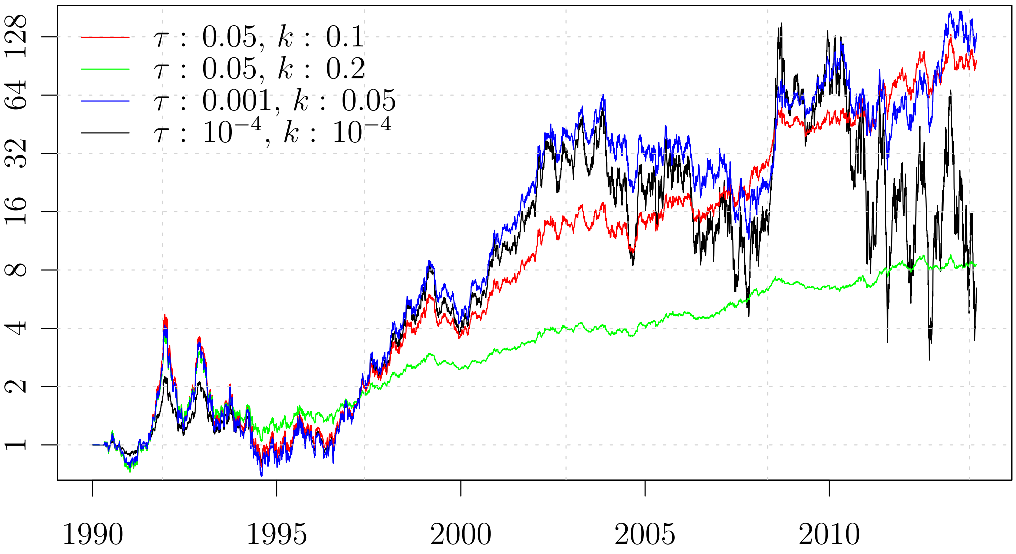

Figure 3.10: Wealth process associated with the optimal allocations for different parameters τ and k. The risk aversion parameter is $\gamma = 5$.

γ	τ : 0.05, k : 0.1	τ : 0.05, k : 0.2	τ : 0.001, k : 0.05	τ : 10^{-4}, k : 10^{-4}
2	0.19	0.29	-0.08	$-\infty$
5	0.50	0.45	0.36	0.07
10	0.59	0.50	0.50	0.33
20	0.64	0.52	0.57	0.45

Table 3.2: Sharpe ratios of the optimal allocations for different values of τ, k and γ. The Sharpe ratios are computed using daily log returns and annualised.

3.A The scalar Riccati equation

Suppose that $f(\cdot)$ solves the Riccati equation

$$\dot{f}(t) = -af(t)^2 - 2bf(t) + c,$$

with $f(t_0) = f_0$ and $ac \geq 0$. We define

$$\phi^{\pm}(t) = \frac{af_0 + b \pm \alpha}{2\alpha} e^{\pm\alpha(t-t_0)},$$

where $\alpha = \sqrt{b^2 + ac}$. Then,

$$f(t) = -\frac{b}{a} + \frac{\alpha}{a} \frac{\phi^+(t) + \phi^-(t)}{\phi^+(t) - \phi^-(t)}.$$

Now, suppose that $f(\cdot)$ and $g(\cdot)$ solve the system of equations

$$\begin{cases} \dot{f}(t) = -af(t)^2 - 2bf(t) + c, \\ \dot{g}(t) = -af(t)g(t) - bg(t) + df(t), \end{cases}$$

with $f(t_0) = f_0$ and $g(t_0) = g_0$. Then, $f(\cdot)$ is given as above and

$$g(t) = -\frac{bd}{\alpha^2} f(t) + \left(g_0 + f_0 \frac{bd}{\alpha^2} - \frac{dc}{\alpha^2} \right) \frac{1}{\phi^+(t) - \phi^-(t)} + \frac{dc}{\alpha^2}.$$

Bibliography

S. Basak and G. Chabakauri. Dynamic mean–variance asset allocation. *Review of Financial Studies*, 23(8):2970–3016, 2010.

S. L. Beach and A. G. Orlov. An application of the Black–Litterman model with EGARCH–M–derived views for international portfolio management. *Financial Markets and Portfolio Management*, 21(2):147–166, 2007.

R. Bellman. *Dynamic programming.* Princeton University Press, Princeton, N. J., 1957.

F. Berrier, L. C. G. Rogers, and M. Tehranchi. A characterization of forward utility functions. 2009.

P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 1968.

T. Björk and A. Murgoci. A General Theory of Markovian Time Inconsistent Stochastic Control Problems. *SSRN eLibrary*, 2010.

T. Björk, M. Davis, and C. Landén. Optimal investment under partial information. *Math. Methods Oper. Res.*, 71(2):371–399, 2010.

T. Björk, A. Murgoci, and X. Y. Zhou. Mean–variance portfolio optimization with state–dependent risk aversion. *Mathematical Finance*, 24(1):1–24, 2014.

F. Black. Estimating expected return. *Financial Analysts Journal*, pages 36–38, 1993.

F. Black. Investment and consumption through time. Financial Note 6B, 1968.

F. Black and R. Litterman. Asset allocation: combining investor views with market equilibrium. *The Journal of Fixed Income*, 1(2):7–18, 1991.

F. Black and R. Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48(5):pp. 28–43, 1992.

B. Bouchard and N. Touzi. Weak Dynamic Programming Principle for Viscosity Solutions. *SIAM Journal on Control and Optimization*, 49 (3):948–962, 2009.

S. Brendle. Portfolio selection under incomplete information. *Stochastic Process. Appl.*, 116(5):701–723, 2006.

S. Browne and W. Whitt. Portfolio choice and the Bayesian Kelly criterion. *Adv. in Appl. Probab.*, 28(4):1145–1176, 1996.

A. Danilova, M. Monoyios, and A. Ng. Optimal investment with inside information and parameter uncertainty. *Math. Financ. Econ.*, 3(1):13–38, 2010.

M. Davis and S. Lleo. Black–Litterman in continuous time: the case for filtering. *Quantitative Finance Letters*, 1(1):30–35, 2013.

N. El Karoui and M. Mrad. An exact connection between two solvable SDEs and a nonlinear utility stochastic PDE. *SIAM J. Financial Math.*, 4(1):697–736, 2013.

N. El Karoui and M. Mrad. Stochastic utilities with a given optimal portfolio: approach by stochastic flows. *arXiv preprint arXiv:1004.5192*, 2010.

N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

R. Frey, A. Gabih, and R. Wunderlich. Portfolio optimization under partial information with expert opinions. *Int. J. Theor. Appl. Finance*, 15(1), 2012.

A. Friedman. *Partial differential equations of parabolic type.* Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

I. Gyöngy. A Note on Euler's Approximations. *Potential Analysis*, 8:205–216, 1998.

I. Gyöngy and N. V. Krylov. Existence of strong solutions for Itô's stochastic equations via approximations. *Probab. Theory Related Fields*, 105(2):143–158, 1996.

G. He and R. Litterman. The intuition behind Black–Litterman model portfolios. *SSRN eLibrary*, 1999.

V. Henderson and D. Hobson. Horizon–unbiased utility functions. *Stochastic Process. Appl.*, 117(11):1621–1641, 2007.

S. Källblad, J. Obłój, and T. Zariphopoulou. Time–consistent investment under model uncertainty: the robust forward criteria. *arXiv preprint arXiv:1311.3529*, 2013.

I. Karatzas and X. Zhao. Bayesian adaptive portfolio optimization. In *Option pricing, interest rates and risk management*, Handb. Math. Finance, pages 632–669. Cambridge Univ. Press, Cambridge, 2001.

I. Karatzas, J. Lehoczky, and S. Shreve. Optimal portfolio and consumption decisions for a "small investor" on a finite horizon. *SIAM J. Control Optim.*, 25(6): 1557–1586, 1987.

T. S. Kim and E. Omberg. Dynamic nonmyopic portfolio behavior. *Review of Financial Studies*, 9(1):141–161, 1996.

F. Knight. *Risk, uncertainty and profit.* Houghton Mifflin, Boston, 1921.

Vladimír Kučera. A review of the matrix Riccati equation. *Kybernetika (Prague)*, 9:42–61, 1973. ISSN 0023-5954.

P. Lakner. Utility maximization with partial information. *Stochastic Process. Appl.*, 56(2):247–273, 1995.

P. Lakner. Optimal trading strategy for an investor: the case of partial information. *Stochastic Process. Appl.*, 76(1):77–97, 1998.

T. Leung, R. Sircar, and T. Zariphopoulou. Forward indifference valuation of American options. *Stochastics*, 84(5-6):741–770, 2012.

D. Li and W. Ng. Optimal dynamic portfolio selection: multiperiod mean–variance formulation. *Math. Finance*, 10(3):387–406, 2000.

A. E. B. Lim. Quadratic hedging and mean–variance portfolio selection with random parameters in an incomplete market. *Math. Oper. Res.*, 29(1):132–161, 2004.

A. E. B. Lim and X. Y. Zhou. Mean–variance portfolio selection with random parameters in a complete market. *Math. Oper. Res.*, 27(1):101–120, 2002.

R. S. Liptser and A. N. Shiryaev. *Statistics of random processes I*, volume 5 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, expanded edition, 2001. ISBN 3-540-63929-2. General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.

F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):pp. 77–91, 1952.

H. Markowitz. Portfolio selection: efficient diversification of assets. *Cowles Foundation Monograph*, 1959.

E. J. McShane. Extension of range of functions. *Bulletin of the American Mathematical Society*, 40(12):837–842, 1934.

R. C. Merton. Lifetime portfolio selection under uncertainty: The continuous–time case. *The Review of Economics and Statistics*, 51(3):pp. 247–257, 1969.

R. C. Merton. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics*, 8(4):323–361, 1980.

M. Mureşan. *A concrete approach to classical analysis*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2009. ISBN 978-0-387-78932-3.

M. Musiela and T. Zariphopoulou. Portfolio choice under dynamic investment performance criteria. *Quant. Finance*, 9(2):161–170, 2009.

M. Musiela and T. Zariphopoulou. Portfolio choice under space–time monotone performance criteria. *SIAM Journal on Financial Mathematics*, 1:326–365, 2010a.

M. Musiela and T. Zariphopoulou. Stochastic partial differential equations and portfolio choice. *Contemporary Quantitative Finance*, Carl Chiarella and Alexander Novikov (eds.):195–215, 2010b.

M. Musiela and T. Zariphopoulou. Optimal asset allocation under forward exponential performance criteria. In *Markov processes and related topics: a Festschrift for Thomas G. Kurtz*, volume 4 of *Inst. Math. Stat. Collect.*, pages 285–300. Inst. Math. Statist., Beachwood, OH, 2008.

S. Nadtochiy and M. Tehranchi. Optimal investment for all time horizons and martin boundary of space-time diffusions. 2012.

S. Nadtochiy and T. Zariphopoulou. A class of homothetic forward investment performance processes with non-zero volatility. In *Inspired by Finance*, pages 475–504. Springer, 2014.

Shi Ge Peng. Stochastic Hamilton–Jacobi–Bellman equations. *SIAM J. Control Optim.*, 30(2):284–304, 1992.

U. Rieder and N. Bäuerle. Portfolio optimization with unobservable Markov-modulated drift process. *J. Appl. Probab.*, 42(2):362–378, 2005.

H. Robbins and D. Siegmund. Statistical tests of power one and the integral representation of solutions of certain partial differential equations. *Bull. Inst. Math. Acad. Sinica*, 1(1):93–120, 1973.

S. Ross. The arbitrage theory of capital asset pricing. *Journal of economic theory*, 13(3):341–360, 1976.

P. A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3):pp. 239–246, 1969.

J. Sass and U. G. Haussmann. Optimizing the terminal wealth under partial information: the drift process as a continuous time Markov chain. *Finance Stoch.*, 8(4):553–577, 2004.

L. J. Savage. *The foundations of statistics.* John Wiley & Sons, Inc., New York; Chapman & Hill, Ltd., London, 1954.

A. Schied. Optimal investments for risk– and ambiguity–averse preferences: a duality approach. *Finance Stoch.*, 11(1):107–129, 2007.

A. Schied, H. Föllmer, and S. Weber. Robust preferences and robust portfolio choice. *Handbook of Numerical Analysis*, 15:29–87, 2009.

C. Skiadas. Dynamic porfolio choice and risk aversion. In J. R. Bridge and V. Linetsky, editors, *Financial Engineering*, volume 15, pages 789–843. Elsevier, 2008.

C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, pages 197–206, Berkeley and Los Angeles, 1956. University of California Press.

N. Touzi. Deterministic and stochastic control, application to finance. URL `http://www.cmap.polytechnique.fr/~touzi/Master-LN.pdf`. Lecture Notes, 2010.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, N. J., 2nd edition, 1947.

H. Whitney. Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89, 1934.

D. V. Widder. *The heat equation.* Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975. Pure and Applied Mathematics, Vol. 67.

J. Xiong and X. Y. Zhou. Mean–variance portfolio selection under partial information. *SIAM J. Control Optim.*, 46(1):156–175 (electronic), 2007.

X. Y. Zhou and D. Li. Continuous–time mean–variance portfolio selection: a stochastic LQ framework. *Appl. Math. Optim.*, 42(1):19–33, 2000.

G. Žitković. A dual characterization of self-generation and exponential forward performances. *Ann. Appl. Probab.*, 19(6):2176–2210, 2009.