Portfolio Optimization in Financial Markets with Partial Information

Sébastien Roland

This Version: September 23, 2007

Université d'Evry Val d'Essonne

THESE

pour obtenir

le grade de DOCTEUR de l'Université d'Evry Val d'Essonne
Discipline : Mathématiques

par

Sébastien ROLAND

Optimisation de Portefeuille sur les Marchés Financiers dans le cadre d'une Information Partielle

Soutenue le 7 Janvier 2008 devant la commission d'examen :

Mme Marie-Claire Quenez-Kammerer	Université Paris VII	Président
M Constantin Mellios	Université Paris I	Rapporteur
M Patrice Poncet	Université Paris I	Rapporteur
Mme Monique Jeanblanc	Université d'Evry	Directeur de thèse
M Vincent Lacoste	ESSEC	Membre du Jury

"Essayer encore, rater encore, rater mieux".
Samuel Beckett, *Cap au Pire*

Résumé

Cette thèse traite - en trois essais - de problèmes de choix de portefeuille en situation d'information partielle, thématique que nous présentons dans une courte introduction. Les essais développés abordent chacun une particularité de cette problématique. Le premier (co-écrit avec M. Jeanblanc et V. Lacoste) traite la question du choix de la stratégie optimale pour un problème de maximisation d'utilité terminale lorsque l'évolution des prix est modélisée par un processus de *Itô-Lévy* dont la tendance et l'intensité des sauts ne sont pas observées. L'approche consiste à réécrire le problème initial comme un problème *réduit* dans la filtration engendrée par les prix. Cela nécessite la dérivation des équations de filtrage non-linéaire, que nous développons pour un processus de Lévy. Le problème est ensuite résolu en utilisant la programmation dynamique par les équations de Bellman et de HJB. Le second essai aborde dans un cadre gaussien la question du coût de l'incertitude, que nous définissons comme la différence entre les stratégies optimales (ou les richesses maximales) d'un agent parfaitement informé et d'un agent partiellement informé. Les propriétés de ce coût de l'information sont étudiées dans le cadre des trois formes standard de fonctions d'utilités et des exemples numériques sont présentés. Enfin, le troisième essai traite la question du choix de portefeuille quand l'information sur les prix de marché n'est disponible qu'à des dates discrètes et aléatoires. Cela revient à supposer que la dynamique des prix suit un processus marqué. Dans ce cadre, nous développons les équations de filtrage et réécrivons le problème initial dans sa forme *réduite* dans la filtration discrète des prix. Les stratégies optimales sont ensuite calculées en utilisant le calcul de Malliavin pour des mesures aléatoires et une extension de la formule de Clark-Ocone-Haussman est à cette fin présentée.

Une second partie de la thèse, indépdendante de la première, présente une méthode de résolution numérique de problèmes d'optimisation stochastique. Dans un essai (co-écrit avec B. Amzal et Y. Ebguy), un problème de calibration jointe d'un modèle d'évalution de type *Itô-Lévy* à des prix d'actions et d'options écrites dessus est discutée et résolue grâce à l'algorithme développée.

Mots Clefs: optimisation de portfeuille, information partielle, processus de Lévy, processus marqué, marché incomplet, filtrage non-linéaire, observations discrètes, programmation dynamique, calcul de Malliavin.

Contents

Part I

Portfolio Optimization

1

Introduction

Agenda

In this introduction, we present the main concepts used in decision theory and study, from the point of view of a decision maker, the distinction between risk and uncertainty. This leads to the setup of decision making under uncertainty in which the robust Savage representation of preferences will be of paramount importance. Finally, we precise the links between learning and uncertainty and detail the setup of portfolio optimization under uncertainty.

1.1. Distinction Between Risk and Uncertainty

In the standard framework of decision theory, the starting point is a specification of a (stochastic) model, a set of future scenarios $(\Omega, \mathcal{F})$ and a probability measure P on these outcomes. However, in many circumstances, the decision maker is not able to attribute a precise probability to future outcomes. This situation has been called *uncertainty*, cf. Keynes (1921) [102], which differs from *risk* when the choice of the probability measure is unique.

Since Knight (1921) [106], it is generally admitted that there exist two kinds of uncertainty. Suppose an investor which may decide to invest in a financial position from the set $\mathcal{X}$. The first uncertainty involves an increase in the dispersion (in terms of the moments) of the state variables distribution (which is unique) that the decision maker thinks he/she faces. Additional information is then required to precise the parameters of the probability distribution. Such a process of information acquisition may take time. Instead, the second uncertainty entails a decrease in his/her confidence about the possibly many state variables model distributions. The process by which this uncertainty is reduced resumes to the choice of one model distribution.

In the lines of Knight (1921) [106], the former type of *uncertainty*, which can be reduced to a single distribution with known parameters, is *risk*, while the latter type of *uncertainty* is true *uncertainty*. While risk and uncertainty are clearly distinct concepts, they have not been treated separately in decision making in an explicit way, at least until recently. This may be due to the celebrated Savage theory of representation of preferences.

Remark 1.1.1 (Uncertainty and Incompleteness) *We note that incompleteness and uncertainty differs. The former concentrates on the precise specification of the historical probability distribution, while the latter concerns the choice of one risk-neutral measure from one historical measure. Therefore, uncertainty and incompleteness are different problems and an increase in uncertainty does not entail an increase in incompleteness.*

1.1.1 Savage Representation

Consider an investor whose attitude in face of uncertainty is described by a preference relation $\succ$ on the space $\mathcal{X}$ of financial positions. It is natural to assume $\succ$ is monotone in the sense that:

$$Y \succeq X \text{ if } Y(\omega) \geqslant X(\omega), \ \forall \omega \epsilon \Omega$$

where $\succeq$ is the weak preference order induced by $\succ$. Under a suitable condition of continuity, one can prove the existence of a numerical representation for the preference relation $\succ$, i.e.: there exists a function $u : \mathcal{X} \rightarrow \mathbb{R}$ s.t.:

$$Y \succ X \text{ i.f.f. } u(Y) > u(X)$$

Savage (1954) [150] introduced a set of additional axioms which guarantee that, when $\mathcal{X}$ represents the set of real random variables, there is a numerical representation of the form:

$$u(X) = E^P [U(X)] = \int_{\omega \epsilon \Omega} U(X(\omega)) \, dP(\omega) \tag{1.1}$$

where U is a continuous function on $\mathbb{R}$. The measure P is the subjective view of the probabilities of events which is implicit in the preference relation $\succ$. When U is an increasing concave function, U is termed a utility function.

Under the Savage representation paradigm, the decision maker's *preferences* are represented by the expectation of some utility function which is computed by means of a single probability measure. The uncertainty that the decision maker faces is thus reduced to risk with some probability measure.

Remark 1.1.2 (Measure of Risk) *The concept of a measure of risk, as introduced by Artzner et al. (1999) [8], bears ressemblance with the representation (1.1) when $u \equiv -l$ where l is a loss function and X is some financial position. In the sense of [8], $E^P[l(X)]$ can then be seen as an extra capital requirement to support the market risk X.*

1.1.2 Ellsberg's Paradox

In Ellsberg (1961) [58], a distinction between aversion to risk and aversion to uncertainty is established. The resulting celebrated Ellsberg's paradox is a paradox in decision theory in which people's choices violate the Savage axioms of expected utility theory. It is generally taken to be evidence for aversion against uncertainty.

Problem 1.1.3 (Ellsberg's Problem) *Suppose that you have an urn containing* 30 *red balls and* 60 *other balls, either black or yellow. You don't know how many black or yellow balls there are, but the total number of black balls plus the total number of yellow balls equals* 60*. The balls are well mixed so that each individual ball is as likely to be drawn as any other. You are now given a choice between two gambles:*

Game	*Gamble A*	*Gamble B*
Ball Pay	*Red:* $100	*Black:* $100

Also you are given the choice between these two gambles:

Game	*Gamble C*	*Gamble D*
Ball Pay	*Red/Yellow:* $100	*Black/Yellow:* $100

Since the profits are exactly the same, it follows that you will prefer Gamble A to Gamble B if and only if you believe that drawing a red ball is more likely than drawing a black ball (according to expected utility theory). Similarly it follows that you will prefer Gamble C to Gamble D if and only if you believe that drawing a red or yellow ball is more likely than drawing a black or yellow ball. If drawing a red ball is more likely than drawing a black ball, then drawing a red or yellow ball is also more likely than drawing a black or yellow ball.

Problem 1.1.4 (Ellsberg's Paradox) *Supposing you prefer Gamble A to Gamble B, it follows that you will also prefer Gamble C to Gamble D. And,* *supposing instead that you prefer Gamble D to Gamble C, it follows that you will also prefer Gamble B to Gamble A. When surveyed, however, most people strongly prefer Gamble A to Gamble B and Gamble D to Gamble C. Therefore, some assumption of expected utility theory is violated.*

Proof. Let R, Y and B denote the estimated probabilities of each color ball s.t. $R = 1/3$. If you prefer Gamble A to Gamble B, by utility theory, it is presumed this preference represents your estimate of expected utility. This is represented in the following inequality:

$$\frac{1}{3}U\left(\$100\right) > B \times U\left(\$100\right)$$

Solving for B gives:

$$\frac{1}{3} > B$$

If you also prefer Gamble D to Gamble C, the following inequality is similarly obtained:

$$B \times U\left(\$100\right) + Y \times U\left(\$100\right) > \frac{1}{3}U\left(\$100\right) + Y \times U\left(\$100\right)$$

Solving for B gives:

$$B > \frac{1}{3}$$

Hence, a contradiction. ■

Ellsberg (1961) [58] experiments have established in a convincing and robust way that decision makers generally prefer to act in settings in which they have better information. In the classic two-urn experiments, agents tend to choose to bet on the color of a ball drawn from an urn whose composition is known rather than on the color of a ball drawn from an urn that contains black and yellow balls in unknown proportion. These experiments have led to numerous models of decision under uncertainty, such as the Choquet expected utility model, cf. Schmeidler (1989) [154], or the multiple prior model, cf. Gilboa and Schmeidler (1989) [77], capturing the fact that agents might have aversion towards uncertainty. However, these models, while referring to the Ellsberg's experiments neglect an important aspect of these experiments: the information available to the decision maker is not part of the modelling.

1.1.3 Robust Savage Representation

As proved by the Ellsberg's paradox, the paradigm of expected utility, formalized in the representation (1.1), has a limited scope. Gilboa (1987) [76], Gilboa and Schmeidler (1989) [77] and Schmeidler (1989) [154] weaken Savage's axioms to settle debates caused by Ellsberg's paradox. They axiomatize the preference which is represented by the minimum among the expected utilities each of which is computed by an element of some set of probability measures. This preference is called the max-min expected utility or the Choquet expected utility. This is a natural extension of preference under uncertainty to the case in which the information is too imprecise to be summarized by a single probability measure. This type of uncertainty is also called the Knightian uncertainty.

Definition 1.1.5 (Robust Savage Representation) *A numerical representation U of the preference order $\succ$ on $\mathcal{X}$ will be called a robust Savage representation if it is of the form:*

$$u(X) = \inf_{P \epsilon \mathcal{P}} E^P [U(X)] \tag{1.2}$$

where $\mathcal{P}$ is a set of probability measures on $(\Omega, \mathcal{F})$ and U a utility function.

Remark 1.1.6 (Robust Measure of Risk) *Following Foellemer and Schied (2002) [68], we note that the concept of a robust measure of risk is in line with the representation (1.2) when $u \equiv -l$ where l is a loss function and X is some financial position. Therefore $\sup_{P \epsilon \mathcal{P}} E^P [l(X)]$ is the maximal, over all models, capital requirement for holding the position X.*

1.2. Discussion on Partial Information

In this section, we present some mathematical and financial material which are pivotal to understand the partial information setup. In particular, we will explore how and why partial information departs from the classical and well-known complete information setup.

1.2.1 Classical Setup

The classical financial economical models - Markowitz (1952) [123], Merton (1969, 1971, 1973) [125], [126], [127], Cox-Ingersoll-Ross (1985) [33] - usually specify equilibrium quantities of interest: asset prices, interest rates and portfolio rules, in terms of the moments of distribution of returns. The moments, however, are unobservable to both real-world investors and empiricists, cf. Detemple (1986) [48], Dothan and Feldman (1986) [53] or Gennotte (1986) [74]. Implementing and testing the so-called complete information models thus requires the estimation of these moments. From a statistical perspective, this is called the estimation procedure. The statistical production of these moment estimates outside of the theoretical model raises various questions and problems, cf. Liptser and Shiryaev (2001) [118] and Frydman and Lakner (2003)

[72]. One central issue is the consistency of the econometric assumptions of the estimation procedure with the structure and hypothesis of the original model.

In financial economics, the issue of parameter estimation first arose from an empirical and statistical perspective in the context of portfolio problems, cf. Merton (1971) [126]. The portfolio problem under unknown mean and variance was usually called the parameter uncertainty problem, cf. Detemple (1986) [48], referring to the real-world situation where moments of asset returns distributions are not observable. To better suit this situation, one solution is to take into account the uncertainty in the determination of these quantities by adapting the mathematical and statistical procedures. This is referred to as the filtering problem, when one needs to compute estimators of unknown quantities based on his/her available information in order to produce the best estimator in the least squares sense, cf. Liptser et al. (2001) [118].

Mathematical Framework

Let a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ and $T > 0$ a fixed time horizon. Consider an economy which consists of two assets. The riskless asset is denoted by B with return[1] r, while the risky price process, henceforth the stock, is an adapted positive process $Y = (Y_t)_{t\in[0,T]}$ which is a $P-$semimartingale. The filtration $(\mathcal{F}_t)_{t\in[0,T]}$ denotes the model filtration, while:

$$\mathcal{G}_t = \sigma\left(Y_s, s \in [0, t]\right)$$

represents the stock price filtration. Recalling the previously mentioned distinction between risk and uncertainty, we note that having access to $\mathcal{F}$ is equivalent to be subject to risk while having access to $\mathcal{G}$ is the situation where uncertainty matters, as $\mathcal{G}$ is less informative than $\mathcal{F}$. In the following, we will be interested in studying financial decisions when having access to $\mathcal{G}$.

Departure from Complete Observation

In general, filtration $\mathcal{F}$ is not available to the agent who can only access the stock price filtration $\mathcal{G}$. Therefore, the uncertainty resumes to a situation of limited information, as $\mathcal{G} \subset \mathcal{F}$. Within this setup, the investor cannot disentangle - for example - the drift term from other sources of randomness. More specifically, in the economy we consider, growth rates are altered by infrequent large shocks and continuous small shocks. Investors observe changes in returns but cannot perfectly distinguish their dynamics. Instead, he/she needs to solve a signal extraction problem. This works as follows: As investors do not have a perfect knowledge of the process associated with stock price dynamics (uncertainty in the drift term, for example), they need to make

[1] We note that the return r can be deterministic or stochastic.

proper estimates using all available data. As more data become available, new information is incorporated into existing beliefs with a certain weight to form posterior beliefs, through a Bayesian updating scheme. To be more precise, the investors lack knowledge, not on the model distribution, but on the parameters of the distribution. Therefore, the problem of reducing the uncertainty by learning is equivalent to solve a filtering problem, cf. Liptser and Shiryaev (2001) [118].

Portfolio Optimization

Thereafter, we consider the next problem: the investor wants to solve the following expected utility problem:

$$u(x) = \sup_{\phi \in \mathcal{A}(x,\mathcal{G})} E^P \left[U \left(X_T^{\phi} \right) \right] \tag{1.3}$$

where U is some utility function, X is the wealth process, ϕ is the investment policy process which is defined on $\mathcal{A}(x, \mathcal{G})$, the set of $\mathcal{G}-$admissible strategies with initial capital $x > 0$ and P a probability measure on $(\Omega, \mathcal{G})$.

Remark 1.2.1 *Problem (1.3) when moments are fully observable, i.e.: when P is defined on $(\Omega, \mathcal{F})$ and the set of investment policies is given by $\mathcal{A}(x, \mathcal{F})$ is commonly termed the Merton problem, cf. Merton (1971) [126].*

1.2.2 Partial Information Setup

At each time, investors form estimates based on their available information and use them to endogeneously determine moments of returns, conditional on their observations. In other words, investors are engaged in continuous Bayesian revisions, cf. Karni (2005) [101], using all available historical information to determine, at each time, the posterior distributions of the unobservable market factors. Their dependency on historical information, negates the Markovian property. However, under certain conditions, we can recapture the Markovian structure. These are related to the *efficiency* of the presentation of the historical information through the conditional moments. If the posterior distribution has a finite number of moments and if these moments can be updated recursively, we can recapture a Markovian structure and identify a proper state vector solution. When the filter is not *compact*, that is, the posterior distribution has infinitely many moments, one might use an approximation, cf. Benes and Karatzas (1983) [16] and Chiarella et al. (2001) [27].

Parameterized Bayesian Family

We now define in a more concise way how uncertainty matters. For this, we consider a random variable $\theta : \Omega \to \mathbb{R}$ such that Y writes as follows:

$$Y_t = \theta t + W_t$$

where W is a $P-$Brownian motion, independent of θ. Thereafter, the filtration $\mathcal{F} = \{\mathcal{F}_t, t \in [0,T]\}$ denotes the $P-$augmentation of the enlarged filtration:

$$\mathcal{F}_t = \sigma\left(W_s, s \in [0,t]\right) \vee \sigma\left(\theta\right)$$

Definition 1.2.2 (Enlarged Measure) *Consider the* $(P, \mathcal{F})-$*martingale:*

$$\Lambda_t = \mathcal{E}\left(\theta W\right)_t$$

where $\mathcal{E}$ denotes the Doléans-Dade exponential and define:

$$\widetilde{P}_T\left(A\right) \triangleq E^P\left[\mathbf{1}_A \cdot \Lambda_T\right], \ A \in \mathcal{G}_T$$

a probability measure equivalent to P on $\mathcal{G}_T$.

Let us also consider the $\left(\widetilde{P}_T, \mathcal{G}\right)-$martingale:

$$\widetilde{\Lambda}_t = E^{\widetilde{P}_T}\left[\frac{dP}{d\widetilde{P}_T}|\mathcal{G}_t\right] = E^{\widetilde{P}_T}\left[\Lambda_T|\mathcal{G}_t\right] = E^{\widetilde{P}_T}\left[E^{\widetilde{P}_T}\left[\Lambda_T|\mathcal{F}_t\right]|\mathcal{G}_t\right] = E^{\widetilde{P}_T}\left[\Lambda_t|\mathcal{G}_t\right]$$

and suppose given a prior-probability distribution α on θ, so that:

$$\alpha\left(A\right) = P\left[\theta \in A\right] = \widetilde{P}^T\left[\theta \in A\right], \ A \in \mathbb{R}$$

is the known distribution of the random variable θ under P and so $\widetilde{P}^T$. Then, at a given time $t \in [0,T]$, the posterior distribution of θ under P, given the observations $\mathcal{G}_t$ up to time t, is given by the Bayes' rule, namely:

$$\alpha_t\left(A\right) = P\left[\theta \in A|\mathcal{G}_t\right] = \frac{\nu_t\left(A\right)}{\nu_t\left(\mathbb{R}\right)}$$

with:

$$\nu_t\left(A\right) = E^{\widetilde{P}^T}\left[\mathbf{1}_{\theta \in A}\Lambda_T|\mathcal{G}_t\right] = E^{\widetilde{P}^T}\left[\mathbf{1}_{\theta \in A}\Lambda_t|\mathcal{G}_t\right]$$

The mean of the conditional distribution $\alpha_t\left(\cdot\right)$ is then given by:

$$\widehat{\theta}_t = E\left[\theta|\mathcal{G}_t\right] = \int_{\mathbb{R}} x\alpha_t\left(dx\right)$$

which is the Bayes estimator of θ on the interval $[0,t]$ w.r.t. the prior distribution α, given the observations $\mathcal{G}_{[0,t]}$. Then, we may check that:

$$\widehat{Y}_t = Y_t - \int_0^t \widehat{\theta}_s ds$$

is a $(R, \mathcal{G})-$Brownian motion, which is called the innovation process in filtering theory, cf. Liptser and Shiryaev (2001) [118].

Related Aspects

Absolute Continuity Hypothesis As it will be studied in the next Chapters, the previous result derives (extensively) from the Girsanov's theorem. When considering stochastic processes in continuous time, it is only possible to consider changes of probability measures between measures which are absolutely continuous between each others. Therefore, it will not be possible to learn on singular models. This happens, for example, when one considers a Itô-Lévy model with a parameterization of the volatility coefficient which entails a singular family of models of the form:

$$\mathcal{P} = \{P^{\sigma} : \sigma \epsilon \left[\underline{\sigma}, \overline{\sigma}\right]\}$$

Time Discretization of Information As previously said, the class of models on which learning will be applicable is restricted to the ones where a property of absolute continuity between the reference and the parameterized measures, R and P^{θ} respectively, holds. This entails that learning on the volatility or jump amplitude of a Lévy process is not granted. As it will be thoroughly developped in Chapter 2, this limitation is due to probabilistic arguments related to the Girsanov's theorem and in particular to its (most commonly used) continuous-time version. Nevertheless, this restriction can be bypassed when working with discrete-time observations. Consider the filtration:

$$\mathcal{G}\left(n\right) = \sigma\left(\left(Y_{\tau_i}, \tau_i\right), \tau_i \leqslant n\right)$$

which is the discrete (randomly sampled) version of $\mathcal{G}_t$. Within such a setting, one may assume, for example, that market data arrive according to a marked-point process, cf. Chapter 4, thus allowing to learn on all dimensions of Y.

1.2.3 Properties of Partial Information

Reduction to the Full Observation Case

The previous discussion will be made more precise by deriving the mathematical foundation of the *equivalence principle* via filtering. From Definition 1.2.2:

$$\Lambda_T = \frac{dR}{d\widetilde{P}^T}|_{\mathcal{F}_T}$$

We then have the following lemma, which - in a more general setup - will be proved in Chapter 2 for the case of a Itô-Lévy process or in Chapter 4 for the case of a marked point process.

Lemma 1.2.3 *The random variable $\widetilde{\Lambda}_t = 1/\Lambda_t$ satisfies:*

$$E^R\left[\widetilde{\Lambda}_t|\mathcal{G}_T\right] = 1,\ a.s.,\ t \in [0,T]$$

where $\mathcal{G}$ is the market information available to the investor.

Let us now define the stochastic control problem:

$$\widetilde{u}^T(x) = \sup_{\phi \epsilon \mathcal{A}(x,\mathcal{G})} E^{\widetilde{P}^T}\left[U\left(X_T^{\phi}\right)\right] \tag{1.4}$$

and we have the proposition.

Proposition 1.2.4 (Filtering Equivalence Principle) *The optimal value functions $u(x)$ (1.3) and $\widetilde{u}^T(x)$ (1.4) admit the similar formal solution.*

The problem (1.4) will be called the investor reformulation problem via filtering. It expresses how and why the construction of a $\sigma-$algebra equivalent economy, which is observationally equivalent to the original one, allows to derive a solution to the optimization problem in a similar way.

Mathematical treatment of this question, in the case of a continuous (Wiener) economy, was treated by Karatzas and Xue (1991) [99], Lakner (1998) [111] or Pham and Quenez (2001) [139], using the martingale approach of the optimization problem through duality, and Lasry and Lions (1999) [113] via the equipement of the dynamic programming approach.

From Continuous to Discrete Observations

As previously mentioned, the partial information framework requires the statistical estimation of the unobservable moments estimates inside of the theoretical model. When turning to the choice of an inference strategy, an important debate in this area concerns the question of what sampling scheme to use, if any is available and in any case what to do with the sampling times, cf. Ait-Sahalia and Mykland (2003) [3]. The traditional answer, and the usual procedure done in empirical finance, is to view the sampling as occuring at fixed discrete time intervals, such as a day, a week or a month. However, this situation is, in some circumstances, not realistic. In fact, all transaction-level data are available at irregularly and randomly spaced times, as shown in Figure 1.1.

When a theorerical model is chosen to describe the factors of an economy, it is spelled out in continuous time, even if the data are randomly spaced in time. However, its estimation necessarily relies on discretely sampled data, as they are the only kind of data available to the empiricists. By now, the financial implications of the effects of the sampling discreteness are well-known, cf. Ait-Sahalia (2002a) [1]. Moreover, various estimation methods that take into account this effect have been designed, such as the ones of Ait-Sahalia (2002b) [2], Ait-Sahalia and Mykland (2004) [4] or Duffie and Glynn (2004) [57] which are based on modified versions of maximum likelihood or moments-based methods where the sampling depends on an arrival intensity. Another proposition is based on models of high-frequency data, see Andersen et al. [6],

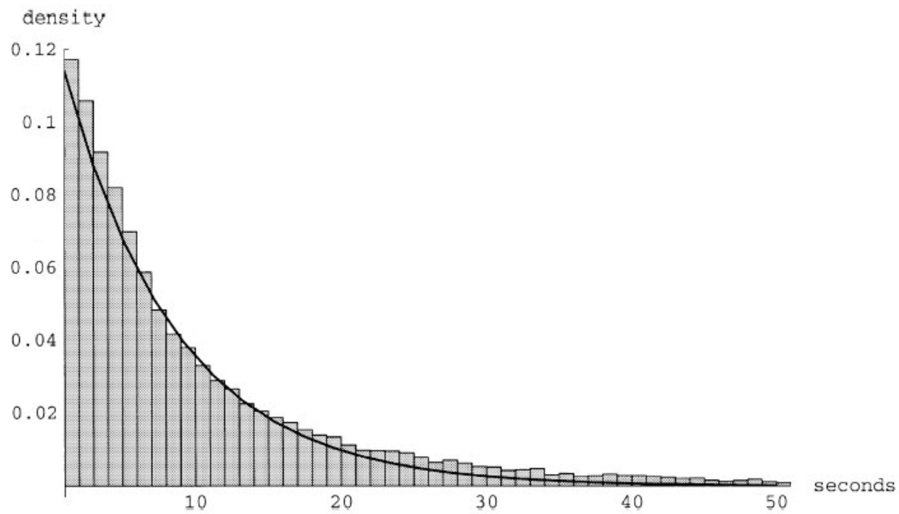

Figure 1.1: Distribution of the sampling intervals for Nokia, January 2000.

where the vector of stock prices is supposed to be observed only discretely at random times, which in turn is modelized by a marked point process, cf. Brémaud (1981) [20], as in Frey and Runggaldier (2001) [71].

In the following, we will examine and explain the additional effect that the randomness of the sampling intervals may have when estimating a continuous-time model with discrete data. From this, we will study, from the maximum expected utility problem, the gain effects related to this *additional* information.

2

Partially Observed Jump Diffusions

Agenda

Note: This chapter is an adapted version of a co-authored paper with M. Jeanblanc and V. Lacoste.

Abstract. This paper studies the question of maximizing terminal wealth from expected utility in a multidimensional jump-diffusion model. The special feature of our approach is that the investor only observes the vector of stock prices, therefore leading to a partial information framework. Resorting to non-linear filtering and change of measure techniques, we show that the optimization problem can be rewritten such that coefficients depend only on past history of observed prices. Through duality approach of the problem, we derive the optimal value function. Then, by resorting to the Bellman and the Hamilton-Jacobi-Bellman equations, we characterize the optimal investment policy. As examples, special attention is given to the three more standard utility functions for which the optimal value functions, investment strategies and risk-premia processes are elicited.

2.1. Introduction

In this paper, we study the question of maximizing terminal wealth from expected utility in a partial information framework. This situation appears when investors only observe the vector of stock prices and cannot disentangle the drift term from the other sources of uncertainty. More specifically, in the economy we consider, growth rates are altered by infrequent large shocks and continuous small shocks. Investors observe changes in returns but cannot perfectly distinguish their dynamics. Instead, they solve a signal extraction problem.

Motivated by recent findings which indicate the importance of jumps in returns to fully capture the empirical features of equity index returns (Eraker et al. 2003 [60]), we consider a market model where the stock price processes follow a mixed jump-diffusion equation where the growth rate and the jump time intensity are unobservable. This seems reasonable, since jumps are often generated by various external sources whose impact cannot completely be analyzed. Also, the changing conditions on the drift and jump intensity are modeled via a strong Markov process; see Duffie et al. (2003) [56]. This unobservable process may be interpreted as an environment process which collects factors which are relevant for the stock price dynamics, like economical news, political situations, technical progress.

The optimization problem with full information goes back to Merton (1971) [126] who solved the question via the Bellman equation of dynamic programmig. For the case of complete markets, we refer to Karatzas et al. (1987) [98] or Cox and Huang (1989) [34]. Models with incomplete information have been investigated by Detemple (1986) [48], Dothan and Feldman (1986) [53] and Gennotte (1986) [74] and deal with the case of a unobserved appreciation rate, by using dynamic programming methods in the linear Gaussian filtering case. Lakner (1995, 1998) [122], [111] solved the partial optimization problem via a martingale approach, provided characterization of the optimal strategy via Malliavin calculus and worked out the special case of the linear Gaussian model. In this essence, Pham and Quenez (2001) [139] treat the case of an incomplete stochastic volatility model and Sass and Haussmann (2004) [149] the case of hidden Markov model filtering. Any of these papers consider the case of a jump-diffusion model. In fact, it would be more realistic to assume that the stock price dynamics follow a jump-diffusion model. In fact, empirical work has shown that log-returns are not generally normally distributed and that the stock price process should contain a jump component. Recent references on dynamic optimization with jumps include Jeanblanc-Picqué and Pontier (1990) [94], Shirakawa (**??**), Bellamy (**??**) or Liu et al. (2003) [120]. Unlike the present setup, the proposed market models are generally supposed to be *completely* observable, in the sense where both the appreciation rate and the jump intensity are observable. In this paper, we consider the situation where both of them are unobservable and seek to optimize the Merton's problem of maximizing expected utility from terminal wealth.

In order to solve this problem, the common way is to use the filtering theory, so as to reduce the stochastic control problem with partial information to one with complete observation. It is then possible to solve this problem either with the martingale approach, cf. Kramkov and Schachermayer (1999) [107] , or via stochastic control methods, cf. Framstad et all (1999) [70]. In this paper, we combine stochastic filtering techniques and a martingale duality approach to characterize the value function and the optimal portfolio of the optimization problem. Nevertheless, as the reduced market model is not complete, we complement the martingale approach by using the theory of stochastic control to solve the problem explicitly. Section 2 states the framework and recall some known results on portfolio optimization. Then, in Section 3, we show that conditioning arguments can be used to replace the original partial information problem by a full information one which depend only on past history of the observed prices. Section 4 present derivation of the optimal wealth and value function within partial information, as well as a formal proof of a solution of the HJB equation of the problem. The special cases of power, logarithmic and exponential utility functions are studied and explicit formulae for the value functions and risk-premia processes are obtained.

2.2. Formulation of the Problem

2.2.1 The Economy

We consider an economy defined on the complete probability space $(\Omega, \mathcal{A}, P)$ for a finite time span $[0, T]$ with $T \in (0, \infty)$, equipped with a filtration $\mathcal{A} =$

$(\mathcal{A}_t)_{t\in[0,T]}$ satisfying the usual conditions and on which are defined all the stochastic processes involved in this paper. In this market, $1+m$ assets are defined. The first one is a non-risky asset which pays no dividends. We refer to it as the bank account and suppose that its price satisfies:

$$dS_t^0 = r_t S_t^0 dt,\ \ S_0^0 = 1$$

where $(r_t, t \in [0,T])$ is uniformly bounded. The m other assets are risky and we refer to them as the stocks. Letting S_t^i be the positive price at time t of the i^{th} asset, we define its *return process by* $dR_t^i = \left(S_t^i\right)^{-1} dS_t^i$ and assume that its evolution is modeled, for $i \in \|1;m\|$, through the following equation:

$$dR_t^i = \mu_t^i\left(\theta_t\right)dt + \sum_{j=1}^{m}\left\{\sigma^{ij}\left(t,R_t^i\right)dW_t^j + \omega^{ij}\left(t,R_t^i\right)\left(dM_t^j + \lambda_t^j\left(\theta_t\right)dt\right)\right\}$$

where $R_0^i = r_i^0$ is a constant and the process $(\theta_t, t\in[0,T])$ stands for an economic factor process. Here, W is a $m-$dimensional Brownian motion, M is the compensated martingale of a $m-$ dimensional inhomogeneous Poisson process N, whose components have no common jumps: each N^j is such that $M_t^j = N_t^j - \int_0^t \lambda_s^j\left(\theta_s\right)ds$ with a $\mathbb{R}-$valued intensity $\lambda^j\left(\theta\right)$. Also, processes W and N are independent of each other and $\mathcal{A}-$adapted. We denote by $\mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}$ the *model* information, which is the filtration generated by the random processes W, N and θ, so:

$$\mathcal{F}_t = \sigma\left(W_s, s\in[0,t]\right) \vee \sigma\left(N_s, s\in[0,t]\right) \vee \sigma\left(\theta_s, s\in[0,t]\right),\ \mathcal{F}_t \subset \mathcal{A}_t$$

In general, the economic factor process θ is not directly observable, which resumes to: $\mathcal{F}$ is not available. This situation therefore entails that the *means* of the continuous and counting processes are not observed.

Thereafter, we shall assume that the following assumptions are satisfied.

Assumption 2.2.1 *For $(i,j) \in \|1;m\|$ and $t \in [0,T]$:*

1. *$r_t > 0,\ \mu^i\left(t,x\right) \in \mathbb{R}$, a.s.,*
2. *$\mu_t^i\left(x\right)$ and $\lambda_t^i\left(x\right)$ are $\mathcal{F}-$adapted and uniformly bounded processes,*
3. *$\sigma^{ij}\left(t,x\right)$ and $\omega^{ij}\left(t,x\right)$ are known Borel functions satisfying Lispschitz and growth conditions, such that $\sigma^{ij}\left(t,x\right) > 0$ and $\omega^{ij}\left(t,x\right) > -1$.*
4. *θ is a $\mathbb{R}-$valued càdlàg homogeneous and $\mathcal{A}-$adapted Markov process,*
5. *To avoid technical difficulties[1], W and N are assumed to be independent of θ and the possible jumps of θ are disjoint from those of N.*

[1] If one assumes common jumps between the state process θ and the observations S, these cannot be made independent under a transformation of measures, which is the method used thereafter to derive the filtering equation; this is the method of the probability of reference.

Using vector notations, the process R_t satisfies:

$$dR_t = b_t(\theta_t)\, dt + \sigma_t dW_t + \omega_t dM_t \tag{2.1}$$

where:

$$b_t(\theta_t) = \left(b_t^1(\theta_t), ..., b_t^m(\theta_t)\right)^\intercal, \ \chi_t = \left(\left(\chi_t^1\right)^\intercal, ..., \left(\chi_t^m\right)^\intercal\right)^\intercal, \ \chi = \{\sigma, \omega\}$$

with $b_t^i(\theta_t) \equiv b^i\left(t, \theta_t, R_t^i\right) = \mu^i\left(t, \theta_t, R_t^i\right) + \omega^i\left(t, R_t^i\right)\lambda^\intercal(t, \theta_t)$ and:

$$\lambda_t(\theta_t) = \left(\lambda_t^1(\theta_t), ..., \lambda_t^m(\theta_t)\right), \ \sigma_t^i = \left(\sigma_t^{i,1}, ..., \sigma_t^{i,m}\right), \ \omega_t^i = \left(\omega_t^{i,1}, ..., \omega_t^{i,m}\right)$$

Under Assumption 2.2.1, the coefficients in the SDE (2.1) are such that a unique $\mathcal{A}-$adapted solution exists that does not explode until time T, cf. Protter (1990) [140]. At this stage, we shall also assume that the $m \times m-$matrix σ_t is *invertible* for all $t \in [0, T]$ and that $\mu(\theta)$ and σ satisfy:

$$\int_0^T E\left[|\mu_t(\theta_t)|\right] dt < \infty, \ \int_0^T E\left[|\sigma_t \sigma_t^\intercal|\right] dt < \infty, \ P - a.s. \tag{2.2}$$

Finally, the $m \times m-$matrix ω_t is *invertible,* for all $t \in [0, T]$. This implies that, since there are no common jumps among the components of the counting process N, there exists a one-to-one correspondence between the observation of the size of a jump of S and the knowledge of which of the process N^i, $i \in \|1; m\|$, has jumped at that time.

The *excess return process* $dY_t = dR_t - r_t \mathbf{1}_m dt$, $t \in [0, T]$, reads:

$$dY_t = \left[b_t(\theta_t) - r_t \mathbf{1}_m\right] dt + \sigma_t dW_t + \omega_t dM_t \tag{2.3}$$

and we note that Y_0 and Y_{0-} are both constants.

Remark 2.2.2 *Note that the aforementioned framework is incomplete due to the random jumps of prices modeled by a jump-diffusion process. In the special case where the total number of randomness (Wiener and Poisson processes) is equal to the number of risky assets, then the market will be complete.*

Partial Information Setup

We now consider, as in Detemple (1986) [48] or Lakner (1995, 1998) [122], [111], that some agents in the financial market have not access to the filtration $\mathcal{F}$ and can only observe the assets prices. Thus, the observations are given by the sequence $\left(S_t^0, S_t^i\right)_{t \geqslant 0}^{i \in \|1;m\|}$ and we denote by $\mathcal{G} = (\mathcal{G}_t, t \in [0, T])$, with $\mathcal{G} \subsetneq \mathcal{F}$, the $P-$augmentation of the *market* filtration generated by the $1 + m$ assets:

$$\mathcal{G}_t = \sigma\left(\left(S_s^0, S_s^i\right), i \in \|1; m\|, s \in [0, t]\right)$$

and we assume that θ_0 is independent of $\mathcal{G}_\infty$. From a filtering perspective, θ_0 is a random variable (which is a-priori fixed) and so $\mathcal{F}_0$ is not trivial.

Assumption 2.2.3 *The interest rate r is $\mathcal{G}-$adapted.*

Under this assumption, we have:

$$\mathcal{G}_t = \sigma\left(Y_s^i, i \in \|1; m\|, s \in [0,t]\right)$$

Remark 2.2.4 *By definition, the processes σ_t and ω_t are $\mathcal{G}_t-$adapted.*

2.2.2 The Optimization Problem

Trading Strategies

By denoting $\phi_s = \left(\phi_s^1, ..., \phi_s^m\right)^\intercal$ the vector of fraction of wealth invested in the m risky assets at time s, a self-financing trading strategy is a pair (x_0, ϕ) where $x_0 \geqslant 0$ is the initial investment and ϕ is an $\mathbb{R}^m-$valued and $\mathcal{G}-$adapted process such that the value/wealth process:

$$dX_t^\phi = X_{t-}^\phi \left(\left(r_t + \phi_t^\intercal \left(b_t\left(\theta_t\right) - r_t \mathbf{1}_m\right)\right)dt + \phi_t^\intercal \sigma_t dW_t + \phi_t^\intercal \omega_t dM_t\right) \tag{2.4}$$

with $X_0^\phi = x_0$, is $P-$a.s. well defined. In the following, the notation $X_t^{x_0,\phi}$ is a shorthand for $X_0^\phi = x_0$ and the wealth process X_t^ϕ, for $t \geqslant 0$, satisfies (2.4). The class of admissible strategies (at time t) reads:

$$\begin{aligned}\mathcal{S}\left(t\right)=&\left\{\phi : [0,T] \times \Omega \to \mathbb{R}^m, \mathcal{G}-predictable\right.\\ &\left., \exists K > -\infty, \forall s \geqslant t, \int_t^s \phi_u^\intercal dR_u \geqslant -K\right\}\end{aligned}$$

Remark 2.2.5 *The numéraire investment is given by $\phi_t^0 = X_t^{x_0,\phi} - \sum_{i=1}^m \phi_t^i$.*

Optimizing Terminal Wealth

A function $U : \mathbb{R} \to \mathbb{R}$ is called a utility function if it is strictly increasing, strictly concave, of class $\mathcal{C}^2$ and satisfies:

$$U^{'}\left(0^+\right) = \infty,\ U^{'}\left(\infty\right) = 0 \tag{2.5}$$

and has reasonable asymptotic elasticity:

$$AE_{0^+}\left(U\right) := \liminf_{x \to 0^+} \frac{xU^{'}\left(x\right)}{U\left(x\right)} > 1,\ AE_{+\infty}\left(U\right) := \liminf_{x \to \infty} \frac{xU^{'}\left(x\right)}{U\left(x\right)} < 1 \tag{2.6}$$

The optimization problem the investor faces is to maximize the expected utility of his/her terminal wealth over the class of admissible policies.

Definition 2.2.6 *Let U be a utility function. We then define:*

$$J_{\phi}^{\mathbb{G}}(t,x) \triangleq E^{P}\left[U(X_T^{x,\phi})|\mathcal{G}_t\right]$$

$$u^{\mathbb{G}}(t,x) = \sup_{\phi\epsilon\mathcal{S}(t)} J_{\phi}^{\mathbb{G}}(t,x) \tag{2.7}$$

Then, for $x_0 > 0$, a portfolio strategy $\phi^* \in \mathcal{S}(0)$ is optimal if:

$$u^{\mathbb{G}}(0,x_0) = J_{\phi^*}^{\mathbb{G}}(0,x_0)$$

From this, we note that $J_{\phi}^{\mathbb{G}}(t,x)$ is a $\mathcal{G}_t-$measurable random variable, so that $J_{\phi}^{\mathbb{G}}(0,x)$ and $u^{\mathbb{G}}(0,x)$ depend on the random variable θ_0. Also, the control problem (2.7) is stated under partial information. In order to be solved, we need to reduce it to a control problem with complete observation.

2.3. Mechanics of the Learning Process

2.3.1 Setup

Since the Markov process $(\theta_t, t \in [0,T])$ is not $\mathcal{G}-$adapted, it is natural to introduce the $\mathcal{G}-$conditional law of the random variable θ_t, say:

$$\pi_t(f) = E^{P}\left[f(\theta_t)|\mathcal{G}_t\right] \tag{2.8}$$

for any $\mathbb{R}-$valued measurable function f such that $E^{P}\left[|f(\theta_t)|\right] < \infty$. By construction, $\pi_t(f)$ is $\mathcal{G}_t-$adapted.

We now recall Itô's formula for jump-diffusions.

Definition 2.3.1 *For a semimartingale x_t and a twice differentiable function f, Itô's formula yields that the semimartingale $f(x_t)$ equals:*

$$\begin{aligned} f(x_t) = f(x_0) + \int_0^t f'(x_{s-})\,dx_s + \frac{1}{2}\int_0^t f''(x_{s-})\,d\left[x^c, x^c\right]_s \\ + \sum_{s\leqslant t}\Big(f(x_s) - f(x_{s-}) - f'(x_{s-})\,\Delta x_s\Big) \end{aligned} \tag{2.9}$$

where x^c denotes the continous part of x.

Reference Measure

Following Kallianpur (1980) [96], which discusses the Wiener case, we introduce a new measure P^0, termed the reference measure[2]. To this end, let:

$$H_t = 1 - \int_0^t H_{s-} \left[a_s^\intercal (\theta_s) \, dW_s + \left(\mathbf{1}_m^\intercal - c_s^\intercal (\theta_s) \right) dM_s \right] \tag{2.10}$$

where [3,4]:

$$a_s (\theta_s) = \sigma_s^* \left(b_s (\theta_s) - r_s \mathbf{1}_m \right), \; c_s (\theta_s) = \lambda_s^* (\theta_s) \tag{2.11}$$

and the Poisson integral does not need to be taken in a predictable way, i.e.: with a $\left(\mathbf{1}_m^\intercal - c_s^\intercal (\theta_{s-}) \right)$ integrand, as M and θ have no common jumps.

Proposition 2.3.2 *Assume Assumption 2.2.1 and* $\ln c_s (\theta_s)$ *is (component-wise) bounded on* $[0, T]$*. Then* H_t *is a strictly positive* $(P, \mathcal{F})$ *−martingale.*

Proof. Let H_t^W and H_t^M be the unique solutions of:

$$dH_t^W = -H_t^W a_t dW_t, \; dH_t^M = -H_t^M (1 - c_s) \, dM_t, \; H_0 = 1$$

(in the one-dimensional case). By Itô's formula, we have $H_t = H_t^W H_t^M$. From Assumption 2.2.1, H_t^W is a $(P, \mathcal{F})$ −martingale. In addition, we can show that:

$$H_t^M = e^{\int_0^t (1 - c_s) \lambda_s ds} \prod\nolimits_{0 \leqslant s \leqslant t} c_s \Delta N_s$$

And as H^W and H^M are orthogonal, the conclusion follows. ∎

From Proposition 2.3.2, we then have:

Proposition 2.3.3 *On* $(\Omega, \mathcal{F})$*, we define the measure* $P^0 \sim P$ *by:*

$$\frac{dP^0}{dP} |_{\mathcal{F}_T} = H_T \tag{2.12}$$

The Girsanov transformation ensures that:

$$W_t^0 = W_t + \int_0^t a_s (\theta_s) \, ds \textit{ is a } \left(P^0, \mathcal{F} \right) - \textit{Brownian motion} \tag{2.13}$$

$$M_t^0 = M_t + \int_0^t D_s^\theta \left[\mathbf{1}_m - c_s (\theta_s) \right] ds \textit{ is a } \left(P^0, \mathcal{F} \right) - \textit{martingale} \tag{2.14}$$

with $D_s^\theta =$*Diag*$(\lambda_s (\theta_s))$*. Thus,* N_t *is an* $\mathcal{F}_t$*−Poisson process with* P^0*−intensity Diag*$(\mathbf{1}_m)$*, so that* $M_t^0 = N_t - \mathbf{1}_m^\intercal t$ *is the* P^0*−compensated martingale of* N*.*

---

[2] The measure P^0 will be subsequently used to derive filtering equations via the reference probability approach, cf. Kallianpur (1980) [96] for a presentation in the Wiener case.

[3] We note $a_s (\theta_s)$ and $c_s (\theta_s)$ for all $s \in [0, T]$ to express the dependence of the processes a and c on both the time variable s and on the unknown process θ_s.

[4] In the following, x^* denotes the component-wise inverse of $x \epsilon \mathbb{R}^k, k > 1$.

Then, the excess return process (2.3) is a P^0-local martingale of the form:

$$dY_t = \sigma_t dW_t^0 + \omega_t dM_t^0 \tag{2.15}$$

and one sees that the processes Y and θ get *decoupled* under the measure P^0, i.e.: the unknown economic factor θ vanishes from the dynamics of Y.

We begin by proving a lemma which will be of paramount importance in the following, extending a result of Pham and Quenez (2001) [139].

Lemma 2.3.4 *The filtration $\mathcal{G}$ is the augmented filtration of (W^0, M^0).*

Proof. We rewrite (2.15) in the more convenient form:

$$\omega_t^* \left(dY_t + \omega_t dt\right) = \sigma_t \omega_t^* dW_t^0 + dN_t \triangleq dY_t^0 \tag{2.16}$$

from which we get, using the fact that the predictable covariance process is $\mathcal{G}-$adapted, that $\left(\sigma_t \omega_t^*\right)^2 dt + dt$ is $\mathcal{G}-$adapted. Besides, as right brackets are adapted, it follows that $\left(\sigma_t \omega_t^*\right)^2 dt + dN_t$ is $\mathcal{G}-$adapted which leads to N_t is $\mathcal{G}_t-$adapted. Then, by (2.16), W_t^0 is $\mathcal{G}-$adapted. Noting $\mathcal{F}^0$ the augmented filtration of the processes (W^0, N), this implies that $\mathcal{F}^0 \subseteq \mathcal{G}$. Conversely, following Protter (1990) [140], as σ and ω are $\mathcal{G}-$adapted processes and under Assumption 2.2.1, we get that the unique solution of (2.16), say Y_t^0, is $\mathcal{F}_t^0-$adapted, so that $\mathcal{G} \subseteq \mathcal{F}^0$, hence yielding $\mathcal{G} = \mathcal{F}^0$. ∎

This result allows then to prove a theorem, which extends those of Liptser and Shiryaev (2001) [118] or Brémaud (1981) [20] to the case of jump-diffusions. By Itô's calculus applied to (2.10), the $\left(P^0, \mathcal{F}\right)-$martingale H_t^{-1} reads:

$$H_t^{-1} = 1 + \int_0^t H_{s-}^{-1} \left[a_s^{\intercal}\left(\theta_s\right) dW_s^0 + \left(\left(c_s^{\intercal}\left(\theta_s\right)\right)^* - \mathbf{1}_m^{\intercal}\right) dM_s^0\right] \tag{2.17}$$

Theorem 2.3.5 *We have:*

$$\begin{aligned} E^{P^0}\left[H_t^{-1} | \mathcal{G}_t\right] = 1 &+ \int_0^t E^{P^0}\left[H_{s-}^{-1} a_s^{\intercal}\left(\theta_s\right) | \mathcal{G}_s\right] dW_s^0 \\ &+ \int_0^t E^{P^0}\left[H_{s-}^{-1}\left(\left(c_s^{\intercal}\left(\theta_s\right)\right)^* - \mathbf{1}_m^{\intercal}\right) | \mathcal{G}_s\right] dM_s^0 \end{aligned} \tag{2.18}$$

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

Proof. In order to prove (2.18), it is sufficient to show that:

$$E^{P^0}\left[H_t^{-1} \mathbf{1}_A\right] = E^{P^0}\left[\mathcal{R}_t \mathbf{1}_A\right] \tag{2.19}$$

for any $A \in \mathcal{G}_t$, where $\mathcal{R}_t$ denotes the right-hand side of (2.18), so that $\mathcal{R}_t = E^{P^0}\left[H_t^{-1}|\mathcal{G}_t\right]$. Via the Martingale Representation theorem for $\mathcal{G}-$martingales w.r.t. Brownian motion W^0 and compensated Poisson process M^0, as quoted in Runggaldier (2003) [145], $\mathbf{1}_A$ is, for $t \geqslant 0$, of the form:

$$\mathbf{1}_A = \mathcal{M}_0 + \int_0^t U_s dW_s^0 + \int_0^t V_s dM_s^0 \triangleq \mathcal{M}_0 + \mathcal{M}_t^0 \tag{2.20}$$

where U, V are $\mathcal{G}-$predictable processes and $\mathcal{M}_0 \in \mathcal{G}_0$ is not a constant. Hence, from the definition of $\mathcal{R}$, representation (2.20) for $\mathbf{1}_A$, (2.19) will follows from:

$$E^{P^0}\left[H_t^{-1}\mathcal{M}_0\right] + E^{P^0}\left[H_t^{-1}\mathcal{M}_t^0\right] = E^{P^0}\left[\mathcal{R}_t\mathcal{M}_0\right] + E^{P^0}\left[\mathcal{R}_t\mathcal{M}_t^0\right]$$

or equivalently from:

$$E^{P^0}\left[\int_0^t H_s^{-1} d\nu_s^0 \cdot \mathcal{M}_0\right] = E^{P^0}\left[\mathcal{R}_t^0\mathcal{M}_0\right] \tag{2.21}$$

$$E^{P^0}\left[\int_0^t H_s^{-1} d\nu_s^0 \cdot \mathcal{M}_t^0\right] = E^{P^0}\left[\mathcal{R}_t^0\mathcal{M}_0^t\right] \tag{2.22}$$

with:

$$d\nu_s^0 = a_s^{\intercal}(\theta_s)\, dW_s^0 + \left(\left(c_s^{\intercal}(\theta_s)\right)^* - \mathbf{1}_m^{\intercal}\right) dM_s^0$$

and:

$$\mathcal{R}_t^0 = \int_0^t E^{P^0}\left[H_{s-}^{-1} a_s^{\intercal}(\theta_s)\,|\mathcal{G}_s\right] dW_s^0 + \int_0^t E^{P^0}\left[H_{s-}^{-1}\left(\left(c_s^{\intercal}(\theta_s)\right)^* - \mathbf{1}_m^{\intercal}\right)|\mathcal{G}_s\right] dM_s^0$$

Using the definition of $\mathcal{M}_t^0$, (2.22) reads:

$$\begin{aligned} &E^{P^0}\left[\int_0^t \left(H_{s-}^{-1} a_s^{\intercal} U_s + H_{s-}^{-1}\left(\left(c_s^{\intercal}\right)^* - \mathbf{1}_m^{\intercal}\right) V_s\right) ds\right] \\ &= E^{P^0}\left[\int_0^t \left(E^{P^0}\left[H_{s-}^{-1} U_s a_s^{\intercal}|\mathcal{G}_s\right] + E^{P^0}\left[H_{s-}^{-1} V_s\left(\left(c_s^{\intercal}\right)^* - \mathbf{1}_m^{\intercal}\right)|\mathcal{G}_s\right]\right) ds\right] \end{aligned}$$

(covariation terms being null) hence yielding:

$$E^{P^0}\left[\int_0^t E^{P^0}\left[H_{s-}^{-1} U_s a_s^{\intercal}|\mathcal{G}_s\right] ds\right] = E^{P^0}\left[\int_0^t H_{s-}^{-1} U_s a_s^{\intercal} ds\right]$$

$$E^{P^0}\left[\int_0^t E^{P^0}\left[H_{s-}^{-1} V_s\left(\left(c_s^{\intercal}\right)^* - \mathbf{1}_m^{\intercal}\right)|\mathcal{G}_s\right] ds\right] = E^{P^0}\left[\int_0^t H_{s-}^{-1} V_s\left(\left(c_s^{\intercal}\right)^* - \mathbf{1}_m^{\intercal}\right) ds\right]$$

by definition of conditional expectation, so that (2.22) holds true. Also, in (2.21), we note that $\mathcal{M}_0 \in \mathcal{G}_0$ and similar arguments yield:

$$\begin{aligned} E^{P^0}\left[\mathcal{R}_t^0 \mathcal{M}_0\right] &= E^{P^0}\left[\int_0^t E^{P^0}\left[H_{s-}^{-1} a_s^{\intercal}\left(\theta_s\right) \mathcal{M}_0 | \mathcal{G}_s\right] dW_s^0\right. \\ &\quad \left. + \int_0^t E^{P^0}\left[H_{s-}^{-1}\left(\left(c_s^{\intercal}\left(\theta_s\right)\right)^* - \mathbf{1}_m^{\intercal}\right) \mathcal{M}_0 | \mathcal{G}_s\right] dM_s^0\right] \\ &= E^{P^0}\left[\int_0^t H_{s-}^{-1} a_s^{\intercal}\left(\theta_s\right) \mathcal{M}_0 dW_s^0 + \int_0^t H_{s-}^{-1}\left(\left(c_s^{\intercal}\left(\theta_s\right)\right)^* - \mathbf{1}_m^{\intercal}\right) dM_s^0\right] \end{aligned}$$

which concludes the proof. ∎

2.3.2 Resolution of Uncertainty

We now construct the restriction of P to $\mathcal{G}$ equivalent to P^0 on $(\Omega, \mathcal{G})$. In this regard, consider the conditional version of the Bayes formula: for any $\mathcal{F}_t-$measurable and P^0-integrable random variable X, we have:

$$E^{P^0}\left[X|\mathcal{G}_t\right] = \frac{E^P\left[H_t X|\mathcal{G}_t\right]}{E^P\left[H_t|\mathcal{G}_t\right]} \tag{2.23}$$

By setting $X = 1/H_t$ in (2.23), we get:

$$Z_t \triangleq E^{P^0}\left[\frac{1}{H_t}|\mathcal{G}_t\right] = \frac{1}{E^P\left[H_t|\mathcal{G}_t\right]} \tag{2.24}$$

and we have the following result for the representation of Z_t.

Proposition 2.3.6 *The process Z is a positive $\mathcal{G}-$martingale under P^0 and*[5]*:*

$$Z_t = 1 + \int_0^t Z_{s-}\left[\pi_s^{\intercal}\left(a\right) dW_s^0 + \left(\pi_s^{\intercal}(c^*) - \mathbf{1}_m^{\intercal}\right) dM_s^0\right] \tag{2.25}$$

with processes (a, c) given by (2.11).

Proof. Theorem 2.3.5 ensures that Z is a $(P^0, \mathcal{G})-$positive martingale. Then, an application of (2.23) yields the desired result. ∎

Then, from (2.12) and (2.24), we define, at least implicitly, the following measure transformation on $(\Omega, \mathcal{G})$:

$$\frac{dP}{dP^0}|_{\mathcal{G}_T} = Z_T \tag{2.26}$$

and as a consequence of the Girsanov theorem for semimartingales, we have:

[5] Thereafter, we note $\pi_s(f) \equiv E^P\left[f_s\left(\theta_s\right)|\mathcal{G}_s\right]$ for some process f.

Lemma 2.3.7 *For all $t \in [0,T]$:*

$$\overline{W}_t = W_t + \int_0^t \sigma_s^{-1} \left[\mu_s(b_s) - \pi_s(b)\right] ds \text{ is a } (P,\mathcal{G})-\text{Brownian motion} \quad (2.27)$$

$$\overline{M}_t = M_t + \int_0^t \left[\lambda_s(\theta_s) - \pi_s(\lambda)\right] ds \text{ is a } (P,\mathcal{G})-\text{martingale}$$

Proof. Consider W^0 in (2.13) where $a_s(\theta_s)$ is as in (2.11). Then, as W^0 is $\mathcal{G}-$adapted, hence a $\mathcal{G}-$Wiener process, the process:

$$\overline{W}_t \triangleq W_t^0 - \int_0^t \sigma_s^{-1}\left(\pi_s(b) - r_s \mathbf{1}_m\right) ds$$

is a $(P,\mathcal{G})-$martingale with continuous trajectories and $E[\overline{W}_0] = 0$. Thus, we note that $[\overline{W}_t^k, \overline{W}_t^k] = t$ and $[\overline{W}_t^k, \overline{W}_t^l] = 0$ for all $k \neq l$ and $t \in [0,T]$. Then, Levy's Characterization theorem ensures that $\overline{W}$ is a Wiener process. In the same way, consider the process (2.14), termed M^0. We have:

$$\overline{M}_t \triangleq M_t^0 - \int_0^t D_s^\pi \left(\mathbf{1}_m - \pi_s(\lambda)\right) ds$$

where D_s^π =Diag$(\pi_s(\lambda))$, is a $(P,\mathcal{G})-$martingale. ∎

Remark 2.3.8 *The processes $\overline{W}$ and $\overline{M}$ are called innovation processes in filtering theory. To justify this for $\overline{M}$, first note that $d\overline{M}_t = dN_t - \pi_t(\lambda)\,dt$. Then, dN_t is what one observed during the time $[t, t+dt)$ while $\pi_t(\lambda)\,dt = E[dN_t|\mathcal{G}_t]$ is what one would expect to happen in $[t, t+dt)$ conditionally to previous observations in $\mathcal{G}_t$. Thus $dN_t - \pi_t(\lambda)\,dt$ is what is really new.*

Complete Observation Problem

By means of these *innovation* processes, we can describe the dynamics of the partially observable excess return model (2.3) within the framework of a complete observation model:

$$dY_t = \left[\pi_t(b) - r_t \mathbf{1}_m\right] dt + \sigma_t d\overline{W}_t + \omega_t d\overline{M}_t \quad (2.28)$$

with the $\mathcal{G}_t-$conditional counterpart of $b_t(\theta_t)$ given by $\pi_t(b) = \pi_t(\mu) + \omega_t \pi_t(\lambda)$. In the same way, the reduced wealth process (2.4) satisfies:

$$dX_t^\phi = X_{t-}^\phi \left(\left(r_t + \phi_t^\intercal \left(\pi_t(b) - r_t \mathbf{1}_m\right)\right) dt + \phi_t^\intercal \sigma_t d\overline{W}_t + \phi_t^\intercal \omega_t d\overline{M}_t \right) \quad (2.29)$$

and we note $X_t^{\pi,x,\phi}$ as a shortand for $X_0^\phi = x$ and X_t^ϕ follows (2.29).

We have thus reduced the partially observable stochastic control problem (2.30) to a complete observation one. The *reduced* problem is as follows:

Definition 2.3.9 *Let U be a utility function. We then define:*

$$J_\phi(t,x) \triangleq E^{P|_{\mathcal{G}_t}}\left[U(X_T^{\pi,x,\phi})\right]$$

$$u(t,x) = \sup_{\phi\epsilon\mathcal{S}(t)} J_\phi(t,x) \tag{2.30}$$

where $P|_{\mathcal{G}_t}$ is the restriction of P to $\mathcal{G}_t$, cf. (2.26). Then, for an initial endowment $x_0 > 0$, a portfolio strategy $\phi^* \in \mathcal{S}(0)$ is optimal if:

$$u(0,x_0) = J_{\phi^*}(0,x_0)$$

The reduced problem (2.30) solves the original one (2.7). The main difference between these two control problems is that (2.7) depends on the whole history $\mathcal{G}$, while (2.30) depends on $\mathcal{G}$ only through π, i.e.: the filter contains the necessary information to solve the control problem under partial information.

Remark 2.3.10 *In (2.28) or (2.29), we do not have to know the filtering equation for all functions f, but only for b and λ, so as to compute $\pi_t(b)$ and $\pi_t(\lambda)$, cf. Appendix A.1 for the treatment of a particular case.*

2.4. Portfolio Selection Problem

2.4.1 Optimal Value Function

Duality Theory

As shown by Kramkov and Schachermayer (1999) [107], the proof of a solution to problem (2.30) relies upon solving the dual optimization problem:

$$v(y) = \inf_{Q\in\mathcal{Q}} E\left[V\left(y\frac{dQ}{dP}\right)\right],\ y > 0 \tag{2.31}$$

where $\mathcal{Q}$ is the set of equivalent martingale measures given by

$$\mathcal{Q} = \left\{Q \sim P_{|\mathcal{G}} \mid Y\ is\ a\ local\ \ (Q,\mathcal{G}) - martingale\right\}$$

and where the conjugate version of the utility function $U(x)$ is defined by:

$$V(y) = \sup_{x\in\mathbb{R}^+}\left[U(x) - xy\right],\ y > 0$$

Martingale Measures

As a first step, we derive a suitable Martingale Representation theorem for $(P, \mathcal{G})$ −local martingales w.r.t. the innovation processes $\left(\overline{W}, \overline{M}\right)$.

Lemma 2.4.1 *Let M be a $(P, \mathcal{G})$ −local martingale with $M_0 = 0$. Then, there exist $\mathbb{R}^m$−valued and $\mathcal{G}$−adapted processes (A, B) s.t.:*

$$M_t = M_0 + \int_0^t A_s^{\intercal} d\overline{W}_s + \int_0^t B_s^{\intercal} d\overline{M}_s$$

Proof. From Bayes formula, the process M^0 given by $M_t^0 = M_t H_t^{-1}$, with H given by (2.10), is a $(P^0, \mathcal{G})$−martingale. Then, from Lemma 2.3.4, an application of the Martingale Representation Theorem for jump-diffusions, entails that there exist $\mathbb{R}^m$−valued and $\mathcal{G}$−adapted processes (A^0, B^0) s.t.:

$$M_t^0 = \int_0^t \left(A_s^0\right)^{\intercal} dW_s^0 + \int_0^t \left(B_s^0\right)^{\intercal} dM_s^0$$

Then, by applying Itô's formula to $M_t = M_t^0 H_t$ and using Lemma 2.3.7:

$$A_s = H_s(A_s^0 - M_s^0 a_s\left(\theta_s\right)),\ A_s = H_s(B_s - M_s^0 b_s\left(\theta_s\right))$$

which concludes the proof. ■

In the following, we will make the following assumption.

Assumption 2.4.2 *The space $\mathcal{K}$ is such that (γ, ψ) are $\mathbb{R}^m$−valued square integrable processes satisfying $P-a.s.$ the integrability conditions:*

$$\int_0^T \left|\gamma_t \gamma_t^{\intercal}\right| dt < \infty,\ \int_0^T \left|\psi_t\right| \pi_t\left(\lambda\right) dt < \infty,\ \psi_t < 0,\ a.e.\ t \epsilon \left[0, T\right]$$

Then, the process[6]*:*

$$\Lambda_t = 1 - \int_0^t \Lambda_{s-} \left[\left(\gamma_s^{\pi}\right)^{\intercal} d\overline{W}_s + \left(\mathbf{1}_m^{\intercal} - \left(\psi_s^{\pi}\right)^{\intercal}\right) d\overline{M}_s\right],\ \Lambda_0 = 1 \qquad (2.32)$$

is a strictly positive $(P, \mathcal{G})$ −local martingale. Also, when $E^P\left[\Lambda_T\right] = 1$, it is a martingale and then there exists a probability measure Q equivalent to P with:

$$\frac{dQ}{dP}|_{\mathcal{G}_T} = \Lambda_T \qquad (2.33)$$

[6]We note γ_s^{π} and ψ_s^{π} to make apparent the dependence on both time and the filter.

Then, the Girsanov transformation ensures that:

$$\widehat{W}_t = \overline{W}_t + \int_0^t \gamma_s^\pi ds \text{ is a } (Q, \mathcal{G}) - \text{Brownian motion}$$

$$\widehat{M}_t = \overline{M}_t + \int_0^t D_s^\pi \left[\mathbf{1}_m - \psi_s^\pi\right] ds \text{ is a } (Q, \mathcal{G}) - \text{martingale}$$

with $D_s^\pi =$Diag$(\pi_s(\lambda))$ and N is a $(Q, \mathcal{G})-$Poisson process with $\mathcal{G}_t-$intensity $\lambda_t^\pi = D_t^\pi \psi_t^\pi$. We can also elicite the martingale condition satisfied by Y.

Proposition 2.4.3 *When:*

$$\pi_t(b) - r_t \mathbf{1}_m = \sigma_t \gamma_t^\pi + \omega_t D_t^\pi \left(\mathbf{1}_m - \psi_t^\pi\right) \tag{2.34}$$

the process Y is a $(Q, \mathcal{G})-$local martingale.

Value Functions

Thereafter, we note $I(\cdot) = (U^{'}(\cdot))^{-1}$, $\beta_{\cdot} = \exp\left(-\int_0^{\cdot} r_s ds\right)$ and $Q^y \equiv Q$. The following theorem is adapted from Owen (2002) [135].

Theorem 2.4.4 *Let U be a utility function satisfying (4.6) and (4.7). Then:*

1. *There exists a unique measure Q^y solution of the dual problem (2.31),*
2. *There exists a unique number $\widehat{y}$ s.t. $E^{Q^{\widehat{y}}}\left[\beta_T I\left(\widehat{y}\beta_T \Lambda_T^{Q^{\widehat{y}}}\right)\right] = x_0$,*
3. *The optimal terminal wealth is given by $\widehat{X}_T^\phi = I\left(\widehat{y}\beta_T \Lambda_T^{Q^{\widehat{y}}}\right)$,*
4. *The optimal investment policy process is uniquely determined by:*

$$\beta_t \widehat{X}_t^\phi = E^{Q^{\widehat{y}}}\left[\beta_T \widehat{X}_T^\phi | \mathcal{G}_t\right] = x_0 + \int_0^t \beta_s \widehat{\phi}_s^\intercal dR_s \tag{2.35}$$

Proof. The two first points follow from a classical duality criterion for optimality, in the spirit of Kramkov and Schachermayer (1999) [107]. Then, we note that under $Q^{\widehat{y}}$, the wealth process (2.29) satisfies the equation:

$$d(\beta_t X_t^\phi) = \beta_t \phi_t^\intercal \sigma_t d\widehat{W}_t + \beta_t \phi_t^\intercal \omega_t d\widehat{M}_t \tag{2.36}$$

Note that, as $\beta_t X_t^\phi$ is a $\left(Q^{\widehat{y}}, \mathcal{G}\right)-$martingale, Lemma 2.4.1 entails that there exist $\mathbb{R}^m-$valued and square integrable processes (A, B) s.t.:

$$\beta_t X_t^\phi = x_0 + \int_0^t A_s^\intercal d\widehat{W}_s + \int_0^t B_s^\intercal d\widehat{M}_s \tag{2.37}$$

As (2.36) and (2.37) are equal, their difference:

$$\int_0^t \left[\phi_s^\intercal \sigma_s - A_s^\intercal\right] d\widehat{W}_s + \int_0^t \left[\phi_s^\intercal \omega_s - A_s^\intercal\right] d\widehat{M}_s$$

must be null, thus yielding the self-financing condition $\alpha_s^\intercal = \phi_s^\intercal \sigma_s,\ \beta_s^\intercal = \phi_s^\intercal \omega_s$ and so the desired form for $\beta_t \widehat{X}_t^\phi$ under the measure $Q^{\widehat{y}}$. ■

In the following, to lighten the notation, we let $Q^{\widehat{y}} \equiv Q$.

Special Cases

Problem (2.30) will now be solved in the case of the three more standard utility functions: power, logarithmic and exponential, defined by:

$$U(x) = \begin{cases} x^p/p & x \in \mathbb{R}^+, p \in (0,1) \\ \log x\ , & x \in \mathbb{R}^+ \\ -e^{-x} & x \in \mathbb{R} \end{cases} \tag{2.38}$$

These utility functions are of particular interest as, while in general, a solution to (2.31) depends on y. for them, this dependence vanishes. In fact:

$$V(y) = \begin{cases} -y^q/q & y \in \mathbb{R}, q = \frac{p}{p-1} \\ -(1+\ln y)\ , & y \in \mathbb{R}^+ \\ y(\ln y - 1) & y \in \mathbb{R}^+ \end{cases}$$

so that (2.31) reads:

$$v(y) = \begin{cases} y^q/q \inf_{Q\in\mathcal{Q}} E\left[-\left(\frac{dQ}{dP}\right)^q\right] \\ -1 - \ln y + \inf_{Q\in\mathcal{Q}} E\left[-\ln \frac{dQ}{dP}\right] \\ y\ln y + y \inf_{Q\in\mathcal{Q}} E\left[\frac{dQ}{dP}\ln\frac{dQ}{dP}\right] \end{cases} \tag{2.39}$$

The quantity appearing in the r.h.s. of (2.39) under the expectation term can be called martingale distance measures, cf. Goll and Ruschendorf (2001) [78].

For the case of the utility functions (2.38), we then have the useful martingale distance decomposition results[7], which are related to problem (2.39).

Proposition 2.4.5 *Case of power utility:*

$$E\left[-\Lambda_t^q\right] = E\left[\frac{q(1-q)}{2}\int_0^t \gamma_s^\intercal \gamma_s ds + \left(\left(\psi_s^\intercal\right)^q - \mathbf{1}_m^\intercal - q\left(\psi_s^\intercal - \mathbf{1}_m^\intercal\right)\right)\lambda_t^\pi dt\right] \tag{2.40}$$

[7] Proofs of these decompositions are developed in Appendix A.2.

Case of logarithmic utility:

$$E\left[-\ln\Lambda_T\right] = E\left[\frac{1}{2}\int_0^t \gamma_s^{\intercal}\gamma_s ds + \left(\psi_s^{\intercal} - \ln\psi_s^{\intercal} - \mathbf{1}_m^{\intercal}\right)\lambda_t^{\pi} dt\right] \tag{2.41}$$

Case of exponential utility:

$$E\left[\Lambda_T\ln\Lambda_T\right] = E^Q\left[\frac{1}{2}\int_0^t \gamma_s^{\intercal}\gamma_s ds + \left(\psi_s^{\intercal}\ln\psi_s^{\intercal} - \psi_s^{\intercal} + \mathbf{1}_m^{\intercal}\right)\lambda_t^{\pi} dt\right] \tag{2.42}$$

where $\gamma \equiv \gamma^{\pi}$, $\psi \equiv \psi^{\pi}$ and $\lambda_t^{\pi} \equiv \pi_t(\lambda)$.

Resorting to Theorem 2.4.4 and Proposition 2.4.5, we can now determine the optimal value functions in the case of our utility functions (2.38).

Proposition 2.4.6 *The "power" optimal value function is given by:*

$$\begin{aligned} u(T,x) = \frac{x^p}{p}E^P\Bigg[&\int_0^T \left(r_t + \frac{1}{2}p(p-1)\left(\gamma_t^{\pi}\right)^{\intercal}\gamma_t\right)dt \\ &- \int_0^T \left(\left(\psi_t^{\intercal}\right)^{-p} - \mathbf{1}_t^{\intercal} + p\left(\psi_t^{\intercal} - \mathbf{1}_n^{\intercal}\right)\right)\pi_t(\lambda)\,dt\Bigg] \end{aligned}$$

The "logarithmic" optimal value function reads:

$$\begin{aligned} u(T,x) = \ln x + E^P\Bigg[&\int_0^T \left(r_t + \frac{1}{2}\left(\gamma_t^{\pi}\right)^{\intercal}\gamma_t\right)dt \\ &+ \int_0^T \left(\psi_t^{\intercal} - \ln\psi_t^{\intercal} - \mathbf{1}_t^{\intercal}\right)\pi_t(\lambda)\,dt\Bigg] \end{aligned}$$

The "exponential" optimal value function writes:

$$\begin{aligned} u(T,x) = -e^{-x}\exp E^Q\Bigg[&\int_0^T \left(r_t + \frac{1}{2}\left(\gamma_t^{\pi}\right)^{\intercal}\gamma_t\right)dt \\ &+ \int_0^T \left(\psi_t^{\intercal}\ln\psi_t^{\intercal} - \psi_t^{\intercal} + \mathbf{1}_n^{\intercal}\right)\pi_t(\lambda)\,dt\Bigg]\Bigg\} \end{aligned}$$

From these, we note that in the complete observation case, i.e.: when we know b and λ, the optimal value functions write similarly. The filtering equivalence principle holds: the unknown drift b_t and jump intensity λ_t are replaced by the estimates $\pi_t(b)$ and $\pi_t(\lambda)$ in the optimal value functions.

Remark 2.4.7 *In order to be amenable to computation, the optimal value functions exhibited in Proposition 2.4.6 require that the risk-premia processes are elicited. In Appendix A.2.1, we present them for the utilities (2.38).*

2.4.2 Optimal Trading Strategies

Bellman Equation

For the utilities (2.38), we note that an attractive factorization of $J_\phi(t,x)$ is available. For a portfolio strategy $\phi \in \mathcal{S}(t)$, one has:

$$J_\phi(t,x) = \begin{cases} (x^p/p)\,\eta_\phi^{\text{pow}}(t) & \text{power} \\ \log x + \eta_\phi^{\text{log}}(t) \quad , & \text{logarithmic} \\ -e^{-x}\eta_\phi^{\text{exp}}(t) & \text{exponential} \end{cases} \tag{2.43}$$

where the functions η_ϕ does not depend on x and reads:

$$\eta_\phi^{\text{pow}}(t) = E^P\left[\exp\left\{p\int_0^t d_\pi^\phi(s)\,ds + p\int_0^t \left(\phi_s^\intercal \sigma_s d\overline{W}_s + \ln\left(\mathbf{1}_m + \phi_s^\intercal\omega_s\right) dN_s\right)\right\}\right]$$

$$\eta_\phi^{\text{log}}(t) = E^P\left[\int_0^t d_\pi^\phi(s)\,ds - \int_0^t \ln\left(\mathbf{1}_m + \phi_s^\intercal\omega_s\right)\pi_s(\lambda)\,ds\right]$$

$$\eta_\phi^{\text{exp}}(t) = E^P\left[\exp\left\{\exp\left\{\int_0^t d_\pi^\phi(s)\,ds + \int_0^t \left(\phi_s^\intercal \sigma_s d\overline{W}_s + \ln\left(\mathbf{1}_m + \phi_s^\intercal\omega_s\right) dN_s\right)\right\}\right\}\right]$$

where:

$$d_\pi^\phi(s) \triangleq r_s + \phi_s^\intercal\left(\pi_s(b) - r_s\mathbf{1}_m\right) - \frac{1}{2}\phi_s^\intercal\sigma_s\sigma_s^\intercal\phi_s$$

If we define:

$$\eta^\bullet(t) \triangleq \sup_{\delta\in\mathcal{S}(t)} \eta_\phi^\bullet(t)$$

then we have:

$$u(t,x) = \begin{cases} (x^p/p)\,\eta^{\text{pow}}(t) & \text{power} \\ \log x + \eta^{\text{log}}(t) \quad , & \text{logarithmic} \\ -e^{-x}\eta^{\text{exp}}(t) & \text{exponential} \end{cases}$$

The Bellman equation then writes:

Definition 2.4.8 *Let $\tau \in [t,T]$ be a $\mathcal{G}-$measurable stopping time. Then:*

$$u(t,x) = \sup_{\phi\in\mathcal{S}(t)} E^P\left[u(\tau, X_\tau^{x,\phi})\right]$$

Logarithmic Utility This case is always the easiest one and can be solved in a relatively direct way. We collect in next lemma the main results.

Lemma 2.4.9 **(a)** *For all $t \in [0,T]$, $x > 0$, we have:*

$$u(t,x) = \log x + \eta^{log}(t)$$

where $\eta^{log}(t) = \sup_{\phi\in\mathcal{S}(t)} \eta_\phi^{log}(t)$.

(b) *Suppose that δ^* maximizes:*

$$\delta \mapsto \int_0^t \left(r_s + \delta^\intercal \left(\pi_s(b) - r_s \mathbf{1}_m\right) - \frac{1}{2}\delta^\intercal \sigma_s \sigma_s^\intercal \delta - \ln\left(\mathbf{1}_m + \delta^\intercal \omega_s\right) \pi_s(\lambda) \right) ds$$

on $[0,1]^{\otimes m}$, then $\phi^* = (\phi_t^*) \in S(t)$ is an optimal portfolio strategy for the given portfolio problem under logarithmic preferences.

Proof. By definition of $u(t,x)$, cf. (2.30), and expression of $\eta_\phi^{\log}(t)$, we deduce point (a). Point (b) then follows directly. ■

From this, we note that in the case of complete observation, i.e.: when we know b and λ, the optimal portfolio would be to invest a fraction δ^* of the wealth in the stocks, where δ^* is the maximizer of:

$$\delta \mapsto \int_0^t \left(r_s + \delta^\intercal \left(b_s - r_s \mathbf{1}_m\right) - \frac{1}{2}\delta^\intercal \sigma_s \sigma_s^\intercal \delta - \ln\left(\mathbf{1}_m + \delta^\intercal \omega_s\right) \lambda_s \right) ds$$

on $[0,1]^{\otimes m}$. The certainty equivalence principle holds, i.e.: the unknown drift b_t and intensity λ_t are replaced by the estimates $\pi_t(b)$ and $\pi_t(\lambda)$ in the optimal portfolio strategy. This means, in this case, that the uncertainty in the drift and jump intensity terms does not change the optimal portfolio strategy.

CRRA & CARA Utilities As a first step, we derive a semimartingale representation of $u(\cdot)$. We suppose that u is sufficiently differentiable, so that its derivatives w.r.t. X exist. We also introduce the operator $\mathcal{A}^{\mathrm{x}}$ where x={pow,exp} which acts on function $v : [0,T] \to \mathbb{R}$ and $\delta \in [0,1]^{\otimes m}$ by:

$$\begin{aligned} \mathcal{A}^{\mathrm{pow}} v_\delta(t) = v(t) &\left(r_t + \delta^\intercal \left(\pi_t(b) - r_t \mathbf{1}_m\right) + \frac{1}{2}(p-1)\,\delta^\intercal \sigma_s \sigma_s^\intercal \delta \right) \\ &+ \frac{1}{p}\left(v(t_-)\left(1 + \delta^\intercal \omega_t\right)^p - v(t)\right) \pi_t(\lambda) \end{aligned}$$

$$\begin{aligned} \mathcal{A}^{\exp} v_\delta(t) = v(t) &\left(r_t + \delta^\intercal \left(\pi_t(b) - \mathbf{1}_m\right) - \frac{1}{2}\delta^\intercal \sigma_s \sigma_s^\intercal \delta \right) \\ &- \left(v(t_-)\, e^{-(1+\delta^\intercal \omega_t)} - v(t) \right) \pi_t(\lambda) \end{aligned}$$

and we define the operator $\mathcal{H}^{\mathrm{x}}$ for x={pow,exp} as:

$$\mathcal{H}^{\mathrm{x}} v_\delta(t) \triangleq c^{\mathrm{x}} \frac{\partial}{\partial t} v_\delta(t) + \mathcal{A}^{\mathrm{x}} v_\delta(t)\,,\ c^{\mathrm{x}} = \begin{cases} \frac{1}{p} & \text{pow} \\ -1 & \exp \end{cases}$$

Lemma 2.4.10 *Let $\phi \in \mathcal{S}(t)$ be an arbitrary portfolio strategy. Then:*

$$du(t, X_t^{\phi}) = \mathcal{H}^x \eta_\delta^x(t)\, dt + dM_t^{x,\phi}$$

where $M_t^{x,\phi}$ is a $(P, \mathcal{G}_t)-$martingale.

Proof. cf. Appendix A.3. ■

From this, the HJB equation, for x={pow,exp}, reads:

$$0 = c^{\mathrm{x}} \frac{\partial}{\partial t} \eta_\delta^{\mathrm{x}}(t) + \sup_{\delta \in [0,1]} \mathcal{A}^{\mathrm{x}} \eta_\delta^{\mathrm{x}}(t)$$

with boundary condition $\eta_\delta^{\mathrm{x}}(T) = 1$. Then, we present a verification theorem.

Theorem 2.4.11 *Suppose that δ^* maximizes:*

$$\delta \mapsto \mathcal{A}^x \eta_\delta^x(t)$$

on $[0,1]^{\otimes m}$. Then, $\phi^* = (\phi_t^*) \in S(t)$ is an optimal feedback portfolio strategy for the given portfolio problem for CRRA and CARA utilities.

Proof. Let $\phi \in \mathcal{S}(t)$ and J_ϕ be given by (2.43). From Lemma 2.4.10:

$$J_\phi(T, X_T) = J_\phi(t, X_t) + \int_t^T \mathcal{H}^{\mathrm{x}} \eta_{\phi_s}^{\mathrm{x}}(s)\, ds + M_T^{\mathrm{x},\phi} - M_t^{\mathrm{x},\phi}$$

As η satisfies the HJB equation, we get $\mathcal{H}^{\mathrm{x}} \eta_{\phi_s}^{\mathrm{x}}(s) \leqslant 0$, so that:

$$\int_t^T \mathcal{H}^{\mathrm{x}} \eta_{\phi_s}^{\mathrm{x}}(s)\, ds \leqslant 0$$

thus:

$$J_\phi(T, X_T) \leqslant J_\phi(t, X_t) + M_T^{\mathrm{x},\phi} - M_t^{\mathrm{x},\phi} \tag{2.44}$$

Taking expectation in (2.44), and using the definition of J_ϕ at T, yields:

$$E^P\left[U(X_T^{\phi})\right] \leqslant J_\phi(t, x)$$

Then, taking supremum over all admissible strategies gives $u(t,x) \leqslant J_\phi(t,x)$. If ϕ is chosen to be ϕ^* solution of the HJB equation, then $\int_t^T \mathcal{H}^{\mathrm{x}} \eta_{\phi_s^*}^{\mathrm{x}}(s)\, ds = 0$ and so $u(t,x) = J_{\phi^*}(t,x)$, which concludes the proof. ■

The next theorem states the existence of a solution of the HJB equation.

Theorem 2.4.12 *In case of CRRA and CARA utilities, the value functions of problem (2.30) are given by (2.43) where η satisfies the HJB equation:*

$$0 = c^x \eta^x_\delta (t) + \sup_{\delta\in[0,1]} \mathcal{A}^x \eta^x_\delta (t)\,, \;\; x\text{=}\{pow,exp\}$$

with boundary condition $\eta^x_\delta (T) = 1$. Moreover, $\phi^*_t = \delta^*$ with δ^* given by Theorem 2.4.11, is an optimal portfolio strategy.

Proof. For $t \leqslant t^* \leqslant T$, from Definition 2.4.8, we have:

$$u\,(t,x) \geqslant E^P \left[u(\tau \wedge t^*, X^\phi_{\tau\wedge t^*}, \right] \tag{2.45}$$

From Lemma 2.4.10, we get:

$$u\,(\tau \wedge t^*, X_{\tau\wedge t^*}) = u\,(t,x) + \int_t^{\tau\wedge t^*} \mathcal{H}^{\text{x}} \eta^{\text{x}}_{\phi_s} (s)\, ds + M^{\text{x},\phi}_{\tau\wedge t^*} - M^{\text{x},\phi}_t$$

Then, replacement in (2.45) yields:

$$E^P \left[\int_t^{\tau\wedge t^*} \mathcal{H}^{\text{x}} \eta^{\text{x}}_{\phi_s} (s)\, ds \right] \leqslant 0$$

or equivalently $\mathcal{H}^{\text{x}} \eta^{\text{x}}_{\phi^*} (s) \leqslant 0$ for any ϕ^* where $\delta \in [0,1]$. As η is continuous, we obtain:

$$c^{\text{x}} \eta^{\text{x}}_\delta (t) + \mathcal{A}^{\text{x}} \eta^{\text{x}}_\delta (s) \leqslant 0$$

Finally, as δ is arbitrary, we can write:

$$c^{\text{x}} \eta^{\text{x}}_\delta (t) + \sup_{\delta\in[0,1]} \mathcal{A}^{\text{x}} \eta^{\text{x}}_\delta (s) \leqslant 0 \tag{2.46}$$

On the other hand, for $\varepsilon > 0$, it may exist a strategy ϕ^ε s.t.:

$$u\,(t,x) - \varepsilon\,(t^* - t) \leqslant E^P \left[u(\tau \wedge t^*, X^\phi_{\tau\wedge t^*}) \right]$$

Again, from Lemma 2.4.10, we get:

$$-\varepsilon\,(t^* - t) \leqslant E^P \left[\int_t^{\tau\wedge t^*} \mathcal{H}^{\text{x}} \eta^{\text{x}}_{\phi^\varepsilon_s} (s)\, ds \right]$$

or:

$$-\varepsilon \leqslant E^P \left[\frac{1}{t^* - t} \int_t^{\tau\wedge t^*} \mathcal{H}^{\text{x}} \eta^{\text{x}}_{\phi^\varepsilon_s} (s)\, ds \right] \leqslant E^P \left[\frac{1}{t^* - t} \int_t^{\tau\wedge t^*} \sup_{\delta\in[0,1]} \mathcal{H}^{\text{x}} \eta^{\text{x}}_\delta (s)\, ds \right]$$

Since the function η is continuous and the constant ε is arbitrary, we deduce $\sup_{\delta\in[0,1]} \mathcal{H}^{\text{x}} \eta^{\text{x}}_\delta (s) \geqslant 0$, so that:

$$c^{\text{x}} \eta^{\text{x}}_\delta (s) + \sup_{\delta\in[0,1]} \mathcal{A}^{\text{x}} \eta^{\text{x}}_\delta (s) \geqslant 0$$

which in conjunction with (2.46) concludes the proof. ■

2.5. Conclusion

In this article, we have investigated the question of optimal policies in a multi-dimensional jump-diffusion model of incomplete market under the setup of partial information. When the model is only partially observable, we have extended the framework of Lakner (1998) [111] by allowing learning in the intensity of the Poisson process. Moreover, resorting to the martingale approach and the theory of optimal control, we are able to derive the Hamilton-Jacobi-Bellman equation and then to identify an optimal investement strategy. The usefulness of this approach has been proved by computing the optimal investment strategy in a discontinuous jump-diffusion model, extending previous results which only deal with the continuous case.

Apart from the problem of optimal policies, another application of this work are the problems of hedging in incomplete markets and of hedging default risk when default is modelled in the reduced form setting and the market model is incomplete. These will be studied in subsequent researches.

A

Appendix (Chapter 2)

Agenda

A.1. Non-Linear Filtering

In this Appendix, we present a particular characterization of the unknown economic factor process θ. Under the probability measure P, it satisfies:

$$d\theta_t = m(t, \theta_t)\, dt + v(t, \theta_t)\, dB_t + \int_\Gamma w(t, \theta_{t-}, z)\, \mathcal{N}(dt, dz)$$

where m, v and w are bounded and continuous functions on $\mathbb{R}^+ \times \mathbb{R}$ and $\mathbb{R}^+ \times \mathbb{R} \times \Gamma$, B is a standard Brownian motion, $\mathcal{N}$ is a Poisson random measure, independent of B, with mean rate $E[\mathcal{N}(t+\Delta, A)] - E[\mathcal{N}(t, A)] = \varrho \Delta \nu(A)$, where ϱ is a real positive number (jump intensity) and ν is a probability measure (jump amplitude) on the space of jumps Γ. Also, B and $\mathcal{N}$ are independent of W and N, which appear in the dynamics of Y, cf. (2.3). For any twice continuously differentiable function f, Itô's formula yields:

$$\begin{aligned} f(\theta_t) = f(\theta_0) &+ \int_0^t \mathcal{L} f(\theta_s)\, ds + \int_0^t f^{'}(\theta_s)\, v(s, \theta_s)\, dB_s \\ &+ \int_0^t \mathcal{J} f(\theta_s)\, ds + \sum_{s \leqslant t} [f(\theta_s) - f(\theta_{s-})] \end{aligned} \tag{A.1}$$

where the differential operator $\mathcal{L}$ is given by:

$$\mathcal{L}f(x) = m(t,x) f^{'}(x) + \frac{1}{2} v^2(t,x) f^{''}(x)$$

and:

$$\mathcal{J}f(x) = -\varrho \int_{\Gamma} [f(x + w(s,x,z)) - f(x)] \nu(dz)$$

Special attention will also be given to the unnormalized filter:

$$\Pi_t(f) = E^{P^0}\left[f(\theta_t) H_t^{-1} | \mathcal{G}_t\right] \tag{A.2}$$

for all $t \in [0,T]$, where the probability measure P^0 and the $P-$martingale H are defined via (2.10) and (2.12). Then, with the help of the conditional Bayes formula, we explicit the relationship between π_t and Π_t as:

$$\pi_t(f) = \frac{E^{P^0}\left[f(\theta_t) H_t^{-1} | \mathcal{G}_t\right]}{E^{P^0}\left[H_t^{-1} | \mathcal{G}_t\right]} = \frac{\Pi_t(f)}{\Pi_t(1)} \tag{A.3}$$

A.1.1 Preliminary Results

Lemma A.1.1 *For all $t \in [0,T]$, we have:*

$$E^{P^0}\left[\int_0^t f(\theta_{s-}) H_s^{-1}\left[a_s^{\intercal}(\theta_s) dW_s^0 + \left((c_s^{\intercal}(\theta_s))^* - \mathbf{1}_m^{\intercal}\right) dM_s^0\right] | \mathcal{G}_t\right]$$
$$= \int_0^t \Pi_{s-}(af) dW_s^0 + \int_0^t \Pi_{s-}((c^* - \mathbf{1}_m) f) dM_s^0 \tag{A.4}$$

Proof. In order to prove (2.18), we proceed as in Theorem 2.3.5. Letting:

$$L_t = \int_0^t f(\theta_{s-}) H_s^{-1}\left[a_s^{\intercal}(\theta_s) dW_s^0 + \left((c_s^{\intercal}(\theta_s))^* - \mathbf{1}_m^{\intercal}\right) dM_s^0\right]$$
$$R_t = \int_0^t \Pi_{s-}(af) dW_s^0 + \int_0^t \Pi_{s-}((c^* - \mathbf{1}_m) f) dM_s^0$$

so that:

$$L_t \triangleq \int_0^t f(\theta_{s-}) H_s^{-1} d\nu_s^0$$

and noting that R is $\mathcal{G}-$adapted, it is sufficient to show that:

$$E^{P^0}[L_t \mathbf{1}_A] = E^{P^0}[R_t \mathbf{1}_A] \tag{A.5}$$

for any $A \in \mathcal{G}_t$, so that $R_t = E^{P^0}[L_t | \mathcal{G}_t]$. Via the Martingale Representation theorem for $\mathcal{G}-$martingales $\mathbf{1}_A$ is, for $t \geqslant 0$, of the form:

$$\mathbf{1}_A = M_0 + \int_0^t U_s dW_s^0 + \int_0^t V_s dM_s^0 \triangleq M_0 + \overline{M}_t^0$$

where U, V are $\mathcal{G}-$predictable processes and $M_0 \in \mathcal{G}_0$ is not necessary a constant. Hence, (A.5) reads:

$$E^{P^0}\left[L_t M_0\right] + E^{P^0}\left[L_t \overline{M}_t^0\right] = E^{P^0}\left[R_t M_0\right] + E^{P^0}\left[R_t \overline{M}_t^0\right]$$

so that by using the definition of L and R, it is sufficient to prove:

$$E^{P^0}\left[\int_0^t f\left(\theta_{s-}\right) H_s^{-1} d\nu_s^0 \cdot M_0\right] = E^{P^0}\left[R_t^0 M_0\right] \tag{A.6}$$

$$E^{P^0}\left[\int_0^t f\left(\theta_{s-}\right) H_s^{-1} d\nu_s^0 \cdot M_t^0\right] = E^{P^0}\left[R_t^0 \overline{M}_t^0\right] \tag{A.7}$$

with:

$$d\nu_s^0 = a_s^\intercal\left(\theta_s\right) dW_s^0 + \left(\left(c_s^\intercal\left(\theta_s\right)\right)^* - \mathbf{1}_m^\intercal\right) dM_s^0$$

and:

$$\begin{aligned}R_t^0 = &\int_0^t E^{P^0}\left[f\left(\theta_{s-}\right) H_{s-}^{-1} a_s^\intercal\left(\theta_s\right) | \mathcal{G}_s\right] dW_s^0 \\ &+ \int_0^t E^{P^0}\left[f\left(\theta_{s-}\right) H_{s-}^{-1}\left(\left(c_s^\intercal\left(\theta_s\right)\right)^* - \mathbf{1}_m^\intercal\right) | \mathcal{G}_s\right] dM_s^0\end{aligned}$$

As $\overline{M}_t^0 = \int_0^t U_s dW_s^0 + \int_0^t V_s dM_s^0$ and that the intensity of M^0 is 1, (A.7) reads:

$$\begin{aligned}&E^{P^0}\left[\int_0^t \left(f\left(\theta_{s-}\right) H_{s-}^{-1} a_s^\intercal U_s + f\left(\theta_{s-}\right) H_{s-}^{-1}\left(\left(c_s^\intercal\right)^* - \mathbf{1}_m^\intercal\right) V_s\right) ds\right] \\ &= E^{P^0}\left[\int_0^t E^{P^0}\left[f\left(\theta_{s-}\right) H_{s-}^{-1} U_s a_s^\intercal | \mathcal{G}_s\right] ds\right. \\ &\left.+ \int_0^t E^{P^0}\left[f\left(\theta_{s-}\right) H_{s-}^{-1} V_s\left(\left(c_s^\intercal\right)^* - \mathbf{1}_m^\intercal\right) | \mathcal{G}_s\right] ds\right]\end{aligned}$$

(covariation terms between W^0 and M^0 being null), this equality holds true if:

$$E^{P^0}\left[\int_0^t E^{P^0}\left[f\left(\theta_{s-}\right) H_{s-}^{-1} U_s a_s^\intercal | \mathcal{G}_s\right] ds\right] = E^{P^0}\left[\int_0^t f\left(\theta_{s-}\right) H_{s-}^{-1} U_s a_s^\intercal ds\right]$$

and:

$$\begin{aligned}&E^{P^0}\left[\int_0^t E^{P^0}\left[f\left(\theta_{s-}\right) H_{s-}^{-1} V_s\left(\left(c_s^\intercal\right)^* - \mathbf{1}_m^\intercal\right) | \mathcal{G}_s\right] ds\right] \\ &= E^{P^0}\left[\int_0^t f\left(\theta_{s-}\right) H_{s-}^{-1} V_s\left(\left(c_s^\intercal\right)^* - \mathbf{1}_m^\intercal\right) ds\right]\end{aligned}$$

These last equalities are obtained by definition of conditional expectation, so that (A.7) holds true. Also, in (A.6), we note that $M_0 \in \mathcal{G}_0$ and similar arguments yield:

$$\begin{aligned}
E^{P^0}\left[R_t^0 M_0\right] &= E^{P^0}\left[\int_0^t E^{P^0}\left[f\left(\theta_{s-}\right) H_{s-}^{-1} a_s^{\intercal}\left(\theta_s\right) M_0 | \mathcal{G}_s\right] dW_s^0\right. \\
&\quad \left. + \int_0^t E^{P^0}\left[f\left(\theta_{s-}\right) H_{s-}^{-1}\left(\left(c_s^{\intercal}\left(\theta_s\right)\right)^* - \mathbf{1}_m^{\intercal}\right) M_0 | \mathcal{G}_s\right] dM_s^0\right] \\
&= E^{P^0}\left[\int_0^t f\left(\theta_{s-}\right) H_{s-}^{-1} a_s^{\intercal}\left(\theta_s\right) M_0 dW_s^0\right. \\
&\quad \left. + \int_0^t f\left(\theta_{s-}\right) H_{s-}^{-1}\left(\left(c_s^{\intercal}\left(\theta_s\right)\right)^* - \mathbf{1}_m^{\intercal}\right) dM_s^0\right]
\end{aligned}$$

which concludes the proof. ∎

A.1.2 Filtering Equation

The next theorem characterizes Π as the solution of a mixed SDE in a Kallianpur-Striebel type formula, thus augmenting Pardoux (1989).

Theorem A.1.2 *The unnormalized filter $\Pi_t = \left(\Pi_t^1, ..., \Pi_t^m\right)^{\intercal}$ satisfies:*

$$\begin{aligned}
\Pi_t(f) &= \Pi_0(f) + \int_0^t \Pi_s(\mathcal{L}f)\, ds - \int_0^t \Pi_s\left(\varrho \int_\Gamma \left(f\left(\cdot + w(\cdot, z)\right) - f(\cdot)\right) \nu(dz)\right) ds \\
&\quad + \int_0^t \Pi_s(af)\, dW_s^0 + \int_0^t \Pi_{s-}\left(\left(c^* - \mathbf{1}_m\right) f\right) dM_s^0 \qquad \text{(A.8)}
\end{aligned}$$

or in terms of the observations Y:

$$\begin{aligned}
\Pi_t(f) &= \Pi_0(f) + \int_0^t \Pi_s(\mathcal{L}f)\, ds - \int_0^t \Pi_s\left(\varrho \int_\Gamma \left(f\left(\cdot + w(\cdot, z)\right) - f(\cdot)\right) \nu(dz)\right) ds \\
&\quad + \int_0^t \Pi_s(af)\, \sigma_s^{-1} dY_s + \int_0^t \left(\Pi_{s-}\left(\left(c^* - \mathbf{1}_m\right) f\right) - \Pi_{s-}(af)\, \sigma_s^{-1} \omega_s\right) \left(dN_s - ds\right)
\end{aligned}$$

Proof. First, we determine the form of $f(\theta_t) H_t^{-1}$, cf. (2.17) and (A.1). By the Itô's product rule for semimartingales, we have:

$$\begin{aligned}
f(\theta_t) H_t^{-1} &= \int_0^t f(\theta_{s-}) \, dH_s^{-1} + \int_0^t H_{s-}^{-1} df(\theta_s) + \left[f(\theta), H^{-1}\right]_t \\
&= \int_0^t f(\theta_{s-}) H_s^{-1} \left[a_s^{\intercal}(\theta_s) \, dW_s^0 + \left((c_s^{\intercal}(\theta_s))^* - \mathbf{1}_m^{\intercal}\right) dM_s^0\right] \\
&+ \int_0^t \mathcal{L} f(\theta_s) H_{s-}^{-1} ds + \int_0^t f^{'}(\theta_s) \, v(s, \theta_s) H_s^{-1} dB_s \\
&+ \int_0^t H_s^{-1} \mathcal{J}_f(s) \, ds + \int_0^t H_s^{-1} d\Big(\sum_{s \leqslant t} [f(\theta_s) - f(\theta_{s-})]\Big) \\
&+ \left[f(\theta), H^{-1}\right]_t
\end{aligned}$$

From the definition of $\mathcal{J}_f(t)$, we get:

$$\int_0^t H_s^{-1} d\mathcal{J}_f(s) = -\int_0^t H_s^{-1} \varrho \int_{\Gamma} [f(\theta_s + w(s, \theta_s, z)) - f(\theta_s)] \, \nu(dz) \, ds$$

and:

$$\begin{aligned}
\left[f(\theta), H^{-1}\right]_t &= f(\theta_0) H_0^{-1} + \sum_{s \leqslant t} \Delta f(\theta_s) \cdot \Delta H_s^{-1} \\
&\quad + \int_0^t f^{'}(\theta_s) \, v(s, \theta_s) H_s^{-1} d\langle B, W\rangle_s = f(\theta_0)
\end{aligned}$$

Then, it follows:

$$\begin{aligned}
f(\theta_t) H_t^{-1} = f(\theta_0) &+ \int_0^t f(\theta_{s-}) H_s^{-1} \left[a_s^{\intercal}(\theta_s) \, dW_s^0 + \left((c_s^{\intercal}(\theta_s))^* - \mathbf{1}_m^{\intercal}\right) dM_s^0\right] \\
&+ \int_0^t \mathcal{L} f(\theta_s) H_{s-}^{-1} ds + \int_0^t f^{'}(\theta_s) \, v(s, \theta_s) H_s^{-1} dB_s + \int_0^t H_s^{-1} d\mathcal{J}_f(t)
\end{aligned} \tag{A.9}$$

We are now interested in computing the five conditional expectations w.r.t. $(P^0, \mathcal{G})$ on the right-hand side of (A.9). We compute them in their order of appearance. Obviously, the first one yields: $\Pi_0(f) = E^{P^0}[f(\theta_0) | \mathcal{G}_t]$. For the second one, thanks to Lemma A.1.1, we have:

$$\begin{aligned}
&E^{P^0}\left[\int_0^t f(\theta_{s-}) H_s^{-1} \left[a_s^{\intercal}(\theta_s) \, dW_s^0 + \left((c_s^{\intercal}(\theta_s))^* - \mathbf{1}_m^{\intercal}\right) dM_s^0\right] | \mathcal{G}_t\right] \\
&= \int_0^t \Pi_{s-}(af) \, dW_s^0 + \int_0^t \Pi_{s-}((c^* - \mathbf{1}_m) f) \, dM_s^0
\end{aligned}$$

Similarly, we get:

$$E^{P^0}\left[\int_0^t \mathcal{L}f(\theta_s)\, H_{s-}^{-1} ds|\mathcal{G}_t\right] = \int_0^t \Pi_{s-}(\mathcal{L}f)\, ds$$

Under P^0, the process θ_t is independent of $\mathcal{G}_t$ and so for B_t, yielding:

$$E^{P^0}\left[\int_0^t f^{'}(\theta_s)\, v(s,\theta_s)\, H_s^{-1} dB_s|\mathcal{G}_t\right] = 0$$

Finally, we derive the last term:

$$E^{P^0}\left[\int_0^t H_s^{-1} d\mathcal{J}_f(s)\,|\mathcal{G}_t\right] = -\int_0^t \Pi_{s-}\left(\varrho\int_\Gamma (f(\cdot + w(\cdot,z)) - f(\cdot))\,\nu(dz)\right) ds$$

Collecting all these terms yields the desired result. ■

Useful in the following is the next lemma.

Lemma A.1.3 *We have:*

$$\begin{aligned}\Pi_t^{-1}(1) = 1 &- \int_0^t \frac{\Pi_{s-}(a)}{\Pi_{s-}^3(1)}\left(\Pi_{s-}(1)\, dW_s^0 - \Pi_{s-}(a)\, ds\right)\\ &- \int_0^t \frac{\Pi_{s-}(c^*) - 1}{\Pi_{s-}(1)\,\Pi_{s-}(c^*)}\left(dN_s - \frac{\Pi_{s-}(c^*)}{\Pi_{s-}(1)} ds\right)\end{aligned}$$

Proof. From Theorem A.1.2, it follows that:

$$\Pi_t(1) = 1 + \int_0^t \Pi_{s-}(a)\, dW_s^0 + \int_0^t \Pi_{s-}(c^* - \mathbf{1}_m)\, dM_s^0$$

Using Itô's formula for semimartingales, cf. (2.9), we take $f(x) = x^{-1}$ and $x_t = \Pi_t(1)$. For the first term in (2.9), we have $f(x_0) = \Pi_0^{-1}(1) = 1$. For the second, we write $\int_0^t f^{'}(x_{s-})\, dx_s = -\int_0^t \Pi_{s-}^{-2}(1)\,\Pi_{s-}(a)\, dW_s^0 - \int_0^t \Pi_{s-}^{-2}(1)\,\Pi_{s-}(c^* - \mathbf{1}_m)\, dM_s^0$. The third reads as $\frac{1}{2}\int_0^t f^{''}(x_{s-})\, d[x^c, x^c]_s = \int_0^t \Pi_{s-}^{-3}(1)\,\Pi_{s-}^2(a)\, ds$. Finally, for the fourth term, we note that $\Pi_s^{-1}(1) - \Pi_{s-}^{-1}(1) = \Delta\Pi_s^{-1}(1) = -\Pi_{s-}^{-1}(1)\,\frac{\Pi_s(c^*)-1}{\Pi_{s-}(c^*)}\Delta N_s$, so that:

$$\begin{aligned}&\sum_{s\leqslant t}\left(f(x_s) - f(x_{s-}) - f^{'}(x_{s-})\,\Delta x_s\right)\\ &= -\int_0^t \frac{\Pi_{s-}(c^*) - 1}{\Pi_{s-}(1)\,\Pi_s(c^*)} dN_s + \int_0^t \Pi_{s-}^{-2}(\Pi_s(c^*) - 1)\, dN_s\end{aligned}$$

Collecting all these terms yields the desired result. ■

The characterization of π resorts then to Theorem A.1.2 and Lemma A.1.3.

Theorem A.1.4 *The normalized filter* $\pi_t = \left(\pi_t^1, ..., \pi_t^m\right)^\intercal$ *satisfies:*

$$d\pi_t\left(f\right) = \left(\pi_t(\mathcal{L}f) - \pi_{t-}\left(\varrho\int_\Gamma \left(f\left(\cdot + w\left(\cdot, z\right)\right) - f\left(\cdot\right)\right)\nu\left(dz\right)\right)\right)dt \qquad \text{(A.10)}$$
$$+\left(\pi_{t-}\left(af\right) - \pi_{t-}\left(a\right)\pi_{t-}\left(f\right)\right)d\overline{W}_t + \left(\frac{\pi_{t-}\left(c^* f\right)}{\pi_{t-}\left(c^*\right)} - \pi_{t-}\left(f\right)\right)d\overline{M}_t$$

Proof. As $\pi_t\left(f\right) = \Pi_t\left(f\right)\Pi_t^{-1}\left(1\right)$, an application of Itô's formula yields:

$$\pi_t\left(f\right) = \int_0^t \Pi_{s-}\left(f\right)d\Pi_s^{-1}\left(1\right) + \int_0^t \Pi_s^{-1}\left(1\right)d\Pi_s\left(f\right) + \left[\Pi\left(f\right), \Pi^{-1}\left(1\right)\right]_t$$

Then, rearrangements yield the desired result. ■

Related Aspects

From Theorem A.1.4, $\pi_t\left(f\right)$ is a $\mathcal{G}-$semimartingale with paths in the Skorokhod space $\mathbb{D}_{[0,\infty)}\left(\mathbb{R}\right)$, so that $\pi_t\left(f\right)$ is a right continuous process with limits from the left (càdlàg). We then have the next uniqueness result.

Lemma A.1.5 $\pi_t\left(f\right)$ *admits a unique solution with values in* $\mathbb{D}_{[0,\infty)}\left(\mathbb{R}\right)$.

Proof. To prove uniqueness of the normalized filter (A.10), we may follow Kliemann, Koch and Marchetti (1990) [104]. By formulating the filtering problem as a filtered martingale problem, it is equivalent to a problem over the space of measures. Then, we can apply Kurtz and Ocone (1988) [108] results about the unique solution for a martingale problem. Finally, the uniqueness of the filtered martingale problem gives the uniqueness of the filtering equation. ■

As the market is observed in continuous time, it is possible to completely distinguish the continuous and discontinuous parts of the stock price process.

Corollary A.1.6 *For jump time* t_i*, the jumping filter equation reads:*

$$\pi_{t_i}\left(f\right) = \frac{\pi_{t_i-}\left(c^* f\right)}{\pi_{t_i-}\left(c^*\right)} - \sigma_{t_i}^{-1}\pi_{t_i-}\left(\mathcal{B}f\right) \triangleq \pi_{t_i}^J\left(f\right) \qquad \text{(A.11)}$$

For $t \in [t_i, t_{i+1})$*, the diffusing filter equation satisfies:*

$$\pi_t\left(f\right) = \pi_{t_i}(f) + \int_{t_i}^t \pi_{s-}\left(\mathcal{D}f\right)ds + \int_{t_i}^t \sigma_s^{-1}\pi_{s-}\left(\mathcal{B}f\right)dY_s \triangleq \pi_t^D\left(f\right) \qquad \text{(A.12)}$$

where:

$$\pi_{t-}\left(\mathcal{P}f\right) \triangleq \frac{\pi_{t-}\left(c^* f\right)}{\pi_{t-}\left(c^*\right)} - \pi_{t-}\left(f\right)$$

$$\pi_{t-}\left(\mathcal{B}f\right) \triangleq \pi_{t-}\left(af\right) - \pi_{t-}\left(a\right)\pi_{t-}\left(f\right)$$

$$\pi_{t-}\left(\mathcal{D}f\right) \triangleq \pi_t(\mathcal{L}^0 f) - \sigma_t^{-1}\left\{\pi_{t-}^*\left(\mathcal{B}f\right)\left\{\pi_t(\mu) + \pi_t\left(\lambda\right) - r_s\mathbf{1}_m\right\} - \pi_{t-}^*\left(\mathcal{P}f\right)\pi_t\left(\lambda\right)\right\}$$

A.2. Martingale Decompositions

Via the martingale approach of optimization, a solution to the control problem (2.30) resorts to solve the dual (2.31), which is also equivalent to solve:

$$\inf_{Q\in\mathcal{Q}} E\left[f\left(\frac{dQ}{dP}\right)\right],\ f(x)=\begin{cases}-x^q & \text{power}\\ -\ln x & \text{logarithmic}\\ x\ln x & \text{exponential}\end{cases}$$

cf. (2.39).

Power Utility

The $q-$Kakutani-Hellinger distance $f(x)=x^q$ admits the decomposition.

Proposition A.2.1 *For $q=p/(p-1)$ with $p\in(0,1)$:*

$$E\left[-\Lambda_t^q\right]=E\left[\frac{1}{2}q(1-q)\int_0^t\gamma_s^\intercal\gamma_s ds+\left(\left(\psi^\intercal\right)^q-\mathbf{1}_m^\intercal-q\left(\psi^\intercal-\mathbf{1}_m^\intercal\right)\right)\lambda_t dt\right]$$

Proof. The $q-$Kakutani-Hellinger distance is related to the Hellinger process of order q as shown by Jacod and Shiryaev (2003) [93]. The result is then a direct consequence of their Corollary IV.1.37. ∎

Logarithmic Utility

The reverse entropy quantity $f(x)=-\ln x$ admits the decomposition.

Proposition A.2.2 *We have:*

$$E\left[-\ln\Lambda_T\right]=E\left[\frac{1}{2}\int_0^t\gamma_s^\intercal\gamma_s ds+\left(\psi^\intercal-\ln\psi^\intercal-\mathbf{1}_m^\intercal\right)\lambda_t dt\right]$$

Proof. The proof relies on the computation of the canonical decomposition of the $P-$submartingale $\ln\Lambda_T=U_t$. An application of Itôs' formula gives:

$$\ln\Lambda_T=U_t-\frac{1}{2}\left\langle U^c\right\rangle_t+\sum_{s\leqslant t}\left(\ln\left(1+\Delta U_s\right)-\Delta U_s\right)=M_t+A_t+V_t \quad \text{(A.13)}$$

where $M_t=U_t$ is a $P-$local martingale, $A_t=-1/2\left\langle U^c\right\rangle_t$ is continuous and increasing and $V_t=\sum_{s\leqslant t}\left(\ln\left(1+\Delta U_s\right)-\Delta U_s\right)$ is increasing but not predictable. Hence, (A.13) is not the canonical decomposition of $\ln\Lambda_T$. Noting that $\Delta U_s=\left(\psi_s^\intercal-\mathbf{1}_n^\intercal\right)\mathbf{1}_{n\times m}\mathbf{1}_{\{\Delta Y_s\neq 0\}}$, we have:

$$\ln\left(1+\Delta U_s\right)-\Delta U_s=\ln\left(\psi_s^\intercal\right)\mathbf{1}_{\{\Delta Y_s\neq 0\}}-\left(\psi_s^\intercal-\mathbf{1}_m^\intercal\right)\mathbf{1}_{\{\Delta Y_s\neq 0\}}$$

which yields:

$$V_t = \left(\ln \psi^\intercal - \psi^\intercal + \mathbf{1}_m^\intercal\right) \lambda_t$$

or equivalently:

$$\begin{aligned} V_t = & \left(\ln \psi^\intercal - \psi^\intercal + \mathbf{1}_m^\intercal\right) \lambda_t \;\; \{= K_t\} \\ & + \left(\ln \psi^\intercal - \psi^\intercal + \mathbf{1}_m^\intercal\right) * (N_t - \lambda_t) \;\; \{= L_t\} \end{aligned}$$

such that V_t has been decomposed into a $P-$local martingale K_t and a predictable process of finite variation L_t. Hence, it follows:

$$\begin{aligned} \ln \Lambda_T &= (M_t + K_t) + (A_t + L_t) \\ &= (U_t + K_t) + \left(-\frac{1}{2} \left\langle U^c \right\rangle_t + L_t\right) \end{aligned} \tag{A.14}$$

Taking conditional expectation w.r.t. P on both sides of (A.14) and noting that the first parenthesis is a $P-$local martingale yields:

$$\begin{aligned} E\left[-\ln \Lambda_T\right] &= E\left[\frac{1}{2} \left\langle U^c \right\rangle_t - \left(\ln \psi^\intercal - \psi^\intercal + \mathbf{1}_m^\intercal\right) \lambda_t dt\right] \\ &= E\left[\frac{1}{2} \int_0^t \gamma_s^\intercal \gamma_s ds + \left(\psi^\intercal - \ln \psi^\intercal - \mathbf{1}_n^\intercal\right) \lambda_t dt\right] \end{aligned}$$

which concludes the proof. ■

Exponential Utility

The entropic term $f(x) = x \ln x$ can be computed as follows.

Proposition A.2.3 *We have:*

$$E\left[\Lambda_T \ln \Lambda_T\right] = E^Q\left[\frac{1}{2} \int_0^t \gamma_s^\intercal \gamma_s ds + \left(\psi^\intercal \ln \psi^\intercal - \psi^\intercal + \mathbf{1}_m^\intercal\right) \lambda_t dt\right]$$

Proof. As in Proposition A.2.2, the aim is to compute the canonical decomposition of the $P-$submartingale $\Lambda_T \ln \Lambda_T$. An application of the integration by parts formula yields:

$$d\left(\Lambda_T \ln \Lambda_T\right) = \Lambda_{t-} d \ln \Lambda_T + \ln \Lambda_{t-} dH_t + d\left[\Lambda, \ln \Lambda\right]_t$$

where the first and second right-hand side term are computed thanks to:

$$\ln \Lambda_T = U_t - \frac{1}{2} \left\langle U^c \right\rangle_t + D_t$$

with $D_t = \sum_{s\leqslant t}(\ln(1+\Delta U_s) - \Delta U_s)$. For the third right-hand side term, noting that $dH_t = \Lambda_{t-}dU_t$, yields:

$$d[\Lambda, \ln\Lambda]_t = \Lambda_{t-}d[U, \ln\Lambda]_t = \Lambda_{t-}\left(d[U]_t - \frac{1}{2}d[U, \langle U^c\rangle]_t + d[U, D]_t\right)$$

where:

$$[U, D]_t = \sum_{s\leqslant t}\Delta U_s \Delta D_s = \sum_{s\leqslant t}\Delta U_s \ln(1+\Delta U_s) - \sum_{s\leqslant t}(\Delta U_s)^2$$

and as $\langle U^c\rangle$ is continuous so $[U, \langle U\rangle^c]$ vanishes, we obtain:

$$\begin{aligned}[U, \ln\Lambda]_t &= [U]_t - \sum_{s\leqslant t}(\Delta U_s)^2 + \sum_{s\leqslant t}\Delta U_s \ln(1+\Delta U_s)\\ &= \langle U^c\rangle_t + \sum_{s\leqslant t}\Delta U_s \ln(1+\Delta U_s)\end{aligned}$$

which in turn gives:

$$[\Lambda, \ln\Lambda]_t = \int_0^t \Lambda_{s-}d\langle U^c\rangle_s + \sum_{s\leqslant t}\Lambda_{s-}\Delta U_s \ln(1+\Delta U_s)$$

Hence, the computation of $\Lambda_T \ln\Lambda_T$ reduces to:

$$\begin{aligned}\Lambda_T \ln\Lambda_T &= \int_0^t \Lambda_{s-}dH_s - \frac{1}{2}\int_0^t \Lambda_{s-}d\langle U^c\rangle_s + \int_0^t \Lambda_{s-}dD_s\\ &\quad + \int_0^t \Lambda_{s-}\ln\Lambda_{s-}dH_s + \int_0^t \Lambda_{s-}d\langle U^c\rangle_s + \sum_{s\leqslant t}\Lambda_{s-}\Delta U_s \ln(1+\Delta U_s)\\ &= \int_0^t \Lambda_{s-}(1+\ln\Lambda_{s-})\,dU_s + \frac{1}{2}\int_0^t \Lambda_{s-}d\langle U^c\rangle_s + \sum_{s\leqslant t}\Lambda_{s-}f(1+\Delta U_s)\\ &= M_t + A_t + V_t\end{aligned}$$

where $f(\psi) = \psi\ln\psi - (\psi - 1)$ and V_t is increasing but not predictable. By the same reasoning as in Proposition A.2.2, we decompose V_t into a local $P-$martingale K_t and a predictable process of finite variation L_t. Noting that $f(1+\Delta U_s) = f(\psi_s^{\intercal})\,\mathbf{1}_{n\times m}\mathbf{1}_{\{\Delta Y_s\neq 0\}}$, we have:

$$\begin{aligned}V_t &= \Lambda_- f(\psi^{\intercal})\,\lambda_t = K_t + L_t\\ &= \Lambda_- f(\psi^{\intercal}) * (N_t - \lambda_t) + \Lambda_- f(\psi^{\intercal})\,\lambda_t\end{aligned}$$

Hence the canonical decomposition of $\Lambda_T \ln\Lambda_T$ follows:

$$\Lambda_T \ln\Lambda_T = (M_t + K_t) + (A_t + L_t) \tag{A.15}$$

Taking conditional expectation w.r.t. P on both sides of (A.15) and noting that the first parenthesis is a $P-$local martingale yields:

$$\begin{aligned} E\left[\Lambda_T \ln \Lambda_T\right] &= E\left[\frac{1}{2}\int_0^t \Lambda_{s-} d\left\langle U^c\right\rangle_s - \Lambda_- f\left(\psi^\intercal\right)\lambda_t dt\right] \\ &= E^Q\left[\frac{1}{2}\int_0^t \gamma_s^\intercal \gamma_s + f\left(\psi^\intercal\right)\lambda_t dt\right] \end{aligned}$$

which concludes the proof. ■

A.2.1 Minimal Distance Premiums

A solution to (2.39) is equivalent to elicite the risk-premia processes subject to the martingale condition (2.34). Thanks to previous Propositions, we are now in place to specialize Girsanov parameters (γ, ψ).

Proposition A.2.4 *Case of power utility:*

$$\gamma_s = \frac{\sigma_s \Upsilon_s}{q\left(1-q\right)},\ \ \psi_s = \left(\mathbf{1}_m - \frac{\Upsilon_s}{q}\right)^{\frac{1}{q-1}}$$

Case of logarithmic utility:

$$\gamma_s = \sigma_s \Upsilon_s,\ \ \psi_s = \frac{1}{\mathbf{1}_m - \Upsilon_s}$$

Case of exponential utility:

$$\gamma_s = \sigma_s \Upsilon_s,\ \ \psi_s = \exp\left(\Upsilon_s\right)$$

where Υ_s solves (2.34).

Proof. Solving (2.31) reduces to optimize the canonical decomposition of a constrained utility-distance based functional. Relying on Proposition 2.4.5, we need to minimize a concave function subject to a convex constraint. Following Rockafellar (1970) [143], it is enough to consider the Lagrangian function and resort to the saddle point theorem where we let Υ to be a $\mathbb{R}^m-$valued Lagrange multiplier. From the first order conditions, we obtain the optimal (γ, ψ) in terms of the Lagrange multiplier Υ which in turn satisfies the martingale condition (2.34), say $\mathcal{L}\left(\Upsilon\right)$. Furthermore, it can be verified that in each case, the function $\mathcal{L}\left(\Upsilon\right)$ is continuous and stricly increasing, thus admitting a unique solution yielding the unicity of the elicited risk premiums. ■

A.3. Optimization Results

In this Appendix, we collect results on optimization results, cf. Section 2.4.2.

Proof of Lemma 2.4.10. Noting $Y_t^\phi = f(X_t^\phi)$ with $f(x) = x^p/p$ and $f(x) = -e^{-x}$ in the CRRA and CARA cases respectively, Itô's lemma yields:

$$dY_t^\phi = f^{'}(X_t^\phi)dX_t^\phi + \frac{1}{2}f^{''}(X_t^\phi)d\left[X^{\phi,c}\right]_t + \sum_{s\leqslant t}\left(f(X_s^\phi) - f(X_{s-}^\phi) - f^{'}(X_{s-}^\phi)\Delta X_{s-}^\phi\right)$$

so that:

$$dY_t^{\text{pow},\phi} = (X_t^\phi)^p\left(r_t + \phi_t^\intercal\left(\pi_t\left(b\right) - r_t\mathbf{1}_m\right) + \frac{1}{2}\left(p-1\right)\phi_t^\intercal\sigma_t\sigma_t^\intercal\phi_t\right)dt + (X_t^\phi)^p\phi_t^\intercal\sigma_t d\overline{W}_t + (X_{t-}^\phi)^p\left((1+\phi_t^\intercal\omega_t)^p - 1\right)dN_t$$

and:

$$dY_t^{\text{exp},\phi} = e^{-X_t^\phi}\left(r_t + \phi_t^\intercal\left(\pi_t\left(b\right) - r_t\mathbf{1}_m\right) - \frac{1}{2}\phi_t^\intercal\sigma_t\sigma_t^\intercal\phi_t\right)dt + e^{-X_t^\phi}\phi_t^\intercal\sigma_t d\overline{W}_t - e^{-X_t^\phi}\left(e^{-(1+\phi_t^\intercal\omega_t)} + 1\right)dN_t$$

As $u(t, X_t^\phi) = f(X_t^\phi)\eta_{\phi_t}^{\text{x}}(t)$, the integration by parts formula yields:

$$u(t, X_t^\phi) = u\left(0, x\right) + \int_0^t \eta_{\phi_s}^{\text{x}}\left(s\right)dY_s^\phi + \int_0^t Y_s^\phi\frac{\partial}{\partial s}\eta_{\phi_s}^{\text{x}}\left(s\right)ds + \sum_{s\leqslant t}\left(Y_s^\phi\eta_{\phi_s}^{\text{x}}\left(s\right) - Y_{s-}^\phi\eta_{\phi_s}^{\text{x}}\left(s_-\right)\right)$$

Let us define:

$$M_t^{\phi,1} = \int_0^t Y_s^\phi\eta_{\phi_s}^{\text{x}}\left(s\right)\phi_s^\intercal\sigma_s d\overline{W}_s$$

and:

$$M_t^{\text{x},\phi,2} = \sum_{s\leqslant t}\left(Y_s^\phi\eta_{\phi_s}^{\text{x}}\left(s\right) - Y_{s-}^\phi\eta_{\phi_s}^{\text{x}}\left(s_-\right)\right) - m_t^{x,\phi}$$

with:

$$m_t^{\text{pow},\phi} = \frac{1}{p}\int_0^t (X_s^\phi)^p\eta_{\phi_s}^{\text{x}}\left(s\right)\left((1+\phi_s^\intercal\omega_s)^p - 1\right)\pi_s\left(\lambda\right)ds$$

and:

$$m_t^{\text{exp},\phi} = \int_0^t e^{-X_s^\phi}\eta_{\phi_s}^{\text{x}}\left(s\right)\left(e^{-(1+\phi_s^\intercal\omega_s)} + 1\right)\pi_s\left(\lambda\right)ds$$

Then, as $M_t^{\phi,1}$ and $M_t^{\mathrm{x},\phi,2}$ are $\mathcal{G}_t-$martingales (boundedness conditions), we note $M_t^{\mathrm{x},\phi} = M_t^{\phi,1} + M_t^{\mathrm{x},\phi,2}$ and the statement follows. ■

3

Financial Value of Information

Agenda

Abstract. This chapter deals with the problem of the cost of uncertainty associated with the utility maximization problem in a complete market with multiple risky assets and unobservable dividends in a Gaussian framework. This leads naturally to a partial information setup from which filtering techniques can be applied. Using a distortion power solution for the primal problem, we prove that the value function can be expressed in terms of the solution of a semilinear PDE, which is suggested by the dynamic programming approach. An explicit solution is obtained for HARA utilities, which we treat in a unified manner. Under these general results, the links between both the optimal investment strategy and the value function under full and partial information are explicited. This allows then to draw some insights on the financial value of information: the minimal initial endowment an investor with partial information must hold in order to attain the same expected utility as under full information. We apply our approach to a two-assets market model and discuss the numerical results in terms of optimal investment strategy.

3.1. Introduction

Information acquisition is an irreversible process. Once apprised any given fact, one cannot return to the state of ignorance. Therefore, information is a crucial ingredient in the financial context of derivative pricing and of portfolio optimization. Nevertheless, information is also a poorly understood concept. The present work attempts to investigate the question of the financial value of information in a utility-based manner. To this end, we consider the problem of maximizing expected utility from terminal wealth in a Gaussian financial market model under the respective hypotheses of full and partial information.

The typical example where our setup is applicable is when one considers the case of a data provider (say Bloomberg, Datastream or Reuters) who wants to assess the monetary value of its database, thus the price at which it should be sold. Turning the problem from an agent with full information (having access to the databases) to one with partial information (not having this access), one has to quantify the minimal initial endowment an investor with partial information must hold in order to attain the same expected utility as under full information. In formula, the problem to solve reads:

Problem 3.1.1 *Let $\mathbb{F}$ and $\mathbb{G}$ represent market model under full and partial information, respectively; with associated filtration $\mathcal{F}$ and $\mathcal{G}$ with $\mathcal{G}_t \subsetneq \mathcal{F}_t$. Also, θ_t represent an unobservable factor process, while $m_t = E^P\left[\theta_t|\mathcal{G}_t\right]$.*

$$\text{Find } P_t \text{ s.t. } E^P\left[u^{\mathbb{F}}\left(t, x, \theta_t\right)|\mathcal{G}_t\right] = u^{\mathbb{G}}\left(t, x + P_t, m_t\right)$$

The *extra* initial endowment P_t will subsequently be called the financial value of information. This is the *fair* price at which the data provider grant subscription access to its database and at which the non-informed agent is willing to acquire new information. The aim of this article is to provide explicit characterization of P_t in the special case of a Gaussian market model.

The partial information situation appears when investors only observe the vector of stock prices and cannot disentangle the drift term from other sources of uncertainty. It goes back to Gennotte (1986) [74]. More specifically, in the model economy we consider, dividend yields, represented by the stochastic process θ_t, are altered by shocks and investors observe movements in the returns level but cannot perfectly distinguished their sources. Instead, they solve a signal extraction problem and, in our diffusion context, an estimator of the dividend process, say m_t, is given by the Kalman filter, cf. Liptser and Shiryaev (2001) [118]. The optimization problem with full information was pioneered and solved by Merton (1971) via the Bellman equation of dynamic programming. Using the same approach and linear Gaussian filtering technique, models with incomplete information has been investigated by Detemple (1986) [48]. Lakner (1998) [111] solved the optimization problem via the martingale approach and duality results and provided characterization of optimal strategy on the special case of the linear Gaussian model.

Pham (2002) [138] or Zariphopoulou (2001) [158] provide an explicit representation of the value function in terms of a *distortion* solution. In relation to this litterature, in the present paper, we solved the optimization primal problem via dynamic programming using *distortion* solutions and prove that the value function can be expressed in terms of the solution of a semilinear parabolic differential equation (PDE). An explicit solution is obtained for HARA (Hyperbolic Absolute Risk Aversion) utilities, which we treat in a unified manner. This solution is characterized by three ordinary differential equations (ODE) and we investigate how one can pass from their representation under full information to their version under partial information. This allows then to precisely investigate the links between the optimal investment strategies, the wealth processes and the value functions under the full and partial information frameworks and then to quantity two crucial terms: the hedging demand and the precautionary demand for uncertainty. Eventually, we draw some insights on the financial value of information in a utility-based manner for CARA (Constant Absolute Risk Aversion) i.e.: exponential, CRRA (Constant Relative Risk Aversion) i.e.: power and logarithmic risk preferences.

This paper is organized as follows. Section 2 states the partial information framework, the filtering result and the utility maximization problem. Then, in Section 3, the primal problem is solved via the dynamic programming approach for both full and partial information and optimal portfolios are exhibited and compared. The question of the link between value functions under complete and limited information is studied in Section 4, from which we give some insights on the question of the financial value of information. Having obtained rather general results, we study in Section 5 as a special case a two-assets market model for which formulae are specialized and numerical aspects are investigated. Section 6 concludes.

3.2. Preliminaries and Definitions

3.2.1 The Market

Let $(\Omega, \mathcal{A}, P)$ be a complete probability space. We assume throughout that all stochastic processes are defined on a finite time horizon $[0, T]$. Suppose that $W = (W_t)_{t \geqslant 0}$ and $W^* = (W_t^*)_{t \geqslant 0}$ are $m-$dimensional independent Wiener processes and that $\mathcal{A} = (\mathcal{A}_t)_{t \geqslant 0}$ is the filtration generated by W and W^* which satisfies the usual conditions. Here, $\mathcal{A}_t$ may be regarded as the *model* information available at time $t \in [0, T]$.

Consider from now on a market with m tradable assets with prices $S^i = \left(S_t^i\right)_{t \geqslant 0}$ for $i \in \|1; m\|$. Here, S^i -the price of the $i-$th risky asset - is assumed to satisfy the differential equation:

$$\frac{dS_t^i}{S_t^i} = \left(\mu^i - \theta_t^i\right) dt + \sigma_S^i dW_t,\ S_0^i = s_0^i$$

where the *volatilities* $\sigma_S^i = (\sigma_S^{i,1}, ..., \sigma_S^{i,m})^\intercal$ for $\sigma_S^{i,j}$ with $j \in \|1; m\|$ and the *growth rate* μ^i are constants. Moreover, the dividend yield $\theta^i = \left(\theta_t^i\right)_{t \geqslant 0}$ for $i \in \|1; m\|$ is assumed to follow a mean reverting process:

$$d\theta_t^i = \lambda^i \left(\delta^i - \theta_t^i\right) dt + \sigma_\theta^i dW_t + \omega_\theta^i dW_t^*,\ \theta_0^i \in \mathbb{R}$$

with $\sigma_\theta^i = (\sigma_\theta^{i,1}, ..., \sigma_\theta^{i,m})^\intercal$, $\omega_\theta^i = (\omega_\theta^{i,1}, ..., \omega_\theta^{i,m})^\intercal$ where, for $(i, j) \in \|1; m\|$, $\sigma_\theta^{i,j}$, $\omega_\theta^{i,j}$, λ^i and δ^i are constants. For any $\mathcal{C}^2-$function $f : \mathbb{R} \to \mathbb{R}$, we associate with θ^i the infinitesimal generator $\mathcal{L}_{\theta^i}$, given by:

$$\mathcal{L}_{\theta^i} f(y) = \lambda^i \left(\delta^i - y\right) \frac{\partial f}{\partial y}(y) + \frac{1}{2}\left(\left(\sigma_\theta^i\right)^2 + \left(\omega_\theta^i\right)^2\right) \frac{\partial^2 f}{\partial y^2}(y),\ y \in \mathbb{R}$$

To ease the presentation, we introduce additional notation. First, the instantaneous variances of the risky assets and of the dividends, respectively:

$$E\left[dS_t/S_t \cdot dS_t/S_t\right] = \sigma_S \sigma_S^\intercal dt \triangleq \Sigma^S dt$$

$$E\left[d\theta_t \cdot d\theta_t\right] = \left(\sigma_\theta \sigma_\theta^\intercal + \omega_\theta \omega_\theta^\intercal\right) dt \triangleq \left(\Sigma^\theta + \Omega^\theta\right) dt$$

where the $m \times m$ *volatility* matrices $\sigma_k = (\sigma_k^1, ..., \sigma_k^m)^\intercal$ for $k = \{S, \theta\}$ and $\omega_\theta = (\omega_\theta^1, ..., \omega_\theta^m)^\intercal$ are constant and non-singular, hence invertible. Second, the instantaneous covariances between the assets and the dividends read:

$$E\left[d\theta_t \cdot dS_t/S_t\right] = \sigma_\theta \sigma_S^\intercal dt \triangleq \Psi^{\theta S} dt$$

We also denote by $dR_t = \left(Diag\left(S_t\right)\right)^{-1} dS_t$ the *return process* and by $\theta_t = \left(\theta_t^1, ..., \theta_t^m\right)^\intercal$ the *dividend process*, for all $t \in [0, T]$, i.e.:

$$dR_t = (\mu - \theta_t)\, dt + \sigma_S dW_t,\ R_0 = r_0 \tag{3.1}$$

$$d\theta_t = \Lambda\left(\Delta - \theta_t\right) dt + \sigma_\theta dW_t + \omega_\theta dW_t^*,\ \theta_0 \in \mathbb{R} \tag{3.2}$$

with $\mu = \left(\mu^1, ..., \mu^m\right)^\intercal$, $\Lambda = \text{Diag}\left(\lambda^1, ..., \lambda^m\right)$, $\Delta = \left(\delta^1, ..., \delta^m\right)^\intercal$. Then, we associate with the process θ the infinitesimal generator $\mathcal{L}_\theta$ which is given, for any $\mathcal{C}^2-$function $f : \mathbb{R}^m \to \mathbb{R}$, by:

$$\mathcal{L}_\theta f\left(y\right) = \mathcal{L}_\theta^0 f\left(y\right) + \frac{1}{2}\sum_{i,j=1}^{m}\left(\Sigma_{ij}^\theta + \Omega_{ij}^\theta\right)\frac{\partial^2 f}{\partial y_i \partial y_j}\left(y\right),\ y \in \mathbb{R}^m \tag{3.3}$$

with:

$$\mathcal{L}_\theta^0 f\left(y\right) = \sum_{i,j=1}^{m} \Lambda_{ij}\left(\Delta_j - y_j\right)\frac{\partial f}{\partial y_i}\left(y\right) \tag{3.4}$$

3.2.2 Information Structure

In this economy, we outline two ways to model investor's information:

- *full* information: The investor observes both the asset prices and the dividend yields. Thus, the observations are given by the process $(S_t, \theta_t)_{t \geqslant 0}$ and we denote by $\mathcal{F} = (\mathcal{F}_t)_{t \in [0,T]}$ the $P-$augmentation of the filtration generated by prices and returns and we write:

$$\mathcal{F}_t = \sigma\left(S_s, \theta_s, s \in [0, t]\right)$$

- *partial* information: The investor does not have access to the filtration $\mathcal{F}$ and can only observe the vector of stock prices. Thus, we denote by $\mathcal{G} = (\mathcal{G}_t)_{t \in [0,T]}$ the $P-$augmentation of the filtration generated by the prices and we write:

$$\mathcal{G}_t = \sigma\left(S_s, s \in [0, t]\right)$$

Remark 3.2.1 *We clearly have the inclusions $\mathcal{G}_t \underset{\neq}{\subset} \mathcal{F}_t$. Also, we note that this setup is the same as Detemple (1986) [48] or Lakner (1998) [111].*

In the following, we will study investment decisions in the context of each of these information structures, which will be denoted respectively $\mathbb{F}$ and $\mathbb{G}$ financial markets. The fact is that when working under the $\mathbb{F}$ (resp. $\mathbb{G}$) financial market, all stochastic processes have to be $\mathcal{F}$ (resp. $\mathcal{G}$) adapted.

3.2.3 The Filtering Setup

Since, under the $\mathbb{G}-$financial market, the Markov process $\theta_t,\ t \in [0,T]$, is not observable, it is natural to introduce its $\mathcal{G}_t-$conditional estimator vector by:

$$m_t = E\left[\theta_t | \mathcal{G}_t\right]$$

and its conditional covariance matrix:

$$\gamma_t = E\left[\left(\theta_t - m_t\right)\left(\theta_t - m_t\right)^{\intercal} | \mathcal{G}_t\right]$$

Then, as the system (3.1)-(3.2) is conditionally Gaussian, the conditional law of θ_t w.r.t. $\mathcal{G}_t$ is also Gaussian with mean m_t and covariance γ_t and the pair (m, γ) satisfies a system of linear equations given by the so-called Kalman filter. Results from Liptser and Shiryaev (2001) [118] give[1]:

$$dm_t = \Lambda\left(\Delta - m_t\right) dt + \left(\Psi^{\theta S} - \gamma_t\right) \Sigma^{S*}\left(dR_t - \left(\mu - m_t\right) dt\right) \tag{3.5}$$

$$\frac{d\gamma_t}{dt} = \Sigma^{\theta} + \Omega^{\theta} - 2\Lambda\gamma_t - \left(\Psi^{\theta S} - \gamma_t\right) \Sigma^{S*}\left(\Psi^{\theta S} - \gamma_t\right)^{\intercal} \tag{3.6}$$

and we note that (3.6) is a deterministic equation.

Remark 3.2.2 *The filtering equation (3.5) bears an appealing structure. While dR_t is what one observed during the time $[t, t+dt)$, $\left(\mu - m_t\right) dt$ is what one would expect to happen during $[t, t+dt)$ conditionally to previous observations $\mathcal{G}_t$. Thus $dR_t - \left(\mu - m_t\right) dt$ is what is new. Also, $\left(\Psi^{\theta S} - \gamma_t\right)$ represents the amount of variance that one would expect to happen during $[t, t+dt)$ Therefore, $\left(\Psi^{\theta S} - \gamma_t\right)\left(dR_t - \left(\mu - m_t\right) dt\right)$ represents what is really new.*

The process $\overline{W}_t = \sigma_S^{*}(R_t - \int_0^t \left(\mu - m_s\right) ds)$ which also reads:

$$\overline{W}_t \triangleq W_t - \sigma_S^{*} \int_0^t \left(\theta_s - m_s\right) ds \text{ is a } (P, \mathcal{G}) - \text{Brownian motion}$$

This is the so-called *innovation* process in filtering theory (cf. Liptser and Shiryaev (2001) [118]). From this, we can restate the *return* and *dividend processes* dynamics w.r.t. the filtration $\mathcal{G}$ as:

$$dR_t = \left(\mu - m_t\right) dt + \sigma_S d\overline{W}_t$$

$$dm_t = \Lambda\left(\Delta - m_t\right) dt + \left(\Psi^{\theta S} - \gamma_t\right) \sigma_S^{*} d\overline{W}_t$$

Hence, consider the differential operator $\mathcal{L}_m$ associated with the process m:

$$\mathcal{L}_m f(y) = \mathcal{L}_m^0 f(y) + \frac{1}{2} \sum_{i,j,k,l=1}^{m} \left(\Psi_{ki}^{\theta S} - \gamma_{ki}(t)\right) \Sigma_{kl}^{S*} \left(\Psi_{lj}^{\theta S} - \gamma_{lj}(t)\right) \frac{\partial^2 f(y)}{\partial y_i \partial y_j} \tag{3.7}$$

[1] In the following, $*$ denotes the inverse operation, hence $\Sigma^{S*} \equiv (\Sigma^S)^{-1}$.

for any $\mathcal{C}^2-$function $f:\mathbb{R}^m\rightarrow\mathbb{R}$, and $\mathcal{L}^0_m$ of the same form than (3.4) with θ replaced by m. Finally, the conditional distribution of θ_t for all $t\in[0,T]$, say $P\left(\theta_t\in A|\mathcal{G}_t\right)=\int_A\rho\left(t,x\right)dx$, is of the Gaussian type and writes:

$$\rho\left(t,x\right)=\left(2\pi\right)^{-m/2}\left(\det\gamma_t\right)^{-1/2}\exp\left(-\frac{1}{2}\left(x-m_t\right)\gamma_t^*\left(x-m_t\right)^{\intercal}\right)\tag{3.8}$$

In the following, we suppose that the variance-covariance matrix γ_t is a symmetrical positive definite one and so γ_t^* exists in (3.8).

3.2.4 The Optimization Problem

By denoting $\phi_s=\left(\phi_s^1,...,\phi_s^m\right)^{\intercal}$ the vector of money amounts invested in the m risky assets at time s, a self-financing strategy for the $\mathbb{F}$ (resp. $\mathbb{G}$) financial market is a pair (x,ϕ) where $x\geqslant 0$ is the initial investment and ϕ is a $\mathbb{R}^m-$valued and $\mathcal{F}$ (resp. $\mathcal{G}$) adapted process s.t. the value process:

$$X_t^{\phi}=X_0^{\phi}+\int_0^t\phi_s^{\intercal}dR_s,\ X_0^{\phi}=x\tag{3.9}$$

is $P-$a.s. well defined, which is ensured as long as the integrals involved in (3.9) are. Then, associated with a pair (x,ϕ), the wealth process (3.9) satisfies the differential representation:

$$dX_t^{\phi}=\phi_t^{\intercal}\left[\left(\mu-\vartheta_t\right)dt+\sigma_S dW_t^{\vartheta}\right]\tag{3.10}$$

with:

$$\left(\mathcal{H},\vartheta_t,W_t^{\vartheta}\right)=\begin{cases}\left(\mathcal{F},\theta_t,W_t\right) & \mathbb{F}-\text{market}\\ \left(\mathcal{G},m_t,\overline{W}_t\right) & \mathbb{G}-\text{market}\end{cases}\tag{3.11}$$

Besides, the class of $\mathcal{H}-$admissible strategies for a given initial endowment $x\geqslant 0$ is defined by:

$$\begin{aligned}\mathcal{S}\left(\mathcal{H},x\right)=&\left\{\phi:[0,T]\times\Omega\rightarrow\mathbb{R}^m,\mathcal{H}-\textit{predictable}\right.\\ &\left.,\exists K>-\infty,\forall t,\ P(X_t^{\phi}\geqslant K)=1,X_0^{\phi}=x\right\}\end{aligned}$$

Thus, associated with the process (3.10), we introduce, for any $\mathcal{C}^2-$function $f:\mathbb{R}\rightarrow\mathbb{R}$ and fixed $y\in\mathbb{R}^m$, the differential operator $\mathcal{L}_{X,y}$:

$$\mathcal{L}_{X,y}f\left(x\right)=\sum_{i=1}^m\phi_i\left(\mu_i-y_i\right)\frac{\partial f}{\partial x}\left(x\right)+\frac{1}{2}\sum_{i,j=1}^m\phi_i\Sigma_{ij}^S\phi_j\frac{\partial^2 f}{\partial x^2}\left(x\right)\tag{3.12}$$

for $x\in\mathbb{R}$.

The Problem

A function $U : \mathbb{R} \to \mathbb{R}$ will be called a utility function if it is strictly increasing, strictly positive, of class $\mathcal{C}^2$ and satisfies:

$$U^{'}\left(0^{+}\right) = \infty,\ U^{'}\left(\infty\right) = 0 \tag{3.13}$$

The optimization problem the investor faces is to maximize the expected utility from his/her terminal wealth over the class of admissible policies.

Definition 3.2.3 *Let U be a utility function. Determine:*

$$u\left(t,x,y\right) = \sup_{\phi \in \mathcal{S}(\mathcal{H},x)} E^{P}\left[U(X_T^{\phi}) | X_t^{\phi} = x, \vartheta_t = y\right] \tag{3.14}$$

and find $\widehat{\phi}$ which satisfies $u\left(t,x,y\right) = E^{P}[U(X_T^{\widehat{\phi}})]$. Then, we call $\widehat{\phi}$ an optimal investment strategy and $\widehat{X} = X^{\widehat{\phi}}$ the optimal wealth process.

3.3. Portfolio Selection Problem

3.3.1 Distortion Solution

In this economy, the investor has risk preferences expressed via a concave utility function. In the following, we consider the three more standard cases:

$$U\left(x\right) = \left\{ \begin{array}{lccc} x^a/a & a \in \left(-\infty, 1\right),\ x \in \mathbb{R}^{+} & \text{power (CRRA)} \\ -e^{-ax} & , \quad a > 0,\ x \in \mathbb{R} & \text{exponential (CARA)} \\ \log x & x \in \mathbb{R}^{+} & \text{logarithmic (Myopic)} \end{array} \right. \tag{3.15}$$

and we introduce the parameter $q \in (-\infty, 1]$ varying with the utility function:

$$q = \left\{ \begin{array}{ll} a/\left(a-1\right), q < 1, q \neq 0 & \text{power} \\ 1 & \text{exponential} \\ 0 & \text{logarithmic} \end{array} \right. \tag{3.16}$$

As is clear from (3.3) and (3.7), the operators associated with the processes m and θ are formally close and so we can solve the optimization problem (3.14) for $\mathbb{F}$ and $\mathbb{G}$ financial markets in a really general way. To this end, we note:

$$\mathcal{L}_{\vartheta} f\left(y\right) = \mathcal{L}_{\vartheta}^{0} f\left(y\right) + \frac{1}{2}\sum_{i,j=1}^{m} \Upsilon_{ij} \frac{\partial^2 f}{\partial y_i \partial y_j}\left(y\right),\ \vartheta = \{\theta, m\}$$

and noting $\Phi = E\left[dR_t \cdot d\vartheta_t\right]/dt$ the correlation between (R, ϑ), we have[2]:

$$\left(\Upsilon, \Phi\right) = \left\{ \begin{array}{cc} \left(\Sigma^{\theta} + \Omega^{\theta}, \Psi^{\theta S}\right) & \mathbb{F} - \text{market} \\ \left(\left(\Psi^{\theta S} - \gamma_t\right)\Sigma^{S*}\left(\Psi^{\theta S} - \gamma_t\right), \Psi^{\theta S} - \gamma_t\right) & \mathbb{G} - \text{market} \end{array} \right. \tag{3.17}$$

[2] To lighten the notations, we omit the lower-script t in γ when writing (Υ, Φ).

We can now state the *distortion* power solution for the primal optimization problem (3.14). Similar idea was used to solve rather general problems by Pham (2002) [138] or Zariphopoulou (2001) [158].

Proposition 3.3.1 *Let $q \leqslant 1$ be given by (3.16). Then, the value function (3.14) of the primal optimization problem is given by:*

$$u(t,x,y) = \begin{cases} U(x)\exp(v(t,y)) & , \quad q \leqslant 1, q \neq 0 \\ U(x) + v(t,y) & , \quad q = 0 \end{cases} \tag{3.18}$$

where $v \equiv v(t,y) : [0,T] \times \mathbb{R}^m \to \mathbb{R}$ satisfies the semilinear PDE:

$$v_t + \mathcal{L}_\vartheta v = \begin{cases} q/2 \sum_{i,j=1}^m \Sigma_{ij}^{S*} (\mu_i - y_i)^2 + w(y, v_y, v_{yy}) & , \quad q \leqslant 1, q \neq 0 \\ -1/2 \sum_{i,j=1}^m \Sigma_{ij}^{S*} (\mu_i - y_i)^2 & , \quad q = 0 \end{cases} \tag{3.19}$$

where:

$$w(y, z^1, z^2) = \sum_{i,j,k=1}^m q\Phi_{ki}\Sigma_{kj}^{S*}(\mu_j - y_j) z_i^1 + \frac{1}{2}\sum_{i,j,k,l=1}^m \left(q\Phi_{ki}\Sigma_{kl}^{S*}\Phi_{lj} - \Upsilon_{ij}\right) z_i^2$$

and boundary condition $v(T,y) = 0$.

Proof. The HJB equation for the optimization problem (3.14) reads as:

$$u_t + \max_{\phi \in \mathbb{R}^m} \left\{ \mathcal{L}_X u + \sum_{i,j=1}^m \phi_i \Phi_{ij} \frac{\partial^2 u}{\partial x \partial y_j} \right\} + \mathcal{L}_\vartheta u = 0 \tag{3.20}$$

with $u(T,x,y) = U(x)$ for $(t,x,y) \in [0,T] \times \mathbb{R} \times \mathbb{R}^m$. The first order condition for an optimal investment strategy is given by:

$$\widehat{\phi}_i = -u_{xx}^* \sum_{j=1}^m \Sigma_{ij}^{S*} \left((\mu_i - y_i) u_x + \sum_{k=1}^m \Phi_{jk} \frac{\partial^2 u}{\partial x \partial y_k} \right) \tag{3.21}$$

Inserting $\widehat{\phi}_i$ into the HJB equation (3.20) yields the semilinear PDE:

$$u_t - \frac{1}{2} u_{xx}^* \left\{ \sum_{i,j=1}^m (\mu_i - y_i) \Sigma_{ij}^{S*} u_x + \sum_{i,j=1}^m \Phi_{ij} \Sigma_{ij}^{S*} \frac{\partial^2 u}{\partial x \partial y_j} \right\}^2 + \mathcal{L}_\vartheta u = 0 \tag{3.22}$$

Then, we seek a solution to (3.22) in the form of (3.18). By direct substitution, it is easy to see that $v(t,y)$ satisfies the semilinear PDE given by (3.19). ■

Remark 3.3.2 *By following Pham (2002) [138] or Zariphopoulou (2001) [158], the proof can be made more precise. First, by establishing that the proposed solution is a viscosity solution of the HJB equation and second by ensuring that the value function is a classical solution of the HJB equation.*

3.3.2 Specializations

In the following, we present solutions of (3.18) in the case of full and partial information and discuss how one can pass from one to another one.

Solution for a $\mathbb{F}$-Market

A solution to (3.19) is derived in the next lemma.

Lemma 3.3.3 *The $\mathbb{F}$-solution v to (3.19) is given by:*

$$v(t,y) = y^{\intercal} a_t y + b_t^{\intercal} y + c_t \tag{3.23}$$

for all $(t,y) \in [0,T] \times \mathbb{R}^m$, with boundary condition $v(T,y) = 0$ and where:

- $q \leqslant 1, q \neq 0$:

$$\begin{aligned}\frac{da_t}{dt} &= 2a_t\zeta_1 a_t + 2\zeta_2 a_t + \frac{q}{2}\Sigma^{S*} \\ \frac{db_t}{dt} &= 2a_t\zeta_1 b_t + \zeta_2 b_t + 2\zeta_3 a_t - q\Sigma^{S*}\mu \\ \frac{dc_t}{dt} &= \frac{1}{2} b_t\zeta_1 b_t + \zeta_3 b_t - Tr(\zeta_4 a_t) + \frac{q}{2} Tr(\Sigma^{S*}\mu^2)\end{aligned} \tag{3.24}$$

  where:

$$\zeta_1 = q\Phi\Sigma^{S*}\Phi^{\intercal} - \zeta_4,\ \zeta_2 = \Lambda - q\Phi\Sigma^{S*},\ \zeta_3 = q\Phi\Sigma^{S*}\mu - \Lambda\Delta,\ \zeta_4 = \Upsilon$$

- $q = 0$:

$$\begin{aligned}\frac{da_t}{dt} &= 2\Lambda a_t - \frac{1}{2}\Sigma^{S*} \\ \frac{db_t}{dt} &= \Lambda b_t - 2\Lambda\Delta a_t + \Sigma^{S*}\mu \\ \frac{dc_t}{dt} &= -\Lambda\Delta b_t - Tr(\Upsilon a_t) - \frac{1}{2} Tr(\Sigma^{S*}\mu^2)\end{aligned} \tag{3.25}$$

with $a(T) = b(T) = c(T) = 0$. One may note that a is a $m \times m-$matrix, b is a $m-$vector while c is a scalar. Going back to (3.18), we have:

$$u^{\mathbb{F}}(t,x,\theta) = U(x) \exp(\theta_t^{\intercal} a_t \theta_t + b_t^{\intercal}\theta_t + c_t) \tag{3.26}$$

Proof. By plugging (3.23) into (3.19), we derive the desired ordinary differential equations for the coefficients a_t, b_t and c_t, with terminal conditions $a(T) = b(T) = c(T) = 0$ since $v(T,y) = 0$ for all $y \in \mathbb{R}^m$. ■

Remark 3.3.4 *In the logarithmic case, $q = 0$, the ordinary differential equations a and b are invariant under full and partial information while c is affected by learning through the covariance matrix Υ, cf. (3.25) and notation (3.17). The case $q \leqslant 1, q \neq 0$ will be deeply studied in subsequent sections.*

Solution for a $\mathbb{G}$-Market

In the following, we will denote by $\overline{a}_t, \overline{b}_t$ and $\overline{c}_t$ the ODE's given in Lemma 3.3.3, (3.24), for a $\mathbb{G}-$financial market, cf. notation (3.17), and we note:

$$u^{\mathbb{G}}(t,x,m) = U(x)\exp\left(m_t^{\intercal}\overline{a}_t m_t + \overline{b}_t^{\intercal} m_t + \overline{c}_t\right)$$

Lemma 3.3.5 *The equations $\overline{a}_t, \overline{b}_t$ and $\overline{c}_t$ satisfy:*

$$\begin{aligned}
\frac{d\overline{a}_t}{dt} &= 2\overline{a}_t\zeta_1^m\overline{a}_t + 2\zeta_2^m\overline{a}_t + \frac{q}{2}\Sigma^{S*} \\
\frac{d\bar{b}_t}{dt} &= 2\overline{a}_t\zeta_1^m\overline{b}_t + \zeta_2^m\overline{b}_t + 2\zeta_3^m\overline{a}_t - q\Sigma^{S*}\mu \\
\frac{d\overline{c}_t}{dt} &= \frac{1}{2}\overline{b}_t\zeta_1^m\overline{b}_t + \zeta_3^m\overline{b}_t - Tr\left(\zeta_4^m\overline{a}_t\right) + \frac{q}{2}Tr(\Sigma^{S*}\mu^2)
\end{aligned} \tag{3.27}$$

where:

$$\zeta_1^m = (q-1)\,\zeta_4^m,\ \ \zeta_2^m = \Lambda - q\left(\Psi^{\theta S} - \gamma_t\right)\Sigma^{S*}$$

$$\zeta_3^m = q\left(\Psi^{\theta S} - \gamma_t\right)\Sigma^{S*}\mu - \Lambda\Delta,\ \ \zeta_4^m = \left(\Psi^{\theta S} - \gamma_t\right)\Sigma^{S*}\left(\Psi^{\theta S} - \gamma_t\right)^{\intercal}$$

Next, we present a result which will thereafter be of capital importance.

Lemma 3.3.6 *The following identities, for $q \leqslant 1, q \neq 0$, hold:*

$$\begin{aligned}
\overline{a}_t &= \left(\mathbf{1}_m - 2\gamma_t a_t\right)^* a_t \\
\overline{b}_t &= \left(\mathbf{1}_m - 2\gamma_t a_t\right)^* b_t
\end{aligned} \tag{3.28}$$

Proof. cf. Appendix B.1.2. ■

This gives the link between the $\mathbb{F}$ and $\mathbb{G}$ financial markets.

3.3.3 Optimal Policy

Eventually, the optimal investment policy can be further explicited.

Lemma 3.3.7 *The optimal strategy $\widehat{\phi}$ to problem (3.14), is given by:*

$$\widehat{\phi}_t = A(x)\,\Sigma^{S*}\left\{\mu_t - \vartheta_t\left(\mathbf{1}_m - \mathbf{1}_{q\neq 0}\left\{2\Phi a_t\right\}\right) + \mathbf{1}_{q\neq 0}\left\{\Phi b_t\right\}\right\} \tag{3.29}$$

with q as in (3.16) and Arrow-Pratt relative risk-aversion:

$$A(x) = -\frac{U^{'}(x)}{U^{''}(x)} = \begin{cases} x/\left(1-a\right) & \textit{power} \\ 1/a & \textit{exponential} \\ x & \textit{logarithmic} \end{cases} \tag{3.30}$$

or similarly:

$$\widehat{\phi}_t = A(x)\,\Sigma^{S*}\left\{\left(\mu_t - \vartheta_t\right) + \mathbf{1}_{q\neq 0}\Phi v_{\vartheta}\left(t,\vartheta\right)\right\} \tag{3.31}$$

Proof. The result follows from direct substitution of (3.18) with v given by (3.23) into (3.21). When $q = 0$, the formula simplifies (see below). ■

In Appendix B.1, we specialize (3.29) to the case of full and partial information. This lead us to quantity two crucial terms which appear in the partial case: the hedging and the precautionary demands for uncertainty.

3.4. Value of Information

As pointed out by several authors, cf. Lakner (1998) [111] or Pham and Quenez (2001) [139], the optimal investment policy under partial information cannot be derived from the full case by just replacing θ by its conditional version m, cf. Propositions B.1.1 and B.1.4. This holds true only in the myopic logarithmic case, cf. Remark 3.3.4, a fact that is known as the certainty equivalence principle, cf. Kuwana (1995) [110]. But what about value functions?

3.4.1 CARA and CRRA Utilities

The next result gives somes insights on this question.

Lemma 3.4.1 *When $q \leqslant 1, q \neq 0$, we have:*

$$E\left[\exp\left(\theta_t^\intercal a_t \theta_t + b_t^\intercal \theta_t\right)|\mathcal{G}_t\right] = \exp\left(d_t\right)\exp\left(m_t^\intercal \overline{a}_t m_t + \overline{b}_t^\intercal m_t\right)$$

with d_t given by:

$$d_t = \frac{1}{2}\left(b_t^\intercal D_t^* \gamma_t b_t - \log\det D_t\right) \tag{3.32}$$

where $D_t = \mathbf{1}_m - 2\gamma_t a_t$.

Proof. From Section 3.2.3, we know that the $\mathcal{G}_t-$conditional law of θ_t is Gaussian with mean vector m_t and covariance matrix γ_t and is given by (3.8). Noting $D_t = \mathbf{1}_m - 2\gamma_t a_t$ and $\Gamma_t = (2\pi)^{-m/2}(\det\gamma_t)^{-1/2}$, it follows that:

$$\begin{aligned}
&e^{\mathbb{F}/\mathbb{G}} \triangleq E\left[\exp\left(\theta_t^\intercal a_t \theta_t + b_t^\intercal \theta_t\right)|\mathcal{G}_t\right] \\
&= \int_{\mathbb{R}^m} \exp\left(x^\intercal a_t x + b_t^\intercal x\right)\rho\left(t,x\right)dx \\
&= \Gamma_t \int_{\mathbb{R}^m} \exp\left(x^\intercal a_t x + b_t^\intercal x\right)\exp\left(-\frac{1}{2}\left(x - m_t\right)\gamma_t^*\left(x - m_t\right)^\intercal\right)dx \\
&= \Gamma_t \int_{\mathbb{R}^m} \exp\left(m_t^\intercal D_t^* a_t m_t + D_t^{*\intercal} b_t^\intercal m_t\right) \times \exp\left(\frac{1}{2} b_t^\intercal D_t^* \gamma_t b_t\right) \\
&\times \exp\left(-\frac{1}{2}\left(x - D_t^*\left(m_t + \gamma_t b_t\right)\right)D_t \gamma_t^*\left(x - D_t^*\left(m_t + \gamma_t b_t\right)\right)^\intercal\right)dx
\end{aligned}$$

Thus, we have:

$$e^{\mathbb{F}/\mathbb{G}} = \exp\left(\frac{1}{2}b_t^\intercal D_t^* \gamma_t b_t - \frac{1}{2}\log\det D_t\right) \times \exp\left(m_t^\intercal D_t^* a_t m_t + D_t^{*\intercal} b_t^\intercal m_t\right)$$

which is the desired result. ■

The things are made clearer owing to the following useful lemma, which is close in essence to Lemma 3.3.6 and is motivated by Lemma 3.4.1. For the rest of this chapter, the upper-scripts cr and ca denote respectively the Constant Relative and Constant Absolute (Risk Aversion) cases, corresponding respectively to power and exponential preferences.

Lemma 3.4.2 *The following identity, for $q \leqslant 1, q \neq 0$,holds:*

$$\bar{c}_t - c_t - d_t = \begin{cases} Tr\left(\frac{q}{1-q}\left(\frac{1}{2}\log\left(\mathbf{1}_m + 2D_t^*\gamma_t e_t^{cr}\right) + f_t\right)\right) & q < 1, q \neq 0 \\ Tr\left(D_t^*\gamma_t e_t^{ca} + f_t\right) & q = 1 \end{cases} \tag{3.33}$$

with d_t as in (3.32) and e_t^{cr}, e_t^{ca} and f_t satisfying the ODE's:

$$\frac{de_t^{cr}}{dt} = -\frac{1-q}{2}e_t^1 + e_t^{2,cr} - 2e_t^{cr}e^0 e_t^{cr} \tag{3.34}$$

and:

$$\frac{de_t^{ca}}{dt} = -\frac{1}{2}e_t^1 + e_t^{2,ca} \tag{3.35}$$

where for $x = \{cr, ca\}$:

$$e^0 = \Psi^{\theta S}\Sigma^{S*}\Psi^{\theta S} - \left(\Sigma^\theta + \Omega^\theta\right)$$

$$e_t^1 = \left(\mathbf{1}_m - 2\Psi^{\theta S}a_t\right)\Sigma^{S*}\left(\mathbf{1}_m - 2\Psi^{\theta S}a_t\right)^\intercal$$

$$e_t^{2,x} = 4a_t e^0 e_t^x + 2\left(\Lambda - \Psi^{\theta S}\Sigma^{S*}\right)e_t^x$$

and:

$$\frac{df_t}{dt} = e^0 e_t^x \tag{3.36}$$

Proof. The proof can be done in the same line as the one in Lemma 3.3.6, although the computations are largely more involved and tedious. ■

We are now in position to define the financial value of information under partial information and learning in a utility-based manner. The next result is a ramification of Lemma 3.4.1 with the help of Lemma 3.4.2.

Proposition 3.4.3 (CARA & CRRA cases) *Let $u^{\mathbb{F}}$ be given by (3.26), θ by (3.2) and its $\mathcal{G}-$version m by (3.5) and $q \leqslant 1, q \neq 0$, then we have:*

$$E\left[u^{\mathbb{F}}(t,x,\theta)\,|\mathcal{G}_t\right] = \exp\left(-\varrho_t^{cr,ca}\right) u^{\mathbb{G}}(t,x,m)$$

with $\varrho_t^{cr,ca}$ given by:

$$\varrho_t^{cr,ca} = \begin{cases} Tr\left(\frac{q}{1-q}\left(\frac{1}{2}\log\left(\mathbf{1}_m + 2D_t^*\gamma_t e_t^{cr}\right) + f_t\right)\right), & q < 1, q \neq 0 \\ Tr\left(D_t^*\gamma_t e_t^{ca} + f_t\right) & q = 1 \end{cases} \tag{3.37}$$

where e_t^{cr}, e_t^{ca} and f_t as in (3.34), (3.35) and (3.36) and so:

$$\frac{d\varrho_t^{cr,ca}}{dt} = -\frac{q}{2}Tr\left\{\left(\mathbf{1}_m - 2\gamma_t a_t\right)^* \gamma_t \left(\mathbf{1}_m - 2\Psi^{\theta S} a_t\right)\Sigma^{S*}\left(\mathbf{1}_m - 2\Psi^{\theta S} a_t\right)^{\intercal}\right\}$$

Proof. This is a direct application of Lemmas 3.4.1 and 3.4.2. ■

The quantity (3.37) can clearly be interpreted as the minimal monetary amount that when added to the initial capital makes the investor indifferent in terms of maximal terminal expected utility between learning about the dividend yields or not. We precise this point by noting:

$$E\left[u^{\mathbb{F}}(t,x,\theta)\,|\mathcal{G}_t\right] = u^{\mathbb{G}}(t,\widehat{x}_t,m)$$

where:

$$\widehat{x}_t = \begin{cases} x\exp\left(-a^*\varrho_t^{cr}\right) & \text{power (CRRA)} \\ x + a^*\varrho_t^{ca} & \text{exponential (CARA)} \end{cases},$$

Then, the difference $\widehat{x}_t - x \geqslant 0$ can be interpreted as the *cost* of uncertainty. This is the endowment an investor with partial information must hold in order to attain the same expected utility as one with full information.

Remark 3.4.4 *Under exponential utility, the financial value of information $\widehat{x}$ may have an interpretation as an utility-based price. Following Hodges and Neuberger (1989) [89] or Davis (1997) [37], the idea of utility-based valuation is the following: a contingent claim is priced so that the investor's utility remains the same whether the optimal portfolio includes a marginal amount of the derivative security or not. If there exists a unique solution ρ^s to:*

$$u(t,x,\cdot) = u(t,x+\rho^s,\cdot)$$

then, ρ^s is called the utility indifference selling (ask) price. Therefore, we have the identification $\rho^s = a^*\varrho_t^{ca}$ and ρ^s is the reservation price of an option which pays off if m, the estimator of θ, is in fact equal to θ.

3.4.2 Logarithmic Utility

When $q = 0$, the things are more simple as shown by Proposition 3.3.1 and Lemma 3.3.3. Therefore, by mimicking the previous analysis, it follows.

Lemma 3.4.5 *The following identity, for $q = 0$, holds:*

$$\bar{c}_t = c_t + Tr(\gamma_t a_t + g_t)$$

with g_t satisfying the ODE:

$$\frac{dg_t}{dt} = \frac{1}{2}\Sigma^{S*}\gamma_t \tag{3.38}$$

Proof. The proof can be done as in Lemma 3.4.2. ■

We are now in position to define the financial value of information in the *myopic* logarithmic case by following Proposition 3.4.3.

Proposition 3.4.6 (Logarithmic case) *Let $u^{\mathbb{F}}$ be given by (**??**), θ by (3.2) and its $\mathcal{G}-$version m by (3.5) and $q = 0$, then we have:*

$$E\left[u^{\mathbb{F}}(t, x, \theta)\,|\mathcal{G}_t\right] = -\varrho_t^{\log} + u^{\mathbb{G}}(t, x, m)$$

with $\varrho_t^{\log}$ given by:

$$\varrho_t^{\log} = Tr(\gamma_t a_t + g_t) \tag{3.39}$$

where g_t is given by (3.38) and so:

$$\frac{d\varrho_t^{\log}}{dt} = -Tr\left\{\left(\left(\Psi^{\theta S} - \gamma_t\right)\Sigma^{S*}\left(\Psi^{\theta S} - \gamma_t\right)^{\intercal} - \left(\Sigma^{\theta} + \Omega^{\theta}\right)\right) a_t\right\}$$

Similarly as for the case of CARA and CRRA utilities, the quantity (3.39) can be interpreted as the *cost* of uncertainty that when added to the initial endowment yields the same terminal expected utility for an agent with partial information than one with full information. We note:

$$E\left[u^{\mathbb{F}}(t, x, \theta)\,|\mathcal{G}_t\right] = u^{\mathbb{G}}(t, \widehat{x}_t, m)$$

where:

$$\widehat{x}_t = x\exp(-\varrho_t^{\log}), \text{ logarithmic (Myopic)}$$

Then, the difference $\widehat{x}_t - x \geqslant 0$ is termed the financial value of information.

3.5. Application

In the following, we conduct a simulation study to demonstrate the main characteristics of the previously derived "value of information" in the case of a two-assets market model, cf. Appendix B.3 for more details and notations.

Table 3.1 presents the parameters values for four reference models. Using this set of parameters for the two-asset model, we simulate stock price and dividend yield and then use the Kalman filter to retrieve the true values, cf. cf. Appendix B.3. Figure 3.1 presents some filtering results.

	Case 1	Case 2	Case 3	Case 4
μ	0.08	0.08	0.08	0.08
σ	0.5	0.8	0.8	0.6
λ	3.1	3.1	4.1	5.1
δ	0.05	0.02	0.05	0.05
ω	0.1	0.6	1.2	0.6
ρ	0.7	0.7	-0.4	0.4

Table 3.1: Parameters values - Four reference models.

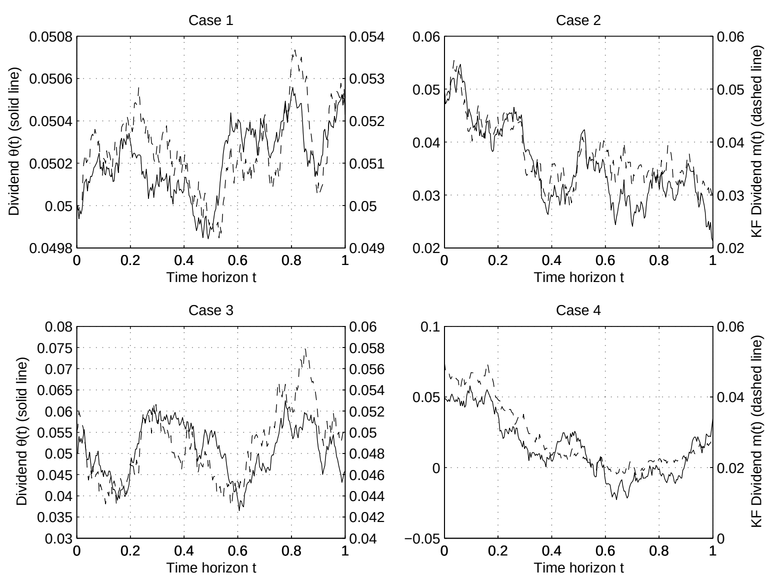

Figure 3.1: Dividends, true and filtered - Four reference models.

Expressions for (a, b, c)

Next, the solutions of the ODE's (a, b, c) are derived.

Lemma 3.5.1 *The equations* (a, b, c) *admit the representations:*

- $q \leqslant 1, q \neq 0$:

$$\begin{aligned}
a(t) &= \frac{a_2}{a_3} \frac{\exp(2a_2(T-t)) - a_4}{\exp(2a_2(T-t)) + a_4} + \frac{a_1}{a_3} \\
b(t) &= \int_t^T \left(2a_5 a_s - q\xi^2 \widetilde{\mu}\right) \exp\left(-a_1(s-t) + a_3 \int_t^s a_u du\right) ds \\
c(t) &= -\frac{1}{4} a_3 \int_t^T b_s^2 ds + a_5 \int_t^T b_s ds - \rho\sigma\omega \int_t^T a_s ds + \frac{q}{2} \xi^2 \widetilde{\mu}^2 (T-t)
\end{aligned}$$

where:

$$a_1 = \lambda - q\rho\xi\omega, \ a_2 = \sqrt{\lambda^2 + q\omega^2\xi^2 - 2q\lambda\rho\xi\omega}$$

$$a_3 = 2\omega^2\left(1 - q\rho^2\right), \ a_4 = \frac{a_2 + a_1}{a_2 - a_1}, \ a_5 = q\rho\xi\omega\widetilde{\mu} - \lambda\delta$$

- $q = 0$:

$$\begin{aligned}
a(t) &= \frac{\xi^2}{4\lambda} (\exp(2\lambda(T-t)) - 1) \\
b(t) &= \int_t^T \left(-2\lambda\delta a_s + \widetilde{\mu}\xi^2\right) \exp(-\lambda s) ds \\
c(t) &= -\lambda\delta \int_t^T b_s ds - \omega^2 \int_t^T a_s ds - \frac{\xi^2 \widetilde{\mu}^2}{2} (T-t)
\end{aligned}$$

Remark 3.5.2 *To ease the presentation, we did not present here the exact solutions of the ODE's. Those are detailed in Appendix B.3.2.*

From Sections 3.3 and 3.4, the investment strategy and the value function (for both $\mathbb{F}$ and $\mathbb{G}$-financial markets) are computed in terms of the triplet of ODE's (a, b, c). As they are of paramount importance, we present some numerical aspects relative to their computation and shape.

Shape Properties As is clear from previous lemma, the functions (a, b, c) bear similarities in the power and exponential cases, only varying in the $q-$parameter, while the logarithmic case can be analyzed separately. For power risk preferences, (a, b, c) are influenced by the agent's risk attitude which is expressed via the $a-$parameter and in turn via q as $q = a/(a-1)$.

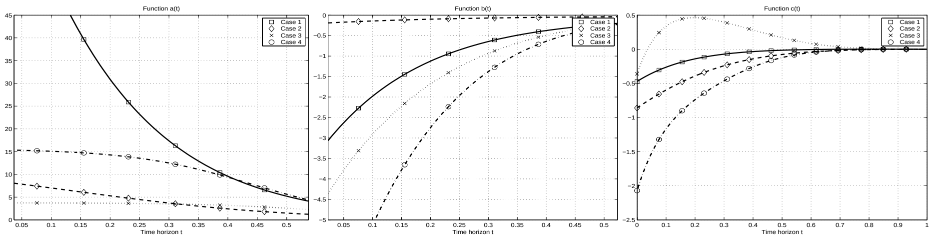

Figure 3.2: Functions (a,b,c) - Four cases - Power utility ($a = -5$).

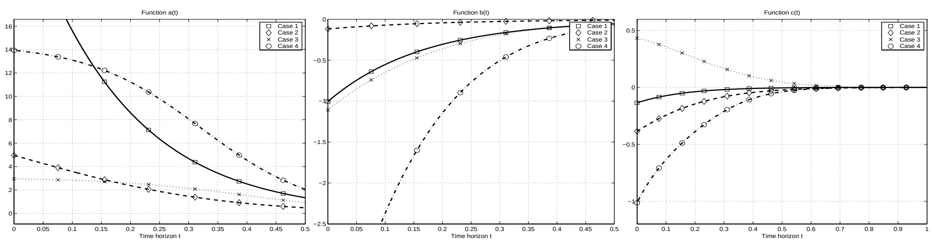

Figure 3.3: Functions (a,b,c) - Four cases - Power utility ($a = 0.8$).

The investor is said to be *conservative* if $a \in (-\infty, 0)$, cf. Figure 3.2, and *agressive* if $a \in (0, 1)$, cf. Figure 3.3. Even if the shape of the (a, b, c) functions are similar in these two cases, their levels are quite different. The *conservative* agent always obtain higher levels (in absolute terms) for all of these ODE's whatever the model used to compute the functions. Then, we note that an agent with exponential preferences, cf. Figure 3.4, yields similar levels pattern as a *conservative* power investor (at least in the case $a = -5$). Considering the four reference models, we observe that the function a decreases as the model becomes more volatile: case 3 with the highest variance (σ, ω) parameters always yields the lowest $a-$level. The rate of mean-reversion (λ) also plays an important role: it causes increase in the $a-$level, cf. case 4. The same ordering can be observed for the b function and the variance and mean reversion parameters show importance again.

The things are quite different for the c function; which only enters in the computation of the optimal primal and dual value functions. For power and exponential risk preferences, the highest volatile model (case 4) yields higher levels than in the other cases, but in all cases we observe a smaller dispersion than in the situation of the (a, b) functions.

In the logarithmic case, cf. Figure 3.5, the same shape analysis can be conducted. The main divergence lies in the relative gaps between the four considered models. The difference in levels between each ones seem to be more marked for this utility function than the two others. This may be due to the

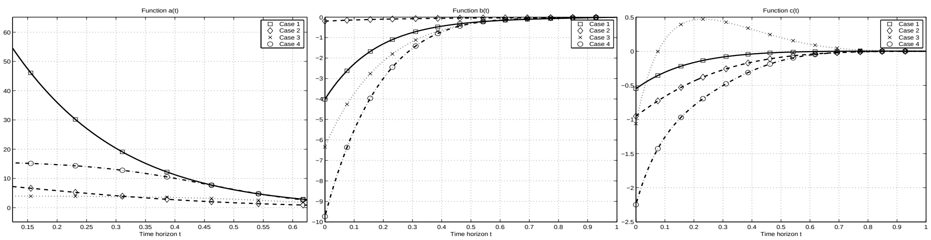

Figure 3.4: Functions (a,b,c) - Four cases - Exponential utility.

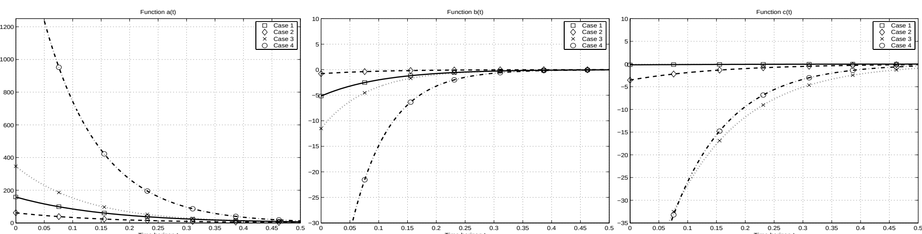

Figure 3.5: Functions (a,b,c) - Four cases - Logarithmic utility.

myopic behaviour of the investor who as being less strategic towards his/her environment, then suffers more from his/her financial model.

Remark 3.5.3 *From Lemma B.2.4, we know that (a, b) enter in the computation of the dual optimizer, which is the risk neutral Radon-Nikodym derivative. Therefore, these can be interpreted as factors for the market price of risk (ψ) due to the incompleteness introduced by the dividend process. In the logarithmic case, the agent acts as if $\rho = 0$, complete market, and so $\psi = 0$.*

Value of Information

In order to be able to compute the financial value of information, cf. Proposition 3.4.3, we need analytical solutions to the equations (e^{cr}, e^{ca}, f, g).

Lemma 3.5.4 *The solutions of (e^{cr}, e^{ca}, f, g) are given by:*

$$e_t^{cr} = a_t + \frac{1}{2\omega^2(\rho^2 - 1)} \left(\tilde{e}^2 \frac{\exp\left(2\tilde{e}^2 (T - t)\right) - \tilde{e}^3}{\exp\left(2\tilde{e}^2 (T - t)\right) + \tilde{e}^3} - \tilde{e}^1 \right)$$

$$e_t^{ca} = -\frac{1}{2}\xi^2 \int_t^T (1 - 2\rho\sigma\omega a_s) \exp\left(-2e^2 (s - t) - 4e^1 \int_t^s a_u du \right) ds$$

with:

$$\widetilde{e}^1 = \rho\xi\omega - \lambda,\ \widetilde{e}^2 = \sqrt{\xi^2\omega^2\left(2\rho^2 - 1\right) + \lambda\left(\lambda - 2\rho\xi\omega\right)},\ \widetilde{e}^3 = \frac{\widetilde{e}^2 - \widetilde{e}^1}{\widetilde{e}^2 + \widetilde{e}^1}$$

and:

$$f_t = -\omega^2\left(1 - \rho^2\right)\int_t^T e_s ds,\ g_t = \frac{1}{2}\xi^2\int_t^T \gamma_s ds$$

Proof. The only difficulty is in solving e_t^{cr}. In fact, it does not admit a direct solution as for e_t^{ca}. Instead, noting $\widetilde{e}_t = a_t - e_t^{cr}$, we make use of the relation:

$$\frac{d\widetilde{e}_t}{dt} = -2\omega^2\left(1 - \rho^2\right)\widetilde{e}_t^2 + 2\left(\lambda - \rho\xi\omega\right)\widetilde{e}_t - \frac{1}{2}\xi^2$$

whose solution is given by:

$$\widetilde{e}_t = \frac{1}{2\omega^2\left(\rho^2 - 1\right)}\left(\widetilde{e}^1 - \widetilde{e}^2\frac{\exp\left(2\widetilde{e}^2\left(T - t\right)\right) - \widetilde{e}^3}{\exp\left(2\widetilde{e}^2\left(T - t\right)\right) + \widetilde{e}^3}\right)$$

Hence, using the identity $\widetilde{e}_t = a_t - e_t^{cr}$ yields the desired result. ■

Following Proposition 3.4.3 and 3.4.6, the financial value of information for power, exponential and logarithmic utilites is given by:

$$\varrho_t = \left\{\begin{array}{cc} \frac{q}{1-q}\left(\frac{1}{2}\log\left(1 - 2D_t^*\gamma_t e_t^{cr}\right) + f_t\right) & q < 1, q \neq 0 \\ D_t^*\gamma_t e_t^{ca} + f_t \qquad , & q = 1 \\ \gamma_t a_t + g_t & q = 0 \end{array}\right.$$

which is easily calculable thanks to Lemmas 3.5.1 and 3.5.4.

Simulation Study We now present some numerical results relative to the computation of the *relative* cost of uncertainty $\widehat{x}_t/x$:

$$\widehat{x}_t/x = \left\{\begin{array}{cc} \exp\left(-\varrho_t^{cr}a^*\right) & \text{power} \\ 1 + \varrho_t^{ca}a^*/x \quad , & \text{exponential} \\ \exp(-\varrho_t^{\log}) & \text{logarithmic} \end{array}\right.$$

This represents the factor which allow to pass from the initial wealth under full information to this under partial information. Figure 3.6 represents this quantity in the case of exponential risk preferences when $x = 1$ and for various risk aversion parameter $a \in (0, 1)$. We observe that as time evolves, the cost of uncertainty (logically) decreases and increases when the risk aversion is near zero which is then close to the myopic (logarithmic) case.

It is well-known that the portfolio selection problem under exponential preferences is independent of the investor's initial wealth. To a certain extent, our

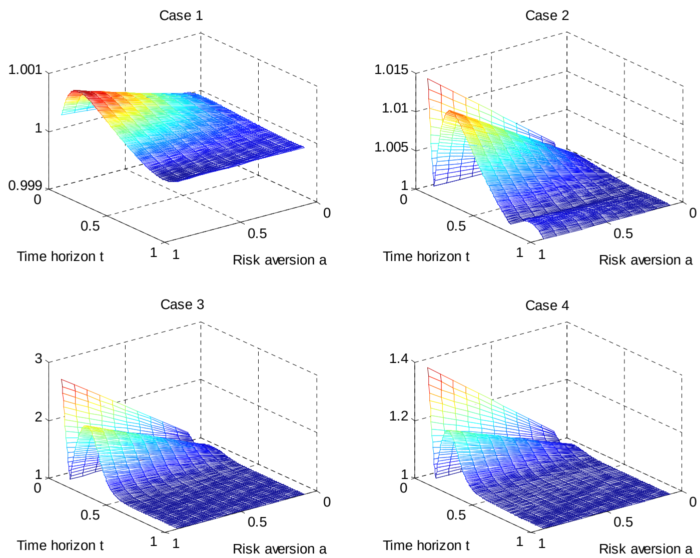

Figure 3.6: Relative cost of uncertainty $\widehat{x}_t/x$ - Exponential utility.

simulation study rediscovers this finding. In fact, the low cost of uncertainty for exponential investors may be due to the wealth invariance property of this utility function.

In the case of logarithmic utility, cf. Figure 3.7, we observe the same pattern but at a higher level. It seems that this higher cost of uncertainty is directly linked to the non-strategic behavior of the investor; a stylized fact which is typical of these risk preferences and which is recovered under the exponential utility setup in the limiting case $a \to 0$.

Finally, for the case of power utility, we consider again the four reference models and vary the risk aversion parameter in the interval $[-5, 0.8]$ which in the extremes correspond respectively to a *conservative* agent and an *aggressive* one. We note that the cost of uncertainty is higer than in the exponential case but lower than for logarithmic preferences.

3.6. Conclusion

This chapter has investigated the question of the financial value of information in a utility-based manner. Working with a two-factor correlated model for

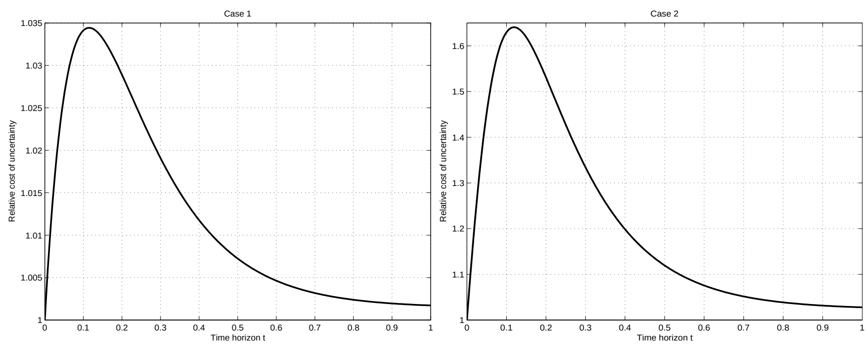

Figure 3.7: Relative cost of uncertainty $\widehat{x}_t/x$ - Logarithmic utility.

stock price and dividend, we demonstrate how one can analyze the cost of uncertainty when using a partially observable version of this model relatively to an investor with full information. From an economical perspective, our results are of interest for data providers which aim to assess the monetary value of their databases and for practitioners who have to decide between the relative costs of learning or subscribing. The financial value of information is then the *fair* price at which the two agents agree to transact.

From a theoretical perspective, the *distortion* technique has been used to derive a solution of the primal and dual problems and we have proved that the related value functions can be expressed in terms of the solution of a semilinear PDE. Moreover, we are able to treat different preferences in a unified framework using the $q-$parameter which is related to the definition of the optimal measure in the dual formulation of the portfolio optimization problem. Also, the convenient framework of our linear Gaussian model allow us to obtain tractable versions of the filter equations from the so-called Kalman filter. Eventually, we completely characterize the optimal investment strategy, the wealth process and the risk-neutral martingale measure in terms of the elicited semilinear PDE. On a fundamental perspective, as our approach is largely dependend on these hypotheses, it cannot be extended to other specifications. In fact, it is not clear how to simplify the HJB equation in the presence of more factors (such as a stochastic volatility) and to obtain easy-to-use filtering equations in the presence of nonlinear dynamics (such as a CEV model). The same remarks hold true in the more complicated case of a Lévy model.

Nevertheless, to obtain a more satisfactory model for empirical applications, further extensions would be necessary. For example, time-dependent parameters would surely more precisely describe the time-inhomogeneity of the series and stochastic interest rates could also be considered.

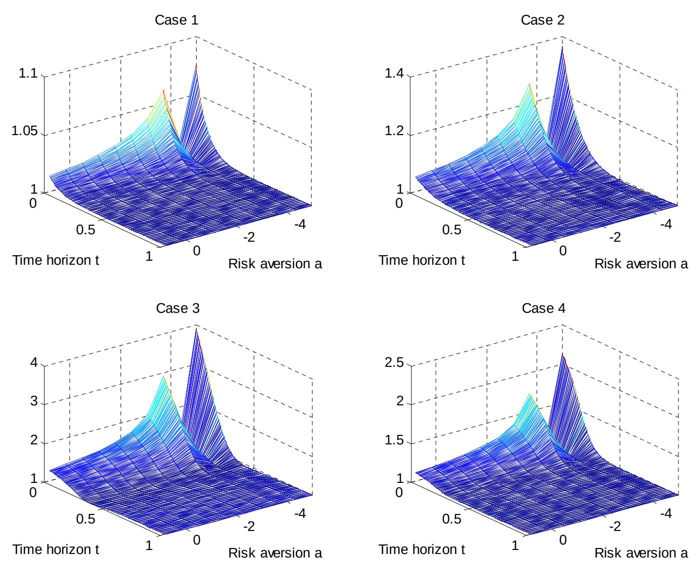

Figure 3.8: Relative cost of uncertainty $\widehat{x}_t/x$ - Power utility.

B

Appendix (Chapter 3)

Agenda

B.1. Optimal Strategies

In this Section, we derive, from (3.31), optimal strategies in the case of full and partial information. Eventually, this will allow to quantity two crucial terms: the hedging demand and the precautionary demand for uncertainty.

To motivate this, we note that, under logarithmic utility, direct substitution of m for θ in (3.29) does not change the formula for the optimal portfolio:

$$\widehat{\phi}_t^{\log} = A\left(x\right)\Sigma^{S*}\left(\mu_t - \vartheta_t\right)$$

This does not hold true for CRRA ($q < 1, q \neq 0$) and CARA ($q = 1$) utilities as learning affects strategies via the covariance matrices Φ and Υ, cf. (3.17).

B.1.1 Full Information Case

This corresponds to the case of a $\mathbb{F}-$financial market.

Proposition B.1.1 *For a $\mathbb{F}-$financial market and $q \leqslant 1, q \neq 0$, we have:*

$$\widehat{\phi}_t^{\mathbb{F}} = A\left(x\right)\Sigma^{S*}\left\{\mu_t - \theta_t\left(\mathbf{1}_m - 2\Psi^{\theta S}a_t\right) + \Psi^{\theta S}b_t\right\}$$

It one wants to compare the optimal strategy when $q \leqslant 1, q \neq 0$ from the *myopic* case $q = 0$, one notes from (3.31) that:

$$\widehat{\phi}_t^{\mathbb{F}} = \widehat{\phi}_t^{\log} + A(x)\,\Sigma^{S*}\Psi^{\theta S} v_\theta\,(t, \mu - \theta)$$

where the first term $(\mu_t - \theta_t)$ (in $\widehat{\phi}_t^{\log}$) is the (traditional) *myopic* mean-variance term while the second (new) term $\Psi^{\theta S} v_\theta\,(t, \mu - \theta)$ represents the hedging demand against fluctuations in the market price of risk.

To complete the picture, the ordinary differential equations (a_t, b_t, c_t), $t \in [0, T]$, given in Lemma 3.3.3, (3.24), are specialized using the notation (3.17):

$$\begin{aligned}
\frac{da_t}{dt} &= 2a_t\zeta_1^\theta a_t + 2\zeta_2^\theta a_t + \frac{q}{2}\Sigma^{S*} \\
\frac{db_t}{dt} &= 2a_t\zeta_1^\theta b_t + \zeta_2^\theta b_t + 2\zeta_3^\theta a_t - q\Sigma^{S*}\mu \\
\frac{dc_t}{dt} &= \frac{1}{2}b_t\zeta_1^\theta b_t + \zeta_3^\theta b_t - \mathrm{Tr}(\zeta_4^\theta a_t) + \frac{q}{2}\mathrm{Tr}(\Sigma^{S*}\mu^2)
\end{aligned} \tag{B.1}$$

where:

$$\zeta_1^\theta = q\Psi^{\theta S}\Sigma^{S*}\Psi^{\theta S} - \zeta_4^\theta,\ \zeta_2^\theta = \Lambda - q\Psi^{\theta S}\Sigma^{S*}$$

$$\zeta_3^\theta = q\Psi^{\theta S}\Sigma^{S*}\mu - \Lambda\Delta,\ \zeta_4^\theta = \Sigma^\theta + \Omega^\theta$$

Remark B.1.2 *When $q = 0$, (a, b) are given by (3.25) and c satisfies:*

$$\frac{dc_t}{dt} = -\Lambda\Delta b_t - Tr(\zeta_4^\theta a_t) - \frac{1}{2}Tr(\Sigma^{S*}\mu^2)$$

B.1.2 Partial Information Case

By following the same line of arguments, one should yield similar results for a $\mathbb{G}-$financial market. Nevertheless, we are here interested in a slightly different question: How one can pass from solutions of a $\mathbb{F}-$financial market to those of a $\mathbb{G}-$financial market? This will aim to study the impact of partial information.

We begin by presenting the proof of Lemma 3.3.6 that we recall here:

Lemma B.1.3 *The following identities, for $q \leqslant 1, q \neq 0$, hold:*

$$\begin{aligned}
\overline{a}_t &= (\mathbf{1}_m - 2\gamma_t a_t)^* a_t \\
\overline{b}_t &= (\mathbf{1}_m - 2\gamma_t a_t)^* b_t
\end{aligned}$$

Proof. Let $D_t = \mathbf{1}_m - 2\gamma_t a_t$. By direct computation, we have:

$$\begin{aligned}\frac{d\bar{a}_t}{dt} &= D_t^* \left[\frac{da_t}{dt} + 2a_t \frac{d\gamma_t}{dt} a_t\right] D_t^* \\ &= D_t^*[2a_t\{\zeta_1^\theta + \dot{\gamma}_t\}a_t + 2\zeta_2^\theta a_t + q/2\Sigma^{S*}]D_t^* \end{aligned} \tag{B.2}$$

We proceed by computing each term in (B.2). The third one gives:

$$\begin{aligned}&D_t^* \left\{q/2\Sigma^{S*}\right\} D_t^* \\ &= D_t^* \left\{q/2\Sigma^{S*}\left(1 + D_t^2 - D_t^2\right)\right\} D_t^* \\ &= D_t^* \left\{q/2\Sigma^{S*}\left(1 - D_t^2\right)\right\} D_t^* + q/2\Sigma^{S*} \\ &= D_t^* \left\{2q\Sigma^{S*}\gamma_t \left(1 - \gamma_t a_t\right) a_t\right\} D_t^* + q/2\Sigma^{S*} \triangleq T_3\end{aligned}$$

Noting that $\zeta_2^\theta - D_t\zeta_2^m = 2\gamma_t\zeta_2^m a_t - q\gamma_t\Sigma^{S*}$, the second term reads as:

$$\begin{aligned}&D_t^*\{2\zeta_2^\theta a_t\}D_t^* \\ &= D_t^*\{2(\zeta_2^\theta - D_t\zeta_2^m)a_t + D_t 2\zeta_2^m a_t\}D_t^* \\ &= D_t^*\{2(2\gamma_t\zeta_2^m a_t - q\gamma_t\Sigma^{S*})a_t\} + 2\zeta_2^m a_t D_t^* \triangleq T_2\end{aligned}$$

and similarly as $\zeta_1^\theta + \dot{\gamma}_t - \zeta_1^m = -2\gamma_t\zeta_2^m + q\gamma_t\Sigma^{s*}\gamma_t$, the first term yields:

$$\begin{aligned}&D_t^*\{2a_t(\zeta_1^\theta + \dot{\gamma}_t)a_t\}D_t^* \\ &= D_t^*\{2a_t(\zeta_1^\theta + \dot{\gamma}_t - \zeta_1^m)a_t + 2a_t\zeta_1^m)a_t\}D_t^* \\ &= D_t^*\{2a_t(-2\gamma_t\zeta_2^m + q\gamma_t\Sigma^{s*}\gamma_t)a_t\}D_t^* + D_t^* a_t\zeta_1^m a_t D_t^* \triangleq T_1\end{aligned}$$

Hence, we conclude:

$$T_1 + T_2 + T_3 = 2D_t^* a_t \zeta_1^m a_t D_t^* + 2\zeta_2^m a_t D_t^* + q/2\Sigma^{S*} = \frac{d\bar{a}_t}{dt}$$

The same calculations hold for $\bar{b}_t$ which conclude the proof. ■

We can now state the $\mathbb{G}-$optimal investment policy.

Proposition B.1.4 *For a $\mathbb{G}-$financial market and $q \leqslant 1, q \neq 0$, we have:*

$$\widehat{\phi}_t^{\mathbb{G}} = A(x)\,\Sigma^{S*}\left\{\mu_t - m_t\left(\mathbf{1}_m - 2\left(\Psi^{\theta S} - \gamma_t\right)\bar{a}_t\right) + \left(\Psi^{\theta S} - \gamma_t\right)\bar{b}_t\right\}$$

or using Lemma 3.3.6 and noting $D_t = 1 - 2\gamma_t a_t$:

$$\widehat{\phi}_t^{\mathbb{G}} = A(x)\,\Sigma^{S*}\left\{\mu_t - (m_t + \gamma_t b_t)\,D_t^*\left(\mathbf{1}_m - 2\Psi^{\theta S} a_t\right) + \Psi^{\theta S} b_t\right\}$$

Proof. The proof can be done by a direct substitution of Lemma 3.3.6. ■

This proposition allows then to separate the optimal investment strategy between a (traditional) *myopic* part which does not take into account the effect of learning and which is essentially equal to the one under full information when θ is replaced by its conditional version m, i.e. the mean-variance term, and a (new) *precautionary* part for uncertainty. So:

$$\widehat{\phi}_t^{\mathbb{G}} = A\left(x\right)\Sigma^{S*}\left\{\mu_t - m_t\left(\mathbf{1}_m - 2\Psi^{\theta S}a_t\right) + \Psi^{\theta S}b_t - \phi_t^h\right\}$$

where ϕ_t^h represents the *precautionary* term:

$$\phi_t^h = \left(2m_t a_t + b_t\right)D_t^*\gamma_t\left(\mathbf{1}_m - 2\Psi^{\theta S}a_t\right)$$

Remark B.1.5 *When $q = 0$, as $\overline{a}_t = a_t$ and $\overline{b}_t = b_t$, we have:*

$$u^{\mathbb{G}}\left(t, x, m\right) = U\left(x\right) + \left(m_t^{\intercal} a_t m_t + b_t^{\intercal} m_t + \overline{c}_t\right)$$

where (a, b) are given by (3.25) and $\overline{c}$ satisfies:

$$\frac{d\overline{c}_t}{dt} = -\Lambda\Delta b_t - Tr(\zeta_4^m a_t) - \frac{1}{2}Tr(\Sigma^{S*}\widetilde{\mu}^2)$$

B.2. Optimal Terminal Wealth

Relying on (3.18) for the solution of the primal problem, we can state the solution of the dual one via by resorting to arguments from duality theory, cf. Kramkov and Schachermayer (1999) [107] or Owen (2002) [135]. To this end, the conjugate $\widetilde{U}\left(\cdot\right)$ of the utility function $U\left(\cdot\right)$ satisfies:

$$\widetilde{U}\left(z\right) = \sup_{x\in\mathrm{dom}(U)}\left(U\left(x\right) - xz\right),\ z > 0 \tag{B.3}$$

For the three considered utility functions (3.15), $\widetilde{U} : \mathbb{R}^+ \to \mathbb{R}$ is given by:

$$\widetilde{U}\left(z\right) = \left\{\begin{array}{ccc} -z^q/q & & \text{power} \\ \left(z/a\right)\left(\log\left(z/a\right) - 1\right) & , & \text{exponential} \\ -\left(1 + \log z\right) & & \text{logarithmic} \end{array}\right. \tag{B.4}$$

and the dual value function of (3.14) is defined by:

$$\widetilde{u}\left(t, z, y\right) = \inf_{Q\in\mathcal{M}(\mathcal{H})} E^P\left[\widetilde{U}\left(z\frac{dQ}{dP}\right)|\vartheta_t = y\right] \tag{B.5}$$

where $\mathcal{M}\left(\mathcal{H}\right)$ is the set of equivalent martingale measures defined as:

$$\mathcal{M}\left(\mathcal{H}\right) = \left\{Q \sim P \mid S \text{ is a } \left(Q, \mathcal{H}\right) - \text{local martingale}\right\}$$

where $\mathcal{H} = \{\mathcal{F}, \mathcal{G}\}$ depending on the informational type of the considered financial market ($\mathbb{F}$ or $\mathbb{G}$), with density processes $dQ/dP|_{\mathcal{H}_t} = M_t,\ t \in [0,T]$. The following result is a ramification of Proposition 3.3.1.

Proposition B.2.1 *The dual value function $\widetilde{u}(t,z,y)$ is given by:*

$$\widetilde{u}(t,z,y) = \begin{cases} \widetilde{U}(z)\exp(v(t,y))^{(1-q)} & q<1, q\neq 0 \\ \widetilde{U}(z) - (z/a)\,v(t,y) \quad , & q=1 \\ \widetilde{U}(z) + v(t,y) & q=0 \end{cases} \tag{B.6}$$

Proof. From duality theory, we know that the dual value function is the convex conjugate of the primal value function, so:

$$\widetilde{u}(t,z,y) = \sup_{x\in \mathrm{dom}(U)} \left(u(t,x,y) - xz\right),\ z>0 \tag{B.7}$$

In particular, if x^* attains the supremum in (B.7), then $u_x(t,x^*,y) = z$. Resorting now to the solution of the primal problem (3.18), it follows:

$$z = \begin{cases} U'(x^*)\exp(v(t,y)) & q\leqslant 1, q\neq 0 \\ U'(x^*) + v(t,y) \quad , & q=0 \end{cases}$$

Or, equivalently:

$$x^* = \begin{cases} I(z/v(t,y)) & q\leqslant 1, q\neq 0 \\ I(z - v(t,y)), & q=0 \end{cases}$$

where $I(\cdot) = (U'(\cdot))^{-1}$. Inserting x^* into (B.7) and using the identity $\widetilde{U}(z) = U(I(z)) - zI(z)$, yields:

$$\widetilde{u}(t,z,y) = \begin{cases} \widetilde{U}(z/\exp(v(t,y)))\exp(v(t,y)) & q\leqslant 1, q\neq 0 \\ \widetilde{U}(z) + v(t,y) \quad , & q=0 \end{cases}$$

Eventually, the specific form of $\widetilde{U}$, cf. (B.4), yields the desired result. ∎

For all $t\in[0,T]$ and $y\in\mathbb{R}^m$, consider:

$$\psi_t \equiv \psi(t,y) = \begin{cases} \Sigma^{S*}\{(\mu_t - y) + \Phi v_y(t,y)\} & q\leqslant 1, q\neq 0 \\ \Sigma^{S*}(\mu_t - y) \quad , & q=0 \end{cases}$$

Then, the optimal portfolio process $\widehat{\phi}_t = \widehat{\phi}(t,\widehat{X}_t,\vartheta_t)$ is given by:

$$\widehat{\phi}_t = A(\widehat{X}_t)\psi_t \triangleq A(\widehat{X}_t)\psi(t,\vartheta_t),\ \vartheta = \{\theta, m\} \tag{B.8}$$

We are now in position to characterize the optimal terminal wealth.

Lemma B.2.2 *The optimal terminal wealth is given by:*

$$\widehat{X}_T = \begin{cases} x \exp\left\{ \overline{q} \left(\int_0^T \psi_t dR_t - \frac{\overline{q}}{2} \int_0^T \Sigma^S \psi_t \Sigma^S \psi_t^\intercal dt \right) \right\} \\ x + \frac{1}{a} \int_0^T \psi_t dR_t \\ \exp\left(\int_0^T \psi_t dR_t \right) \end{cases}, \quad \begin{matrix} q < 1, q \neq 0 \\ q = 1 \\ q = 0 \end{matrix} \tag{B.9}$$

with $\overline{q} = 1 - q$.

Proof. From (3.10) and (B.8), we have:

$$d(\widehat{X}_t) = A(\widehat{X}_t)\psi_t dR_t$$

where we have used the relation $dR_t = dS_t/S_t = (\mu_t - \vartheta_t)\, dt + \sigma_S dW_t^\vartheta$, with the notation (3.11). The Arrow-Pratt risk-aversion is given by:

$$A\,(x) = \begin{cases} x\overline{q} \\ 1/a \end{cases}, \begin{matrix} q < 1, q = 0 \\ q = 1 \end{matrix} \quad \text{with } \overline{q} = \begin{cases} 1 - q \\ 1 \end{cases}, \begin{matrix} q < 1, q \neq 0 \\ q = 0 \end{matrix}$$

Hence, for $q < 1, q = 0$, by using Itô's formula, it follows:

$$\log(X_T) = \log(X_0) + \overline{q} \left(\int_0^T \psi_t dR_t - \frac{\overline{q}}{2} \int_0^T \psi_t \psi_t^\intercal d\,[R]_t \right)$$

which is the desired result since $d\,[R]_t = \Sigma^S \Sigma^S dt$. When $q = 1$, we have:

$$X_T = X_0 + \frac{1}{a} \int_0^T \psi_t dR_t$$

from which we conclude the proof. ■

Remark B.2.3 *The formula (B.9) holds true for both $\mathbb{F}$ and $\mathbb{G}$ financial markets. To specialize the result for each one, we can rely on Proposition B.1.1 for the case of full information and Proposition B.1.4 for the case of partial information. These propositions also allow to pass from $\psi^{\mathbb{F}}$ to $\psi^{\mathbb{G}}$, thus proving the existing link between the terminal optimal wealths $\widehat{X}_T^{\mathbb{F}}$ and $\widehat{X}_T^{\mathbb{G}}$.*

To complete the picture, we compute the likelihood ratio of the dual martingale measure optimizer of problem (B.5). To do so, we rely on the classical relation between the dual optimizer and the optimal terminal wealth:

$$\widehat{X}_T = I\left(z \frac{d\widehat{Q}}{dP} \right), \; z > 0 \tag{B.10}$$

where $I(\cdot) = (U^{'}(\cdot))^{-1}$ is given by:

$$I(z) = \begin{cases} z^{-(1-q)} & \text{power} \\ -\log(z/a)/a & , \text{ exponential} \\ 1/z & \text{logarithmic} \end{cases}$$

and z is related to the initial wealth x by $z = u_x(0, x, y)$ or $x = -\widetilde{u}_z(0, z, y)$.

Lemma B.2.4 *The Radon-Nikodym derivative of the dual optimizer satisfies:*

$$\log \frac{d\widehat{Q}}{dP} = \begin{cases} -v(0, y) - \int_0^T \psi_t dR_t + \frac{1-q}{2} \int_0^T \Sigma^S \psi_t \Sigma^S \psi_t^{\intercal} dt & q < 1, q \neq 0 \\ -\frac{1}{a} v(0, y) - \frac{1}{a} \int_0^T \psi_t dR_t & , \quad q = 1 \\ -\int_0^T \psi_t dR_t & q = 0 \end{cases} \tag{B.11}$$

Proof. We demonstrate the result for $q < 1, q \neq 0$ (power utility). The proofs for other preferences follow the same lines of reasoning. Using (B.6) and as $x = -\widetilde{u}_z(0, z, y)$, we have $x = z^{-(1-q)} \exp(-v(0, y))^{-(1-q)}$, so that (B.9) becomes:

$$\widehat{X}_T = z^{-(1-q)} \exp\left\{-v(0, y) - \int_0^T \psi_t dR_t + \frac{1-q}{2} \int_0^T \psi_t \psi_t^{\intercal} d[R]_t\right\}^{-(1-q)}$$

which is equal to (B.10) since $I(z) = z^{-(1-q)}$. ∎

Remark B.2.5 *As for the terminal optimal wealth, the Radon-Nikodym derivative (B.11) holds true for both $\mathbb{F}$ and $\mathbb{G}$ financial markets. The specialization for each one is ensured by $\psi^{\mathbb{F}}$ and $\psi^{\mathbb{G}}$ and rely on Propositions B.1.1 and B.1.4.*

B.3. Practical Considerations

In this Section, we present some computations and considerations which are related to the application presented in Section 3.5.

B.3.1 Expressions for Filters

To cast previous notation, we note:

$$\Sigma^S = \sigma^2, \ \Sigma^\theta + \Omega^\theta = \omega^2, \ \Psi^{\theta S} = \rho\sigma\omega, \ \widetilde{\mu} = \mu - r$$

Then, the filter equations are given by:

$$dm_t = \lambda(\delta - m_t) dt + (\rho\sigma\omega - \gamma_t) \xi^2 (dR_t - (\mu - m_t) dt), \ m_0 \in \mathbb{R} \tag{B.12}$$

$$\frac{d\gamma_t}{dt} = \omega^2 - 2\lambda\gamma_t - \xi^2 \left(\rho\sigma\omega - \gamma_t\right)^2, \ \gamma_0 = 0 \tag{B.13}$$

where $\xi = \sigma^*$.

We obtain m_t as the solution of (B.12). Letting:

$$\Phi(t) = \exp\left(-\lambda t - \int_0^t \upsilon_s ds\right), \ \Phi(0) = 1$$

then, m_t is determined, for $t \in [0, T]$, cf. Kloeden and Platen (1999) [105], via:

$$m(t) = \Phi_t \left[m_0 + \int_0^t \Phi_s^* \upsilon_s \left(dY_s - (\mu - \frac{1}{2}\sigma^2) ds \right) + \lambda\delta \int_0^t \Phi_s^* ds \right] \tag{B.14}$$

with $\upsilon_s = \rho\xi\omega - \xi^2\gamma_s$. Also, γ_t, cf. (B.13), has an explicit solution which is:

$$\gamma(t) = \frac{\gamma_2}{\xi^2} \frac{\exp(2\gamma_2 t) - \gamma_3}{\exp(2\gamma_2 t) + \gamma_3} - \frac{\gamma_1}{\xi^2} \tag{B.15}$$

where:

$$\gamma_1 = \lambda - \rho\xi\omega, \ \gamma_2 = \sqrt{\lambda^2 + \omega^2\xi^2 - 2\lambda\rho\xi\omega}, \ \gamma_3 = \frac{\gamma_2 - \gamma_1}{\gamma_2 + \gamma_1} \tag{B.16}$$

We also note that Φ_t can be written in the more convenient form:

$$\Phi(t) = \frac{(1 + \gamma_3)\exp(\gamma_2 t)}{\exp(2\gamma_2 t) + \gamma_3}, \ t \in [0, T]$$

B.3.2 Solutions for ODE's

Expressions for (a, b, c)

For the case of full information, from (B.1), ODE's (a, b, c) satisfy:

- $q \leqslant 1, q \neq 0$:

$$\dot{a}_t = 2\omega^2 \left(q\rho^2 - 1\right) a_t^2 + 2\left(\lambda - q\rho\xi\omega\right) a_t + \frac{q}{2}\xi^2$$
$$\dot{b}_t = 2\omega^2 \left(q\rho^2 - 1\right) a_t b_t + \left(\lambda - q\rho\xi\omega\right) b_t + 2\left(q\rho\xi\omega\widetilde{\mu} - \lambda\delta\right) a_t - q\xi^2\widetilde{\mu}$$
$$\dot{c}_t = \frac{1}{2}\omega^2 \left(q\rho^2 - 1\right) b_t^2 + \left(q\rho\xi\omega\widetilde{\mu} - \lambda\delta\right) b_t - \left(\rho\sigma\omega\right) a_t + \frac{q}{2}\xi^2\widetilde{\mu}^2$$

- $q = 0$:

$$\dot{a}_t = 2\lambda a_t - \frac{1}{2}\xi^2, \ \dot{b}_t = \lambda b_t - 2\lambda\delta a_t + \xi^2\widetilde{\mu}, \ \dot{c}_t = -\lambda\delta b_t - \omega^2 a_t - \frac{1}{2}\xi^2\widetilde{\mu}^2$$

Also, the equations (e^{cr}, e^{ca}, f, g) satisfy:

$$\begin{aligned}\dot{e}_t^{cr} &= -\frac{1-q}{2}\xi^2\left(1-2\rho\sigma\omega a_t\right)^2 + 4a_t e^1 e_t^{cr} + 2e^2 e_t^{cr} - 2e_t^{cr} e^1 e_t^{cr} \\ \dot{e}_t^{ca} &= -\frac{1}{2}\xi^2\left(1-2\rho\sigma\omega a_t\right)^2 + 4a_t e^1 e_t^{ca} + 2e^2 e_t^{ca}\end{aligned}$$

with:

$$e^1 = -\omega^2\left(1-\rho^2\right),\; e^2 = \lambda - \rho\xi\omega$$

and:

$$\dot{f}_t = -\omega^2\left(1-\rho^2\right)e_t,\; \dot{g}_t = \frac{1}{2}\xi^2\gamma_t$$

Case $q \leqslant 1, q \neq 0$

Subsequently, we give more insights on Lemma 3.5.1.

Function a Another representation for a_t is given by:

$$a\left(t\right) = \frac{a_2}{a_3}\tanh\left(a_2\left(T-t\right)+\widetilde{a}_4\right) + \frac{a_1}{a_3}$$

with:

$$\widetilde{a}_4 = \frac{1}{2}\log\left(\frac{a_2-a_1}{a_2+a_1}\right)$$

Then, we note that $a_2 > a_1$ as long as $q\rho^2 < 1$. Thus, for the case $q < 1, q \neq 0$, we apply the modification:

$$q = \begin{cases} 1-a & , \quad a \in (0,1) \\ a/\left(a-1\right) & \quad a \in (-\infty, 0)\end{cases}$$

Therefore, $a_2 > |a_1|$ and thus $\widetilde{a}_4$ is well defined for all choices of the parameters. Also, as $a_t^{'} < 0$ for all $t \in [0,T]$ and $a\left(T\right) = 0$, a_t is non-negative and monotonically non-increasing. It is then clear that b_t and c_t will be non-positive and monotonically non-decreasing since $b\left(T\right) = 0$ and $c\left(T\right) = 0$.

Remark B.3.1 *In the case $q = 0$, the same analysis can be conducted for functions (a, b, c) and the conclusions are similar.*

Function b We begin with three results useful in computations.

Lemma B.3.2 *For $u \in [t, s]$, let $A_t^s = \exp\left(-a_1\left(s-t\right) + a_3\int_t^s a_u du\right)$, then:*

$$A_t^s = \exp\left(a_2\left(t-s\right)\right)\frac{\exp\left(2a_2\left(T-t\right)\right)+a_4}{\exp\left(2a_2\left(T-s\right)\right)+a_4}$$

Lemma B.3.3 *For $t \in [0,T]$, let $A_2(t) = \int_t^T A_t^s ds$, then:*

$$A_2(t) = \frac{\exp(a_2(t-T))}{a_2\sqrt{a_4}}(n(0,0) - n(t,T))(a_4 + \exp(2a_2(T-t)))$$

with:

$$n(x,y) \triangleq \arctan\left(\sqrt{a_4}\exp(a_2(x-y))\right)$$

Lemma B.3.4 *For $t \in [0,T]$, let $A_1(t) = \int_t^T A_t^s a_s ds$, then:*

$$\begin{aligned} A_1(t) = &\frac{\exp(a_2(3T-t)) + a_4\exp(a_2(T+t))}{d(T,T)} - \frac{\exp(2a_2T) + a_4\exp(2a_2t)}{d(t,T)} \\ &+ \frac{a_1(n(0,0) - n(t,T))}{a_2a_3}\left(\frac{\exp(a_2(T-t))}{\sqrt{a_4}} + \exp(a_2(t-T))\sqrt{a_4}\right) \end{aligned}$$

with:

$$d(x,y) \triangleq a_3(a_4\exp(2a_2x) + \exp(2a_2y))$$

With the help of the previous lemmas, we can now derive an explicit solution for b_t. For $t \in [0,T]$, we have:

$$b(t) = 2a_5A_1(t) - q\xi^2\widetilde{\mu}A_2(t)$$

Expressions for $\left(\overline{a}, \overline{b}, \overline{c}\right)$

To complete the picture, for the case of partial information from (3.27), we have when $q \leqslant 1, q \neq 0$:

$$\begin{aligned} \dot{\overline{a}}_t &= 2(q-1)\Upsilon^m\overline{a}_t^2 + 2(\lambda - \Lambda^m)\overline{a}_t + \frac{q}{2}\xi^2 \\ \dot{\overline{b}}_t &= 2(q-1)\Upsilon^m\overline{a}_t\overline{b}_t + (\lambda - \Lambda^m)\overline{b}_t + 2(\Lambda^m\widetilde{\mu} - \lambda\delta)\overline{a}_t - q\xi^2\widetilde{\mu} \\ \dot{\overline{c}}_t &= \frac{1}{2}(q-1)\Upsilon^m\overline{b}_t^2 + (\Lambda^m\widetilde{\mu} - \lambda\delta)\overline{b}_t - \Upsilon^m\overline{a}_t + \frac{q}{2}\xi^2\widetilde{\mu}^2 \end{aligned}$$

with:

$$\Upsilon^m = (\rho\sigma\omega - \gamma_t)^2\xi^2, \ \Lambda^m = q(\rho\sigma\omega - \gamma_t)\xi^2$$

while when $q = 0$, (a, b) are invariant and we have:

$$\dot{\overline{c}}_t = -\lambda\delta b_t - \Upsilon^m a_t - \frac{1}{2}\xi^2\widetilde{\mu}^2$$

Nevertheless, solutions for $\left(\overline{a}, \overline{b}, \overline{c}\right)$ cannot be derived as easily as for (a, b, c). Thus, we use the identities elicited in Lemmas 3.3.6, 3.4.2 and 3.4.5.

Lemma B.3.5 *The equations* $\left(\overline{a},\overline{b},\overline{c}\right)$ *satisfy:*

$$\overline{a}_t \ / \ \overline{b}_t = \begin{cases} D_t^* a_t \ / \ D_t^* b_t & , \quad q \leqslant 1, q \neq 0 \\ a_t \ / \ b_t & \quad\quad q = 0 \end{cases}$$

and:

$$\overline{c}_t - c_t - d_t = \begin{cases} \frac{q}{q-1}\left(\frac{1}{2}\log\left(1 + 2D_t^*\gamma_t e_t^{cr}\right) + f_t\right) & q < 1, q \neq 0 \\ D_t^*\gamma_t e_t^{ca} + f_t & , \quad q = 1 \\ c_t + \gamma_t a_t + g_t & q = 0 \end{cases}$$

with $D_t = \mathbf{1}_m - 2\gamma_t a_t$ *and* $d_t = \frac{1}{2}\left(D_t^*\gamma_t b_t^2 - \log D_t\right)$.

4

Optimal Policies from Discrete Prices

Agenda

Abstract. We study the questions of optimal portfolios and hedging strategies in a market where the asset log-price Y follows a diffusion model whose coefficients are unobservable and are given in terms of a Markov process θ. This leads naturally to a partial information setup, where the strategies are measurable w.r.t. the observations. Nevertheless, the proposed framework departs from the usual one as the stock price is observed only discretely at random times $(\tau_n)_{n\geqslant 1}$. While quite natural, these hypotheses lead to special optimization and filtering approaches. In the above setting, the investor problem can be approached by considering that the wealth process is subject to shocks produced by a multivariate point process given by the observations $(\tau_n, \Delta Y_{\tau_n})_{n\geqslant 1}$. And from a filtering perspective, the optimal filter for θ based on these point observations requires a thorough treatment. Eventually, by resorting to the martingale approach of the stochastic control problem through duality and use of Malliavin calculus for random measures, we fully characterize the investment policy process. To this end, we derive an extension of Clark's formula for random measures under an equivalent change of measure. As examples, optimal portfolios are given for power, logarithmic and exponential utility functions and hedging strategies for contingent claims will be derived.

4.1. Introduction

Understanding the joint dynamics of the stock price and of the investment policy processes is of paramount importance. In fact, the increasing availability of high frequency data allows investors to analyze how the complex interactions among various market participants affect investment decisions. Empirical researchers, however, face challenges when analyzing irregularly and frequently-sampled data, cf. Ait-Sahalia and Mykland (2004) [4]. From a statistical perspective, tools employed in time-series analysis are not well-suited while the lack of a theoretical framework applicable to utility maximization under transaction level data is more serious. The present work attempts to reconcile continuous-time modeling and discrete-time observations in tackling the utility maximization problem.

This objective will be meet by putting two hypotheses. First, we propose a continuous-time model of stock prices where the coefficients - drift, volatility, jumps times and sizes - of the asset log-price Y are unobservable and driven by a strong Markov process θ. Under this setting, θ represents the latent market factors that affect Y and that are not captured by its modeling, cf. Runggaldier (2004) [146]. This difference matters as while the randomness generated by Y captures the systematic risk attached to the asset price, the random shocks induced by θ represent the idiosyncratic risk that cannot be perfectly hedged. Second, as in Frey and Runggaldier (2001) [71], we suppose that the vector of stock prices is observed only discretely at random times $(\tau_n)_{n \geqslant 1}$. This assumption is prompted so as to reflect the discrete nature of high-frequency data, see Andersen et al. (2003) [6].

From now on, we note that the present paper is in sharp contrast to most of the previous litterature. The optimization problem with full information goes back to Merton (1971) [126] who solved the question via the Bellman equation of dynamic programming. For the case of complete markets, a rigorous mathematical treatment is presented in Karatzas et al. (1987) [98]. Models with incomplete information have been investigated by Detemple (1986) [48] or Lakner (1995, 1998) [122], [111] who solved the optimization problem via a martingale approach, provided characterization of optimal strategy via the Malliavin calculus on the Wiener space and worked out the special case of the linear Gaussian model. Under this framework, partially observable market models are only uncertain in the growth rate which is altered by shocks whose magnitude cannot be distinguished from other sources of randomness, as deeply investigated by Pham and Quenez (2001) [139]. The key mathematical tool of this approach is the Girsanov theorem for semimartingales (cf. Jacod and Shiryaev, 2003 [93]) which allows to perform absolutely continuous changes of measures, thus avoiding learning in the volatility or jump components of the stock price process, cf. Jeanblanc, Lacoste and Roland (2005) [95] for more details on this approach.

While quite natural, the previous hypotheses lead to non standard optimization and filtering approaches. In the above setting, the investor problem can be approached by considering that the wealth process is subject to shocks produced by a multivariate point process given by the observations $(\tau_n, \Delta Y_{\tau_n})_{n \geqslant 1}$. From a financial perspective, this implies that the market becomes incomplete. Moreover, the optimal filter for θ based on these point observations is out of the scope of classical filtering techniques, cf. Liptser and Shiryaev (2001) [118], and has to be solved as a non-linear filtering problem. To this end, we prove a Kallianpur-Striebel type equation for both the normalized and unnormalized filters.

As the market we consider is incomplete, we follow the common approach used in derivative pricing or hedging to base the prices or hedges on a minimal distance martingale measure, as proposed by Delbaen and Schachermayer (1996) [44]. Further, Kramkov and Schachermayer (1999) [107] show that an optimal portfolio can be expressed by the solution of a dual variational problem which is related to those of finding a minimal distance martingale measure. In the present paper, we exploit this *martingale technique of portfolio optimization* to derive explicit representations of the investment policy and hedging strategy processes in terms of stochastic integrals and Malliavin derivatives for random measures, following a version of this calculus introduced and developped by Lokka (2003) [121]. Central in obtaining such formulae is an extension of the Clark-Ocone-Haussmann formula (Clark formula henceforth), Clark (1970) [30], Ocone (1984) [133], Haussmann (1979) [87], for random measures. Moreover, in the context of portfolio optimization problems, we have to look at the Clark formula under an equivalent change of measure. This derivation has previously been done by Ocone and Karatzas (1991) [134] for Itô processes. In this paper, we derive such a formula for random measures. As examples, we provide solutions of the Girsanov quantities appearing in the determination of the minimal distance martingale measures for the three more standard utility functions: power, logarithmic and exponential.

The remainder of this paper is organized as follows. Section 2 states the framework and gives some preliminary results on information structure, martingales measures and portfolio optimization. Section 3 provides the filtering setup in a multivariate point process framework and detail the filter equations attached with it. Section 4 then gives our main results which are the extension of Clark formula and the derivations of investment policy and hedging strategy processes. Special attention will be given to the three standard utility fonctions and quadratic hedging of Asian options will be considered as a case study. We conclude in Section 5.

4.2. Model and Assumptions

4.2.1 The Economy

On the filtered probability space $\left(\Omega, \mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}, P\right)$ - all stochastic processes being $(\mathcal{F}_t)_{t\geqslant 0}$ −adapted, we consider, on $[0,T]$ with $T \in (0,\infty)$, a financial market consisting of two assets. The riskless security reads:

$$B_t = \exp\left(\int_0^t r_s ds\right)$$

where $(r_t, t\in[0,T])$ is a uniformly bounded process. The risky asset satisfies:

$$\bar{S}_t = \bar{S}_0 e^{\bar{Y}_t},\ \bar{S}_0 \in \mathbb{R}^+ \tag{4.1}$$

with $\bar{Y}_t$ of the form:

$$\bar{Y}_t = \int_0^t \left(b\left(u,\theta_u\right) - \frac{1}{2}\sigma^2\left(u,\theta_u\right)\right) du + \int_0^t \sigma\left(u,\theta_u\right) dW_u \tag{4.2}$$

where $(\theta_t, t\in[0,T])$ stands for an economic factor process, which is in general not observable. W is a standard Brownian motion, while b and σ are bounded and continuous functions on $\mathbb{R}^+\times\mathbb{R}$ with values in $\mathbb{R}$ and $\mathbb{R}^+$ respectively.

From this setup, the filtration $\mathcal{F}$ will denote the *model* information, i.e.: the filtration on which are defined all the stochastic processes.

4.2.2 The Information Structure

We now consider the situation where the agent has not access to the filtration $\mathcal{F}$ but only observes the risky price $\bar{S}$. This is the classical partial information framework, as studied by Detemple (1986) [48] or Lakner (1995, 1998) [122], [111]. Furthermore, we depart from this setup by considering as in Frey and Runggaldier (2001) [71] that the prices $\bar{S}$ are only available at random times $(\tau_n)_{n\geqslant 1}$. As $\mathcal{F}^{\bar{S}} = \mathcal{F}^{\bar{Y}}$, we will denote by $(\mathcal{G}(n))_{n\geqslant 1}$ the *market* filtration:

$$\mathcal{G}(n) = \left(\sigma\left(\tau_k, \Delta\bar{Y}_{\tau_k}\right), \tau_k \leqslant \tau_n\right)$$

where $\Delta\bar{Y}_{\tau_n} = \bar{Y}_{\tau_k} - \bar{Y}_{\tau_{k-1}}$. The observations process $\left(\tau_n, \Delta\bar{Y}_{\tau_n}\right)_{n\geqslant 1}$ is a multivariate marked point process, cf. Proposition 4.2.2, with the counting measure:

$$\mu(dt, dy) = \sum_{n\geqslant 1} \delta_{\{\tau_n, \Delta\bar{Y}_{\tau_n}\}}(dt, dy)$$

From this, we introduce the *continuous* counterpart of $(\mathcal{G}(n))_{n\geqslant 1}$, say:

$$\mathcal{G}_t = \sigma\left(\mu\left([0,s], \mathcal{X}\right), s \leqslant t, \mathcal{X} \in \mathcal{B}(\mathbb{R})\right)$$

and we note that $\mathcal{G}_{\tau_n} = \mathcal{G}(n)$ for any $n \geqslant 1$. Besides, we shall consider the filtration $\mathcal{F}^\theta_t = \sigma(\theta_s, s \leqslant t)$ generated by the state variable θ as well as:

$$\mathcal{H}_t = \mathcal{F}^\theta_t \vee \mathcal{G}_t$$

As only the filtration $\mathcal{G}$ is available, θ is not observable. Therefore, we introduce the $\mathcal{G}-$conditional density of the random variable θ by:

$$\pi_t(f) = E^P\left[f(\theta_t)\,|\mathcal{G}_t\right]$$

for any $\mathbb{R}-$valued measurable function f s.t. $E^P\left[|f(\theta_t)|\right] < \infty$.

Remark 4.2.1 *In Appendix C.1, we present a possible parameterization for the economic factor process θ. And in Appendix C.1.2, we derive the semi-martingale representation of π for this setup. This is of some interest if one is willing to handle the numerical/statistical part of the problem.*

Observations Process

Let $(N_t, t \in [0,T])$ be the counting process with interarrival times $(\tau_n - \tau_{n-1})_{n\geqslant 1}$ with $\tau_0 = 0$. We suppose that N_t admits an $\mathcal{H}_t-$measurable intensity $\lambda(t, \theta_t)$ being a bounded and non-negative function. Let:

$$N_t = \sum_{n\geqslant 1} \mathbf{1}_{\{\tau_n \leqslant t\}}$$

and note that $\mathcal{G}_t = \mathcal{G}(N_t)$. We then define the *discrete* log asset price by:

$$Y_t \triangleq \bar{Y}_{\tau_{N_t}} \tag{4.3}$$

Suppose now that the process N jumps at t. By (4.3), Y has a jump of size $\bar{Y}_t - \bar{Y}_{\tau_{N_{t-}}}$. Then, by (4.2), we know that, conditionally on $\mathcal{H}_t$, this quantity is normally distributed with mean $\widehat{b}_t \equiv \widehat{b}\left(\tau_{N_{t-}}, t\right)$ and variance $\widehat{\sigma}_t^2 \equiv \widehat{\sigma}^2\left(\tau_{N_{t-}}, t\right)$, what we note $\phi_t(dx) \equiv \phi_{\tau_{N_{t-}},t}(dx) \equiv \phi\left(\widehat{b}_t, \widehat{\sigma}_t^2; dx\right)$, where:

$$\widehat{b}(s,t) = \int_s^t \left(b(u,\theta_u) - \frac{1}{2}\sigma^2(u,\theta_u)\right) du, \;\; \widehat{\sigma}^2(s,t) = \int_s^t \sigma^2(u,\theta_u)\, du$$

We now define in a more concise way the observations process $\left(\tau_n, \Delta\bar{Y}_{\tau_n}\right)_{n\geqslant 1}$. Following Brémaud (1981) [20], it can be viewed as a multivariate point process.

Proposition 4.2.2 *Denote by μ the integer valued random measure associated to the observations $\left(\tau_n, \Delta\bar{Y}_{\tau_n}\right)_{n\geqslant 1}$, s.t. $\forall t \in [0,T]$:*

$$\mu(dt, dy) = \sum_{n\geqslant 1} \delta_{\{\tau_n, \Delta\bar{Y}_{\tau_n}\}}(dt, dy)$$

Also, the $(P, \mathcal{H}_t)-$predictable compensator of μ is given by:

$$\nu_{\mathcal{H}}^P(dt, dy) = \lambda_t(\theta_t)\, \phi_t(dy)\, dt$$

where $\phi_t(dy)$ is the mark of the marked point process and $\lambda_t(\theta_t) = \lambda(t, \theta_{t-})$ is the $(P, \mathcal{H}_t)-$ intensity of the point process N_t.

From these, we recast (4.3) in the line of (4.1), say $S_t = S_0 e^{Y_t}$, so that:

$$S_t = S_0 + \int_0^t S_{u-} \int_{\mathcal{X}} (e^x - 1)\, \mu(du, dx) \tag{4.4}$$

$\mathcal{G}-$Characteristics

The joint distribution of $\left(\tau_{n+1}, \bar{Y}_{\tau_{n+1}}\right)$, given $\mathcal{F}_t^{\theta} \vee \mathcal{G}(n)$ with $\tau_{n+1} < t$, writes:

$$P\left(\tau_{n+1} \leqslant t, \bar{Y}_{\tau_{n+1}} \leqslant y | \mathcal{F}_t^{\theta} \vee \mathcal{G}(n)\right) = \left(1 - e^{-\Lambda_n(t)}\right) I_n(y) \tag{4.5}$$

with:

$$\Lambda_n(t) = \int_{\tau_n}^{t} \lambda_s(\theta_s)\, ds,\ \ I_n(x) = \int_{-\infty}^{x} \phi_{\tau_n,t}(z - Y_{\tau_n})\, dz$$

We introduce:

$$\psi_n(f; t, x) = E^P\left[f(\theta_t)\, e^{-\Lambda_n(t)} \phi_{\tau_n,t}(x - Y_{\tau_n})\, |\mathcal{F}_{\tau_n}^{\theta}\right]$$

$$\overline{\psi}_n(f; t) = \int_{\mathbb{R}} \psi_n(f; t, x)\, dx = E^P\left[f(\theta_t)\, e^{-\Lambda_n(t)} |\mathcal{F}_{\tau_n}^{\theta}\right]$$

Lemma 4.2.3 *The $(P, \mathcal{G})-$compensator of μ is given by:*

$$\nu_{\mathcal{G}}^P(dt, dy) \triangleq \gamma(\lambda; t, y)\, dtdy = \sum_{n \geqslant 0} \mathbf{1}_{]\tau_n, \tau_{n+1}]}(t)\, \frac{\pi_{\tau_n}(\psi_n(\lambda; t, y))}{\pi_{\tau_n}\left(\overline{\psi}_n(\lambda; t)\right)} dtdy$$

where $\pi_t(f) = E^P[f(\theta_t)\, |\mathcal{G}_t]$, cf. Section C.1.1.

Proof. Owing to (4.5), the regular conditional distribution of $\left(\tau_{n+1}, \bar{Y}_{\tau_{n+1}}\right)$ given $\mathcal{G}(n)$, say $G_n(t, y) = \frac{1}{dt}\frac{1}{dy} P\left(\tau_{n+1} \leqslant t, \bar{Y}_{\tau_{n+1}} \leqslant y | \mathcal{G}(n)\right)$ is an increasing function. From Proposition 3.4.1 of Lipster and Shiryaev (2001) [117], we have:

$$\nu_{\mathcal{G}}^P(dt, dy) = \sum_{n \geqslant 0} \mathbf{1}_{]\tau_n, \tau_{n+1}]}(t)\, \frac{G_n(dt, dy)}{G_n([t, \infty), \mathbb{R})}$$

from which the desired result follows. ■

We then have the following characterization.

Proposition 4.2.4 *Under the condition:*

$$\int_0^T \int_{\mathcal{X}} (e^x - 1)^2\, \nu_{\mathcal{G}}^P(dt, dx) < \infty,\ \ P - a.s.$$

S is a $(P, \mathcal{G})-$semimartingale with decomposition $S_t = S_0 + M_t + A_t$ where:

$$A_t = \int_0^t \int_{\mathcal{X}} S_{u-}(e^x - 1)\, \nu_{\mathcal{G}}^P(du, dx)$$

is a predictable process with bounded variation and:

$$M_t = \int_0^t \int_{\mathcal{X}} S_{u-}(e^x - 1)\left(\mu(du, dx) - \nu_{\mathcal{G}}^P(du, dx)\right)$$

is a locally square integrable local martingale.

Remark 4.2.5 *Equivalently, we may write $dS_t = S_{t-} \int_{\mathcal{X}} (e^x - 1)\, \mu(dt, dx)$.*

4.2.3 The Optimization Problem

Since the agent has only access to the filtration $\mathcal{G}$, we modify the classical definition by restricting our attention to strategies $(\delta_t, t \in [0,T])$ adapted to the filtration $\mathcal{G}$. Then, the wealth process reads:

$$dX_t^\delta = \left(X_t^\delta - \delta_t\right) r_t dt + \delta_t \frac{dS_t}{S_{t-}},\ X_0^\delta = x_0 \in \mathbb{R}^+$$

and the class of admissible policies is given by:

$$\mathcal{A}(x) = \left\{\delta : [0,T] \times \Omega \to \mathbb{R}, \exists K > -\infty, \forall t,\ P(X_t^\delta \geqslant K) = 1,\ X_0^\delta = x\right\}$$

In the following, let $\beta_t = B_t^{-1} = \exp\left(-\int_0^t r_s ds\right)$.

Definition 4.2.6 *A self-financing strategy $(\delta_t, t \in [0,T])$ is called $(Q,\mathcal{G})-$ admissible - for a probability measure $Q \sim P$ - if it is a $\mathcal{G}-$predictable process and if the discounted wealth $\beta_t X_t^\delta$ is a $(Q,\mathcal{G})-$martingale.*

Optimzing Terminal Wealth

A fonction $U : \mathbb{R} \to \mathbb{R}$ will be called a utility function if it is strictly increasing, strictly concave, of class $\mathcal{C}^2$ and satisfies:

$$U^{'}\left(0^+\right) = \infty,\ U^{'}\left(\infty\right) = 0 \tag{4.6}$$

and has reasonable asymptotic elasticity:

$$AE_{0^+}\left(U\right) \triangleq \liminf_{x \to 0^+} \frac{xU^{'}\left(x\right)}{U\left(x\right)} > 1,\ AE_{+\infty}\left(U\right) \triangleq \liminf_{x \to \infty} \frac{xU^{'}\left(x\right)}{U\left(x\right)} < 1 \tag{4.7}$$

The optimization problem the investor faces is then as follows.

Definition 4.2.7 *Let U be a utility function. Determine:*

$$u\left(x\right) = \sup_{\delta \in \mathcal{A}(x)} E^P\left[U\left(X_T^\delta\right)\right] \tag{4.8}$$

and find $\widehat{\delta}$ which satisfies $u(x) = E^P[U(X_T^{\widehat{\delta}})]$. Then we call $\widehat{\delta}$ the optimal investment strategy and $\widehat{X} = X(\widehat{\delta})$ the optimal wealth process.

Associated Dual Problem

From Kramkov and Schachermayer (1999) [107], we know that a solution to problem (4.8) relies upon solving the dual optimization problem:

Definition 4.2.8 *The dual version of the primal problem (4.8) writes:*

$$v(y) = \inf_{Q \in \mathcal{Q}} E^P \left[V \left(y \frac{dQ}{dP} \right) \right], \ y > 0 \tag{4.9}$$

with $V(y)$ the conjugate version of the utility function $U(x)$ given by:

$$V(y) = \sup_{x \in \mathbb{R}^+} \left[U(x) - xy \right], \ y > 0$$

and $\mathcal{Q}$ the set of equivalent martingale measures:

$$\mathcal{Q} = \left\{ Q \sim P : \beta_t X_t^\delta \text{ is a local } (Q, \mathcal{G}) - \text{martingale} \right\}$$

Martingale Measures

Let us state the following proposition.

Proposition 4.2.9 *Let Ψ be a $\mathcal{G}-$predictable, $\mathcal{X}-$marked process, s.t.:*

$$\int_0^t \int_{\mathcal{X}} |\Psi_u(x)| \, \nu(dx) \, du < \infty, \ P - a.s., \ t \in [0, T]$$

When $\Psi_u(x) > 0$, $P - a.s.$, the process:

$$\Lambda_t = 1 + \int_0^t \int_{\mathcal{X}} \Lambda_{u-} \left(\Psi_u(x) - 1 \right) \left(\mu(du, dx) - \nu_{\mathcal{G}}^P(du, dx) \right) \tag{4.10}$$

is a strictly positive $\mathcal{G}-$local martingale. Also, when $E^P[\Lambda_T] = 1$ and:

$$\int_0^t \int_{\mathcal{X}} \Psi_u(x) \, \nu_{\mathcal{G}}^P(du, dx) < \infty, \ P - a.s.$$

there exists a probability measure $Q \sim P$ with:

$$\frac{dQ}{dP}|_{\mathcal{G}_t} = \Lambda_t$$

and the $(Q, \mathcal{G}_t)-$compensator of μ is given by:

$$\nu_{\mathcal{G}}^Q(du, dx) = \Psi_u(x) \, \nu_{\mathcal{G}}^P(du, dx)$$

From this, we can now characterize the martingale condition for $\beta_t X_t^\delta$.

Proposition 4.2.10 *The probability measure Q belongs to $\mathcal{Q}$ iff:*

$$\int_{\mathcal{X}} (e^x - 1) \, \Psi_t(x) \, \nu_{\mathcal{G}}^P(dt, dx) = r_t, \ P - a.s. \tag{4.11}$$

and:

$$\int_0^T \int_{\mathcal{X}} (e^x - 1) \, \Psi_t(x) \, \nu_{\mathcal{G}}^P(dt, dx) < \infty, \ P - a.s.$$

Proof. Under Q, we have $d\left(\beta_t X_t^{\delta}\right) = \beta_t \delta_t \left(dS_t / S_{t-} - r_t dt\right)$, so that:

$$\begin{aligned}\frac{dS_t}{S_{t-}} - r_t dt = &\int_{\mathcal{X}} \left(e^x - 1\right)\left(\mu - \nu_{\mathcal{G}}^{Q}\right)(dt, dx) \\ &+ \int_{\mathcal{X}} \left(e^x - 1\right) \Psi_t (x)\, \nu_{\mathcal{G}}^{P} (dt, dx) - r_t dt\end{aligned}$$

Then, as from Proposition 4.2.4, S is a special $P-$semimartingale, then it is a $Q-$local martingale under the above conditions. ∎

By using the Itô's formula, (4.10) can equivalently be written as the solution of the exponential Doléans-Dade equation, say:

$$\Lambda_t = \exp\left(\int_0^t \int_{\mathcal{X}} \ln \Psi_u (x)\, \mu (du, dx) - \int_0^t \int_{\mathcal{X}} \left(\Psi_u (x) - 1\right) \nu_{\mathcal{G}}^{P} (du, dx)\right)$$

Optimal Strategy

The following theorem, adapted to Owen (2002) [135], is central in the derivation of the optimal strategy. We note $I(\cdot) = \left(\partial U(\cdot)\right)^{-1}$ and $\beta_t = B_t^{-1}$.

Theorem 4.2.11 *Let U be a utility function satisfying (4.6) and (4.7). Then:*

1. *There exists a unique solution Q^y to the dual problem (4.9),*
2. *There exists a unique number $\widehat{y}$ s.t. $E^{Q^{\widehat{y}}}\left[\beta_T I\left(\widehat{y}\beta_T \Lambda_T\right)\right] = x_0$,*
3. *The optimal terminal wealth is given by $\widehat{X}_T^{\delta} = I\left(\widehat{y}\beta_T \Lambda_T\right)$,*
4. *The optimal investment policy process is uniquely determined by:*

$$\beta_t \widehat{X}_t^{\delta} = E^{Q^{\widehat{y}}}\left[\beta_T \widehat{X}_T^{\delta} | \mathcal{G}_t\right] = x_0 + \int_0^t \int_{\mathcal{X}} \beta_u \widehat{\delta}_u \frac{dS_u}{S_{u-}} \tag{4.12}$$

where, under $Q^{\widehat{y}}$:

$$\frac{dS_u}{S_{u-}} = \int_{\mathcal{X}} \left(e^x - 1\right)\left(\mu - \nu_{\mathcal{G}}^{Q^{\widehat{y}}}\right)(du, dx)$$

In general, the optimal martingale measure Q^y depends on the parameter $\widehat{y}$, which therefore reads $Q^{\widehat{y}}$. Nevertheless, to lighten the notation, we will omit it in the following and write $Q^{\widehat{y}} \equiv Q$ and $\nu_{\mathcal{G}}^{Q^{\widehat{y}}} \equiv \nu_{\mathcal{G}}^{Q}$.

4.3. Optimization Results

4.3.1 Extension of Clark Formula

Following Appendix C.2 which presents essential material on Malliavin calculus for random measures, we recall the Clark-Ocone-Haussmann formula.

Theorem 4.3.1 (Lokka [121]) *Let $F \in L^2(P, \mathcal{G}_T) \cap \mathbb{D}_{1,2}$, then:*

$$F = E^P[F] + \int_0^T \int_{\mathcal{X}} E^P[D_{t,x}F | \mathcal{G}_{t-}] \left(\mu - \nu_{\mathcal{G}}^P\right)(dt, dx)$$

where $E[D_{t,x}F|\mathcal{G}_{t-}]$ is the $\mathcal{G}-$predictable projection of $D_{t,x}F$.

Our aim is now to represent $Q-$random variables as stochastic integrals w.r.t. the compensated measure $(\mu - \nu_{\mathcal{G}}^Q)$. Following lemma will be helpful.

Lemma 4.3.2 *Let $F \in L^2(P, \mathcal{G}_T) \cap \mathbb{D}_{1,2}$, then:*

$$D_{t,x}(\Lambda_T F) = \Lambda_T \left(D_{t,x}F \int_{\mathcal{X}} \Psi_t(x)\, \eta_t(dx) + F \int_{\mathcal{X}} (\Psi_t(x) - 1)\, \eta_t(dx) \right)$$

where $\eta_t(dx) \triangleq (\mu - \nu_{\mathcal{G}}^Q)(dt, dx)$.

Proof. By the chain rule (C.8), we have:

$$D_{t,x}(\Lambda_T F) = \Lambda_T \cdot D_{t,x}(F) + F \cdot D_{t,x}(\Lambda_T) + D_{t,x}(\Lambda_T) \cdot D_{t,x}(F)$$

Then, we get:

$$D_{t,x}\Lambda_T = \Lambda_T \int_{\mathcal{X}} (\Psi_t(x) - 1)\, \eta_t(dx)$$

which concludes the proof. ■

We now state our Clark's formula extension result.

Theorem 4.3.3 *Let $F \in L^2(Q, \mathcal{G}_t) \cap \mathbb{D}_{1,2}$, then $\Lambda_T F \in \mathbb{D}_{1,2}$ and:*

$$\begin{aligned} F = {} & E^P[\Lambda_T F] + \int_0^T \int_{\mathcal{X}} E^Q[F|\mathcal{G}_{t-}] \left(\Psi_t^{-1}(x) - 1\right) (\mu - \nu_{\mathcal{G}}^Q)(dt, dx) \\ & + \int_0^T \int_{\mathcal{X}} \frac{1}{\Lambda_t} E^Q[D_{t,x}(\Lambda_T F)|\mathcal{G}_{t-}] (\mu - \nu_{\mathcal{G}}^Q)(dt, dx) \end{aligned} \tag{4.13}$$

Proof. For a $\mathcal{G}-$measurable random variable F and a $\mathcal{G}-$martingale Λ given by (4.10), the conditional version of the Bayes formula reads:

$$E^Q\left[F|\mathcal{G}_t\right] = \frac{1}{\Lambda_t} E^P\left[\Lambda_T F|\mathcal{G}_t\right]$$

Applying Clark's formula on $\Lambda_T F$, cf. Theorem 4.3.1, yields:

$$E^Q\left[F|\mathcal{G}_t\right] \triangleq Z_t = \frac{1}{\Lambda_t} U_t = \Lambda_t^{-1} U_t$$

with:

$$U_t = E^P\left[\Lambda_T F\right] + \int_0^t \int_{\mathcal{X}} E^P\left[D_{u,x}\left(\Lambda_T F\right)|\mathcal{F}_{u-}\right]\left(\mu - \nu_{\mathcal{G}}^P\right)(du, dx)$$

$$d\left(\Lambda_t^{-1}\right) = \Lambda_{t-}^{-1} \int_{\mathcal{X}} \left(\Psi_t^{-1}(x) - 1\right)\left(\mu - \nu_{\mathcal{G}}^Q\right)(dt, dx)$$

By Itô's product rule, we get $d\left(Z_t\right) = \Lambda_t^{-1} dU_t + U_t d\Lambda_t^{-1} + d\left[U, \Lambda^{-1}\right]_t$, so that:

$$\begin{aligned} dZ_t = \Lambda_t^{-1} & \int_{\mathcal{X}} E^P\left[D_{t,x}\left(\Lambda_T F\right)|\mathcal{F}_{t-}\right]\left(\mu - \nu_{\mathcal{G}}^Q\right)(dt, dx) \\ & + U_t \Lambda_{t-}^{-1} \int_{\mathcal{X}} \left(\Psi_t^{-1}(x) - 1\right)\left(\mu - \nu_{\mathcal{G}}^Q\right)(dt, dx) \end{aligned}$$

Then, integrating over $[0, T]$ and noting that:

$$Z_T = E^Q\left[F|\mathcal{G}_T\right] = F$$

$$Z_0 = E^Q\left[F|\mathcal{G}_0\right] = E^Q\left[F\right]$$

yield the desired result. ■

4.3.2 The General Result

Consider a $\mathcal{G}_T-$measurable random variable B s.t.:

$$E^Q\left[\beta_T B\right] = E^P\left[\Lambda_T \beta_T B\right] = x_0$$

where $x_0 > 0$. Furthermore, recalling notation from Section 4.2.3, we assume that there exists a unique portfolio process $\delta \in \mathcal{A}(x)$ s.t. the wealth process X^δ satisfies $X_0^\delta = x_0$ and $X_T^\delta = B$, $P-$a.s. Then:

$$\beta_t X_t^\delta = E^Q\left[\beta_T B|\mathcal{G}_t\right] \tag{4.14}$$

This setup arises in the theory of hedging of contingent claims where B corresponds to the payoff of a derivative (written on the underlying S) and so the portfolio X^{δ} is the one that attains the given level of wealth B. Using (4.12) and if the martingale condition (4.11) is satisfied, then (4.14) reads:

$$E^{Q}\left[\beta_T B|\mathcal{G}_t\right] = E^{P}\left[\Lambda_T \beta_T B\right] + \int_0^t \int_{\mathcal{X}} \beta_s \delta_u \left(e^x - 1\right)\left(\mu - \nu_{\mathcal{G}}^{Q}\right)(du, dx)$$

Then, we derive the hedging portfolio δ under partial information.

Proposition 4.3.4 *We have:*

$$\widehat{\delta}_t = \frac{\beta_t^{-1}}{\int_{\mathcal{X}} \left(e^x - 1\right)\eta_t(dx)} \left\{ \int_{\mathcal{X}} E^{Q}\left[\beta_T B|\mathcal{G}_{t-}\right]\left(\Psi_t^{-1}(x) - 1\right)\eta_t(dx) \right. \tag{4.15}$$
$$\left. + \int_{\mathcal{X}} E^{Q}\left[\Psi_t(x) D_{t,x}\left(\beta_T B\right) + \left(\Psi_t(x) - 1\right)\left(\beta_T B\right)|\mathcal{G}_{t-}\right]\eta_t(dx) \right\}$$

where $\eta_t(dx) \triangleq \left(\mu - \nu_{\mathcal{G}}^{Q}\right)(dt, dx)$.

Proof. We impose the conditions of Theorem 4.3.3 and let $F \triangleq \beta_T B$. Then, we obtain the desired formula for $\widehat{\delta}$ by identification with (4.13). ■

In (4.15), the first integral term may be viewed as a mean-variance one, as in a continuous-time setting, while the second integral term accounts for intertemporal hedging terms. It can be decomposed in two parts: the first takes into account the variations in the discount factor, while the second comes from variations in the risk-premia process.

Remark 4.3.5 *Portfolio (4.15) is the hedging strategy for the contingent claim B when prices are only observed at random times. As the market we consider is incomplete, this result needs again to be specialized by a proper choice of the equivalent martingale measure and so the elicitation of the Girsanov quantity Ψ. This result can be connected to this of Frey and Runggaldier (2001) [71] which deal with the same type of problem but under a different framework. Special attention is given, in the litterature, on the existence of the minimal entropy martingale measure, which is defined as follows:*

$$H(Q|P) = E^{P}\left[\frac{dQ}{dP}\log\frac{dQ}{dP}\right],\ Q \ll P$$

so that, cf. Definitions 4.2.7 and 4.2.8:

$$\sup_{\delta \in \mathcal{A}(x)} E^{P}\left[U\left(X_T^{\delta}\right)\right] = -\exp\left(-\inf_{Q \in \mathcal{Q}} H(Q|P)\right) = -\exp\left(-H\left(Q^*|P\right)\right)$$

4.3.3 The Optimal Strategy

Proposition 4.3.4 is now specialized to the case where $B = \widehat{X}_T^\delta$, i.e.: optimal wealth from expected terminal utility, thanks to Theorem 4.2.11.

Theorem 4.3.6 *If $\widehat{y}$ is the unique number such that $E^{Q^{\widehat{y}}}\left[\beta_T I\left(\widehat{y}\beta_T\Lambda_T\right)\right] = x_0$, then $\widehat{X}_T^\delta = I\left(\widehat{y}\beta_T\Lambda_T\right)$ and the optimal investment policy is given by:*

$$
\begin{aligned}
\widehat{\delta}_t^{\widehat{y}} = {} & \frac{\beta_t^{-1}}{\int_{\mathcal{X}}\left(e^x - 1\right)\eta_t\left(dx\right)}\left\{\int_{\mathcal{X}} E^Q\left[\beta_T I\left(\widehat{y}\beta_T\Lambda_T\right)|\mathcal{G}_{t-}\right]\left(\Psi_t^{-1}\left(x\right) - 1\right)\eta_t\left(dx\right)\right. \\
& + \int_{\mathcal{X}} E^Q\left[\Psi_t\left(x\right)\beta_T\left(\mathcal{I}_T^\Psi\left(x\right) - I\left(\widehat{y}\beta_T\Lambda_T\right)\right)|\mathcal{G}_{t-}\right]\eta_t\left(dx\right) \\
& + \int_{\mathcal{X}} E^Q\left[\left(\Psi_t\left(x\right) - 1\right)\beta_T I\left(\widehat{y}\beta_T\Lambda_T\right)|\mathcal{G}_{t-}\right]\eta_t\left(dx\right) \\
& \left. + \int_{\mathcal{X}} E^Q\left[\mathcal{I}_T^\Psi\left(x\right) D_{t,x}\beta_T|\mathcal{G}_{t-}\right]\eta_t\left(dx\right)\right\}
\end{aligned} \tag{4.16}
$$

where:

$$
\mathcal{I}_T^\Psi\left(x\right) = I\left(\widehat{y}\beta_T\Lambda_T + \widehat{y}\Lambda_T\left(\left\langle\Psi_t\left(x\right) - 1\right\rangle_{\eta_t}\beta_T + \left\langle\Psi_t\left(x\right)\right\rangle_{\eta_t} D_{t,x}\beta_T\right)\right)
$$

with:

$$
\left\langle g\right\rangle_{\eta_t} = \int_{\mathcal{X}} g\left(x\right)\eta_t\left(dx\right)
$$

Proof. From Proposition 4.3.4, let $B_T \triangleq I\left(\widehat{y}\beta_T\Lambda_T\right)$ and $C_T \triangleq \widehat{y}\beta_T\Lambda_T$. Clearly $B, C \in \mathbb{D}_{1,2}$ and I is differentiable. By the chain rule (C.8):

$$
D_{t,x}\left(\beta_T B_T\right) = \beta_T \cdot D_{t,x} I\left(C_T\right) + I\left(C_T\right)\cdot D_{t,x}\beta_T + D_{t,x} I\left(C_T\right)\cdot D_{t,x}\beta_T
$$

and as:

$$
\begin{aligned}
D_{t,x} C_T &= \widehat{y}\left(\beta_T \cdot D_{t,x}\left(\Lambda_T\right) + \Lambda_T \cdot D_{t,x}\left(\beta_T\right) + D_{t,x}\left(\Lambda_T\right)\cdot D_{t,x}\left(\beta_T\right)\right) \\
&= \widehat{y}\Lambda_T\left(\left\langle\Psi_t\left(x\right) - 1\right\rangle_{\eta_t}\beta_T + \left\langle\Psi_t\left(x\right)\right\rangle_{\eta_t} D_{t,x}\left(\beta_T\right)\right)
\end{aligned}
$$

we get:

$$
D_{t,x} I\left(C_T\right) = I\left(C_T + \widehat{y}\Lambda_T\left(\left\langle\Psi_t\left(x\right) - 1\right\rangle_{\eta_t}\beta_T + \left\langle\Psi_t\left(x\right)\right\rangle_{\eta_t} D_{t,x}\beta_T\right)\right) - I\left(C_T\right)
$$

so that:

$$
D_{t,x}\left(\beta_T B_T\right) = \left(\mathcal{I}_T^\Psi\left(x\right) - I\left(C_T\right)\right)\beta_T + \mathcal{I}_T^\Psi\left(x\right) D_{t,x}\beta_T
$$

thus yielding the desired result. ■

We now make the assumption.

Assumption 4.3.7 *Let r follows the dynamics:*

$$r_t = r_0 + \int_0^t \int_{\mathcal{X}} r_u h(u,x)\left(\mu - \nu_{\mathcal{G}}^P\right)(du, dx)$$

where h is a $\mathcal{G}-$measurable, $\mathcal{X}-$integrable function. Then, for $t \leqslant u$:

$$D_{t,x} r_u = \int_{\mathcal{X}} r_t h(t,x)\eta_t(dx) + \int_t^u \int_{\mathcal{X}} D_{t,x} r_s h(s,x)\eta_s(dx)$$

From this, we then have:

$$D_{t,x}\beta_T = \beta_T\left(e^{-\int_t^T D_{t,x} r_u du} - 1\right)$$

Proposition 4.3.8 *Under Assumption 4.3.7, (4.16) reads:*

$$\widehat{\delta}_t^{\widehat{y}} = \frac{\beta_t^{-1}}{\int_{\mathcal{X}}(e^x - 1)\eta_t(dx)}\left\{\int_{\mathcal{X}} E^Q\left[\beta_T I\left(\widehat{y}\beta_T\Lambda_T\right)|\mathcal{G}_{t-}\right]\Psi_t^{-1}(x)\eta_t(dx)\right. \tag{4.17}$$
$$\left.+\int_{\mathcal{X}} E^Q\left[\Psi_t(x)\beta_T\mathcal{I}_T^{\Psi}(x)\left(e^{-\int_t^T D_{t,x} r_u du} - 1\right)|\mathcal{G}_{t-}\right]\eta_t(dx)\right\}$$

where:

$$\mathcal{I}_T^{\Psi}(x) = I\left(\widehat{y}\beta_T\Lambda_T + \widehat{y}\Lambda_T\left(\langle\Psi_t(x) - 1\rangle_{\eta_t}\beta_T + \langle\Psi_t(x)\rangle_{\eta_t}\left(e^{-\int_t^T D_{t,x} r_u du} - 1\right)\right)\right)$$

The same interpretation than in (4.15) applies. The first integral term is a traditional mean-variance one, while the second accounts of both the variations in the interest rate and risk-premia processes and may be interpreted as hedging terms. Interestingly, this decomposition is of the same nature than in the continuous-time setting, cf. Detemple et al. (2003) [49].

Remark 4.3.9 *In the case of the three more standard utility functions: power, logarithmic and exponential risk preferences, (4.17) can be specialized. Let:*

$$U(x) = \begin{cases} x^p/p & x\epsilon\mathbb{R}^+, p\epsilon(0,1) \\ \ln x \quad , & x\epsilon\mathbb{R}^+ \\ -e^{-x} & x\epsilon\mathbb{R} \end{cases}, I(y) = \begin{cases} y^q & y\epsilon\mathbb{R}, q = \frac{p}{p-1} \\ y^{-1} \quad , & y\epsilon\mathbb{R}^* \\ -\ln y & y\epsilon\mathbb{R}^+ \end{cases}$$

Then, direct replacements in (4.17) yields the desired investment policy. The only thing that remains to be done is to specialize the risk-premia process Ψ.

Extension to Non-Zero Consumption

We conclude this section by extending the previous results to the case where we maximize expected utility from consumption, thus considering the problem[1]:

$$u^c(x) = \sup_{(\delta,c)\in\mathcal{A}(x)} E^P\left[\int_0^T U(t, c_t)\,dt\right] \tag{4.18}$$

[1] For more details on the optimization setup, cf. Section 4.2.3.

for a utility function $U : [0, T] \times \mathbb{R} \to \mathbb{R}$ of class $\mathcal{C}^{0,2}$ such that $U(t, \cdot)$ satisfies the properties of a utility function, a portfolio process δ_t and a consumption process c_t which is assumed to be $\mathbb{R}^+-$valued and $\mathcal{G}_t-$adapted. Assume that the following integrability condition is satisfied:

$$\int_0^T (\delta_t^2 + c_t)dt < \infty, \ P - a.s.$$

This problem has been addressed by Karatzas et al. (1987) [98].

Theorem 4.3.10 *Problem (4.18) admits a unique solution $(\widehat{\delta}, \widehat{c}) \in \mathcal{A}(x)$ s.t.:*

$$\widehat{c}_t = I(t, \widehat{y}\beta_t\Lambda_t)$$

where $\widehat{y}$ is determined via:

$$x_0 = E^Q\left[\int_0^T \beta_t I(t, \widehat{y}\beta_t\Lambda_t)\, dt\right]$$

and associated wealth process is given by:

$$\beta_t\widehat{X}_t + \int_0^t \beta_s\widehat{c}_s ds = E^Q\left[\beta_T\widehat{X}_T|\mathcal{G}_t\right] = x_0 + \int_0^t \int_{\mathcal{X}} \beta_u\widehat{\delta}_u (e^x - 1)\, \eta_u(dx)$$

Then, analogously to Theorem 4.3.6, we have.

Theorem 4.3.11 *The optimal portfolio from consumption is given by:*

$$\begin{aligned}\widehat{\delta}_t^{\widehat{y}} = \frac{\beta_t^{-1}}{\int_{\mathcal{X}} (e^x - 1)\, \eta_t(dx)} &\left\{\int_{\mathcal{X}} E^Q\left[\int_t^T \beta_u I(\widehat{y}\beta_u\Lambda_u)\, ds|\mathcal{G}_{t-}\right] \Psi_t^{-1}(x)\, \eta_t(dx)\right. \\ &\left. + \int_{\mathcal{X}} E^Q\left[\int_t^T \Psi_u(x)\, \beta_u \mathcal{I}_u^{\Psi}(x) \left(e^{-\int_t^u D_{t,x} r_s ds} - 1\right) du|\mathcal{G}_{t-}\right] \eta_t(dx)\right\}\end{aligned} \tag{4.19}$$

Proof. The computation is in the same vein as in Theorem 4.3.6 where we have set $F = \int_0^T \beta_t I(t, \widehat{y}\beta_t\Lambda_t)\, dt$ and apply the same computations. ■

4.3.4 Options Hedging

In this section, we will prove the usefulness of the extension of the Clark's formula and of the general trading strategy, cf. (4.15). We will build the risk-minimizing hedging under partial information of various options written on the underlying S, which is assumed to follow the dynamics (4.4).

We will consider the case where $r = 0$, so $\beta = 1$, so that (4.15) reads:

$$\widehat{\delta}_t = \frac{1}{\int_{\mathcal{X}} (e^x - 1)\, \eta_t(dx)} \left\{ \int_{\mathcal{X}} E^Q [B|\mathcal{G}_{t-}] \left(\Psi_t^{-1}(x) - 1\right) \eta_t(dx) \right.$$
$$\left. + \int_{\mathcal{X}} E^Q [\Psi_t(x) D_{t,x} B + (\Psi_t(x) - 1) B|\mathcal{G}_{t-}]\, \eta_t(dx) \right\}$$

where Ψ, cf. Proposition 4.2.10, satisfies:

$$\int_{\mathcal{X}} (e^x - 1)\, \Psi_t(x)\, \nu_{\mathcal{G}}^P(dt, dx) = 0$$

Before going further, we note that the time $t \leqslant T$ price of an option with strike K written on S with payoff f is given by:

$$V_f(0, T, K) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{L}_f(R + iu)\, \phi_{S_{T-t}}(iR - u)\, du$$

where $\mathcal{L}_f$ is the Laplace transform of f (with real part R) and ϕ is the characteristic function of the process S, cf. Carr and Madan (1998) [24].

Example 4.3.12 *Let $B = (S_T - K)^+$. Then:*

$$D_{t,x}B = (e^x S_T - K)^+ - (S_T - K)^+$$
$$= e^x S_t E\left[\left(S_{T-t} - \frac{K}{e^x S_t}\right)^+ |\mathcal{G}_t\right] - S_t E\left[\left(S_{T-t} - \frac{K}{S_t}\right)^+ |\mathcal{G}_t\right]$$

The problem reduces then to the computation of the prices of two options. Noting by $f_1(S_T) = S_t \left(S_{T-t} - \frac{K}{S_t}\right)^+$ and $f_2(S_{T-t}) = \left(S_{T-t} - \frac{K}{e^x S_t}\right)^+$ their respective payoffs, their prices read:

$$V_1\left(0, T-t, \frac{K}{S_t}\right) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{L}_1(R + iu)\, \phi_{S_T}(iR - u)\, du$$
$$V_2\left(0, T-t, \frac{K}{e^x S_t}\right) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{L}_2(R + iu)\, \phi_{S_{T-t}}(iR - u)\, du$$

so that:

$$E^Q[B|\mathcal{G}_{t-}] = S_{t-} V_1\left(0, T-t, \frac{K}{S_{t-}}\right)$$

$$E^Q[D_{t,x}B|\mathcal{G}_{t-}] = e^x S_{t-} V_2\left(0, T-t, \frac{K}{e^x S_{t-}}\right) - S_{t-} V_2\left(0, T-t, \frac{K}{S_{t-}}\right)$$

Thus, we have:

$$\widehat{\delta}_t = \frac{S_{t-}}{\int_{\mathcal{X}} (e^x - 1)\, \eta_t(dx)} \left\{ \int_{\mathcal{X}} V_1\left(0, T-t, \frac{K}{S_{t-}}\right) \left(\Psi_t^{-1}(x) - 2\right) \eta_t(dx) \right.$$
$$\left. + \int_{\mathcal{X}} \left(e^x V_2\left(0, T-t, \frac{K}{e^x S_{t-}}\right) - V_1\left(0, T-t, \frac{K}{S_{t-}}\right)\right) \Psi_t(x)\, \eta_t(dx) \right\}$$

Example 4.3.13 *Let* $B = \left((S_T - K)^+\right)^n$, $n \in \mathbb{N}^*$. *Then:*

$$D_{t,x}B = \left((e^x S_T - K)^+\right)^n - \left((S_T - K)^+\right)^n$$

which can be be computed as in previous example.

4.4. Conclusion

This article has investigated the question of optimal portfolios and hedging strategies from prices observed at random times. Turning the problem into one w.r.t. the random measure associated with the observations process, we have been able then to apply the *martingale technique of portfolio optimization*. The originality of the approach is to extend the traditional, continuous time, partial information framework which allow only to learn about the drift parameters of Wiener or Poisson processes, cf. Jeanblanc, Lacoste and Roland (2005) [95]. Under the proposed framework, it becomes possible to learn about stochastic volatility or jump amplitude components.

Our general result, the derivation of the optimal investment for a portfolio that attains a certain given level of wealth, is not reduced to the computation of optimal policies. We can use it to obtain hedging strategies for contigent claims for general payoff functions, thus enriching the result of Frey and Runggaldier (2001) [71] to a more general setup. Moreover, for power, logarithmic and exponential risk preferences, we may compute risk-premia processes in a fairly straightforward fashion, thus turning them to be solved by a numerical scheme.

From an estimation perspective, the proposed model presents some challenges. Usual inference and simulation methods may fail to properly estimate the parameters of the model and to compute the filtering equations. Relying on Crisan et all (1998) [35], particle methods may be well suited. Finally, the study of the convergence of the investment policy computed from discrete observations to this with continuous ones may help to understand the *discrete trading effect*. We will treat these questions on future researches.

C

Appendix (Chapter 4)

Agenda

C.1. Filtering Results

In this Appendix, we study a special dynamics for the process θ and present the semimartingale decomposition of the filter equation. To this end, we assume that the economic factor process θ satisfies, under P, the equation:

$$d\theta_t = m\left(t, \theta_t\right) dt + v\left(t, \theta_t\right) dB_t + \int_\Gamma w\left(t, \theta_{t-}; x\right) \mathcal{N}\left(dt, dx\right)$$

where m, v and w are bounded and continuous functions on $\mathbb{R}^+ \times \mathbb{R}$ and $\mathbb{R}^+ \times \mathbb{R} \times \Gamma$, B is a standard Brownian motion, independent of W, $\mathcal{N}$ is a Poisson random measure, independent of B and of W, with mean rate $E\left[\mathcal{N}\left(t+dt, A\right)\right] - E\left[\mathcal{N}\left(t, A\right)\right] = \varrho dt \nu\left(A\right)$, where ϱ is a real positive number (jump intensity) and ν is a probability measure (jump amplitude) on the space of jumps $\Gamma \subset \mathbb{R}$. We also suppose that habitual Lipschitz and growth conditions for existence of a unique $\mathcal{F}-$adapted solution, cf. Protter (1990) [140], are verified. For a bounded C^2 function f, the generator $\mathcal{L}$ of θ reads:

$$\begin{aligned} \mathcal{L}f\left(x\right) &= m\left(t, x\right) f^{'}\left(x\right) + \frac{1}{2} v^2\left(t, x\right) f^{''}\left(x\right) \\ &\quad + \int_{\mathcal{X}} \left(f\left(x + w\left(u, x, z\right)\right) - f\left(x\right) - f^{'}\left(x\right) w\left(u, x, z\right)\right) \varrho\nu\left(dz\right) \end{aligned} \tag{C.1}$$

C.1.1 The Problem

We introduce the $\mathcal{G}-$conditional density of the random variable θ by:

$$\pi_t(f) = E^P\left[f(\theta_t)\,|\mathcal{G}_t\right]$$

for any $\mathbb{R}-$valued measurable function f s.t. $E^P\left[|f(\theta_t)|\right] < \infty$.

From Section 4.2.2, this setup can be solved as a non-linear filtering problem with measurements generated by $\left(\tau_n, \bar{Y}_{\tau_n}\right)_{n\geqslant 1}$, so that the *discrete* asset log price Y is a pure-jump process with unobservable factor process θ.

We now introduce a further definition related to random measures and conditional expectation, cf. Chapter 3 in Liptser and Shiryaev (1989) [117].

Definition C.1.1 *To each random measure μ and probability measure P, one can relate the associated Doléans measure M_μ^P with:*

$$M_\mu^P(d\omega; dt, dx) = P(d\omega)\,\mu(\omega; dt, dx)$$

Then, for any non-negative $\mathcal{G}-$measurable function $X = X(\omega; t, x)$:

$$M_\mu^P(X) = E\left[X * \mu_\infty\right] = E\left[\int_0^\infty \int_{\mathcal{X}} X(\omega; u, x)\,\mu(\omega; du, dx)\right]$$

Denote by $\mathcal{P}(\mathcal{G})$ the predictable $\sigma-$algebra on $\Omega \times [0,\infty)$ w.r.t. $\mathcal{G}$ and set:

$$\overline{\mathcal{P}}(\mathcal{G}) = \mathcal{P}(\mathcal{G}) \otimes \mathcal{B}(\mathbb{R})$$

The non-negative $\overline{\mathcal{P}}(\mathcal{G})-$measurable function $M_\mu^P\left(X|\overline{\mathcal{P}}(\mathcal{G})\right)$ is called the conditonal expectation of X w.r.t. the measure M_μ^P and the $\sigma-$algebra $\overline{\mathcal{P}}(\mathcal{G})$ if for any bounded non-negative $\overline{\mathcal{P}}(\mathcal{G})-$measurable function $Z = Z(\omega; t, x)$:

$$M_\mu^P(ZX) = M_\mu^P\left(Z M_\mu^P\left(X|\overline{\mathcal{P}}(\mathcal{G})\right)\right) \tag{C.2}$$

When the random measure μ possesses a compensator ν, we have $M_\mu^P(X) = M_\nu^P(X)$, so that (C.2) then reads $M_\nu^P(ZX) = M_\mu^P\left(Z M_\mu^P\left(X|\overline{\mathcal{P}}(\mathcal{G})\right)\right)$.

The next version of the Martingale Representation theorem will be useful.

Lemma C.1.2 *Let M be a $(P,\mathcal{G})-$local martingale with $M_0 = 0$. Then, there exists an integrable $\mathcal{G}-$predictable, $\mathcal{X}-$marked process $H(u,x)$ s.t.:*

$$M_t = \int_0^t \int_{\mathcal{X}} H(u,x)\,(\mu - \nu)(du, dx)$$

From Lemma 4.8.1 in Lipster and Shiryaev (1989) [117], one can take:

$$H(u,x) = M_\mu^P\left(\Delta M_u|\overline{\mathcal{P}}(\mathcal{G})\right)(u,x)$$

Besides, if $f^{'}$ and $f^{''}$ exist and are bounded, Itô's formula yields:

$$f(\theta_t) = f(\theta_0) + \int_0^t \mathcal{L}f(\theta_u)\,du + \int_0^t f^{'}(\theta_u)\,v(u,\theta_u)\,dB_u \\ + \int_0^t \int_{\mathcal{X}} (f(x + w(u,x,z)) - f(x))\,(\mathcal{N} - \varrho)(du,dz) \tag{C.3}$$

where $\mathcal{L}$ is given by (C.1).

C.1.2 The Filtering Equation

In the following, we derive a semimartingale representation of the filter.

Theorem C.1.3 *The conditional filter* $\pi_t(f)$ *satisfies:*

$$d\pi_t(f) = \pi_t(\mathcal{L}f)\,dt \\ + \int_{\mathcal{X}} \left(\sum_{n \geqslant 0} \mathbf{1}_{]\tau_n,\tau_{n+1}]}(t)\,\frac{\pi_{\tau_n}(\psi_n(f;t,x))}{\pi_{\tau_n}(\psi_n(1;t,x))} - \pi_{t-}(f) \right) \left(\mu - \nu_{\mathcal{G}}^{P}\right)(dt,dx)$$

Proof. Denote by L_t the martingale defined by:

$$L_t = f(\theta_t) - f(\theta_0) - \int_0^t \mathcal{L}f(\theta_u)\,du$$

cf. (C.3), so that:

$$\pi_t(f) = E[f(\theta_0)|\mathcal{G}_t] + E\left[\int_0^t \mathcal{L}f(\theta_u)\,du|\mathcal{G}_t\right] + E[L_t|\mathcal{G}_t]$$

Set:

$$M_t = E[L_t|\mathcal{G}_t] + \{E[f(\theta_0)|\mathcal{G}_t] - \pi_0(f)\} \\ + \left\{E\left[\int_0^t \mathcal{L}f(\theta_u)\,du|\mathcal{G}_t\right] - \int_0^t \pi_s(\mathcal{L}f)\,ds\right\}$$

and note that each of the three terms entering in the definition of M is a $\mathcal{G}-$martingale. From this, we then note that $\Delta M_t = \pi_t(f) - \pi_{t-}(f)$. Then, as M is a $(P,\mathcal{G})-$martingale, Lemma C.1.2 applies and we obtain:

$$M_{\mu}^{P}\left(\Delta M_u|\overline{\mathcal{P}}(\mathcal{G})\right)(u,x) = M_{\mu}^{P}\left(\pi_u(f) - \pi_{u-}(f)\,|\overline{\mathcal{P}}(\mathcal{G})\right)(u,x)$$

From Lemma 4.10.2 in Liptser and Shiryaev (1989) [117], we have:

$$M_{\mu}^{P}\left(\pi_u(f)\,|\overline{\mathcal{P}}(\mathcal{G})\right)(u,x) = M_{\mu}^{P}\left(f|\overline{\mathcal{P}}(\mathcal{G})\right)(u,x)$$

and as $\pi_{u-}(f)$ is $\overline{\mathcal{P}}(\mathcal{G})$ −measurable, it follows that:

$$M_{\mu}^{P}\left(\Delta M_u|\overline{\mathcal{P}}(\mathcal{G})\right)(u,x)=M_{\mu}^{P}\left(f|\overline{\mathcal{P}}(\mathcal{G})\right)(u,x)-\pi_{u-}(f)$$

From Definition C.1.1, the conditional expectation $M_{\mu}^{P}\left(f|\overline{\mathcal{P}}(\mathcal{G})\right)(u,x)$ is defined, for any $\overline{\mathcal{P}}(\mathcal{G})$ −measurable function ϕ, via:

$$M_{\mu}^{P}\left(\phi(u,x)f(\theta_u)\right)=M_{\mu}^{P}\left(\phi(u,x)M_{\mu}^{P}\left(f(\theta_u)|\overline{\mathcal{P}}(\mathcal{G})\right)\right)$$

As:

$$M_{\mu}^{P}\left(\phi(u,x)f(\theta_u)\right)=M_{\nu}^{P}\left(\phi(u,x)f(\theta_u)\right)$$

and noting, thanks to Lemma 4.2.3, that:

$$\begin{aligned}
&\int_0^\infty\int_{\mathcal{X}}\phi(u,x)f(\theta_u)\nu_{\mathcal{G}}^{P}(du,dx)\\
&=\sum_{n\geqslant 0}\int_{\tau_n}^{\tau_{n+1}}\int_{\mathcal{X}}\phi(u,x)f(\theta_u)\frac{G_n(u,x)}{G_n([u,\infty),\mathbb{R})}dudx\\
&=\sum_{n\geqslant 0}\int_{\tau_n}^{\tau_{n+1}}\int_{\mathcal{X}}\phi(u,x)\frac{E\left[f(\theta_u)\lambda_u(\theta_u)e^{-\Lambda_n(u)}\phi_{\tau_n,u}(y-Y_{\tau_n})|\mathcal{G}(n-1)\right]}{E\left[\lambda_u e^{-\Lambda_n(u)}|\mathcal{G}(n-1)\right]}dudx\\
&=\int_0^\infty\int_{\mathcal{X}}\phi(u,x)\frac{\gamma(f\lambda;u,x)}{\gamma(\lambda;u,dx)}\nu_{\mathcal{G}}^{P}(du,dx)
\end{aligned}$$

Taking expectations, this means that:

$$M_{\mu}^{P}\left(f(\theta_u)|\overline{\mathcal{P}}(\mathcal{G})\right)=\frac{\gamma(f\lambda;u,x)}{\gamma(\lambda;u,x)}$$

thus yielding:

$$H(u,x)=\frac{\gamma(f\lambda;u,x)}{\gamma(\lambda;u,x)}-\pi_{u-}(f)$$

which, thanks to Lemma 4.2.3, concludes the proof. ■

In practical situations, the following Corollary will be of pivotal importance.

Corollary C.1.4 *The filter presents a more convenient recursive structure which follows the jump times $(\tau_n)_{n\geqslant 1}$. In fact, for any jump time τ_n:*

$$\pi_{\tau_n}(f)=\frac{\pi_{\tau_{n-1}}\left(\psi_{n-1}(f;\tau_n,Y_{\tau_n})\right)}{\pi_{\tau_{n-1}}\left(\psi_{n-1}(1;\tau_n,Y_{\tau_n})\right)}\tag{C.4}$$

and for $t\in[\tau_n,\tau_{n+1})$:

$$\pi_t(f)=\pi_{\tau_n}(f)+\int_{\tau_n}^{t}\pi_s(\mathcal{L}f)\,ds\tag{C.5}$$

Remark C.1.5 *The filter $\pi_t(f)$ may be solved numerically via two ways. The first one uses the recursive structure of the filter and so equations (C.4)-(C.5), which can be computed in a relatively easy way. The second path resorts to branching particle system, cf. Crisan et all (1998) [35], which consists in constructing weighted empirical measures which converge to the optimal filter.*

C.2. Primer on Malliavin Calculus

This appendix is inspired by Lokka (2003) [121] who derived a Clark-Ocone-Haussmann formula for random measures with a derivative (Malliavin) operator D defined by its action on the chaos representation of L^2 functionals.

We begin with the following useful lemma.

Lemma C.2.1 (Lokka [121]) *The linear span of random variables:*

$$\exp\left(\int_0^t \int_{\mathcal{X}} \ln \Psi_u(x)\, \mu(du, dx) - \int_0^t \int_{\mathcal{X}} (\Psi_u(x) - 1)\, \nu_{\mathcal{G}}^P(du, dx)\right) \tag{C.6}$$

where Ψ is a $\mathcal{G}-$predictable process, is dense in $L^2(P, \mathcal{G}_T)$.

Crucial in the following is the definition of the integral:

$$I_n(f_n) \triangleq \int_{[0,T]^n \times \mathcal{X}^n} f_n(t_1, ..., t_n; x_1, ..., x_n)\, d(\mu - \nu)^{\otimes n}$$

for constants $f_0 \in \mathbb{R}$ s.t. $I_0(f_0) = f_0$ and for each $n \in \mathbb{N}^*$ and $f \in L^2([0,T]^n \times \mathcal{X}^n)$. And we introduce the set $\mathbb{D}_{1,2} \subset L^2(P)$ by:

$$\mathbb{D}_{1,2} \triangleq \left\{ F = \sum_{n=0}^{\infty} I_n(f_n) : \sum_{n=0}^{\infty} n \cdot n!\, \|f_n\|_n^2 < \infty \right\}$$

Then, we can define a linear operator $D : \mathbb{D}_{1,2} \to L^2([0,T]^n \times \mathcal{X} \times \Omega)$ by:

$$D_{t,x}F \triangleq \sum_{n=1}^{\infty} n I_{n-1}(f_n(\cdot, t, x)) \tag{C.7}$$

for $F = \sum_{n=0}^{\infty} I_n(f_n)$ where $f_n(\cdot, t, x) = f_n(t_1, ..., t_{n-1}, t, x_1, ..., x_{n-1}, x)$. On Ω, we define the canonical random measure:

$$\omega(A \times B) \triangleq \mu(\omega, A \times B)$$

and define the transformations:

$$\varepsilon_{(t,x)}^{-}\omega(A \times B) = \omega(A \times B \cap (t,x)^c)$$
$$\varepsilon_{(t,x)}^{-}\omega(A \times B) = \varepsilon_{(t,x)}^{-}\omega(A \times B) + \mathbf{1}_A(t)\, \mathbf{1}_B(x)$$

which consists of removing or adding a mass at (t, x), respectively.

Proposition C.2.2 *For $F \in \Omega$, $D_{t,x}F = F \circ \varepsilon^{+}_{(t,x)} - F$.*

The following chain rules will be helpful in computations.

Lemma C.2.3 *If f is differentiable and $F, G, FG \in \mathbb{D}_{1,2}$ and $f(G) \in \mathbb{D}_{1,2}$:*

$$D_{t,x}(FG) = D_{t,x}F \cdot G + F \cdot D_{t,x}G + D_{t,x}F \cdot D_{t,x}G \tag{C.8}$$

and:

$$D_{t,x}f(G) = f(G + D_{t,x}G) - f(G) \tag{C.9}$$

Part II

Simulation-Based Optimization

5

Introduction

Agenda

In this introduction, we discuss numerical solutions for a class of Maximization of Expected Utility (MEU) problems. Except in some special cases, such problems cannot be solved numerically. One approach is to turn them into their probabilistic formulation and then resort to simulation-based algorithms. Relying on Monte-Carlo Markov Chain and particle system methods, we present an original alternative for MEU problems.

5.1. Maximization of Expected Utility

Subsequently, we recall the general framework in which are embedded expected utility problems. The maximization is done over some design (or control) space $d \in \mathcal{D}$. The experiment is defined by a model $p_d\left(y|\theta\right)$, i.e.: a distribution of the vector $y \in \mathcal{Y}$ of observables, conditional on some unknown parameter vector $\theta \in \Theta$. The model may depend on the design (or control) parameter d, hence the subscript d. From a Bayesian perspective, the model is completed by a prior distribution $p\left(\theta\right)$ over the parameter vector. Utility is a function of the form $u(d, y, \theta)$. Since the design (or control) parameter d has to be chosen before observing the experiment, we need to maximize the expectation of $u(\cdot)$ w.r.t. (y, θ). Therefore, the maximum expected utility (MEU) problem reads as:

$$d^* = \arg\max_{d \in \mathcal{D}} U(d) \tag{5.1}$$

where:

$$U(d) = \int_{\mathcal{Y} \times \Theta} u(d, y, \theta)\, p_d(y, \theta)\, d\theta dy$$

and so $U(d)$ is the expected utility for action d.

Remark 5.1.1 *Note that even if the optimization of interest is interpreted in terms of the maximization of a utility function U, it is straightforward to adapt it to the case of the minimization of a loss function L, for example by writting* $\min U(d) = \max(K - U(d))$ *for a properly chosen constant K.*

A traditional method to evaluate problems of the form of (5.1) is to use backward induction via the dynamic programming principle, cf. De Groot (1970) [38] or hybrid ones combining forward simulation and backward induction. More precisely, if the structure of the problem satisfies some assumptions, cf. Fleming and Soner (1993) [65], a dynamic programming principle (DPP) holds. From this, we can derive the DPP (under general hypothesis) or Hamilton-Jacobi-Bellman (HJB) equation (under Markov hypothesis). When no closed-form solution is available, the DPP allows a direct numerical evaluation of the HJB equation. Since the incepit of the theory, numerous and powerful methods have been developped. Those rely either on Markov chain methods, cf. Kushner and Dupuis (1992) [109], or finite difference methods, cf. Lapeyre et al. (2006) [112]. But these can perform poorly in high dimension. Another recently borned proposition for solving such problems is quantization algorithm, cf. Pagès et al. (2004) [136].

5.1.1 Joint Integration Issue

Problems of the form of (5.1) require first to integrate the uncertainty to compute expected utility and then to integrate. When turning to numerical aspects, a joint issue arises. Expect in some special cases, neither the integration nor the maximization can be solved analytically. Also, the aforementionned algorithms: grid approximation, finite difference or quantization, cannot handle this joint uncertainty-utility integration in a same loop. The two integrations have to be separated and done at different times, resulting in an additional amount of computational time and greater complexity. To cope with this, we propose a methodology based on simulation-based algorithms.

5.1.2 Monte-Carlo Method

As previously said, the utility function $U(d)$ may be costly to compute and so (5.1) does not allow a closed-form solution and classical optimization methods are not practicable. An attempt to cope with this is to resort to simulation-based methods. Typically, these are based on the observation that the integral $U(d)$ is easily evaluated by Monte Carlo simulation. In fact, in most setups, the model distribution $p_d(y,\theta) = p_d(y|\theta)\,p(\theta)$ is available for sampling, then allowing an approximation of $U(d)$ by:

$$\widehat{U}(d) = \frac{1}{M}\sum_{i=1}^{M} u(d, y_i, \theta_i) \underset{M\to\infty}{\to} U(d)$$

where $\{(y_i,\theta_i), i = 1, ..., M\}$ is a sample from $y_i \sim p_d(y|\theta_i)$ and $\theta_i \sim p(\theta)$.

Muller and Parmigiani (1996) [130] were the first to propose a fully numerical optimal Bayesian scheme to solve (5.1). We briefly outline the procedure. In a first step, one selects some designs (or controls) $d_i \in \mathcal{D}$, either in a simulation context or in a grid. Then, one simulates experiments $(y_i,\theta_i) \sim p_{d_i}(y,\theta)$, one for each chosen design. For each chosen experiment (d_i, y_i, θ_i), we evaluate the observed utility $u_i = u(d_i, y_i, \theta_i)$. Then, in a scatterplot of d_i and u_i, the integration in (5.1) can be replaced by a simple scatterplot smoothing $\widehat{U}(d)$ and the optimal design can be read off as the mode of the smooth curve, cf. Figure 5.1 (taken from Muller and Parmigiani (1996) [130]).

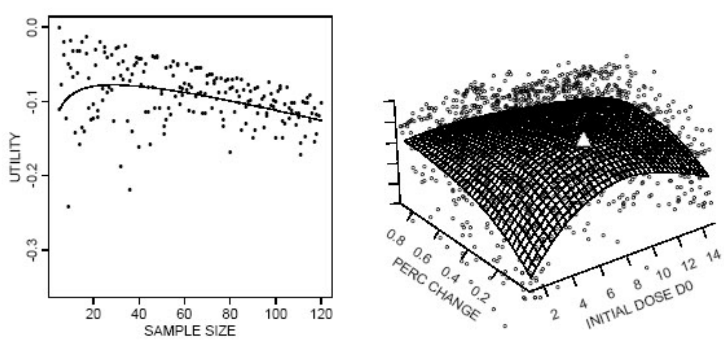

Figure 5.1: Simulated utilities u_i. Left chart: one-dimensional problem; right chart: bivariate design parameter d.

Nevertheless, this procedure generally fails, principally in high dimension, when the number of designs d is numerous and that the problem may trap in local minimums. Another approach may be developped.

5.2. MCMC and Particle Methods

We propose one which resort to Markov-Chain Monte Carlo methods, cf. Appendix D, and particle systems. The key idea is that MCMC are well suited to approximate $U(d)$ in high dimension and particle systems, by generating a large number of designs d, may circumvert local minima issues.

5.2.1 Stochastic Optimization

The key idea, originally developed by Muller (1998) [131], is to consider U as a probability density function, up to a multiplicative constant, over the decision space $\mathcal{D}$ and to generate a sample from this distribution, the mode of which corresponds to the optimal decision d^*. This is made possible by the fact that the *utility* function u is bounded, positive and continuous. As U is costly to compute in practice, we need to introduce an augmented (artificial) probability distribution h (omitting the variable θ to lighten the notation) over $(\mathcal{D}, \mathcal{Y})$:

$$h(d, y) \propto u(d, y)p(y|d)$$

We can compute h explicitly since $p(y|d)$ and $u(d, y)$ are both easy to compute. By definition, the marginal of h on d is proportional to U, so:

$$\int_{\mathcal{Y}} h(d, y)dy \propto U(d)$$

We are now in the context where MCMC simulation methods, cf. Robert and Casella (2004) [142], apply. From these, we can generate a sample from h and then from U. As we are interested in the mode of U, this method is not very efficient especially since dimensionality of our problem may be relatively high ($\dim \mathcal{D}$ can be up to 10), cf. Chapter 6, and since our surface may be quite complex. As a consequence, we need to improve the mode search.

5.2.2 Simulated Annealing

A classical improvement of this approach is given by the simulated annealing algorithm that we can easily adapt. The seminal idea, used in Brooks and Morgan (1995) [21] or in Muller (1998) [131], is to simulate a sample from U^J, where J is a large integer. This will obviously sharpen the top of utility surface and concentrate simulations closer to the mode, as shown in Figure 5.2.

Theoretically, we could use the same fixed value of J for all iterations, but this is not efficient in high dimensional cases. We can also use a *cooling* schedule that makes $J(n)$ increase up to $+\infty$ when $n \to +\infty$. If the simulation method is of Metropolis-Hastings type, we know, from Geman and Geman (1984) [73],

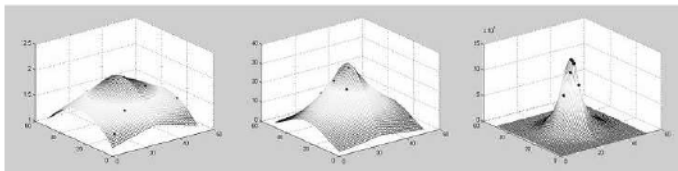

Figure 5.2: Concentration of Simulated Samples around the Mode of Utility Surface.

that the asymptotic condition that ensures convergence to the mode is:

$$J(n) < \frac{\log n}{m\left(\sup_d U(d) - \inf_d U(d)\right)}$$

where $m = \dim \mathcal{D}$. Following a similar idea as in our MEU context, we introduce a new joint augmented distribution h_J on $\mathcal{D} \times \mathcal{Y}^J$ defined as:

$$h_J(d, d_1, .., d_J) \propto \prod_{j=1}^{J} u(d, y_j) p(y_j|d)$$

Assuming independence on the y_j variables, we keep the key property:

$$\int_{\mathcal{Y}} ... \int_{\mathcal{Y}} h_J(d, y_1, ..., y_J)\, dy_1 ... dy_J \propto U^J(d)$$

so that a sample from h_J would marginally gives us a sample from U^J. The limits of MCMC methods, especially in high dimension and when implemented with simulated annealing, are definitely problems of local modes which can trap Markov chains, cf. Andrieu et al. (2000) [7]. In most of the practical implementations, the utility surface is rather complicated and high-dimensional and so we are very concerned by such limits. We propose here an original algorithm based on Interacting Particle System (IPS) methods, cf. Del Moral et al. (2001) [41], exploiting all the improvements exposed above, for a better exploration of $\mathcal{D}$, and then for a more efficient marginal mode search of h_J.

5.2.3 A Particle Approach

This section describes an original alternative to standard MCMC for MEU problems. The method is based on recent developments on particle filters, cf. Doucet et al. (2001) [54], and population Monte Carlo simulations, cf. Cappé et al. (2004) [23]. To simulate a sample from h_J, we no longer produce one

Markov chain $(d^{(n)}, y_j^{(n)})_{j=1..J}$ like in Muller's (1998) [131] algorithm, but we generate instead N *parallel* chains $(d_i^{(n)}, y_{i,j}^{(n)})_{i=1..N,j=1..J}$. Using the vocabulary from sequential MCMC theory, each couple $(d_i^{(n)}, y_{i,j}^{(n)})$ is called a particle, and the set of N Markov chains is called an Interacting Particle System (IPS). Our interest is not to produce a sample to approximate the target distribution but rather to simulate particles close to the mode(s). Figure 5.3 illustrates the mechanism of the proposed MCMC with IPS algorithm.

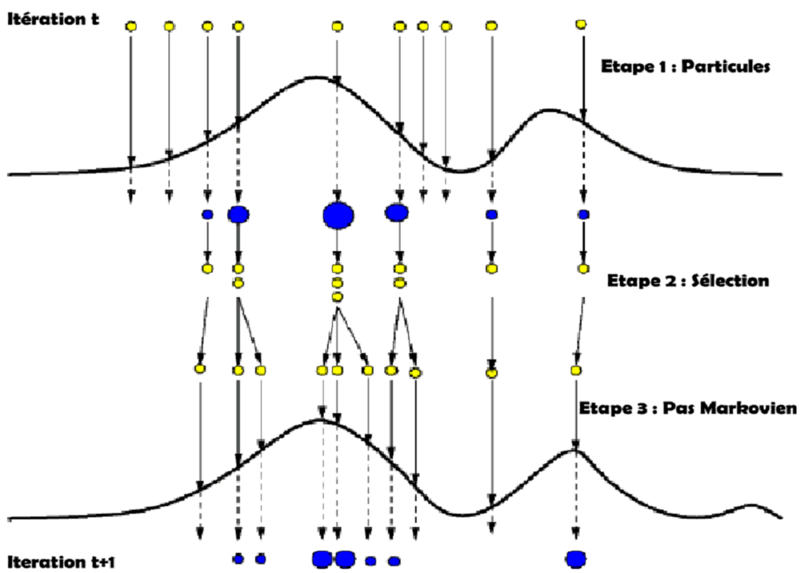

Figure 5.3: Mechanism of the MCMC with IPS Algorithm.

This algorithm is done in three steps:

- **importance sampling**: the idea, cf. Geweke (1989) [75], is to generate, at each iteration n, an approximated weighted sample from h_J,
- **selection procedure**: it is performed to duplicate particles closer to the modes of the target distribution while eliminating the others. A standard selection procedure can be a *sampling importance resampling* scheme, as described in Rubin (1988) [144] or Smith and Gelfand (1992) [155]. This selection procedure has been widely studied and applied in the literature about particles methods for sequential MCMC, cf. Doucet et al. (2001) [54]. Note that other ways of selection are possible, cf. Carpenter et al. (1999) [25] or Liu et al. (1998) [119], but we won't discuss them here,

- **Markov step**: an independent random walk step with h_J as target distribution is also added for each particle, to avoid degeneracy problems.

At each step (importance sampling, selection procedure, Markov step) and for any iteration n, this would generate a sample $(\xi_i^{(n)})_{i=1..N}$ from h_J such that as $N \to +\infty$, the following Monte Carlo approximation:

$$\frac{1}{N}\sum_{i=1}^{N}\phi(\xi_i^{(n)}) \underset{N\to\infty}{\to} \int \phi(\xi)h_J(\xi)d\xi$$

for any measurable and bounded function ϕ, holds. The interest of this approach is to get a rich sample from h_J. The obvious drawback is that this iterative scheme for fixed J and N would cumulate noise, so that the approximation would worsen with iterations. This point has been underlined in developments of non-sequential population MCMC algorithms, cf Cappé et al. (2004) [23] or Chopin (2002) [28]. However the interest of such iterative algorithms is fully recovered when the target distribution changes with iterations, like in sequential MCMC or particle filter algorithms. In our case, this holds since we add a simulated annealing effect, so that J grows with iterations n. Note that in this situation, we don't have anymore constraints about the form of the cooling function $J(n)$ to get the convergence of the algorithm, as we will see it in the convergence Appendix E. When no confusion is possible, we will keep writing J instead of $J(n)$, for lighter notation.

5.3. Particle Algorithm with Simulated Annealing

Previous developements lead to an original algorithm for stochastic optimization that encompasses simulated annealing into an interacting particle approach. As Markov step, a common and convenient choice will be to use a Metropolis-Hastings step[1], since it is easy to implement in practice, cf. Robert and Casella (2004) [142]. Furthermore, we need to choose two random walk jump functions over $\mathcal{D}$:

- $q_{1,n}$ as transition kernel for the importance sampling step,
- $q_{2,n}$ for the Metropolis-Hastings jump.

These functions can be widely different. For example, $q_{1,n}$ could have a smaller variance to allow a narrow scale exploration at the importance sampling step, and a larger scale one at the Metropolis-Hastings step[2]. With these

[1] We will afford a simplification in the theoretical part, cf. Appendix E.
[2] We will precise these choices in Chapters ?? and 6.

notations, the weights for importance resampling (multinomial law) are given by:

$$w_i^{(n)} \propto \frac{\prod_{j=1}^{J} u(\widetilde{d}_i^{(n)}, \widetilde{y}_{ij}^{(n)})}{q_{1,n}(d_i^{(n-1)}, \widetilde{d}_i^{(n)})}$$

The detailed algorithm can be written as follow:

Algorithm 5.3.1 (General Particle Optimization Algorithm) *We have:*

1. *Initialization: Start with a sample* $(d_i^{(0)})_{i=1..N}$ *at* $t = 0$. *Set* $J = J(0)$.
2. *Importance Sampling Step: For each* $i = 1...N$ *and* $j = 1...J$*:*
   (a) *simulate* $\widetilde{d}_i^{(1)}$ *from* $K_{1,1}(d_i^{(0)}, .)$ *and* $(\widetilde{y}_{ij}^{(1)})$ *from* $p(y|\widetilde{d}_i^{(1)})$,
   (b) *compute* $\widetilde{u}_i^{(1)} = \prod_{j=1}^{J} u(\widetilde{d}_i^{(1)}, \widetilde{y}_{ij}^{(1)})$ *and* $w_i^{(1)} \propto \widetilde{u}_i^{(1)} / K_{1,1}(d_i^{(0)}, \widetilde{d}_i^{(1)})$.
3. *Selection Step: Resample* $(\widehat{d}_1^{(1)}, .., \widehat{d}_N^{(1)})$ *from* $(\widetilde{d}_1^{(1)}, .., \widetilde{d}_N^{(1)})$ *from a multinomial distribution with weights* $w_i^{(1)}$, *note* $\widehat{u}_i^{(1)}$ *the corresponding utilities.*
4. *Metropolis-Hastings Step: For each* $i = 1...N$ *and* $J = 1...J$*:*
   (a) *simulate* $\overline{d}_i^{(1)}$ *from* $K_{2,1}(\widehat{d}_i^{(1)}, .)$ *and* $(\overline{d}_{ij}^{(1)})$ *from* $p(y|\overline{d}_i^{(1)})$,
   (b) *compute* $\overline{u}_i^{(1)} = \prod_{j=1}^{J} u(\overline{d}_i^{(1)}, \overline{d}_{ij}^{(1)})$ *and the acceptance rates* $\alpha_i = \min(1, (\overline{u}_i^{(1)} K_{2,1}(\overline{d}_i^{(1)}, \widehat{d}_i^{(1)}))/(\widehat{u}_i^{(1)} K_{2,1}(\widehat{d}_i^{(1)}, \overline{d}_i^{(1)})))$,
   (c) *set* $d_i^{(1)} = \overline{d}_i^{(1)}$ *with probability* α_i *and* $d_i^{(1)} = \widehat{d}_i^{(1)}$ *elsewhere.*

- *We loop the last three steps until* J *is sufficiently large to allow mode determination.*

5.3.1 A Resampling Markov Algorithm

From now on, we can notice an important particular case of this general algorithm. By removing the importance sampling step, we can only consider loops of multinomial resampling and Markov steps. Indeed, if one gets, at the beginning of loop n, an approximated sample from π_n, it becomes an approximated sample from π_{n+1} after resampling and Markov rejuvenating. We assume that $J(n) > J(n-1)$ to give a sense to the resampling, which means that, in this case, the cooling schedule is at least linear. In this setup, the resampling weights have a much simpler form:

$$w_i^{(n)} \propto \prod_{j=J(n-1)+1}^{J(n)} u(d_i^{(n)}, y_{ij}^{(n)})$$

where $y_{ij}^{(n)}$ are additional independent draws from $p(y|d_i^{(n)})$. The Markov transition, denoted K_n in this case, is the same as in the general algorithm. Below, we take the Metropolis-Hastings Markov step to describe the algorithm. An obvious advantage of this resampling algorithm is a substantial saving of computation time, which may be used to improve the Markov kernel (like using adaptive Markov transitions). A practical drawback lies in the exploration of the utility surface in the most complex multimodal cases. However, we will see that this algorithm is also theoretically more *stable*, as we will prove an uniform convergence theorem for it, when a linear form for the cooling schedule is provided, cf. Appendix E.

Algorithm 5.3.2 (Resampling-Markov Algorithm) *We have:*

1. *Initialization: Start at* $t = 0$ *with a sample* $(d_i^{(0)}, y_{i1}^{(0)})_{i=1..N}$ *drawn by importance sampling like in Step* 2 *of Algorithm 5.3.1. Set* $J = J(0) = 1$.
2. *Reweighting: For each* $i = 1...N$ *and* $j = 1...J$:
   (a) *simulate independent additional data* $(y_{ij})_{j=J(0)-1..J(1)}$ *from* $p(y|d_i^{(0)})$,
   (b) *compute the new weights* $w_i^{(1)} \propto \prod_{j=J(0)+1}^{J(1)} u(d_i^{(1)}, y_{ij}^{(1)})$,
3. *Selection Step: Resample* $(\widehat{d}_1^{(1)}, .., \widehat{d}_N^{(1)})$ *from* $(d_1^{(0)}, .., d_N^{(0)})$ *with a multinomial distribution with weights* $w_i^{(1)}$, *note* $\widehat{u}_i^{(1)}$ *the corresponding utilities.*
4. *Metropolis-Hastings Step: For each* $i = 1...N$ *and* $j = 1...J$:
   (a) *simulate* $\overline{d}_i^{(1)}$ *from* $K_1(\widehat{d}_i^{(1)}, .)$ *and* $(\overline{d}_{ij}^{(1)})$ *from* $p(y|\overline{d}_i^{(1)})$,
   (b) *compute* $\overline{u}_i^{(1)} = \prod_{j=1}^{J} u(\overline{d}_i^{(1)}, \overline{y}_{ij}^{(1)})$ *and the acceptance rates* $\alpha_i = \min(1, (\overline{u}_i^{(1)} K_1(\overline{d}_i^{(1)}, \widehat{d}_i^{(1)}))/(\widehat{u}_i^{(1)} K_1(\widehat{d}_i^{(1)}, \overline{d}_i^{(1)})))$,
   (c) *set* $d_i^{(1)} = \overline{d}_i^{(1)}$ *with probability* α_i *and* $d_i^{(1)} = \widehat{d}_i^{(1)}$ *elsewhere.*

- *We loop the last three steps along* n.

D

Appendix (Chapter 5)

In this Appendix, we review Markov Chain Monte-Carlo (MCMC) algorithms, their theoretical underpinnings and convergence properties. More details on this can be found in Robert and Casella (2004) [142].

D.1. MCMC Methods

Central in the study of MCMC methods is the ergodic theorem which is the analog for Markov chains of the law of large numbers for random variables.

Theorem D.1.1 (Ergodic Theorem) *Let (X_n) be a Markov chain on the finite space E, homogeneous and irreducible. Let also μ be its unique invariant law. Then, for all initial law for X_0 and any function $f : E \to \mathbb{R}$, we have:*

$$\frac{1}{n+1}\left(f\left(X_0\right)+...+f\left(X_n\right)\right) \underset{n\to\infty}{\overset{p.s.}{\longrightarrow}} \int_E f\left(x\right)\mu\left(dx\right)$$

Monte-Carlo algorithms allow to give an approximation of the integral $\int_E f\left(x\right)\mu\left(dx\right)$ by sampling a series (X_n) of random variables, independent and with same distribution μ, via the law of large numbers approximation:

$$\int_E f\left(x\right)\mu\left(dx\right) \simeq \frac{1}{n}\left(f\left(X_1\right)+...f\left(X_n\right)\right)$$

In practice, there may exist situations where sampling from μ is not easy at all, especially in high dimension problems. Instead, it is relatively easier to simulate a Markov chain with stationary distribution μ. If this chain is irreducible, the ergodic Theorem D.1.1 then ensures the convergence, in the place of the law of large numbers approximation. A general scheme to construct such a Markov chain is given by the Metropolis-Hastings algorithm, cf. Robert and Casella (2004) [142] for a more complete presentation.

D.1.1 Metropolis-Hastings Algorithm

The general principle is as follows. We aim to construct a Markov chain with reversible measure μ, i.e.: with transition matrix $q(x,y)$ satisfying $\mu(x)\,q(x,y) = \mu(y)\,q(y,x)$, where $\mu(x) > 0, \forall x \in E$. The easier way to proceed is to use a positive symetric function $k(x,y)$ s.t.:

$$q(x,y) = \frac{k(x,y)}{\mu(x)},\ x \neq y \tag{D.1}$$

and $q(x,x) = 1 - \sum_{x \neq y} q(x,y)$. To be sure that (D.1) is always positive, we can choose a irreducible Markovian matrix $p(x,y)$ and let:

$$k(x,y) = \mu(x)\,p(x,y) \wedge \mu(y)\,p(y,x),\ x \neq y$$

From this, $q(x,y) \leqslant p(x,y)$ and:

$$\sum_{x \neq y} q(x,y) \leqslant 1$$

Another advantage of this situation is that it comes:

$$q(x,y) = p(x,y) \wedge \frac{\mu(y)}{\mu(x)} p(y,x)$$

and so:

$$q(x,y) \propto \frac{\mu(y)}{\mu(x)}$$

which may be largely easier to compute than $\mu(x)$. The choice of the matrix $q(x,y)$ depends on the structure of the problem. Some simplifactions arises when $p(x,y) = p(y,x)$, cf. Metropolis et al. (1953) [129], and so:

$$q(x,y) = \left(\frac{\mu(y)}{\mu(x)} \wedge 1\right) p(x,y) \tag{D.2}$$

The Metropolis-Hastings algorithm consists in a two-step procedure. The construction of the Markov chain (X_n) can be described as follows.

Algorithm D.1.2 (Metropolis-Hastings Algorithm) *We have:*

1. *If $X_n = x$, sample $Y_n \sim p(x,y)$. When $Y_n = y$, let:*

$$p = \frac{\mu(x)\,p(y,x)}{\mu(y)\,p(x,y)} \wedge 1$$

2. *Sample Z_n s.t. $P(Z_n = 1) = p$, then:*

$$X_{n+1} = \begin{cases} X_n = x & \text{if } Z_n < 1 \\ Y_n = y & \text{elsewhere} \end{cases}$$

Remark D.1.3 *Algorithm D.1.2 when (D.2) do not hold, i.e.: $p(x,y) \neq p(y,x)$, is referred as the Hastings (1970) [86] algorithm.*

6

Joint Calibration of Option Pricing Models

Agenda

Note: This chapter is an adapted version of a co-authored paper with B. Amzal and Y. Ebguy.

Abstract. This chapter develops a new joint calibration procedure where both implied and historical distributions are simultaneously used. We closely link it to the minimization of measures of risk under uncertainty or robust problems. We focus on the class of affine jump-diffusion models to derive a parametric formulation of the problem. From both theoretical and practical points of view, we are concerned with the quantitative assessment of the financial risk to build a robust and efficient pricing system. As a loss function, we choose the sum of a classical least-square cost and of a regularizing term of relative entropy. Then, in order to develop a computationally efficient methodology, we translate it into its probabilistic counterpart in a general context of maximization of expected utility (MEU). This leads us to the development of simulation-based algorithms in line with Monte-Carlo Markov Chains (MCMC) methods. To avoid their traditional shortcomings like local mode trapping, we consider an original alternative derived from Interacting Particle Systems (IPS). A new theoretical framework for this method is provided and convergence results are established. This algorithm is applied to simulated data and to a EuroStoxx 50 data set. We extensively discuss these results and interpret them notably in terms of risk aversion and models perception.

6.1. Introduction

Stocks or interest-rate derivatives are priced under the risk neutral measure while the statistical measure is used to model the underlying on which the option is written. Classical calibration approaches such as those of Avellaneda (1998) [11] or Bates (1996) [14] only resort to the implied measure to fit the observed smile of volatility ignoring information available in the historical measure. Thus, one may wish to build a joint calibration procedure where both implied and historical distributions are simultaneously used. It is worth noting that several attempts have been made to answer these questions as in Chernov and Ghysels (2000) [26] or in Eraker (2004) [61], but these papers didn't display satisfying results: the risk premiums were erratic, and their interpretation not obvious. In this article, we make a new attempt to build a bridge between these two worlds. We propose a calibration method using the joint information from market prices of some options as well as of the corresponding stocks on which the positions were written.

We consider an original prediction/calibration problem. For example, at time t, we want to predict some option prices on the coming days. It consists of *predicting* the underlying value with its dynamics under the historical probability and then calculating the corresponding option prices with the risk-neutral actualized expectation. In that sense, the *a posteriori* problem to solve reads as follows. Given a set of data and a chosen day in this time series, what are the historical and risk-neutral measures that would have implied the best prediction for the option prices quoted on the days following the chosen one? This calibration is a kind of prediction in-sample. Formally, this is expressed in the following inverse problem:

Definition 6.1.1 (Joint Calibration) *Given a set of stock prices and liquid call options written on this stock, say* $\{Y_t^*, C^*(Y_t, T_i, K_i)\}$ *for* $i = \{1...I\}$ *and* $t = \{1...T\}$*, find two random measures* $\mathbb{Q} \in \mathbf{Q}$*, the risk-neutral one, and* $\mathbb{P} \in \mathbf{P}$*, the historical one, such that the observed option prices are given by the* $\mathbb{P}-$*expectation of their theoretical prices on stock values simulated under* $\mathbb{P}$*:*

$$C^*(Y_t^*, T_i, K_i) = B(t, T_i)\mathbb{E}_{t-1}^{\mathbb{P}}\left[C^{\mathbb{Q}}(Y_t^{\mathbb{P}}, T_i, K_i)|Y_{t-1}\right] \tag{6.1}$$

- $Y_t^{\mathbb{P}}$ *is simulated under* $\mathbb{P}$ *from* Y_{t-1}^* *at time* $t-1$*,*
- $B(t, T_i)$ *is the discount factor at time* t *with maturity* T_i*,*
- $C^{\mathbb{Q}}(Y_t^{\mathbb{P}}, T_i, K_i)$ *is the theoretical price under the risk-neutral measure* $\mathbb{Q}$ *of the option considered with strike* K_i*, maturity* T_i *and spot value* $Y_t^{\mathbb{P}}$*.*

The spaces of probability measure $\mathbf{P}$ and $\mathbf{Q}$ need to be specified so that both measures are equivalent and that arbitrage opportunities are avoided. This goal is normally met by assigning market price of risk process(es) to the dynamics of the state variable(s) as described by Harrison and Kreps (1979) [84] and Harrison and Pliska (1981) [85]. This results in the fact that if a process is within the class of affine jump-diffusion, cf. Duffie et al. (2003) [56], under the objective probability measure, the market price of risk specification ensures that it is within the same class under the equivalent martingale measure and vice-versa. The bridge alluded to earlier will also be achieved through the assignment of market prices of risk. This enables one to preserve equivalence of measures and precludes arbitrage opportunities. Therefore, our joint calibration approach will succeed if we manage to separate the impact of the historical parameters and that of the risk-neutral ones. It could help one both to calibrate models and to determine parameters usually estimated roughly from historical data with a better understanding of their impact and of their evolution.

But, at this stage, the inverse problem (6.1) is ill-posed, cf. Engl et al. (1996) [59]. Indeed, it is under-determined: the knowledge of a finite numbers of option prices is not enough to characterize the risk neutral measure. There might be no solution or an infinite number of solutions, not necessarily in the class of models we were expecting it to be. Traditionally, in the calibration issue, this obstacle is bypassed by minimizing the in-sample quadratic pricing error, cf. Christoffersen and Jacobs (2001) [29]. Inspiring ourselves from that idea to obtain a practical solution, we resort to minimizing a *measure of risk*, or more precisely to minimize the expectation of a loss-function L, defined on $\mathbf{Q}$, the probability measure space to be chosen. The concept of a measure of risk advanced in Artzner et al. (1999) [8] and refined in Foellmer and Schied (2002) [68] and in a large subsequent literature has gained widespread acceptance in the option pricing industry. This approach focuses on the measure of the quantitative risk involved in a financial position. The goal is now to find:

$$(\mathbb{P}, \mathbb{Q}) = arg \min_{(\mathbf{P}, \mathbf{Q})} \mathbb{E}^{\mathbb{P}} \left[\mathbb{L} \left(\mathbb{Q} \right) \right] \tag{6.2}$$

This formulation has a mathematical sense if the loss function has good properties. But it could seem numerically difficult to solve. Usual ways of optimizing a function like BFGS gradient-descent method would be here non-efficient because of the too numerous calculations of the expectation needed. We thus have to resort to other numerical methods. Following the fast development of computers, simulation techniques have appeared to be a more and more interesting alternative for analytic or algebraic approach of optimization problems. MCMC algorithms have been widely developed and applied this last decade for Bayesian problems, cf. Robert and Casella (2004) [142], in finance as well as in many other fields of application. The optimization issue is transformed into a simulation one, cf. Section 5.1, from which particle methods appear to be adapted to our problem. A recently born algorithm, presented in Section 5.3, inspired by Interacting Particle Systems which doesn't demand any calculation of the expectation is used. New powerful convergence properties which help us for the practical implementation are proved in Appendix E. We note that this departs from the methodology recently presented in Ben Hamida and Cont (2004) [17].

The remainder of the chapter is organized as follows. Section 2 sets the jump-diffusion models considered and defines $\mathbf{P}$ and $\mathbf{Q}$, which gives a parametric formulation of the optimization problem. In Section 2, we also explain the choice of our loss-function, adding to the traditional least-squares cost, an entropy term, regularizing the problem. In Section 3, this algorithm applied to some specific and largely used option pricing models derived from Section 2, and the numerical results obtained with simulated data are discussed. In section 4, a real-world application is given and risk premiums analysis is derived. At end, section 5 concludes.

6.2. Pricing Models Driven by Affine Jump Diffusions

We begin with some definitions and properties of affine jump-diffusion models. Moreover, we precise the spaces of probability measures **P** and **Q** on which the optimization is made and we discuss market price of risk specifications and absence of arbitrage for this class. Finally, we parametrize the joint calibration problem for a specific family of stochastic volatility with jumps models. Throughout this section, $\mathbb{P}$ corresponds to the historical probability, and $\mathbb{Q}$ to the risk-neutral one.

6.2.1 Affine Jump Diffusions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which are defined the two following independent random processes: a $d-$dimensional standard Brownian motion $(W_t)_{t \geqslant 0}$ and a $d-$dimensional compound Poisson process $(N_t)_{t \geqslant 0}$. We also suppose that there is a Markov process Y taking values in some open subset $\mathcal{D}$ of $\mathbb{R}^d$ and satisfying the following theorem, cf. Duffie et al. (2003) [56]. Let an affine jump-diffusion (AJD) be given by:

$$\begin{aligned} dY_t &= \mu\left(Y_t, t\right) dt + \Sigma\left(Y_t, t\right) dW_t + d\widetilde{N}_t \\ &= \mu\left(Y_t, t\right) dt + \Sigma\left(Y_t, t\right) dW_t + \int_{\mathcal{D}} N\left(dy, dt\right) - \nu\left(dy, dt\right) \end{aligned} \tag{6.3}$$

where $\mu : \mathcal{D} \to \mathbb{R}^d$ and $\Sigma : \mathcal{D} \to \mathbb{R}^{d \times d}$ are deterministic functions, while $\widetilde{N}_t$ is a compensated Poisson process with compensator $\nu\left(dy, dt\right)$. The Lévy measure ν dictates how jumps occur. In a finite activity setup, jumps arrive with intensity $\lambda : \mathcal{D} \to \mathbb{R}^d$ with $\lambda < \infty$ and are distributed according to a fixed probability distribution m on $\mathbb{R}^d$, with $\nu\left(dy, dt\right) = \lambda(y) dm\left(y, t\right)$.

Theorem 6.2.1 (AJD Characterization) *An AJD process satisfies:*

- *(affinity) Drift, squared volatility and intensity are all affine such that the determining triplet of characteristics $\varkappa = (\mu, \Sigma, \lambda)$ writes as follows:*

$$\begin{aligned} \mu &= k_0\left(t\right) + k_1\left(t\right) \cdot Y \\ \Sigma\Sigma^{\intercal} &= h_0\left(t\right) + h_1\left(t\right) \cdot Y \\ \lambda &= l_0\left(t\right) + l_1\left(t\right) \cdot Y \end{aligned} \tag{6.4}$$

- *(ode-s) Coefficients are such that solutions β and α to the following system of ordinary differential equations exist:*

$$\beta'\left(t\right) = -k_1\left(t\right)^{\intercal} \beta\left(t\right) - \frac{1}{2}\beta\left(t\right)^{\intercal} h_1\left(t\right) \beta\left(t\right) - l_1\left(t\right) \int_{\mathcal{D}} e^{\beta(t) \cdot y} dm\left(y\right) - 1$$

$$\alpha'(t) = -k_0(t)^\intercal \beta(t) - \frac{1}{2}\beta(t)^\intercal h_0(t)\beta(t) - l_0(t)\int_{\mathcal{D}} e^{\beta(t)\cdot y} dm(y) - 1$$

with boundary conditions $\beta(T) = u$ *and* $\alpha(T) = 0$*, where* $u \in \mathbb{R}^d$.

- *Then, for each* $u \in \mathbb{R}^d$*, the discounted characteristic function process*

$$\phi^{\varkappa}(u, Y_t, t, T) = \mathbb{E}^{\mathbb{Q}}\left(e^{-r(T-t)}e^{u\cdot Y_T}|\mathcal{F}_t\right) \tag{6.5}$$

has exponential affine form in X*, namely*

$$\phi^{\varkappa}(u, Y_t, t, T) = \exp(\alpha(t) + \beta(t)\cdot Y_t)$$

Let us consider a contingent claim written on Y whose profile is given by a final cash amount $g(Y_T) \equiv g(Y_T, \cdot)$ and whose value at time t is:

$$P_t = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}g(Y_T, T, K)\right]$$

The general result stated above allows a direct pricing of such claim, in the case of affine pay-off functions in the factor Y, via Fast Fourier transform, cf. Carr and Madan (1998) [24], methods. In fact, we can use the general result in Theorem 6.2.1 to compute characteristic function distributions and then numerically invert them to recover the corresponding density function. This leads to a general call option formula for AJD stock price models:

Proposition 6.2.2 (Option Pricing) *Let us consider a call option on a function* g *of the underlying factor expiring at time* T *and striking at price* K*. For any non-decreasing function*[1] h*, the price of the call is then:*

$$\begin{aligned} P_t &= \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}(g(Y_T) - K)^+\right] \\ &= f_g(-h(K)) - Kf_1(-h(K)) \end{aligned} \tag{6.6}$$

with 1 *the identity function and where:*

$$f_g(y) = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}g(Y_T)\mathbf{I}_{\{-hog(Y_T)\leqslant y\}}\right] \tag{6.7}$$

can be calculated by inverting its Fourier transform as

$$\begin{aligned} F[f_g](u) &= \int_{\mathbb{R}} e^{iux}\mathbb{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}g(Y_T)\mathbf{I}_{\{-hog(Y_T)\leqslant x\}}\right]dx \\ &= \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}g(Y_T)\int_{\mathbb{R}} e^{iux}dx\mathbf{I}_{\{-hog(Y_T)\leqslant x\}}\right] \\ &= \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}g(Y_T)e^{iu(-hog(Y_T))}\right] \end{aligned}$$

where h *has to be chosen such that the expectation (6.7) is of the form (6.5).*

[1] h is s.t. $g(Y_T) \leqslant K \Leftrightarrow h(g(Y_T)) \geqslant h(K)$.

6.2.2 Market Price of Risk Specifications

The following proposition is a consequence of Girsanov transformation for semi-martingales, cf. Jacod and Shiryaev (2003) [93]. In all the sequel, we use the notation $f * \nu$ meaning that we integrate the function f w.r.t. the measure ν (both being defined on $\mathbb{R}^+ \times \mathbb{R}^d$).

Proposition 6.2.3 (Girsanov Change of Measure) *Let Y be an AJD with $\mathbb{P}-$characteristics $\left(\mu^{\mathbb{P}}, \Sigma^{\mathbb{P}}, \nu^{\mathbb{P}}\right)$. For any probability measure $\mathbb{Q} \ll \mathbb{P}$, there exist a predictable function $\Lambda_3 > 0$ and a predictable $\mathbb{R}^d-$valued process Λ_1 such that $\mathbb{Q}-$characteristics of Y are given by:*

$$\begin{aligned}
\mu^{\mathbb{Q}} &= \mu^{\mathbb{P}} + \Sigma^{\mathbb{P}} \cdot \Lambda_1 + (\Lambda_3 - 1) * \nu^{\mathbb{P}} \\
\Sigma^{\mathbb{Q}} &= \Sigma^{\mathbb{P}} \\
\nu^{\mathbb{Q}} &= \Lambda_3 \cdot \nu^{\mathbb{P}}
\end{aligned} \tag{6.8}$$

Λ_1 and Λ_3 are called the Girsanov quantities of $\mathbb{Q}$ with respect to $\mathbb{P}$ relative to Y. Intuitively, Λ_3 describes how the jump distribution of Y, cf. Esche (2003) [62], changes when we turn from the historical measure $\mathbb{P}$ to the risk-neutral one $\mathbb{Q}$ and Λ_1 together with Λ_3 determines the change in drift.

To give a deeper insight of market price of risk specifications, one can express the density process $Z^{\mathbb{Q}}$ of $\mathbb{Q}$ with respect to $\mathbb{P}$ explicitly in terms of Λ_1 and Λ_3. If we denote the Doleans-Dade exponential $\mathcal{E}$, we have the following proposition based on weak representation property for semi-martingales, cf. Jacod and Shiryaev (2003) [93].

Proposition 6.2.4 *Let Y be an AJD. If $\mathbb{Q} \ll \mathbb{P}$ with Girsanov quantities Λ_1 and Λ_3, the density process of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ is given by $Z^{\mathbb{Q}} = \mathcal{E}\left(N^{\mathbb{Q}}\right)$ with:*

$$N_t^{\mathbb{Q}} = \left(\Lambda_1 \cdot Y^c + (\Lambda_3 - 1) * \left(m^{\mathbb{P}} - \nu^{\mathbb{P}}\right)\right)_t \tag{6.9}$$

If $\mathbb{P}$ exists and is solution of a martingale problem for the process (6.3), existence of the quantities Λ_1 and Λ_3 is sufficient neither for the existence of the implied probability measure $\mathbb{Q}$ nor for its equivalence to $\mathbb{P}$, cf. Jacod and Shirayev (2003) [93]. However, these two properties are jointly entailed by the following necessary and sufficient conditions:

$$\mathcal{E}\left(N^{\mathbb{Q}}\right)_{.} > 0, \ \mathbb{E}^{\mathbb{P}}\left[\mathcal{E}\left(N^{\mathbb{Q}}\right)_t\right] = 1 \tag{6.10}$$

The existence of an equivalent martingale measure implies the absence of arbitrage opportunity, cf. Delbaen and Schachermayer (1994) [43]. Our specification for market price of risk in AJD models precludes arbitrage opportunities and the models under both measures are of the same class.

Theorem 6.2.5 (Equivalence between AJDs) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space containing a $d-$Brownian motion $\left(W_t^{\mathbb{P}}\right)_{t\geqslant 0}$ and a $d-$compensated Poisson process $(\widetilde{N}_t^{\mathbb{P}})_{t\geqslant 0}$ such that there exists a stochastic process $(Y_t)_{t\geqslant 0}$ satisfying (6.3). Then, there exists a measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that:*

$$dY_t = \mu^{\mathbb{Q}}(Y_t, t)\, dt + \Sigma(Y_t, t)\, dW_t^{\mathbb{Q}} + d\widetilde{N}_t^{\mathbb{Q}} \tag{6.11}$$

such that $(W_t^{\mathbb{Q}})_{t\geqslant 0}$ is a $\mathbb{Q}-$Brownian motion and $(\widetilde{N}_t^{\mathbb{Q}})_{t\geqslant 0}$ a $\mathbb{Q}-$compensated Poisson process.

6.2.3 Affine Pricing Jump Diffusions

In $(\Omega, \mathcal{F}, \mathbb{P})$ we suppose that Y is the price process, strictly positive, of a security that pays no dividend. The state process is $Z = (Y, V)^{\intercal}$ where V is the volatility process. Under the statistical measure $\mathbb{P}$, the dynamics of Z characterized by the triplet $\left(\mu^{\mathbb{P}}, \Sigma, \nu^{\mathbb{P}}\right)$ is given by:

$$\frac{dZ_t}{Z_{t-}} = \mu_t^{\mathbb{P}}(\gamma)\, dt + \Sigma_t(\gamma)\, dW_t^{\mathbb{P}} + \int N^{\mathbb{P}}(dz, dt) - \nu^{\mathbb{P}}(\gamma)(dz, dt) \tag{6.12}$$

where $W^{\mathbb{P}}$ and $N^{\mathbb{P}}$ are respectively a Brownian motion and a Poisson process, while $\nu^{\mathbb{P}}(dz, dt) = \lambda^{\mathbb{P}} dm^{\mathbb{P}}(z, t)$ is the random measure attached with jumps which arrive with intensity λ and mark according to the probability distribution $m^{\mathbb{P}}$. The probability measure $\mathbb{P}$ is completely defined by the parameters $\gamma^{\mathbb{P}}$ of the Brownian motion and of the Lévy measure, so that we have the set of $\mathbb{P}-$parameters:

$$\gamma^{\mathbb{P}} = \left(\gamma\left(W^{\mathbb{P}}\right), \gamma\left(\nu^{\mathbb{P}}\right)\right)$$

Given a specification of $\left(\mu^{\mathbb{P}}, \Sigma, \nu^{\mathbb{P}}\right)$ such that a solution of (6.12) exists, we may consider the existence of an equivalent martingale measure $\mathbb{Q}$ by specifying market price of risk Λ_1 and Λ_3 which satisfies equations (6.8). Under this measure, Z is an affine jump diffusion with triplet $\left(\mu^{\mathbb{Q}}, \Sigma, \nu^{\mathbb{Q}}\right)$ such that:

$$\frac{dZ_t}{Z_{t-}} = \mu_t^{\mathbb{Q}}(\gamma)\, dt + \Sigma_t(\gamma)\, dW_t^{\mathbb{Q}} + \int N^{\mathbb{Q}}(dz, dt) - \nu^{\mathbb{Q}}(\gamma)(dz, dt) \tag{6.13}$$

Similarly, the measure $\mathbb{Q}$ can be specified by a set of Brownian motion and Lévy measure parameters, so that we define the set of $\mathbb{Q}-$parameters:

$$\gamma^{\mathbb{Q}} = \left(\gamma\left(W^{\mathbb{Q}}\right), \gamma\left(\nu^{\mathbb{Q}}\right)\right)$$

By Girsanov's theorem for AJD processes, cf. Proposition 6.2.3, we define the market price of risk $\Lambda = (\Lambda_1, \Lambda_3)$ as solutions of the following system:

$$\begin{aligned} \mu_t^{\mathbb{Q}} &= \mu_t^{\mathbb{P}}(\gamma) + \Sigma_t(\gamma) \cdot \Lambda_1 + \int (\Lambda_3(z) - 1)\, dm^{\mathbb{P}}(z,t) \\ W^{\mathbb{Q}} &= W^{\mathbb{P}} - \int \Lambda_1 dt \\ \lambda^{\mathbb{Q}} dm^{\mathbb{Q}}(z,t) &= \Lambda_3(z) \cdot \lambda^{\mathbb{P}} dm^{\mathbb{P}}(z,t) \end{aligned}$$

Remark 6.2.6 *This presentation of a specific affine jump-diffusion framework raises attention to a number of common pricing models. The general formulation (6.12), or (6.13), includes the pioneering models of Black and Scholes (1973) [19] or Merton (1976) [128] in the univariate case. Another widely studied specification is the class of stochastic volatility models with the derivation of Heston (1993) [88], Bates (1996) [14] and in a jump-diffusion approach those of Eraker (2004) [61].*

Finally, to cast specifications (6.12) and (6.13) into an estimation framework, we define the set of parameters involved in our optimization procedure:

$$\gamma = \left(\gamma^{\mathbb{Q}}, \Lambda\right) \in \mathbb{R}^m$$

At this stage, the objective is to estimate the model-dependent set of parameters γ. What remains to be done to have a full formulation of the problem is to choose a relevant loss function.

6.2.4 Towards a Well-Posed Optimization Problem

Regularizing by Relative Entropy

A natural loss function $L(\gamma)$ interesting for the risk minimization (6.2) is the quadratic pricing error, cf. Christoffersen and Jacobs (2001) [29]. The problem would then become. Given a data set $\overline{\mathcal{D}} = \{Y_t^*, C^*(Y_t^*, T_i, K_i)\}$, find:

$$\gamma^* = arg \min_{\gamma \in \mathbb{R}^m} L(\gamma)$$

with:

$$L(\gamma) = \sum_{i,t} \omega_i \mathbb{E}_{Y_t|\gamma, Y_{t-1}^*} \left| C_t^{\gamma}(Y_t, T_i, K_i) - C^*(Y_t^*, T_i, K_i) \right|^2$$

where:

- ω_i are the weights, to be chosen by the decision maker,
- Y_t is simulated under $\mathbb{P}^{\gamma}$, the objective measure given γ knowing Y_{t-1}^*,
- $C_t^{\gamma}(Y_t, T_i, K_i)$ is the $t-$time theoretical price with spot Y_t, maturity T_i and strike K_i under $\mathbb{Q}^{\gamma}$, the pricing measure induced by γ.

But, as Cont and Tankov (2004) [32] noticed, from which this subsection is widely inspired, this formulation, though giving a statistical sense to the problem, doesn't resolve the uniqueness and stability issues. Indeed, L is not convex, so many local minima might exist or some flat directions might make the solution get unstable. It is an ill-posed problem, cf. Engl et al. (1996) [59]. The usual way to cope with that issue is to introduce, as in Cont and Tankov (2004) [32], a penalization term, namely the relative entropy of the pricing measure $\mathbb{Q}^\gamma$ with respect to some prior model $\mathbb{Q}_0$ explicitly given by:

$$\mathbf{I}_t\left(\mathbb{Q}^\gamma|\mathbb{Q}_0\right) = \mathbb{E}^{\mathbb{Q}^\gamma}\left[\log\frac{d\mathbb{Q}^\gamma}{d\mathbb{Q}_0}|\mathcal{F}_t\right]$$

The relative entropy has a few interesting properties which make it relevant to use as a penalization term:

- Financial issue: if $\mathbb{Q}^\gamma$ is not absolutely continuous w.r.t. the prior, $\mathbf{I}_t\left(\mathbb{Q}^\gamma|\mathbb{Q}_0\right)$ becomes infinite. Thus, if the prior is well chosen, we can impose good properties to $\mathbb{Q}^\gamma$. For example, we might take an auxiliary simpler diffusion model, which will be easier to calibrate.
- Numerical aspect: $\mathbf{I}_t\left(\mathbb{Q}^\gamma|\mathbb{Q}_0\right)$ is convex in the different parameters, so the penalization term has a *convexification impact* on the surface to optimize which brings stability to the solution(s).
- Information-theoretic foundation: minimizing $\mathbf{I}_t\left(\mathbb{Q}^\gamma|\mathbb{Q}_0\right)$ corresponds to adding the least possible information to the prior to fit in the best way with the option prices (and implicitly with the historical evolution of the underlying). So it introduces a tradeoff between the accuracy of the fit (information contained in option prices) and the numerical stability of the results (information contained in the prior). It could therefore also be interesting to take the objective measure as a prior.

The following result shows that in the case where the measures are generated by affine jump-diffusions, the relative entropy can be expressed in terms of the Girsanov parameters and $\mathbb{Q}_0-$characteristics of $X = \log Y$.

Proposition 6.2.7 (Relative Entropy) *If $\mathbb{Q}^\gamma \ll \mathbb{Q}_0$ with Girsanov quantities Λ_1 and Λ_3, the entropy process of $\mathbb{Q}^\gamma$ w.r.t. $\mathbb{Q}_0$ is explicitly given by:*

$$\mathbf{I}_t\left(\mathbb{Q}^\gamma|\mathbb{Q}_0\right) = \frac{1}{2}\mathbb{E}^{\mathbb{Q}^\gamma}\left[\Lambda_1\cdot\Sigma\right] + \mathbb{E}^{\mathbb{Q}^\gamma}\left[f\left(\Lambda_3\right)*\nu^{\mathbb{Q}_0}\right] \tag{6.14}$$

where $f(y) = y\log(y) - (y-1)$.

Proof. Denote by $T = T^{\mathbb{Q}^\gamma} = \mathcal{E}(N)$ the density process with respect to $\mathbb{Q}_0$. The canonical decomposition of the $\mathbb{Q}_0-$submartingale $T\log(T)$ is $T\log(T) =$

$M + A$ with:

$$M = \int Z_{-} \left(1 + \log T_{-}\right) dN + \left(T_{-} f\left(\Lambda_3\right)\right) * \left(m^{\mathbb{Q}_0} - \nu^{\mathbb{Q}_0}\right)$$

$$A = \frac{1}{2} \int T_{-} d\left\langle N^c \right\rangle + \left(T_{-} f\left(\Lambda_3\right)\right) * \nu^{\mathbb{Q}_0}$$

where M is a local $\mathbb{Q}_0$-martingale and A is predictable and of finite variation, cf. Appendix **??** and especially Proposition A.2.3. The quadratic variation $\left\langle N^c \right\rangle_t = \Lambda_1 \cdot \Sigma$ is the same under both measures $\mathbb{Q}_0$ and $\mathbb{Q}^\gamma$. Hence:

$$\begin{aligned} \mathbf{I}_t\left(\mathbb{Q}^\gamma | \mathbb{Q}_0\right) &= \mathbb{E}^{\mathbb{Q}_0}\left[T_t \log T_t\right] = \mathbb{E}^{\mathbb{Q}_0}\left[A_t\right] \\ &= \frac{1}{2} \mathbb{E}^{\mathbb{Q}^\gamma}\left[\Lambda_1 \cdot \Sigma\right] + \mathbb{E}^{\mathbb{Q}^\gamma}\left[f\left(\Lambda_3\right) * \nu^{\mathbb{Q}_0}\right] \end{aligned}$$

which gives the desired result. ■

The Well-Posed Optimization Problem

Let us consider the following regularized problem, where α is the weight given to the accuracy (or to the stability):

$$\gamma^* = arg \min_{\gamma \in \mathbb{R}^m} L^\alpha(\gamma)$$

with:

$$L^\alpha(\gamma) = \sum_{i,t} \omega_i \mathbb{E}_{Y_t | \gamma, Y^*_{t-1}} \left| C^\gamma_t\left(Y_t, T_i, K_i\right) - C^*\left(Y^*_t, T_i, K_i\right)\right|^2 + \alpha \mathbf{I}_t\left(\mathbb{Q}^\gamma | \mathbb{Q}_0\right)$$

The role of α is clearly important and its value should depend on the data used (which governs the shape of the function to be optimized) and on the loss of precision due to the introduction of the entropy term. This corresponds to what is called an *a posteriori* choice of α. We won't detail here the way to determine a good value for α, the interested reader will refer to Cont and Tankov (2004) [32]. Their determination is based on the Morozov discrepancy principle, as described in Engl et al. (1996) [59]. This way to determine α gives the convergence of the solution towards a minimum entropy least squares solution when the error level allowed alluded to earlier tends to zero. There are two main advantages for this new formulation:

- The first one is that it transforms the ill-posed problem into a well-posed one. The existence of the solution is easy to prove. Let us give here an idea of the proof. We assume simple conditions on the jumps and on the normal laws of the Brownians to prevent the underlying level from being multiplied by more than 100 from one day to another (which is always satisfied in practice). Then the corresponding prices of the options are bounded. So the least squares term is bounded. Besides, the entropy term explodes as soon as the parameters are far from the prior ones. So the infimum of L^α is in a compact set. As L^α is continuous with respect to γ, this infimum is a minimum. For α big enough, we could show that this solution is unique. We now seek γ in a compact set $\Gamma \subset \mathbb{R}^m$

- In the neighborhood of the minimum, the surface to explore is more convex which helps one to locate the minimum sought.

Now, we can say that the problem is well-posed in the two senses of the term: theoretically because it has an *admissible* solution and practically because it will be easier to solve.

Casting the MEU Setup

Subsequently, we cast the general framework in which we embed our optimization problem into the general context of maximization of expected utility, cf. Section 5.1. Our idea is based on a practical point of view of calibration: we want to interpret such calibration or inference problem as a decision one. Indeed, given our data set $\overline{\mathcal{D}} = \{Y_t^*, C^*(Y_t^*, T_i, K_i)\}$ we define a utility function to be maximized, namely $U_{\overline{\mathcal{D}}}(\gamma)$ or more simply $U(\gamma)$, such that:

- if we denote u the joint utility with:

$$u(\gamma, y) = M - \left(\sum_{i,t} \omega_i \left| C_t^\gamma (Y_t, T_i, K_i) - C^* (Y_t^*, T_i, K_i) \right|^2 + \alpha \mathbf{I}_t (\mathbb{Q}^\gamma | \mathbb{Q}_0) \right)$$

- if the parameter γ is the chosen one for calibration in Γ,
- if $Y \in \mathcal{Y}$ is the vector of *predictive* data $(Y_t\,,\, t = 1..T)$ drawn from density distribution $p_{\overline{\mathcal{D}}}(Y\,|\gamma) \equiv p(Y|\gamma)$ and defined as:

$$p(Y|\gamma) = \delta_{Y_1^*} \otimes p(Y_2\,|\, Y_1^*, \gamma) \otimes ... \otimes p(Y_T|\, Y_{T-1}^*, \gamma)$$

where $\mathcal{Y}$, the space of underlying variables, is taken without loss of generality as $\mathbb{R}^+$ (for stochastic volatility models, the dimension is 2)

- if the constant M is such that u is strictly positive, then:

$$\gamma^* = \arg\max_{\gamma \in \Gamma} U(\gamma) = \arg\max_{\gamma \in \Gamma} \mathbb{E}_{p\left(Y_t | \gamma, Y_{t-1}^*\right)} \left[u(\gamma, y) \right]$$

where u is bounded and continuous over Γ.

We are now facing a utility maximization problem, in the statistical sense of it. However, the utility function $U(\gamma)$ is very costly to compute in our case, and classical optimization methods (like gradient methods) would not be practicable. We then propose to use an alternative based on a Markov Chain Monte Carlo with Interacting Particle System algorithm, cf. Section 5.3.

6.3. Reliability of the Joint Procedure

In this section, we present practical applications of the particle algorithm to the introductory joint calibration Problem 6.1, in order to test on simple examples the effectiveness of the method. For this purpose, we calibrate the historical and risk-neutral parameters of two affine pricing models: Black-Scholes and Heston model (with latent stochastic volatility).

Remark 6.3.1 *We termed estimation the determination of the parameters of interest, even if we work within a decision analysis framework. Nevertheless, it does not refer to the estimation in the statistical sense, and confidence intervals do not really make sense in such a context.*

6.3.1 Implementation Issues

Two Markov kernels $K_{1,n}$ and $K_{2,n}$ are required, as well as the number N of particles and the cooling schedule $J(n)$. As Markov proposals, we simply choose truncated normal random walks as evoked in Section E.5. As for sample size N, central limit type inequalities would lead us to typical values of order $N = 10,000$ particles to give accurate Monte-Carlo integral approximations. However, as we are not interested by integral approximations but only by the mode of target distributions, this number might be too large for our purpose.

In practice, it appears that N below $1,000$ gives good results in most cases. Specifically, we take $N = 500$. Note that this sample size should obviously be larger for the RM Algorithm 5.3.2, since in practice, it suffers more from degeneracy. As cooling schedule, our theoretical arguments suggest a logarithmic form for small values of J, and a linear form for large ones, when the algorithm is almost *reduced* to the RM Algorithm 5.3.2. For computational convenience, we simply take a linear cooling scheme from 1 to 31 by a time-step of 2.

In the computation of the objective function L defined in Section 6.2.4, we choose weights of the form:

$$\omega_i = \frac{1}{\text{Vega}\left(T_i, K_i\right)}$$

where Vega$(\cdot)$ is the Black-Scholes vega of the option computed using the market implied volatility. The interest of this weight, as noted by Cont and

Tankov (2004) [32], is that it converts errors in price in errors in implied volatility, thus rescaling all terms entering in L to the same order of magnitude.

For each day of our sample, we need to produce an estimate of the spot volatility level. A practical way to obtain this was to evaluate this volatility from an historical analysis. The statistical estimate used was a moving empirical average variance in order to capture day-to-day effects and both persistence and clustering of volatility.

Regarding the entropy term, it caused numerical problems linked to the choice of the constant added to make U positive. Indeed, after 3 iterations, the entropy term explodes and the calculation of the coefficient α gets hazardous so we could not use in a satisfying way the advantages of the entropy. However, the expected convexification effect was actually observed these 3 first iterations.

6.3.2 Application to a Simulated Data Set

Using Black-Scholes Model

In the first simulation study, we generated a panel of stocks and option prices using the well-known Black and Scholes (1973) [19] model. One main advantage of this specification is its simplicity which allows to precisely test the cross-sectional effect. Following market price of risk specification enlightened in Section 6.2, we fix $\mathbb{P}$ and $\mathbb{Q}$ parameters respectively to:

$$(\mu;\sigma) = (0.08; 0.4)\,,\ (\mu^* = r;\sigma) = (0.037; 0.4)$$

A 30−days long returns series was simulated under $\mathbb{P}-$parameters and option prices were computed under the measure $\mathbb{Q}$ using the Black-Scholes formula. The maturity of theses options was one year and we used 21 equidistant strikes from 30 to 50 for a spot being at 40.

Our approach to capture cross-sectional effect is made to produce independent estimations with different stocks returns length, beginning from a short horizon of 5 days and iteratively by a time-step of size 5 reaching a 30 days long horizon. We estimate volatility σ, (historical) drift $\mu^{\mathbb{P}}$ and market price of risk λ parameters. The estimates are reported in Table 6.1. They correspond to optimal parameter estimates, i.e.: the value corresponding to the minimum of our utility functional (overall option fit).

A number of broad points emerge from Table 6.1. Longer time horizons provide significant pricing improvement than shorter horizons, as the LSE's (least square error) of the pricing errors are relatively stable even though the parameters from time $T = 5$ to time $T = 30$ days are progressively larger and more difficult to identify. Once again, let us recall that we constrain the risk-neutral parameters to be time-consistent with the objective measure dynamics

	Horizon (days)					
Parameters	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 25$	$T = 30$
$\mu^{\mathbb{P}}$ (0.08)	0.068	0.074	0.086	0.072	0.073	0.088
$\sigma^{\mathbb{P},\mathbb{Q}}$ (0.4)	0.344	0.355	0.372	0.391	0.429	0.433
λ (0.1)	0.073	0.076	0.144	0.102	0.097	0.129
LSE	0.58	1.03	1.75	1.85	2.13	2.81

Table 6.1: Black-Scholes Model Estimates.

and we use both stocks and option prices spanning a long time period so that our results are not driven by a specific episode.

To assess the quality (both on the historical and risk neutral sides) of our particle algorithm, the final sample can be plotted, cf. Figure 6.1.

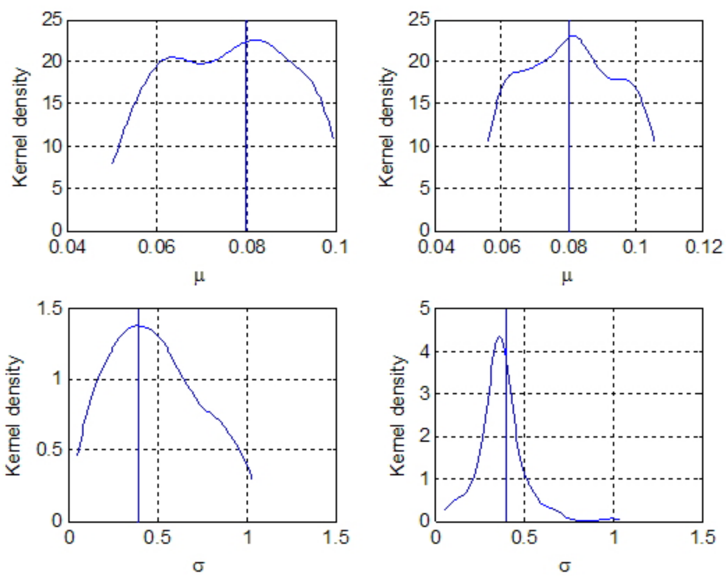

Figure 6.1: Distributions of Simulated Particles.

We also illustrate, in this particular case, the interest of the simulated annealing. In the Figure 6.2, one can observe the convexification/discrimination effects with the evolution of the scales.

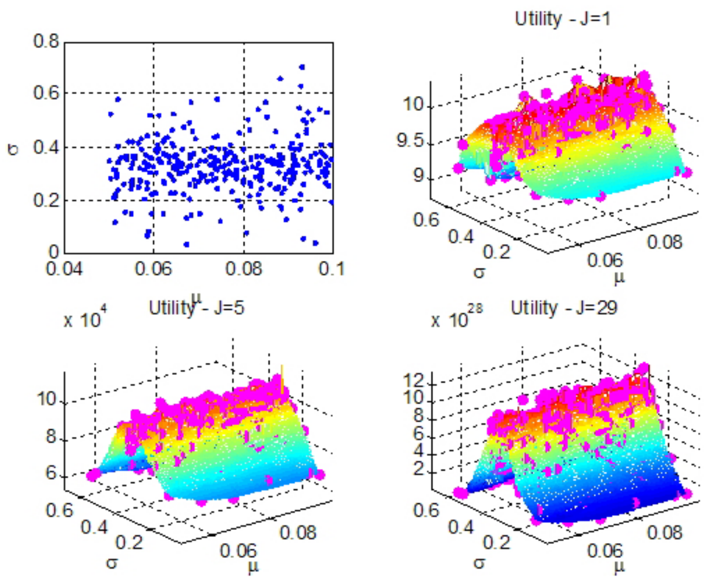

Figure 6.2: Simulated Particles on the Utility Surface.

Using Heston Model

In the second series of simulation, our goal is two-fold. First, we examine how risk premiums affect option prices and then test the efficiency of our algorithm to recover the historical latent stochastic volatility. The analysis has been conducted in the Heston (1993) [88] framework. Stock returns and volatility paths were generated under a specified historical measure $\mathbb{P}$, while option prices were computed under an appropriate measure $\mathbb{Q}$ for 21 equidistant strike values. The system for (S, V) can be written as:

$$dS_t = \mu S_t dt + S_t \sqrt{V_t} dW_t^s - S_t \overline{\mu}_s \lambda dt \tag{6.15}$$

$$dV_t = \kappa_v (\theta_v - V_t) dt + \sigma_v \sqrt{V_t} dW_t^v \tag{6.16}$$

where W_t^k, $k = (s, v)$ are two Brownian motions with $E[dW_t^s dW_t^v] = \rho dt$. As Eraker (2004) [61], we chose the following shape of market prices of risk:

- return market price of risk:

$$\lambda^S(t) = \frac{\mu^{\mathbb{P}} - \mu^{\mathbb{Q}}}{\sqrt{V(t)}}$$

- volatility market price of risk:

$$\lambda^V(t) = -\frac{\kappa^{\mathbb{P}} - \kappa^{\mathbb{Q}}}{\sigma_v \sqrt{1-\rho^2}}$$

We led the same estimation strategy that for Black-Scholes model and produced separate estimates for time horizon ranging from $T = 5$ to $T = 30$. The estimated parameters for both measures are reported in Table 6.2.

	Horizon					
Parameters	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 25$	$T = 30$
$\mu^{\mathbb{P}}$ (0.08)	0.077	0.075	0.075	0.075	0.075	0.076
$\kappa_v^{\mathbb{P}}$ (3.21)	3.317	3.322	3.263	3.330	3.427	3.515
$\theta_v^{\mathbb{P}}$ (0.25)	0.19	0.283	0.208	0.201	0.239	0.192
$\sigma_v^{\mathbb{P},\mathbb{Q}}$ (0.81)	0.706	0.720	0.762	0.821	0.879	0.856
$\rho^{\mathbb{P},\mathbb{Q}}$ (-0.2)	-0.11	-0.32	-0.431	-0.525	-0.224	-0.184
$\kappa_v^{\mathbb{Q}}$ (4.41)	4.249	4.425	4.464	4.49	4.556	4.533
$\theta_v^{\mathbb{Q}}$ (0.35)	0.302	0.26	0.353	0.39	0.327	0.26
λ^S (0.048)	0.045	0.043	0.044	0.043	0.043	0.044
λ^V (-1.2)	-1.383	-1.565	-1.14	-1.212	-1.129	-1.018
LSE	0.0017	0.0012	0.0024	0.0036	0.0037	0.0068

Table 6.2: Heston Model Estimates.

Once again, the results are satisfying and bear some interesting comments. Indeed, without using a huge number of simulations and particles, a good level of accuracy has been reached. Therefore, it is far from being absurd to rely on the results produced by our procedure. Though, we can notice that the joint calibration is more cumbersome for long time horizon data which was predictable because of the more complex and numerous effects to catch. It was not true with Black Scholes: it can be explained by the difficulty entailed by the diffusion of the volatility. But in practice, this procedure will not be of any interest for horizons larger than 10 days. Besides, a kind of stability over time seems to emerge from the results. This is an important point to examine. This could confirm one of the motivations of the resolution of this problem.

6.4. Real-World Risk Premiums Investigation

In order to assess the usefulness of our joint calibration approach, we address two option pricing issues: selecting the appropriate model and quantifying the risk premiums of the various underlying factors. In this attempt, we used the information from the cross-section of EuroStoxx 50 stocks and options series for the period June 2003 to April 2004.

6.4.1 Models Specification

We consider the general framework displayed in Section 6.2 and study models incorporating stochastic volatility and jumps in both returns and volatility. On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the equity price index S_t and its spot variance V_t are assumed to jointly verify:

$$dS_t = \mu S_t dt + S_t \sqrt{V_t} dW_t^s + d\left(\sum_{n=1}^{N_t} S_{\tau_{n-}} \left[e^{Z_n^s} - 1\right]\right) - S_t \overline{\mu}_s \lambda dt \tag{6.17}$$

$$dV_t = \kappa_v \left(\theta_v - V_t\right) dt + \sigma_v \sqrt{V_t} dW_t^v + d\left(\sum_{n=1}^{N_t} Z_n^v\right) \tag{6.18}$$

where $W_t^k,\ k = (s, v)$ are two Brownian motions with $E\left[dW_t^s dW_t^v\right] = \rho dt$ and N_t is a Poisson process with intensity λ, independent of the two diffusion processes. $Z_n^s | Z_n^v \sim N\left(\mu_s + \rho_s Z_n^v, \sigma_s^2\right)$ are the jumps in returns and $Z_n^v \sim \exp\left(\mu_v\right)$ are the jumps in volatility and $\overline{\mu}_s = \exp\left(\mu_s + \frac{1}{2}\left(\sigma_s\right)^2\right)$.

From this very general formulation, augmenting Bates (1996) [14] and close to Eraker (2004) [61], we derive four specifications: SV and SVJ models assume that there are respectively no jumps at all and no jumps in volatility, SVIJ and SVCJ models allow both types of jumps and consider respectively that jumps sizes in returns are independent, correlated (parameter ρ_s) with those in volatility.

The market generated by the structure (6.17) and (6.18) is incomplete. Therefore, there is a multiplicity of equivalent martingale measures, corresponding to the absence of arbitrage. For the highest flexibility of the equivalent martingale measure, we assume a very general change of measures. According to the generalized Girsanov theorem recalled in Section 6.2, the measure transformation for Brownian motions only shifts the drift of the stochastic differential equations, while measure transformation for jump processes are more flexible. As we choose a specification with constant intensity and time-independent jump sizes, we only require that the two distributions are absolutely continuous with respect to the other one. Finally, under the risk-neutral probability measure $\mathbb{Q}$, the equity index and its variance verify:

$$dS_t = r S_t dt + S_t \sqrt{V_t} dW_t^s\left(\mathbb{Q}\right) + d\left(\sum_{n=1}^{N_t(\mathbb{Q})} S_{\tau_{n-}} \left[e^{Z_n^s(\mathbb{Q})} - 1\right]\right) - S_t \overline{\mu}_s^{\mathbb{Q}} \lambda^{\mathbb{Q}} dt \tag{6.19}$$

$$dV_t = \kappa_v^{\mathbb{Q}} \left(\theta_v^{\mathbb{Q}} - V_t\right) dt + \sigma_v \sqrt{V_t} dW_t^v\left(\mathbb{Q}\right) + d\left(\sum_{n=1}^{N_t(\mathbb{Q})} Z_n^v\left(\mathbb{Q}\right)\right) \tag{6.20}$$

where $\overline{\mu}_s^{\mathbb{Q}} = \exp\left(\mu_s^{\mathbb{Q}} + \frac{1}{2}\left(\sigma_s^{\mathbb{Q}}\right)^2\right)$. We define the sets of structural:

$$\gamma^{\mathbb{P}} = (\mu, \kappa_v, \theta_v, \lambda, \mu_s, \rho_s, \sigma_s, \mu_v)$$

and implicit:

$$\gamma^{\mathbb{Q}} = \left(\kappa_v^{\mathbb{Q}}, \theta_v^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \mu_s^{\mathbb{Q}}, \rho_s^{\mathbb{Q}}, \sigma_s^{\mathbb{Q}}, \mu_v^{\mathbb{Q}}\right)$$

parameters. We also note that the change of measures constrains σ_v and ρ to be the same under both measures.

6.4.2 Parameters Estimation using Cross-Section

The methodology has been applied to the cross-section of stock returns and option prices. Nevertheless, the extreme computational burden generated when using both sources of data severely constrained how much and what type of data could be used. Therefore, we focused on three sub-periods of different horizons of our data set and discussed the option pricing implications of our results. These three Eurostoxx 50 scenarios are denoted by S_i, $i = (1, 2, 3)$.

Parameter Estimates

Table 6.3 reports parameter estimates under both measures $\mathbb{P}$ and $\mathbb{Q}$ for the SV and SVJ models and the three scenarios. The first scenario lasts from 02/06/03 to 06/06/03, the second from 09/06/03 to 13/06/03 and the third one is the joint of the two preceding, spanning from 02/06/03 to 13/06/03. Maturity of all the scenarios is 183 days. The estimates are quoted in annualized form in order to be comparable with existing results in option pricing literature, e.g. Bates (1996) [14] and Pan (2002) [137].

There are several interesting features to point out from these estimates. First, we note that for a specified model, reported values are relatively stable across scenarios. Moreover, the joint scenario S_3 produces estimates close to those obtained in sub-samples S_1 and S_2. These results give evidence of the stabilization effect of our procedure in estimating model parameters across time. By incorporating both stocks and options cross-sectional aspects in the calibration, our joint approach may capture trend effects and produces also a sort of benchmark model. Second, the difference between estimates under the two measures (historical minus risk-neutral) is the risk premium associated with Brownian, volatility or jump risk. All of these premiums were estimated to be positive across all models and scenarios. This implies that investors are averse to changes in Brownian, volatility or jump dimensions. This remark is of particular interest from the perspective of option prices: when the market is more (resp. less) volatile or jumpy, the options are more (resp. less) expensive than those implied by the objective measure as investors require higher (resp. lower) premiums.

	SV			SVJ		
	S_1	S_2	S_3	S_1	S_2	S_3
μ	0.072	0.078	0.069	0.069	0.075	0.064
κ_v	7.009	6.874	7.109	7.127	6.938	7.261
θ_v	0.028	0.027	0.023	0.028	0.026	0.022
λ	.	.	.	0.06	0.07	0.06
μ_s (%)	.	.	.	-3.75	-4.01	-4.66
σ_s (%)	.	.	.	4.07	3.17	6.63
$\sigma_v^{\mathbb{P},\mathbb{Q}}$	0.541	0.558	0.472	0.542	0.612	0.529
$\rho^{\mathbb{P},\mathbb{Q}}$	-0.306	-0.381	-0.540	-0.379	-0.431	-0.582
$\mu^{\mathbb{Q}} = r$	0.03	0.03	0.03	0.03	0.03	0.03
$\kappa_v^{\mathbb{Q}}$	4.521	4.756	4.781	4.368	4.679	4.804
$\theta_v^{\mathbb{Q}}$	0.031	0.024	0.022	0.028	0.025	0.026
$\lambda^{\mathbb{Q}}$	.	.	.	0.002	0.002	0.002
$\mu_s^{\mathbb{Q}}$ (%)	.	.	.	-2.59	-2.89	-3.23
$\sigma_s^{\mathbb{Q}}$ (%)	.	.	.	2.89	2.21	4.91

Table 6.3: Joint P and Q Parameters Estimates.

Market Smiles Fit

Having estimated the $\mathbb{P}$ and $\mathbb{Q}$ parameters, we now discuss the empirical performance of the various models in fitting the historical implied volatility smiles. Figure 6.3 plots the model and market Black-Scholes implied volatilities (IVs) and presents evidence on the fit of the SV and SVJ models. We can notice that the IVs curves are very similar from one model to another one and on average they fit quite well the data.

The upper left and right graphs in Figure 6.3 superimposes the five daily IVs curves for respectively scenarios S_1 and S_2. As moneyness is different from one day to another, we represent volatility curves with respect to options instead of moneyness, so that to each X-axis point corresponds an option whose daily implied volatilities are reported along the Y-axis ranging from the first to the last day. Besides, for each day, market curve goes through the big points, while model smile is represented by a solid line. We remark that for each day and scenario, SV or SVJ model smiles fit relatively well market values. The lower graphs present how the presented methodology allows to reconstruct term-structure of implied volatilities across options and time. While the reconstructed volatility surface (right) is more perturbed than the original model surface (left), the error is reasonably small.

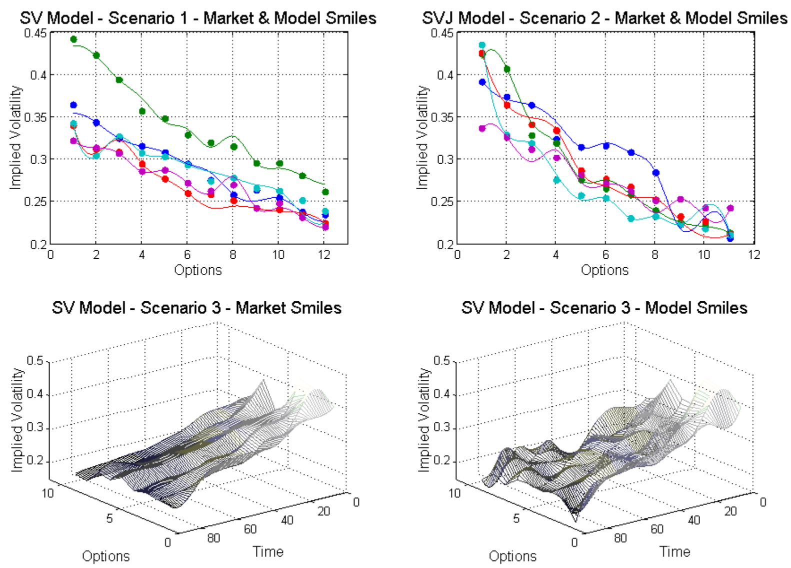

Figure 6.3: Calibrated Black-Scholes Implied Volatility Curves.

6.4.3 Premiums Estimates and Option Prices Effects

Our joint calibration approach attempts to simultaneously capture the effect of both the historical and implied probability measures in a one-stage procedure. Since the methodology requires the absence of arbitrage, σ_v and ρ should be the same under $\mathbb{P}$ and $\mathbb{Q}$. Therefore, the estimation of risk premium parameters is central to our methodology. A thorough analysis of risk premium estimates and effects is therefore of interest.

Inference on Risk Premiums

To see how risk premiums affect conditional moments of returns and volatility, Table 6.4 provides the instantaneous first and second moments of $Y_t = \ln(S_t)$ and V_t for the SV and SVJ models.

For example, risk premiums affect both the level and mean-reversion of volatility, which implies that a positive volatility risk premium generates a structural volatility higher than its risk-neutral counterpart. Besides, jumps in returns and volatility generate different patterns of conditional non-normalities. Jumps in returns result in decreasing amounts of excess skewness and kurtosis while jumps in volatility provide a factor that combines features from both jumps in volatility and diffusive stochastic volatility.

	SV	SVJ
$\mathbb{E}[Y_t]$	$\mu - 0.5\mathbb{E}[V_t]$	$\mu - 0.5\mathbb{E}[V_t] - \lambda(\overline{\mu} - \mu_s)$
$Var[Y_t]$	V_t	$V_t + \lambda(\mu_s^2 + \sigma_s^2)$
$\mathbb{E}[V_t]$	θ_v	$\theta_v + (\kappa_v)^{-1}\lambda\mu_v$
$Var[V_t]$	$\sigma_v^2 V_t$	$\sigma_v^2 V_t$

Table 6.4: Conditional First and Second Moments.

An analysis of conditional moments provides a theoretical description of why risk premiums are difficult to estimate. In fact, variations in these parameters slightly affect the first moment of Y_t and V_t and should have a small effect on the cross-section of returns and options series. A satisfactory identification of these parameters would thus require derivative contracts which solely depend on how quickly the conditional moments of the state variables Y_t and V_t fluctuate.

To assess further the estimation challenge of risk premiums, we illustrate in a smiles variation study why such parameters are difficult to identify. Figure 6.4 shows how variations of risk premiums affect implied volatility.

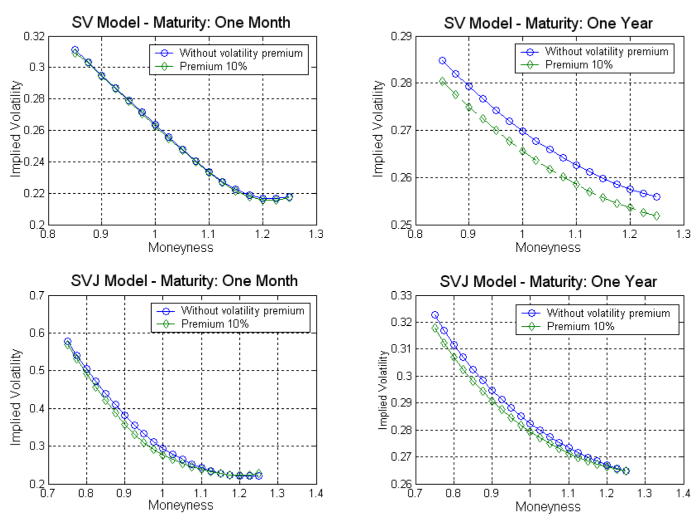

Figure 6.4: Effects of Variations of Volatility and Jump Risk Premiums on IVCs for Two Maturities in the SV and SVJ Models.

From $\mathbb{P}$ to $\mathbb{Q}$ implied volatility

The effects of risk premiums can also be assessed on the basis of option prices. Figure 6.5 displays Black-Scholes implied volatility curves for the SV and SVJ models for two maturities and prices computed under $\mathbb{P}$ and $\mathbb{Q}$.

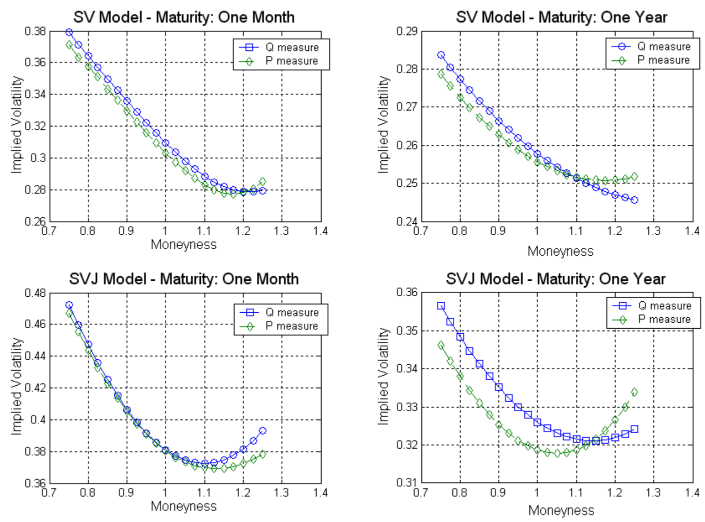

Figure 6.5: IVCs for SV and SVJ Models Based on Parameters from $\mathbb{P}$ and $\mathbb{Q}$.

The first smile is based on $\mathbb{Q}$ parameters while the second one is based on $\mathbb{P}$ parameters which include the effects of risk premiums estimates. One can notice that the two models generated quite similar implied volatility curves. Besides, spreads between the two measures are more severe as maturity increases and assess for uncertainty and risk averse behaviour of investors. Therefore, given a sufficient number of parameters and allowing them to change from one measure to another without constrains, one cannot distinguish different models in an analysis only based on option prices. And quite naturally, a simpler model will always be preferred. From a classical calibration perspective where ease and reliability are of first importance, this is not a problem. However, if one wishes to jointly fit both cross-section of returns and options series and to determine a relevant benchmark model, another strategy has to be set up. As a consequence, our joint calibration procedure might help to select the best model(s) to take into account all aspects of the data set.

6.4.4 Aversion to Model Misspecification

Given our joint calibration procedure, we next investigate the time homogeneous property of the models specification and the behaviour of investors towards Brownian, volatility and jump risk, described by risk premiums, across time. To answer this question in a satisfactory way, we set up a procedure which consists of estimating risk premiums parameters for SV model for each month of our data set. The Table 6.5 reports relevant parameter estimates.

	SV				Options		Returns	
	$\kappa_v^{\mathbb{P}}$	$\theta_v^{\mathbb{P}}$	$\kappa_v^{\mathbb{Q}}$	$\theta_v^{\mathbb{Q}}$	Skew	Kurt	Skew	Kurt
jun 03	7.35	0.028	3.91	0.033	1.15	3.90	0.06	2.27
jul 03	7.99	0.028	4.69	0.030	1.42	4.88	−0.36	3.06
aug 03	7.89	0.023	3.94	0.028	1.18	3.94	−0.34	1.69
sep 03	7.46	0.033	3.61	0.051	1.18	4.16	−0.92	3.38
oct 03	8.09	0.024	3.46	0.035	1.11	4.09	−0.52	2.09
nov 03	6.97	0.031	4.16	0.032	1.45	5.77	−0.73	2.53
dec 03	8.76	0.024	5.63	0.024	1.60	5.81	0.63	2.37
jan 04	8.23	0.030	5.76	0.033	1.47	5.79	−0.01	1.57
feb 04	8.76	0.025	6.69	0.024	1.51	5.69	−0.16	1.97
mar 04	7.74	0.030	5.59	0.029	1.15	4.13	0.28	1.59
apr 04	7.49	0.027	4.86	0.025	0.92	3.13	−0.59	3.25

Table 6.5: Volatility Risk Premiums Estimates Across Time.

An important conclusion emerges from this result. The divergence between the information embedded in the measures $\mathbb{P}$ and $\mathbb{Q}$ is indicative of time-varying risk premiums. More interestingly, returns and options series displayed opposite effect in terms of moments properties as measured by the skewness and kurtosis: while returns displayed negative skewness and slight kurtosis, options are characterized by strong and positive values of these moments. This effect could be related to investor's anticipations towards model misspecification which lead to overprice Brownian, volatility and jump risk by putting high risk premiums. In fact, if the models are misspecified along certain dimensions and agents are both risk and uncertainty-averse to this model failure, then the historical and implied distributions present a form of time-varying aversion. Here again, in a classical calibration perspective, this aspect is not crucial as practitioners preclude model indeterminacy by frequently rebalancing model parameters. On the contrary, our joint and cross-sections calibration procedure may benefit from studying models that allow time-varying parameters in their specification and under this setup might help to investigate time inhomogeneity of the historical and implied distributions.

6.5. Conclusion

We developed an innovative and practicable methodology for a joint calibration challenge: from a time series of stock returns and option prices, we managed to determine the characteristic parameters of both historical and risk-neutral measures. This was made possible by a sophisticated particle sampler where usual methods would have failed. Our theoretical study of it led us to efficiently tune the algorithmic preferences.

The interest of this approach is two-fold: first, it stands for a quest for information embedded in real data, complementary to the one led through classical calibration (embedded in the stochastic model). It operates a kind of average of this calibration and might be used as a benchmark. Moreover, it takes advantage of more information which could be notably used to calibrate parameters usually determined by rough calculations or intuition or a simple historical study. For instance, the idea briefly alluded to in the introduction remains a good one: the determination of unobservable parameters like correlations in the multi-dimensional case is often achieved through historical statistics independently from other parameters. With this algorithm, it could be done in a more sophisticated way and since the at-stake of these correlation products is tremendous, it is not a pointless remark.

The second interest is directly linked to the theoretical and more general information brought by the joint calibration. In particular, the last study on risk premiums is quite precious, and must be deepened. In fact, our work sheds light on the relationship between the investors and the model used: leading this study through different periods of time could be very enriching regarding our understanding of the models, their limitations and the trust investors have in them.

Then, this work gives way to a lot of interesting tracks to be explored. The first one is to optimize the conception of the algorithm. It would be very helpful to try a more dynamical optimization with less rough approximations as suggested in Section E.5. It could allow us to establish a clear and refined dependency between the cooling schedule, the shape of the utility function and the size of the compact considered. Besides we could exploit in a better way the impact of entropy with a dynamic determination of the constant imposing the utility to be positive. It would prevent it from exploding in the calculation of the entropy and so the entropy could bring its convexification effect as suggested by our first attempts. After these improvements the convergence could be faster and the algorithm could give quicker accurate results.

E

Convergence Results

Agenda

In this Appendix, we present convergence results for particle systems. Especially, we provide a uniform convergence theorem for our Markov Chain Monte-Carlo with Interacting Particle System algorithm, cf. Section 5.3.

E.1. Existing Convergence Theorems

Let π_n be the probability measure over the Borel set $\mathcal{D}$ with Lebesgue density U^J, up to a normalization constant. In the following, we will extensively resort on the next notation for some sets of test functions:

- $C_b(\mathcal{D})$ is the set of continuous bounded functions on $\mathcal{D}$,
- $C_{b,1}(\mathcal{D})$ is the subset of $C_b(\mathcal{D})$ of the functions ϕ with $\|\phi\| \leq 1$ where:

$$\|\phi\| = \sup_{x \in \mathcal{D}} |\phi(x)|$$

- $C^{0,\infty}_{b,1}(\mathcal{D})$ is the subset of $C_b(\mathcal{D})$, continuous and infinitely differentiable.

We first recall a simple convergence result shown in Amzal et al. (2005) [5], for the two previously presented algorithms, cf. Section 5.3. To this end, for any fixed iteration n, we introduce a set of assumptions.

Assumption E.1.1 *Assume that:*

- u is continuous, positive and bounded,
- for all d and n, $K_{1,n}(d,\,.\,) > 0$ on the support of h_J,
- $(1/N)\sum_{i=1}^{N} var_{K_{1,n}(d_i^{(n-1)},\,.\,)\otimes^n p}(w_i^{(n)})$ is bounded independently of N,
- there exists $\delta > 0$ s.t. conditionally to formerly drawn particles $(d_i^{(n-1)})$:
$$\left(1/\sum_{i=1}^{N} var_{K_{1,n}\otimes^n p}(w_i^{(n)})\right)^{2+\delta} \sum_{i=1}^{N} \mathbb{E}_{K_{1,n}\otimes^n p}\left[(w_i^{(n)})^{2+\delta}\right] \longrightarrow 0, N \to +\infty$$
- $d \mapsto var\,[K_{2,n}(d,\,.\,)]$ is bounded.

Assumption E.1.1 corresponds to regularity hypothesis on Markov kernels that are easy to meet in practice. Under these conditions, one can prove the following step-by-step theorem:

Theorem E.1.2 (Step-by-Step Convergence) *Under Assumption E.1.1, for any iteration n, there exists a constant a_n s.t. for any measurable and bounded function ϕ over $\mathcal{D}$, we have:*

$$E\left[\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{d_i^{(n)}}(\phi) - \pi_n(\phi)\right)^2 |\mathcal{F}_{n-1}\right] \le \frac{a_n||\phi||_\infty^2}{N} \tag{E.1}$$

where $\mathcal{F}_{n-1}$ stands for the conditioning on the previous sample $(d_i^{(n-1)})_{i=1..N}$.

A detailed proof of this result can be found in Amzal et al. (2005) [5], using Lindeberg's types of arguments. It states that, for each iteration n, the simulated sample generated by our algorithm will get closer to the searched optimum for a big enough sample size N. The main limit of this result is that we have no clue about how a_n might grow with n. We then aim to establish an explicit dependency of a_n on n. Once this obtained, it may be used to know the required number of particles to be used to achieve a certain level of accuracy of the optimization algorithm. Indeed, this level imposes a minimal value of $J(n)$ through the convergence of the target law towards the Dirac measure on the mode of the utility function. Then we only have to find the optimal cooling schedule to deduce a minimal value of n.

In the particular case of the Resampling-Markov (RM) Algorithm 5.3.2, a stronger result can be shown, as presented in the following theorem.

Theorem E.1.3 (Convergence of the RM Algorithm) *Assuming that u is bounded by strictly positive constants, for any iteration n, there exists a constant c_n such that for any ϕ in $C_b(\mathcal{D})$, we have:*

$$\mathbb{E}\left[\left(\pi_n^N(\phi) - \pi_n(\phi)\right)^2\right] \leq c_n \frac{||\phi||_\infty^2}{N}$$

Let B be a neighborhood of the targeted optimum. Let $f_n^N(B)$ be the frequency of visiting B for the RM Algorithm, and $f_n(B)$ be the frequency of visiting B for a theoretical trajectory from $\pi_0 \otimes ... \otimes \pi_n$. If the cooling schedule is linear:

$$E\left[\left(f_n^N(B) - f_n(B)\right)^4\right]^{1/4} \leq \frac{c}{\sqrt{N}}$$

A simple proof of it is also available in Amzal et al. (2005) [5]. Recalling that our goal is the optimum determination, this convergence of the visiting frequencies around optimum's neighborhoods brings a valuable evidence for the effectiveness of the RM Algorithm with a linear cooling schedule. We are now concerned by extending those convergence results.

E.2. The Feynman-Kac Formalism

Usual central limit type of statistical theorems, cf. Jacod-Shiryaev (2003) [93], are not refined enough for our purpose. That's why we should analyze our particle systems algorithm in terms of measure processes, as pioneered by Del Moral (1998) [39]. Our demonstration will be strongly inspired by Del Moral-Guionnet (2001) [40] in which an uniform convergence result is brought up for sequential particle filtering. Let us introduce some useful notations related to measure theory.

E.2.1 Notation

On the space of signed measures on the Borel set $\mathcal{D}$, we define, for a signed measure μ, its total variation norm by:

$$\|\mu\|_{\mathrm{TV}} = \sup\{\mu(f), f \in \mathcal{C}_{b,1}^{0,\infty}(\mathcal{D})\}$$

We also introduce $M_1(\mathcal{D})$ as the space of all probability measures on $\mathcal{D}$, equipped with the weak topology. Then, if K is a Markov transition on $\mathcal{D}$ with set of Borelians given by $B(\mathcal{D})$, we define the Dobrushin ergodic coefficient, cf. Dobrushin (1970) [52], by:

$$\mathrm{Dob}(K) = 1 - \sup_{x,z \in \mathcal{D}} |K(x, A) - K(z, A)|\,, A \in B(\mathcal{D})$$

At this point, we can notice that if K is independent from the starting point, its Dobrushin coefficient will be equal to 1. To interpret our algorithm in terms of discrete-time measure-valued stochastic process, we consider the following measure-valued dynamical system:

$$\pi_n = \phi_n(\pi_{n-1})$$

where $\pi_0 \in M_1(\mathcal{D})$ and $\phi_n : M_1(\mathcal{D}) \to M_1(\mathcal{D})$ is a continuous function s.t.:

$$\phi_n(\pi) = \psi_n(K^{\otimes}_{1,n}\pi)K^{\otimes}_{2,n} \tag{E.2}$$

$$\psi_n(\pi)f = \frac{(\mathbf{1}\otimes\mathbf{1})(\pi)(g_n f)}{(\mathbf{1}\otimes\mathbf{1})(\pi)(g_n)} \tag{E.3}$$

where:

- $1 \otimes 1$ is the identity tensor $M_1(\mathcal{D}) \mapsto M_1(\mathcal{D}\times\mathcal{D})$ s.t. $(1\otimes 1)(\pi) = \pi\otimes\pi$,
- $\{K^{\otimes}_{1,n} = \delta\otimes K_{1,n}, n \geq 1\}$ and $\{K^{\otimes}_{2,n} = \delta\otimes K_{2,n}, n \geq 1\}$ are two sequences of Markov transition kernels on $\mathcal{D}\times\mathcal{D}$, where δ is the identity kernel on $\mathcal{D}$ and $K_{1,n}$ et $K_{2,n}$ are sequences of transition kernels on $\mathcal{D}$ corresponding respectively to the importance sampling steps and to the Markov renewal steps. We shall underline that, for time, the Markov kernel is not anymore of Metropolis-Hastings type, we will re-discuss this issue in Section E.5. We will slightly abuse of the notations by identifying $K_{i,n}$ and $K^{\otimes}_{i,n}$ as well as any test function f (on $\mathcal{D}$) and $1\otimes f$ (on $\mathcal{D}\times\mathcal{D}$),
- $\{g_n, n \geq 1\}$ is a sequence of bounded positive random functions on $\mathcal{D}\times\mathcal{D}$. In our case, they stand for the non-normalized weights and can be written as a deterministic function g_n:

$$g_n(d,d') = \frac{\prod_{j=1}^{J(n)} u(d', F^{-1}(d', V_j))}{K_{1,n}(d,d')}$$

where $F^{-1}(d',\,.\,)$ is the vector of inverse cumulative distribution function of $p(.\,|\,d')$ taken on as an independent uniform random vector V_j.

By construction, the equation $\tilde{\pi}_n = \pi_n \otimes U^{\otimes J(n)}_{[0,1]}$ is solution of the transport equation $\mu_n = \phi_n(\mu_{n-1})$. This is also true for the empirical measure $\pi^N_n = \frac{1}{N}\sum_{i=1}^N \delta_{d^n_i}$. Indeed, recalling that $\widehat{d^n_i}$ is the state of the $i-$th particle after n iterations of the importance sampling step, and d^n_i the state of the $i-$th particle after n iterations, we can write:

$$\psi_n\Big(\frac{1}{N}\sum_{q=1}^N \delta_{\widehat{d^n_q}}/d^n_1, ..., d^n_N\Big) = \sum_{i=1}^N \frac{g_n(\widehat{d^n_i}, d^n_i)K_{2,n}(\widehat{d^n_i}, .)}{\sum_{j=1}^{j=N} g_n(\widehat{d^n_i}, d^n_i)}$$

Similarly, we note by $\tilde{\pi}^N_n$, the joint empirical measure over the augmented probability space, namely $\frac{1}{N}\sum_{i=1}^N \delta_{d^n_i, V_{i1}, .., V_{iJ(n)}}$.

E.2.2 Relations

Before giving the main theorem, let us introduce the fundamental relations that will help us to both formulate and prove it. We have:

$$\phi_{n/p} = \phi_n \circ ... \circ \phi_{p+1}, 0 \leq p \leq n.$$

with convention $\phi_{n/n}$ =Id. A simple but careful induction gives the relation:

$$\phi_{n/p}(\pi)f = \frac{(\mathbf{1} \otimes \mathbf{1})(\pi)\left(K_{1,p}g_{n/p}\left(K_{2,n/p}f\right)\right)}{(\mathbf{1} \otimes \mathbf{1})(\pi)(K_{1,p}g_{n/p})}, \ \forall f \in C_b(\mathcal{D})$$

where:

$$K_{2,n/p-1}f = \frac{K_{2,p}K_{1,p}\left(g_{n/p}\left(K_{2,n/p}f\right)\right)}{K_{2,p}K_{1,p}(g_{n/p})}, \ g_{n/p-1} = g_p K_{2,p}K_{1,p}(g_{n/p}) \quad \text{(E.4)}$$

with the conventions $g_{n/n} = 1$ and $K_{2,n/n}$ =Id.

Another useful notation will be to call $S_{n/p}$ the operator:

$$S_{n/p}f = \frac{K_{2,p}K_{1,p}(g_{n/p}f)}{K_{2,p}K_{1,p}(g_{n/p})}, \ 0 \leq p \leq n \quad \text{(E.5)}$$

so that one can easily check the following equality

$$K_{2,n/p-1} = S_{n/p}S_{n/p+1}...S_{n/n}$$

Finally, we denote:

$$K_{2,p,n} = K_{2,p}K_{2,p+1}...K_{2,p+n}$$

E.3. A General Convergence Result

Assumption E.1.1 are clearly checked by the problem we propose to solve. Indeed, a sufficient reason is that we work on truncated functions, cf. Section E.5 for practical developments on this. So the step-by-step convergence Theorem E.1.2 holds, but we are aiming to prove a stronger result.

The convergence theorem we sought must prove the convergence of the empirical laws towards the Dirac measure in the mode of our utility function and must establish a dependency between the number of simulations needed and the number of particles used to achieve a certain level of precision.

We will denote by D the largest diameter over all the coordinates of $\mathcal{D}$ and $K_{3,k} = K_{2,k}K_{1,k}$. Also, let σ be such that:

$$var\left(K_{3,k}\right) \geq \frac{\sigma^2}{J(k)} \tag{E.6}$$

In the following demonstration, we consider a Markov step which is continuous relatively to Lebesgue measure. The constants displayed are computed for Gaussian random walks, without any loss of generality. For further details on practical aspects, the reader can refer to Section E.5.

Our first useful, in its own, result is the control on the Dobrushin coefficient of the operators $S_{p+T,k}$ given by (E.5).

Lemma E.3.1 *For every k, p, T verifying $p+1 \leq k \leq p+T$, we get:*

$$Dob(S_{p+T/k}) \geq \epsilon_k^2, \ a.s.$$

with:

$$\epsilon_k = \exp\left(-\frac{J(k)}{2}\left(\frac{mD^2}{\sigma^2} + ln\left(\frac{U_{\max}}{U_{\min}}\right)\right)\right)$$

Proof. Computations with rough inequalities give, that, for all borelians $A \subset \mathcal{D}$ and for all $d \in \mathcal{D}$:

$$\epsilon_k \pi_k(A) \leq K_{3,k}(d, A) \leq \frac{1}{\epsilon_k}\text{Leb}(A)$$

with Leb the Lebesgue measure. Then, as

$$S_{p+T,k}f = \frac{K_{3,k}(g_{p+T/k}f)}{K_{3,k}(g_{p+T/k})}$$

with $f \in C_{b,1}(\mathcal{D})$, we easily get that

$$S_{p+T,k}f \geq \epsilon_k^2 \frac{\text{Leb}(g_{p+T/k}f)}{\text{Leb}(g_{p+T/k})}$$

which gives the expected result. ■

Now we can state the main result of this section.

Theorem E.3.2 (General Convergence Result) *For any $f \in C_{b,1}(E)$, $n \in N$, $T \leq n$, we have:*

$$\mathbb{E}\left(\left|\pi_n^N f - \pi_n f\right|\right) \leq \frac{2T}{\sqrt{N}} \prod_{k=n-T+1}^{n} \frac{1}{\varepsilon_k^2} + \prod_{k=n-T+1}^{n} (1 - \varepsilon_k^2)$$

When:

$$J(k) = c\ln(k)\left(\frac{mD^2}{\sigma^2} + ln\left(\frac{U_{\max}}{U_{\min}}\right)\right)^{-1} \tag{E.7}$$

with $c \leq 1$ and $T \propto n^b$ with $c < b < 1$, this inequality gives us the dependency sought between n and N for a given required level of accuracy.

Proof. With the convenient notation introduced in Section E.2, the demonstration relies on the following, telescopic sum, decomposition:

$$\begin{aligned}\tilde{\pi}_n^N f - \tilde{\pi}_n f = &\sum_{p=n-T+1}^{n} \left(\phi_{n/p}(\tilde{\pi}_p^N)f - \phi_{n/p}(\phi_p(\tilde{\pi}_{p-1}^N))f\right) \\ &+ \left(\phi_{n/n-T}(\tilde{\pi}_{n-T}^N)f - \phi_{n/n-T}(\tilde{\pi}_{n-T})f\right)\end{aligned} \tag{E.8}$$

By the triangular inequality, we get:

$$\begin{aligned}\left|\tilde{\pi}_n^N f - \tilde{\pi}_n f\right| \leq &\sum_{p=n-T+1}^{n} \left|\phi_{n/p}(\tilde{\pi}_p^N)f - \phi_{n/p}(\phi_p(\tilde{\pi}_{p-1}^N))f\right| \\ &+ \left|\phi_{n/n-T}(\tilde{\pi}_{n-T}^N)f - \phi_{n/n-T}(\tilde{\pi}_{n-T})f\right|\end{aligned} \tag{E.9}$$

Let us notice here that we can control the g_n and the $g_{n/p}$ corresponding to the weights involved in the selection step:

$$C^{n-p}\,\beta^{\sum_{k=p}^{n} J(k)} \leq g_{n/p} \leq C^{n-p}\,\alpha^{\sum_{k=p}^{n} J(k)},\ a.s.$$

with:

$$\alpha = U_{\max} e^{\frac{mD^2}{2\sigma^2}},\ \beta = U_{\min} e^{\frac{-mD^2}{2\sigma^2}},\ C = Vol(\mathcal{D})$$

Using (E.4), we can bound $a.s.$ each term of the sum by:

$$\frac{b_{n/p}}{a_{n/p}}\left(\left|\tilde{\pi}_p^N f_1 - \phi_p(\tilde{\pi}_{p-1}^N)f_1\right| + \left|\tilde{\pi}_p^N f_2 - \phi_p(\tilde{\pi}_{p-1}^N)f_2\right|\right)$$

with:

$$f_1 = K_{1,p} g_{n/p} K_{2,n/p} f,\ f_2 = K_{1,p} g_{n/p}$$

so that $f_1, f_2 \in C_{b,1}(\mathcal{D})$. This can be classically proved by introducing a third appropriate term, $\phi_p(\tilde{\pi}_{p-1}^N)(K_{1,p}g_{n/p}K_{2,n/p}f)\tilde{\pi}_p^N(K_{1,p}g_{n/p})^{-1}$ and using the triangular inequality. Since $\tilde{\pi}_p^N$ is the empirical measure associated to N independent random variables, with common law $\phi_p(\tilde{\pi}_{p-1}^N)$, the central limit theorem gives us that:

$$\mathbb{E}\left(\left|\tilde{\pi}_p^N f_1 - \phi_p(\tilde{\pi}_{p-1}^N)f_1\right|\right) \leq \frac{1}{\sqrt{N}}\mathbb{E}\left(\left|\tilde{\pi}_p^N f_2 - \phi_p(\tilde{\pi}_{p-1}^N)f_2\right|\right) \leq \frac{1}{\sqrt{N}}$$

Summing these inequalities in (E.9), and using Jensen inequality, we get:

$$\mathbb{E}\left(\left|\pi_n^N f - \pi_n f\right|\right) \leq \frac{2T}{\sqrt{N}} \left(\frac{\alpha}{\beta}\right)^{\sum_{k=n-T+1}^{n} J(k)}$$

which is exactly equals to:

$$\mathbb{E}\left(\left|\pi_n^N f - \pi_n f\right|\right) \leq \frac{2T}{\sqrt{N}} \prod_{k=n-T+1}^{n} \frac{1}{\varepsilon_k^2} \tag{E.10}$$

The second term is bounded. Indeed Lemma E.3.1 entails, *a.s.*, that:

$$\begin{aligned}
&\left\|\phi_{n/n-T}(\mu) - \phi_{n/n-T}(\nu)\right\|_{\mathrm{TV}} \\
&= \left\|\mu_{n/n-T} K_{2,n/n-T} - \nu_{n/n-T} K_{2,n/n-T}\right\|_{\mathrm{TV}} \\
&\leq \prod_{k=n-T+1}^{n} \left(1 - Dob(S_{n/k})\right) \left\|\mu_{n/n-T} - \nu_{n/n-T}\right\|_{\mathrm{TV}} \\
&\leq \prod_{k=n-T+1}^{n} (1 - \epsilon_k^2)
\end{aligned}$$

This gives us the announced result which was the first part of the theorem. Then, if $J(k)$ is given by (E.7) with $c \leq 1$, we get $\varepsilon_k^2 = \frac{1}{k^c}$, and (E.10) becomes:

$$\mathbb{E}\left(\left|\pi_n^N f - \pi_n f\right|\right) \leq \frac{2T}{\sqrt{N}} (n!)^c + e^{\frac{-T}{n^c}} \tag{E.11}$$

If $T = E(n^b)$ with b as above, we only have to take n large enough to have the second term in (E.11) small enough. Then, the first term in (E.10) gives us the adjustment needed over N. This ends the proof of the theorem. ∎

Theorem E.3.2 proves the effectiveness of our optimization algorithm. According to the proof of this theorem, both the numbers of particles and iterations needed to achieve a good level of accuracy are relatively huge, a fact that already appeared in the field of particle filter convergence, cf Doucet et al. (2001) [54]. Martingale arguments can be used to improve the derived constants, cf. Del Moral and Jacod (2002) [42]. In Section E.5, we explain how in practice these constants are not so big.

E.4. A Uniform Convergence Theorem

Now, we want to analyze the convergence issue in the particular case of the Resampling-Markov Algorithm 5.3.2. We can notice that if we choose a linear cooling schedule for J, then g_n are uniformly bounded, which means that one can find a constant $0 < a < 1$ such that, for any n:

$$a \leq g_n \leq 1/a, \ a.s.$$

Moreover, it has appeared that regularity conditions on the Markov kernel are needed to control the convergence. We will therefore assume that the Markov step K_n verifies the following *mixing property*. For each n, there exists $0 < \epsilon < 1$ and a probability measure μ_n s.t., for all measurable A:

$$\mu_n(A)\epsilon \leq K_n(x, A) \leq 1/\epsilon \mu_n(A) \tag{E.12}$$

This, in turn, leads to the following uniform convergence theorem.

Theorem E.4.1 (Uniform Convergence Result for RM Algorithm) *Let the cooling schedule be linear and K_n verifying the mixing property (E.12) with $\epsilon = 1 - e^{-\theta}$. Then, for any $f \in C_{b,1}(\mathcal{D})$, we have:*

$$\sup_{n \in \mathbb{N}} \mathbb{E}\left(\left|\pi_n^N f - \pi_n f\right|\right) \leq \frac{C}{N^\alpha}$$

with:

$$C = \frac{5}{a^2}, \ \alpha \geq \frac{\theta}{2\theta - 2\log a} > 0$$

Proof. The proof of this result is an obvious particular case of the proof of Theorem E.3.2, with $\epsilon_k = \epsilon$ in Lemma E.3.1. The same calculations lead to:

$$\sup_{n \in \mathbb{N}} E\left[\left|\pi_n^N f - \pi_n f\right|\right] \leq \frac{4T}{\sqrt{N}} \frac{1}{a^T} + (1 - \epsilon)^T$$

Choosing the linking function:

$$T(N) = 1 + \left[\frac{1}{4} \frac{-\log N}{\log a(1 - \epsilon)}\right]$$

where [.] stands for the integer part, it comes:

$$\sup_{n \in \mathbb{N}} \mathbb{E}\left[\left|\pi_n^N f - \pi_n f\right|\right] \leq \frac{C}{N^\alpha}$$

which concludes the proof. ■

E.5. A Practical Point of View

In this section, we want to precise the choices of jump functions, according to convergence aspects presented in Sections E.3 and E.4. Also, we discuss the hypothesis used to prove our main convergence Theorem E.3.2.

E.5.1 On the Importance Sampling Proposal

As formerly said, a major concern is the choice for the range of the variance of the transition functions over $\mathcal{D}$. Let us first analyze how one can propose a smart choice for the variance of the importance sampling proposal $K_{1,n}$. For this purpose, we should recall the normalized weights at iteration n:

$$w_i^{(n)} = \frac{\prod_{j=1}^{J} u(\widetilde{d}_i^{(n)}, \widetilde{y}_{i,j}^{(n)})/K_{1,n}(d_i^{(n-1)}, \widetilde{d}_i^{(n)})}{\sum_{i=1}^{N}[\prod_{j=1}^{J} u(\widetilde{d}_i^{(n)}, \widetilde{y}_{i,j}^{(n)})/K_{1,n}(d_i^{(n-1)}, \widetilde{d}_i^{(n)})]}$$

Considering that N and J are big enough to allow Large Number Law's type of approximations, it comes:

$$\log(w_i^{(n)}) = J\left(\frac{1}{J}\sum_{j=1}^{J}\log(u(\widetilde{d}_i^{(n)}, \widetilde{y}_{i,j}^{(n)}))\right) - \log(K_{1,n}(d_i^{(n-1)}, \widetilde{d}_i^{(n)}))$$
$$- \log(\frac{1}{N}\sum_{i=1}^{N}[\prod_{j=1}^{J} u(\widetilde{d}_i^{(n)}, \widetilde{y}_{i,j}^{(n)})/K_{1,n}(d_i^{(n-1)}, \widetilde{d}_i^{(n)})]) - \log(N)$$

Then, the following approximation for the weights holds:

$$w_i^{(n)} = \frac{\exp\left(J\, I(\widetilde{d}_i^{(n)})\right) U^J(\widetilde{d}_i^{(n)})}{NK_{1,n}(d_i^{(n-1)}, \widetilde{d}_i^{(n)}) \int U^J(d)d\gamma}\left(1 + \mathcal{O}\left(\frac{1}{\sqrt{J}}\right)\right) \tag{E.13}$$

where $I(d) = E_{y|d}[\log(u(d,y)/U(d))]$.

From (E.13), we can first notice that:

$$\mathbb{E}_{K_{1,n}\left(d_i^{n-1},\cdot\right)}\left[\omega_i^{(n)}\right] = \frac{1}{N}\int e^{JI(d)}d\pi_n\left(d\right)\left(1 + \mathcal{O}\left(\frac{1}{\sqrt{J}}\right)\right)$$

So, for our order of approximation, we get that:

$$\int e^{JI(d)}d\pi_n(d) = 1 + O\left(\frac{1}{\sqrt{J}}\right) \tag{E.14}$$

Moreover, assuming that $K_{1,n}$ is chosen as a Gaussian random walk with variance $\sigma_{1,n}^2$, and using (E.13), it comes:

$$w_i^{(n)} = \left(\frac{\sqrt{2\pi}\sigma_{1,n}}{N} \exp\left(J\, I(\widetilde{d}_i^{(n)}) + \frac{1}{2}X^2 \right) \frac{U^J(\widetilde{d}_i^{(n)})}{\int U^J(d) d\gamma} \right) \left(1 + \mathcal{O}\left(\frac{1}{\sqrt{J}} \right) \right)$$

where $X \rightsquigarrow N(0,1)$. Considering we want to achieve $\mathbb{E}_{K_{1,n}(d_i^{(n-1)},.)}\left[w_i^{(n)} \right] = 1/N$ and that $\exp(\sqrt(var(X))/2) = \left(1 + \mathcal{O}\left(\frac{1}{\sqrt{J}}\right)\right)$, we can obtain an equation that $\sigma_{1,n}$ should verified, say:

$$\sigma_{1,n} = \frac{1 + \mathcal{O}\left(\frac{1}{\sqrt{J}}\right)}{e^{1/2}\sqrt{2\pi}\mathbb{E}_{K_{1,n}(d_i^{(n-1)},.)}\left[e^{JI(d)}\, U^J\,(d) / \int U^J\,(d')\, d\gamma' \right]} \tag{E.15}$$

We want now to get an approximation of $U^J(\widetilde{d}_i^{(n)}) / \int U^J(d) d\gamma$ in (E.15). Assuming that the points $\widetilde{d}_i^{(n)}$ are relatively close to the optimum, which arrives when J is *large* enough as supposed, we are allowed to use a Taylor development of U around its mode:

$$U^J(\widetilde{d}_i^{(n)}) = \left(U^* - \frac{1}{2} H^* (\widetilde{d}_i^{(n)} - d^*)^2 + o(\sigma_{1,n}^2) \right)^J \tag{E.16}$$

And using the fact that:

$$\left(1 - \frac{H^*}{2U^*} (\widetilde{d}_i^{(n)} - d^*)^2 \right)^J = \exp\left(-\frac{JH^*}{2U^*} (\widetilde{d}_i^{(n)} - d^*)^2 \right) (1 + o(\sigma_{1,n}^2))$$

it comes:

$$\int U^J(d) d\gamma = \sqrt{\frac{2\pi\, U^*}{JH^*}} (1 + o(\sigma_{1,n}^2)) \tag{E.17}$$

In addition, as we want to reach d^* from $\widetilde{d}_i^{(n)}$ by $K_{1,n}$, we need $(\widetilde{d}_i^{(n)} - d^*)^2 = \sigma_{1,n}^2 + o(\sigma_{1,n}^2)$, so, using (E.16) and (E.17), we can write:

$$\frac{U^J(\widetilde{d}_i^{(n)})}{\int U^J(d) d\gamma} = \sqrt{\frac{JH^*}{2\pi U^*}} \exp\left(-\frac{JH^*}{2U^*} \sigma_{1,n}^2 \right) (1 + o(\sigma_{1,n}^2))$$

This last approximation represents the fact that, for a sufficiently large J, the density of π_n can be locally approximated around the mode d^* by a Gaussian distribution with mean d^* and with variance U^*/JH^*. Then, injecting this approximation in (E.15), we can easily check that:

$$\sigma_{1,n}^2 = \frac{U^*}{JH^*} \tag{E.18}$$

is a solution of (E.15), noticing that, with this choice of $\sigma_{1,n}$ and using (E.14):

$$\mathbb{E}_{K_{1,n}\left(d_i^{n-1},\cdot\right)}\left[\exp\left(JI\left(d\right)\right)\right] = \int e^{JI(d)} d\pi_n\left(d\right) + \mathcal{O}\left(\frac{1}{\sqrt{J}}\right) = 1 + \mathcal{O}\left(\frac{1}{\sqrt{J}}\right)$$

Following the same line of computations, it is easy to show that we can generalize this result for $dim(\mathcal{D}) > 1$ choosing:

$$\sigma_{1,n}^2 = \det\left(\frac{U^*}{JH^*}\right) \tag{E.19}$$

where H^* is now the Hessian matrix of U at the optimum.

E.5.2 On the Markov Renewal Step

As we saw in the convergence analysis, a key property of the Markov kernels for fast convergence is the *mixing property* (E.12), which can be typically met when $K_{2,n}$ has a continuous Lebesgue derivative, as supposed in the demonstration. However, as we introduced it, a convenient choice to build such Markov steps which converge adequately to the true target distribution is the Metropolis-Hastings kernel, which is definitely not Lebesgue-derivable. A practical and heuristic conclusion from this may be: the higher is the acceptance probability in the Metropolis-Hastings step, the faster is the convergence of the algorithm. As a consequence, using an adaptive Markov step, in which this acceptance rate is made to be larger, see Haario et al. (2003) [81], could significantly improve the convergence. More case-specific Markov steps could be used, such as Gibbs kernels which are then Lebesgue-derivable.

As for the variance of this Markov proposal, we choose it lower bounded. Combined with the choice of the variance of the importance sampling proposal, cf (E.19), this would roughly ensure the variance assumption, cf. (E.6). It also allows us a larger scale of exploration of $\mathcal{D}$, as desired. Note that this condition on the variance proposal does not imply any bounding condition on the resulting Metropolis-Hastings kernel.

Empirical Remarks on the Convergence

We end this practical point of view by considerations on the effective convergence speed in practice. Indeed, the constant involved in the upper bound of Theorem E.3.2 can be very large in all generality, as noticed in a particle filter environment in Doucet et al. (2001) [54]. As a matter of fact, the determination of this constant was made by rough inequalities which could be obviously improved, as managed in Jacod-Del Moral (2002) [42] by martingale arguments. Indeed, the total variation distance between the empirical and the target distributions was roughly bounded by 1 although it should empirically tend towards 0 with iterations. More subtle bounds might be found by linking this distance to iteration n, but this should be either mathematically more complicated, or less general (only for specific kernels). Another consequence of this fact is that the utility function could be adapted by proper dynamic truncatures when samples get closer to the mode as iterations increase, in order to reduce the ratio $U_{\max}/U_{\min}$ (which obviously improves the convergence).

Let us conclude with the following remark. Using the logarithmic cooling schedule of Theorem E.3.2 would require $\left(mD^2/\sigma^2 + ln\left(U_{\max}/U_{\min}\right)\right)$ to be small in order to quickly reach the final J determined by the wanted level of accuracy. However, this is equivalent to reduce the compact size and/or the ratio $U_{\max}/U_{\min}$, driving to an over-simplified optimization problem or to a hopeless algorithm. Let us recall that the constants displayed are not optimal, but they give an intuition of the different variables impacts.

Bibliography

[1] Ait-Sahalia, Y.: Telling From Discrete Data Whether the Underlying Continuous-Time Model is a Diffusion, Journal of Finance, 57, 2075-2112 (2002a).

[2] Ait-Sahalia, Y.: Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach. Econometrica, 70, 223-262 (2002b).

[3] Ait-Sahalia, Y., Mykland, P.: The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions, Econometrica, 71, 483-549 (2003).

[4] Ait-Sahalia, Y., Mykland, P.: Estimating Diffusions with Discretely and Possibly Randomly Spaced Data: A General Theory, Annals of Statistics, 32, 2186-2222 (2004).

[5] Amzal, B., Bois, F. Y.,Parent, E., Robert, C. P.: Bayesian Optimal Design via Interacting MCMC. Journal Of the American Statistical Association, Theory and Methods, *Forthcoming* (2005).

[6] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P.: Modeling and Forecasting Realized Volatility. Econometrica, 71, 579–625 (2003).

[7] Andrieu, C., de Freitas, N., Doucet, A.: Reversible Jump MCMC Simulated Annealing for Neural Networks. Uncertainty in Artificial Intelligence (2000).

[8] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent Measure of Risk. Mathematical Finance, 9, 203-228 (1999).

[9] Aubin, J-P., Ekeland, I.: Applied Nonlinear Analysis. John Wiley and Sons (1984).

[10] Avellaneda, M., Levy, A., Paras, A.: Pricing and Hedging Derivative Securities in Markets with Uncertain Volatilities. Applied Mathematical Finance, 1(2) 73-88 (1995).

[11] Avellaneda, M.: Minimum Entropy Calibration of Asset Pricing Models. International Journal of Theoretical and Applied Finance, 1, 47-72 (1998).

[12] Becherer, D.: Rational Hedging and Valuation with Utility-Based Preferences. Technical University of Berlin, PhD Thesis (2001).

[13] Basak, S.: Asset Pricing with Heterogeneous Beliefs. London Business School, Centre for Economic Policy Research, *Working Paper* (2004).

[14] Bates, D.S.: Jumps and Stochastic Volatility: The Exchange Rate Processes Implicit in Deutschmark Options. Reviews of Financial Studies, 9, 69-107 (1996).

[15] Bellini, F., Frittelli, M.: On the Existence of Minimax Martingale Measures. Mathematical Finance, 12(1), 1-21 (2002).

[16] Benes, V., Karatzas, I.: Estimation and Control for Linear Partially Observable Systems with non-Gaussian Initial Distribution. Stochastic Processes and their Applications, 14, 233-248 (1983).

[17] Ben Hamida, S., Cont, R.: Recovering Volatility from Option Prices by Evolutionary Optimization. Ecole Polytechnique, *Working Paper* (2004).

[18] Berrada, T.: Incomplete Information, Heterogeneity and Asset Pricing. HEC Montréal, *Working Paper* (2004).

[19] Black, F., Scholes, M.: The Pricing of Options and Corporate Liabilites. Journal of Political Economy, 81, 637-654 (1973).

[20] Brémaud, P.: Point Processes and Queues: Martingale Dynamics. Springer Verlag, New York (1981).

[21] Brooks, S. P., Morgan, B. J. T.: Optimization using Simulated Annealing. The Statistician, 44, 241-257 (1995).

[22] Cagetti, M., Hansen, L. P., Sargent, T., Williams, N.: Robustness and Pricing with Uncertain Growth. The Review of Financial Studies, 15(2), 363-404 (2002).

[23] Cappé, O., Guillin, A., Marin, J.M., and Robert, C.P.: Population Monte Carlo. Journal of Computational and Graphical Statistics, 13(4), 907-929 (2004).

[24] Carr, P., Madan, D.: Option Valuation Using the Fast Fourier Transform. Journal of Computational Finance, 2, 61-73 (1998).

[25] Carpenter, J., Clifford, P., Fearnhead, P.: An Improved Particle Filter for Nonlinear Problems. IEEE Proceedings: Radar, Sonar and Navigation, 146(1), 2-7 (1999).

[26] Chernov, M., Ghysels, E.: A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Option Valuation. Journal of Financial Economics, 56, 407-458 (2000).

[27] Chiarella, C., Pasquali, S., Runggaldier, W. J.: On Filtering in Markovian Term Structure Models (an approximation approach). Advances in Applied Probability, 33, 794-809 (2001).

[28] Chopin, N.: A Sequential Particle Filter Method for Static Models. Biometrika, 89, 539-552 (2002).

[29] Christoffersen, P., Jacobs, K.: The Importance of the Loss Function in Option Pricing. Cirano, *Working Paper* (2001).

[30] Clark, J. M. C.: The Representation of Functionals of Brownian Motion as Stochastic Integrals. Annals of Mathematical Statistics 41, 1282-1295 (1970).

[31] Clarke, F.H.: Optimization and Non-Smooth Analysis. New York: Springer-Verlag.

[32] Cont, R., Tankov, P.: Non-Parametric Calibration of Jump-Diffusion Option Pricing Models. Journal of Computational Finance, 7 (2004).

[33] Cox, J. C., Ingersoll, J. E., Ross, S. A.: An Intertemporal General Equilibrium Model of Asset Prices. Econometrica, 51, 363-383 (1985).

[34] Cox, J., Huang, C. F.: Optimal Consumption and Portfolio Policies when Asset Process follow a Diffusion Process. Journal of Economic Theory 49, 33–83 (1989).

[35] Crisan, D., Gaines, J., Lyons, T.: Convergence of a Branching Particle Method to the Solution of the Zakai Equation. SIAM Journal of Applied Mathematics, 58/5, 1568–1590 (1998).

[36] Cvitanic, J.: Minimizing Expected Loss of Hedging in Incomplete and Constraint Markets. SIAM Journal of Control and Optimization, 38(4), 1050-1066 (2000).

[37] Davis, M. H. A.: Option Pricing in Incomplete Markets. in Mathematics of Derivative Securities, 216–226, Publications of the Newton Institute, Cambridge University Press (1997).

[38] De Groot, M.: Optimal Statistical Decisions. New York, McGraw Hill (1970).

[39] Del Moral, P.: Measure Valued Processes and Interacting Particle Systems: Application to Nonlinear Filtering Problems. Annals of Applied Probability, 2, 438-495 (1998).

[40] Del Moral, P., Guionnet, A.: On the Stability of Measure Valued Processes: Applications to Nonlinear Filtering Problems. Annales de l'Institut Henri Poincaré, 37(2), 155-194 (2001).

[41] Del Moral, P., Kouritzin, M., Miclo, L.: On a Class of Discrete-Generation Interacting Particle Systems. Electronic Journal of Probability, 6(16), 1-26 (2001).

[42] Del Moral, P., Jacod J.: The Monte-Carlo Method for Filtering with Discrete Time Observations: Central Limit Theorems. The Fields Institute Communications, Numerical Methods and Stochastics, Ed. T.J. Lyons, T.S. Salisbury, American Mathematical Society (2002).

[43] Delbaen, F., Schachermayer, W.: A General Version of the Fundamental Theorem of Asset Pricing. Mathematische Annalen, 300, 463–520 (1994).

[44] Delbaen, F., Schachermayer, W.: The Variance Optimal Martingale Measure for Continuous Processes. Bernoulli 9, 81-105 (1996).

[45] Dembo, A., Zeintouni, O.: Parameter Estimation of Partially Observed Continuous Time Stochastic Processes. Stochastic Processes and their Applications 23, 91–113 (1986).

[46] Denis, L.: Theory of Risk-Measure in a Non-Dominated Model. *Working Paper* (2000).

[47] Denis, L., Martini, C.: A Theorical Framework for the Pricing of Continent Claims in the Presence of Model Uncertainty. *Working Paper* (2001).

[48] Detemple, J.: Asset Pricing in a Production Economy with Incomplete Information. The Journal of Finance, 41, 383-391 (1986).

[49] Detemple, J. B., Garcia, R., Rindisbacher, M.: A Monte Carlo Method for Optimal Portfolios. The Journal of Finance 58/1, 401-446 (2003).

[50] Detemple, J., Murthy, S.: Intertemporal Asset Pricing with Heterogeneous Beliefs. Journal of Economic Theory, 62, 294-320 (1994).

[51] Detemple, J.: Asset Pricing in an Intertemporal Partially-Revealing Rational Expectations Equilibrium. Journal of Mathematical Economics, 38, 219-248 (2002).

[52] Dobrushin, R. L.: Prescribing a System of Random Variables by Conditional Distributions. Theory of Probability and Applications, 15, 458-486 (1970).

[53] Dothan, M. U., Feldman, D.: Equilibrium Interest Rates and Multiperiod Bonds in a Partially Observable Economy. Journal of Finance, 41, 369-382 (1986).

[54] Doucet, A., de Freitas, N., Gordon, N.: Sequential Monte-Carlo Methods in Practice. Springer-Verlag (2001).

[55] Duffie, D.: Dynamic Asset Pricing Theory. Princeton University Press, Princeton, New Jersey (1992).

[56] Duffie, D., Filipovic, D., Schachermayer, W.: Affine Processes and Applications in Finance. Annals of Applied Probability, 13, 984-1053 (2003).

[57] Duffie, D., Glynn, P.: Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals. Econometrica, 72(6), 1773-1808 (2004).

[58] Ellsberg, D.: Risk, Ambiguity, and the Savage Axioms," Quarterly Journal of Economics 75, 643-669 (1961).

[59] Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer Academinc Publishers (1996).

[60] Eraker, B., Johannes, M. S., Polson, N. G.: The Impact of Jumps in Returns and Volatility. Journal of Finance, 58/3, 1269–1300 (2003).

[61] Eraker, B.: Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. Journal of Finance, 59, 1367-1403 (2004).

[62] Esche, F.: Two Essays on Incomplete Markets. PhD Thesis, Technical University of Berlin (2003).

[63] Frittelli M.: The Minimal Entropy Martingale Measure and the Valuation Problem in Incomplete Markets. Mathematical Finance 10, 39-52 (2000).

[64] Fleming, W.H., Rishel, R.: Deterministic and Stochastic Optimal Control. Springer-Verlag, New York (1975).

[65] Fleming, W.H., Soner, H.M.: Controlled Markov Processes and Viscosity Solutions. Springer-Verlag, New York (1993).

[66] Foellmer, H.: Calcul d'itô sans Probabilités. Séminaire de Probabilités de Strasbourg, XV, Lecture Notes in Mathematics 850, 143-150 (1981).

[67] Foellmer, H., Leukert, P.: Efficient Hedging: Balancing Cost versus Shortfall Risk. Finance and Stochastics, 4, 117-146 (2000).

[68] Foellmer, H., Schied, A.: Robust Preferences and Convex Measures of Risk. Advances in Finance and Stochastics, Springer-Verlag (2002).

[69] Foellmer, H., Gundel, A.: Robust Projections in the Class of Martingale Measures. Humboldt Universität zu Berlin, *Working Paper* (2006).

[70] Framstad, N. C., Oksendal, B., Sulem, A.: Optimal Consumption and Portfolio in a Jump Diffusion Market. Workshop on Mathematical Finance (2001).

[71] Frey, R., Runggaldier, W. J.: A Nonlinear Filtering Approach to Volatility Estimation with a View Towards High Frequency Data. International Journal of Theoretical and Applied Finance, 4, 199-210 (2001).

[72] Frydman, H., Lakner, P.: Maximum Likelihood Estimation of Hidden Diffusions. The Annals of Applied Probability, 13, 1296-1312 (2003).

[73] Geman, S., Geman, D.: Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(6), 721-741 (1984).

[74] Gennotte, G.: Optimal Portfolio Choice under Incomplete Information. Journal of Finance, 41, 733-746 (1986).

[75] Geweke, J.: Bayesian Inference in Econometric Models using Monte Carlo Integration. Econometrica, 24, 1317-1399 (1989).

[76] Gilboa, I.: Expected Utility with Purely Subjective Non-Additive Probabilities. Journal of Mathematical Economics, 16, 65-88 (1987).

[77] Gilboa, I., Schmeidler, D.: Maxmin Expected Utility with Non-Unique Prior. Journal of Mathematical Economy, 18, 141-153 (1989).

[78] Goll, T., Ruschendorf, L.: Minimax and Minimal Distance Martingale Measures and Their Relationship to Portfolio Optimization. Finance and Stochastics, 5, 557-581 (2001).

[79] Green, R., Jarrow, R. A.: Spanning and Completeness in Markets with Contingent Claims. Journal of Economic Theory, 41, 202-210 (1987).

[80] Gundel, A.: Robust Utility Maximization for Complete and Incomplete Market Models. Finance and Stochastics, 9 (2), 151-176 (2005).

[81] Haario, D., Laine, M., Mira, A., Saksman E.: Efficient Adaptive MCMC. University of Helsinki, Technical Report (2003).

[82] Hansen, P. L., Sargent, T. J.: Robust Control and Model Uncertainty. The American Economic Review, 2, 60-66 (2001).

[83] Hansen, P. L., Sargent, T. S., Turmuhambetova, G. A., Williams, N.: Robust Control and Model Misspecification. Journal of Economic Theory *forthcoming* (2005).

[84] Harrison, M., Kreps, D.: Martingales and Arbitrage in Multiperiod Securities Markets. Journal of Economic Theory, 20, 381-408 (1979).

[85] Harrison, M., Pliska, S.: Martingales and Stochastic Integrals in the Theory of Continuous Trading. Stochastic Processes and Their Applications, 11, 215-260 (1981).

[86] Hastings, W.K.: Monte Carlo Sampling Methods using Markov Chains and their Applications. Biometrika, 57, 97-109 (1970).

[87] Haussmann, U. G.: On the Integral Representation of Functionals of Itô Processes. Stochastics 3, 17-28 (1979).

[88] Heston, S.L.: A Closed-Form Solution for Option with Stochastic Volatility with Applications to Bond and Currency Options. Review of Financial Studies, 6(2), 327-343 (1993).

[89] Hodges, S. D., Neuberger, A.: Optimal Replication of Contingent Claims under Transaction Costs. Review of Futures Markets 8, 222-239 (1989).

[90] Huber, P.: Robust Statistics. Wiley, New York (1981).

[91] Itô, K., Spectral type of the Shift Transformation of Differential Processes with Stationary Increments. Transactions of the American Mathematical Society, 81, 252-263 (1956).

[92] Jacod, J.: Multivariate Point Processes: Predictable Projection, Radon-Nikodym Derivatives, Representation Martingales. Z. Wahrsch. Verw. Gebiete, 31, 235-53 (1975).

[93] Jacod, J., Shiryaev, A. : Limit Theorems for Stochastic Processes. Berlin: Springer (2003).

[94] Jeanblanc-Picqué, M., Pontier, M.: Optimal Portfolio for a Small Investor in a Market Model with Discontinuous Prices. Appl. Math. Opt., 22, 287–310, (1990).

[95] Jeanblanc, M., Lacoste, V., Roland, S.: Utility Maximization under Partial Information in Jump-Diffusions Models. Université d'Evry and Essec Business School. Preprint (2005).

[96] Kallianpur, G.: Stochastic Filtering Theory, New York: Springer Verlag (1980).

[97] Kallsen, J.: Duality Links Between Portfolio Optimization and Derivative Pricing. Technical Report, Mathematische Fakult at Universit at Freiburg i Br (1998).

[98] Karatzas, I., Lehoczky, J. P., Shreve, S.: Optimal Portfolio and Consumption Decisions for a Small Investor on a Finite Horizon. Journal of Control and Optimization 25, 1557–1586 (1987).

[99] Karatzas, I. Xue, X..: A Note on Utility Maximization under Partial Observations. Mathematical Finance, 1, 57-70 (1991).

[100] Karatzas, I., Shreve, S.E.: Methods of Mathematical Finance. Springer-Verlag, New York (1998).

[101] Karni, E.: Foundations of Bayesian Theory. Department of Economics, Johns Hopkins University, *Working Paper* (2005).

[102] Keynes, J.M.: A Treatise on Probability, Macmillan, London (1921).

[103] Kirch, M., Runggaldier, W.J.: Efficient Hedging when Asset Prices Follow a Geometric Poisson Process with Unknown Intensities. SIAM Journal on Control and Optimization, 43 1174-1195 (2004).

[104] Kliemann, W., Koch, G., Marchetti, F.: On the Unnormalized Solution of the Filtering Problem with Counting Process Observations. IEEE Transactions on Information Theory 36, 1415-1425 (1990).

[105] Kloeden, P. E., Platen, E.: Numerical Solution of Stochastic Differential Equations, vol. 23 of Applications of Mathematics. Springer Verlag, 3 edn. 1999.

[106] Knight, F.: Risk, Uncertainty and Profit. Houghton Mifflin, Boston (1921).

[107] Kramkov, D., Schachermayer, W.: The Asymptotic Elasticity of Utility Functions and Optimal Investment in Incomplete Markets. Annals of Applied Probability 9, 904–950 (1999).

[108] Kurtz, T., Ocone, D.: Unique Characterization of Conditional Distributions in Nonlinear Filtering. Annals of Probability 16, 80-107 (1988).

[109] Kushner, H., Dupuis, P.: Numerical Methods for Stochastic Control Problems in Continuous-Time. Springer Verlag, New-York (1992).

[110] Kuwana, Y.: Certainty Equivalence and Logarithmic Utilities in Consumption/Investment Problems. Mathematical Finance 5, 297-310 (1995).

[111] Lakner, P.: Optimal Trading Strategy for an Investor - The Case of Partial Information. Stochastic Processes and their Applications, 76, 77-97 (1998).

[112] Lapeyre, B., Sulem, A., Talay, D.: Understanding Numerical Analysis for Financial Models. Cambridge University Press (2006).

[113] Lasry, J. M., Lions, P.-L.: Stochastic Control under Partial Information and Applications To Finance. Comptes Rendus De L'Academie Des Sciences Serie I - Mathematique, 328, 1003-1010 (1999).

[114] Lehmann, E., L.: Testing Statistical Hypotheses. Wiley: New York (1986).

[115] Léon, J.A., Solé, J. L., Utzet, F., Vives, J.: On Lévy processes, Malliavin Calculus and Market Models With Jumps. Finance and Stochastics 6/2, 197-225 (2002).

[116] Liese, F., Vajda, I.: Convex Statistical Distances. Teubner: Leipzig (1987).

[117] Liptser, R., Shiryayev, A. N.: Theory of Martingales. Kluwer (1986).

[118] Liptser, R. S., Shiryaev, A. N.: Statistics of Random Processes I&II, 2nd Edition, Springer, New York (2001).

[119] Liu, J.S., Chen, R.: Sequential Monte Carlo Methods for Dynamic Systems. Journal of American Statistical Association, 113-119 (1998).

[120] Liu, J., Longstaff, F., Pan, J.: Dynamic Asset Allocation with Event Risk. The Journal of Finance 58/1, 231–259 (2003).

[121] Lokka, A.: White Noise Claculations for Pure Jump Processes with Application to Mathematical Finance. PhD thesis, Department of Mathematics, University of Oslo (2003).

[122] Lyons, T.: Uncertain Volatility and the Risk Free Synthesis of Derivatives. Applied Mathematical Finance, 1(2) 117-133 (1995).

[123] Markowitz, H.: Portfolio Selection. Journal of Finance, 7, 77-91 (1952).

[124] Maybeck, P. S.: Stochastic Models: Estimation and Control. London: Academic Press 1982.

[125] Merton, R.: Lifetime Portfolio Selection under Uncertainty: the Continuous Case. The Review of Economics and Statistics, 51, 247-257 (1969).

[126] Merton, R. C.: Optimum Consumption and Portfolio Rules in a Continuous Time Model. Journal of Economic Theory, 3, 373-413 (1971).

[127] Merton, R. C.: An Intertemporal Capital Asset Pricing Model. Econometrica, 41, 867-887 (1973).

[128] Merton, R. C.: Option Pricing when Underlying Stock Returns are Discontinuous. Journal of Financial Economics, 3(2), 125-144 (1976).

[129] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E.: Equations of State Calculations by Fast Computing Machines. Journal of Chemical Physics, 21(6) 1087-1092 (1953).

[130] Muller, P., Parmigiani, G.: Optimal Design via Curve Fitting of Monte Carlo Experiments. Journal of the American Statistical Association, 90, 1322-1330 (1996).

[131] Muller, P.: Simulation-Based Optimal Design. Bayesian Statistics, 6, 459-474 (1999).

[132] Nualart, D.: The Malliavin Calculus and Related Topics. Springer, Berlin (1995).

[133] Ocone, D.: Malliavin's Calculus and Stochatic Integral Representation of Functionals of Diffusion Processes. Stochastics 12, 161-185 (1984).

[134] Ocone, D., Karatzas, I.: A Generalized Clark Representation Formula, with Application to Optimal Portfolios. Stochastics and Stochastic Reports 34, 187-220 (1991).

[135] Owen, M. P.: Utility Based Optimal Hedging in Incomplete Markets. Annals of Applied Probability 12/2, 691–709 (2002).

[136] Pagès, G., Pham, H., Printems, J.: Optimal Quantization Methods and Applications to Numerical Problems in Finance. Handbook of Computational and Numerical Methods in Finance, Springer Verlag (2004).

[137] Pan, J.: The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study. Journal of Financial Economics, 63, 3-50 (2002).

[138] Pham, H.: Smooth Solutions to Optimal Investment Models with Stochastic Volatilities and Portfolio Constraints. Applied Mathematics and Optimization 46/1, 55-78 (2002).

[139] Pham, H., Quenez, M.-C.: Optimal Portfolio in Partially Observed Stochastic Volatility Models. The Annals of Applied Probability 11(1), 210-238 (2001).

[140] Protter, P.: Stochastic Integration and Differential Equations. New York: Springer Verlag 1990.

[141] Quenez, M.-C.: Optimal Portfolio in a Multiple-Priors Model. Seminar on Stochastic Analysis, Random Fields and Applications IV, 291–321 (2004).

[142] Robert, C. P., Casella, G.: Monte Carlo Statistical Methods., Springer-Verlag (2004).

[143] Rockafellar, R. T.: Convex Analysis. Princeton University Press, New York: Princeton (1970).

[144] Rubin, D.B.: Bayesian Statistics, volume, 3, chapter Using the SIR Algorithm to Simulate Posterior Distributions, 395-402. Oxford University Press (1988).

[145] Runggaldier, W. J.: Jump Diffusion Models. in Handbook of Heavy Tailed Distributions in Finance, Handbooks in Finance Book 1 169-209. North-Holland: Elsevier, 2003.

[146] Runggaldier, W. J.: Estimation via Stochastic Filtering in Financial Market Models. In Mathematics of Finance (G.Yin and Q.Zhang eds.). Contemporary Mathematics 351, 309-318 (2004).

[147] Roland, S.: Optimal Policies from Prices Observed at Random Times. Université d'Evry & Essec Business School, Preprint (2005a).

[148] Roland, S.: Financial Value of Information in Complete Market Models. Université d'Evry & Essec Business School, Preprint (2005b).

[149] Sass, J., Haussmann, U.: Optimizing the Terminal Wealth under Partial Information: The Drift Process as a Continuous Markov Chain. Finance and Stochastics 8, 553-577 (2004).

[150] Savage, L.: The Foundations of Statistics. John Wiley, New York (1954).

[151] Schachermayer, W.: Utility Maximisation in Incomplete Markets. Lecture Notes in Mathematics, Vol. 1856, 225-288 (2004).

[152] Scheinkman, J. A., Xiong, W.: Heterogeneous Beliefs, Speculation and Trading in Financial Markets. Paris-Princeton Lectures on Mathematical Finance 2003, Lecture Notes in Mathematics 1847, Springer-Verlag, Berlin (2004).

[153] Schied, A.: Optimal Investments for Robust Utility Functionals in Complete Market Models. Mathematics of Operations Research, 30 (3), 750-764 (2005).

[154] Schmeidler, D.: Subjective Probability and Expected Utility without Additivity. Econometrica, 57, 571-587 (1989).

[155] Smith, A.F.M., Gelfand, A.E.: Bayesian Statistics without Tears: a Sampling-Resampling Perspective. The American Statistician, 46, 84-88 (1992).

[156] Solé, J. L., Utzet, F., Vives, J.: Canonical Lévy Process and Malliavin Calculus. Universitat Autonoma de Barcelona, *Working Paper* (2005).

[157] Strasser, H.: Mathematical Theory of Statistics. Studies in Mathematics: de Gruyter (1985).

[158] Zariphopoulou, T.: A Solution Approach to Valuation with Unhedgeable Risks. Finance and Stochastics 5/1, 61-82 (2001).