HEC MONTRÉAL
École affiliée à l'Université de Montréal

Couverture globale de risques en marché incomplet

par

Frédéric Godin

Thèse présentée en vue de l'obtention du grade de Ph. D en administration
(option Méthodes Quantitatives)

Mai 2014

HEC MONTRÉAL

École affiliée à l'Université de Montréal

Cette thèse intitulée :

Couverture globale de risques en marché incomplet

Présentée par :

Frédéric Godin

a été évaluée par un jury composé des personnes suivantes :

Chantal Labbé
HEC Montréal
Président-rapporteur

Geneviève Gauthier
HEC Montréal
Directrice de recherche

Bruno Rémillard
HEC Montréal
Membre du jury

Peter Tankov
Université Paris-Diderot
Examinateur Externe

Nicolas A. Papageorgiou
HEC Montréal
Représentant du directeur de HEC Montréal

Résumé

Une approche de couverture de risques de type globale est développée dans différents contextes de marchés incomplets. Une telle méthode utilise l'optimisation stochastique afin de minimiser une mesure de risque appliquée à l'erreur terminale de couverture. La présente thèse est séparée en trois articles. L'article *«Optimal hedging when the underlying asset follows a regime-switching Markov process»* illustre l'application de la couverture globale dans le contexte d'un sous-jacent suivant un modèle à changements de régime. La programmation dynamique est utilisée afin de minimizer l'espérance d'une fonction de pénalité quelconque appliquée à l'erreur de couverture. Le second article, «*Minimizing CVaR in global dynamic hedging with transaction costs*», considère la présence de frais de transactions dans l'optimisation de la stratégie de couverture et minimise la Valeur-à-Risque conditionnelle (CVaR) de l'erreur de couverture. Finalement, le troisième article «*Short-term hedging for an electricity retailer*» applique le concept de couverture global au contexte concret des marchés de l'électricité afin de réduire les risques de prix et de volume d'un détaillant du marché scandinave Nord Pool. La dynamique stochastique de la consommation d'électricité et du prix des contrats à termes sur l'électricité est modélisée dans l'article.

Mots clés : Gestion de risques, couverture de risques, ingénierie financière, changements de régime, électricité, programmation dynamique, modélisation du risque, coûts de transaction, modélisation mathématique, recherche quantitative.

Summary

A global risk hedging approach in different incomplete market contexts is developed. This approach uses stochastic optimization to minimize a risk measure applied to terminal hedging errors. This thesis is separated into three articles. The article *«Optimal hedging when the underlying asset follows a regime-switching Markov process»* applies the global hedging methodology in a regime-switching context. Dynamic programming is used to minimize the expectation of a penalty function applied to hedging errors. The second article, «*Minimizing CVaR in global dynamic hedging with transaction costs*» considers the presence of transaction costs during the optimization of the hedging strategy and minimizes the hedging conditional Value-at-Risk (CVaR). The third article, «*Short-term hedging for an electricity retailer*», applies the global hedging methodology to the concrete context of electricity markets in order to mitigate price and load risks of a retailer operating in the Nord Pool Scandinavian market. The stochastic dynamics of electricity consumption and electricity futures prices is modeled in the article.

Keywords : Risk management, hedging, financial engineering, regime-switching, electricity, dynamic programming, risk modeling, transaction costs, mathematic modeling, quantitative research.

Table des matières

Liste des tableaux

Liste des figures

Remerciements

Je tiens à remercier Geneviève Gauthier, ma superviseure, pour toutes les connaissances qu'elle m'a transmises, tous les conseils qu'elle m'a prodigués et pour les nombreuses heures au cours de mon doctorat où elle a travaillé afin de me guider et m'aider à m'améliorer. Je souhaite aussi souligner sa contribution notoire aux articles contenus dans cette thèse.

Je veux aussi remercier les deux autres co-auteurs de mes articles, Pascal François et Debbie Dupuis, qui ont contribué à la complétion de mes projets de recherche et m'ont permis de bénéficier de leur expertise et leurs judicieux conseils.

Je suis reconnaissant envers les membres de mon comité de thèse, Peter Tankov, Bruno Rémillard et Chantal Labbé, qui ont pris le temps de s'intéresser à mes travaux et d'effectuer plusieurs recommandations afin de les améliorer.

Je remercie aussi le Conseil de recherche en sciences naturelles et en génie (CRSNG), la Bourse de Montréal, le HEC Montréal et Geneviève Gauthier pour m'avoir octroyé le soutien financier nécessaire à la complétion de mes doctorales.

Je veux aussi remercier les nombreux professeurs qui m'ont enseigné au cours de mes études, m'ont transmis leur passion pour le travail de chercheur et m'ont influencé dans mon parcours académique, dont Manuel Morales, Charles Dugas, Marlène Frigon, Michèle Breton et Bruno Rémillard.

Je tiens tout d'abord à remercier ma famille et mes amis dont ma copine Anne-Marie, mon père Robert, ma mère Monique et mon frère Sébastien. Ils ont été présents tout au long de mon parcours scolaire, et je n'aurais pu me rendre où je suis aujourd'hui si je n'avais pu bénéficer de leur amour, de leurs encouragements et de leur support, autant financier que moral.

Chapitre 1

Introduction générale

La crise financière de 2008 a démontré l'importance de la gestion des risques au sein des institutions financières. Une manière de réduire les risques associés à la possession de produits dérivés est de couvrir ceux-ci en achetant d'autres actifs financiers qui contrebalancent le risque encouru. Il est important de pouvoir réaliser cette tâche de manière optimale afin d'éliminer le mieux possible la portion de risque que l'institution financière ne désire pas garder.

L'avènement du modèle de Black-Scholes (1973) en finance a amené un large pan de la litérature à considérer les modèles de marchés complets où n'importe quel actif contingent peut être répliqué à l'aide d'un portefeuille dynamique contenant les actifs transigés au marché. Ceci permet la neutralisation complète des risques associés à n'importe quel produit dérivé en utilisant la stratégie de couverture auto-financée qui mène à la réplication de celui-ci.

Or, dans la réalité, la réplication exacte de produits dérivés est entravée par plusieurs caractérisques des marchés réels : présence de coûts de transaction, rebalancements en temps discret plutôt qu'en temps continu, sauts dans la trajectoire du prix des actifs financiers, etc. L'impossibilité de répliquer certains produits dérivés est ce qui caractérise les marchés dits incomplets, et les articles de la présente thèse s'inscrivent dans un tel contexte. Dans cette situation, faute de pouvoir complètement éliminer le risque, il est souhaitable de pouvoir le réduire de manière optimale.

Différents critères sont proposés dans la littérature afin de déterminer ce en quoi consiste une stratégie de couverture optimale. La plupart des méthodes se basent sur des critères de type local. La méthode la plus connue, le *delta-hedging*, permet d'obtenir un portefeuille dont les variations de valeur répliquent approximativement celles du prix du produit dérivé couvert lorsque de petites variations du sous-jacent se produisent. Cette méthode peut être étendue en recherchant un portefeuille de couverture qui réplique la sensibilité du produit dérivé par rapport à d'autres facteurs de risque tels que les taux d'intérêt (*rho-hedging*), la volatilité sur les marchés (*vega-hedging*) ou le temps (*theta-hedging*). Ces méthodes sont dites locales par rapport au mouvement (*move-based*) ; elles cherchent à répliquer le mieux possibles les mouvements du prix du produit dérivé lorsque de petits variations des facteurs de risque se produisent. D'autres méthodes sont dites locales par rapport au temps (*time-based*). Elles cherchent à minimiser une mesure de risque appliquée aux pertes dues au produit dérivé et au portefeuille de couverture d'ici le prochain rebalancement de ce dernier (i.e. pour la période de temps courante).

Un problème associé aux méthodes de couverture locales est qu'elles sont myopes : elles ignorent tout ce qui peut se passer après que l'incrément de temps ou le léger mouvement du sous-jacent se soient produits. L'approche de couverture globale vient corriger cette lacune : le critère utilisé par celle-ci est la minimisation d'une mesure de risque appliquée à l'erreur terminale de couverture i.e. à l'écart entre le paiement à maturité d'un produit dérivé et la valeur du portefeuille de couverture à cette date. La présente thèse développe cette méthode dans différentes situations : lorsque le sous-jacent suit un modèle à changement de régime, lorsque des frais de transactions sont encourus lors des rebalancements du portefeuille de couverture et dans le cadre des marchés de l'électricité.

Le développement d'un algorithme de couverture global pour couvrir les risques associés à un produit dérivé donné nécessite la réalisation des étapes suivantes :

- Sélection des actifs à utiliser pour la couverture,
- Modélisation de la dynamique stochastique des actifs utilisés dans la couverture et des variables d'états caractérisant le paiement à maturité du produit dérivé,
- Développement de l'algorithme théorique identifiant la stratégie de couverture optimale,
- Développement de l'algorithme numérique calculant la stratégie de couverture optimale.

La présente thèse illustre la réalisation de chacune de ces étapes dans différents contextes. De plus, dans chacun des articles, la performance de la méthode de couverture globale est comparée avec celle d'autres méthodes alternatives proposées dans la littérature à l'aide de simulations numériques. La méthode de couverture globale se compare avantageusement aux autres méthodes en terme de capacité à réduire les risques associés au produit dérivé qui est couvert.

Cette thèse est séparée en trois chapitres principaux qui présentent le contenu de trois articles de recherche. Le chapitre 2 traite de la couverture globale dans le cadre des modèles à changement de régime. Le chapitre 3 illustre la minimisation de la mesure de risque CVaR dans un contexte de couverture globale en présence de frais de transaction. Le chapitre 4 illustre la situation d'un détaillant du marché d'électricité Nord Pool qui effectue une couverture globale de ses risques de prix et de volume à l'aide de contrats à terme. La section 5 conclut.

Chapitre 2

Couverture et changements de régime

Optimal hedging when the underlying asset follows a regime-switching Markov process[1]

By Pascal François[2], Geneviève Gauthier[3] and Frédéric Godin[4]

Abstract

We develop a flexible discrete-time hedging methodology that minimizes the expected value of any desired penalty function of the hedging error within a general regime-switching framework. A numerical algorithm based on backward recursion allows for the sequential construction of an optimal hedging strategy. Numerical experiments comparing this and other methodologies show a relative expected penalty reduction ranging between 0.9% and 12.6% with respect to the best benchmark.

JEL classification : G32, C61

Keywords : Dynamic programming, hedging, risk management, regime switching.

---

1. Earlier versions of this paper were presented at the IFM2 Mathematical Finance Days 2012, CORS Annual Conference 2012 and 25th Annual Australasian Finance and Banking Conference 2012. We thank the conference participants for their helpful feedback. Financial support from SSHRC (François), NSERC (Gauthier, Godin) and the Montreal Exchange (Godin) is gratefully acknowledged.

2. Professor at HEC Montréal, Department of Finance and CIRPÉE Fellow, 3000 ch. Côte-Ste-Catherine, Montréal, Québec, Canada, H3T 2A7. pascal.francois@hec.ca.

3. Professor at HEC Montréal, Department of Management Sciences, 3000 ch. Côte-Ste-Catherine, Montréal, Québec, Canada, H3T 2A7, tel. 001 (514) 340-5627. genevieve.gauthier@hec.ca

4. Ph.D. student at HEC Montréal, Department of Management Sciences, 3000 ch. Côte-Ste-Catherine, Montréal, Québec, Canada, H3T 2A7, frederic.godin@hec.ca.

2.1 Introduction and literature review

For a derivatives trading and risk management activity to be sustainable, hedging is paramount. In practice, portfolio rebalancing is performed in discrete time and the market is typically incomplete, implying that most contingent claims cannot be replicated exactly. Thus, to implement a hedging policy, the challenge is twofold : a model must be specified and hedging strategy objectives must be set.

From a modelling perspective, this article adopts a regime-switching environment. One widely studied class of regime-switching models views log-returns as a mixture of Gaussian variables. These models, introduced in finance by Hamilton (1989), have been shown to improve the statistical fit and forecasts of financial returns. They reproduce widely documented empirical properties such as heteroskedasticity, autocorrelation and fat tails. In this framework, the option pricing problem must deal with incomplete markets and requires the specification of a risk premium. Among significant contributions, Bollen (1998) presents a lattice algorithm to compute the value of European and American options. Hardy (2001) finds a closed-form formula for the price of European options. The continuous-time version of the Gaussian mixture model is studied by Mamon & Rodrigo (2005) who find an explicit value for European options by solving a partial differential equation. Elliott et al. (2005) price derivatives by means of the Esscher transform under the same continuous-time model. Buffington & Elliott (2002) derive an approximate formula for American option prices. Beyond the Gaussian mixture models, extensions address GARCH effects (Duan et al., 2002) and jumps (Lee, 2009a), for example.

Several authors study the problem of hedging an underlying asset with its futures under regime-switching frameworks. Alizadeh & Nomikos (2004) and Alizadeh et al. (2008) base their hedging strategy on minimal variance hedge ratios. Lee et al. (2006), Lee & Yoder (2007), Lee (2009a) and Lee (2009b) extend the dynamics of the underlying asset in Alizadeh & Nomikos (2004) to incorporate a time-varying correlation between the spot and futures returns, GARCH-type feedback from returns on the volatilty, jumps and copulas for the dependence between futures and spot returns. Lien (2012) provides conditions under which minimal variance ratios taking into account the existence of regimes overperform their unconditional counterparts.

Option hedging under regime-switching models has recently raised interest in the literature. Rémillard & Rubenthaler (2013) adapt the work of Schweizer (1995) to a regime-switching framework and identify the hedging strategy that minimizes the squared error of hedging in both discrete-time and continuous-time for European options. The implementation of this methodology is present in Rémillard et al. (2010). Rémillard et al. (2012) extend the hedging procedure to American options.

Another strand of literature discusses self-financing hedging policies [5] under general model assumptions. A widely known methodology is delta hedging. It consists in building a portfolio whose value variations mimick those of the hedged contingent claim when small changes in the underlying asset's value occur. In continuous-time complete markets, delta hedging is the cornerstone of any hedging strategy since it allows for perfect replication. Based on the first derivative of the option price with respect to the underlying asset price, it requires a full characterization of the risk-neutral measure. Many authors discuss the implementation of delta hedging in discrete-time and/or incomplete markets (Duan, 1995, among others). It should be stressed, however, that delta hedging is subject to model misspecification. Nevertheless, it stands as a relevant benchmark when it comes to assessing the performance of a hedging strategy.

Another approach is super-replication (e.g. El Karoui & Quenez, 1995, and Karatzas 1997). It identifies the cheapest trading strategy whose terminal wealth is at least equal to the derivative's payoff. Since the option buyer alone carries the price of the hedging risk, the initial capital required is often unacceptably large. Eberlein & Jacod (1997) show that, under many models, the initial capital required to super-replicate a call option is the price of the underlying asset itself.

An alternative to super-replication is Global Hedging Risk Minimization (GHRM), which consists in identifying trading strategies that replicate the derivative's payoff as closely as possible, or alternatively, minimize the risk associated with terminal hedging shortfalls. Xu (2006) proposes to minimize general risk measures applied to hedging errors. Several authors choose more specific risk measures : quantiles of the hedging shortfall (Föllmer & Leukert, 1999, Cvitanić & Spivak, 1999), expected hedging shortfall (Cvitanić & Karatzas, 1999),

5. By contrast, local risk-minimization, which considers hedging strategies that are not self-financing, selects one that minimizes a measure of the costs related to non-initial investments in the portfolio (Schweizer, 1991).

expected powers of the hedging shortfall (Pham, 2000), Tail Value-at-Risk (Sekine, 2004), expected squared hedging error (Schweizer, 1995, Motoczyński, 2000, Cont et al., 2007 and Rémillard & Rubenthaler, 2013) and the expectation of general loss functions (Föllmer & Leukert, 2000). Theoretical existence of optimal hedging strategies under those risk measures and their characterization are studied in a general context. However, explicit solutions exist only for some particular cases of market setups and risk measures. The implementation of the preceding methodologies in the case of incomplete markets is often not straightforward, and tractable algorithms computing the optimal strategies have yet to be identified. The presence of regimes adds an additional layer of difficulty in applying those methods.

This paper's contributions are twofold. First, on a theoretical level, we develop a discrete-time hedging methodology with the GHRM objective that miminizes the expected value of any desired penalty function of the hedging error within a general regime-switching framework (possibly including time-inhomogeneous regime shifts). This methodology is highly flexible and generalizes the quadratic hedging approach. It incorporates a large class of penalty functions encompassing usual risk measures such as Value-at-Risk and expected shortfall. The proposed framework can accommodate portfolio restrictions such as no short-selling. Portfolios can be rebalanced more frequently than the regime-switch timeframe. Second, from an implementation perspective, a numerical algorithm based on backward recursion allows for the sequential construction of an optimal hedging strategy. Numerical experiments challenge our model with existing methodologies. The relative expected penalty reduction obtained with this paper's optimal hedging approach, in comparison with the best benchmark, ranges between 0.9% and 12.6% in the different cases exposed.

This paper is organized as follows. In Section 2.2, the market model and the hedging problem are described. In Section 2.3, the hedging problem is solved. Section 2.4 presents a numerical scheme to compute the solution to the hedging problem. Section 2.5 presents the market model used for the simulations and provides numerical results. Section 2.6 concludes the paper.

2.2 Market specifications and hedging

2.2.1 Description of the market

Transactions take place in a discrete-time, arbitrage-free financial market. Denote by Δ_t the constant time elapsing between two consecutive observations. Two types of assets are traded. The risk-free asset is a position in the money market account with a nominal amount normalized to one monetary unit. The time$-n$ price of the risk-free asset is

$$S_n^{(1)} = \exp\left(rn\Delta_t\right), \ n \in \{0, 1, 2, ...\}$$

where r is the annualized risk-free rate. The price of the risky asset, starting at $S_0^{(2)}$, evolves according to

$$S_n^{(2)} = S_0^{(2)} \exp\left(Y_n\right),$$

where Y_n is the risky asset's cumulative return over the time interval $[0, n]$. $\vec{S}_n$ denotes the column vector $\left(S_n^{(1)}, S_n^{(2)}\right)^\top$ and $\vec{S}_{0:n}$ stands for the whole price process up to time n.

The financial market is subject to various regimes that affect the dynamics of the risky asset's price. These regimes are represented by an integer-valued process $\{h_n\}_{n=0}^N$ taking values in $\mathcal{H} = \{1, 2..., H\}$ where h_n is the regime prevailing during time interval $]n, n+1]$. The joint process (Y, h) has the Markov property [6] with respect to the filtration $\{\mathcal{F}_n\}_{n=0}^N$ satisfying the usual conditions, where

$$\mathcal{F}_n = \sigma\left(\vec{S}_{0:n}, h_{0:n}\right) = \sigma\left(Y_{0:n}, h_{0:n}\right),$$

meaning that the distribution of (Y_{n+1}, h_{n+1}) conditional on information $\mathcal{F}_n$ is entirely determined by Y_n and h_n. [7] This assumption is consistent with Hamilton (1989) and Duan et al. (2002), among others. Transition probabilities of the regime process h are denoted by

$$P_{i,j}^{(n)}(y) = \mathbb{P}(h_{n+1} = j | h_n = i, Y_n = y) \quad i, j \in \mathcal{H}.$$

Because regimes h are not observable, a coarser filtration $\{\mathcal{G}_n\}_{n=0}^N$ modelling the information available to investors is required, that is, $\mathcal{G}_n = \sigma\left(Y_{0:n}\right)$.

6. A stochastic process $\{X_n\}$ has the Markov property with respect to filtration $\mathcal{F}$ if $\forall n, x$,

$$\mathbb{P}(X_{n+1} \leq x | \mathcal{F}_n) = \mathbb{P}(X_{n+1} \leq x | X_n).$$

7. Equivalently, the process $(\vec{S}, h)$ has the Markov property with respect to filtration $\mathcal{F}$.

2.2.2 The hedging problem

A market participant (referred to as the "hedger") wishes to replicate (or "hedge") the payoff $\phi(S_N^{(2)})$ of a European contingent claim written on the risky asset and maturing at time N, where $\phi(\cdot)$ is some positive Borel function $\phi : [0, \infty) \to \mathbb{R}$. Alternatively, the payoff can be written as a function of the risky asset return

$$\phi(S_N^{(2)}) = \tilde{\phi}(Y_N),$$

for some function $\tilde{\phi}(\cdot)$.

To implement the replication, the hedger adopts $\mathcal{G}-$predictable self-financing [8] hedging strategies $\theta = \left\{\vec{\theta}_n\right\}_{n=1}^{N}$ with time$-n$ value [9] $V_n(v_0, Y_{0:n}, \vec{\theta}_{1:n}) := \vec{\theta}_n^{\top} \vec{S}_n$ and initial value $V_0 := v_0 = \vec{\theta}_1^{\top} \vec{S}_0$. This ensures that all trading decisions are made based on up-to-date price information, regardless of the unobserved regime. Below, $\theta_n^{(k)}$ represents the number of shares of asset k held during period $]n-1, n]$ and $\vec{\theta}_n$ is the column vector $\left(\theta_n^{(1)}, \theta_n^{(2)}\right)^{\top}$ that characterizes the hedging portfolio.

Definition 2.2.1 *The set of all $\mathcal{G}-$predictable self-financing hedging strategies satisfying possible additional requirements (such as no short-selling constraints [10]) is denoted by Θ. We refer to Θ as the set of* admissible hedging strategies.

Unobservable regimes and discrete-time trading make perfect replication of the European contingent claim impossible to achieve. The hedger therefore aims to best replicate the payoff $\tilde{\phi}(Y_N)$ according to a certain metric. This justifies the use of a penalty function that sanctions departure of the hedging portfolio's terminal value V_N from $\phi(S_N^{(2)})$. Let $g(\cdot)$ be a Borel function $g : \mathbb{R} \to \mathbb{R}$ representing a penalty function. For a given amount of initial wealth v_0, the hedger wishes to find an admissible hedging strategy solving

$$\min_{\theta \in \Theta} \mathbb{E}\left[g(\phi(S_N^{(2)}) - V_N)\right]. \tag{2.1}$$

The solution is referred to as the optimal hedging strategy. Admittedly, g, ϕ, θ and $S^{(2)}$ need to be well-behaved and integrable enough for this expectation to exist.

Defining the hedging problem at the terminal date does not require a pricing function for the derivatives, and in particular a characterization of the risk premium. By contrast, hedging

8. $\theta = \left\{\vec{\theta}_n\right\}_{n=1}^{N}$ is a self-financing hedging strategy if $\forall n \geq 1$, $\vec{\theta}_n^{\top} \vec{S}_n = \vec{\theta}_{n+1}^{\top} \vec{S}_n$.
9. To ease notation, $V_n(v_0, Y_{0:n}, \vec{\theta}_{1:n})$ is denoted by V_n.
10. Or a weaker version of it asking for V_n to be positive.

strategies considering intermediate dates (option tracking) rely on additional assumptions about the martingale measure.

Schweizer (1995) and Rémillard & Rubenthaler (2013) work with the quadratic penality function $g(x) = x^2$. However, this specification entails that gains and losses on the hedge are penalized equally. In practice, the hedger might be interested in treating gains and losses on the hedge differently. Among asymmetric penalty functions, Pham (2000) investigates the case $g(x) = x^p \mathbf{1}_{\{x>0\}}$ for a positive constant p, where $\mathbf{1}_{\{.\}}$ denotes the indicator variable. Another possibility is to choose $g(x) = \mathbf{1}_{\{x \geq z\}}$ where z is a constant. Such a penalty function induces the minimization of the probability that the hedging shortfall is greater than z. Föllmer & Leukert (1999) and Cvitanić & Spivak (1999) study the hedging problem in continuous time with a similar hedging goal. In this paper, we opt for a general asymmetric penalty function of the form

$$g(x) = \alpha_1 |x|^p \mathbf{1}_{\{x \leq \gamma_1\}} + \alpha_2 |x|^q \mathbf{1}_{\{x > \gamma_2\}}, \tag{2.2}$$

for some constants α_1, α_2, γ_1, γ_2, $p \geq 0$ and $q \geq 0$. This specification encompasses both symmetric and asymmetric penalties and allows different penalty weights to be put on the under- and over-replication of the terminal payoff. If $q = \alpha_1 = 0$ and $\alpha_2 = 1$, the penalty reduces to a Value-at-Risk type of measure. If $q = \alpha_2 = 1$ and $\alpha_1 = 0$, the penalty becomes an Expected shortfall type of measure. The case $p = q = 2$, $\alpha_1 = \alpha_2 = 1$ and $\gamma_1 = \gamma_2 = 0$ leads to the quadratic penalty.

2.3 Solving the hedging problem

2.3.1 From path-dependence to the Markov property

The tools of dynamic programming and the Bellman equation are tailor-made to solve problems of the Equation (2.1) type if one can invoke the Markov property for the state variables process. However, the observable process Y does not necessarily have the Markov property with respect to the filtration $\mathcal{G}$, because the cumulative returns depend on the regimes. Indeed, all past values of the cumulative returns path Y give information about the current value of the unobservable regime h. This obstacle is circumvented by defining additional state variables that summarize all the relevant information of Y's previous path. Those variables allow for the definition of a process that has the Markov property with respect to information flow $\mathcal{G}$.

Below, $f_{\vec{X}}(\vec{x})$ denotes the joint probability density function (pdf) of a random vector $\vec{X}$. In some cases, if some components of $\vec{X}$ are discrete-type random variables, $f_{\vec{X}}(\vec{x})$ is a mixed pdf. Similarly, $f_{\vec{X}|\vec{Y}}(\vec{x}|\vec{y})$ denotes the pdf of $\vec{X}$ conditional upon $\vec{Y}=\vec{y}$. All proofs are provided in Appendix 2.7.

Definition 2.3.1 *The conditional probability $\eta_{i,n}$ of being in regime i at time n given the cumulative returns $Y_{0:n}$ is the $\mathcal{G}_n-$measurable function*

$$\eta_{i,n} := \mathbb{P}\left(h_n = i|\mathcal{G}_n\right) = f_{h_n|Y_{0:n}}\left(i|Y_{0:n}\right),\ i \in \mathcal{H}.$$

As a special case, $\eta_{i,0}=\mathbb{P}(h_0=i)=f_{h_0}(i)$. The $\mathcal{G}_n-$measurable vector $\vec{\eta}_n=(\eta_{1,n},...,\eta_{H,n})$ denotes the set of conditional probabilities at time n.

Those η are the state variables required in the construction of a Markov process with respect to filtration $\mathcal{G}$. Theorem 2.3.1 provides a recursion formula allowing for an efficient computation of those probabilities. [11]

Theorem 2.3.1 *The conditional probabilities are given recursively by*

$$\eta_{i,n+1} = \frac{\sum_{j=1}^{H} f_{h_{n+1},Y_{n+1}|h_n,Y_n}\left(i,Y_{n+1}\left|j,Y_n\right.\right)\eta_{j,n}}{\sum_{j=1}^{H}\sum_{\ell=1}^{H} f_{h_{n+1},Y_{n+1}|h_n,Y_n}\left(j,Y_{n+1}\left|\ell,Y_n\right.\right)\eta_{\ell,n}}.$$

Moreover, if Y_{n+1} and h_{n+1} are conditionally independent upon $\mathcal{F}_n$, then

$$f_{h_{n+1},Y_{n+1}|h_n,Y_{0:n}}\left(i,Y_{n+1}\left|j,Y_{0:n}\right.\right) = P_{j,i}^{(n)}(Y_n) f_{Y_{n+1}|h_n,Y_{0:n}}\left(Y_{n+1}\left|j,Y_{0:n}\right.\right).$$

Corollary 2.3.1 states that those conditional probabilities are the natural extension for the cumulative returns to retrieve the Markov property.

Corollary 2.3.1 *$\{Y_n,\vec{\eta}_n\}_{n=0}^{N}$ has the Markov property with respect to $\mathcal{G}$.*

Finally, the next corollary extends the previous one to include the hedging portfolio value. In the general case of predictable hedging strategies, this inclusion unfortunately destroys the Markov property. However, if asset reallocation is solely determined by the information about current cumulative return and portfolio value as well as the recursive conditional probabilities (as defined in Theorem 2.3.1), then the Markov property can be retrieved. This property is crucial, from a numerical point of view, to obtaining an implementable algorithm.

11. An alternative recursion formula is presented in Rémillard et al. (2010). However, the current formula is preferred for two main reasons. First, $\eta_{i,n}$ lying in $[0,1]$ makes it numerically more stable. Second, it benefits from a dimension reduction since $\eta_{H,n}=1-\sum_{j=1}^{H-1}\eta_{j,n}$.

Corollary 2.3.2 *For any admissible hedging strategy* $\theta \in \Theta$, *the conditional distribution of* $(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1})$ *given* $\mathcal{G}_n$ *is the same as if it is conditioned upon* $\sigma\left(Y_n, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}\right)$. *Moreover, if the condition that* $\vec{\theta}_{n+1}$ *is* $\sigma(Y_n, \vec{\eta}_n, V_n)$ *−measurable for any* n *is added, then* $\{Y_n, \vec{\eta}_n, V_n\}_{n=0}^{N}$ *has the Markov property with respect to* $\mathcal{G}$.

2.3.2 A recursive construction

In this section, an optimal hedging strategy is constructed. Let Ψ_N^* be the hedging penalty at time N,

$$\Psi_N^* := g\left(\tilde{\phi}(Y_N) - V_N\right) \tag{2.3}$$

and for any $n \in \{0, 1, ..., N-1\}$, let Ψ_n^* be the smallest possible expected hedging penalty

$$\Psi_n^* := \min_{\vec{\theta}_{n+1:N}} \mathbb{E}\left[\Psi_N^* | \mathcal{G}_n\right] \tag{2.4}$$

where $\vec{\theta}_{n:N} = \left(\vec{\theta}_n, ..., \vec{\theta}_N\right)$.

Remark 2.3.1 *One assumes sufficient regularity in* g, ϕ *and the distribution of* $\{Y_n\}_{n=0}^{N}$ *such that, for all* n, *the minimum in (2.4) is attained.*

Equation (2.4) is stated as a minimization over $N-n$ portfolio vectors. Theorem 2.3.2 presents a way to optimize these portfolios one at a time.

Theorem 2.3.2 *For any* $n \in \{0, 1, ..., N-1\}$, *the smallest expected penalty at time* n *may be computed using a recursive argument :*

$$\Psi_n^* = \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}^* \middle| \mathcal{G}_n\right]. \tag{2.5}$$

Furthermore, let $\vec{\theta}^*_{(n+2):N}$ *denote one of the possible admissible hedging strategies that minimize the expected penalty at time* $n+1$, *that is,*

$$\vec{\theta}^*_{(n+2):N} = \arg\min_{\vec{\theta}_{n+2:N}} \mathbb{E}[\Psi_N^* | \mathcal{G}_{n+1}].$$

Then,

$$\vec{\theta}^*_{(n+1):N} := \left(\arg\min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}^* \middle| \mathcal{G}_n\right], \vec{\theta}^*_{(n+2):N}\right), \tag{2.6}$$

is a solution to the following equation :

$$\vec{\theta}^*_{(n+1):N} = \underset{\vec{\theta}_{n+1:N}}{\arg\min}\, \mathbb{E}[\Psi^*_N | \mathcal{G}_n].$$

This means that the optimal admissible hedging strategy may be built up using a backward induction construction.

Equations (2.5) and (2.6) involve conditional expectations with respect to all past return realizations. Theorem 2.3.3 shows that it is possible to remove path-dependence and appeal only to conditional expectations with respect to the current state variables $\{Y_n, \vec{\eta}_n, V_n\}_{n=0}^N$.

Theorem 2.3.3 *Assume that for all n, constraints on the portfolio $\vec{\theta}_{n+1}$ depend only on the value of $(Y_n, \vec{\eta}_n, V_n)$. Then, $\forall n \leq N$, Ψ^*_n is $\sigma(Y_n, \vec{\eta}_n, V_n)-$measurable. Moreover, there exists an optimal self-financing hedging strategy $\left\{\vec{\theta}^*_n\right\}$ that solves (2.1) such that $\forall n \geq 1$, $\vec{\theta}^*_{n+1}$ is $\sigma(Y_n, \vec{\eta}_n, V_n)-$measurable. Furthermore,*

$$\vec{\theta}^*_{n+1} = \underset{\vec{\theta}_{n+1} \in \sigma(Y_n, \vec{\eta}_n, V_n)}{\arg\min} \mathbb{E}\left[\Psi^*_{n+1} | Y_n, \vec{\eta}_n, V_n\right]. \tag{2.7}$$

Since Ψ^*_n is $\sigma(Y_n, \vec{\eta}_n, V_n)-$measurable, one can write $\Psi^*_n = \Psi_n(Y_n, \vec{\eta}_n, V_n)$. Finally, the next theorem combines Theorems 2.3.2 and 2.3.3 to optimize one portfolio vector at a time, searching on the space of hedging strategies for which $\{Y_n, \vec{\eta}_n, V_n\}_{n=0}^N$ has the Markov property with respect to $\mathcal{G}$. These two features make the algorithm numerically tractable.

Theorem 2.3.4 ***The Bellman Equation*** *There exists a self-financing hedging strategy $\left\{\vec{\theta}^*_n\right\}$ that solves problem (2.1) and the following set of recursive equations :*

$$\forall n,\ \vec{\theta}^*_{n+1} = \underset{\vec{\theta}_{n+1} \in \sigma(Y_n, \vec{\eta}_n, V_n)}{\arg\min} \mathbb{E}\left[\Psi_{n+1}(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}(\vec{\theta}_{n+1}))\middle| Y_n, \vec{\eta}_n, V_n\right].$$

Furthermore, the minimal expected penalty can be computed as follows :

$$\Psi_N(Y_N, \vec{\eta}_N, V_N) = g(\phi(S^{(2)}_N) - V_N) = g(\tilde{\phi}(Y_N) - V_N) \tag{2.8}$$

$$\Psi_n(Y_n, \vec{\eta}_n, V_n) = \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}(\vec{\theta}_{n+1}))\middle| Y_n, \vec{\eta}_n, V_n\right],\ n \in \{0, 1, ..., N-1\} \tag{2.9}$$

Finally, $\displaystyle\min_{\{\vec{\theta}_n\} \in \Theta} \mathbb{E}\left[g(\phi(S^{(2)}_N) - V_N(\vec{\theta}_{1:N}))\right] = \Psi_0(Y_0, \vec{\eta}_0, V_0).$

The proof of Theorem 2.3.4 is a direct consequence of Theorems 2.3.2 and 2.3.3 and the definition of Ψ_n.

2.4 Lattice implementation

Analytical solutions to Theorem 2.3.4's equations are unlikely to be found for general penalties. Therefore, numerical approximations must be considered in order to implement the algorithm. The numerical application of the hedging algorithm is discussed in this section.

2.4.1 Dimensionality reduction

Since $\sum_j^H \eta_{j,n} = 1$, the variable $\eta_{H,n}$ provides no additional information. Therefore, $\vec{\eta}_n = (\eta_{1,n}, ..., \eta_{H,n})$ can be replaced with $\vec{\eta}_n := (\eta_{1,n}, ..., \eta_{H-1,n})$ in Theorem 2.3.4. This reduces the dimension of the problem, which is an important numerical issue. Similarly, since for self-financing strategies $\sum_{k=1}^2 \theta_{n+1}^{(k)} S_n^{(k)} = V_n$, the optimization over $\vec{\theta}_{n+1}$ is in fact equivalent to optimizing only over $\theta_{n+1} := \theta_{n+1}^{(2)}$.

2.4.2 Grid values

To compute the minimal expected penalty Ψ_n and optimal portfolio position $\vec{\theta}_{n+1}$ from Theorem 2.3.4, one resorts to a grid whose nodes correspond to a discrete subsample of all possible values of (Y_n, η_n, V_n). For each state variable, the largest and smallest values in the grid must be set. One can use the $[0, 1]$ bounds for $\vec{\eta}$ since it contains probabilities. Variables V_n and Y_n are unbounded. Therefore, grid bounds for V_n and Y_n are found numerically using a Monte-Carlo simulation. To this end, 10^5 sample paths of cumulative returns $Y_{0:N}$ are simulated. This yields the approximate distribution of Y_n for all n. The case of the portfolio value V_n is different since the optimal hedging strategy is not yet known. However, a proxy $V_n^{(BS)}$ is built for V_n using the Black-Scholes delta hedging as described in Section 2.5.4. Let $Y_{n,\alpha}$, and $V_{n,\alpha}^{(BS)}$ be the α^{th} sample quantiles. Define

$$Y_{n,mid} := \frac{1}{2}(Y_{n,0.25\%} + Y_{n,99.75\%}) \text{ and } V_{n,mid}^{(BS)} := \frac{1}{2}(V_{n,0.25\%}^{(BS)} + V_{n,99.75\%}^{(BS)})$$

as the mid-points of two extreme quantiles. For some positive stretching factors ($\lambda_Y^{(small)}$, $\lambda_Y^{(large)}, \lambda_V^{(small)}, \lambda_V^{(large)}$), the largest and smallest values for the grid at time n are set

$$\begin{aligned}
Y_n^{(small)} &:= \quad (1 + \lambda_Y^{(small)})(Y_{n,0.25\%} - Y_{n,mid}) + Y_{n,mid} \\
Y_n^{(large)} &:= \quad (1 + \lambda_Y^{(large)})(Y_{n,99.75\%} - Y_{n,mid}) + Y_{n,mid} \\
V_n^{(small)} &:= \quad (1 + \lambda_V^{(small)})(V_{n,0.25\%}^{(BS)} - V_{n,mid}^{(BS)}) + V_{n,mid}^{(BS)} \\
V_n^{(large)} &:= \quad (1 + \lambda_V^{(large)})(V_{n,99.75\%}^{(BS)} - V_{n,mid}^{(BS)}) + V_{n,mid}^{(BS)}.
\end{aligned}$$

2.4.3 Algorithm solving the Bellman equation

A numerical algorithm allowing for the computation of the minimal expected penalty and the optimal portfolio position at each time step is given in this section. First, define two grids of different sizes (one finer and one coarser) containing a discrete subset of values for $(Y_n, \vec{\eta}_n, V_n)$.

On the coarse grid

Assume that $(Y_n, \vec{\eta}_n, V_n) = (y, \vec{\eta}, v)$. According to Theorem 2.3.4, the goal is to evaluate Equation (2.9) at each node $(y, \vec{\eta}, v)$ of the grid :

$$\Psi_n^{y,\vec{\eta},v} = \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}\left(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}(\vec{\theta}_{n+1})\right) \Big| (Y_n, \vec{\eta}_n, V_n) = (y, \vec{\eta}, v)\right].$$

From Theorem 2.3.1, $\vec{\eta}_{n+1}$ is a function of $(Y_{n+1}, Y_n, \vec{\eta}_n)$. Seen from node $(y, \vec{\eta}, v)$, it may be denoted $\vec{\eta}_{n+1}^{y,\vec{\eta}}(Y_{n+1})$. Because the amount invested in the riskless asset is the value of the portfolio minus the investment in the risky asset, the time-$(n+1)$ value of the hedging portfolio, seen from the grid point $(Y_n, \vec{\eta}_n, V_n) = (y, \vec{\eta}, v)$, is

$$\begin{aligned} V_{n+1}\left(\vec{\theta}_{n+1}\right) &= \theta_{n+1}^{(1)} \exp\left(r\,(n+1)\,\Delta_t\right) + \theta_{n+1}^{(2)} S_0^{(2)} \exp\left(Y_{n+1}\right) \\ &= \exp\left(r\Delta_t\right)\left(v - \theta_{n+1}^{(2)} S_0^{(2)} \exp\left(y\right)\right) + \theta_{n+1}^{(2)} S_0^{(2)} \exp\left(Y_{n+1}\right) \\ &= V_{n+1}^{y,v}\left(\vec{\theta}_{n+1}, Y_{n+1}\right). \end{aligned}$$

Therefore, the expected penalty at time n and at grid point $(y, \vec{\eta}, v)$ satisfies

$$\begin{aligned} \Psi_n^{y,\vec{\eta},v} &= \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}\left(Y_{n+1}, \vec{\eta}_{n+1}^{y,\vec{\eta}}(Y_{n+1}), V_{n+1}^{y,v}\left(\vec{\theta}_{n+1}, Y_{n+1}\right)\right)\Big| (Y_n, \vec{\eta}_n, V_n) = (y, \vec{\eta}, v)\right] \\ &= \min_{\vec{\theta}_{n+1}} \sum_{j=1}^{H} \eta_{j,n} \mathbb{E}\left[\Psi_{n+1}\left(Y_{n+1}, \vec{\eta}_{n+1}^{y,\vec{\eta}}(Y_{n+1}), V_{n+1}^{y,v}(\vec{\theta}_{n+1}, Y_{n+1})\right)\Big| (h_n, Y_n, \vec{\eta}_n, V_n) = (j, y, \vec{\eta}, v)\right] \\ &\quad \text{(from Equation (2.20))} \\ &= \min_{\vec{\theta}_{n+1}} \sum_{j=1}^{H} \eta_{j,n} \int_{-\infty}^{\infty} \Psi_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z\right) f_{Y_{n+1}|Y_n,\vec{\eta}_n,V_n,h_n}\left(z|y, \vec{\eta}, v, j\right) dz \\ &= \min_{\vec{\theta}_{n+1}} \sum_{j=1}^{H} \eta_{j,n} \int_{-\infty}^{\infty} \Psi_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z\right) f_{Y_{n+1}|Y_n,h_n}(z|y,j) dz \;\; \text{(Markov property and Lemma 2.7.1)} \end{aligned}$$

where

$$\Psi_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z\right) = \Psi_{n+1}\left(z, \vec{\eta}_{n+1}^{y,\vec{\eta}}(z), V_{n+1}^{y,v}\left(\vec{\theta}_{n+1}, z\right)\right).$$

In general, there is no closed-form solution for this integral and it is evaluated numerically. Therefore, the support of Y_{n+1} is partioned in M intervals with boundaries $-\infty = z_0 < z_1 <$

... $< z_{M-1} < z_M = \infty$ and $z_i^* \in [z_{i-1}, z_i]$ acts as a representative of the interval $[z_{i-1}, z_i]$. Let $\omega_i^{y,j,n}$ be the following quadrature weights :

$$\omega_i^{y,j,n} = F_{Y_{n+1}|Y_n,h_n}(z_i\,|y,j) - F_{Y_{n+1}|Y_n,h_n}(z_{i-1}\,|y,j),$$

$F_{Y_{n+1}|Y_n,h_n}$ being the cumulative distribution function of Y_{n+1} given (Y_n, h_n). Then,

$$\begin{aligned}
\int_{-\infty}^{\infty} \Psi_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z\right) f_{Y_{n+1}|Y_n,h_n}(z\,|y,j)\,dz &= \sum_{i=1}^{M} \int_{z_{i-1}}^{z_i} \Psi_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z\right) f_{Y_{n+1}|Y_n,h_n}(z\,|y,j)\,dz \\
&\cong \sum_{i=1}^{M} \Psi_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z_i^*\right) \int_{z_{i-1}}^{z_i} f_{Y_{n+1}|Y_n,h_n}(z\,|y,j)\,dz \\
&= \sum_{i=1}^{M} \Psi_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z_i^*\right) \omega_i^{y,j,n} \qquad (2.10)
\end{aligned}$$

In general, the approximation (2.10) is good if the distances between the z_i are small and the Ψ_{n+1} function is relatively smooth. The z_i are chosen to be quantiles of the conditional distribution $F_{Y_{n+1}|Y_n,h_n}$. To better capture the impact of extreme events, particular attention is paid to the tails of the distribution. The left (right) tail is defined as the smallest (largest) 5% values of the distribution. The $M_{(1)}$ smallest z_i's correspond to quantiles of level $k\frac{5\%}{M_{(1)}}$, $k \in \{1, 2, ..., M_{(1)}\}$. The central part of the distribution is proxied by $M_{(2)}$ quantiles of level $k\frac{90\%}{M_{(2)}} + 5\%$, $k \in \{1, 2, ..., M_{(2)}\}$, while the right tail is represented by $M_{(3)}$ quantiles whose level lies in $]95\%, 100\%]$. Consequently, the weights $\omega_i^{y,j,n}$ are $\frac{5\%}{M_{(1)}}$, $\frac{90\%}{M_{(2)}}$ or $\frac{5\%}{M_{(3)}}$ depending on which part of the distribution z_i belongs to. Among possible specifications, z_i^* are chosen as quantiles whose level is the mean between the levels of z_{i-1} and z_i.

Because the maximization is time-consuming, especially if it must be done at all nodes of the lattice, the research area is reduced to a discrete set $\mathcal{O}$ of values :

$$\Psi_n(y, \vec{\eta}, v) \cong \min_{\vec{\theta}_{n+1} \in \mathcal{O}} \sum_{i=1}^{M} \Psi_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z_i^*\right) \omega_i^{y,\vec{\eta},n}. \qquad (2.11)$$

Since the backward induction on time leads to a numerical approximation $\widehat{\Psi}_{n+1}$ of Ψ_{n+1}, the latter is replaced by former in Equation (2.11) in applications.

<u>Step 1 : Rough estimate of optimal hedging strategy</u>

A rough estimate of the optimal hedging strategy is

$$\hat{\theta}_{n+1}^{y,\vec{\eta},v} = \operatorname*{arg\,min}_{\theta_{n+1} \in \mathcal{O}} \sum_{i=1}^{M} \widehat{\Psi}_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z_i^*\right) \omega_i^{y,\vec{\eta},n}.$$

By construction, the z_i^* do not match the grid's discretization of next period return Y_{n+1}. For this reason, interpolation is required to evaluate each of the $\widehat{\Psi}_{n+1}^{y,\vec{\eta},v}\left(\vec{\theta}_{n+1}, z_i^*\right)$ whose arguments most likely lie between the grid nodes. This step proceeds with multivariate linear interpolation. [12]

On the finer grid

Step 2 : Smoothing of the hedging strategy

From step 1, one gets an approximate portfolio position $\hat{\theta}_{n+1}^*$ for every node of the coarse grid at time n. For every value of $(y, \vec{\eta}, v)$ on the finer grid, one computes the hedging portfolio position $\hat{\vartheta}_{n+1}^{y,\vec{\eta},v}$ using smoothing splines based on $\hat{\theta}_{n+1}$. $\hat{\vartheta}_{n+1}$ is now used as the final estimation of the optimal hedging portfolio position.

Step 3 : Recalculation of the value function

A finer partition of the distribution of Y_{n+1} and the corresponding weights, denoted by $\tilde{z}_i^*$ and $\tilde{\omega}_i$, $i = 1, ..., \tilde{M}$, serve for the approximation of the minimal expected penalty function with the new portfolio position $\hat{\vartheta}_{n+1}$. Thus, mimicking Equation (2.11),

$$\widehat{\Psi}_n(y, \vec{\eta}, v) = \sum_{i=1}^{\tilde{M}} \widehat{\Psi}_{n+1}^{y,\vec{\eta},v}\left(\hat{\vartheta}_{n+1}^{y,\vec{\eta},v}, \tilde{z}_i^*\right) \tilde{\omega}_i^{y,\vec{\eta},n}.$$

The subsequent iteration of the three-step algorithm will call this new approximation for $\hat{\Psi}_n$. Thus, to minimize the accumulation of errors, the interpolation is performed with natural splines. [13]

2.5 Numerical results

2.5.1 The model

As in Hamilton (1989), the regime process is assumed to be a Markov chain, implying that the conditional distribution of h_{n+1} given $\mathcal{F}_n$ is the same as if it were conditioned upon h_n. The model can accomodate a regime shift timeframe which is coarser than the rebalancing schedule. In that context, τ represents the number of periods between two possible

12. This approximation of $\widehat{\Psi}_{n+1}$ is not involved in further iterations. Therefore, while high precision is not a crucial issue at this step, computational speed is.

13. Natural splines in three dimensions are implemented through the *interp3* matlab function.

regime transitions and $\{h_n\}_{n=0}^N$ becomes a time-inhomogeneous Markov chain with probability transition matrix

$$P^{(n)}(y) = \begin{cases} P \text{ if } (n+1) \bmod \tau = 0 \\ I_{H\times H} \text{ otherwise,} \end{cases}$$

where $I_{H\times H}$ is the identity matrix.

A basic model based on two regimes ($H = 2$) serves as benchmark to test the proposed algorithm. Conditioned on the actual regime $h_n = i$, the one-period log-return $\epsilon_{n+1} = Y_{n+1} - Y_n$ has a Gaussian distribution with mean $\mu_i \Delta_t$ and variance $\sigma_i^2 \Delta_t$.

The application of Theorem 2.3.4 relies on the following relations :

$$\begin{aligned} Y_{n+1} &= Y_n + \epsilon_{n+1} \\ V_{n+1} &= V_n e^{r\Delta_t} + \theta_{n+1} S_0 e^{Y_n}(e^{\epsilon_{n+1}} - e^{r\Delta_t}) \\ \eta_{1,n+1} &= \frac{\sum_{j=1}^H \; P_{j,1}^{(n)} \eta_{j,n} f_{\epsilon_{n+1}|h_n}(\epsilon_{n+1}|j)}{\sum_{u=1}^H \sum_{j=1}^H \; P_{j,u}^{(n)} \eta_{j,n} f_{\epsilon_{n+1}|h_n}(\epsilon_{n+1}|j)}, \end{aligned}$$

where $\eta_{2,n} = 1 - \eta_{1,n}$ and $f_{\epsilon_{n+1}|h_n}(\epsilon_{n+1}|j)$ is the Gaussian density function

$$f_{\epsilon_{n+1}|h_n}(\epsilon_{n+1}|j) = \frac{1}{\sqrt{2\pi\Delta_t}\sigma_j} \exp\left(-\frac{1}{2}\frac{(\epsilon_{n+1} - \mu_j \Delta_t)^2}{\sigma^{(j)2}\Delta_t}\right). \tag{2.12}$$

The conditional distribution of $\epsilon_{n+1}|(Y_n, \eta_n, V_n)$ is a mixture of two Gaussian distributions :

$$\begin{aligned} \mathbb{P}(\epsilon_{n+1} \le x|Y_n, \eta_n, V_n) &= \mathbb{P}(\epsilon_{n+1} \le x|h_n = 1)\eta_{1.n} + \mathbb{P}(\epsilon_{n+1} \le x|h_n = 2)(1 - \eta_{1,n}) \\ &= \Phi\left(\frac{x - \mu_1}{\sigma_1}\right)\eta_{1,n} + \Phi\left(\frac{x - \mu_2}{\sigma_2}\right)(1 - \eta_{1,n}), \end{aligned}$$

where Φ is the standard normal cumulative distribution function.

Moreover, the following boundaries can be used for η in the algorithm of Section 2.4 :

Proposition 2.5.1 *For all* j, n, $\min\limits_{i\in\mathcal{H}} P_{i,j}^{(n)} \le \eta_{j,n+1} \le \max\limits_{i\in\mathcal{H}} P_{i,j}^{(n)}$.

The proof is in Appendix 2.7.

2.5.2 Estimation

Regime switches potentially occur each week and rebalancing is performed weekly ($\Delta_t = 1/52, \tau = 1$) or daily ($\Delta_t = 1/260, \tau = 5$). Maximum likelihood with the Baum-Welch algorithm (Baum et al., 1970), a particular case of the EM algorithm of Dempster et al. (1997), is applied to a sample of S&P 500 weekly log-returns from January 1, 2000 to December 31, 2010. Parameter estimates are reported in Table 2.1.

A p-value of 34.4% for the Cramer-Von-Mises parametric bootstrap goodness-of-fit test (see Genest & Rémillard, 2008) for the regime-switching process indicates that the model is not rejected. The first (second) regime represents an economy in expansion (recession) : returns exhibit a positive (negative) mean with a low (high) volatility. The risk-free rate is set to $r = 2\%$.

TABLE 2.1 – Estimated parameters of the Gaussian regime switching model

Parameter	Regime 1	Regime 2
μ_j	.0718	$-.2884$
σ_j	.1283	.3349
$P_{j,j}$	.9736	.9091

2.5.3 Hedging strategies

The option to be hedged is a European at-the-money call option with payoff $\phi(S_N) = \max(0, S_N - E)$. The initial index value is $S_0 = 1{,}257.64$, which is the value of the S&P 500 on December 31, 2010. The option strike is $E = 1{,}257$. The maturity of the option is 12 weeks.[14]

The initial probability of being in regime 1 is set to $\eta_0 = 0.2318$. This value is chosen instead of the estimated value on the S&P 500 time series because it leads to the same call option price under the Black-Scholes and Hardy models (see Sections 2.5.4 and 2.5.4, respectively). Thus, both these hedging methodologies use the same initial capital, which makes the numerical results comparable. The initial hedging capital, which is the option price under those models, is $V_0 = 62.4316$.

14. That is, $N = 60$ periods for daily rebalancing and $N = 12$ for weekly rebalancing.

The following penalty functions are under consideration :

$$
\begin{aligned}
g(x) &= x^2 && \text{quadratic,} && (2.13)\\
g(x) &= x^2\mathbf{1}_{\{x>0\}} && \text{short quadratic,} && (2.14)\\
g(x) &= x^2\mathbf{1}_{\{x<0\}} && \text{long quadratic,} && (2.15)
\end{aligned}
$$

where x represents the hedging error $\tilde{\phi}(Y_N)-V_N$. The quadratic penalty sanctions departures from the option payoff. The short (long) quadratic penalty is designed for the option seller (buyer), since it does not penalize profits ; only losses are sanctioned.

The restrictions considered on the portfolio positions are that $\forall n$, $\theta_n \in [0,1]$, thereby preventing short sales and excessive leverage.

2.5.4 Benchmarks

In order to compare the hedging model presented in this paper, benchmarks must be set. In the following, the optimal hedging strategy presented in Section 2.3 is referred to as "minimal expected penalty hedging" (MEPH).

The most common hedging strategy relies on delta hedging. [15] In this case, a pricing kernel is required to compute the deltas. The first two benchmarks examine two pricing models.

Black-Scholes delta hedging (BSDH)

The classic Black-Scholes delta with a modified volatility determines the position held in the underlying asset : [16]

$$\theta_{n+1}^{(\mathrm{BSDH})} = \Phi\left(\frac{\log(S_n/E) + (r + .5\zeta^2)\Delta_t(N-n)}{\zeta\sqrt{\Delta_t(N-n)}}\right),$$

where ζ is the asymptotic stationary volatility of log-returns ϵ_n in the case $\tau = 1$:

$$\zeta = \sqrt{\left(\sum_{j=1}^{H} P_j^*(\sigma^{(j)2} + \mu^{(j)2})\right) + \left(\sum_{j=1}^{H} P_j^*\mu^{(j)}\right)^2}, \qquad (2.16)$$

15. The empirical performance of delta hedging under different option pricing models is investigated by De Giovanni et al. (2008).

16. Black-Scholes delta hedging under model misspecification is studied in Augustyniak & Boudreault (2012).

P^* is the stationary distribution associated with the transition matrix P. In the case $\tau > 1$, the stationary distribution for the regimes does not exist in general because of the cyclical nature of the Markov chain transition probabilities. Nevertheless, Equation (2.16) is used as the presumed market volatility. The characterization of the hedging position is explicit and does not require a lattice approximation.

The initial capital used for hedging is the option price given by the Black-Scholes formula with the volatility given by (2.16). The Black-Scholes hedging methodology can be seen as a naive benchmark that would be applied by a hedger who ignores the presence of regimes in the market.

Hardy delta hedging (HDH)

In Hardy (2001)'s two-regime model, the risk-neutral dynamics of one-period log-returns ϵ_{n+1} follow a mixture of Gaussian distributions. The delta hedging strategy commands that :

$$\theta_{n+1}^{(HDH)} = \sum_{j=1}^{2} \eta_{j,n}^{\mathbb{Q}} \sum_{R=0}^{N-n} \Phi\left(\frac{\log(S_n/E) + (N-n)r\Delta_t + (R\sigma_1^2/2 + (N-n-R)\sigma_2^2/2)\Delta_t)}{\sqrt{(R\sigma_1^2 + (N-n-R)\sigma_2^2)\Delta_t}}\right) f_n^j(R),$$

where $f_n^j(R)$ is the probability, given current regime i, that the number of periods between times n and N spent in the first regime is R. Probabilities $f_n^j(R)$ can be computed recursively (see Hardy, 2001). $\eta_{j,n}^{\mathbb{Q}}$ is the risk-neutral probability of being in regime j during time interval $]n, n+1]$ given the current partial information $\mathcal{G}_n$. With this benchmark, the initial capital used for hedging is the option price. The hedger acknowledges the existence of regimes, but assigns an arbitrary risk premium to price options.

Forecast regime quadratic hedging (FRQH)

Besides delta hedging, Rémillard & Rubenthaler (2013) propose a global hedging risk minimization approach. The hedging strategy θ minimizes the expected terminal squared error of hedging with respect to complete information $\mathcal{F}$. This implies perfect knowledge of the current and all past regimes. Since in practice the states h are not observable, Rémillard et al. (2010) forecast them with the most likely regime.

Let $\bar{\Theta}$ be the set of all $\mathcal{F}$-predictable self-financing strategies.[17] The FRQH strategy solves

$$\min_{\theta \in \bar{\Theta}} \mathbb{E}\left[(\phi(S_N) - V_N)^2\right].$$

17. By contrast, the MEPH strategy is $\mathcal{G}$-predictable.

With this benchmark, the hedging problem is based on the terminal date. Therefore, no assumption related to the risk premium is needed, which implies in particular that this strategy works with any initial capital. However, it comes at the price of using a lattice approach to compute the strategy. The hedger acknowledges the existence of regimes. However, the hedging objective is restricted to the quadratic penalty. Furthermore, the uncertainty surrounding regime forecasts is not taken into account.

2.5.5 Lattice parameters

The grid's stretching factors are $(\lambda_Y^{(small)}, \lambda_Y^{(large)}, \lambda_V^{(small)}, \lambda_V^{(large)}) = (.6, .6, 1, 1)$. For step 1 of the algorithm in Section 2.4.3, $M_{(1)} = M_{(3)} = 100$ and $M_{(2)} = 200$. For step 3, $\tilde{M}_{(1)} = \tilde{M}_{(3)} = 200$ and $\tilde{M}_{(2)} = 300$. Putting more points near the tails is used to better capture the extreme events which contribute more heavily to the hedging penalty. The discrete set $\mathcal{O}$ over which the θ_n are optimized in step 1 of the algorithm is $\mathcal{O} = \{j/99 \,|\, j = 0, ..., 99\}$.

The number of grid nodes for each variable on the finer grid (step 3) is :

$$(\#Y_n, \#\eta_n, \#V_n) = \begin{cases} (200, 100, 200) \text{ if } n = N-1 \\ (150, 100, 150) \text{ otherwise} \end{cases} \tag{2.17}$$

More nodes are put on the first step of the recursion as it can be computed faster because of explicit formulas.[18] For the coarse grid in step 1, only a subset of the nodes of the finer grid in step 3 are retained. The proportion of nodes kept in the coarse grid from the finer grid across dimensions Y_n, η_n and V_n is 1/3, 1/3 and 1/4.

2.5.6 A simulation study

The numerical efficiency of the current paper's hedging algorithm is validated by means of Monte-Carlo simulations. The MEPH and FQRH strategies are implemented through a lattice. Hedging errors $\tilde{\phi}(Y_N) - V_N$ and hedging penalties $g(\tilde{\phi}(Y_N) - V_N)$ are computed for $I = 10^6$ simulated paths of the underlying returns.

Table 2.2 reports estimates of the expected penalty and their standard error for each hedging methodology. Note that only the MEPH strategy is affected by the choice of penalty function. For the three benchmarks, the hedging strategy remains the same, but the

18. An explicit expression for Ψ_{N-1} exists for the quadratic penalty. For the short (long) quadratic penalty, an explicit expression for $\mathbb{E}[\Psi_N^*(\theta_N)|\mathcal{G}_{N-1}]$ also exists. Details are available on request.

calculated penalty differs.

A first observation is that the MEPH grid estimate is relatively close to the simulated expected penalty. This confirms the accuracy of the numerical implementation.

TABLE 2.2 – Estimated expected penalties

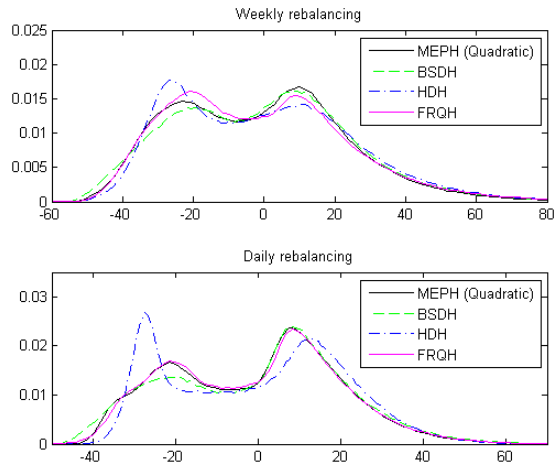
	MEPH	BSDH	HDH	FRQH	MEPH	BSDH	HDH	FRQH
Rebalacing frequency	weekly	weekly	weekly	weekly	daily	daily	daily	daily
Quadratic Penalty								
Grid Estimate	622.57	-	-	-	417.88	-	-	-
Expected Penalty	622.70	662.13	669.10	664.87	418.77	457.16	442.28	422.57
Standard Error	1.010	1.082	1.051	1.148	0.5804	0.5605	0.4555	0.5191
Short Quadratic Penalty								
Grid Estimate	325.15	-	-	-	191.01	-	-	-
Expected Penalty	325.39	372.16	391.15	374.75	193.71	222.15	220.21	202.90
Standard Error	0.9493	1.065	1.064	1.149	0.5321	0.4980	0.4343	0.4726
Long Quadratic Penalty								
Grid Estimate	268.11	-	-	-	214.87	-	-	-
Expected Penalty	267.35	289.97	277.95	290.12	211.62	235.01	222.07	219.66
Standard Error	0.4365	0.5043	0.4346	0.4640	0.3577	0.4129	0.3416	0.3678

In all six cases considered, the MEPH strategy significantly reduces the expected penalty. The magnitude of the penalty dispersion is comparable across all hedging strategies.

As for the quadratic penalty, the MEPH reduces the expected penalty by 6.0% in the weekly case and by 0.9% in the daily case with respect to the best benchmark, namely BSDH for weekly and FRQH for daily. The short (long) quadratic penalty is specifically designed for the call option seller (buyer). The MEPH reduces the expected penalty by 12.6% (3.8%) in the weekly case and by 4.5% (3.7%) in the daily case with respect to the best benchmark. The latter differs across penalties and rebalancing frequencies. For the weekly case, the second best strategy is HDH for the long quadratic penalty and BSDH otherwise. In the daily case, as regime forecasts are more accurate, the FRQH method performs better than the other two benchmarks.

Hedging errors drive hedging penalties and are therefore worthy of investigation. However, descriptive statistics about hedging errors should not be the sole basis on which to judge the performance of hedging strategies. Nevertheless, analyzing those quantities sheds light on how the penalty performance is achieved. Figure 2.1 displays the hedging error distributions for the quadratic MEPH and the three benchmarks. All distributions exhibit bimodality with similar mode locations but different frequencies.

FIGURE 2.1 – Density plot of hedging errors for MEPH versus benchmarks



The distribution behaviour, especially in the tails, is better described by Tables 2.3 and 2.4. In terms of RMSE, the quadratic MEPH strategy slightly dominates all other benchmarks for both weekly and daily rebalancing. This is consistent with the quadratic objective of reducing the occurrence of large deviations of the hedging portfolio from the derivative. As far as Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) are concerned, the picture is not as clear. The short quadratic MEPH used by a call seller performs slightly better than the other hedging strategies.[19] For the call option buyer, MEPH is the second best behind HDH for both VaR and TVaR risk measures.

Hedgers using the short quadratic penalty perceive the left tail as positive outcomes. Therefore, it is not surprising that left tail VaRs and TVaRs are smaller for the HDH than for the short quadratic MEPH. Furthermore the quadratic MEPH aims at shrinking both tails, but it does not imply that it should beat the HDH on the two tails simultaneously. The right tail is much slimmer for the quadratic MEPH than for the HDH. Therefore the HDH can have a slimmer left tail on the extremes and still be beaten by the quadratic MEPH when the symmetric penalty is considered.

More interestingly, the HDH also displays a lower TVaR than the MEPH with long

19. The 95th and 99th percentiles and TVaR are smaller than those of all benchmarks (except for the 99% TVaR of the HDH with daily rebalancing).

quadratic penalty (although the difference is rather small). This may sound surprising since, by construction, the MEPH with long quadratic penalty aims at reducing the left tail of the hedging error. However, inspection of Figure 2.1 reveals that this phenomenon is due to the presence of a hump in the left tail of the HDH hedging errors. Although this hump represents negative outcomes for the hedger using the long quadratic MEPH, it is not located in the extreme part of the left tail and is therefore not taken into account by the VaR and TVaR measures.

TABLE 2.3 – Descriptive statistics for hedging errors (weekly rebalancing)

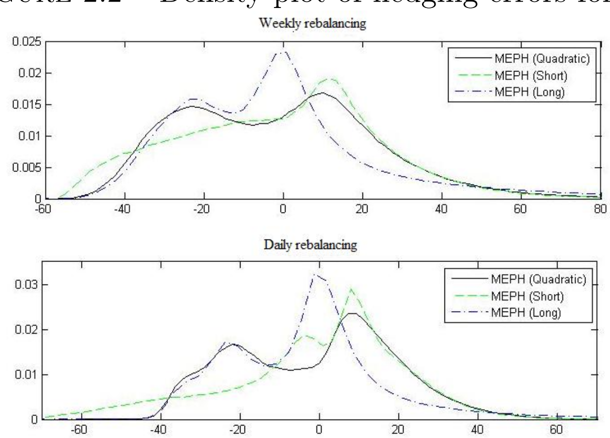
	MEPH (Quadratic)	MEPH (Long)	MEPH (Short)	BSDH	HDH	FRQH
Mean	−0.6787	0.2152	−0.7834	0.1199	0.1492	−0.7741
Standard Deviation	24.945	31.543	25.686	25.732	25.867	25.774
RMSE	**24.954**	31.544	25.698	25.732	25.867	25.785
Skewness	0.5824	2.3438	0.2844	0.5479	0.6516	0.7566
Excess Kurtosis	0.6937	9.7892	0.4501	0.6640	0.4505	1.0720
99th Percentile	67.595	128.67	**65.117**	70.887	71.283	72.869
95th Percentile	41.431	58.540	**40.272**	43.762	45.252	44.219
Median	−0.9649	−3.7258	0.7948	−0.0248	−1.4879	−2.7079
5th Percentile	−36.555	−35.124	−42.559	−38.133	**−34.734**	−36.428
1st Percentile	−43.988	−42.834	−50.357	−46.463	**−41.938**	−43.614
Upper TVaR 99%	84.481	163.48	**81.574**	87.608	86.821	91.001
Upper TVaR 95%	57.768	98.176	**55.795**	60.591	61.370	62.128
Lower TVaR 5%	−41.047	−39.790	−47.352	−43.183	**−39.056**	−40.840
Lower TVaR 1%	−46.788	−45.883	−52.414	−49.051	**−44.941**	−46.575

If a specific risk measure is the ultimate objective, the penalty function should be designed accordingly. Indeed, our methodology precisely permits to adapt the hedging strategy to the desired performance criterion. To illustrate this flexibility, Figure 2.2 shows the effect of the penalty choice on the MEPH hedging error distribution.

TABLE 2.4 – Descriptive statistics for hedging errors (daily rebalancing)

	MEPH (Quadratic)	MEPH (Long)	MEPH (Short)	BSDH	HDH	FRQH
Mean	−0.6067	0.7418	−0.8713	−0.0877	0.0025	−0.6306
Standard Deviation	20.455	31.951	22.800	21.381	21.031	20.547
RMSE	**20.464**	31.960	22.816	21.381	21.031	20.557
Skewness	0.1459	3.5405	−0.5902	0.0477	0.0771	0.1436
Excess Kurtosis	−0.0621	20.905	0.8132	−0.4963	−0.9394	−0.4742
99th Percentile	46.337	139.87	**44.619**	49.132	45.309	47.335
95th Percentile	31.617	50.569	**31.011**	33.540	32.554	32.138
Median	2.1453	−2.0499	3.3012	3.1464	2.5206	1.6114
5th Percentile	−32.851	−32.322	−45.839	−34.544	**−30.522**	−32.475
1st Percentile	−37.966	−37.809	−62.017	−41.267	**−34.508**	−38.363
Upper TVaR 99%	55.126	195.84	53.215	57.680	**52.190**	55.742
Upper TVaR 95%	40.850	105.89	**39.585**	43.142	40.413	41.529
Lower TVaR 5%	−36.063	−35.801	−55.737	−38.626	**−32.949**	−36.182
Lower TVaR 1%	−39.745	−39.767	−68.407	−43.477	**−36.396**	−41.247

FIGURE 2.2 – Density plot of hedging errors for MEPH



2.6 Conclusion

A flexible and tractable methodology is presented for the hedging of contingent claims in the presence of regimes. It accommodates various hedging objectives through the penalty function specification. Constraints on trading strategies, such as no short-selling, can be incorporated.

Path dependency issues are tackled by the addition of a state variable, making the hedging problem suitable for dynamic programming. The approach is implemented with the standard Gaussian two-regime model estimated from weekly S&P 500 returns. Based on the hedging of an at-the-money call option, the current methodology compares favourably with three relevant alternatives.

Since the current paper's algorithm involves lattices, the curse of dimensionality prevents the use of a large number of underlying assets and regimes. The addition of a single dimension (transaction costs, a three-regime model or stochastic interest rates) remains feasible at a substantial numerical cost.

The modeling design voluntarily avoids the identification of the pricing measure. Nevertheless, if one wishes to determine the pricing kernel (at the cost of a specification error), several extensions become feasible : option tracking, hedging American options and hedging with other derivatives.

2.7 Appendix

Lemma 2.7.1 is used in the proofs of Corollaries 2.3.1 and 2.3.2.

Lemma 2.7.1 *Let $\mathcal{I} \subseteq \mathcal{J} \subseteq \mathcal{M}$ be sigma-algebras and Z be a random variable.*

$$\textit{If}\ \mathbb{E}\left[Z|\mathcal{I}\right] = \mathbb{E}\left[Z|\mathcal{M}\right],\ \textit{then}\ \mathbb{E}\left[Z|\mathcal{J}\right] = \mathbb{E}\left[Z|\mathcal{M}\right] = \mathbb{E}\left[Z|\mathcal{I}\right].$$

Proof of Lemma 2.7.1

$$\begin{aligned} \mathbb{E}\left[Z|\mathcal{J}\right] &= \mathbb{E}\left[\mathbb{E}\left[Z|\mathcal{M}\right]\big|\mathcal{J}\right] \quad \text{(Law of iterated expectations)} \\ &= \mathbb{E}\left[\mathbb{E}\left[Z|\mathcal{I}\right]\big|\mathcal{J}\right] \\ &= \mathbb{E}\left[Z|\mathcal{I}\right]. \quad \text{(Law of iterated expectations)} \end{aligned}$$

QED

Proof of Theorem 2.3.1

$$\begin{aligned}
\eta_{i,n+1} &= f_{h_{n+1}|Y_{0:n+1}}\left(i\,|Y_{0:n+1}\right) \\
&= \frac{f_{h_{n+1},Y_{0:n+1}}\left(i,Y_{0:n+1}\right)}{f_{Y_{0:n+1}}\left(Y_{0:n+1}\right)} \\
&= \frac{\sum_{j=1}^{H} f_{h_{n+1},h_n,Y_{0:n+1}}\left(i,j,Y_{0:n+1}\right)}{\sum_{k=1}^{H}\sum_{\ell=1}^{H} f_{h_{n+1},h_n,Y_{0:n+1}}\left(k,\ell,Y_{0:n+1}\right)} \\
&= \frac{\sum_{j=1}^{H} f_{h_{n+1},Y_{n+1}|h_n,Y_{0:n}}\left(i,Y_{n+1}\,|j,Y_{0:n}\right) f_{h_n|Y_{0:n}}\left(j\,|Y_{0:n}\right) f_{Y_{0:n}}\left(Y_{0:n}\right)}{\sum_{k=1}^{H}\sum_{\ell=1}^{H} f_{h_{n+1},Y_{n+1}|h_n,Y_{0:n}}\left(k,Y_{n+1}\,|\ell,Y_{0:n}\right) f_{h_n|Y_{0:n}}\left(\ell\,|Y_{0:n}\right) f_{Y_{0:n}}\left(Y_{0:n}\right)} \\
&= \frac{\sum_{j=1}^{H} f_{h_{n+1},Y_{n+1}|h_n,Y_{0:n}}\left(i,Y_{n+1}\,|j,Y_{0:n}\right)\eta_{j,n}}{\sum_{k=1}^{H}\sum_{\ell=1}^{H} f_{h_{n+1},Y_{n+1}|h_n,Y_{0:n}}\left(k,Y_{n+1}\,|\ell,Y_{0:n}\right)\eta_{\ell,n}}.
\end{aligned}$$

The Markov property of (Y,h) and Lemma 2.7.1 complete the proof since

$$f_{h_{n+1},Y_{n+1}|h_n,Y_{0:n}}\left(i,Y_{n+1}\,|j,Y_{0:n}\right) = f_{h_{n+1},Y_{n+1}|h_n,Y_n}\left(i,Y_{n+1}\,|j,Y_n\right).$$

QED

Proof of Corollary 2.3.1. Applying the Law of iterated expectations,

$$\mathbb{P}\left(h_n = i\,|\vec{\eta}_n\right) = \mathbb{E}\left(\mathbb{P}\left(h_n = i|\mathcal{G}_n\right)|\vec{\eta}_n\right) = \mathbb{E}\left(\eta_{i,n}\,|\vec{\eta}_n\right) = \eta_{i,n} = \mathbb{P}\left(h_n = i|\mathcal{G}_n\right). \tag{2.18}$$

By Lemma 2.7.1, since $\sigma(\vec{\eta}_n) \subseteq \sigma(Y_n, \vec{\eta}_n) \subseteq \mathcal{G}_n$,

$$\mathbb{P}\left(h_n = i\middle|\mathcal{G}_n\right) = \mathbb{P}\left(h_n = i\middle|Y_n, \vec{\eta}_n\right). \tag{2.19}$$

Moreover, since $\{Y_n, h_n, \vec{\eta}_n\}_{n=0}^{N}$ has the Markov property with respect to $\mathcal{F}$, then for any Borel set $\mathcal{D} \subseteq \mathbb{R}\times[0,1]^H$,

$$\begin{aligned}
& \mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}) \in \mathcal{D}|\,\mathcal{G}_n\right] \\
=\;& \mathbb{E}\left[\mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}) \in \mathcal{D}|\,\mathcal{F}_n\right]|\,\mathcal{G}_n\right] \text{ (Law of iterated expectations)} \\
=\;& \mathbb{E}\left[\mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}) \in \mathcal{D}|\,Y_n, h_n, \vec{\eta}_n\right]|\,\mathcal{G}_n\right] \text{ (Markov property)} \\
=\;& \sum_{j=1}^{H} \mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}) \in \mathcal{D}|\,Y_n, h_n = j, \vec{\eta}_n\right] \mathbb{P}\left[h_n = j|\,\mathcal{G}_n\right] \\
=\;& \sum_{j=1}^{H} \mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}) \in \mathcal{D}|\,Y_n, h_n = j, \vec{\eta}_n\right] \mathbb{P}\left[h_n = j|\,Y_n, \vec{\eta}_n\right] \text{ (Eq. (2.19))} \\
=\;& \mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}) \in \mathcal{D}|\,Y_n, \vec{\eta}_n\right] \text{ (Bayes' Law).}
\end{aligned}$$

QED

Proof of Corollary 2.3.2. For any admissible strategy, because of the self-financing restriction, its time-$(n+1)$ value satisfies

$$
\begin{aligned}
V_{n+1} &= V_n + \vec{\theta}_{n+1}^{\top}\left(\vec{S}_{n+1} - \vec{S}_n\right) \\
&= V_n + \theta_{n+1}^{(1)}\left(\exp(r(n+1)\Delta_t) - \exp(rn\Delta_t)\right) + \theta_{n+1}^{(2)} S_0^{(2)}\left(\exp(Y_{n+1}) - \exp(Y_n)\right).
\end{aligned}
$$

Hence, V_{n+1} is $\sigma\left(Y_{n+1}, Y_n, V_n, \vec{\theta}_{n+1}\right)$ −measurable. Furthermore, by Equation (2.18), Lemma 2.7.1 and the fact that $\sigma(\vec{\eta}_n) \subseteq \sigma(Y_n, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}) \subseteq \mathcal{G}_n$,

$$
\mathbb{P}\left(h_n = i \middle| \mathcal{G}_n\right) = \mathbb{P}\left(h_n = i \middle| Y_n, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}\right). \tag{2.20}
$$

Therefore, for any Borel set $\mathcal{D} \subseteq \mathbb{R} \times [0,1]^H \times \mathbb{R}$,

$$
\begin{aligned}
& \mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}) \in \mathcal{D} \middle| \mathcal{G}_n\right] \\
= \; & \mathbb{E}\left[\mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}) \in \mathcal{D} \middle| \mathcal{F}_n\right] \middle| \mathcal{G}_n\right] \text{ (Law of iterated expectations)} \\
= \; & \mathbb{E}\left[\mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}) \in \mathcal{D} \middle| Y_n, h_n, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}\right] \middle| \mathcal{G}_n\right] \text{ (Corollary 2.3.1)} \\
= \; & \sum_{j=1}^{H} \mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}) \in \mathcal{D} \middle| Y_n, h_n = j, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}\right] \mathbb{P}\left[h_n = j \middle| \mathcal{G}_n\right] \\
= \; & \sum_{j=1}^{H} \mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}) \in \mathcal{D} \middle| Y_n, h_n = j, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}\right] \mathbb{P}\left[h_n = j \middle| Y_n, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}\right] \text{ (Eq. (2.20))} \\
= \; & \mathbb{P}\left[(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}) \in \mathcal{D} \middle| Y_n, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}\right] \text{ (Bayes' Law).}
\end{aligned}
$$

QED

Proof of Theorem 2.3.2. In the following, all minimizations are performed over the set of admissible hedging strategies Θ. The hedging penalty Ψ_N^* depends on the initial prices $\vec{S}_0$, the initial portfolio value V_0, the cumulative returns $Y_{1:N}$, and the portfolio position $\vec{\theta}_{1:N}$, that is,

$$
\Psi_N^* = g(\phi(\vec{S}_N) - V_N) = \Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:N}\right).
$$

For any $n \in \{0, ..., N-2\}$, define

$$
\vec{\vartheta}_{(n+2):N} := \underset{\vec{\theta}_{n+2:N}}{\arg\min}\, \mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:N}\right) \middle| \mathcal{G}_{n+1}\right], \tag{2.21}
$$

implying that

$$
\begin{aligned}
\Psi_{n+1}^* &= \min_{\vec{\theta}_{n+2:N}} \mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:N}\right) \middle| \mathcal{G}_{n+1}\right] \\
&= \mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:n+1}, \vec{\vartheta}_{(n+2):N}\right) \middle| \mathcal{G}_{n+1}\right].
\end{aligned} \tag{2.22}
$$

First direction. For any admissible strategy $\vec{\theta}_{1:N}$,

$$\Psi_{n+1}^* = \mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:n+1}, \vec{\vartheta}_{n+2:N}\right)\middle|\mathcal{G}_{n+1}\right] \leq \mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:N}\right)\middle|\mathcal{G}_{n+1}\right].$$

Consequently, by monotonicity of the conditional expectation operator,

$$\begin{aligned}
\min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}^*\middle|\mathcal{G}_n\right] &\leq \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:N}\right)\middle|\mathcal{G}_{n+1}\right]\middle|\mathcal{G}_n\right] \\
&= \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:N}\right)\middle|\mathcal{G}_n\right] \text{ (Law of iterated expectations).}
\end{aligned}$$

Because the left-hand side of the previous inequality does not depend on $\vec{\theta}_{n+2:N}$, then

$$\min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}^*\middle|\mathcal{G}_n\right] = \min_{\vec{\theta}_{n+1:N}} \mathbb{E}\left[\Psi_{n+1}^*\middle|\mathcal{G}_n\right] \leq \min_{\vec{\theta}_{n+1:N}} \mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:N}\right)\middle|\mathcal{G}_n\right] = \Psi_n^*$$

where the last equality arises from Definition (2.4).

Second direction.

$$\begin{aligned}
\Psi_n^* &= \min_{\vec{\theta}_{n+1:N}} \mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:N}\right)\middle|\mathcal{G}_n\right] \text{ (Definition (2.4))} \\
&= \min_{\vec{\theta}_{n+1:N}} \mathbb{E}\left[\mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:N}\right)\middle|\mathcal{G}_{n+1}\right]\middle|\mathcal{G}_n\right] \text{ (Law of iterated expectations)} \\
&\leq \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:n+1}, \vec{\vartheta}_{n+2:N}\right)\middle|\mathcal{G}_{n+1}\right]\middle|\mathcal{G}_n\right] \text{ (Reducing optimization domain)} \\
&= \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}^*\middle|\mathcal{G}_n\right] \text{ (Definition (2.22)).}
\end{aligned} \tag{2.23}$$

Therefore, $\Psi_n^* = \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}^*\middle|\mathcal{G}_n\right]$, establishing Equation (2.5).

Furthermore, define $\vec{\theta}_{n+1}^*$ for any $n \in \{0, ..., N-1\}$ as a solution of

$$\vec{\theta}_{n+1}^* := \arg\min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}^*\middle|\mathcal{G}_n\right]. \tag{2.24}$$

Then,

$$\begin{aligned}
\Psi_n^* &= \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}^*|\mathcal{G}_n\right] \text{ (Equation (2.5))} \\
&= \mathbb{E}\left[\Psi_{n+1}^*(\vec{\theta}_{n+1}^*)\middle|\mathcal{G}_n\right] \text{ (Equation (2.24))} \\
&= \mathbb{E}\left[\mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:n}, \vec{\theta}_{n+1}^*, \vec{\vartheta}_{(n+2):N}\right)\middle|\mathcal{G}_{n+1}\right]\middle|\mathcal{G}_n\right] \text{ (Equation (2.22))} \\
&= \mathbb{E}\left[\Psi_N^*\left(\vec{S}_0, V_0, Y_{1:N}, \vec{\theta}_{1:n}, \vec{\theta}_{n+1}^*, \vec{\vartheta}_{(n+2):N}\right)\middle|\mathcal{G}_n\right] \text{ (Law of iterated expectations).}
\end{aligned}$$

Therefore, by Equation (2.4),

$$\vec{\vartheta}_{(n+1):N} := \left(\arg\min_{\vec{\theta}_{(n+1)}} \mathbb{E}\left[\Psi_{n+1}^*\middle|\mathcal{G}_n\right], \vec{\vartheta}_{(n+2):N}\right)$$

is a solution (possibly not the only one) to the following equation :

$$\vec{\vartheta}_{(n+1):N} := \underset{\vec{\theta}_{n+1:N}}{\arg\min}\, \mathbb{E}\left[\Psi_N^* \middle| \mathcal{G}_n \right]. \tag{2.25}$$

Hence, if Equation (2.24) is satisfied $\forall n \in \{0, ..., N-1\}$, a recursive argument implies that $\theta^* := \left(\vec{\theta}_1^*, ..., \vec{\theta}_N^*\right)$ solves Problem (2.1).

QED

Proof of Theorem 2.3.3.

The proof hinges on a backward induction over time. Clearly, $\Psi_N^* = g\left(\tilde{\phi}(Y_N) - V_N\right)$ is $\sigma\left(Y_N, \vec{\eta}_N, V_N\right)$ −measurable. Assume that Ψ_{n+1}^* is $\sigma\left(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}\right)$ −measurable. From Corollary 2.3.2 , there is a Borel-measurable function φ such that

$$\begin{aligned}
\Psi_n^* &= \min_{\vec{\theta}_{n+1} \in \mathcal{G}_n} \mathbb{E}\left[\Psi_{n+1}^* \middle| \mathcal{G}_n \right] \text{ (Equation (2.5))} \\
&= \min_{\vec{\theta}_{n+1} \in \mathcal{G}_n} \mathbb{E}\left[\Psi_{n+1}^* \middle| Y_n, \vec{\eta}_n, V_n, \vec{\theta}_{n+1} \right] \text{ (Corollary 2.3.2)} \\
&= \min_{\vec{\theta}_{n+1} \in \mathcal{G}_n} \varphi\left(Y_n, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}\right).
\end{aligned}$$

Therefore, a necessary and sufficient condition for $\vec{\theta}_{n+1}^* = \underset{\vec{\theta}_{n+1}}{\arg\min} \mathbb{E}\left[\Psi_{n+1}^* | \mathcal{G}_n\right]$ is to minimize $\xi(\cdot) := \varphi\left(Y_n, \vec{\eta}_n, V_n, \cdot\right)$ which only depends on $(Y_n, \vec{\eta}_n, V_n)$. Consequently, there exists $\vec{\theta}_{n+1}^*$ which is $\sigma\left(Y_n, \vec{\eta}_n, V_n\right)$ −measurable. Hence, $\Psi_n^* = \varphi\left(Y_n, \vec{\eta}_n, V_n, \vec{\theta}_{n+1}^*\right)$ is also $\sigma\left(Y_n, \vec{\eta}_n, V_n\right)$ −measurable and

$$\vec{\theta}_{n+1}^* = \underset{\vec{\theta}_{n+1} \in \sigma(Y_n, \vec{\eta}_n, V_n)}{\text{argmin}} \mathbb{E}\left[\Psi_{n+1}^* \left| \mathcal{G}_n \right.\right] = \underset{\vec{\theta}_{n+1} \in \sigma(Y_n, \vec{\eta}_n, V_n)}{\text{argmin}} \mathbb{E}\left[\Psi_{n+1}^* \left| Y_n, \vec{\eta}_n, V_n \right.\right],$$

that is, the set of admissible hedging strategies Θ may be restricted to keep only strategies that also satisfy that θ_{n+1} is $\sigma\left(Y_n, \vec{\eta}_n, V_n\right)$ −measurable.

QED

Proof of Proposition 2.5.1.

$$\begin{aligned}
\eta_{j,n+1} &= \mathbb{P}(h_{n+1} = j | \mathcal{G}_n) \\
&= \mathbb{E}\left[\mathbb{P}(h_{n+1} = j | h_n) \middle| \mathcal{G}_n\right] \\
&= \sum_{i=1}^{H} \mathbb{P}(h_{n+1} = j | h_n = i) \mathbb{P}(h_n = i | \mathcal{G}_n) \\
&= \sum_{i=1}^{H} P_{i,j}^{(n)} \eta_{i,n} \geq \sum_{i=1}^{H} \min_{u \in \mathcal{H}} P_{u,j}^{(n)} \eta_{i,n} = \min_{u \in \mathcal{H}} P_{u,j}^{(n)}
\end{aligned}$$

The case $\eta_{j,n+1} \leq \max_{u \in \mathcal{H}} P_{u,j}$ is similar.

QED

2.8 Technical Report (not part of the paper)

2.8.1 Explicit formulas for the last time step

Theorem 2.8.1 *Let the market dynamics be those described in Section 2.5. Let also the penalty be* $g(x) = x^2$ *and the payoff be the one of a call option :* $\phi(S_N) = \max\{S_N - K, 0\}$. *Suppose there are no constraints on* θ^*_N. *Then, in theorem 2.3.4,* θ^*_N *and* Ψ_{N-1} *are given by :*

$$\begin{aligned}
\theta^*_N &= \frac{\sum_{j=1}^2 \eta_{j,N-1}\left[e^{r\Delta_t}V_{N-1}(Q_j^{(1)} - e^{r\Delta_t}) + e^{r\Delta_t}EC_j - ESC_j/S_{N-1}\right]}{S_{N-1}\sum_{j=1}^2 \eta_{j,N-1}\left[2e^{r\Delta_t}Q_j^{(1)} - Q_j^{(2)} - e^{2r\Delta_t}\right]} \\
\Psi_{N-1} &= \sum_{j=1}^2 \eta_{j,N-1}\left[EV_j^2 - 2ECV_j + EC_j^2\right],
\end{aligned}$$

where

$$\begin{aligned}
ECV_j &= \theta^*_N ESC_j + e^{r\Delta_t}(V_{N-1} - \theta^*_N S_{N-1})EC_j \\
ESC_j &= Q_j^{(4)} - KQ_j^{(3)} \\
EC_j &= Q_j^{(3)} - KQ_j^{(5)} \\
EC_j^2 &= Q_j^{(4)} - 2KQ_j^{(3)} + K^2Q_j^{(5)} \\
EV_j^2 &= (\theta^*_N S_{N-1})^2 Q_j^{(2)} + e^{2r\Delta_t}(V_{N-1} - \theta^*_N S_{N-1})^2 + 2\theta^*_N S_{N-1}e^{r\Delta_t}(V_{N-1} - \theta^*_N S_{N-1})Q_j^{(1)} \\
Q_j^{(1)} &= \exp(\mu_j\Delta_t + .5\sigma_j^2\Delta_t) \\
Q_j^{(2)} &= \exp(2\mu_j\Delta_t + 2\sigma_j^2\Delta_t) \\
Q_j^{(3)} &= S_{N-1}\exp(\mu_j\Delta_t + .5\sigma_j^2\Delta_t)\Phi\left(\frac{\log(S_{N-1}/K) + \mu_j\Delta_t + \sigma_j^2\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right) \\
Q_j^{(4)} &= S_{N-1}^2\exp(2\mu_j\Delta_t + 2\sigma_j^2\Delta_t)\Phi\left(\frac{2\log(S_{N-1}/K) + 2\mu_j\Delta_t + 2\sigma_j^2\Delta_t}{\sigma_j\sqrt{2\Delta_t}}\right) \\
Q_j^{(5)} &= \Phi\left(\frac{\log(S_{N-1}/K) + \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right)
\end{aligned}$$

Proof of Theorem 2.8.1 : Define Let $C_N := \max(S_N - K, 0)$. For an arbitrary θ_N,

$$\begin{aligned}
\mathbb{E}\left[(C_N - V_N)^2|\mathcal{G}_{N-1}\right] &= \mathbb{E}\left[C_N^2|\mathcal{G}_{N-1}\right] - 2\mathbb{E}\left[C_NV_N|\mathcal{G}_{N-1}\right] + \mathbb{E}\left[V_N^2|\mathcal{G}_{N-1}\right] \\
&= \sum_{j=1}^H \eta_{j,N-1}\left(\mathbb{E}\left[C_N^2|h_{N-1}=j\right] - 2\mathbb{E}\left[C_NV_N|h_{N-1}=j\right] + \mathbb{E}\left[V_N^2|h_{N-1}=j\right]\right)
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left[V_N^2|h_{N-1}=j\right] &= \mathbb{E}\left[(\theta_N S_N + (V_{N-1}-\theta_N S_{N-1})e^{r\Delta_t})^2|h_{N-1}=j\right] \\
&= \theta_N^2 S_{N-1}^2 \mathbb{E}\left[e^{2\epsilon}|h_{N-1}=j\right] + (V_{N-1}-\theta_N S_{N-1})^2 e^{2r\Delta_t} \\
&\quad +2\theta_N S_{N-1}(V_{N-1}-\theta_N S_{N-1})e^{r\Delta_t}\mathbb{E}\left[e^{\epsilon}|h_{N-1}=j\right] \\
&= \theta_N^2 S_{N-1}^2 Q_j^{(2)} + (V_{N-1}-\theta_N S_{N-1})^2 e^{2r\Delta_t} \\
&\quad +2\theta_N S_{N-1}(V_{N-1}-\theta_N S_{N-1})e^{r\Delta_t} Q_j^{(1)}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[C_N^2|h_{N-1}=j\right] &= \mathbb{E}\left[\max\left(S_{N-1}e^{\epsilon}-K,0\right)^2|h_{N-1}=j\right] \\
&= \mathbb{E}\left[\left(S_{N-1}e^{\epsilon}-K\right)^2 \mathbf{1}_{\{S_{N-1}e^{\epsilon}>K\}}|h_{N-1}=j\right] \\
&= \mathbb{E}\left[S_{N-1}^2 e^{2\epsilon}\mathbf{1}_{\{S_{N-1}^2 e^{2\epsilon}>K^2\}}|h_{N-1}=j\right] - 2K\mathbb{E}\left[S_{N-1}e^{\epsilon}\mathbf{1}_{\{S_{N-1}e^{\epsilon}>K\}}|h_{N-1}=j\right] \\
&\quad +K^2\mathbb{P}\left(S_{N-1}e^{\epsilon}>K|h_{N-1}=j\right) \\
&= Q_j^{(4)} - 2KQ_j^{(3)} + K^2 Q_j^{(5)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\quad [C_N V_N|h_{N-1}=j] \\
&= \mathbb{E}\left[C_N(\theta_N S_N + (V_{N-1}-\theta_N S_{N-1})e^{r\Delta_t})|h_{N-1}=j\right] \\
&= \theta_N \mathbb{E}\left[S_N C_N|h_{N-1}=j\right] + (V_{N-1}-\theta_N S_{N-1})e^{r\Delta_t}\mathbb{E}\left[C_N|h_{N-1}=j\right] \\
&= \theta_N \mathbb{E}\left[S_N(S_N-K)\mathbf{1}_{\{S_N>K\}}|h_{N-1}=j\right] + (V_{N-1}-\theta_N S_{N-1})e^{r\Delta_t}\mathbb{E}\left[(S_N-K)\mathbf{1}_{\{S_N>K\}}|h_{N-1}=j\right] \\
&= \theta_N \mathbb{E}\left[S_N^2\mathbf{1}_{\{S_N^2>K^2\}}|h_{N-1}=j\right] - K\theta_N\mathbb{E}\left[S_N\mathbf{1}_{\{S_N>K\}}|h_{N-1}=j\right] \\
&\quad +(V_{N-1}-\theta_N S_{N-1})e^{r\Delta_t}\left(\mathbb{E}\left[S_N\mathbf{1}_{\{S_N>K\}}|h_{N-1}=j\right] - K\mathbb{P}\left(S_N>K|h_{N-1}=j\right)\right) \\
&= \theta_N\left(Q_j^{(4)} - KQ_j^{(3)}\right) + (V_{N-1}-\theta_N S_{N-1})e^{r\Delta_t}\left(Q_j^{(3)} - KQ_j^{(5)}\right)
\end{aligned}
$$

To obtain the optimal θ_N, one applies the first order condition :

$$
\begin{aligned}
&\frac{d\mathbb{E}\left[(C_N-V_N)^2|\mathcal{G}_{N-1}\right]}{d\theta_N} = 0 \\
\Leftrightarrow\ & \sum_{j=1}^{H}\eta_{j,N-1}\left[-2\left(Q_j^{(4)} - KQ_j^{(3)} + S_{N-1}e^{r\Delta_t}\left(Q_j^{(3)} - KQ_j^{(5)}\right)\right)\right] \\
&+\sum_{j=1}^{H}\eta_{j,N-1}\left[2\theta_N S_{N-1}^2 Q_j^{(2)} - 2S_{N-1}(V_{N-1}-\theta_N S_{N-1})e^{2r\Delta_t}\right] \\
&+\sum_{j=1}^{H}\eta_{j,N-1}\left[2S_{N-1}V_{N-1}e^{r\Delta_t}Q_j^{(1)} - 4\theta_N S_{N-1}^2 e^{r\Delta_t}Q_j^{(1)}\right] = 0 \\
\Leftrightarrow\ & \theta_N^* = \frac{\sum_{j=1}^{2}\eta_{j,N-1}\left[e^{r\Delta_t}V_{N-1}(Q_j^{(1)} - e^{r\Delta_t}) + e^{r\Delta_t}(Q_j^{(3)} - KQ_j^{(5)}) - (Q_j^{(4)} - KQ_j^{(3)})/S_{N-1}\right]}{S_{N-1}\sum_{j=1}^{2}\eta_{j,N-1}\left[2e^{r\Delta_t}Q_j^{(1)} - Q_j^{(2)} - e^{2r\Delta_t}\right]}
\end{aligned}
$$

Define now

$$\begin{aligned}
ECV_j &= \theta_N^* ESC_j + e^{r\Delta_t}(V_{N-1} - \theta_N^* S_{N-1}) EC_j \\
ESC_j &= Q_j^{(4)} - KQ_j^{(3)} \\
EC_j &= Q_j^{(3)} - KQ_j^{(5)} \\
EC_j^2 &= Q_j^{(4)} - 2KQ_j^{(3)} + K^2 Q_j^{(5)} \\
EV_j^2 &= (\theta_N^* S_{N-1})^2 Q_j^{(2)} + e^{2r\Delta_t}(V_{N-1} - \theta_N^* S_{N-1})^2 + 2\theta_N^* S_{N-1} e^{r\Delta_t}(V_{N-1} - \theta_N^* S_{N-1}) Q_j^{(1)}
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
\Psi_{N-1}^* &= \mathbb{E}\left[(C_N - V_N(\theta_N^*))^2 | \mathcal{G}_{N-1}\right] \\
&= \sum_{j=1}^{2} \eta_{j,N-1}\left[\mathbb{E}\left[C_N^2 | h_{N-1} = j\right] - 2\mathbb{E}\left[C_N V_N(\theta_N^*) | h_{N-1} = j\right] + \mathbb{E}\left[V_N^2(\theta_N^*) | h_{N-1} = j\right]\right] \\
&= \sum_{j=1}^{2} \eta_{j,N-1}\left[EC_j^2 - 2ECV_j + EV_j^2\right]
\end{aligned}$$

QED

Theorem 2.8.2 *Let the market dynamics be those described in Section 2.5. Let also the penalty be* $g(x) = x^2 \mathbf{1}_{\{x>0\}}$ *or* $g(x) = x^2 \mathbf{1}_{\{x<0\}}$. *Define events* $F^{(1)} := \{\phi(S_N) > V_N\}$, $F^{(2)} := \{\phi(S_N) < V_N\}$. *Define* $F := F^{(1)}$ *if* $g(x) = x^2 \mathbf{1}_{\{x>0\}}$ *and* $F := F^{(2)}$ *if* $g(x) = x^2 \mathbf{1}_{\{x<0\}}$. *Define also* $A := \{S_N > K\}$ *Let the payoff be the one of a call option :* $\phi(S_N) = \max\{S_N - K, 0\}$. *Then,* $\rho_{X_{N-1}}(\theta_N) := \mathbb{E}\left[\Psi_N(\theta_N) \middle| Y_{N-1}, \eta_{N-1}, V_{N-1}\right]$ *is given by :*

$$\rho_{X_{N-1}}(\theta_N) = \sum_{j=1}^{2} \eta_{j,N-1}\left[EV_j^2 - 2ECV_j + EC_j^2\right],$$

where

$$\begin{aligned}
EV_j^2 &= S_{N-1}^2 \theta_N^2 E^j[e^{2\epsilon} \mathbf{1}_F] + 2S_{N-1}\theta_N e^{r\Delta_t}(V_{N-1} - S_{N-1}\theta_N) E^j[e^{\epsilon} \mathbf{1}_F] \\
&\quad + e^{2r\Delta_t}(V_{N-1}^2 - 2V_{N-1}S_{N-1}\theta_N + S_{N-1}^2 \theta_N^2) E^j[\mathbf{1}_F] \\
EC_j^2 &= S_{N-1}^2 E^j[e^{2\epsilon} \mathbf{1}_F \mathbf{1}_A] - 2KS_{N-1} E^j[e^{\epsilon} \mathbf{1}_F \mathbf{1}_A] + K^2 E^j[\mathbf{1}_F \mathbf{1}_A] \\
ECV_j &= S_{N-1}^2 \theta_N E^j[e^{2\epsilon} \mathbf{1}_F \mathbf{1}_A] + S_{N-1}(V_{N-1} e^{r\Delta_t} - S_{N-1}\theta_N e^{r\Delta_t} - K\theta_N) E^j[e^{\epsilon} \mathbf{1}_F \mathbf{1}_A] \\
&\quad + K e^{r\Delta_t}(S_{N-1}\theta_N - V_{N-1}) E^j[\mathbf{1}_F \mathbf{1}_A]
\end{aligned}$$

$$
\begin{aligned}
G^{(1)} &= K/S_{N-1} \\
G^{(2)} &= K + (V_{N-1} - S_{N-1}\theta_N)e^{r\Delta_t} \\
G^{(3)} &= \frac{K + (V_{N-1} - S_{N-1}\theta_N)e^{r\Delta_t}}{S_{N-1}(1-\theta_N)} \\
G^{(4)} &= \frac{K - G^{(2)}}{S_{N-1}\theta_N}
\end{aligned}
$$

Define $\zeta(x,y,z) := \exp(x + .5y^2)\Phi\left(\frac{-\log(z)+x+y^2}{y}\right)$

$$
\begin{aligned}
&E^j\left[\mathbf{1}_A\right] = \Phi\left(\frac{-\log(G^{(1)}) + \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right) \\
&E^j\left[e^{\epsilon}\mathbf{1}_A\right] = \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, G^{(1)}) \\
&E^j\left[e^{2\epsilon}\mathbf{1}_A\right] = \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, (G^{(1)})^2)
\end{aligned}
$$

For the case of the short quadratic penalty :

If $\theta_N < 0$,

$$
\begin{aligned}
&E^j\left[\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \Phi\left(\frac{-\log(\max(G^{(1)}, G^{(3)})) + \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right) \\
&E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, \max(G^{(1)}, G^{(3)})) \\
&E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, \max(G^{(1)}, G^{(3)})^2) \\
&E^j\left[\mathbf{1}_{F^{(1)}}\right] = E[\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \max\left(0, \Phi\left(\frac{-\log(G^{(4)}) + \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right) - \Phi\left(\frac{-\log(G^{(1)}) + \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right)\right) \\
&E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \max(0, \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, G^{(4)}) - \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, G^{(1)})) \\
&E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \max(0, \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, (G^{(4)})^2) - \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, (G^{(1)})^2))
\end{aligned}
$$

If $\theta_N = 0$,

$$
\begin{aligned}
&E^j\left[\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \Phi\left(\frac{-\log(\max(G^{(1)}, G^{(3)})) + \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right) \\
&E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, \max(G^{(1)}, G^{(3)})) \\
&E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, \max(G^{(1)}, G^{(3)})^2) \\
&E^j\left[\mathbf{1}_{F^{(1)}}\right] = E[\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \mathbf{1}_{\{V_{N-1}<0\}}\Phi\left(\frac{\log(G^{(1)} - \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right) \\
&E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \mathbf{1}_{\{V_{N-1}<0\}}(\exp(\mu_j\Delta_t + .5\sigma_j^2\Delta_t) - E^j[e^{\epsilon}\mathbf{1}_A]) \\
&E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \mathbf{1}_{\{V_{N-1}<0\}}(\exp(2\mu_j\Delta_t + 2\sigma_j^2\Delta_t) - E^j[e^{\epsilon}2\mathbf{1}_A])
\end{aligned}
$$

If $0 < \theta_N < 1$,

$$E^j\left[\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \Phi\left(\frac{-\log(\max(G^{(1)}, G^{(3)})) + \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right)$$
$$E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, \max(G^{(1)}, G^{(3)}))$$
$$E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, \max(G^{(1)}, G^{(3)})^2)$$
$$E^j\left[\mathbf{1}_{F^{(1)}}\right] = E[\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \Phi\left(\frac{\log(\min(G^{(1)}, G^{(4)})) - \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right)$$
$$E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \exp(\mu_j\Delta_t + .5\sigma_j^2\Delta_t) - \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, \min(G^{(1)}, G^{(4)}))$$
$$E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \exp(2\mu_j\Delta_t + 2\sigma_j^2\Delta_t) - \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, \min(G^{(1)}, G^{(4)})^2)$$

If $\theta_N = 1$,

$$E^j\left[\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \mathbf{1}_{\{G^{(2)}<0\}}E^j[\mathbf{1}_A]$$
$$E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \mathbf{1}_{\{G^{(2)}<0\}}E^j[e^{\epsilon}\mathbf{1}_A]$$
$$E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \mathbf{1}_{\{G^{(2)}<0\}}E^j[e^{2\epsilon}\mathbf{1}_A]$$
$$E^j\left[\mathbf{1}_{F^{(1)}}\right] = E[\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \Phi\left(\frac{\log(\min(G^{(1)}, G^{(4)})) - \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right)$$
$$E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \exp(\mu_j\Delta_t + .5\sigma_j^2\Delta_t) - \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, \min(G^{(1)}, G^{(4)}))$$
$$E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \exp(2\mu_j\Delta_t + 2\sigma_j^2\Delta_t) - \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, \min(G^{(1)}, G^{(4)})^2)$$

If $\theta_N > 1$,

$$E^j\left[\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \max\left(0, \Phi\left(\frac{-\log(G^{(1)}) + \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right) - \Phi\left(\frac{-\log(G^{(3)}) + \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right)\right)$$
$$E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \max(0, \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, G^{(1)}) - \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, G^{(3)}))$$
$$E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A\right] = \max(0, \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, (G^{(1)})^2) - \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, (G^{(3)})^2))$$
$$E^j\left[\mathbf{1}_{F^{(1)}}\right] = E[\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \Phi\left(\frac{\log(\min(G^{(1)}, G^{(4)})) - \mu_j\Delta_t}{\sigma_j\sqrt{\Delta_t}}\right)$$
$$E^j\left[e^{\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \exp(\mu_j\Delta_t + .5\sigma_j^2\Delta_t) - \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}, \min(G^{(1)}, G^{(4)}))$$
$$E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\right] = E^j[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A] + \exp(2\mu_j\Delta_t + 2\sigma_j^2\Delta_t) - \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}, \min(G^{(1)}, G^{(4)})^2)$$

For the case of the long quadratic penalty :

$$E^j\left[\mathbf{1}_{F^{(2)}}\mathbf{1}_A\right] = E^j[\mathbf{1}_A] - E^j[\mathbf{1}_{F^{(1)}}\mathbf{1}_A]$$

$$E^j\left[e^{\epsilon}\mathbf{1}_{F^{(2)}}\mathbf{1}_A\right] = E^j[e^{\epsilon}\mathbf{1}_A] - E^j[e^{\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A]$$

$$E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(2)}}\mathbf{1}_A\right] = E^j[e^{2\epsilon}\mathbf{1}_A] - E^j[e^{2\epsilon}\mathbf{1}_{F^{(1)}}\mathbf{1}_A]$$

$$E^j\left[\mathbf{1}_{F^{(2)}}\right] = 1 - E^j[\mathbf{1}_{F^{(1)}}]$$

$$E^j\left[e^{\epsilon}\mathbf{1}_{F^{(2)}}\right] = \exp(\mu_j\Delta_t + .5\sigma_j^2\Delta_t) - E^j[e^{\epsilon}\mathbf{1}_{F^{(1)}}]$$

$$E^j\left[e^{2\epsilon}\mathbf{1}_{F^{(2)}}\right] = \exp(2\mu_j\Delta_t + 2\sigma_j^2\Delta_t) - E^j[e^{2\epsilon}\mathbf{1}_{F^{(1)}}]$$

Proof of Theorem 2.8.2 :

Define $C_N := \phi(S_N)$. Then, $\{g(C_N - V_N) > 0\} = F$.

Therefore,

$$\begin{aligned}
\rho_{X_{N-1}}(\theta_N) &= \sum_{j=1}^{H}\eta_{j,n}\mathbb{E}\left[g(C_N - V_N(\theta_N))\middle|S_{N-1}, V_{N-1}, h_{N-1} = j\right] \\
&= \sum_{j=1}^{H}\eta_{j,n}\mathbb{E}\left[(C_N - V_N(\theta_N))^2\mathbf{1}_F\middle|S_{N-1}, V_{N-1}, h_{N-1} = j\right] \\
&= \sum_{j=1}^{H}\eta_{j,n}\mathbb{E}\left[C_N^2\mathbf{1}_F - 2C_NV_N(\theta_N)\mathbf{1}_F + V_N^2(\theta_N)\mathbf{1}_F\middle|S_{N-1}, V_{N-1}, h_{N-1} = j\right]
\end{aligned}$$

Define

$$\begin{aligned}
EV_j^2 &:= \mathbb{E}\left[C_N^2\mathbf{1}_F\middle|S_{N-1}, V_{N-1}, h_{N-1} = j\right] \\
ECV_j &:= \mathbb{E}\left[C_NV_N(\theta_N)\mathbf{1}_F\middle|S_{N-1}, V_{N-1}, h_{N-1} = j\right] \\
EV_j^2 &:= \mathbb{E}\left[V_N^2(\theta_N)\mathbf{1}_F\middle|S_{N-1}, V_{N-1}, h_{N-1} = j\right]
\end{aligned}$$

Denote $A := \{C_N > 0\} = \{e^{\epsilon_N} > K/S_{N-1}\}$. Therefore, $C_N = (S_{N-1}e^{\epsilon_N} - K)\mathbf{1}_A$. Also, $V_N(\theta_N) = V_{N-1}e^{r\Delta_t} + \theta_NS_{N-1}(e^{\epsilon_N} - e^{r\Delta_t})$. Combining the above formulas yields and using notation shortcuts $E^j[\bullet] := \mathbb{E}\left[\bullet\middle|S_{N-1}, V_{N-1}, h_{N-1} = j\right]$ and $\epsilon := \epsilon_N$, one gets :

$$\begin{aligned}
EV_j^2 &= S_{N-1}^2\theta_N^2E^j[e^{2\epsilon}\mathbf{1}_F] + 2S_{N-1}\theta_Ne^{r\Delta_t}(V_{N-1} - S_{N-1}\theta_N)E^j[e^{\epsilon}\mathbf{1}_F] \\
&\quad + e^{2r\Delta_t}(V_{N-1}^2 - 2V_{N-1}S_{N-1}\theta_N + S_{N-1}^2\theta_N^2)E^j[\mathbf{1}_F] \\
EC_j^2 &= S_{N-1}^2E^j[e^{2\epsilon}\mathbf{1}_F\mathbf{1}_A] - 2KS_{N-1}E^j[e^{\epsilon}\mathbf{1}_F\mathbf{1}_A] + K^2E^j[\mathbf{1}_F\mathbf{1}_A] \\
ECV_j &= S_{N-1}^2\theta_NE^j[e^{2\epsilon}\mathbf{1}_F\mathbf{1}_A] + S_{N-1}(V_{N-1}e^{r\Delta_t} - S_{N-1}\theta_Ne^{r\Delta_t} - K\theta_N)E^j[e^{\epsilon}\mathbf{1}_F\mathbf{1}_A] \\
&\quad + Ke^{r\Delta_t}(S_{N-1}\theta_N - V_{N-1})E^j[\mathbf{1}_F\mathbf{1}_A].
\end{aligned}$$

Define

$$\zeta(x,y,z) := \begin{cases} \exp(x+.5y^2)\Phi\left(\frac{-\log(z)+x+y^2}{y}\right) & \text{if } z>0 \\ \exp(x+.5y^2) \quad \text{if } z \le 0. \end{cases} \tag{2.26}$$

Since

$$\epsilon_N | h_{N-1} = j \text{ is Gaussian}(\mu_j \Delta_t, \sigma_j \sqrt{\Delta_t}), \tag{2.27}$$

one gets that

$$E^j[e^\epsilon \mathbf{1}_{\{e^\epsilon > Z\}}] = \zeta(\mu_j\Delta_t, \sigma_j\sqrt{\Delta_t}), Z) \quad \text{if } Z>0 \tag{2.28}$$

$$E^j[e^{2\epsilon} \mathbf{1}_{\{e^\epsilon > Z\}}] = \zeta(2\mu_j\Delta_t, 2\sigma_j\sqrt{\Delta_t}), \text{sign}(Z)Z^2) \tag{2.29}$$

Define

$$\begin{aligned} G^{(1)} &= K/S_{N-1} \\ G^{(2)} &= K + (V_{N-1} - S_{N-1}\theta_N)e^{r\Delta_t} \\ G^{(3)} &= \frac{K + (V_{N-1} - S_{N-1}\theta_N)e^{r\Delta_t}}{S_{N-1}(1-\theta_N)} \\ G^{(4)} &= \frac{K - G^{(2)}}{S_{N-1}\theta_N} \end{aligned}$$

The rest of the proof is a direct consequence of (2.27), (2.28)-(2.29), $\mathbf{1}_{F^{(1)}} + \mathbf{1}_{F^{(2)}} = 1$ and the following equations :

$$\begin{aligned} A &= \{e^\epsilon > G^{(1)}\} \\ A \cap F^{(1)} &= \begin{cases} \{e^\epsilon > \max(G^{(1)}, G^{(3)})\} \text{ if } \theta_N < 1 \\ \{e^\epsilon > G^{(1)}\} \cap \{G^{(2)} < 0\} \text{ if } \theta_N = 1 \\ \{e^\epsilon \in [G^{(1)}, G^{(3)}]\} \text{ if } \theta_N > 1 \end{cases} \\ F^{(1)} &= (A \cap F^{(1)}) + \begin{cases} \{e^\epsilon \in [G^{(4)}, G^{(1)}] \text{ if } \theta_N < 0 \\ \{e^\epsilon \le G^{(1)}\} \cap \{V_{N-1} < 0\} \text{ if } \theta_N = 0 \\ \{e^\epsilon < \min(G^{(1)}, G^{(4)}) \text{ if } \theta_N > 0 \end{cases} \end{aligned}$$

with $[a,b] = \emptyset$ if $a > b$, and the $+$ operation between sets denoting disjoint union.

QED

2.8.2 The tight hedging extension

Assume one seeks a replicating portfolio that not only replicates the final derivative's payoff accurately, but that also remains close to the contingent claim's price C_n at all times during the hedging period (i.e. a tight hedging). Then, one could search for a hedging strategy which minimizes $E[\sum_{n=1}^{N} g_n(C_n - V_n)]$ where $\{g_n\}$ is a set of penalty functions. Problem (2.1) is the particular case $g_N = g$ and $g_n \equiv 0 \, \forall n < N$.

Assume the price C_n of the derivative only depends on $(Y_n, \vec{\eta}_n, V_n)$. This assumption encompasses many option pricing methodologies (c.f. amongst others Rémillard (2009)). Then, instead of having a Bellman equation of the form given in Theorem 2.3.4, one gets an equation of the form

$$\Psi_n(Y_n, \vec{\eta}_n, V_n) = g_n(C_n - V_n) + \min_{\vec{\theta}_{n+1}} \mathbb{E}\left[\Psi_{n+1}(Y_{n+1}, \vec{\eta}_{n+1}, V_{n+1}) \middle| Y_n, \vec{\eta}_n, V_n\right].$$

For this methodology, an arbitrary martingale measure must be selected to compute C_n. In conclusion, solving the tight hedging problem is not harder than solving (2.1), provided one has an efficient way of computing C_n given $(Y_n, \vec{\eta}_n, V_n)$.

References

[1] Alizadeh, A., Nomikos, N. (2004). A Markov regime switching approach for hedging stock indices. *Journal of Futures Markets*. **(24)**. pp. 649-674

[2] Alizadeh, A.H., Nomikos, N.K., Pouliasis, P.K. (2008). A Markov regime switching approach for hedging energy commodities. *Journal of Banking and Finance*. **(32)**. pp. 1970-1983

[3] Augustyniak, M., Boudreault, M. (2012). An out-of-sample analysis of investment guarantees for equity-linked products : lessons from the financial crisis of the late 2000s. *North American Actuarial Journal*. **(16)**. pp. 183-206

[4] Baum, L.E., Petrie, T., Soules, G., Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*. **(41)**. pp. 164-171

[5] Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*. **(81)**. pp. 637-654

[6] Bollen, N.P.B. (1998). Valuing options in regime-switching models. *Journal of Derivatives*. **(6)**. pp. 38-49

[7] Buffington, J., Elliott, R.J. (2002). American options with regime switching. *International Journal of Theoretical and Applied Finance*. **(5)**. pp. 497-514

[8] Cont, R., Tankov, P., Voltchkova, E. (2007). Hedging with options in models with jumps. *Stochastic Analysis and Applications*. **(2)**. pp. 197-217

[9] Cvitanić, J., Spivak, G. (1999). Maximizing the probability of a perfect hedge. *The Annals of Applied Probability*. **(9)**. pp. 1303-1328

[10] Cvitanić, J., Karatzas, I. (1999). On dynamic measures of risk. *Finance and Stochastics*. **(3)**. pp. 451-482

[11] De Giovanni, D., Ortobelli, S., Rachev, S. (2008). Delta hedging strategies comparison. *European Journal of Operational Research.* **(185)**. pp. 1615-1631

[12] Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society.* **(39)**. pp. 1-38

[13] Duan, J.C. (1995). The GARCH Option Pricing Model. *Mathematical Finance.* **(5)**. pp. 13-32

[14] Duan, J.C., Popova, I. and Ritchken, P. (2002). Option pricing under regime switching. *Quantitative Finance.* **(2)**. pp. 116-132

[15] Eberlein, E., Jacod, J. (1997). On the range of option prices. *Finance and Stochastics.* **(1)**. pp. 131-140

[16] El Karoui, N., Quenez, M.-C. (1995). Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM Journal on Control and Optimization.* **(33)**. pp. 29-66

[17] Elliott, R.J., Chan, L.L., Siu, T.K., (2005). Option pricing and Esscher transform under regime switching. *Annals of Finance.* **(1)**. pp. 423-432

[18] Föllmer, H., Leukert, P. (2000). Efficient hedging : cost versus shortfall risk. *Finance and Stochastics.* **(4)**. pp. 117-146

[19] Föllmer, H., Leukert, P. (1999). Quantile hedging. *Finance and Stochastics.* **(3)**. pp. 251-273

[20] Genest, C., Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques.* **(44)**. pp. 1096-1127

[21] Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica.* **(57)**. pp. 357-384

[22] Hardy, M. (2001). A regime switching model of long term stock returns. *North American Actuarial Journal.* **(5)**. pp. 41-53

[23] Karatzas, I. (1997). Lectures in mathematical finance. *Providence : American Mathematical Society*

[24] Lee, H.T., Yoder, J.K., Mittelhammer, R.C., McCluskey, J.J. (2006). A random coefficient autoregressive Markov regime switching model for dynamic futures hedging. *Journal of Futures Markets*. **(26)**. pp. 103-129

[25] Lee, H.T., Yoder, J.K. (2007). Optimal hedging with a regime-switching time-varying correlation GARCH model. *Journal of Futures Markets*. **(27)**. pp. 495-516

[26] Lee, H.T. (2009a). Optimal futures hedging under jump switching dynamics. *Journal of Empirical Finance*. **(16)**. pp. 446-456

[27] Lee, H.T. (2009b). A copula-based Markov regime switching GARCH model for optimal futures hedging. *Journal of Futures Markets*. **(29)**. pp. 946-972

[28] Lien, D. (2012). A note on the performance of regime switching hedge strategy. *The Journal of Futures Markets*. **(32)**. pp. 389-396

[29] Mamon, R.S., Rodrigo, M.R. (2005). Explicit solutions to European options in a regime-switching economy. *Operations Research Letters*. **(33)**. pp. 581-586

[30] Motoczyński, M. (2000). Multidimensional variance-optimal hedging in discrete-time model - a general approach. *Mathematical Finance*. **(10)**. pp. 243-257

[31] Pham, H. (2000). Dynamic L^p-hedging in discrete time under cone constraints. *SIAM Journal on Control and Optimization*. **(38)**. pp. 665-682

[32] Rémillard, B., Hocquard, A., and Papageorgiou, N. (2010). Option pricing and dynamic discrete time hedging for regime-switching geometric random walks models. Working paper series, HEC Montréal.

[33] Rémillard, B., Hocquard, A., Langlois, H., Papageorgiou, N. (2012). Optimal hedging of American Options in discrete time. *Numerical Methods in Finance, Springer Proceeding in Mathematics*. **(12)**. pp. 145-170

[34] Rémillard, B., Rubenthaler, S. (2013). Optimal hedging in discrete time. *Quantitative Finance*. **(13)**. pp. 819-825.

[35] Schweizer, M. (1991). Option hedging for semimartingales. *Stochastic Processes and their Applications*. **(37)**. pp. 339-363

[36] Schweizer, M. (1995). Variance-optimal hedging in discrete time. *Mathematics of Operation Research*. **(20)**. pp. 1-32

[37] Sekine, J. (2004). Dynamic minimization of worst conditional expectation of shortfall. *Mathematical Finance*. **(14)**. pp. 605-618

[38] Xu, M. (2006). Risk measure pricing and hedging in incomplete markets. *Annals of Finance*. **(2)**. pp. 51-71

Chapitre 3

Couverture avec des coûts de transaction

Minimizing CVaR in global dynamic hedging with transaction costs [1]

By Frédéric Godin [2]

Abstract

This study develops a global derivatives hedging methodology which takes into account the presence of transaction costs. It extends the Hodges & Neuberger (1989) framework in two ways. First, to reduce the occurrence of extreme losses, the expected utility is replaced by the conditional Value-at-Risk (CVaR) coherent risk measure as the objective function. Second, the normality assumption for the underlying asset returns is relaxed : general distributions are considered to improve the realism of the model and to be consistent with fat tails observed empirically. Dynamic programming is used to solve the hedging problem. The CVaR minimization objective is shown to be part of a time-consistent framework. Simulations with parameters estimated from the S&P 500 financial time series show the superiority of the proposed hedging method over multiple benchmarks of the literature in terms of tail risk reduction.

JEL classification : G32, C61

Keywords : Hedging, risk management, risk measures, Value-at-Risk, Tail Value-at-Risk, Conditional Value-at-Risk, dynamic programming, transaction costs.

---

1. I thank Geneviève Gauthier for her helpful feedback. Financial support from NSERC is gratefully acknowledged.

2. Ph.D. student at HEC Montréal, Department of Management Sciences, 3000 ch. Côte-Ste-Catherine, Montréal, Québec, Canada, H3T 2A7, frederic.godin@hec.ca.

3.1 Introduction

The complete market paradigm has become widespread across the financial literature since the development of the Black & Scholes (1973) option pricing framework and the seminal work of Harrison & Pliska (1981). The possibility of perfectly replicating contingent claims in this context leads to a natural procedure for pricing and hedging derivatives. However, several properties of financial markets prevent achieving exact replication in practice, which results in market incompleteness. Amongst them are transaction costs, discrete-time portfolio rebalancing and jumps in stock price paths. To improve the performance and accuracy of financial models, it is desirable that they take these features into account.

In incomplete markets, the total removal of risk can be achieved through super-replication. However, this procedure often requires an amount of capital that is prohibitively large and alternative hedging methods must therefore be considered. [3] Most of the alternative methods can be classified into the local hedging and global hedging subcategories.

Local hedging procedures can be move-based or time-based. Delta-hedging falls into the first category since it neutralizes the risks associated with a portfolio for small movements of the underlying asset price. Time-based local hedging strategies involve minimizing the risk associated with a portfolio until the next time step. Local variance minimization, as suggested by Ederington (1979), falls into this category.

A flaw associated with local hedging procedures is their myopia ; they only consider outcomes for small increments of price or time and disregard the results over the full hedging period. The global hedging methodology, which aims at optimizing the risk associated with the terminal hedging error, does not share this drawback. This approach is therefore pursued in the current study.

For global hedging procedures, several objective functions are proposed in the literature. Some papers choose to maximize the expected utility of the hedger, as in Hodges & Neuberger (1989). This leads to the identification of an optimal trade-off between risk and return. Another possibility is to minimize the risk. Global quadratic hedging, pioneered by Schweizer (1995), follows this objective and attempts an optimal replication of derivatives. A drawback

3. For example, Jacod & Eberlein (1997) show that when the underlying asset follows a purely discontinuous process, the super-replication portfolio price for a European call option is generally the price of the underlying asset itself.

of using the quadratic penalty function is that it sanctions both gains and losses. Föllmer & Leukert (2000) and François et al. (2014) suggest using general expected penalties as the objective function, including the semi-quadratic penalty that only penalizes losses and therefore does not share the pitfall associated with the quadratic penalty. Other studies use objective functions which focus on the tail of the hedging loss distribution. Föllmer & Leukert (1999) propose the quantile hedging scheme which maximizes the probability that the hedging loss does not exceed a certain threshold. Under certain conditions, this problem is equivalent to the minimization of the Value-at-Risk (VaR) of the hedging loss. Several pitfalls are associated with the use of such an objective. The most severe one is that the VaR completely disregard the most extreme losses i.e. those which occur with a probability lesser than the VaR confidence level. Such a feature is not consistent with the prime concern in hedging which is to avoid extreme losses. Using the CVaR instead of the VaR as the objective function allows the magnitude of the most extreme losses to be taken into account.

The CVaR risk measure described in Rockafellar & Uryasev (2002) possesses many desirable properties and resolves many problems associated with the use of the VaR risk measure. The CVaR averages all worst-case losses beyond a certain threshold and therefore considers all extreme losses. Furthermore, it is a coherent risk measure, which makes it convenient to use in risk management frameworks. Because of its favorable properties, the CVaR risk measure is used in the current paper to measure risk. Several papers use this risk measure in global risk management problems. The first to propose a dynamic global CVaR minimization scheme were Boda & Filar (2006) ; they minimize the CVaR of the terminal value of an investment portfolio. Melnikov & Smirnov (2012) minimize the terminal hedging error's CVaR in continuous-time by adapting a technique applied in statistical hypothesis testing previously used in Föllmer & Leukert (1999) for quantile hedging. CVaR minimization is also studied in the context of static hedging with multiple assets by Alexander et al. (2003).

Boda & Filar (2006) raise the issue that using a sequence of conditional CVaR risk measures as objective functions for the hedging procedure yields time-inconsistency, which means that the optimal solution identified initially becomes suboptimal with respect to the sequence's other risk measures as time passes and additional information is received. Fortunately, the CVaR risk-measure can be included in a time-consistent framework, as shown herein.

Considering transaction costs in a global hedging framework is important because this feature reduces the efficacy of high-frequency portfolio rebalancing procedures which are typical of many hedging schemes found in the literature. The seminal paper by Hodges & Neuberger (1989) tackles the problem of hedging in the presence of transaction costs by presenting a continuous-time global hedging framework where the hedger maximizes its utility by offsetting a derivatives payoff with a self-financing hedging portfolio. Hodges & Neuberger (1989) assume that the underlying asset follows a geometric Brownian motion. The Hamilton-Jacobi-Bellman partial differential equation is solved to obtain the optimal hedging strategy.

Various authors propose approximations or simplifications of the solution to the problem defined in Hodges & Neuberger (1989), still in the context of assets being represented by a geometric Brownian Motion. For instance, Zakamouline (2006) proposes a parametric representation of the no-transaction region in the presence of proportional transaction costs. The no-transaction region is an interval surrounding the option delta such that it is optimal not to rebalance the portfolio when its delta lies within this interval. Whalley & Wilmott (1997) and Barles & Soner (1998) provide an asymptotic description of the no-transaction region when transaction fees are small. A strategy described in Martellini & Priaulet (2002) assumes a constant width for the no-transaction region.

The normality assumption for log-returns of the underlying asset in Hodges & Neuberger (1989) can be improved. It is well documented fact that returns of financial markets are often non-Gaussian and exhibit semi-heavy tails ; for example, see Fama (1965) and Eberlein & Keller (1995). Allowing fat tail distributions instead of the Gaussian distribution for the underlying asset returns adds realism to the model and better represents the reality of financial markets. Global hedging problems in the absence of transaction costs have been solved for general models of the underlying asset allowing for the presence of regimes and jumps (Rémillard & Rubenthaler, 2013). Log-returns under such models exhibit fail tails, and such a feature should also be applied to hedging procedures that consider transaction costs.

We modify the Hodges & Neuberger (1989) framework for global hedging in the presence of transaction costs and incorporate two novel features into their model. First, the expected utility objective function is replaced by the CVaR risk measure, which is better suited for risk management purposes since it tracks extreme losses. The second added feature is to allow non-Gaussian asset returns that better match the fat tails observed on financial market return data. To keep our model tractable for general return distributions, the model is developed in discrete-time. Another contribution of the current paper is to show that the CVaR minimization objective can be incorporated into a time-consistent framework. Finally, the performance of the hedging method presented here is compared to the performance achieved by several literature benchmarks through simulation experiments.

The paper is structured as follows. Section 3.2 describes the market in which the current work takes place and defines the hedging problem. A rigorous proof guaranteeing the existence of a solution and an algorithm allowing its computation are then given. The relationship between the time-consistency property and the CVaR minimization objective is discussed in Section 3.3. Section 3.4 illustrates the application of the proposed global hedging scheme in a numerical example and compares its performance to several literature benchmarks. Section 3.5 concludes.

3.2 The global hedging methodology

3.2.1 Market setup

The market is in discrete time and time steps are denoted by $t = 0, 1, \ldots, T$. There are two liquid and tradable assets on the market. The first one is a deterministic risk-free asset and its non-decreasing price at time t is denoted by B_t. The second asset is risky and stochastic. Its non-negative price at time t is denoted by S_t. The filtration $\mathcal{F}$ driving the information on the market contains all observations about past asset prices : $\mathcal{F}_t = \sigma\left(S_u | u \leq t\right)$.

Consider a European-type derivative maturing at time T whose payoff is given by $C_T = C(S_T)$. A financial institution, referred to as the hedger, seeks to hedge this derivative with a self-financing investment strategy involving the traded assets. A hedging strategy is denoted by $\theta = \theta_{1:T}$,[4] where $\theta_t = \left(\theta_t^{(B)}, \theta_t^{(S)}\right)$ represents the number of shares of each respective asset contained in the portfolio within the time interval $(t-1, t]$ for $t \geq 1$. θ_0 is the initial allocation of assets in the portfolio at time $t = 0$ before the first trade is applied. Transaction costs K_t incurred at time t for trades on the risky asset are composed of a fixed and a variable component proportional to the total transaction amount :

$$K_t = k_1 \mathbb{I}_{\{\theta_{t+1}^{(S)} \neq \theta_t^{(S)}\}} + k_2 |\theta_{t+1}^{(S)} - \theta_t^{(S)}| S_t$$

4. The shorthand notation $\theta_{t_1:t_2} = \{\theta_t | \, t = t_1, \ldots, t_2\}$ is used.

with $k_1, k_2 \geq 0$.[5] For any self-financing hedging strategy θ,[6] the portfolio value at time t (before transaction costs at time t are incurred) is given by :

$$\begin{aligned} V_t^{(\theta)} &= V_0 + \sum_{i=1}^{t} \theta_i^{(B)}(B_i - B_{i-1}) + \theta_i^{(S)}(S_i - S_{i-1}) - K_{i-1} \\ &= \frac{B_t}{B_{t-1}}(V_{t-1}^{(\theta)} - K_{t-1}) + \theta_t^{(S)}\left(S_t - S_{t-1}\frac{B_t}{B_{t-1}}\right) \end{aligned} \tag{3.1}$$

where the initial portfolio value V_0 is a given constant.[7] The portfolio is completely invested in the risk-free asset at the initial time : $\theta_0^{(S)} = 0$ and $\theta_0^{(B)} = \frac{V_0}{B_0}$.

The set of admissible strategies, Θ, contains all self-financing trading strategies that the hedger is allowed to use. Trading strategies must be predictable : θ_t is $\mathcal{F}_{t-1}-$measurable for all t. The current paper uses the following admissible strategies :

$$\Theta = \left\{\mathcal{F}\text{-predictable, self-financing } \theta \text{ that satisfy } \forall t,\ \theta_t^{(S)} \in \mathcal{B}_t\right\} \tag{3.2}$$

where $\mathcal{B}_t \subseteq \mathbb{R}$. Sets $\mathcal{B}_t$ allow constraints to be put on trading strategies. For example, using $\mathcal{B}_t = [0, \infty)$ prevents short sales.

To avoid path-dependence issues in the hedging framework to be described in a subsequent section below, the risky asset price process is assumed to be Markovian.

Assumption 3.2.1 *S has the Markov property with respect to the filtration $\mathcal{F}$.*

3.2.2 The global hedging problem

To perform the hedge, the hedger must select a particular hedging strategy from among all admissible strategies. Global hedging procedures minimize a suitable risk measure applied to the terminal hedging error. In the current paper, the CVaR is considered :

$$\min_{\theta \in \Theta} \mathrm{CVaR}_\alpha \left(C_T - V_T^{(\theta)}\right). \tag{3.3}$$

The CVaR is defined and discussed in Rockafellar & Uryasev (2002). The CVaR is shown by Kusuoka (2001) to be the smallest law-invariant coherent risk measure that is larger than the VaR.

5. Transaction fees of the form $K_t = k_1 \mathbb{I}_{\{\theta_t^{(S)} \neq \theta_{t-1}^{(S)}\}} + k_2 |\theta_t^{(S)} - \theta_{t-1}^{(S)}|$ whose variable component is proportional to the number of shares traded could also be considered.

6. A hedging strategy θ is self-financing if $\forall t$, $S_t\theta_{t+1}^{(S)} + B_t\theta_{t+1}^{(B)} + K_t = S_t\theta_t^{(S)} + B_t\theta_t^{(B)}$.

7. Since $V_t^{(\theta)}$ does not depend on $\theta_{t+1:T}$, the notation $V_t^{(\theta_{1:t})} = V_t^{(\theta)}$ is used.

Defining the CVaR requires the definition of the VaR. The VaR of any random variable Y corresponds to a quantile of its distribution :

$$\begin{aligned}
\mathrm{VaR}_\alpha(Y) &= \min_y \{y \,|\, \mathbb{P}[Y \leq y] \geq \alpha\} \\
&= \min_y \{y \,|\, \mathbb{E}[\mathbb{I}_{\{Y \leq y\}}] \geq \alpha\}. \qquad (3.4)
\end{aligned}$$

The value of α is typically 0.95 or 0.99. The CVaR of an integrable random variable Y is defined as the expectation of its conditional tail distribution whose cumulative distribution function (cdf) is $h^{(\alpha)}(y)$:

$$\mathrm{CVaR}_\alpha(Y) = \int_{\mathbb{R}} y \, dh^{(\alpha)}(y), \quad \text{where} \qquad (3.5)$$

$$h^{(\alpha)}(y) = \begin{cases} \frac{\mathbb{P}[Y \leq y] - \alpha}{1-\alpha} \text{ if } y \geq \mathrm{VaR}_\alpha(Y), \\ 0 \text{ otherwise.} \end{cases} \qquad (3.6)$$

Rockafellar & Uryasev (2002) give the following useful characterization of the CVaR :

Theorem 3.2.1 *Define the following auxiliary function :*

$$f_{c,\alpha}^{(CVaR)}(Y) = c + \frac{1}{1-\alpha}(Y - c)\mathbb{I}_{\{Y > c\}}. \qquad (3.7)$$

Then the CVaR *at level* α *of* Y *is given by*

$$\mathrm{CVaR}_\alpha(Y) = \min_{c \in \mathbb{R}} \mathbb{E}[f_{c,\alpha}^{(CVaR)}(Y)]. \qquad (3.8)$$

The existence of such a representation is helpful since the problem of dynamically minimizing the expectation of some random variable is well understood and can be solved with the Bellman Equation of dynamic programming.

3.2.3 Solving the hedging problem

This section describes the conditions guaranteeing the existence of a solution to problem (3.3). The theoretical algorithm allowing for the computation of the solution is also described. All proofs are presented in Appendix 3.6.

Equation (3.8) provides an alternative representation of the CVaR minimization problem, which is also used in Boda & Filar (2006) :

$$\begin{aligned}
\inf_{\theta \in \Theta} \mathrm{CVaR}_\alpha(C_T - V_T^{(\theta)}) &= \inf_{\theta \in \Theta} \min_{c \in \mathbb{R}} \mathbb{E}\left[f_{c,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)})\right] \qquad (3.9) \\
&= \inf_{c \in \mathbb{R}} \inf_{\theta \in \Theta} \mathbb{E}\left[f_{c,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)})\right]. \qquad (3.10)
\end{aligned}$$

Solving (3.9) therefore requires solving the problem

$$\inf_{\theta\in\Theta} \mathbb{E}\left[f_{c,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)})\right] \tag{3.11}$$

for different values of c. Lemma 3.2.1 provides sufficient conditions for the minimum to be attained in (3.11). Lemma 3.2.2 provides the algorithm to compute the minimum and minimizing arguments. Those Lemmas are adapted from François *et al.* (2014) to take into account transaction costs. The minimum attainable expected penalty conditioned upon available information at time t is

Definition 3.2.1

$$\begin{aligned} &\Psi_{T,c}^{*,CVaR}(\theta) = f_{c,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)}) \\ \forall t < T,\ &\Psi_{t,c}^{*,CVaR}(\theta_{1:t}) = \inf_{\theta_{t+1:T}}\left\{\mathbb{E}\left[f_{c,\alpha}^{(CVaR)}(C_T - V_T^{(\theta_{1:t},\theta_{t+1:T})})\middle|\mathcal{F}_t\right] : (\theta_{1:t},\theta_{t+1:T}) \in \Theta\right\} \end{aligned} \tag{3.12}$$

where $f_{c,\alpha}^{(CVaR)}$ is defined by (3.7).

The function Ψ^* is known as the value function within the dynamic programming terminology. The following lemma states certain properties of the value function :

Lemma 3.2.1 *Assume that*

$$\forall t,\ \mathcal{B}_t \text{ are compact,} \tag{3.13}$$

$$C_T \text{ and } S_t \text{ are integrable for all } t. \tag{3.14}$$

Then for all $t \leq T$, $\theta_{1:t}$,

- *If $t < T$, the infimum in (3.12) is attained (the infimum is a minimum).* (3.15)
- *There exist functions $\Psi_{t,c}^{CVaR}$ such that $\Psi_{t,c}^{*,CVaR}(\theta_{1:t}) = \Psi_{t,c}^{CVaR}(S_t, \theta_t, V_t^{(\theta_{1:t})})$* (3.16)
- *$\Psi_{t,c}^{CVaR}(S_t, \vartheta, v)$ is continuous with respect to (c, ϑ, v)* (3.17)

Remark 3.2.1 *Equation (3.16) indicates that given $(\theta_t, V_t^{(\theta_{1:t})})$, $\Psi_{t,c}^{*,CVaR}(\theta_{1:t})$ does not depend on $\theta_{1:t-1}$.*

The compactness of sets $\mathcal{B}_t$ implied by (3.13) is not restrictive in practice because financial institutions do not have unlimited access to credit. The following theorem, known as the Bellman Equation, gives a recursive scheme to compute the value function at all time steps and identify the optimal hedging strategy that solves the hedging problem (3.9).

Lemma 3.2.2 (The Bellman Equation) *Assume (3.13)-(3.14) hold. Define*

$$SF_t = \left\{ (\theta_t^{(B)}, \theta_t^{(S)}) \,|\, \theta_t^{(S)} \in \mathcal{B}_t \text{ and } S_{t-1}\theta_t^{(S)} + B_{t-1}\theta_t^{(B)} + K_{t-1} = V_{t-1}^{(\theta_{1:t-1})} \right\}. \tag{3.18}$$

Then,

$$\Psi_{t,c}^{CVaR}(S_t, \theta_t, V_t^{(\theta_{1:t})}) = \min_{\theta_{t+1} \in SF_{t+1}} \mathbb{E}\left[\Psi_{t+1,c}^{CVaR}(S_{t+1}, \theta_{t+1}, V_{t+1}^{(\theta_{1:t}, \theta_{t+1})}) \Big| S_t, \theta_t, V_t^{(\theta_{1:t})} \right]. \tag{3.19}$$

Moreover, define for all t,

$$\theta_{t+1}^*(\theta_{1:t}) = \underset{\theta_{t+1} \in SF_{t+1}}{\arg\min}\ \mathbb{E}\left[\Psi_{t+1,c}^{CVaR}(S_{t+1}, \theta_{t+1}, V_{t+1}^{(\theta_{1:t}, \theta_{t+1})}) \Big| S_t, \theta_t, V_t^{(\theta_{1:t})} \right], \tag{3.20}$$

which is $(S_t, \theta_t, V_t^{(\theta_{1:t})})$-measurable.[8] *Then $\theta^* = \left(\theta_1^*, \ldots, \theta_T^*(\theta_{1:T-1}^*)\right)$ solves (3.11).*

Lemmas 3.2.1 and 3.2.2 lead to the following characterization of the solution to problem (3.9).

Theorem 3.2.2 *If (3.13)-(3.14) hold, then*

$$\min_{\theta \in \Theta} \mathrm{CVaR}_\alpha \left(C_T - V_T^{(\theta)} \right) = \min_{c \in \mathbb{R}} \Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0) = \Psi_{0,c^*}^{CVaR}(S_0, \theta_0, V_0) \tag{3.21}$$

for some $c^* \in \mathbb{R}$. Furthermore, there exists a solution $\theta^{*,CVaR}$ to (3.21) which also solves

$$\theta^{*,CVaR} \in \underset{\theta \in \Theta}{\arg\min}\ \mathbb{E}\left[f_{c^*,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)}) \right]. \tag{3.22}$$

Theorem 3.2.2 indicates that minimizing the hedging CVaR involves calculating $\Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0)$ for multiple values of c. However, the following result indicates that it is sufficient to evaluate that function for a single c but for multiple V_0, which can be done through a single run of the dynamic program solving (3.11) :

Theorem 3.2.3 $\forall x, c \in \mathbb{R}$, $\Psi_{0,x}^{CVaR}(S_0, \theta_0, V_0) = x - c + \Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0 + (x-c)\frac{B_0}{B_T})$.

3.3 Time-consistency of the CVaR

Solutions to problem (3.3) minimize the CVaR measure at the initial time step. However, it is relevant to ask whether or not those solutions will remain optimal through time, in the sense that they continue minimizing a similar version of the same risk measure at further

8. If many θ_{t+1} solve (3.20), any of the solutions can be used in defining $\theta_{t+1}^*(\theta_{1:t})$.

time steps. Indeed, one may argue that there is no point in identifying an optimal solution at a particular point in time if it becomes non-optimal later on and is therefore not pursued. Boda & Filar (2006) investigate this question by introducing the time-consistency principle for sequences of risk measures. Informally, a sequence of risk measures is said to be time-consistent if the solution obtained by minimizing one of the risk measures also minimizes all other risk measures of the sequence, each one associated with one of the time steps. The current section recalls the definition of time-consistency found in Boda & Filar (2006). It is also shown how CVaR minimizing strategies can be incorporated in a sequence of consistent risk measures.

3.3.1 Defining time-consistency of risk measures

Definition 3.3.1 *A $\mathcal{F}_t-$risk measure ρ_t is a mapping that takes a random variable as input and outputs a $\mathcal{F}_t-$measurable random variable.*

Definition 3.3.2 *A random variable Y is the minimal element of a set $\mathbb{Y}$ if $\forall \tilde{Y} \in \mathbb{Y}$, $Y \leq \tilde{Y} a.s.$*

Definition 3.3.3 *The projection operator $\mathcal{P}_{i,j}$ selects a subset of any trading strategies :*

$$\forall \theta \in \Theta, \mathcal{P}_{i,j}(\theta) = (\theta_i, \dots, \theta_j) \,. \tag{3.23}$$

Let $\{\rho_t\}_{t=0}^{T-1}$ be a sequence of $\mathcal{F}_t-$risk measures (each one can be thought of as a hedging objective at time t). The hedger therefore seeks the solutions to the following set of problems (one problem for each t and past trading positions $\theta_{1:t}$) :

$$\min_{\theta_{t+1:T}} \rho_t \left(V_T^{(\theta)} - C_T \right), \quad \theta \in \Theta, \theta_{1:t} \text{ given.} \tag{3.24}$$

Denote the set of solutions to problems (3.24) by $\tilde{\Theta}_t^*(\theta_{1:t})$ and define the following set of optimal trading strategies :

$$\Theta_t^* = \{\theta_{1:t} \times \tilde{\Theta}_t^*(\theta_{1:t}) | \, \exists \theta \in \Theta \text{ such that } \mathcal{P}_{1,t}(\theta) = \theta_{1:t}\} \tag{3.25}$$

where $\times$ is the cartesian product. Therefore, $\Theta_t^* \subseteq \Theta$ for all t.

The following defines the time-consistency property for a set of risk measures. The definition is based on Boda & Fillar (2006), but adapted to account past policies which can influence endogenous variables. They base their definition on the principle of optimality of dynamic programming (see Bellman & Dreyfus, 1962).

Definition 3.3.4 *The sequence of risk measures* $\{\rho_t\}_{t=0}^{T-1}$ *is said to be* time-consistent *if all the following conditions are satisfied :*

- $\Theta_0^* \neq \varnothing$,
- $\forall t$, $\forall \theta^* \in \Theta_0^*$, $\theta^* \in \tilde{\Theta}_t^*(\mathcal{P}_{1,t}(\theta^*))$,
- *If* $\theta^* = (\theta_1^*, \ldots, \theta_T^*)$ *is any hedging strategy which satisfies*

$$\forall t,\, \theta_{t+1} \in \underset{\theta_{t+1}}{\arg\min} \left\{ \rho_t \left(V_T^{(\theta^*_{1:t}, \theta^*_{t+1}, \theta^*_{t+2:T})} - C_T \right) \middle| (\theta^*_{1:t}, \theta_{t+1}, \theta^*_{t+2:T}) \in \Theta \right\},$$

then $\theta^* \in \Theta_0^*$.

3.3.2 Time-consistency and the CVaR

This section defines a conditional version of the CVaR. Furthermore, the current paper demonstrates that CVaR minimizing solutions can be encompassed in a time-consistent framework.

Define the conditional VaR and CVaR, which are $\mathcal{F}_t-$risk measures, by

$$\begin{aligned} \mathrm{VaR}_{t,\alpha}(Y) &= \min_y \{ y \,|\, \mathbb{P}[Y \leq y | \mathcal{F}_t] \geq \alpha \}, && (3.26) \\ \mathrm{CVaR}_{t,\alpha}(Y) &= \int_{\mathbb{R}} y \, dh^{(t,\alpha)}(y), \quad \text{where} && (3.27) \\ h^{(t,\alpha)}(y) &= \begin{cases} \frac{\mathbb{P}[Y \leq y | \mathcal{F}_t] - \alpha}{1-\alpha} \text{ if } y \geq \mathrm{VaR}_{t,\alpha}(Y), \\ 0 \text{ otherwise.} \end{cases} && (3.28) \end{aligned}$$

Boda & Fillar (2006) give a counter-example to show that the sequence of $\mathcal{F}_t-$risk measures

$$\{\rho_t\} = \{\mathrm{CVaR}_{t,\alpha}\} \tag{3.29}$$

is not time-consistent in general. Their example, in an investment framework, only considers open loop solutions i.e. solutions where the portfolio choice only depends on information $\mathcal{F}_0$. An example in a hedging context showing that the set of conditional CVaRs is not time-consistent in general even when feedback policies are considered is given in Appendix 3.8.

However, assume (3.13)-(3.14) hold and let $c^*(\alpha) = \underset{c \in \mathbb{R}}{\arg\min} \Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0)$. Then, from the optimality principle of dynamic programming, the sequence of $\mathcal{F}_t-$risk measures

$$\{\rho_t\} = \left\{ \mathbb{E}\left[f^{(CVaR)}_{c^*(\alpha),\alpha}(\bullet) | \mathcal{F}_t \right] \right\} \tag{3.30}$$

is time-consistent [9], which leads to the following result :

Corollary 3.3.1 *The sequence characterized by*

$$\rho_t := \begin{cases} \mathrm{CVaR}_{0,\alpha} \ \text{if } t = 0, \\ \mathbb{E}\left[f^{(CVaR)}_{c^*(\alpha),\alpha}(\bullet)|\mathcal{F}_t\right] \ \text{if } t > 0 \end{cases} \tag{3.31}$$

is time-consistent.

Proof of Corollary 3.3.1 : From Theorem 3.2.2

$$\arg\min_{\theta\in\Theta} \mathrm{CVaR}_\alpha\left(C_T - V_T^{(\theta)}\right) = \arg\min_{\theta\in\Theta} \mathbb{E}\left[f^{(CVaR)}_{c^*(\alpha),\alpha}\left(C_T - V_T^{(\theta)}\right)\right],$$

which means that sequence (3.30) remains time-consistent if its first element is substituted by $\mathrm{CVaR}_{0,\alpha}$. ■

In summary, the hedging strategy which minimizes the hedging CVaR_α will not necessarily minimize its conditional version of the $\mathrm{CVaR}_{t,\alpha}$ at later time steps $t > 0$. However, it will minimize the expected penalties $\mathbb{E}\left[f^{(CVaR)}_{c^*(\alpha),\alpha}(\bullet)|\mathcal{F}_t\right]$ at all further time steps. This time-consistency property of the sequence (3.31) illustrates that CVaR minimizing solutions of hedging problems can be incorporated in a time-consistent framework and can therefore be considered suitable for hedging, this even if the sequence (3.29) is not time-consistent in general.

3.4 A numerical example

In this section, the performance of hedging procedures obtained by solving the global hedging problem (3.3) is assessed by comparing them with benchmarks on simulated data.

3.4.1 Market specifications

An at-the-money European call option ($C_T = \max[0, S_T - K]$) is hedged with a portfolio rebalanced at every time step, each one corresponding to a week. The maturity of the option is $T = 12$ weeks. The initial price of the stochastic asset and the option strike price are both $S_0 = K = 1000$. The two tradable assets have the following dynamics :

$$B_t = e^{rt/52} \tag{3.32}$$

$$S_t = S_0 e^{\sum_{k=1}^t z_k} \tag{3.33}$$

9. See Theorem 3.2 in François et al. (2014) for a proof.

where $r = 2\%$ is the annualized risk-free rate and where z is a strong white noise. The Gaussian distribution might not be the most appropriate one to represent financial returns because of the fat tails and the asymmetry they exhibit. The Normal Inverse Gaussian (NIG) distribution is a natural extension of the Gaussian distribution which allows for positive excess kurtosis and skewness. For the current example, log-returns z have the NIG distribution with parameters $(\alpha^{(NIG)}, \beta^{(NIG)}, \delta^{(NIG)}, \mu^{(NIG)})$. A random variable has a NIG$(\alpha, \beta, \delta, \mu)$ distribution if its density function (pdf) is given by

$$f(x) = \frac{\alpha \delta K_1\left(\alpha\sqrt{\delta^2 - (x-\mu)^2}\right)}{\pi\sqrt{\delta^2 - (x-\mu)^2}} e^{\gamma\delta + \beta(x-\mu)}$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$, $\alpha > 0, \beta \in (-\alpha, \alpha), \delta > 0, \mu \in \mathbb{R}$ and K_ν is the modified Bessel function of the second kind :

$$K_\nu(y) = \frac{1}{2}\int_0^\infty u^{\nu-1} e^{-\frac{1}{2}y(u+u^{-1})} du, \quad y > 0.$$

This distribution, introduced by Barndorff-Nielsen (1977), is part of the generalized hyperbolic distribution family. This family has been recently successfully applied to model stock returns : see Eberlein & Keller (1995), Rydberg (1997), Prause (1999) and Rydberg (1999). Barndorff-Nielsen & Halgreen (1977) show the infinite divisibility of this family of distributions, and consequently the existence of Lévy processes having NIG distributed increments.

From a statistical perspective, additional features could be added to the geometric random walk model (3.33) to increase its realism, such as the presence of regimes, a stochastic volatility or autocorrelation of returns. However, this would increase the number of dimensions in the hedging problem, severely increasing the numerical burden associated with its solution and hindering the feasibility of the hedging approach described in this paper.

Under model (3.33), the market is incomplete. Incompleteness is caused by three features. The first is discrete-time rebalancing. The second is the presence of transaction costs. This feature severely penalizes high-frequency portfolio rebalancing which is characteristic of typical derivatives replication algorithms. The third is the risky asset's returns model. The continuous-time version of the NIG geometric random walk (3.33) is the exponential of a pure-jump Lévy process with NIG-distributed increments. Trajectories of the NIG Lévy process contain an infinite number of small jumps, which cause market incompleteness.

To obtain realistic parameters, model (3.33) was estimated on a time series of weekly S&P 500 closing prices between January 1, 2000, and August 20, 2013. The estimation was performed by maximum likelihood with the EM algorithm for NIG random variables described in Karlis (2002). The estimated parameters are presented in Table 3.1.

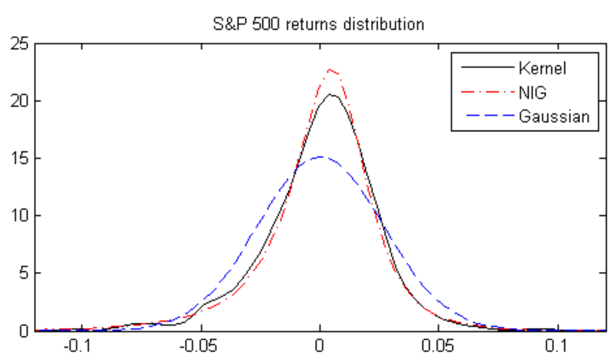
TABLE 3.1 – Estimated parameters for weekly NIG log-returns

	$\alpha^{(NIG)}$	$\beta^{(NIG)}$	$\delta^{(NIG)}$	$\mu^{(NIG)}$
Estimate	35.7	-10.8	2.04×10^{-2}	6.7×10^{-3}
Standard error	6.6	3.7	0.23×10^{-2}	1.6×10^{-3}

Estimated parameters and standard errors for weekly NIG log-returns. Estimation performed by maximum likelihood with the EM algorithm on weekly closing of the S&P 500 between January 1, 2000, and August 20, 2013. Standard errors are obtained by simulation.

Figure 3.4.1 illustrates the kernel estimator of the S&P 500 log-returns distribution and compares it to fitted NIG and Gaussian distributions. This figure shows that the NIG distribution captures the peakedness and asymmetry better than its Gaussian counterpart. Cramer-Von-Mises tests with simulated p-values, which are described in Genest & Rémillard (2008), were run to assess the distribution's goodness-of-fit. The null hypothesis of this test is that log-returns are independent, identically distributed and Gaussian or NIG-distributed. The p-value is smaller than 0.01% for the Gaussian distribution whereas it is 63.2% for the NIG, confirming the superiority of the latter in terms of model adequacy.

FIGURE 3.1 – S&P 500 returns distribution



Distribution of weekly S&P 500 log-returns between January 1, 2000, and August 20, 2013 and comparison with the NIG and Gaussian distribution.

3.4.2 Benchmarks

In this section, the different benchmarks with which the hedging methodology presented herein are compared are given. The strategy solving (3.3) is called MCVaR (minimizing CVaR). The constraint $\theta_t^{(S)} \in [0,1]$ for all t is imposed to prevent short sales and excessive leverage.

Black-Scholes delta-hedging (BSDH)

This benchmark uses delta-hedging where the Black-Scholes formula is used to price the option. It consists in setting

$$\theta_{t+1}^{(S)} = \frac{\partial C_t^{BS}}{\partial S_t} = \Phi\left(\frac{\ln\left(\frac{S_0}{K}\right) + (r/52 + \sigma_{BS}^2/2)\,(T-t)}{\sigma_{BS}\sqrt{T-t}}\right), \tag{3.34}$$

where C_t^{BS} is the Black-Scholes option price at time t and Φ is the standard Gaussian cdf. The weekly volatility used the compute the option price is the standard deviation of the weekly log-returns : $\sigma_{BS} = \frac{\delta^{(NIG)}\alpha^{(NIG)2}}{(\alpha^{(NIG)2}-\beta^{(NIG)2})^{3/2}} = 0.0263$. This hedging method is used by a hedger who presumes that log-returns are normally distributed.

Delta move-based hedging (DMBH)

A strategy described in Martellini & Priaulet (2002) involves rebalancing the hedging portfolio every time its delta gets outside a certain interval surrounding the derivative's delta. The portfolio is then rebalanced so that its delta is brought back to the interval bound; if the portfolio delta is larger (smaller) than the higher (lower) bound of of the interval, it is brought back to the higher (lower) bound. The motivation for such a strategy is that the solution to the Hodges & Neuberger (1989) problem under proportional transaction costs has this form. As in Martellini & Priaulet (2002), the size of the interval b is kept constant through time. In the current study, the selected b is the value that yields the lowest possible hedging error semi-RMSE, as defined by (3.42). This value is identified through simulation.

NIG delta-hedging (NDH)

This benchmark is a version of delta-hedging in which the hedger takes into account that log-returns have the NIG distribution. Because the market is incomplete in that case, several risk-neutral measures are available to price the option (Schoutens, 2003). A natural extension of the Girsanov transform to the case of exponential Lévy processes is the mean-correcting measure, which only shifts the mean of returns in a risk-neutral world. Godin et al. (2012) give the closed-form option price under that risk-measure. The NIG delta-hedging therefore consists in setting

$$\theta_{t+1}^{(S)} = \frac{\partial C_t^{NIG}}{\partial S_t} = 1 - \Phi^{NIG}\left(\ln\left(\frac{K}{S_t}\right); (\alpha^{(NIG)}, \beta^{(NIG)} + 1, \delta^{(NIG)}(T-t), \tilde{\mu}(T-t))\right), \tag{3.35}$$

where C_t^{NIG} is the option price at time t obtained through the mean-correcting measure, Φ^{NIG} is the cdf associated with the NIG distribution and

$$\tilde{\mu} = r/52 + \delta^{(NIG)}\left(\sqrt{\alpha^{(NIG)2} - (\beta^{(NIG)} + 1)^2} - \sqrt{\alpha^{(NIG)2} - \beta^{(NIG)2}}\right).$$

Global quadratic hedging (GQH)

This benchmark uses the semi-explicit formulas found in Schweizer (1995) to obtain the solution to problem $\min_{\theta\in\Theta}\mathbb{E}\left[(C_T - V_T^{(\theta)})^2\right]$. A drawback of the quadratic penalty is that it penalizes gains and losses equally. Semi-explicit formulas accelerate the computation of the solution. However, they do not allow transaction costs to be taken into account.

Global Semi-Quadratic Hedging (GSQH)

This benchmark provides the solution to problem $\min_{\theta\in\Theta}\mathbb{E}\left[g(C_T - V_T^{(\theta)})\right]$ where $g(x) = x^2\mathbb{I}_{\{x>0\}}$. The solution is obtained by the Bellman Equation, as in Lemma 3.2.2, using the terminal condition $\Psi_T(S_T, \theta_T, V_T) = g(C(S_T) - V_T^{(\theta)})$. This benchmark is proposed by François et al. (2014).

Minimizing VaR (MVaR)

Despite of the drawbacks associated with the VaR measure, it can still be used as the objective function to be minimized. The MVaR benchmarks solves the problem

$$\min_{\theta\in\Theta} \mathrm{VaR}_\alpha\left(C_T - V_T^{(\theta)}\right). \tag{3.36}$$

The procedure to obtain the solution to this problem is analogous to the CVaR case. Define

$$\Psi_{0,c}^{*,VaR} = \inf_{\theta\in\Theta}\mathbb{E}[f_c^{(VaR)}(C_T - V_T^{(\theta)})] \tag{3.37}$$

$$f_c^{(VaR)}(x) = \mathbb{I}_{\{x>c\}}. \tag{3.38}$$

Problem (3.37) [10] can be solved with Lemma 3.2.2 by using the terminal condition $\Psi_{T,c}^{VaR}(\theta) = f_c^{(VaR)}(C_T - V_T^{(\theta)})$. The VaR minimization problem is linked to Problem (3.37) :

Theorem 3.4.1 *If*

$$\forall t,\ \mathcal{B}_t \text{ is compact,} \tag{3.39}$$

$$\text{the infimum is attained by some } \theta \text{ in (3.36) for all } c, \tag{3.40}$$

$$\Psi_{0,c}^{*,VaR} \text{ is continuous with respect to } c, \tag{3.41}$$

then

$$\min_{\theta\in\Theta} \mathrm{VaR}_\alpha\left(C_T - V_T^{(\theta)}\right) = \min \mathbb{C}$$

where $\mathbb{C} = \{c|\Psi_{0,c}^{*,VaR} = 1-\alpha\}$. *Furthermore, trading strategies solving (3.36) are those solving (3.37) with* $c = \min\mathbb{C}$ *:*

$$\arg\min_{\theta\in\Theta} \mathrm{VaR}_\alpha\left(C_T - V_T^{(\theta)}\right) = \arg\min_{\theta\in\Theta}\mathbb{E}[f_{\min\mathbb{C}}^{(VaR)}(C_T - V_T^{(\theta)})]$$

Proofs for these results are found in Appendix 3.7.

Risk-free hedging (RFH)

In this benchmark, the totality of the capital allocated for hedging is invested in the risk-free asset : $\theta_t^{(S)} = 0$ for all t.

3.4.3 Numerical algorithm for global hedging procedures

For the MCVaR, MVaR and GSQH hedging strategy, a numerical algorithm is required to compute the solution of the Bellman Equation. The François et al. (2014) backward induction algorithm based on a set of three-dimensional lattices and splines interpolation is used herein. A modification to their algorithm is applied to account for the presence of transaction costs and the possibility that, for some nodes of the grid, it might be suboptimal to rebalance the portfolio.

10. This problem is studied in continuous-time in Föllmer & Leukert (1999) and is known as quantile hedging.

3.4.4 Performance assessment

This section reports the results of the simulation study performed to compare the MCVaR hedging method to several literature benchmarks. To isolate the impact of the presence of the transaction costs from the choice of hedging methodology, two different experiments are performed : transaction costs are applied only in the second experiment and are set to zero in the first.

In each experiment, 10^6 paths of the underlying asset are simulated, and each hedging method is applied on all paths. The hedging errors $C_T - V_T^{(\theta)}$ are calculated for each path, and descriptive statistics about those errors are provided. Note that the Semi-RMSE of a sample $x_1, \ldots, x_n$ is given by

$$\text{Semi-RMSE} = \sqrt{n^{-1} \sum_{i=1}^{n} x_i^2 \mathbb{I}_{\{x_i > 0\}}} \tag{3.42}$$

The MCVaR and MVaR are calculated at the $\alpha = 95\%$ level. In the second experiment, proportional transaction fees are applied : $k_1 = 0$ and $k_2 = 1\%$.[11] The initial portfolio value is set to the Black-Scholes price of the option : $V_0 = 38.63$. This choice is made to give a chance to delta-hedging to perform properly. However, our method does not prescribe this choice and other initial portfolio values could have been chosen.

Simulation without transaction costs

In this experiment, no transaction fees are applied, and the optimal hedging solutions are computed by taking that assumption into account. Table 3.2 gives the simulation results.

Results show that the hedging performance is greatly affected by the hedging methodology (which includes the choice of the risk measure to minimize in global hedging schemes). As expected the MCVaR gives the lowest hedging CVaR, achieving the best risk reduction in tails at both the 95% and 99% levels. It reduces the $\text{CVaR}_{95\%}$ by 5.1% with respect to GSQH, the best benchmark. The MCVaR also exhibits the third best semi-RMSE behind the GQH and GSQH, meaning that the MCVaR also has a respectable performance outside the tail. The GSQH also performs well ; it is the second best at minimizing the CVaRs and as expected the best for the semi-RMSE minimization. Delta-hedging-based procedures (BSDH

11. The case of proportional transaction costs under the Hodges & Neuberger framework is studied in Whalley & Wilmott (1997), Barles & Soner (1998), Davis et al. (1993) and Zakamouline (2006).

and NDH) perform worse than the global MCVaR, GSQH and GQH procedures in terms of both the CVaRs and the semi-RMSE. It is somewhat surprising that the NDH gives higher CVaRs and semi-RMSE than the BSDH, as the former takes into account that returns are NIG-distributed while the latter assumes that returns are Gaussian.

TABLE 3.2 – Descriptive statistics on simulated hedging errors without transaction costs.

Model	MCVaR	MVaR	BSDH	NDH	GQH	GSQH	RFH
Mean	-1.436	-1.501	-1.536	**-1.547**	-1.467	-1.455	-0.6814
RMSE	15.52	18.75	14.40	14.91	**13.78**	14.26	53.92
Semi-RMSE	10.74	15.76	11.19	11.79	10.34	**9.996**	45.93
$\text{VaR}_{95\%}$	21.00	**18.86**	23.72	25.28	21.72	20.64	112.1
$\text{VaR}_{99\%}$	**36.76**	82.84	48.72	51.62	44.18	41.34	178.3
$\text{CVaR}_{95\%}$	**32.10**	57.01	39.65	41.96	36.07	33.84	153.0
$\text{CVaR}_{99\%}$	**56.58**	120.0	67.87	71.24	61.69	57.52	215.0

Descriptive statistics on 10^6 simulated hedging errors $C_T - V_T$ for each hedging method. Transaction fees are not applied in simulations. Best results are in boldface. MCVaR and MVaR minimize the 95% CVaR and 95% VaR of hedging errors. BSDH and NDH are delta-hedging procedure using the Black-Scholes price and the NIG price with the mean-correcting measure, respectively. GQH and GSQH minimize the RMSE and Semi-RMSE of hedging errors, respectively. RFH invests the totality of the hedging capital in the risk-free asset.

A striking result is the atrocious performance of the MVaR method. It is successful at minimizing the VaR at 95%, which is its objective. However, the CVaRs and the semi-RMSE are much bigger than for all other benchmarks (except the no-hedging case). This can be explained by the fact that the MVaR encourages gambling when it expects to be unsuccessful ; when $V_{T-1} \ll C(S_{T-1})$, using a $\mathbb{I}_{\{C_T - V_T^{(\theta)} > c\}}$ penalty encourages increasing the volatility of the hedging portfolio to maximize the probability that $C_T - V_T^{(\theta)} < c$. In those cases, extremely high losses become highly probable. Such a phenomenon is contrary to what a hedger tries to achieve : minimizing extreme losses.

The expected hedging error are close for all hedging methods, except it is slightly higher for the RFH (which involves no hedging). This implies a negative cost for hedging a short position on the option.

Simulation with transaction costs

The first step consists in understanding the importance of taking into account transaction costs in global hedging strategies. To achieve this objective, global strategies MCVaR, MVaR and GSQH are computed twice ; first with transaction fees set to their real value and then by setting fees to zero. MCVaR*, MVaR* and GSQH* denote the strategies where transaction fees were set to zero in the optimization scheme of Lemma 3.2.2. All those strategies are then applied to simulations with the presence of transaction fees. The MCVaR, MVaR and GSQH methods are therefore aware that transaction fees occur whereas MCVaR*, MVaR* and GSQH* are not. The performance of both sets of strategies are compared in Table 3.3.

Table 3.3 – Simulated hedging errors, taking vs not taking into account transaction costs.

Model	MCVaR	MCVaR*	MVaR	MVaR*	GSQH	GSQH*
Mean	7.992	11.00	8.394	16.23	**7.764**	12.12
RMSE	18.84	19.50	21.64	32.98	**17.51**	19.18
Semi-RMSE	17.39	18.88	20.54	32.71	**16.40**	18.82
$\text{VaR}_{95\%}$	32.68	33.34	**29.87**	75.92	33.63	37.37
$\text{VaR}_{99\%}$	**48.73**	60.37	91.78	139.1	54.25	60.59
$\text{CVaR}_{95\%}$	**43.50**	51.20	67.38	113.2	46.96	52.16
$\text{CVaR}_{99\%}$	**68.63**	87.77	127.2	169.4	71.20	78.15

Descriptive statistics on 10^6 simulated hedging errors $C_T - V_T$ for each hedging method. Transaction fees are applied in simulations. MCVaR, MCVaR*, MVaR and MVaR* use the $\alpha = 95\%$ confidence level. GSQH and GSQH* minimizes the hedging error Semi-RMSE. MCVaR, MVaR and GSQH take into account that transaction costs will be incurred, while MCVaR*, MVaR* and GSQH* do not.

We notice that not taking transaction costs into account in global hedging procedures provokes a systematic deterioration of the hedging performance (for example, the MCVaR reduces the $\text{CVaR}_{95\%}$ by 15% when compared to the MCVaR*). For all presented metrics in Table 3.3, the hedging version which takes transaction costs into account outperforms their counterpart that does not consider those costs.

As a second step, all hedging benchmarks are compared with the global hedging strategies that take transaction costs into account. The results are presented in Table 3.4.

TABLE 3.4 – Descriptive statistics on simulated hedging errors with transaction costs.

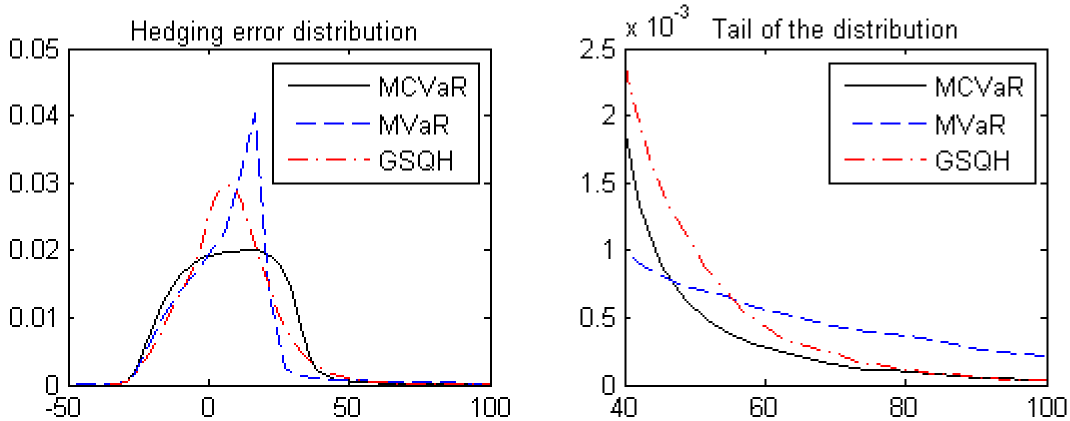
Model	MCVaR	MVaR	BSDH	DMBH	NDH	GQH	GSQH	RFH
Mean	7.992	8.394	13.02	7.575	13.51	12.52	7.764	**-0.6814**
RMSE	18.84	21.64	20.13	18.06	21.01	19.28	**17.51**	53.92
Semi-RMSE	17.39	20.54	19.96	16.84	20.88	19.06	**16.40**	45.93
$\text{VaR}_{95\%}$	32.68	**29.87**	40.74	35.12	43.20	38.21	33.63	112.1
$\text{VaR}_{99\%}$	**48.73**	91.78	66.51	56.10	70.24	61.76	54.25	178.3
$\text{CVaR}_{95\%}$	**43.50**	67.38	57.19	48.57	60.33	53.19	46.96	153.0
$\text{CVaR}_{99\%}$	**68.63**	127.2	85.96	72.81	89.99	79.49	71.20	215.0

Descriptive statistics on 10^6 simulated hedging errors $C_T - V_T$ for each hedging method. Transaction fees are applied in simulations. Best results are in boldface. MCVaR and MVaR minimize the 95% CVaR and 95% VaR of hedging errors. BSDH and NDH are delta-hedging procedure using the Black-Scholes price and the NIG price with the mean-correcting measure respectively. DMBH rebalances the portfolio when the distance between its delta and the derivative delta is larger than a given constant. GQH and GSQH minimize respectively the RMSE and Semi-RMSE of hedging errors. RFH invests the totality of the hedging capital in the risk-free asset.

Once again, the MCVaR outperforms all other methods in reducing the CVaRs (the $\text{CVaR}_{95\%}$ and $\text{CVaR}_{99\%}$ are respectively reduced of 7.4% and 3.6% with respect to GSQH, the best benchmark), which is expected. It is the second best at minimizing the semi-RMSE, which implies that the CVaR performance is good even away from the error distribution tail. The GSQH also performs well, being the second best at reducing CVaRs and the best at minimizing the Semi-RMSE. For the same reason as explained above, the performance of the MVaR is disastrous and yields extremely large CVaRs (an increase of 46% in the $\text{CVaR}_{99\%}$ with respect to the MCVaR). The RMSE is smaller for the GSQH than for the GQH, which is explained by the fact that the semi-explicit solution used to compute the quadratic hedging does not take transaction costs into account. Once again delta-hedging-based procedures under-perform the global hedging methods. It is worth noting that the difference in the mean error between the no-transaction cost case in Table 3.2 and the transaction cost case in Table 3.4. The negative impact of transaction costs on the hedging performance is apparent : the mean error for the MCVaR ranges from -1.43 to 7.99 when transaction costs are added. The DMBH method, which can be seen as an adaptation of the delta-hedging when transaction costs are present, only slightly under-performs the global GSQH hedging (the semi-RMSE is 16.84 for the former and 16.40 for the latter).

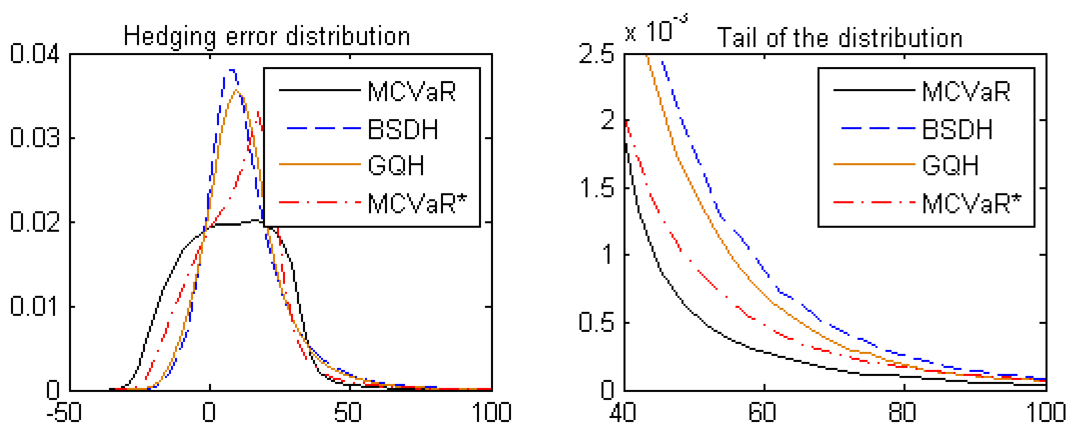
Kernel plots representing the hedging error distribution across the different hedging models are shown in Figures 3.2 and 3.3, which indicate that, of all methods considered, the MCVaR strategy yields the thinnest hedging error distribution tail.

FIGURE 3.2 – Hedging error distribution for global hedging methods



Kernel plots of hedging errors for the MCVaR, MVaR and GSQH global hedging methods.

FIGURE 3.3 – Hedging error distribution for benchmarks



Kernel plots of hedging errors for the MCVaR, MCVaR*, BSDH and GQH methods.

3.5 Conclusion

This paper presents a discrete-time global hedging methodology which takes into account transaction costs. It adapts the work of Hodges & Neuberger (1989) to consider non-Gaussian returns and replace the expected utility metric by the CVaR, a coherent risk measure that allows for the reduction of risk associated with all worst-case outcomes. A rigorous proof of the existence of the solution and a dynamic programming algorithm to compute such a solution are given. The CVaR risk measure is shown to be the first element of a time-consistent sequence of risk measures, implying that the solution to the hedging problem with the CVaR objective function remains optimal through time with respect to that sequence of risk measures. Simulation experiments show the presented hedging methodology compares favorably with other literature benchmarks in terms of risk reduction.

Some aspects of our approach could be improved in further work. Possible alternative numerical schemes based on simulation, spectral interpolation or parametric approximations of the value function could be investigated to ease the numerical burden associated with the computation of the solution to the hedging problem. Such an improvement would accelerate computations and potentially allow for additional dimensions in the hedging problem such as regimes, a stochastic volatility, a stochastic interest rate or additional hedging assets. Moreover, a continuous-time version of our model could be developed, using the Hamilton-Jacobi-Bellman equation to solve the problem.

3.6 Appendix : Proofs for the CVaR minimization problem

Lemma 3.6.1 *Let $f : \mathbb{R}^n \times K \to \mathbb{R}$ be a continuous function where $K \subset \mathbb{R}^d$ is compact. Then $\tilde{f}$ defined by $\tilde{f} := \min_{\theta \in K} f(v, \theta)$ is continuous.*

Proof of Lemma 3.6.1 : Let $v_0 \in \mathbb{R}^n, \theta_0 \in K, \epsilon > 0$. Since f is continuous, it is uniformly continuous on the compact subdomain $B(v_0, 1) \times K$ where $B(x, y)$ is the closed ball of radius y around the point x. Therefore $\exists \delta < 1$ such that $\forall v \in B(v_0, 1), \theta \in K$ satisfying $||(v, \theta) - (v_0, \theta)|| < \delta$ then $|f(v, \theta) - f(v_0, \theta)| < \epsilon$. In other words, if $||v - v_0|| < \delta$, then $\forall \theta, f(v_0, \theta) - \epsilon < f(v, \theta) < f(v_0, \theta) + \epsilon$. This implies by continuity that if $||v - v_0|| < \delta$, $\min_{\theta \in K} f(v_0, \theta) - \epsilon \leq \min_{\theta \in K} f(v, \theta) \leq \min_{\theta \in K} f(v_0, \theta) + \epsilon$, which is $|\tilde{f}(v) - \tilde{f}(v_0)| \leq \epsilon$. ■

Lemma 3.6.2

$$\frac{V_T^{(\theta)}}{B_T} = \frac{V_t^{(\theta)}}{B_t} + \sum_{j=t+1}^{T} \theta_j^{(S)} \left(\frac{S_j}{B_j} - \frac{S_{j-1}}{B_{j-1}} \right) - \frac{K_{j-1}}{B_{j-1}}$$

Proof of Lemma 3.6.2 :

By (3.1),

$$\frac{V_T}{B_T} = \frac{V_{T-1}}{B_{T-1}} + \theta_T^{(S)}\left(\frac{S_T}{B_T} - \frac{S_{T-1}}{B_{T-1}}\right) - \frac{K_{T-1}}{B_{T-1}} \tag{3.43}$$

The proof is completed by repeating the previous iteration $T-t-1$ more times. ■

Lemma 3.6.3 *If assumption (3.13) holds, $\exists M_1, M_2 > 0$ such that $\forall t, \theta$, $|V_T^{(\theta)}| \leq \bar{V}_t^{(\theta_{1:t})}$ a.s., where*

$$\bar{V}_t^{(\theta_{1:t})} := \frac{B_T}{B_t}|V_t^{(\theta_{1:t})}| + M_1 + M_2\sum_{i=0}^{T} S_i. \tag{3.44}$$

Proof of Lemma 3.6.3 : From (3.13), $\exists \tilde{M} \in \mathbb{R}$ such that $\max_t(|\theta_t^{(S)}|) < \tilde{M}$ a.s. From Lemma 3.6.2,

$$\begin{aligned}
\frac{|V_T^{(\theta)}|}{B_T} &\leq \frac{|V_t^{(\theta_{1:t})}|}{B_t} + \sum_{i=t+1}^{T}\tilde{M}(\frac{S_i}{B_0} + \frac{S_{i-1}}{B_0}) + \frac{K_{i-1}}{B_0} \\
&\leq \frac{|V_t^{(\theta_{1:t})}|}{B_t} + \frac{1}{B_0}\sum_{i=0}^{T}\tilde{M}(S_i + S_{i-1}) + k_1 + k_2 2\tilde{M}S_{i-1} \qquad (3.45)\\
&\leq \frac{|V_t^{(\theta_{1:t})}|}{B_t} + \frac{1}{B_0}\sum_{i=0}^{T}\tilde{M}(S_i + S_{i-1}) + (k_1 + k_2 2\tilde{M}S_{i-1}). \blacksquare \qquad (3.46)
\end{aligned}$$

Corollary 3.6.1 *If assumption (3.13)-(3.14) hold, there exists $M_1, M_2 \in \mathbb{R}$ such that*

$$|\Psi_{t,c}^{*,CVaR}(\theta_{1:t})| \leq \frac{|c|(2-\alpha)}{1-\alpha} + \frac{\mathbb{E}[C_T|\mathcal{F}_t] + \frac{B_T}{B_0}V_0 + M_1 + M_2\sum_{j=0}^{T}\mathbb{E}[S_j|\mathcal{F}_t]}{1-\alpha} \; a.s., \tag{3.47}$$

which is integrable.

Proof of Corollary 3.6.1 : By Lemma 3.6.3, $\exists\, M_1, M_2 \in \mathbb{R}$

$$|f_{c,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)})| \leq |c|\left(1 + \frac{1}{1-\alpha}\right) + \frac{|C_T| + \frac{B_T}{B_0}V_0 + M_1 + M_2\sum_{j=0}^{T}\mathbb{E}[S_j|\mathcal{F}_t]}{1-\alpha}. \blacksquare$$

Proof of Lemmas 3.2.1 and 3.2.2 :

The first objective is to show that (3.15)-(3.17) hold for all $t \leq T$. The proof proceeds by induction. When, $t = T$, (3.15)-(3.17) are trivially satisfied because of the continuity of $f_{c,\alpha}^{(CVaR)}$. Assume (3.15)-(3.17) hold for $t = n+1$. We show that they also hold for $t = n$.

$\forall\theta_{1:n}$, $\forall\theta_{n+1} \in SF_n$, $\Psi_{n+1,c}^{CVaR}(S_{n+1}, \theta_{n+1}, v_{n+1})$ is a non-decreasing function of v_{n+1}. Thus,

$$\begin{aligned}
\tilde{\Psi}_{n,c,\theta_{1:n}}^{*,CVaR} &:= \inf_{\theta_{n+1}\in SF_{n+1}} \mathbb{E}\left[\Psi_{n+1,c}^{CVaR}(S_{n+1}, \theta_{n+1}, V_{n+1}(\theta_{1:n+1}))\middle|\mathcal{F}_n\right] \qquad (3.48)\\
&= \min\left\{\inf_{\theta_{t+1}\in SF_{t+1}}\tilde{\rho}_n^{(1)},\ \tilde{\rho}_n^{(2)}\right\} \qquad (3.49)
\end{aligned}$$

where

$$
\begin{aligned}
&\tilde{\rho}_n^{(1)} := \mathbb{E}\left[\Psi_{n+1,c}^{CVaR}(S_{n+1},\theta_{n+1},\mathcal{V}_{n+1}^{(1)}(V_n(\theta_{1:n}),\theta_n,\theta_{n+1},S_n,S_{n+1}))\big|\mathcal{F}_n\right] \\
&\tilde{\rho}_n^{(2)} := \mathbb{E}\left[\Psi_{n+1,c}^{CVaR}(S_{n+1},\theta_{n+1},\mathcal{V}_{n+1}^{(2)}(V_n(\theta_{1:n}),\theta_n,S_n,S_{n+1}))\big|\mathcal{F}_n\right] \\
&\mathcal{V}_{n+1}^{(1)}(V_n(\theta_{1:n}),\theta_n,\theta_{n+1},S_n,S_{n+1}) \\
&\qquad := \frac{B_{n+1}}{B_n}(V_n(\theta_{1:n})-k_1-k_2|\theta_n^{(S)}-\theta_{n-1}^{(S)}|S_{n-1}) + \theta_{n+1}^{(S)}\left(S_{n+1}-S_n\frac{B_{n+1}}{B_n}\right), \\
&\mathcal{V}_{n+1}^{(2)}(V_n(\theta_{1:n}),\theta_n,S_n,S_{n+1}) := \frac{B_{n+1}}{B_n}V_n(\theta_{1:n}) + \theta_n^{(S)}\left(S_{n+1}-S_n\frac{B_{n+1}}{B_n}\right).
\end{aligned}
$$

$\mathcal{V}_{n+1}^{(1)}$ represents the next-period value of the portfolio if the fixed transaction fees k_1 are applied, even if no rebalancing occurs. $\mathcal{V}_{n+1}^{(2)}$ represents the next-period value of the portfolio if no rebalancing occurs.

From Assumption 3.2.1, $\tilde{\rho}_n^{(1)} = \rho_n^{(1)}(c,V_n(\theta_{1:n}),\theta_n,\theta_{n+1},S_n)$ and $\tilde{\rho}_n^{(2)} = \rho_n^{(2)}(c,V_n(\theta_{1:n}),\theta_n,S_n)$. Because of the continuity of $\mathcal{V}_{n+1}^{(1)}(v_n,\vartheta_n,\vartheta_{n+1},S_n,S_{n+1})$ with respect to $(v_n,\vartheta_n,\vartheta_{n+1})$ and because $\Psi_{n+1,c}^{CVaR}(S_{n+1},\vartheta_{n+1},v_{n+1})$ is continuous with respect to $(c,\vartheta_{n+1},v_{n+1})$ by the induction hypothesis,

$$
\begin{aligned}
& \lim_{(\bar{v}_n,\bar{c},\bar{\theta}_n,\bar{\theta}_{n+1})\to(v_n,c,\theta_n,\theta_{n+1})} \mathbb{E}\left[\Psi_{n+1,\bar{c}}^{CVaR}\left(S_{n+1},\theta_{n+1},\mathcal{V}_{n+1}^{(1)}(\bar{v}_n,\bar{\theta}_n,\bar{\theta}_{n+1},S_n,S_{n+1})\right)\big|\mathcal{F}_n\right] \\
= \; & \mathbb{E}\left[\lim_{(\bar{v}_n,\bar{c},\bar{\theta}_n,\bar{\theta}_{n+1})\to(v_n,c,\theta_n,\theta_{n+1})} \Psi_{n+1,\bar{c}}^{CVaR}\left(S_{n+1},\theta_{n+1},\mathcal{V}_{n+1}^{(1)}(\bar{v}_n,\bar{\theta}_n,\bar{\theta}_{n+1},S_n,S_{n+1})\right)\big|\mathcal{F}_n\right] \\
& \qquad\qquad \text{By dominated convergence with bound (3.47)} \\
= \; & \mathbb{E}\left[\Psi_{n+1,c}^{CVaR}\left(S_{n+1},\theta_{n+1},\mathcal{V}_{n+1}^{(1)}(v_n,\theta_n,\theta_{n+1},S_n,S_{n+1})\right)\big|\mathcal{F}_n\right]. \quad \text{by continuity}
\end{aligned}
$$

This implies $\rho_n^{(1)}(c,v_n,\theta_n,\theta_{n+1},S_n)$ is continuous with respect to $(c,v_n,\theta_n,\theta_{n+1})$. Similarly, $\rho_n^{(2)}(c,v_n,\theta_n,S_n)$ is continuous with respect to (c,v_n,θ_n). The compacity of $\mathcal{B}_{t+1}$ implies that the minimum is attained in

$$
\underline{\tilde{\rho}}_n^{(1)} := \inf_{\theta_{n+1}\in SF_{n+1}} \tilde{\rho}_n^{(1)} = \min_{\theta_{n+1}\in SF_{n+1}} \rho_n^{(1)}(c,V_n(\theta_{1:n}),\theta_n,\theta_{n+1},S_n)
$$

From Lemma 3.6.1, $\underline{\tilde{\rho}}_n^{(1)} = \underline{\rho}_n^{(1)}(c,V_n(\theta_{1:n}),\theta_n,S_n)$ is continuous with respect to (c,v_n,θ_n). This implies from (3.49) that the minimum is attained in (3.48)

$$
\tilde{\Psi}_{n,c,\theta_{1:n}}^{*,CVaR} = \min_{\theta_{n+1}\in SF_{n+1}} \mathbb{E}\left[\Psi_{n+1,c}^{CVaR}(S_{n+1},\theta_{n+1},V_{n+1}(\theta_{1:n+1}))\big|\mathcal{F}_n\right], \tag{3.50}
$$

This and (3.49) prove $\tilde{\Psi}_{n,c,\theta_{1:n}}^{*,CVaR}$ is $\sigma(S_n,V_n^{(\theta_{1:n})},\theta_n)$-measurable. This, combined with the fact that $\theta_{1:t-1}$ are not used in the computation of $\rho_n^{(1)}$ and $\rho_n^{(2)}$, implies that

$$
\tilde{\Psi}_{n,c,\theta_{1:n}}^{*,CVaR} = \tilde{\Psi}_{n,c}^{CVaR}(S_n,V_n^{(\theta_{1:n})},\theta_n) \tag{3.51}
$$

for some functions $\tilde{\Psi}^{CVaR}_{n,c}$.

From the well-known principle of optimality of dynamic programming (Bellman & Dreyfus, 1962),

$$\Psi^{*,CVaR}_{n,c}(\theta_{1:n}) = \tilde{\Psi}^{*,CVaR}_{n,c,\theta_{1:n}}. \tag{3.52}$$

Furthermore, denote

$$\begin{aligned}
\theta^*_{n+2:T}(\theta_{1:n+1}) &:= \underset{\theta_{n+2:T}}{\arg\min} \left\{ \mathbb{E}\left[f^{(CVaR)}_{c,\alpha}(C_T - V_T^{(\theta_{1:n+1},\theta_{n+2:T})}) \middle| \mathcal{F}_{n+1} \right] : (\theta_{1:n+1}, \theta_{n+2:T}) \in \Theta \right\}, \\
\theta^*_{n+1}(\theta_{1:n}) &:= \underset{\theta_{n+1}\in SF_{n+1}}{\arg\min}\ \mathbb{E}\left[\Psi^{CVaR}_{n+1,c}(S_{n+1}, \theta_{n+1}, V_{n+1}(\theta_{1:n+1})) \middle| \mathcal{F}_n \right]
\end{aligned}$$

Then, from the principle of optimality,

$$\begin{aligned}
&\left(\theta^*_{n+1}(\theta_{1:n}), \theta^*_{n+2:T}(\theta_{1:n}, \theta^*_{n+1}(\theta_{1:n}))\right) \\
&\quad = \underset{\theta_{n+1:T}}{\arg\min} \left\{ \mathbb{E}\left[f^{(CVaR)}_{c,\alpha}(C_T - V_T^{(\theta_{1:n+1},\theta_{n+2:T})}) \middle| \mathcal{F}_n \right] : (\theta_{1:n}, \theta_{n+1:T}) \in \Theta \right\}
\end{aligned} \tag{3.53}$$

We refer to Theorem 3.2 of François et al. (2014) for a proof of statements (3.52) and (3.53) in the hedging context. This proves Lemma 3.2.2. Moreover, combining (3.50), (3.51) and (3.52) proves that (3.15)-(3.17) hold for $t = n$. ■

Lemma 3.6.4 $\Psi^{CVaR}_{0,c}$*, which is defined in Lemma 3.2.1, satisfies* $\lim_{c\to-\infty} \Psi^{CVaR}_{0,c}(S_0, \theta_0, V_0) = \infty$.

Proof of Lemma 3.6.4 :

Let $\mathcal{J} > 0$. For all θ,

$$\begin{aligned}
f^{(CVaR)}_{c,\alpha}(C_T - V_T^{(\theta)}) &= c + \frac{1}{1-\alpha}(C_T - V_T^{(\theta)} - c)\mathbb{I}_{\{C_T - V_T^{(\theta)} > c\}} \\
&\geq c - \frac{c}{1-\alpha}\mathbb{I}_{\{C_T - V_T^{(\theta)} > c\}} - \frac{(C_T + |V_T^{(\theta)}|)}{1-\alpha} \\
&\geq c - \frac{c}{1-\alpha}\mathbb{I}_{\{C_T - V_T^{(\theta)} > c\}} - \frac{\left(C_T + \bar{V}_0\right)}{1-\alpha}
\end{aligned} \tag{3.54}$$

where $\bar{V}_0$ is defined in Lemma 3.6.3.

Define $\mathcal{E} := \frac{1}{1-\alpha}\mathbb{E}\left[C_T + \bar{V}_0\right] < \infty$.

Let $\tilde{c} < -2(\mathcal{J} + \mathcal{E}) < 0$ be small enough such that

$$\mathbb{P}[C_T - \bar{V}_0 > \tilde{c}] > (1-\alpha)/2. \tag{3.55}$$

This gives $\forall \theta \in \Theta$, $\forall c \leq \tilde{c}$,

$$\mathbb{P}[C_T - V_T^{(\theta)} > c] \geq \mathbb{P}[C_T - V_T^{(\theta)} > \tilde{c}] \geq \mathbb{P}[C_T - \bar{V}_0 > \tilde{c}]. \tag{3.56}$$

Thus for all θ, for all $c \leq \tilde{c}$,

$$\begin{aligned}
\mathbb{E}\left[f_{c,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)})\right] &\geq c\left(1 - \frac{\mathbb{P}[C_T - V_T^{(\theta)} > c]}{1-\alpha}\right) - \mathcal{E} \quad \text{from (3.54)} \\
&\geq \tilde{c}\left(1 - \frac{\mathbb{P}[C_T - \bar{V}_0 > \tilde{c}]}{1-\alpha}\right) - \mathcal{E} \quad \text{from (3.55) and (3.56)} \\
&\geq \tilde{c}\left(1 - \frac{(1-\alpha)/2}{1-\alpha}\right) - \mathcal{E} \text{ from (3.55)} \\
&= -\frac{\tilde{c}}{2} - \mathcal{E} \geq \mathcal{J}.
\end{aligned}$$

We have thus shown that $\forall \mathcal{J} \in \mathbb{R}$, $\exists \tilde{c} < 0$ such that $\forall c < \tilde{c}$, $\Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0) \geq \mathcal{J}$. ■

Proof of Theorem 3.2.2 :

From Lemma 3.2.1,

$$\min_{\theta \in \Theta}\left\{\mathbb{E}\left[f_{c,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)})\right]\right\} = \Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0)$$

is continuous with respect to c. Since $f_{c,\alpha}^{(CVaR)}(\bullet) \geq c$, $\Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0) \geq c$. This implies $\lim_{c \to \infty} \Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0) = \infty$. Furthermore, Lemma 3.6.4 shows that $\lim_{c \to -\infty} \Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0) = \infty$. Since $\Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0)$ is continuous with respect to c and converges to infinity when $|c| \to \infty$, its minimum is attained at some point c^*. Thus,

$$\begin{aligned}
\inf_{\theta \in \Theta} \mathrm{CVaR}_\alpha(C_T - V_T^{(\theta)}) &= \inf_{c \in \mathbb{R}} \inf_{\theta \in \Theta} \mathbb{E}\left[f_{c,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)})\right]. \\
&= \inf_{c \in \mathbb{R}} \Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0) \\
&= \Psi_{0,c^*}^{CVaR}(S_0, \theta_0, V_0).
\end{aligned}$$

Furthermore

$$\arg\min_{\theta \in \Theta} \mathrm{CVaR}_\alpha(C_T - V_T^{(\theta)}) = \arg\min_{\theta \in \Theta} \mathbb{E}\left[f_{c^*,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)})\right] \tag{3.57}$$

which exists due to Lemma 3.2.1. ■

Proof of Theorem 3.2.3 :

For all θ, define

$$\tilde{V}_0 := V_0 + (x - c)\frac{B_0}{B_T}, \tag{3.58}$$

$$\tilde{V}_T(\theta) := B_T\left[\frac{\tilde{V}_0}{B_0} + \sum_{j=1}^{T} \theta_j^{(S)}\left(\frac{S_j}{B_j} - \frac{S_{j-1}}{B_{j-1}}\right) - \frac{K_{j-1}}{B_{j-1}}\right]. \tag{3.59}$$

As can be seen from Lemma 3.6.2, $\tilde{V}_T(\theta)$ is the terminal value for the self-financing portfolio invested through the hedging strategy θ and starting with the initial value $\tilde{V}_0$. Using Lemma 3.6.2,

$$
\begin{aligned}
& f_{x,\alpha}^{(CVaR)}(C_T - V_T(\theta)) \\
&= x + \frac{1}{1-\alpha}(C_T - V_T(\theta) - x)\mathbb{I}_{\{C_T - V_T(\theta) > x\}} \\
&= x + \frac{1}{1-\alpha}\left(C_T - \left[\frac{B_T V_0}{B_0} + \sum_{j=1}^{T} B_T \theta_j^{(S)}\left(\frac{S_j}{B_j} + \frac{S_{j-1}}{B_{j-1}}\right) - \frac{B_T K_{j-1}}{B_{j-1}}\right] - x\right) \\
&\qquad \times \mathbb{I}_{\left\{C_T - \left[\frac{B_T V_0}{B_0} + \sum_{j=1}^{T} B_T \theta_j^{(S)}\left(\frac{S_j}{B_j} + \frac{S_{j-1}}{B_{j-1}}\right) - \frac{B_T K_{j-1}}{B_{j-1}}\right] > x\right\}} \qquad \text{by Lemma 3.6.2} \\
&= (x-c) + c + \frac{1}{1-\alpha}(C_T - \tilde{V}_T(\theta) - c)\mathbb{I}_{\{C_T - \tilde{V}_T(\theta) > c\}} \\
&= (x-c) + f_{c,\alpha}^{(CVaR)}(C_T - \tilde{V}_T(\theta)). \quad \blacksquare
\end{aligned}
$$

3.7 Appendix : Proofs for the VaR minimization problem

Consider the problem of minimizing the VaR_α of the hedging error :

$$
\begin{aligned}
\inf_{\theta \in \Theta} \mathrm{VaR}_\alpha(C_T - V_T^{(\theta)}) &= \inf_{\theta \in \Theta} \min_c \{c \,|\, \mathbb{E}[\mathbb{I}_{\{C_T - V_T^{(\theta)} \leq c\}}] \geq \alpha\} \qquad (3.60) \\
&= \inf_{\theta \in \Theta} \min_c \{c \,|\, \mathbb{E}[\mathbb{I}_{\{C_T - V_T^{(\theta)} > c\}}] \leq 1 - \alpha\} \\
&= \inf\{c \,\big|\, \exists \theta \in \Theta \text{ such that } \mathbb{E}[f_c^{(VaR)}(C_T - V_T^{(\theta)})] \leq 1 - \alpha\} \qquad (3.61)
\end{aligned}
$$

where $f_c^{(VaR)}$ is defined by (3.38). In order to solve (3.60), one must therefore solve Problem (3.37) for different values of c. The latter problem can be solved through dynamic programming.

Definition 3.7.1

$$
\begin{aligned}
& \Psi_{T,c}^{*,VaR}(\theta) := f_{c,\alpha}^{(VaR)}(C_T - V_T^{(\theta)}) \qquad (3.62) \\
\forall t < T, \quad & \Psi_{t,c}^{*,VaR}(\theta_{1:t}) := \inf_{\theta_{t+1:T}} \left\{ \mathbb{E}\left[f_{c,\alpha}^{(VaR)}(C_T - V_T^{(\theta_{1:t}, \theta_{t+1:T})}) \big| \mathcal{F}_t \right] : (\theta_{1:t}, \theta_{t+1:T}) \in \Theta \right\} \qquad (3.63)
\end{aligned}
$$

where $f_{c,\alpha}^{(VaR)}$ is defined by (3.38).

Results analogous to Lemmas 3.2.1 and 3.2.2 can be derived.

Lemma 3.7.1 (The Bellman Equation) *If (3.39)-(3.41) hold, then for all $t < T$, $\theta_{1:t}$,*

There exists functions $\Psi^{VaR}_{t,c}$ such that $\Psi^{*,VaR}_{t,c}(\theta_{1:t}) = \Psi^{VaR}_{t,c}(S_t, \theta_t, V_t^{(\theta_{1:t})})$ a.s.

$$\Psi^{VaR}_{t,c}(S_t, \theta_t, V_t^{(\theta_{1:t})}) = \min_{\theta_{t+1} \in SF_{t+1}} \mathbb{E}\left[\Psi^{VaR}_{t+1,c}(S_{t+1}, \theta_{t+1}, V_{t+1}^{(\theta_{1:t},\theta_{t+1})}) \middle| S_t, \theta_t, V_t^{(\theta_{1:t})}\right].$$

The proof is similar to the proof of one of the latter theorems and is therefore not repeated. The following result gives the asymptotic behavior of $\Psi^{VaR}_{0,c}$ when $|c|$ goes to infinity, which is required in order to ensure the existence of a solution to problem (3.36).

Lemma 3.7.2 *If $\mathcal{B}_t$ are compact for all t, $\Psi^{VaR}_{0,c}$ which is defined in Lemma 3.7.1, satisfies $\lim_{c \to -\infty} \Psi_{0,c}(S_0, \theta_0, V_0) = 1$ and $\lim_{c \to \infty} \Psi_{0,c}(S_0, \theta_0, V_0) = 0$.*

Proof of Lemma 3.7.2 :

Let $\epsilon > 0$. Select any $\theta \in \Theta$, and choose $\bar{c}_1$ such that $\mathbb{P}(C_T - V_T^{(\theta)} > \bar{c}_1) < \epsilon$. This implies $\forall c > \bar{c}_1$, $\Psi_{0,c}(S_0, \theta_0, V_0) < \Psi_{0,\bar{c}_1}(S_0, \theta_0, V_0) < \epsilon$.

Furthermore, define $\bar{V}_0$ as in Lemma 3.6.3. Choose $\bar{c}_2$ such that $\mathbb{P}\left[C_T - \bar{V}_0 > \bar{c}_2\right] \geq 1 - \epsilon$. By Lemma 3.6.3, $\forall \theta \in \Theta$, $\forall c < \bar{c}_2$,

$$\Psi_{0,c}(S_0, \theta_0, V_0) = \inf_{\theta \in \Theta} \mathbb{P}\left[C_T - V_T^{(\theta)} > c\right] \geq \mathbb{P}\left[C_T - \bar{V}_0 > \bar{c}_2\right] \geq 1 - \epsilon. \quad \blacksquare$$

Lemmas 3.7.1 and 3.7.2 lead to the following characterization of the solution to problem (3.60).

Theorem 3.7.1 *If (3.39)-(3.41) hold, then*

$$\min_{\theta \in \Theta} VaR_\alpha \left(C_T - V_T^{(\theta)}\right) = c^* \tag{3.64}$$

where $c^* = \min\{c | \Psi^{VaR}_{0,c}(S_0, V_0) = 1 - \alpha\}$. Furthermore, all solutions $\theta^{*,VaR}$ to

$$\theta^{*,VaR} \in \arg\min_{\theta \in \Theta} \mathbb{E}\left[f^{(VaR)}_{c^*,\alpha}(C_T - V_T^{(\theta)})\right], \tag{3.65}$$

which exist, also solve (3.64).

Proof of Theorem 3.7.1 :

From Lemma 3.7.1 and assumption (3.40),

$$\min_{\theta \in \Theta} \mathbb{E}\left[f^{(VaR)}_{c}(C_T - V_T^{(\theta)})\right] = \Psi^{VaR}_{0,c}(S_0, V_0)$$

is continuous with respect to c. Define $\mathbb{C} := \{c|\Psi^{VaR}_{0,c}(S_0,\theta_0,V_0) = 1-\alpha\}$. From Lemma 3.7.2, $\lim_{c\to\infty}\Psi_{0,c}(S_0,\theta_0,V_0) = 0$ and $\lim_{c\to-\infty}\Psi_{0,c}(S_0,\theta_0,V_0) = 1$. This implies that $\mathbb{C}$ is not empty. Since $\Psi^{VaR}_{0,c}(S_0,\theta_0,V_0)$ is continuous with respect to c by (3.41), $\mathbb{C}$ is closed. Define $\tilde{c} := \min\mathbb{C}$.

Define $\theta^*(c) := \arg\min_{\theta\in\Theta}\mathbb{E}\left[f^{(VaR)}_c(C_T - V^{(\theta)}_T)\right]$ which exists from Lemma 3.7.1. First, since $\mathrm{VaR}_\alpha(C_T - V^{(\theta^*(\tilde{c}))}_T) = \tilde{c}$, $\inf_{\theta\in\Theta}\mathrm{VaR}_\alpha(C_T - V^{(\theta)}_T) \leq \tilde{c}$. Second, by the definition of $\tilde{c}$, $\forall c < \tilde{c}$, $\Psi^{VaR}_{0,c}(S_0,\theta_0,V_0) > 1-\alpha$. This implies that $\forall c < \tilde{c}$, $\inf_{\theta\in\Theta}\mathrm{VaR}_\alpha(C_T - V^{(\theta)}_T) \geq c$.

These two inequalities imply that $\min_{\theta\in\Theta}\mathrm{VaR}_\alpha(C_T - V^{(\theta)}_T) = \tilde{c}$ and that the minimum strategy is attained by the hedging strategy $\theta^*(\tilde{c})$. ■

As in the CVaR case, a single run of the dynamic programming leads to the calculation of $\Psi^{VaR}_{0,c}$ all c simulaneously :

Theorem 3.7.2 $\forall x, c \in \mathbb{R}$, $\Psi^{VaR}_{0,x}(S_0,\theta_0,V_0) = \Psi^{VaR}_{0,c}(S_0,\theta_0,V_0 + (x-c)\frac{B_0}{B_T})$.

Proof of Theorem 3.7.2 :

For all θ, define $\tilde{V}_0$ and $\tilde{V}_T(\theta)$ by (3.58)-(3.59).

$$\begin{aligned}
f^{(VaR)}_x(C_T - V_T(\theta)) &= \mathbb{I}_{\{C_T - V_T(\theta) > x\}} \\
&= \mathbb{I}_{\left\{C_T - \left[\frac{B_T V_0}{B_0} + \sum_{j=1}^T B_T\theta^{(S)}_j\left(\frac{S_j}{B_j} - \frac{S_{j-1}}{B_{j-1}}\right) - \frac{K_{j-1}}{B_{j-1}}\right] > x\right\}} \quad \text{by Lemma 3.6.2} \\
&= \mathbb{I}_{\left\{C_T - \left[\frac{B_T \tilde{V}_0}{B_0} + \sum_{j=1}^T B_T\theta^{(S)}_j\left(\frac{S_j}{B_j} - \frac{S_{j-1}}{B_{j-1}}\right) - \frac{K_{j-1}}{B_{j-1}}\right] > c\right\}} \\
&= \mathbb{I}_{\left\{C_T - \tilde{V}_T(\theta) > c\right\}} \\
&= f^{(VaR)}_c(C_T - \tilde{V}_T(\theta)). \quad ■
\end{aligned}$$

3.8 Appendix : Non-time-consistency conditional VaRs and CVaRs

This section shows by means of a simple counter-example that the set of conditional VaR and CVaR risk measures are not necessarily time-consistent.

Consider the case of a two period market : $t = 0, 1, 2$. For simplicity, the risk-free rate is null and the risk-free asset value is therefore constant. The dynamics of the risky asset is represented by an arbitrage-free tree where its log-return between times t and $t+1$ is denoted by $\epsilon_{t+1} = \log \frac{S_{t+1}}{S_t}$. The following values for log-return are possible :

$$\epsilon_1 = -0.1 \text{ or } 0.1, \text{ each with probability } p_1 = 0.5, \tag{3.66}$$

$$\epsilon_2 = -0.1, -0.05, 0, 0.05 \text{ or } 0.1, \text{ each with probability } p_2 = 0.2. \tag{3.67}$$

Log-returns ϵ_1 and ϵ_2 are independent. The initial value of the stock is $S_0 = 100$. Assume the hedged derivative is a European call option with strike $E = 105$: $C_2 = \max\{0, S_2 - E\}$. The admissible strategies are $\Theta := \{(\theta_1, \theta_2) | \forall i,\ \theta_i \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}\}$. Discrete admissible strategies make the illustration of the case easier. Denote by $\theta_2(u)$ the portfolio position in the upper node, i.e. when $\epsilon_1 = -0.1$, and by $\theta_2(d)$ the portfolio position in the lower node ($\epsilon_1 = 0.1$). For this simple market, consider the following sets of risk measures $\forall t,\ \rho_t = \text{VaR}_{t,\alpha}$ (the VaR case) and $\forall t,\ \rho_t = \text{CVaR}_{t,\alpha}$ (the CVaR case). Table 3.5 gives the optimal solutions for problem (3.24). The risk levels considered are $\alpha = 0.7$ for the VaR case and $\alpha = 0.6$ for the CVaR case.

TABLE 3.5 – Optimal trading strategies in the simple tree market

Risk measure	Unique solution at $t = 0$	Solution at $t = 1$
$\text{VaR}_{t,0.7}$	$(\theta_1, \theta_2(u), \theta_2(d)) = (0, 0, 0)$	$\forall \theta_1, (\theta_2(u), \theta_2(d)) = (1, 0)$
$\text{CVaR}_{t,0.6}$	$(\theta_1, \theta_2(u), \theta_2(d)) = (.4, .8, 0)$	$\forall \theta_1, (\theta_2(u), \theta_2(d)) = (1, 0)$

Optimal trading strategies in the simple tree market when the two following risk measures are considered : conditional $\text{VaR}_{t,0.7}$ and $\text{CVaR}_{t,0.6}$.

As seen in Table 3.5, in both the VaR and CVaR cases, $\bigcap_{t=0}^{1} \Theta_t^* = \varnothing$. Therefore, the set of conditional VaRs and CVaRs are time-inconsistent in this example.

3.9 Technical report (not part of the paper)

3.9.1 Additional result on minimizing the CVaR

The following theorem guarantees that the c that minimizes $\Psi_{0,c}^{CVaR}$ is unique.

Theorem 3.9.1 *Define*

$$\mathbb{C} := \left\{ c^* \middle| \Psi_{0,c^*}^{CVaR}(S_0, \theta_0, V_0) = \min_{c \in \mathbb{R}} \Psi_{0,c}^{CVaR}(S_0, \theta_0, V_0) \right\}.$$

If $\forall c^* \in \mathbb{C}, \exists \theta^{c^*}$ *such that* $\theta^{c^*} \in \arg\min_{\theta \in \Theta} \mathbb{E}\left[f_{c^*,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)}) \right]$ *and* $C_T - V_T^{(\theta^{c^*})}$ *has a strictly increasing cdf, then* $\mathbb{C}$ *contains a unique point.*

Proof of Theorem 3.9.1 :

Since $\Psi_{0,c}^{CVaR}(S_0,\theta_0,V_0)$ is continuous with respect to c, $\mathbb{C}$ is closed. Define $c_1 := \min\mathbb{C}$. Suppose there exists $c_2 \in \mathbb{C}$ such that $c_2 > c_1$. Let θ^{c_2} be the hedging strategy that minimizes $\mathbb{E}\left[f_{c_2,\alpha}^{(CVaR)}(C_T - V_T^{(\theta)})\right]$ which is such that $C_T - V_T^{(\theta^{c_2})}$ has a strictly increasing cdf. Then,

$$\begin{aligned}
& (C_T - V_T^{(\theta^{c_2})} - c_1)\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} > c_1\}} - (C_T - V_T^{(\theta^{c_2})} - c_2)\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} > c_2\}} \\
\leq\ & (c_2 - c_1)\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} > c_1\}}\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} > c_2\}} + (C_T - V_T^{(\theta^{c_2})} - c_1)\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} > c_1\}}\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} \leq c_2\}} \\
\leq\ & (c_2 - c_1)\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} > c_1\}}\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} > c_2\}} + (c_2 - c_1)\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} > c_1\}}\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} \leq c_2\}} \\
=\ & (c_2 - c_1)\mathbb{I}_{\{C_T - V_T^{(\theta^{c_2})} > c_1\}}
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathbb{E}\left[f_{c_1,\alpha}^{(CVaR)}(C_T - V_T^{(\theta^{c_2})})\right] - \min_{c\in\mathbb{R}}\Psi_{0,c}^{CVaR}(S_0,\theta_0,V_0) \\
=\ & \mathbb{E}\left[f_{c_1,\alpha}^{(CVaR)}(C_T - V_T^{(\theta^{c_2})})\right] - \mathbb{E}\left[f_{c_2,\alpha}^{(CVaR)}(C_T - V_T^{(\theta^{c_2})})\right] \\
\leq\ & (c_1 - c_2) + \frac{1}{1-\alpha}(c_2 - c_1)\mathbb{P}[C_T - V_T^{(\theta^{c_2})} > c_1] \\
=\ & (c_2 - c_1)\left(\frac{\mathbb{P}[C_T - V_T^{(\theta^{c_2})} > c_1]}{1-\alpha} - 1\right).
\end{aligned} \tag{3.68}$$

Since $c_2 = \mathrm{VaR}_\alpha(C_T - V_T^{(\theta^{c_2})})$ from (Theorem 10 of Rockafellar & Uryasev, 2002), $\mathbb{P}[C_T - V_T^{(\theta^{c_2})} > c_2] = 1-\alpha$. Furthermore $\mathbb{P}[C_T - V_T^{(\theta^{c_2})} \in (c_1, c_2]\,] > 0$, otherwise $\mathrm{VaR}_\alpha(C_T - V_T^{(\theta^{c_2})}) \leq c_1$. This implies

$$\mathbb{P}[C_T - V_T^{(\theta^{c_2})} > c_1] > 1-\alpha. \tag{3.69}$$

Plugging (3.69) in (3.68) implies that

$$\mathbb{E}\left[f_{c_1,\alpha}^{(CVaR)}(C_T - V_T^{(\theta^{c_2})})\right] < \min_{c\in\mathbb{R}}\Psi_{0,c}^{CVaR}(S_0,\theta_0,V_0),$$

which contradicts that $c_1 \in \mathbb{C}$. ■

3.9.2 Additional result on minimizing the VaR

Theorem 3.9.2 *Assume (3.39)-(3.41) hold. Let* $\theta^*(c) := \arg\min_{\theta\in\Theta}\mathbb{E}\left[f_c^{(VaR)}(C_T - V_T^{(\theta)})\right]$ *which exists from Lemma 3.7.1. Define* $\mathbb{C} := \{c|\Psi_{0,c}^{VaR}(S_0,\theta_0,V_0) = 1-\alpha\}$ *and* $\tilde{c} := \min\mathbb{C}$*. If, the distribution* $C_T - V_T^{(\theta^*(\tilde{c}))}$ *has no atom, then* $\mathbb{C}$ *contains a unique point.*

Proof of Theorem 3.9.2 :

The non-emptiness of $\mathbb{C}$ is proved by Theorem 3.7.1. Define $\tilde{c} := \min \mathbb{C}$. Let $c_2 \in \mathbb{C} > \tilde{c}$ such that $c_2 > \tilde{c}$. From the definition of $\theta^*(\tilde{c})$ and since $\tilde{c} \in \mathbb{C}$, $\mathbb{P}\left[C_T - V_T^{(\theta^*(\tilde{c}))} > \tilde{c}\right] = 1 - \alpha$. Since the distribution of $C_T - V_T^{(\theta^*(\tilde{c}))}$ has no atoms, $\mathbb{P}\left[C_T - V_T^{(\theta^*(\tilde{c}))} > c_2\right] < 1 - \alpha$, which contradicts that $c_2 \in \mathbb{C}$.

3.9.3 More on defining time-consistency of risk measures

It could be argued that definition 3.3.4 can be relaxed, and that the existence of one solution which remains optimal through time (instead of the necessity of all solutions to remain so) is sufficient for consistence. This inspires the following more general version of consistency.

Definition 3.9.1 *The set of risk measures $\{\rho_t\}_{t=0}^{T-1}$ are said to be* weakly time-consistent *if $\bigcap_{t=0}^{T-1} \Theta_t^* \neq \varnothing$.*

In other words, risk measures are time-consistent if there exists a common solution θ which solves problems (3.24) for all $t = 0, \ldots, T-1$. Trivially, time-consistency implies weak time-consistency. Furthermore, weak time-consistency and the existence and uniqueness of solutions to all problems (3.24) imply time-consistency.

Remark 3.9.1 *Even in simple cases, the existence of a solution might not exist. Such a situation is illustrated for the expected squared hedging error measure in Example* 4 *of Schweizer (1995). Section 3.9.4 give a simple example where the* VaR *cannot be minimized.*

3.9.4 Example of non-existence of an optimal strategy with VaR

The following example illustrates a situation where the existence of a solution minimizing the hedging VaR does not exist.

Consider the case of a European call option with strike $K = 2$ with maturity $T = 2$ on a stock following a binomial tree. The initial value of the stock is $S_0 = 3$ and its dynamics are characterized by $\mathbb{P}[S_{t+1} = S_t + 1 | S_t] = \mathbb{P}[S_{t+1} = S_t - 1 | S_t] = 1/2$. The risk-free asset has a constant value of $B_t = 1$ for all t. The market is therefore complete and arbitrage-free in this case. The initial capital is the arbitrage-free price of the option ($V_0 = 1.25$) and perfect replication can therefore be achieved. The distribution of the hedging error $C_T - V_T$ is

$$\begin{cases} 3 - \theta_1^{(S)} - \theta_2^{(S,u)} \text{ with probability } 1/4, \\ 1 - \theta_1^{(S)} + \theta_2^{(S,u)} \text{ with probability } 1/4, \\ 1 + \theta_1^{(S)} - \theta_2^{(S,d)} \text{ with probability } 1/4, \\ 0 + \theta_1^{(S)} + \theta_2^{(S,d)} \text{ with probability } 1/4, \end{cases} \tag{3.70}$$

since $V_T = V_0 + \theta_1(S_1 - S_0) + \theta_2(S_2 - S_1)$. $\theta_2^{(S,u)}$ and $\theta_2^{(S,u)}$ are the number of stocks in the portfolio at the respective nodes $S_1 = 5$ and $S_1 = 3$. The set of admissible strategies is set to $\Theta = \{(\theta_1^{(S)}, \theta_2^{(S,u)}, \theta_2^{(S,d)}) \in \mathbb{R}^3\}$. Perfect replication can be achieved by choosing $(\theta_1^{(S)}, \theta_2^{(S,u)}, \theta_2^{(S,d)}) = (0.75, 1, 0.5)$.

Consider the risk measure $\rho = VaR_{0.75}$ to solve problem (3.3). By choosing $\theta_2^{(S,u)} = 1$ and $\theta_2^{(S,d)} = 2\theta_1^{(S)} - 1$, (3.70) becomes

$$\begin{cases} 2 - \theta_1^{(S)} \text{ with probability } 3/4, \\ 3\theta_1^{(S)} - 1 \text{ with probability } 1/4, \end{cases} \tag{3.71}$$

and therefore $VaR_{0.75}(C_T - V_T) = 2 - \theta_1^{(S)}$. Therefore, $VaR_{0.75}(C_T - V_T) \to -\infty$ when $\theta_1^{(S)} \to \infty$ and there exists no solution θ that minimizes $\rho(C_T - V_T)$. Furthermore, the hedging strategy allowing perfect replication is not optimal according to the $VaR_{0.75}$ risk measure.

Remark 3.9.2 *In the previous example, the exact replication of the derivative is sub-optimal. To ensure that exact replication is always optimal if it is achievable in a global hedging problem, the following conditions are sufficient : the risk measure to be minimized is* ρ *where* $\rho(0) = 0$ *and* $\min \rho(C_T - V_T) \geq 0$. *Using a* ρ *such that* $\forall X$, $\rho(X) \geq 0$ *assures the latter condition is satisfied.*

3.9.5 Formulas for the global quadratic hedging

The general solution to the problem $\min_{\theta \in \Theta} \mathbb{E}\left[(C_T - V_T^{(\theta)})^2\right]$ when $\mathcal{B}_t = \mathbb{R}$ for all t can be obtained by applying a recursive scheme. Define the following quantities, starting with

$\nu_{T+1} := 1$ and $\tilde{C}_T := C_T$:

$$\Delta_t := B_t^{-1} S_t - B_{t-1}^{-1} S_{t-1} \tag{3.72}$$

$$A_t := \mathbb{E}\left[\Delta_t^2 \nu_{t+1} | \mathcal{F}_{t-1}\right] \tag{3.73}$$

$$b_t := A_t^{-1} \mathbb{E}\left[\Delta_t \nu_{t+1} | \mathcal{F}_{t-1}\right] \tag{3.74}$$

$$\alpha_t := A_t^{-1} \mathbb{E}\left[B_t^{-1} \tilde{C}_t \Delta_t \nu_{t+1} | \mathcal{F}_{t-1}\right] \tag{3.75}$$

$$\nu_t := \mathbb{E}[(1 - b_t \Delta_t) \nu_{t+1} | \mathcal{F}_{t-1}] \tag{3.76}$$

$$\tilde{C}_{t-1} := \frac{\mathbb{E}\left[B_t^{-1}(1 - b_t \Delta_t) \tilde{C}_t \nu_{t+1} | \mathcal{F}_{t-1}\right]}{B_{t-1}^{-1} \nu_t}. \tag{3.77}$$

Then, the solution of the problem is obtained by setting

$$\theta_{t+1}^{(S)} = \alpha_{t+1} - B_t^{-1} V_t b_{t+1}. \tag{3.78}$$

Rémillard (2013) give more explicit expressions for coefficients (3.73)-(3.77) in the case of geometric random walks, which is the model used for simulations in the current paper. Indeed, assume (3.32)-(3.33) hold with log-returns z_k being i.i.d. Define $D = e^{-r/52}$, $\mu_{(1)} = D\mathbb{E}[e^{z_k}]$ and $\mu_{(2)} = D^2 \mathbb{E}[e^{2z_k}]$. Then,

$$a = 1 - \frac{\mu_{(1)}^2 - 2\mu_{(1)} + 1}{\mu_{(2)} - 2\mu_{(1)} + 1}, \tag{3.79}$$

$$A_t = B_{t-1}^2 S_{t-1}^2 (\mu_{(2)} - 2\mu_{(1)} + 1) a^{T-t}, \tag{3.80}$$

$$b_t = \frac{\mu_{(1)} - 1}{B_{t-1} S_{t-1} \left(\mu_{(2)} - 2\mu_{(1)} + 1\right)}, \tag{3.81}$$

$$\nu_t = a^{T-t+1}. \tag{3.82}$$

3.9.6 The numerical algorithm

The numerical algorithm of François et al. (2014) relies on three steps performed for each time step :

1. Compute the optimal portfolio choice on a coarse grid,
2. Interpolate the optimal portfolio choice on all nodes of the finer grid,
3. Compute the value function Ψ on all nodes of the finer grid, using the interpolated portfolio choice of step 2.

A fourth step is added to the algorithm in the current paper. In the presence of transaction costs, it might be suboptimal to rebalance the portfolio. However, the interpolation approximation of the optimal portfolio choice on the fine grid in step 2 can suggest in some cases a light rebalancing of the portfolio. When the rebalancing suggested by the algorithm is very small, it might generate transaction costs that are more harmful than the benefit obtained by the rebalancing. [12]

The current paper adds a fourth step to the algorithm. The value function approximation obtained in step 3 and the one obtained without rebalancing the portfolio are compared. The optimal strategy then involves rebalancing the portfolio only if the value function obtained from step 3 is lower than the one obtained without rebalancing.

Lattice parameters

Some parameters of the François et al. (2014) algorithm must be specified. The notation for parameters is borrowed from the latter article. The grid stretching factors are $(\lambda_Y^{(small)}, \lambda_Y^{(large)}, \lambda_V^{(small)}, \lambda_V^{(large)}) = (.6, .6, 1, 1)$. For the step 1 of the algorithm, the number of quadrature points for the three regions of the distribution are $M_{(1)} = M_{(3)} = 100$ and $M_{(2)} = 200$. For the step 3, $\tilde{M}_{(1)} = \tilde{M}_{(3)} = 200$ and $\tilde{M}_{(2)} = 300$. Putting more points near the tails is used to better capture the extreme events which contribute more heavily to the hedging penalty. The discrete set $\mathcal{O}$ over which θ_{t+1} is optimized in step 1 of the algorithm is $\mathcal{O} = \{j/99 \,|\, j = 0, ..., 99\}$. The number of grid nodes for each variable on the finer grid (step 3 of the algorithm) is

$$(\#Y_t, \#\theta_t, \#V_t) = \begin{cases} (200, 100, 200) \text{ if } n = N-1 \\ (150, 100, 150) \text{ otherwise} \end{cases} \tag{3.83}$$

More nodes are put on the first step of the recursion since it can be computed faster because of explicit formulas. [13] For the coarse grid in step 1, only a subset of the nodes of the finer grid in step 3 are retained. The proportion of nodes kept in the coarse grid from the finer grid across dimensions Y_t, θ_t and V_t is 1/3, 1/3 and 1/4.

12. As described in Zakamouline (2006), the solution to the hedging problem of Hodges & Neuberger (1989) involves a no-rebalancing region i.e. a subset of state variable values for which the optimal solution is to not rebalance the portfolio.

13. An explicit expression for $\mathbb{E}[\Psi_{T,c}^{*,CVaR} | \mathcal{F}_{T-1}]$ exists.

References

[1] Alexander, S., Coleman, T.-F., Li, Y. (2003). Derivative portfolio hedging based on CVaR. *New Risk Measures in Investment and Regulation*. Wiley.

[2] Barles, G., Soner, H.-M. (1998). Option pricing with transaction costs and a nonlinear Black-Scholes equation. *Finance and Stochastics*. **(2)**. pp. 369-397.

[3] Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London A*. **(353)**. pp. 401-409.

[4] Barndorff-Nielsen, O., Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*. **(38)**. pp. 439-455.

[5] Bellman, R., Dreyfus, S. (1962) Applied dynamic programming. *Princeton University Press*. Princeton

[6] Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*. **(81)**. pp. 637-654.

[7] Boda, K., Filar, J.-A. (2006). Time consistent dynamic risk measures. *Mathematical Methods of Operations Research*. **(63)**. pp. 169-186.

[8] Davis, M.-H.-A., Panas, V.-G., Zariphopoulou, T. (1993). European Option Pricing with Transaction Costs. *SIAM Journal of Control and Optimization*. **(31)**. pp. 470-493.

[9] Eberlein, E., Jacod, J. (1997). On the range of options prices. *Finance and Stochastics*. **(1)**. pp. 131-140.

[10] Eberlein, E., Keller, U. (1995). Hyperbolic distributions in finance. *Bernouilli*. **(1)**. pp. 281-299.

[11] Ederington, L.H. (1979). The hedging performance of the new futures markets. *Journal of Finance*. **(34)**. pp. 157-170.

[12] Fama, E.-F. (1965). The behavior of stock-market prices. *The Journal of Business*. **(38)**. pp. 34-105.

[13] Föllmer, H., Leukert, P. (1999). Quantile hedging. *Finance and Stochastics*. **(3)**. pp 251-273.

[14] Föllmer, H., Leukert, P. (2000). Efficient hedging : cost versus shortfall risk. *Finance and Stochastics*. **(4)**. pp. 117-146.

[15] François, P., Gauthier, G., Godin, F. (2014). Optimal hedging when the underlying asset follows a regime-switching Markov process. *European Journal of Operational Research*. **(237)**. pp. 312-322.

[16] Genest, C., Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*. **(44)**. pp. 1096-1127.

[17] Godin, F., Mayoral, S., Morales, M. (2012). Contingent claim pricing using a normal inverse Gaussian probability distortion operator. *Journal of Risk and Insurance*. **(79)**. pp. 841-866.

[18] Harrison, M., Pliska, S. (1981). Martingales and Stochastic Integrals in the Theory of Continuous Trading. *Stochastic Processes and Their Applications*. **(11)**. pp. 215-260.

[19] Hodges, S., Neuberger, A. (1989). Optimal replication of contingent claims under transactions costs. *Review of Futures Markets*. **(8)**. pp. 222-239.

[20] Karlis, D. (2002). An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. *Statistics & Probability Letters*. **(57)**. pp. 43-52.

[21] Kusuoka, S. (2001). On law invariant risk measures. *Advances in Mathematical Economics*. **(3)**. pp. 83-95.

[22] Martellini, L., Priaulet, P. (2002). Competing methods for option hedging in the presence of transaction costs. *Journal of Derivatives*. **(9)**. pp. 26-38.

[23] Melnikov, A., Smirnov, I. (2012). Dynamic hedging of conditional value-at-risk. *Insurance : Mathematics and Economics*. **(51)**. pp. 182-190.

[24] Prause, K. (1999). The generalized hyperbolic model : estimation, financial derivatives and risk measures. Ph.D. Thesis, Albert Ludwigs University of Freiburg, Germany.

[25] Rémillard, B. (2013). Statistical Methods for Financial Engineering. *CRC Press*. Boca Raton.

[26] Rémillard, B., Rubenthaler, S. (2013). Optimal hedging in discrete time. *Quantitative Finance*. **(13)**. pp. 819-825.

[27] Rockafellar, R.-T., Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*. **(26)**. pp. 1443-1471.

[28] Rydberg, T.-H. (1997). The normal inverse Gaussian Lévy process : simulation and approximation. *Communications in Statistics : Stochastic Models*. **(13)**. pp. 887-910.

[29] Rydberg, T.-H. (1999). Generalized hyperbolic diffusion processes with applications in finance. *Mathematical Finance*. **(9)**. pp. 183-201.

[30] Schoutens, W. (2003). Lévy processes in finance : pricing financial derivatives. *Wiley*.

[31] Schweizer, M. (1995). Variance-optimal hedging in discrete time. *Mathematics of Operation Research*. **(20)**. pp. 1-32.

[32] Whalley, A.-E., Wilmott, P. (1997). An asymptotic analysis of an optimal hedging model for option pricing with transaction costs. *Mathematical Finance*. **(7)**. pp. 307-324.

[33] Zakamouline, V. (2006). Optimal hedging of option portfolios with transaction costs. Working paper. University of Agder. Norway.

Chapitre 4

Couverture sur les marchés d'électricité

Short-term hedging for an electricity retailer

By Debbie Dupuis [1], Geneviève Gauthier [2] and Frédéric Godin [3]

Abstract

A dynamic global hedging procedure making use of futures contracts is developed for a retailer of the electricity market facing price, load and basis risk. Statistical models reproducing stylized facts are developed for the electricity load, the day-ahead spot price and futures prices in the Nord Pool market. These models serve as input to the hedging algorithm, which also accounts for transaction fees. Backtests with market data from 2007 to 2012 show that the global hedging procedure provides considerable risk reduction when compared to hedging benchmarks found in the literature.

JEL classification : G32, L94, C61

Keywords : Risk management, power markets, energy, load modeling, futures contracts.

4.1 Introduction

With the recent liberalization of electricity markets and the disentanglement of the vertical integration in the electricity supply chain in Nordic countries, continental Europe and

1. Professor at HEC Montréal, Department of Management Sciences, 3000 ch. Côte-Ste-Catherine, Montréal, Québec, Canada, H3T 2A7, debbie.dupuis@hec.ca.

2. Professor at HEC Montréal, Department of Management Sciences, 3000 ch. Côte-Ste-Catherine, Montréal, Québec, Canada, H3T 2A7, tel. 001 (514) 340-5627. genevieve.gauthier@hec.ca

3. Ph.D. student at HEC Montréal, Department of Management Sciences, 3000 ch. Côte-Ste-Catherine, Montréal, Québec, Canada, H3T 2A7, frederic.godin@hec.ca.

North America, new risks have arisen for some of the participants of the electricity markets. One such risk-facing participant is the retailer buying from wholesalers to sell to end-users. These retailers [4] sign contracts giving them the obligation to supply electricity to consumers. Retailers often need to supply a quantity of electricity at a fixed price while acquiring it at a variable market price (Von der Fehr & Hansen, 2010 and Johnsen, 2011), exposing the retailers to price risk. Furthermore, as the quantity of electricity which must be supplied to consumers is uncertain, retailers also face load (or volumetric) risk (Deng & Oren, 2006).

Electricity is not easily storable and retailers cannot build up electricity reserves upon which to draw to cover an unexpectedly high load demand or an electricity price increase. The non-storability of electric power fuels extreme price volatility as highly inelastic demand can cause spot prices to skyrocket when shortages occur. For retailers, the volatility can affect profitability since an unexpected high cost of electricity can lead to major losses. *The profit margin for a retailer is so small in relation to the price risk that the profit margin can quickly disappear if the price risk is not hedged* (NordREG, 2010). In some cases, there was eventual bankruptcy as with the Pacific Gas and Electric Company in 2001 and Texas Commercial Energy in 2003. To prevent such events, some government regulatory initiatives were even implemented to force retailers to hedge their obligation to serve electricity loads. For example, the California Public Utility Commission now requires load serving entities (LSE) to use forward contracts and options (with mandatory physical settlement) to reduce their risk exposure State of (California, 2004).

It is clear that deficient risk management can lead to financial hardship for retailers and developing effective hedging methodologies in the electricity market has become paramount. Different approaches, using different electricity derivatives, have been proposed in the literature. Deng & Oren (2006) survey available derivatives and list the papers that implement methods pertaining to each. Hedging procedures can be divided into two main categories : (i) static, and (ii) dynamic. For static hedging, hedging instruments are bought at one point in time and the hedging portfolio is never rebalanced. For dynamic hedging, the composition of the hedging portfolio is adjusted through time as additional information becomes available. Dynamic hedging procedures can be divided into two sub-categories, which we refer to as *local* and *global* hedging. Local hedging procedures minimize the risk associated with

4. Retailers is the term used on Nord Pool for these participants. On the US market, they are referred to as load serving entities.

the portfolio until the next rebalancing whereas global hedging procedures minimize the risk related to the terminal cash flow.

Several papers apply static hedging without considering load uncertainty. Stoft et al. (1998) describe simple hedging strategies with vanilla derivatives. Bessembinder & Lemmon (2002) identify the optimal position in forward contracts for electricity producers and retailers through an equilibrium scheme. Fleten et al. (2010) optimize the static futures contract position of a hydro-power electricity producer in Nord Pool. Other papers studying static hedging incorporate load uncertainty in their model. Wagner et al. (2003) and Woo et al. (2004) investigate static hedges with forward and futures contracts under different risk constraints. Deng & Xu (2009) examine hedging strategies using interruptible contracts in a one-period setting. In a series of papers, Oum et al. (2006), Oum & Oren(2009) and Oum & Oren(2010) propose a static hedging procedure maximizing the expected utility of a LSE using a portfolio of derivatives. Kleindorfer & Li (2005) optimize the expected return of an electricity portfolio corrected by a risk measure (either variance or Value-at-Risk).

To the authors' best knowledge, there exists no paper on dynamic hedging which incorporates load uncertainty. The literature on dynamic hedging strategies includes some local procedures. For example, Ederington (1979) suggests to hedge an underlying asset with its futures in a way to minimize the one-period variance of the total portfolio. Byström (2003), Madaleno & Pinho (2008), Zanotti et al. (2010), Liu et al. (2010), and Torro (2012) adapt this procedure to the electricity market, but with different model specifications for the spot and futures prices. Byström (2003) applies one-week horizon hedges on Nord Pool, comparing conditional and unconditional hedge ratios. The unconditional version of hedge ratios outperforms the conditional models. Madaleno & Pinho (2008) and Zanotti et al. (2010) compare different correlation models for the spot and futures prices to compute optimal hedge ratios on European electricity markets. Liu et al. (2010) use copulas to represent the relationship between the spot and futures prices. Torro (2012) studies the case of early dismantlement of the hedging portfolio in the Nord Pool market.

Alternative dynamic hedging schemes are discussed in Eydeland & Wolyniec (2003). For example, there is delta hedging, a method which consists in building a portfolio with value variations that mimic those of the hedged contingent claim. Eydeland & Wolyniec (2003) apply delta hedging to achieve perfect replication when a LSE hedges the price of a fixed amount of load to be served. When perfect replication cannot be achieved, they propose local mean-variance optimization to tackle hedging problems.

Local procedures are attractive because they are simple to implement. Local risk minimization procedures are myopic however as they do not necessarily minimize the risk through the entire period of exposure (see Rémillard 2013). Global hedging procedures remedy this drawback by taking into account the outcomes of all future time periods at any point in time ; they evaluate the adequacy of a hedge by looking at the terminal hedging error, i.e. at the maturity of the hedged contingent claim. The following is a non-exhaustive list of papers which study this methodology in general financial contexts. Schweizer (1995) minimizes the global quadratic hedging error in a discrete-time framework for European-type securities. Rémillard et al. (2010) extend his work for American-type derivatives. Föllmer & Leukert (1999) minimize the probability of incurring a hedging shortfall. Föllmer & Leukert (2000) minimize an expected function of the terminal hedging error. To the authors' best knowledge, developing global dynamic hedging procedures in electricity markets has not yet been attempted.

The current paper therefore seeks to fill the gap in the literature concerning global hedging procedures for electricity markets and offers three main contributions. First, a dynamic global hedging methodology is developed for a retailer trying to hedge itself with futures contracts by considering both its price and load risks. Obtaining global solutions to hedging problems is non-trivial and it often requires advanced numerical schemes. This could explain why this avenue has not yet been explored in electricity markets. We not only show that the approach is feasible, but present the first dynamic hedging strategy to account for load risk, and one of very few to account for transaction costs. Second, as our global hedging algorithm uses *weekly* futures, we develop the required *weekly* load model. We also present a statistical model for the joint dynamics of the spot and futures prices. A statistical approach using multivariate time series analysis is applied. Third, an empirical study which compares the performance of different hedging procedures on the Nord Pool market is presented.

The non-quadratic global hedging procedure developed outperforms the benchmarks in reducing the risk borne by the retailers. Hedging backtests show a significant reduction in several risk metrics applied to the weekly hedging error. Considering the case of a retailer serving 1% of the Nord Pool load, the $\mathrm{TVaR}_{1\%}$ is reduced from 172,900 Euros to 161,900 Euros if our load-basis model is used in the delta hedging procedure (see Table 4.7). When our global hedging procedure is applied, the $\mathrm{TVaR}_{1\%}$ is further shrunk by a considerable amount to 133,100 Euros.

The remainder of the paper is organized as follows. Section 4.2 presents the price and volumetric risks faced by retailers and describes the hedging procedure. Section 4.3 describes the data used for modeling purposes and presents the models for the electricity load, the spot price and futures prices. Section 4.4 describes the numerical experiments which test the efficacy of the hedging methodology. Section 4.5 presents concluding remarks. Some technical results, estimation details and goodness-of-fit tests are relegated to Appendices.

4.2 Risk exposure and hedging for retailers

In this section, we describe the risks faced by a retailer and the hedging procedure it can undertake to hedge its exposure with futures contracts.

4.2.1 Risks faced by retailers

Consider the case of a retailer forced to supply a quantity of electricity at a fixed price to end-users while buying it at a variable price on the market. Market conditions for a retailer might differ across different electricity markets and we look specifically at the Nordic electricity market Nord Pool. This market is chosen since it is one of the first to operate in a liberalized setup, and the mature markets provide some historical data less likely to include structural changes.

Assume all electricity purchases occur on the day-ahead market. In this market, the spot price is set on an hourly basis by balancing supply and demand. Suppose the retailer needs to serve the load $L_{t,d,h}$ during hour h of day d in week t, while $S_{t,d,h}$ is the Nord Pool system spot price for the corresponding period. The total load to be served during week t is thus

$$L_t = \sum_{d=1}^{7} \sum_{h=1}^{24} L_{t,d,h}. \tag{4.1}$$

The mean price paid by a retailer for the purchase of each unit of load during week t is

$$S_t^* = \frac{\sum_{d=1}^{7} \sum_{h=1}^{24} L_{t,d,h} S_{t,d,h}}{L_t}, \tag{4.2}$$

the load-weighted average of all hourly prices during the week. Assume the retailer charges a constant price Π for each unit of load. If no hedging is implemented, the retailer cash flow for weekly operations during week t is

$$L_t(\Pi - S_t^*). \tag{4.3}$$

A retailer thus faces revenue uncertainty due to (i) price risk caused by the variability of the spot price $S_{t,d,h}$, and subsequently of S_t^*, and (ii) volumetric risk caused by randomness in the total volume L_t of electricity to be served.

4.2.2 Electricity futures contracts

A retailer wishes to hedge its exposure to both price and volumetric risks with derivatives on the electricity markets. Different derivatives are available to hedge those risks : forward and futures contracts, options, weather derivatives and interruptible contracts. With the exception of forward and futures contracts, most derivatives on the Nord Pool market are traded over-the-counter, are illiquid, and are not well-suited for dynamic hedging methodology as they may be unavailable when they are required. More liquid futures and forwards are better suited for dynamic hedging procedures.

For the Nord Pool market, futures and forward contracts are traded on the NASDAQ OMX. Futures contracts provide hedging for shorter horizons (daily and weekly), while forwards cover longer periods (months, quarters and years). The current paper focuses on short-term hedging. Tables 4.1 and 4.2 present the percentage of trading days on which non-null trading volumes occur for weekly and daily base load futures. [5] Liquidity is much higher on weekly contracts with $1-$, $2-$ and $3-$week maturities. These derivatives will be used in this paper.

TABLE 4.1 – Liquidity of weekly futures

Weeks-to-maturity	1	2	3	4	5
Percentage	96%	90%	61%	29%	16%

Percentage of trading days between January 1st, 2007 and December 31th, 2012 with non-null trading volume of Nord Pool weekly futures on NASDAQ OMX.

Futures on Nord Pool are cash-settled ; no exchange of the underlying commodity occurs. The underlying asset S_T of a weekly futures maturing at week T is the arithmetic average

5. Base load means that the contracts deliver electricity during all hours of the day, in opposition to peak load contracts that only deliver electricity between 8 :00 a.m. and 8 :00 p.m.

TABLE 4.2 – Liquidity of daily futures

Days-to-maturity	1	2	3	4	5
Percentage	64%	16%	8.8%	2.6%	2.0%

Percentage of trading days between January 1st, 2007 and December 31th, 2012 with non-null trading volume of Nord Pool daily futures on NASDAQ OMX.

of the Nord Pool system spot price observed during week T :

$$S_t = \frac{1}{7 \times 24} \sum_{d=1}^{7} \sum_{h=1}^{24} S_{t,d,h}, \tag{4.4}$$

where a week starts on Monday and ends on Sunday. Contracts are traded on NASDAQ OMX during weekdays until the Friday of week $T-1$. The futures is thus not traded during its maturity week.

There is a slight mismatch between the average weekly electricity price paid by the retailer, given by (4.2), and the underlying asset of weekly futures given by (4.4). The basis ratio

$$\eta_t = \frac{S_t^*}{S_t} \tag{4.5}$$

links the former and the latter. [6] The basis ratio represents an additional source of risk which must be taken into account by the hedging procedure.

Futures contracts are marked-to-market. This means that (i) their cash flows do not occur strictly at maturity (in opposition to forward contracts) and (ii) the variation of their quote (referred to as the futures price) is reflected by the continuous transfer of funds between the margin accounts of the long and short position holders. Using futures contracts implies having to pay transaction fees and the cost of these will be accounted for in our methods. [7]

6. As shown in Appendix 4.6, the basis ratio η usually evolves between 1 and 1.05. This is explained by a higher spot price during peak hours when electricity consumption is more important.

7. Transaction fees are described at http ://www.nasdaqomx.com/commodities/Marketaccess/feelist/. Fixed annual fees for membership to the Exchange are disregarded in the current study. Variable fees which are proportional to the volume of futures transactions include Exchange fees (for trading positions) and Clearing fees (for clearing positions). Exchange fees are 0.004 EUR/MWh. Clearing fees depend on the volume of futures cleared in the most recent quarter, but they range from 0.0035 EUR/MWh to 0.0085 EUR/MWh. For illustrative purposes, a 0.004 EUR/MWh rate is used. Combining Exchange and Clearing fees, entering or clearing any long or short position is therefore approximated to cost 0.004 EUR/MWh.

4.2.3 Hedging procedure

Throughout this section, the retailer hedges its week T exposure. A self-financing investment portfolio containing a risk-free asset [8] and futures with maturity week T is set up at t_0 and rebalanced weekly until $T-1$. Since futures are traded during weekdays only, rebalancing occurs on Fridays at closing time. The closing price on Friday of week t (or Sunday if $t = T$) of the risk-free asset is $B_t = \exp(rt)$, and the closing price of the futures is $F_{t,T}$, $t = t_0, \ldots, T-1$. Since futures are cash-settled, the last futures quote on Sunday of week T is automatically set by the clearing house to $F_{T,T} = S_T$. The hedging procedure is summarized by the following algorithm :

At week t_0. An initial amount of capital V_{t_0} is allocated for hedging purposes. The retailer enters into θ_{t_0+1} long positions on the futures contract. A portion $\mathcal{M}_{t_0}$ of the initial capital is placed in the margin account required by the clearing house. Another part is used to pay transaction fees $\mathcal{C}_{t_0}$. The remainder $\mathcal{B}_{t_0}$ is invested in the risk-free asset. If V_{t_0} is insufficient to cover the margin call and fees, the money is borrowed. Since entering positions on futures contracts involves no immediate cash flows besides the amount placed in the margin,

$$V_{t_0} = \mathcal{M}_{t_0} + \mathcal{B}_{t_0} + \mathcal{C}_{t_0}.$$

Capital $\mathcal{M}_{t_0} + \mathcal{B}_{t_0} = V_{t_0} - \mathcal{C}_{t_0}$ (both inside and outside the margin) is invested (or borrowed) at the risk-free rate r. [9]

At week $t+1$, $t \in \{t_0, \ldots, T-2\}$. The total capital available for hedging (the sum of the amount placed in the margin account and in the risk-free asset) at week t before transaction costs are paid is V_t. This capital accrues interest up to week $t+1$ and is now worth

$$(V_t - \mathcal{C}_t)\frac{B_{t+1}}{B_t}.$$

The futures margin account of the retailer is adjusted from marking-to-market, [10] the amount $\theta_{t+1}(F_{t+1,T} - F_{t,T})$ is deposited (withdrawn if negative) in the margin account. The total

8. Since this paper focuses on short-term hedging, a constant weekly risk-free rate r is assumed.

9. It is assumed that the retailer can always borrow capital at the risk-free rate. Such an assumption has a limited impact ; hedging errors are very insensitive to interest rates because of the short term horizon of the hedge.

10. To simplify calculations, it is assumed that the futures are marked-to-market weekly. On NASDAQ OMX, marking-to-market is in reality performed daily. However, because maturities are short-term (and therefore accumulation of interest is small), such an approximation has only a minor impact.

capital available for hedging at week $t+1$ is therefore

$$V_{t+1} = (V_t - \mathcal{C}_t)\frac{B_{t+1}}{B_t} + \theta_{t+1}(F_{t+1,T} - F_{t,T}).$$

The retailer modifies its portfolio to hold θ_{t+2} long positions on the futures contract. Transactions fees $\mathcal{C}_{t+1}$ are paid. A margin call might be made, but it does not affect the total amount $V_{t+1} - \mathcal{C}_{t+1}$ invested at the risk-free rate.

<u>At week T</u>. The terminal hedging capital is

$$V_T = (V_{T-1} - \mathcal{C}_{T-1})\frac{B_T}{B_{T-1}} + \theta_T(S_T - F_{T-1,T}) - \mathcal{C}_T,$$

where $\mathcal{C}_T$ are clearing fees. Transaction costs are computed following $\mathcal{C}_T = 0.004|\theta_T|$ (final clearing costs) and $\mathcal{C}_t = 0.004|\theta_{t+1} - \theta_t|$ if $t < T$ with $\theta_{t_0} = 0$.

The retailer is at risk of bearing losses when the price it pays to purchase electricity is higher than the price it charges to its customers. To avoid this situation, the hedging algorithm proposed in this paper aims at minimizing risks related to electricity procurement costs incurred by the retailer. Having reliable procurement costs stabilizes the retailer's profitability.[11] Weekly futures, which are used by the retailer to hedge its exposure, allow locking in the payoff of the variable contingent claim S_T to $F_{t_0,T}$ (see Appendix 4.7).[12] However, the retailer has short positions on S_T^* (instead of S_T) because it needs to buy electricity at that price. Since $S_T^* = \eta_T S_T$, the electricity procurement target price for each unit of load bought by the retailer during week T is set to $(S_T^*/S_T)F_{t_0,T} = \eta_T F_{t_0,T}$. The retailer's cash flow at time T, given by Equation (4.3), can be separated into an unhedged cash flow $L_T(\Pi - \eta_T F_{t_0,T})$, the baseline profit margin, and a more risky component $\mathcal{L}_T(S_T - F_{t_0,T})$, the procurement costs risk :

$$\begin{aligned} L_T(\Pi - S_T^*) &= L_T(\Pi - \eta_T F_{t_0,T}) - L_T(S_T^* - \eta_T F_{t_0,T}) \\ &= L_T(\Pi - \eta_T F_{t_0,T}) - \mathcal{L}_T(S_T - F_{t_0,T}) \end{aligned}$$

where the load-basis $\mathcal{L}_T$ is the product of the load and the basis factor :

$$\mathcal{L}_t = \eta_t L_t. \tag{4.6}$$

11. Hedging procurement costs does not remove all risks ; profits are still proportional to the load. Adequate hedging of procurement costs will however prevent extreme losses.

12. This is true if transaction fees are disregarded.

The procurement costs risk can cause large losses when the price S_T peaks way above the futures price $F_{t_0,T}$. The hedging strategy aims at offsetting the variation of the quantity

$$\Psi_T = \mathcal{L}_T(S_T - F_{t_0,T}) \tag{4.7}$$

while the retailer determines the fixed price Π to extract an expected but uncertain profit. Considering the load-basis $\mathcal{L}$ (instead of the load L and the basis factor η separately) is convenient since only a single model is required for the former quantity (instead of two models for the latter).

Two main reasons motivate the decision to hedge procurement cost risk instead of the full cash flow of the retailer given by Equation (4.3). The first one is the design of futures contracts which allow locking the spot price S_T to the initial futures price $F_{t_0,T}$. It is thus more natural for futures contracts to hedge the procurement cost term in (4.7) which is based on the difference between those two quantities than the full cash flow which is driven by the retail selling price Π. The second reason is that the selling price Π is constant through the year whereas the spot price S_T^* exhibits seasonality. Hedging the full cash flow would imply that the target would be in average larger (smaller) than what is expected to be gained from the futures portfolio at times of the year when $F_{t_0,T}$ is larger (smaller) than Π. Hedging the procurement cost risk term solves this problem since the initial futures price $F_{t_0,T}$ should reflect the seasonal trend of the spot price S_T^*.

The retailer would like the terminal value of the hedging portfolio V_T to be bigger than the target Ψ_T (or at least as close as possible to it) to offset the procurement costs risk. The global hedging problem that must be solved is thus

$$\min_{(\theta_{t_0+1},\ldots,\theta_T)\in\Theta} \mathbb{E}\left[G(\Psi_T - V_T)|\mathcal{G}_{t_0}\right], \tag{4.8}$$

where $V_T = V_T(\theta_{t_0+1}, \ldots, \theta_T)$, $\mathcal{G} = \{\mathcal{G}_t | t = t_0, ..., T\}$ is the filtration that defines the information available to the retailer,[13] Θ is the set of all trading strategies available to the retailer[14] and G is a penalty function which weights and sanctions losses. Some integrability and regularity conditions might need to be satisfied to ensure that the solution exists.

13. The retailer is assumed to consider information $\mathcal{G}$ relative to past and contemporaneous load-basis, spot prices and futures prices : $\mathcal{G}_t = \sigma\{\mathcal{L}_u, S_u, F_{u,u+j} \big| 0 \leq u \leq t, j = 1, 2, 3\}$.

14. In the current paper, this consists of all $\mathcal{G}-$predictable trading strategies, meaning that θ_{t+1} is $\mathcal{G}_t-$measurable for all t.

There are numerous possibilities for the penalty function G. A standard choice in the literature is the quadratic function, $G(x) = x^2$, since it conveniently leads to semi-analytical formulas [36] and therefore enhances the tractability and the computational speed of the solution. This approach has two principal caveats : (i) the semi-analytical formulas do not take transaction fees into account ; (ii) the quadratic penalty is symmetric, such that gains and losses are equally penalized. [15] To remedy the problem of penalized gains, we also consider a semi-quadratic penalty [16]

$$G(x) = x^2 \mathbb{I}_{\{x>0\}}. \tag{4.9}$$

A retailer using this penalty tries to remove losses as much as possible and disregards gains. A drawback of using this penalty is that it leads to a substantial increase in the numerical burden. The computations are however still feasible for the current framework. The computation of solutions for problem (4.8) with penalty (4.9) is discussed in Appendix 4.8. A simulation-based algorithm is proposed to solve the Bellman equation. This algorithm can accommodate a wide class of penalty functions.

4.3 Models for the state variables

To compute the optimal trading strategy, the dynamics of the state variables $\mathcal{L}_t$ and $F_{t,T}$, the key components in the hedging problem, must be modeled. The proposed models are constructed from historical data.

4.3.1 Load-basis

We assume that the load the retailer must supply is proportional to the entire system load on the Nord Pool spot market. [17] This proportionality assumption, which is justified by a high correlation between firm load and market load, is also found in [8] for the Texas electricity market.

15. Ni et al. (2012) add a linear term to the quadratic penalty which makes it asymmetric.
16. This penalty is also considered in a hedging problem by François et al. (2012).
17. If the internal load data of the retailer do not support this assumption, the load model should be reworked. For differences between the load consumption patterns across the four countries that are part of the Nord Pool market, see Huovila (2003).

Load forecasting has been studied in the literature. Weron (2006) surveys different load forecasting methods and divides them in two classes : artificial intelligence models (neural networks, fuzzy logic, support vector machines) and statistical models (regression models, exponential smoothing, Box-Jenkins type time series models). Load forecasting methods are split into three different segments : short-term load forecasting (STLF), medium-term load forecasting (MTLF) and long-term load forecasting (LTLF). STLF is interested in hourly forecasts up to one week ahead, MTLF considers forecasts from one week to one year ahead and LTLF considers even longer horizons. The vast majority of the load forecasting literature considers STLF (Hahn et al., 2009), but MTLF has attracted more attention recently. Gonzalez et al. (2008) use a combination of neural networks and Fourier series to represent respectively the trend and the cyclical fluctuation of the monthly load in the Spanish market. In their paper, Fourier series outperform neural networks in their predictive ability for the cyclical load fluctuations. Abdel-Aal (2008) compares the use of neural and abductive networks to forecast the monthly load supplied by a power utility based in Seattle. Abdel-Aal & Al-Garni (1997) compare the use of univariate ARIMA process, abductive networks and multivariate regression models incorporating demographic, economic and weather related covariates to forecast the monthly domestic energy consumption in the Eastern province of Saudi Arabia. ARIMA processes outperformed their competitors in their study. Barakat & Al-Qasem(1998) propose a regression model with time and temperature as covariates to forecast the weekly load on the Riyadh system (Saudi Arabia).

To the authors' best knowledge, no MTLF has been attempted in the literature for the weekly load in Nord Pool. A parametric statistical model for load dynamics on Nord Pool, that supports our hedging methodology, is now presented.

Load-basis data

Time series of hourly load (in MegaWatt-hours, MWh) and hourly day-ahead spot price (in Euros) on Nord Pool for the January 1st, 2007 to July 29th, 2012 period are obtained through the Nord Pool FTP server. [18] The hourly load is aggregated as shown in (4.1) and yields 291 weekly load observations. The resulting load series L and basis ratio series η defined by Equation (4.5) are then combined to obtain the load-basis series $\mathcal{L}$ in (4.6).

18. Nord Pool uses the expression "turnover" to designate the load.

The most salient feature of the load time series is a seasonal pattern, both in the mean and in the variance. Autocorrelation between consecutive load departures from its trend is also present. The model chosen for the dynamics of the load-basis $\mathcal{L}$ (observed in Figure 1) is thus :

$$
\begin{aligned}
\mathcal{L}_t - g(t) &= \gamma(\mathcal{L}_{t-1} - g(t-1)) + \sqrt{v(t)}\epsilon_t^{(\mathcal{L})} && (4.10)\\
g(t) &= \beta_0 + \sum_{j=1}^{P} \beta_j C_t^{(\sin,j)} + \sum_{j=1}^{P} \beta_{j+P} C_t^{(\cos,j)} && (4.11)\\
\log v(t) &= \alpha_0 + \sum_{j=1}^{Q} \alpha_j C_t^{(\sin,j)} + \sum_{j=1}^{Q} \alpha_{j+Q} C_t^{(\cos,j)} && (4.12)
\end{aligned}
$$

where $\epsilon^{(\mathcal{L})}$ is a strong standardized Gaussian white noise. The g function represents the seasonal trend of the load-basis level and its fitted value is represented by the dashed line in Figure 1. The v function characterizes the trend in the variance of seasonally corrected load-basis observations and the square root of its fitted value is represented by the dashed line in Figure 4.2. Terms of a Fourier expansion

$$
C_t^{(\sin,j)} = \sin\left(\frac{3\pi}{2} + \frac{2\pi jt}{365.25/7}\right), \quad C_t^{(\cos,j)} = \cos\left(\frac{3\pi}{2} + \frac{2\pi jt}{365.25/7}\right)
$$

are used to capture yearly cycles (see also Gonzalez et al., 2008). The γ parameter in Equation (4.10) represents the autocorrelation in seasonally corrected load-basis observations. To preserve the Markov property, only one lag is considered.[19] $\gamma, \beta_0, \ldots, \beta_{2P}, \alpha_0, \ldots, \alpha_{2Q}$ are the parameters to be estimated.

Estimation of model parameters

The model estimation is performed in two steps[20]. The first consists in estimating γ and $\beta_0, \ldots, \beta_{2P}$ by quasi-maximum likelihood under the assumption that $v(t)$ is constant. The optimal number $P = 3$ of Fourier terms in the mean trend is chosen using the cross-validation procedure described in Appendix 4.9.1. Table 4.3 gives estimated parameters and their standard errors for this step. Figure 1 shows the load series L, the load-basis series $\mathcal{L}$ and the estimated load-basis seasonality trend g. Even if the variance v is presumed

19. Otherwise each additional lag would have to be included as a state variable.

20. Results in Appendix 4.9.2 show that the fitted model is good and a more numerically challenging single-step estimation was thus not attempted.

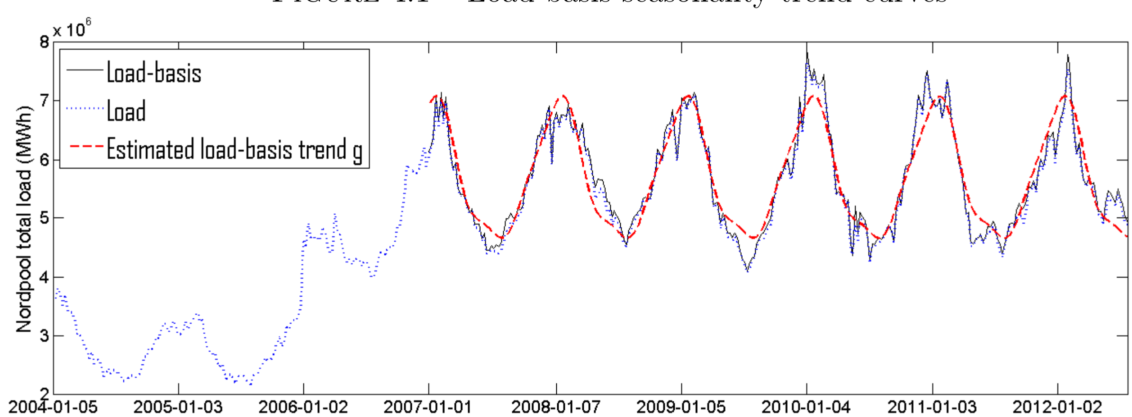
constant during the estimation of the trend parameters [21], the overall trend seems reasonably captured. The corrected load is much larger in winter than in summer ; this is expected given the winter heating requirements for Scandinavian countries. The autocorrelation parameter γ is estimated at 0.68, this large value indicating a high persistence in load deviations from the trend.

TABLE 4.3 – Load-basis seasonality trend parameters

Parameter	γ	$\beta_0 \times 10^{-6}$	$\beta_1 \times 10^{-6}$	$\beta_2 \times 10^{-6}$	$\beta_3 \times 10^{-6}$	$\beta_4 \times 10^{-6}$	$\beta_5 \times 10^{-6}$	$\beta_6 \times 10^{-6}$
Estimated Value	0.68	5.70	-1.13	-0.15	0.06	0.24	0.10	0.11
Standard Error	0.04	0.04	0.06	0.05	0.04	0.06	0.05	0.04

Estimated parameters and their standard error for the load-basis seasonality trend g defined by Equation (4.11). Observations between January 1st, 2007 and July 29th, 2012. Estimated parameter variance is obtained through the inverse of the observed Fisher information matrix.

FIGURE 4.1 – Load-basis seasonality trend curves



Observed total weekly load on the Nord Pool market as defined by (4.1), corresponding load-basis $\mathcal{L}$ and fitted seasonality trend $g(t)$. Observations between January 1st, 2007 and July 29th, 2012. Load data before 2007 are also included to show the shift in the overall system load level and justify the use of data starting from January 2007.

Once the trend parameters are estimated, proxy values for $\sqrt{v(t)}\epsilon_t^{(\mathcal{L})}$, denoted $\sqrt{\hat{v}(t)}\hat{\epsilon}_t^{(\mathcal{L})}$, can be computed using Equation (4.10). Those proxies serve as input in the second step which consists in estimating $\alpha_0, \ldots, \alpha_{2Q}$ by maximum likelihood (ML). The optimal number $Q = 2$

21. At this step, the constant estimated volatility is $\sqrt{\hat{v}} = 2.264 \times 10^5$.

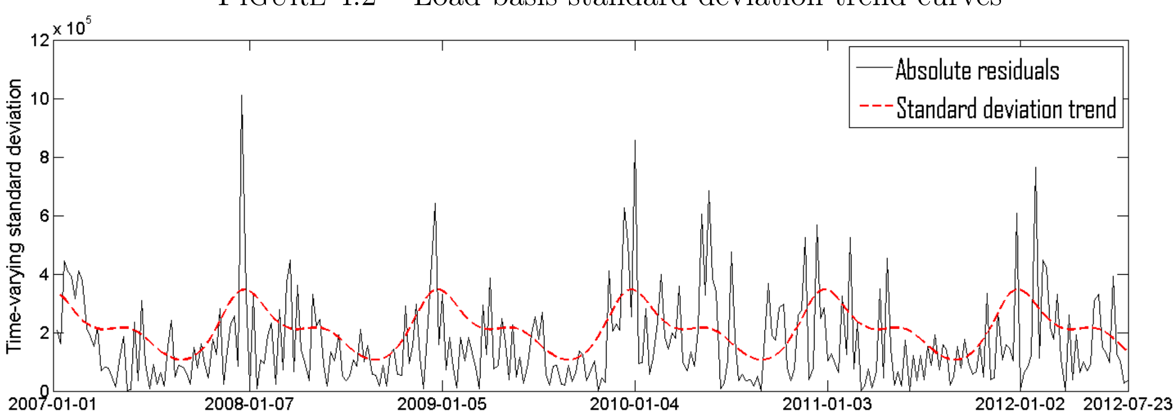
of Fourier terms in the variance trend is selected through the cross-validation procedure described in Appendix 4.9.1. Table 4.4 presents the estimated parameters for the variance model (4.12). Figure 4.2 shows the estimated standard deviation trend $\sqrt{\hat{v}_t}$ (dashed curve) and the absolute value of the $\sqrt{\hat{v}(t)}\hat{\epsilon}_t^{(L)}$ proxies (full curve). The peak in volatility occurs in the beginning of winter, while the lowest volatility is observed during the end of the summer. Goodness-of-fit tests that confirm the adequacy of the load model are found in Appendix 4.9.2.

TABLE 4.4 – Load-basis variance trend

Parameter	α_0	α_1	α_2	α_3	α_4
Estimated Value	24.40	-0.72	-0.38	0.49	-0.32
Standard Error	0.08	0.11	0.11	0.11	0.12

Estimated parameters and their standard error for the load-basis variance trend v defined by Equation (4.12). Observations between January 1st, 2007 and July 29th, 2012. Estimated parameter variance is obtained through the inverse of the observed Fisher information matrix.

FIGURE 4.2 – Load-basis standard deviation trend curves



Realized absolute load-basis volatility $\sqrt{\hat{v}(t)}|\hat{\epsilon}_t^{(\mathcal{L})}|$ and fitted standard deviation trend $\sqrt{\hat{v}}$ as defined by Equation (4.12). Data between January 1st, 2007 and July 29th, 2012.

Load-basis forecasting from incomplete information

The selection of θ_{t+1}, the number of futures shares detained in the portfolio from the Friday of week t until the Friday of week $t+1$, is based on $\mathcal{L}_t$, the weekly load-basis on week

t. However $\mathcal{L}_t$ is only observed at midnight on Sunday of week t, and not at the closing time of markets on Friday. What is observed at the latter time is the sum of hourly loads from the beginning of week t to 4 :00 p.m. on Friday :

$$\tilde{L}_t = \sum_{d=1}^{4} \sum_{h=1}^{24} L_{t,d,h} + \sum_{h=1}^{16} L_{t,5,h}.$$

When θ_{t+1} is selected, the value of $\mathcal{L}_t$ must thus be forecast using $\tilde{L}_t$. The accuracy of several forecasting models were compared through a cross-validation test and the model

$$\hat{\mathcal{L}}_t = \tilde{L}_t \times \left(\ell + c \sum_{j=1}^{q} \frac{\mathcal{L}_{t-j}}{\tilde{L}_{t-j}} \right) \tag{4.13}$$

produced the lowest out-of-sample forecasting RMSE. The out-of-sample mean absolute percentage error (MAPE) is 1.12%. [22] The parameter ℓ drives the long-term average of the ratio $\mathcal{L}_t/\tilde{L}_t$, while the autoregressive coefficient c characterizes the dependence of the current ratio on previous ratios. The estimated parameters obtained when re-estimating with the full dataset are $\hat{\ell} = 0.716$, $\hat{c} = 0.171$ and $\hat{q} = 3$. The long-term average of the $\mathcal{L}_t/\tilde{L}_t$ ratio is given by $\hat{\ell}/(1 - \hat{q}\hat{c}) = 1.47$. [23]

4.3.2 Futures and spot price

In this section, time series of futures prices are modeled. Modeling the relation between the spot and futures prices in the context of electricity markets is complicated by the fact that electricity is not storable. This prevents the use of the usual cash-and-carry scheme to price futures contracts.

Solving problem (4.8) requires a model that completely specifies the stochastic dynamics of futures prices. An important strand of the literature studies the risk premium on electricity futures contracts. [24] Although these papers provide relevant information concerning the relation between the spot and futures prices, their models do not directly fit our needs

22. The out-of-sample MAPE obtained by using the naive benchmark $\hat{\mathcal{L}}_t := \ell \tilde{L}_t$ is 1.21%. Obtaining good load forecasts is crucial to the success of the hedging procedure and the small improvement of model (4.13) over the naive method justifies its use.

23. This is consistent with what is expected ; since $\mathcal{L}$ and $\tilde{L}$ are respectively approximately the sum of 168 and 112 hourly loads, the long-term average of the ratio should revolve around $168/112 = 1.5$.

24. For example, Lucia & Torro (2011) study the behavior of the risk premium on Nord Pool weekly futures with an ex-post econometric model taking into account hydropower reservoir levels.

since they do not characterize the futures prices dynamics. Several approaches are however proposed in the literature for this purpose and they will now be discussed.

Despite the fact that electricity is not storable and not openly traded, some authors follow the risk-neutral approach commonly used in finance. The dynamics of the spot price are modeled and a martingale measure is selected to compute futures prices as an expectation of the discounted cash flows. Benth et al. (2008) use a linear combination of non-Gaussian Ornstein-Uhlenbeck processes to represent the stochastic variability of the spot. They use the Esscher transform to compute futures prices. Coulon et al. (2012) propose a structural factor model encompassing natural gas price and electricity load to characterize spot price dynamics on the ERCOT electricity market. They use the Girsanov transform to compute derivatives prices.

Besides the non-storability of electricity, there is another potential pitfall with the risk-neutral approach to price futures. On the Nord Pool market, a principal component analysis applied to weekly futures returns shows that the spot price might be driven by factors different than those driving futures prices Benth et al. (2008). The martingale measure approach discounting the expected spot price to obtain the futures price might thus be inappropriate. This result is consistent with the study of Koekebakker & Ollmar(2005) who uses principal component analysis to propose a multi-factor model for forward returns. They find that the number of factors necessary to represent the full forward curve is much larger for electricity futures on the Nord Pool market than for other commodities ; the correlation between short-term and long-term electricity forward prices is smaller than in other markets.

Benth et al. (2008) also suggest adapting the Heath-Jarrow-Morton framework to electricity markets. Under such a methodology, the dynamics of forward prices that deliver an infinitesimal volume of electricity are directly specified. However, futures prices, which are really swap prices in the context of electricity markets, suffer from severe intractability issues under this model and we did not retain this approach.

The third method proposed in Benth et al. (2008) is to find a statistical model that reproduces the dynamics of the observed futures returns. This approach is followed in the current paper since it better suits our need to fully specify the distribution and the stochastic dynamics of futures and spot prices of electricity. Furthermore, this approach reproduces stylized facts.

Daily prices of futures on NASDAQ OMX are provided by Bloomberg. Since futures prices vary during the day, closing prices are used.

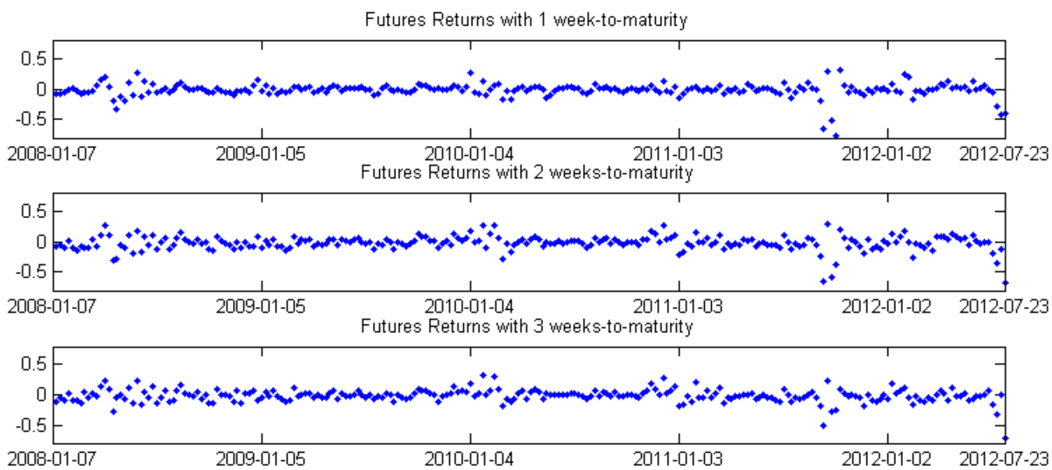
Our model

For a market participant hedging the cost of electricity at maturity week T, the sequence of observed futures prices that must be modeled is $\{F_{T-j,T} | j = 3, 2, 1, 0\}$. We propose a multivariate time series model for the joint dynamics of the spot and futures prices. As with financial assets, futures price returns are modeled (instead of the futures prices) as they are more likely to be stationary. Futures returns defined by

$$\epsilon_{t,T} = \log(F_{t,T}/F_{t-1,T}) \tag{4.14}$$

are shown in Figure 3 for $t = T, T-1, T-2$.

FIGURE 4.3 – Futures price returns for different times-to-maturity



Time series for NASDAQ OMX electricity weekly futures returns (on Nord Pool day-ahead spot price) as defined by (4.14) between January 1st, 2007 and July 29th, 2012. The trivariate time series illustrated contains 290 observations.

Futures price returns exhibit autocorrelation, volatility clustering and fat tails. These features suggest a multivariate AR-GARCH process with innovations drawn from a fat-tail distribution. For the latter, we choose a Normal Inverse Gaussian (NIG) distribution. More specifically, for $i = 0, 1, 2$, the trivariate AR(1)-GARCH(1,1) with NIG innovations is

$$\epsilon_{t,t+i} = \mu_i + a_i\, \epsilon_{t-1,t-1+i} + \sigma_{i,t} z_{i,t} \tag{4.15}$$

$$\sigma^2_{i,t+1} = \min\{\varsigma^2, \kappa_i + \gamma_i \sigma^2_{i,t} + \xi_i \sigma^2_{i,t} z^2_{i,t}\} \tag{4.16}$$

where $\mathbf{z}_t = (z_{0,t}, z_{1,t}, z_{2,t})$ has the following properties :

if $s \neq t$, $\mathbf{z}_t$ and $\mathbf{z}_s$ are independent ;

$z_{i,t}$ are drawn from a standardized [25] NIG(α_i, β_i) distribution ;

$z_{0,t}$, $z_{1,t}$ and $z_{2,t}$ are linked by the Gaussian copula.

A bound ζ is used on the volatility to ensure that futures prices are square-integrable. [26] The a_i parameter represents autocorrelation of futures returns while μ_i adjusts their long-term expected level. The κ_i parameter adjusts the long-term level of futures return volatility, the γ_i characterizes the persistence in returns volatility, and ξ_i determines how shocks associated with current returns affect the current volatility. The NIG parameter α_i drives the tail thickness in the distribution of the futures return while the β_i drives its asymmetry.

Model estimation

A two-step procedure is applied. First, the parameters for the three marginal AR(1)-GARCH(1,1) processes ($\epsilon_{t,t+1}, \epsilon_{t,t+2}$ and $\epsilon_{t,t+3}$) are estimated by ML. [27] Plugging the estimated parameters in (4.15)-(4.16) yields proxy values $\hat{\mathbf{z}}_t$ for $\mathbf{z}_t$. Then, the proxies are used to estimate the parameters of the Gaussian copula. Letting F_{NIG} denote the cdf associated with the NIG distribution and applying the Rosenblatt (1952) transform to the proxy $\hat{\mathbf{z}}_t$ yields a series of approximatively independent uniformly distributed observations

$$\mathbf{U}_t = \left(F_{\text{NIG}(\hat{\alpha}_0,\hat{\beta}_0)}(\widehat{z_{0,t}}), F_{\text{NIG}(\hat{\alpha}_1,\hat{\beta}_1)}(\widehat{z_{1,t}}), F_{\text{NIG}(\hat{\alpha}_2,\hat{\beta}_2)}(\widehat{z_{2,t}})\right)$$

drawn from the Gaussian copula. ML is used and the closed-form solution is

$$\hat{\rho}_{i,j} = corr(\Phi^{(-1)}(U_{i,t}), \Phi^{(-1)}(U_{j,t})),$$

where *corr* is the sample correlation and $\Phi^{(-1)}$ is the inverse cdf of a standard Gaussian variable.

25. A standardized NIG is a NIG distribution with mean 0 and variance 1. Such a distribution only has two free parameters : α and β. Note that these α and β should not be confused with those used in the load-basis model in Section 3.1.1.

26. More precisely, the condition

$$\sigma_{i,t} < \frac{\alpha_i - \beta_i}{2} \text{ a.s.} \tag{4.17}$$

is necessary and sufficient to obtain $\mathbb{E}\left[e^{2\epsilon_{t,t+i}}\right] < \infty$. Thus, the volatility bound ς combined with the additional constraints $\alpha_i > \varsigma$, and $\beta_i \in (-\alpha_i, \alpha_i - 2\varsigma]$ assure (4.17) is satisfied.

27. The proxy for the initial value for the volatility $\hat{\sigma}_{i,0}$ is its long-term stationary average. The bound is set at $\varsigma = 0.6$ since such a constraint is not numerically binding with the available data.

Parameter estimates are shown in Tables 4.5 and 4.6. The negative mean parameters μ_i indicate the futures market is in contango. The GARCH parameters γ_i and ξ_i are highly significant, confirming the presence of volatility clustering in futures returns. The autocorrelation parameters a_i are also all significant and positive, indicating that futures returns are partially predictable. The α_i parameters are all low (smaller than 2) so the kurtosis of futures returns is more pronounced than in a Gaussian distribution (which corresponds to an infinite α). The correlation parameters of the Gaussian copula are all higher than 0.65, indicating a somewhat high correlation between futures returns across the time-to-maturity dimension.

Goodness-of-fit tests that confirm the adequacy of the futures return model are found in Appendix 4.9.3. Futures returns and load-basis innovations are assumed to be independent. Statistical tests in Appendix 4.9.4 validate this assumption.

TABLE 4.5 – Futures return parameters

Parameter	$i = 1$	$i = 2$	$i = 3$
$\mu_i \times 10^2$	−0.722 (0.001)	−1.566 (0.003)	−1.265 (0.002)
a_i	0.215 (0.005)	0.143 (0.005)	0.073 (0.004)
$\kappa_i \times 10^2$	0.177 (3×10^{-5})	0.124 (3×10^{-5})	0.106 (2×10^{-5})
γ_i	0.282 (0.018)	0.577 (0.008)	0.603 (0.010)
ξ_i	0.500 (0.027)	0.373 (0.009)	0.340 (0.010)
α_i	1.097 (0.009)	1.270 (0.089)	1.236 (0.022)
β_i	−0.108 (0.001)	−0.056 (0.099)	0.009 (0.006)

Estimated parameters (standard error) for futures returns model defined in (4.15)-(4.16). Observations between January 1st, 2007 and July 29th, 2012 for futures with $i = 1$, 2 and 3 weeks to maturity.

4.4 Performance assessment

We carry out numerical experiments to assess the performance of the hedging strategy given by solutions of problem (4.8). We propose two different hedging procedures : (i) the hedging methodology which solves problem (4.8) with $G(x) = x^2$ is referred to as quadratic dynamic global hedging (QDGH) ; (ii) the methodology solving that same problem but without penalizing the gains, i.e. using (4.9), is called semi-quadratic dynamic global hedging

Table 4.6 – Futures return copula parameters

Parameter	$\rho_{0,1}$	$\rho_{0,2}$	$\rho_{1,2}$
Estimate (Standard Error)	0.76 (0.03)	0.67 (0.04)	0.88 (0.01)

Estimated parameters (standard errors) for the Gaussian copula linking futures returns. $\rho_{i,j}$ links returns on futures with respectively $i+1$ and $j+1$ weeks to maturity. Observations between January 1st, 2007 and July 29th, 2012.

(SQDGH). The benchmarks are described in Section 4.4.1 while the backtests are explained in Section 4.4.2.

4.4.1 Benchmarks

Delta Hedging.

If the load to be served by the retailer is known with certainty and no transaction fees exist, the delta hedging strategy proposed by Eydeland & Wolyniec (2003) completely eliminates the price risk borne by the retailer by locking in the spot price to $F_{t_0,T}$ (see Appendix 4.7). This strategy is adapted to the case of a stochastic load by hedging the expected load-basis, i.e. the retailer enters into

$$\theta_{t+1} = \frac{B_{t+1}}{B_T}\mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_t\right] \tag{4.18}$$

long positions in the futures contract at time t to cover its exposure at time T. Improved delta hedging (IDH) uses the load-basis model (4.10)-(4.12) to compute $\mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_t\right]$ in (4.18).

To quantify the impact of using the (4.10)-(4.12) load-basis model in the hedging algorithm, alternative load-basis models are also proposed to compute $\mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_t\right]$. For example, one may state that a good prediction of the expected load-basis in a near future is the last observed load-basis. This points to the first alternative, the naive delta hedging (NDH), which uses the naive prediction model

$$\mathbb{E}[\mathcal{L}_{t+1}^{(NDH)}|\mathcal{G}_t] = \mathcal{L}_t^{(NDH)}.$$

The second alternative, referred to as delta hedging (DH), uses a load-basis model inspired from [41] where the latent variable found in their model is removed for simplicity. Their model specifies the load dynamics, but is applied here to the load-basis. More specifically, the load-basis model for DH is

$$\mathcal{L}_{t+1}^{(DH)} = \mathcal{L}_t^{(DH)} + \gamma^{(DH)}(\bar{\mathcal{L}}_{m_{t+1}} - \mathcal{L}_t^{(DH)}) + \mathcal{E}_{t+1}$$

where $\mathcal{E}$ is a Gaussian white noise, $\bar{\mathcal{L}}_m$ is the mean value of the load-basis during the m^{th} month of the year ($m = 1, \ldots, 12$) in the estimation set, m_{t+1} is the month associated with the week $t+1$ and $\gamma^{(DH)}$ is estimated by ML. We find $\hat{\gamma}^{(DH)} = 0.3477$.

Local Minimal Variance Hedging (LMVH).

The objective of this strategy, which is based on the Ederington (1979) scheme, is to construct a portfolio of futures whose variation mimics the variation of the spot price as closely as possible for the current period. More precisely, for each unit of load to serve, the retailer would detain ϑ_{t+1} units of futures at time t, where ϑ_{t+1} minimizes

$$\mathbb{V}ar\left[(S_{t+1} - S_t) - \vartheta_{t+1}(F_{t+1,T} - F_{t,T})|\mathcal{G}_t\right].$$

This yields the solution $\vartheta_{t+1} = \mathbb{C}ov\left[S_{t+1}, F_{t+1,T}|\mathcal{G}_t\right]/\mathbb{V}ar\left[F_{t+1,T}|\mathcal{G}_t\right]$. To adapt this scheme to the case of stochastic load, the retailer hedges its expected load-basis by detaining at time t,

$$\theta_{t+1}^{(LMVH)} = \mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_t\right] \frac{\mathbb{C}ov\left[S_{t+1}, F_{t+1,T}|\mathcal{G}_t\right]}{\mathbb{V}ar\left[F_{t+1,T}|\mathcal{G}_t\right]} \tag{4.19}$$

long positions in the futures contract to cover its exposure at time T. Many different models are used in the literature to compute the conditional variance and covariance in (4.19). We compute these quantities with the futures model (4.15) for consistency and refer to the approach as local minimal variance hedging (LMVH).

Static Hedging.

Since many papers are devoted to static hedging procedures, we include them in our study. To apply static hedging (SH), the retailer identifies the solution to problem (4.8) under the constraint $\theta_{t_0+1} = \ldots = \theta_T$. We use the semi-quadratic penalty (4.9) and identify the optimal trading strategy through simulation.

4.4.2 Backtests

In all tests, the initial value of the portfolio V_{t_0} is set to 0 and the annualized continuously compounded risk free rate is $r = 0.0193$ [28]. The case of a retailer serving 1% of the Nord Pool load is considered.

In-sample backtest

In this experiment, our global hedging and the benchmarks are applied to historical data during the 287 weeks over the January 29th, 2007 to July 23th, 2012 period. Hedging errors $\Psi_T - V_T$ are recorded at the end of week T and the performance of the various approaches are compared through the following metrics :

$$\text{RMSE} = \sqrt{\frac{1}{287}\sum_{T=1}^{287}(\Psi_T - V_T)^2}, \tag{4.20}$$

$$\text{Semi-RMSE} = \sqrt{\frac{1}{287}\sum_{T=1}^{287}\left((\Psi_T - V_T)\mathbb{I}_{\{\Psi_T > V_T\}}\right)^2}, \tag{4.21}$$

$$\text{TVaR}_\alpha = \frac{\sum_{T=1}^{287}(\Psi_T - V_T)\mathbb{I}_{\{\Psi_T - V_T \geq q_{(1-\alpha)}\}}}{\sum_{T=1}^{287}\mathbb{I}_{\{\Psi_T - V_T \geq q_{(1-\alpha)}\}}}. \tag{4.22}$$

where $\text{VaR}_\alpha = q_{(1-\alpha)}$ is the quantile of level $1-\alpha$ of hedging errors $\Psi_T - V_T$. Results are reported in Table 4.7.

The main result is that the semi-quadratic SQDGH outperforms all other methods in terms of risk reduction ; it reduces the semi-RMSE, the $\text{TVaR}_{5\%}$ and the $\text{TVaR}_{1\%}$ by 2,420 Euros, 9,340 Euros and 28,800 Euros , respectively (i.e. by 9.3%, 9.4% and 17.8% in relative measurement), with respect to IDH, the best benchmark. Those improvements can be attributed to using global hedging procedures instead of delta hedging since both approaches share the same load model. To put these numbers in context, the mean weekly procurement costs of electricity (the average of $\mathcal{L}_T S_T$ for the January 2007 to August 2012 period) for the considered retailer is $2.35M$ Euros. Von der Fehr & Hansen (2010) identify a retail price mark-up ranging between 7.2% and 13% over the wholesale price for fixed-price contracts in Norway. Using a 10% mark-up for ballpark calculations, this leaves the retailer with an average weekly margin of $235,000$ Euros to cover expenses and profit ; average profits will be

28. The average overnight EURO LIBOR rate between January 1st, 2007 and July 29th, 2012.

Table 4.7 – In-sample backtest results

Model	SQDGH	QDGH	IDH	DH	NDH	STAH	LMVH	NOH
Mean	6.107	7.641	7.696	8.176	7.069	8.240	−18.95	−70.73
RMSE	26.20	27.65	27.36	31.37	35.03	28.05	274.8	474.1
Semi-RMSE	23.68	26.30	26.10	29.75	32.15	26.99	208.4	319.5
$\text{VaR}_{5\%}$	48.97	52.41	54.53	60.56	73.85	52.93	414.0	688.7
$\text{VaR}_{1\%}$	115.4	130.3	129.2	142.7	153.5	146.9	1206	1471
$\text{TVaR}_{5\%}$	89.97	100.3	99.31	115.3	121.0	103.5	784.5	1201
$\text{TVaR}_{1\%}$	133.1	163.1	161.9	172.9	180.9	159.2	1273	1831

Hedging error risk metrics for the in-sample backtest (in 1000 Euros). Semi-quadratic dynamic global hedging (SQDGH), quadratic dynamic global hedging (QDGH), Delta Hedging (DH), Improved Delta Hedging (IDH), Naive Delta Hedging (NDH), Static Hedging (STAH), Local Minimal Variance Hedging (LMVH) and No hedging (NOH), i.e. $\theta_t^{(NOH)} = 0$ for all t.

a fraction of that amount. SQDGH reduces the 1% worst-scenarios average loss with respect to IDH by $28,800$ Euros, a substantial fraction of average profits.

Note that IDH benefits from our load-basis model (4.10)-(4.12). The added value of the latter model is isolated by comparing IDH with DH and NDH. The $\text{TVaR}_{1\%}$ is reduced from $180,900$ Euros for the NDH to $172,900$ Euros for the DH, and further reduced to $161,900$ Euros for the IDH. This illustrates the importance of having an accurate load-basis model and the benefits provided by the model (4.10)-(4.12) in terms of risk reduction.

The combined reduction in $\text{TVaR}_{1\%}$ due to methodology presented in this paper obtained by comparing SQDGH and NDH is $47,800$ Euros, with a combined reduction in semi-RMSE of $8,470$ Euros.

It is also interesting that the mean hedging error is lower for SQDGH than for all other models except LVMH and NOH. This indicates the risk reduction yielded by the SQDGH method is not obtained at the expense of a lesser profitability. The LMVH and NOH methods are the two most profitable on average, but they yield extremely poor results in terms of risk. The poor performance of the LMVH method is explained by positions in the futures that are significantly too low. Indeed, since the cash-and-carry relationship of futures price and the spot price does not hold in this market, the correlation between spot price and futures price variations are much lower than in other markets. This reduces the $\theta^{(LMVH)}$ position and produces under-hedging. Because the Nord Pool electricity futures market is in contango, [29] under-hedging produces higher average profits than full hedging.

In terms of semi-RMSE, STAH underperforms IDH, QDGH and SQDGH, showing the benefits of dynamic hedging over a static procedure.

Out-of-sample backtest

This experiment is a rolling-window out-of-sample test. This replicates more realistic application conditions where future observations cannot be used to estimate state variable models. The three-week hedging procedure is applied weekly between December 28, 2009 and July 23, 2012. For each iteration of the test (an iteration corresponds to the starting point of the hedging procedure), all load and futures price models are estimated with the historical data contained in the three previous years (the estimation set). The hedging algorithm is then applied out-of-sample on the three weeks following the estimation set, and the terminal hedging error is recorded. There are 132 iterations in total. Descriptive statistics and risk metrics (4.20)-(4.22) applied to hedging errors for the out-of-sample backtest are given Table 4.8.

Once again, SQDGH outperforms all the benchmarks, reducing the Semi-RMSE by 7.0%, the $\text{TVaR}_{5\%}$ by 9.6% and the $\text{TVaR}_{1\%}$ by 20.0% with respect to the best competitors (respectively IDH, IDH and STAH). The SQDGH is therefore the best hedging method among all the proposed methods in both the in-sample and the out-of-sample backtests.

4.5 Conclusion

A dynamic global hedging methodology involving futures contracts is developed to allow retailers to cover their exposure to price and load risk. Global hedging procedures have received little or no attention in the electricity markets literature because they often yield solutions which are computationally more complex than their local counterparts. We show that the approach is not only feasible but easily allows us to account for load uncertainty, basis risk and transaction costs when seeking the optimal trading strategy.

29. The average 3-weeks futures price is 7.7% higher than the arithmetic average spot price for the January 2007 to July 2012 period.

Table 4.8 – Out-of-sample backtest results

Model	SQDGH	QDGH	IDH	DH	NDH	STAH	LMVH	NOH
Mean	13.12	13.88	13.68	14.82	11.29	15.22	-0.608	-59.94
RMSE	34.02	35.64	35.15	42.04	42.64	37.45	347.5	585.5
Semi-RMSE	31.79	34.62	34.19	40.10	40.15	36.52	275.4	409.0
$\text{VaR}_{5\%}$	82.94	88.76	92.51	104.7	97.51	103.9	678.7	907.9
$\text{VaR}_{1\%}$	113.9	121.7	121.5	150.7	155.5	119.9	1208	1633
$\text{TVaR}_{5\%}$	105.1	118.9	116.3	137.0	134.9	121.4	1040	1483
$\text{TVaR}_{1\%}$	119.0	149.1	150.0	166.1	168.8	148.7	1311	2005

Hedging error risk metrics for the out-of-sample backtest (in 1000 Euros). Semi-quadratic dynamic global hedging (SQDGH), quadratic dynamic global hedging (QDGH), Delta Hedging (DH), Improved Delta Hedging (IDH), Naive Delta Hedging (NDH), Static Hedging (STAH), Local Minimal Variance Hedging (LMVH) and No hedging (NOH), i.e. $\theta_t^{(NOH)} = 0$ for all t.

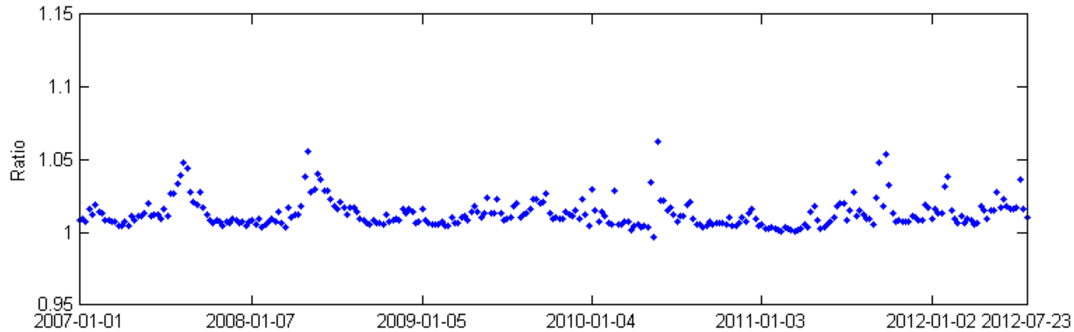
Statistical models were proposed for the load to be served by the retailer, the electricity spot price and futures contract prices on the Nord Pool market. Those models were built from weekly historical data and reproduce their stylized facts. The load basis model accounts for seasonality in the mean and the variance, as well as autocorrelation in seasonally corrected shocks. The proposed model for futures price returns, a multivariate AR(1)-GARCH with NIG innovations, exhibits stochastic volatility, partially predictable returns and fat tails. Multiple goodness-of-fit tests validate the adequacy of all models developed.

Backtests using historical market data show the superiority of the semi-quadratic global hedging procedure compared to various benchmarks of the literature in terms of risk reduction.

4.6 Appendix : basis ratio

The weekly average price S_t^* paid for electricity differs from the weekly arithmetic average price S_t, which is the underlying asset of weekly futures. The extent to which S_t^* and S_t differ is represented by basis ratio η_t in (4.5). Figure 4 shows the observed ratio over the January 1st, 2007 and July 23th, 2012 period.

FIGURE 4.4 – Basis ratio time series



Observed ratio of the load weighted mean spot price to the arithmetic mean spot price as defined by (4.5). Observations between January 1st, 2007 and July 23th, 2012.

As η_t is larger than one in all but one instance, S_t^* overestimates S_t. Such a departure has not yet been considered in the literature. This departure is due to the fact that more electricity is consumed during peak hours when its price is higher.

4.7 Appendix : delta-hedging with futures

If transaction costs are disregarded, the terminal value of the self-financing hedging portfolio with an initial value of 0 is given by $V_T = \sum_{j=t_0+1}^{T} \theta_j (B_T/B_j)(F_{j,T} - F_{j-1,T})$. Setting $\theta_j = B_j/B_T$, the terminal value of the portfolio becomes

$$V_T = \sum_{j=t_0+1}^{T} (F_{j,T} - F_{j-1,T}) = F_{T,T} - F_{T,t_0} = S_T - F_{T,t_0}.$$

Therefore, holding one unit of this portfolio for each unit of load sold (in the case where the load to serve is known with certainty) permits to lock in the price of electricity to F_{T,t_0}.

4.8 Appendix : Solving the global hedging problem

The optimal trading strategy $(\theta_{T-2}^*, \theta_{T-1}^*, \theta_T^*)$ solving problem (4.8) with the semi-quadratic penalty (4.9) is obtained through dynamic programming (Bertsekas, 1995) :

$$\psi_{t,T} = \min_{\theta_{t+1}} \mathbb{E}\left[\psi_{t+1,T} | \mathcal{G}_t\right] \quad \text{with the terminal condition } \psi_{T,T} = G(\Psi_T - V_T), \quad (4.23)$$

$$\theta_{t+1}^* = \arg\min_{\theta_{t+1}} \mathbb{E}\left[\psi_{t+1,T} | \mathcal{G}_t\right]. \quad (4.24)$$

This optimization problem is tackled using backward induction over time. The traditional approach used for solving (4.23) is based on a lattice which includes all state variables of the problem ; these include the current value of the load-basis and futures prices, current futures return volatilities, the current hedging portfolio value, lagged futures returns and the past portfolio composition. Such an approach is not viable due to its large dimension. Our approach is a stochastic tree which is feasible because the hedging portfolio is only rebalanced three times. The optimization of the trading position θ_t is performed numerically by discretizing its possible values.

4.8.1 Simulation of the stochastic tree

Since the terminal condition $\psi_{T,T} = G(\Psi_T - V_T) = G(\mathcal{L}_T\,(F_{T,T} - F_{T-3,T}) - V_T)$ depends on the state variables (the load-basis $\mathcal{L}$ and the futures contracts related variables) and some endogenous variables (the portfolio value V_T and consequently the corresponding portfolio positions θ_{T-1}, θ_{T-2}, and θ_{T-3}), the random tree must account for all these dimensions.

At time $T-3$, M_{T-3} scenarios for the state variables are simulated from Equations (4.10)-(4.12) and (4.15)-(4.16).[30] These scenarios are combined with all the possible portfolio positions[31] $\theta_{T-2} \in \Theta_{T-2}$ to generate $N_{T-3} = M_{T-3}\text{Card}\{\Theta_{T-2}\}$ simulated values for endogenous variables (V_{T-2}, θ_{T-2}).

At time $T-2$, these N_{T-3} scenarios for the state and endogenous variables are subdivided into $N_{T-2} = M_{T-2}\text{Card}\{\Theta_{T-1}\}$ branches corresponding to all combinations of simulated state variables and possible portfolio positions. A similar iteration occurs at time $T-1$, leading to $N_{T-3} \times N_{T-2} \times N_{T-1}$ terminal nodes.

30. Simulating a scenario at time t involves simulating the values of the load-basis and futures price innovations, respectively $\epsilon_{t+1}^{(\mathcal{L})}$ and $\epsilon_{t+1,t+j}$, $j = 1, 2, 3$.

31. A discretize subset Θ_{T-2} of the possible positions is considered. $\text{Card}\{\Theta_{T-2}\}$ represents the number of elements it contains.

4.8.2 Backward induction

The algorithm solving (4.23) starts by computing the final hedging penalty at each terminal node [32] of the tree :

$$\begin{aligned}
&\hat{\psi}_T \binom{m_{T-3},m_{T-2},m_{T-1}}{\theta_{T-2},\theta_{T-1},\theta_T} \\
&= G\left(\mathcal{L}_T \binom{m_{T-3},m_{T-2},m_{T-1}}{\theta_{T-2},\theta_{T-1},\theta_T} (F_{T,T}(m_{T-3},m_{T-2},m_{T-1}) - F_{T-3,T}) - V_T \binom{m_{T-3},m_{T-2},m_{T-1}}{\theta_{T-2},\theta_{T-1},\theta_T}\right).
\end{aligned}$$

Equations (4.23)-(4.24) are then approximated using the following backward recursion for each node of the tree :

$$\begin{aligned}
\hat{\theta}^*_T \binom{m_{T-3},m_{T-2}}{\theta_{T-2},\theta_{T-1}} &= \arg\min_{\theta \in \Theta_T} \frac{1}{M_{T-1}} \sum_{m=1}^{M_{T-1}} \hat{\psi}_T \binom{m_{T-3},m_{T-2},m}{\theta_{T-2},\theta_{T-1},\theta}, \\
\hat{\psi}_{T-1} \binom{m_{T-3},m_{T-2}}{\theta_{T-2},\theta_{T-1}} &= \min_{\theta \in \Theta_T} \frac{1}{M_{T-1}} \sum_{m=1}^{M_{T-1}} \hat{\psi}_T \binom{m_{T-3},m_{T-2},m}{\theta_{T-2},\theta_{T-1},\theta},
\end{aligned}$$

$$\begin{aligned}
\hat{\theta}^*_{T-1} \binom{m_{T-3}}{\theta_{T-2}} &= \arg\min_{\theta \in \Theta_{T-1}} \frac{1}{M_{T-2}} \sum_{m=1}^{M_{T-2}} \hat{\psi}_{T-1} \binom{m_{T-3},m}{\theta_{T-2},\theta}, \\
\hat{\psi}_{T-2} \binom{m_{T-3}}{\theta_{T-2}} &= \min_{\theta \in \Theta_{T-1}} \frac{1}{M_{T-2}} \sum_{m=1}^{M_{T-2}} \hat{\psi}_{T-1} \binom{m_{T-3},m}{\theta_{T-2},\theta}, \\
\hat{\theta}^*_{T-2} &= \arg\min_{\theta \in \Theta_{T-2}} \frac{1}{M_{T-3}} \sum_{m=1}^{M_{T-3}} \hat{\psi}_{T-2} \binom{m}{\theta}, \\
\hat{\psi}_{T-3} &= \min_{\theta \in \Theta_{T-2}} \frac{1}{M_{T-3}} \sum_{m=1}^{M_{T-3}} \hat{\psi}_{T-2} \binom{m}{\theta}.
\end{aligned}$$

In the experiments of Section 4.4, the number of scenarios are $M_{T-3} = M_{T-2} = 1000$ and $M_{T-1} = 100$. Fewer scenarios are required at the final step since the conditional expectations can partially be solved analytically. More precisely, Equations (4.23)-(4.24) involve double integrals (one over the load innovation and the other over the futures return innovation with a one-week maturity). Fortunately, the load innovation is Gaussian, so the first integral can be computed analytically. Therefore, instead of using a regular Monte-Carlo simulation for the futures innovation, a quadrature in a single dimension is applied.

32. The terminal nodes are identified with the set of indices corresponding to the branches constituting the path :

$$\binom{m_{T-3},m_{T-2},m_{T-1}}{\theta_{T-2},\theta_{T-1},\theta_T}.$$

The discrete sets of portfolio positions are $\Theta_{T-2} = \{0.96, 0.965, \ldots, 1.04\}$ and $\Theta_{T-1} = \Theta_T = \{0.93, 0.94 \ldots, 1.07\}$, implying that $\text{Card}\{\Theta_{T-2}\} = 17$ and $\text{Card}\{\Theta_{T-1}\} = \text{Card}\{\Theta_T\} = 15$.

Variance reduction techniques improve the precision of the Monte Carlo estimates and compensate for small sample sizes. Antithetic variables are used in the simulation for load-basis innovations $\epsilon^{(\mathcal{L})}$. The first half of scenarios are simulated by regular Monte-Carlo methods. In the last half of scenarios, the futures return innovations are identical to the ones in the first half. Load innovations are however set equal to their antithetic counterparts.

4.8.3 Re-simulation

The previous algorithm determines the optimal hedging strategy $\left(\theta^*_{T-2}, \theta^*_{T-1}, \theta^*_T\right)$ as seen from time $T-3$. At time $T-2$, the retailer holds θ^*_{T-2} long futures positions and has to select θ^*_{T-1} to perform the rebalancing. The realization of the state variables at time $T-2$ will not exactly fall on one particular node of the random tree. The standard approach used to solve this issue is to interpolate between the nodes of the tree to determine the optimal hedging position θ^*_{T-1}. We opted for a re-simulation to obtain simulated data which incorporates the newly observed realization of state variables. More precisely, a two-period random tree is simulated from time $T-2$ up to time T to update the optimal hedging strategy $\left(\theta^*_{T-1|T-2}, \theta^*_{T|T-2}\right)$. Since this tree is smaller than the previous one, we opted for a thinner discretization of the portfolio positions : $\Theta_{T-1} = \{0.93, 0.9325 \ldots, 1.07\}$, and $\Theta_T = \{0.93, 0.94 \ldots, 1.07\}$ while keeping $M_{T-2} = 1000$ and $M_{T-1} = 100$.

Finally, at time $T-1$, a one-period random tree with $\Theta_T = \{0.900, 0.901 \ldots, 1.100\}$ is simulated to update the final hedging position $\theta^*_{T|T-1}$.

4.9 Appendix : load-basis model estimation

4.9.1 Cross-validation procedure for load model selection

To determine the number P of Fourier terms in step 1 of the load-basis model estimation (or Q in step 2), a cross-validation procedure is implemented. The load-basis data are from 2007 to 2012. Data from year y are removed and retained as out-of-sample, while remaining data are in-sample. For each value of P (or Q), the model is estimated in-sample. Denote

$\mathcal{J}^y_{1,P} = (\gamma, \beta_0, ...\beta_{2P+1})$ and $\mathcal{J}^y_{2,Q} = (\alpha_0, ...\alpha_{2Q+1})$. f denotes the pdf function.

$$\begin{aligned}
\hat{\mathcal{J}}^y_{1,P} &= \underset{\mathcal{J}^y_{1,P}}{\text{argmax}} \sum_{t, year(t) \neq y} \log f_{\mathcal{L}_t|\mathcal{L}_{t-1}}(\mathcal{L}_t|\mathcal{L}_{t-1}) \quad \text{(under assumption that } v(t) \text{ is constant)} \\
\hat{\mathcal{J}}^y_{2,Q} &= \underset{\mathcal{J}^y_{2,Q}}{\text{argmax}} \sum_{t, year(t) \neq y} \log f_{v_t(\mathcal{J}^y_{2,Q})\epsilon_t}(\sqrt{\tilde{v}(t)}\tilde{\epsilon}^{(\mathcal{L})}_t)
\end{aligned}$$

where $\tilde{g}(t)$ and $\tilde{v}(t)$ are obtained by respectively plugging $\hat{\mathcal{J}}^y_{1,P}$ in (4.11) and $\mathcal{J}^y_{2,Q}$ in (4.12). The $\tilde{\epsilon}^{(\mathcal{L})}_t$ are calculated by replacing $g(t)$ and $v(t)$ by $\tilde{g}(t)$ and $\tilde{v}(t)$ in (4.10).

Then, a test statistic assessing the goodness-of-fit (MSE for P, log-likelihood for Q) is calculated out-of-sample :

$$\text{MSE}^P_y = \frac{1}{n_y} \sum_{t, year(t) = y} (\mathcal{L}_t - \text{Pred}(\mathcal{L}_t, \hat{\mathcal{J}}^y_{1,P}))^2 \tag{4.25}$$

$$\text{log-l}^Q_y = \sum_{t, year(t) = y} \log f_{\sqrt{v_t(\hat{\mathcal{J}}^y_{2,Q})}\epsilon^{(\mathcal{L})}_t}(\sqrt{\hat{v}(t)}\hat{\epsilon}^{(\mathcal{L})}_t) \tag{4.26}$$

where n_y is the number of observations in year y. $\hat{g}(t)$ and $\hat{v}(t)$ are obtained by respectively plugging $\hat{\mathcal{J}}^y_{1,P}$ in (4.11) and $\hat{\mathcal{J}}^y_{2,Q}$ in (4.12). The $\hat{\epsilon}^{(\mathcal{L})}_t$ are calculated by replacing $g(t)$ and $v(t)$ by $\hat{g}(t)$ and $\hat{v}(t)$ in (4.10). The predicted load-basis is $\text{Pred}(\mathcal{L}_t, \hat{\mathcal{J}}^y_{1,P}) = \hat{g}(t) + \hat{\gamma}(\mathcal{L}_{t-1} - \hat{g}(t-1))$ where $\hat{g}$ is calculated by plugging $\hat{\mathcal{J}}^y_{1,P}$ in (4.11) and $\hat{\gamma}$ is the first component of $\hat{\mathcal{J}}^y_{1,P}$. The prediction is obtained by applying a conditional expectation on (4.10). This operation is repeated for all years y and the test statistic is aggregated across all years :

$$\text{RMSE}^P_{total} = \sqrt{\frac{\sum_{y=2007}^{2012} n_y \text{MSE}^P_y}{\sum_{2007}^{2012} n_y}} \quad \text{or} \quad \text{log-l}^Q_{total} = \sum_{y=2007}^{2012} \text{log-l}^Q_y.$$

Parameters $\hat{P}$ and $\hat{Q}$ are selected to optimize the corresponding test statistic

$$\hat{P} = \underset{P}{\text{argmin}} \ \text{RMSE}^P_{total} \quad \text{and} \quad \hat{Q} = \underset{Q}{\text{argmax}} \ \text{log-l}^Q_{total}.$$

Results are shown in Tables 9 and 10 and suggest $\hat{P} = 3$ and $\hat{Q} = 2$.

4.9.2 Goodness-of-fit for the load model

In this section, the properties of the standardized residuals $\hat{\epsilon}^{(\mathcal{L})}_t$ are analyzed to determine the adequacy of the load-basis model (4.10)-(4.12). Figure 5 shows a boxplot of residuals by quarter of the year, a QQ-plot and a kernel density plot. Residuals look reasonably uniform across quarters ; there is thus no obvious evidence that the seasonal trend is not properly

Table 4.9 – Cross-validation test results for the load-basis seasonality trend

Value for P	1	2	3	4	5
Cross-validation RMSE ($\times 10^5$)	2.386	2.376	**2.360**	2.364	2.367

Out-of-sample cross-validation prediction root-mean-square-error for the load-basis model with different numbers of Fourier terms P in the load-basis seasonality trend g defined by (4.11).

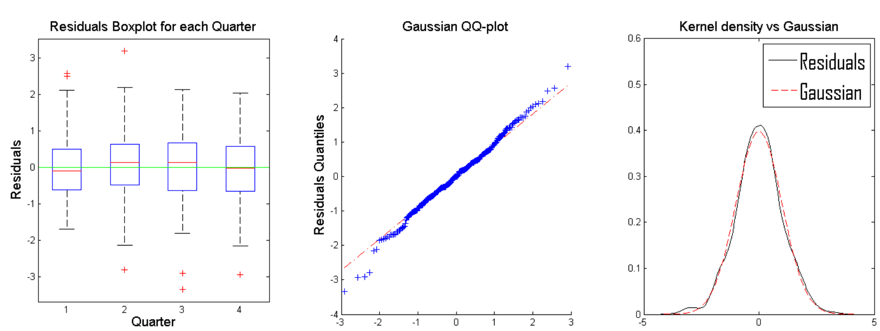
Table 4.10 – Cross-validation test results for the load-basis variance trend

Value for Q	1	2	3	4	5
Cross-validation log-likelihood ($\times 10^{-3}$)	-3.973	$-\mathbf{3.967}$	-3.973	-3.974	-3.976

Out-of-sample cross-validation log-likelihood for the load-basis model with different numbers of Fourier terms Q in the load-basis variance trend v defined by (4.12).

being captured. The Gaussian distribution seems to be a suitable candidate for residuals, even if the empirical left tail of the load residuals is slightly heavier. A bootstrap Cramer-Von-Mises goodness-of-fit test (Genest & Rémillard, 2008) for the adequacy of the Gaussian distribution is applied to the residuals and the p-value is 27%, not rejecting the Gaussian distribution. A Ljung-Box test for autocorrelation of residuals has a p-value of 92% and does not reject $\hat{\epsilon}_t^{(L)}$ as white noise. The presence of a GARCH effect in the residuals is tested through the McLeod-Li test (p-value of 18%) and Lagrange Multiplier test (p-value of 16%) ; there is no significant presence of a GARCH effect. Therefore, the $\epsilon^{(\mathcal{L})}$ load-basis innovations are modeled by a strong Gaussian white noise.

Figure 4.5 – Load-basis model residuals



Boxplot, Gaussian QQ-plot and kernel plot for load-basis residuals $\hat{\epsilon}^{(\mathcal{L})}$ between January 1st, 2007 and July 29th, 2012.

4.9.3 Goodness-of-fit of futures return model

Ljung-Box and McLeod-Li tests for strong white noise are carried out on the scaled residuals $\widehat{z}_{j,t}$, $j = 0, 1, 2$. P-values are obtained through simulation (usual p-value formulas incorrectly assume Gaussianity). P-values are given in Table 11 and none of the tests reject the white noise hypothesis.

Table 4.11 – Autocorrelation tests for futures return innovations

Series	$z_{0,t}$	$z_{1,t}$	$z_{2,t}$
Ljung-Box p-value	0.36	0.35	0.41
McLeod-Li p-value	0.97	0.33	0.72

Bootstrapped $p-$values for the Ljung-Box and McLeod-Li tests applied on futures return innovations. Observations between January 1st, 2007 and July 29th, 2012 for futures with 1, 2 and 3 weeks to maturity.

The choice of the NIG distribution for the innovations must be validated. Figure 6 compares the kernel density of the $\hat{z}_{i,t}$, its fitted NIG distribution and a corresponding Gaussian distribution. The NIG distribution represents more adequately the shape of the empirical residuals distribution than the Gaussian distribution, the latter is unable to capture the peakedness of the empirical futures returns distribution. Cramer-Von-Mises tests with simulated p-value (Genest & Rémillard, 2008) are applied to assess the adequacy of the fit of the NIG distribution for the $z_{i,t}$ innovations. P-values are found in Table 12 for each univariate $z_{i,t}$, $i = 0, 1, 2$ series. The p-value for the joint trivariate series is 0.82. The NIG distribution thus provides an acceptable fit.

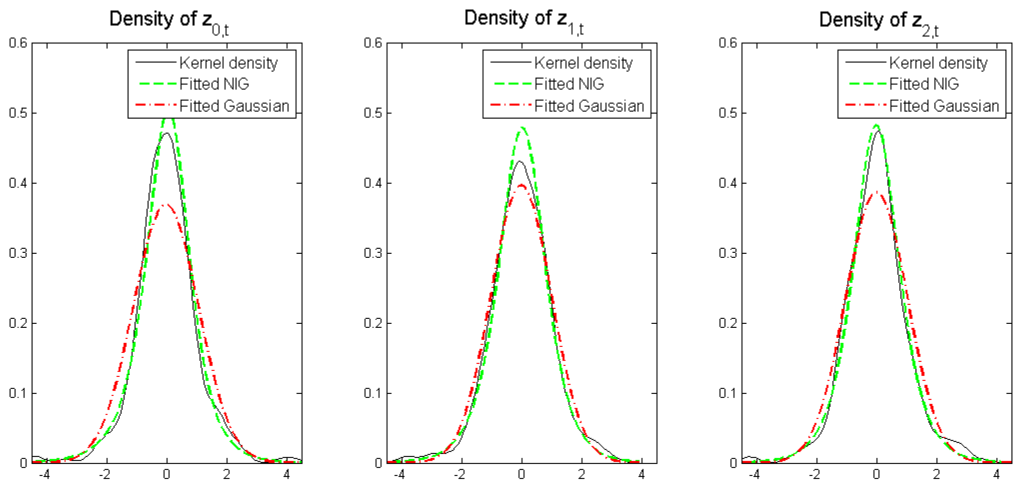
Table 4.12 – Goodness-of-fit of the futures return distribution

Series	$z_{0,t}$	$z_{1,t}$	$z_{2,t}$
p-value	0.09	0.88	0.65

Bootstrapped $p-$values for the Cramer-Von-Mises goodness-of-fit test on the NIG distribution for futures return. Observations between January 1st, 2007 and July 29th, 2012 for futures with 1, 2 and 3 weeks to maturity.

To validate the choice of the copula, Cramer-Von-Mises goodness-of-fit tests are applied for the Gaussian copula on the three following pairs of processes : $(z_{0,t}, z_{1,t})$, $(z_{0,t}, z_{2,t})$ and

FIGURE 4.6 – Futures return distribution



Kernel density plots of futures return innovations and fitted NIG and Gaussian distributions. Observations between January 1st, 2007 and July 29th, 2012 for futures with 1, 2 and 3 weeks to maturity.

$(z_{1,t}, z_{2,t})$.[33] The p-values for the three tests are given in Table 13. Since p-values are all high, the Gaussian copula provides an acceptable fit.

TABLE 4.13 – Goodness-of-fit of the futures return copula

Innovation Pair	$(z_{0,t}, z_{1,t})$	$(z_{0,t}, z_{2,t})$	$(z_{1,t}, z_{2,t})$
p-value	0.90	0.67	0.97

Bootstrapped $p-$values for the Cramer-Von-Mises goodness-of-fit test applied to the Gaussian copula linking futures returns. Tests are applied on pairs of returns instead of the triplet $(z_{0,t}, z_{1,t}, z_{2,t})$. Observations between January 1st, 2007 and July 29th, 2012 for futures with 1, 2 and 3 weeks to maturity.

4.9.4 Independence of futures return and load-basis innovations

Independence tests for load-basis residuals $\hat{\epsilon}_t^{(\mathcal{L})}$ and futures return innovation proxies $\hat{z}_{t,i}$ are applied for each of the three futures return maturities : $i = 0, 1, 2$. We use a Cramer-Von-Mises goodness-of-fit test on the independence copula (Rémillard, 2013). The p-values are

33. The test was not carried on the triplet $(z_{0,t}, z_{1,t}, z_{2,t})$. The numerical burden associated with such a test is very high.

obtained through simulation and are given in Table 14. Large p-values allow us to assume that the load-basis residuals and the futures return innovations are independent.

TABLE 4.14 – Independence test for load-basis and futures return innovations

Futures Returns Series	$i=0$	$i=1$	$i=2$
p-value	0.28	0.66	0.50

Bootstrapped $p-$values for the Cramer-Von-Mises goodness-of-fit test applied to the independence copula linking the load-basis observations and futures returns. Three tests are applied separately for the three maturities of futures returns. Observations between January 1st, 2007 and July 29th, 2012 for futures with 1, 2 and 3 weeks to maturity.

4.10 Technical report (not part of the paper)

4.10.1 Trading Volume by Year

TABLE 4.15 – Trading days with non-nul trading volume each year on Nord Pool weekly futures.

Year/Weeks to maturity	1	2	3	4	5
2007	97%	88%	71%	34%	18%
2008	96%	90%	63%	26%	13%
2009	94%	93%	46%	17%	10%
2010	95%	93%	69%	37%	24%
2011	96%	86%	58%	28%	13%
2012	97%	91%	60%	30%	19%

Percentage of trading days between January 1st, 2007 and December 31th, 2012 on which a non-null trading volume on weekly futures with different maturities occured on NASDAQ OMX.

4.10.2 Out-of-sample tests for the load-basis model

Table 4.16 illustrates the value of $\mathrm{RMSE}_y^P := \sqrt{\mathrm{MSE}_y^P}$ and Table 4.17 illustrates the value of $\sqrt{\text{log-l}_y^Q}$ defined in (4.25) and (4.26) for out-of-sample years $y = 2007, \ldots, 2012$ and different values of P, Q.

Table 4.16 – Cross-validation RMSE ($\times 10^5$) for parameter P

Value for P/Year	2007	2008	2009	2010	2011	2012
1	2.488	**1.930**	2.101	3.000	2.033	2.734
2	2.442	2.150	2.057	2.935	2.013	2.631
3	2.428	2.213	**2.018**	2.935	1.959	**2.548**
4	2.412	2.196	2.046	**2.927**	1.949	2.621
5	**2.402**	2.177	2.116	2.941	**1.941**	2.582

Cross-validation root-mean-square-error for all out-of-sample years for the load-basis model with different numbers of Fourier terms P in the load seasonality trend g defined by (4.11). The smallest RMSE for each out-of-sample year is in boldface.

Table 4.17 – Cross-validation Log-likelihood for parameter Q

Value for Q/Year	2007	2008	2009	2010	2011	2012
1	−708.90	−704.22	−704.24	−740.71	−702.12	−413.06
2	−707.64	**−704.08**	**−701.05**	**−739.76**	−702.11	**−412.75**
3	−707.20	−705.88	−702.36	−742.54	**−702.11**	−413.29
4	**−707.18**	−705.67	−702.11	−741.77	−704.30	−413.36
5	−707.90	−706.93	−701.92	−740.76	−703.93	−414.19

Cross-validation log-likelihood for all out-of-sample years for the load-basis model with different numbers of Fourier terms Q in the load volatility trend v defined by (4.12). The largest log-likelihood for each out-of-sample year is in boldface.

4.10.3 Incomplete information load-basis forecast models

To forecast weekly load from its partial observation as described in Section 4.3.1, the following models were compared :

$$\hat{\mathcal{L}}_t := \tilde{L}_t\left(\sum_{j=1}^{q}\frac{c^j}{\sum_{i=1}^{q}c^i}\frac{\mathcal{L}_{t-j}}{\tilde{L}_{t-j}}\right), \quad (\mathbf{ES}) \tag{4.27}$$

$$\hat{\mathcal{L}}_t := \tilde{L}_t\left(\ell + a\sum_{j=1}^{q}c^j\frac{\mathcal{L}_{t-j}}{\tilde{L}_{t-j}}\right), \quad (\mathbf{CES}) \tag{4.28}$$

$$\hat{\mathcal{L}}_t := \tilde{L}_t\left(\ell + \sum_{j=1}^{q}c_j\frac{\mathcal{L}_{t-j}}{\tilde{L}_{t-j}}\right), \quad (\mathbf{RR}) \tag{4.29}$$

$$\hat{\mathcal{L}}_t := \tilde{L}_t\left(\ell + c\sum_{j=1}^{q}\frac{\mathcal{L}_{t-j}}{\tilde{L}_{t-j}}\right), \quad (\mathbf{CMA}) \tag{4.30}$$

$$\hat{\mathcal{L}}_t := \ell\tilde{L}_t, \quad (\mathbf{CR}) \tag{4.31}$$

$$\hat{\mathcal{L}}_t := g(t) + \hat{\gamma}(\mathcal{L}_{t-1} - g(t-1)). \quad (\mathbf{PL}) \tag{4.32}$$

Their forecasting ability were compared in a cross-validation experiment similar to the one described in Section 4.9.1. The out-of-sample forecasting RMSE and MAPE are given in Table 4.18.

TABLE 4.18 – Cross-validation load-basis forecast from partial observation of load

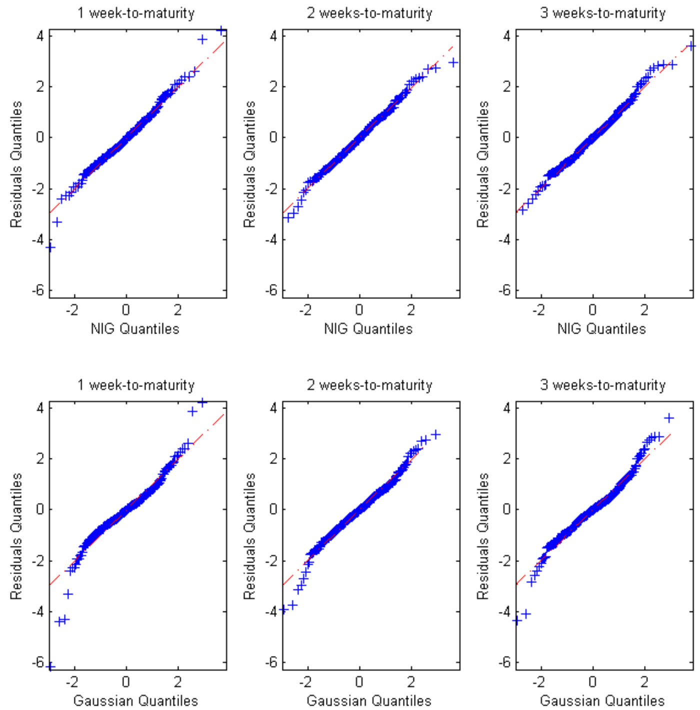
Model	RMSE$\times 10^{-5}$	MAPE
ES	9.29	1.22%
CES	8.84	1.12%
RR	8.90	1.13%
CMA	**8.82**	**1.12**%
CR	9.48	1.21%
PL	22.59	2.86%

Cross-validation load-basis forecast statistics. Load-basis forecasts are made from the partial observation of the load in the same week. Models (4.27)-(4.32) are compared. The best results are in boldface.

4.10.4 More on futures returns diagnostics

Figure 4.7 presents QQ-plots comparing the empirical futures return innovation quantiles and those of the fitted Gaussian and NIG distribution.

FIGURE 4.7 – QQ-plots for futures returns : empirical distribution vs Gaussian and NIG



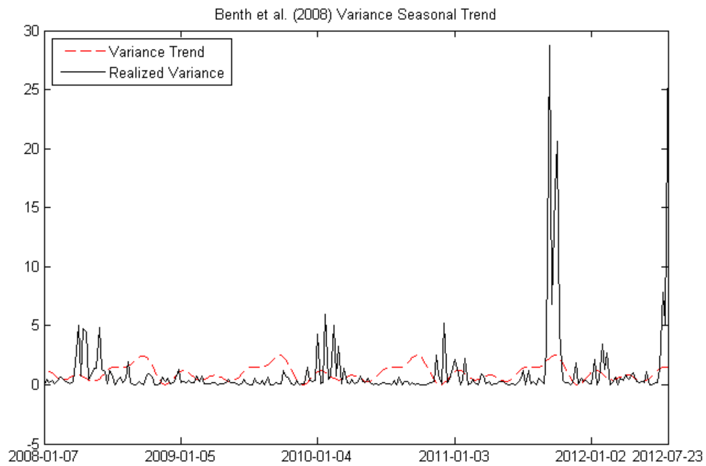
QQ-plots comparing the empirical futures return innovation quantiles and those of the fitted Gaussian and NIG distribution. Observations between January 1st, 2007 and July 23, 2012 for futures with 1, 2 and 3 weeks to maturity.

4.10.5 Seasonal volatility for futures

Benth et al. (2008) propose to use Fourier terms to model the variance seasonality of futures returns.[34] Such a regression is applied to the current futures dataset to see if some seasonality pattern can be extracted. Figure 4.8 presents the variance term and realized variance of the scaled futures returns as described in the step 1 of the estimation procedure in Benth et al. (2008). This approach does not seem to work in the present case. The volatility clusters seem to occur at different times through the year without any regular pattern. This is an advantage of the GARCH feature of the model of the current paper ; volatility clusters occur unexpectedly at random times.

34. Section 8.6

FIGURE 4.8 – QQ-plots for futures returns : empirical distribution vs Gaussian and NIG



The variance trend and realized variance of futures returns obtained through the market model of Benth et al. (2008) applied to the dataset of futures returns used in the rest of this paper.

4.10.6 The volatility bound on futures returns

In order to assess the degree to which the bound on volatility ς in (4.16) is binding, 1000 paths of futures returns, each one lasting 200 weeks, were simulated with the estimated parameters found in Tables 4.5 and 4.6. The percentage of volatility observations exceeding $\varsigma = 0.6$ is given in table 4.19. Those percentages are low enough to consider that the bound $\varsigma = 0.6$ has only minor implications.

TABLE 4.19 – Futures volatility bounds hit rates

Time-to-maturity	1	2	3
Hit rate	0.073%	0.45%	0.27%

Percentage of futures returns volatility observations hitting the bound $\varsigma = 0.6$ in the simulation experiment. 1000 paths of 200 periods each were simulated.

4.10.7 Information on NIG distribution

This appendix presents relevant information about the Normal Inverse Gaussian (NIG) distribution. A random variable has a NIG$(\alpha, \beta, \delta, \mu)$ distribution if its density function (pdf) is given by

$$f(x) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2 - (x-\mu)^2}\right)}{\pi\sqrt{\delta^2 - (x-\mu)^2}} e^{\gamma\delta + \beta(x-\mu)}$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$, $\alpha > 0, \beta \in (-\alpha, \alpha), \delta > 0, \mu \in \mathbb{R}$ and K_ν is the modified Bessel function of the second kind :

$$K_\nu(y) = \frac{1}{2}\int_0^\infty u^{\nu-1} e^{-\frac{1}{2}y(u+u^{-1})} du, \quad y > 0.$$

If X has a NIG$(\alpha, \beta, \delta, \mu)$ distribution, then $aX + b$ has the NIG$(\alpha/a, \beta/a, a\delta, a\mu + b)$ distribution. Furthermore, if X_1 and X_2 are two independent variables respectively from distributions NIG$(\alpha, \beta, \delta_1, \mu_1)$ and NIG$(\alpha, \beta, \delta_2, \mu_2)$, the sum $X_1 + X_2$ has a NIG$(\alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2)$ distribution.

The moment generating function of a NIG$(\alpha, \beta, \delta, \mu)$ variable X is

$$\phi(z) = \mathbb{E}\left[e^{zX}\right] = e^{\mu z + \delta\left(\gamma - \sqrt{\alpha^2 - (\beta+z)^2}\right)}.$$

which exists at z if $\alpha^2 - (\beta + z)^2 > 0$.

The first four moments are characterized by

$$\begin{aligned}
&\mathbb{E}[X] = \mu + \frac{\delta\beta}{\gamma} \\
&\mathrm{Var}[X] = \frac{\delta\alpha^2}{\gamma^3} \\
&\mathrm{Skewness}[X] = \frac{3\beta}{\alpha\sqrt{\delta\gamma}} \\
&\text{Excess Kurtosis}[X] = \frac{3\left(1 + 4\frac{\beta^2}{\alpha^2}\right)}{\delta\gamma}.
\end{aligned}$$

The NIG distribution is flexible since it can exhibit asymmetry and excess kurtosis to various degrees depending on parameters.

<u>Standardized NIG</u>

A standardized NIG distribution is a NIG distribution with mean 0 and variance 1. Therefore, $\mu = -\frac{\delta\beta}{\gamma}$ and $\delta = \frac{\gamma^3}{\alpha^2}$. The only two free parameters left are (α, β).

Definition 4.10.1 *A standardized NIG(α, β) distribution is a NIG$(\alpha, \beta, \frac{\gamma^3}{\alpha^2}, -\frac{\gamma^2\beta}{\alpha^2})$ distribution with $\gamma = \sqrt{\alpha^2 - \beta^2}$.*

4.10.8 Remark on futures price integrability

For $p > 0$, all expectations $\mathbb{E}\left[e^{p\epsilon_{t,t+i}}\right]$, $t = 0, 1, \ldots$ exist if and only if $\mathbb{E}\left[e^{p\varsigma z_{i,t}}\right]$ exists. This last expectation exists if and only if $\alpha^2 - (\beta + p)^2 \geq 0$.

Combining $\alpha^2 - (\beta + p)^2 \geq 0$ with the domain constraint $|\beta| < \alpha$, we obtain the following constraints on parameters :

$$\begin{aligned} \alpha &> \frac{p\varsigma}{2} \\ \beta &\in (-\alpha, \alpha - p\varsigma) \end{aligned}$$

In the current paper, values used are $p = 2$.

4.10.9 The quadratic penalty case

The solution to problem (4.8) when $G(x) = x^2$ is obtained by Schweizer (1995). This methodology disregards transactions costs : it assumes $\forall t, \mathcal{C}_t = 0$. Define the following quantities : starting with $\nu_{T+1} := 1$ and $C_T := \Psi_T$,

$$\Delta_t := B_t^{-1}(F_{t,T} - F_{t-1,T}) \tag{4.33}$$

$$A_t := \mathbb{E}\left[\Delta_t^2 \nu_{t+1} | \mathcal{G}_{t-1}\right] \tag{4.34}$$

$$b_t := A_t^{-1}\mathbb{E}\left[\Delta_t \nu_{t+1} | \mathcal{G}_{t-1}\right] \tag{4.35}$$

$$\alpha_t := A_t^{-1}\mathbb{E}\left[B_t^{-1} C_t \Delta_t \nu_{t+1} | \mathcal{G}_{t-1}\right] \tag{4.36}$$

$$\nu_t := \mathbb{E}[(1 - b_t\Delta_t)\nu_{t+1} | \mathcal{G}_{t-1}] \tag{4.37}$$

$$C_{t-1} := \frac{\mathbb{E}\left[B_t^{-1}(1 - b_t\Delta_t) C_t \nu_{t+1} | \mathcal{G}_{t-1}\right]}{B_{t-1}^{-1}\nu_t}. \tag{4.38}$$

Theorem 4.10.1 *If $G(x) = x^2$, the solution to problem (4.8) is given by :*

$$\theta_t = \alpha_t - B_{t-1}^{-1} V_{t-1} b_t. \tag{4.39}$$

Proof of Theorem 4.10.1

The problem solved in Schweizer (1995) is the following. Let X be a $\mathcal{G}-$adapted stochastic process. Define the following discrete stochastic integral : $M_t := M_{t_0} + \sum_{j=t_0+1}^{t} \theta_j (X_j - X_{j-1})$. Let Y be a $\mathcal{G}_N-$measurable random variable. It is shown that the trading strategy

$(\theta_{t_0+1}, \ldots, \theta_T)$ that minimizes $\mathbb{E}\left[(Y - M_T)^2 | \mathcal{G}_{t_0}\right]$ is given by $\theta_t = \alpha_t - M_{t-1} b_t$ where

$$\Delta_t := X_t - X_{t-1} \tag{4.40}$$

$$A_t := \mathbb{E}\left[\Delta_t \Delta_t^\top P_{t+1} | \mathcal{G}_{t-1}\right] \tag{4.41}$$

$$b_t := A_t^{-1} \mathbb{E}\left[\Delta_t P_{t+1} | \mathcal{G}_{t-1}\right] \tag{4.42}$$

$$\alpha_t := A_t^{-1} \mathbb{E}\left[Y \Delta_t P_{t+1} | \mathcal{G}_{t-1}\right] \tag{4.43}$$

$$P_t := \prod_{j=t}^{T} (1 - b_j^\top \Delta_j) \tag{4.44}$$

In the problem of the current paper, (4.6) directly leads to

$$B_{t+1}^{-1} V_{t+1} = B_t^{-1} V_t + \theta_{t+1} B_{t+1}^{-1} (F_{t+1,T} - F_{t,T}). \tag{4.45}$$

Consequently,

$$B_t^{-1} V_t = B_{t_0}^{-1} V_{t_0} + \sum_{j=t_0+1}^{t} \theta_j B_j^{-1} (F_{j,T} - F_{j-1,T}). \tag{4.46}$$

Since B_T^{-1} is deterministic, minimizing $\mathbb{E}\left[(\Psi_T - V_T)^2 | \mathcal{G}_{t_0}\right]$ is equivalent to minimizing $\mathbb{E}\left[(B_T^{-1}\Psi_T - B_T^{-1} V_T)^2 | \mathcal{G}_{t_0}\right]$. By posing $X_t := \sum_{j=t_0+1}^{t} B_j^{-1}(F_{j,T} - F_{j-1,T})$ and $M_{t_0} := B_{t_0}^{-1} V_{t_0}$, one gets that $M_t = B_t^{-1} V_t$ and $\Delta_t = B_t^{-1}(F_{t,T} - F_{t-1,T})$. The solution to the latter problem is therefore $\theta_t = \alpha_t - B_t^{-1} V_t b_t$ where $Y := B_T^{-1} \Psi_T$ in (4.41)-(4.44). For a demonstration of the equivalence between (4.41)-(4.44) and (4.34)-(4.38), see Rémillard & Rubenthaler (2013).

Remark 4.10.1 *The definition of Δ_t is a bit different than the definition found in Schweizer (1995). This is due to the fact that futures contracts are used instead of stocks, and therefore the dynamics of the hedging portfolio are slightly different.*

4.10.10 Computing Quadratic Hedging Factors

Theorem 4.10.2

$$\forall t, j \geq 0,\ \mathbb{E}\left[\mathcal{L}_t | \mathcal{G}_{t-j}\right] = g(t) + \gamma^j \left(\mathcal{L}_{t-j} - g(t-j)\right) \tag{4.47}$$

Proof of Theorem 4.10.2 :

The proof uses induction. Let t be an integer. The initial step $j = 0$ is trivial. Assume (4.47) holds for j. Then,

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{L}_t|\mathcal{G}_{t-(j+1)}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathcal{L}_t|\mathcal{G}_{t-j}\right]|\mathcal{G}_{t-(j+1)}\right] \\
=\ & \mathbb{E}\left[g(t) + \gamma^j\left(\mathcal{L}_{t-j} - g(t-j)\right)|\mathcal{G}_{t-j-1}\right] \\
=\ & g(t) + \gamma^j\mathbb{E}\left[\gamma(\mathcal{L}_{t-j-1} - g(t-j-1)) + \sqrt{v(t-j)}\epsilon^{(\mathcal{L})}_{t-j}|\mathcal{G}_{t-j-1}\right] \quad \text{by (4.10)} \\
=\ & g(t) - \gamma^{j+1}(\mathcal{L}_{t-j-1} - g(t-j-1)). \square
\end{aligned}$$

Theorem 4.10.3

$$\forall t \in \{t_0, \ldots, T\},\ C_t = \mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_t\right] B_1^{-(T-t)}(F_{t,T} - F_{t_0,T}) \tag{4.48}$$

Proof of Theorem 4.10.3 :

The proof uses backward induction. The initial step $t = T$ is trivial since $C_T = \Psi_T$. Assume (4.48) holds for t. Define $B^{-1} := B_1^{-1} = e^{-r}$. Then,

$$\begin{aligned}
C_{t-1} &:= \frac{\mathbb{E}\left[B_t^{-1}(1-b_t\Delta_t)C_t\nu_{t+1}|\mathcal{G}_{t-1}\right]}{B_{t-1}^{-1}\nu_t} \quad \text{by (4.38)} \\
&= B^{-1}\frac{\mathbb{E}\left[(1-b_t\Delta_t)\mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_t\right]B_1^{-(T-t)}(F_{t,T} - F_{t_0,T})\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} \quad \text{by induction hypothesis} \\
&= B^{-(T-(t-1))}\mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_{t-1}\right]\frac{\mathbb{E}\left[(1-b_t\Delta_t)(F_{t,T} - F_{t_0,T})\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} \text{ by independence} \qquad (4.49)
\end{aligned}$$

In (4.49),

$$\begin{aligned}
&\frac{\mathbb{E}\left[(1-b_t\Delta_t)(F_{t,T} - F_{t_0,T})\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} = \\
&\frac{\mathbb{E}\left[(1-b_t\Delta_t)F_{t,T}\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} - F_{t_0,T}\frac{\mathbb{E}\left[(1-b_t\Delta_t)\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t}
\end{aligned} \tag{4.50}$$

Notice first, that $\mathbb{E}\left[(1-b_t\Delta_t)\nu_{t+1}|\mathcal{G}_{t-1}\right] = \nu_t$ by (4.37). Furthermore, since $F_{t,T}\Delta_t = \frac{\Delta_t^2}{B_t^{-1}} + F_{t-1,T}\Delta_t$,

$$
\begin{aligned}
& \frac{\mathbb{E}\left[(1-b_t\Delta_t)F_{t,T}\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} \\
= & \frac{\mathbb{E}\left[\left(F_{t,T}-b_t\frac{\Delta_t^2}{B_t^{-1}}-b_tF_{t-1,T}\Delta_t\right)\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} \\
= & \frac{\mathbb{E}\left[F_{t,T}\nu_{t+1}|\mathcal{G}_{t-1}\right]-b_t\frac{\mathbb{E}\left[\Delta_t^2\nu_{t+1}|\mathcal{G}_{t-1}\right]}{B_t^{-1}}-b_tF_{t-1,T}\mathbb{E}\left[\Delta_t\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} \\
= & \frac{\mathbb{E}\left[F_{t,T}\nu_{t+1}|\mathcal{G}_{t-1}\right]-\frac{\mathbb{E}\left[\Delta_t\nu_{t+1}|\mathcal{G}_{t-1}\right]}{B_t^{-1}}-b_tF_{t-1,T}\mathbb{E}\left[\Delta_t\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} \text{ by (4.35)} \\
= & \frac{\mathbb{E}\left[F_{t,T}\nu_{t+1}|\mathcal{G}_{t-1}\right]-\mathbb{E}\left[(F_{t,T}-F_{t-1,T})\nu_{t+1}|\mathcal{G}_{t-1}\right]-b_tF_{t-1,T}\mathbb{E}\left[\Delta_t\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} \text{ by (4.33)} \\
= & \frac{F_{t-1,T}\mathbb{E}\left[(1-b_t\Delta_t)\nu_{t+1}|\mathcal{G}_{t-1}\right]}{\nu_t} \\
= & F_{t-1,T} \quad \text{by (4.37)}
\end{aligned}
\tag{4.51}
$$

The proof is completed by putting (4.51) in (4.50), and then putting the result in (4.49). □

Closed formulas for the last time step

Theorem 4.10.4 *Let $\phi_{\alpha,\beta}(z) := \mathbb{E}\left[e^{zX}\right]$ be the moment generating function of a standardized NIG(α,β) distributed variable X. Therefore,*

$$
\begin{aligned}
b_T &= \frac{1}{B_T^{-1}F_{T-1,T}}\frac{e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})-1}{e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(2\sigma_{0,T})-2e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})+1} \\
\nu_T &= 1-\frac{e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}[\phi_{\alpha_1,\beta_1}(\sigma_{0,T})]^2-2e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})+1}{e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(2\sigma_{0,T})-2e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})+1} \\
\alpha_T &= \mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_{T-1}\right]\frac{e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(2\sigma_{0,T})-e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})}{e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(2\sigma_{0,T})-2e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})+1} \\
&\quad +\mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_{T-1}\right]\frac{F_{t_0,T}}{F_{T-1,T}}\frac{(1-e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T}))}{e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(2\sigma_{0,T})-2e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})+1}.
\end{aligned}
$$

Proof of Theorem 4.10.3 :

For b_T and ν_T is suffice to notice that

$$
\begin{aligned}
\mathbb{E}\left[\Delta_T|\mathcal{G}_{T-1}\right] &= B_T^{-1}F_{T-1,T}\left[e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})-1\right], \\
\mathbb{E}\left[\Delta_T^2|\mathcal{G}_{T-1}\right] &= (B_T^{-1}F_{T-1,T})^2\left[e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(2\sigma_{0,T})-2e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})+1\right].
\end{aligned}
$$

For α_T,

$$\begin{aligned}
\alpha_T &= A_T^{-1}\mathbb{E}\left[B_T^{-1}\Psi_T\Delta_T|\mathcal{G}_{T-1}\right] \\
&= A_T^{-1}B_T^{-2}F_{T-1,T}\mathbb{E}\left[\mathcal{L}_T(S_T-F_{t_0,T})(e^{\epsilon_{T,T}}-1)|\mathcal{G}_{T-1}\right] \\
&= A_T^{-1}B_T^{-2}F_{T-1,T}\mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_{T-1}\right]\mathbb{E}\left[(F_{T,T}-F_{t_0,T})(e^{\epsilon_{T,T}}-1)|\mathcal{G}_{T-1}\right] \\
&= \frac{\mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_{T-1}\right]}{\mathbb{E}\left[\Delta_T^2|\mathcal{G}_{T-1}\right]}B_T^{-2}F_{T-1,T}\left(F_{T-1,T}\mathbb{E}\left[e^{2\epsilon_{T,T}}|\mathcal{G}_{T-1}\right]-(F_{T-1,T}+F_{t_0,T})\mathbb{E}\left[e^{\epsilon_{T,T}}|\mathcal{G}_{T-1}\right]+F_{t_0,T}\right) \\
&= \mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_{T-1}\right]\frac{e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(2\sigma_{0,T})-e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})}{e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(2\sigma_{0,T})-2e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})+1} \\
&\quad +\mathbb{E}\left[\mathcal{L}_T|\mathcal{G}_{T-1}\right]\frac{F_{t_0,T}}{F_{T-1,T}}\frac{(1-e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T}))}{e^{2\mu_1+2a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(2\sigma_{0,T})-2e^{\mu_1+a_1\epsilon_{T-1,T-1}}\phi_{\alpha_1,\beta_1}(\sigma_{0,T})+1}.\square
\end{aligned}$$

4.10.11 Explicit formula for the Bellman Equation at the last time step

In the Bellman Equation (4.23)-(4.24), a part of the expectation can be computed explicitly when $t=T-1$ if the semi-quadratic penalty (4.9) is used. Indeed,

$$\begin{aligned}
\mathbb{E}\left[\psi_{T,T}|\mathcal{G}_{T-1}\right] &= \mathbb{E}\left[g(\Psi_{T,T}-V_T)|\mathcal{G}_{T-1}\right] \\
&= \mathbb{E}\left[g(\mathcal{L}_T(F_{T,T}-F_{t_0,T})-V_T)|\mathcal{G}_{T-1}\right] \\
&= \mathbb{E}\left[(\mathcal{L}_T(F_{T,T}-F_{t_0,T})-V_T)^2\mathbb{I}_{\{\mathcal{L}_T(F_{T,T}-F_{t_0,T})>V_T\}}|\mathcal{G}_{T-1}\right] \\
&= \int_{\bar{F}}\int_{\bar{L}}(\bar{L}(\bar{F}-F_{t_0,T})-V_T)^2\mathbb{I}_{\{\bar{L}(\bar{F}-F_{t_0,T})>V_T\}}f_{\mathcal{L}_T}(\bar{L})f_{F_{T,T}}(\bar{F})d\bar{L}d\bar{F} \\
&\approx \sum_{m=1}^{M_{T-1}}\int_{\bar{L}}(\bar{L}(F_{T,T}^{(m)}-F_{t_0,T})-V_T)^2\mathbb{I}_{\{\bar{L}(F_{T,T}^{(m)}-F_{t_0,T})>V_T\}}f_{\mathcal{L}_T}(\bar{L})d\bar{L} \qquad (4.52)
\end{aligned}$$

where $F_{T,T}^{(m)}$ is the simulated value of $F_{T,T}$ in the m^{th} path of a quadrature and $f_X(x)$ denotes the pdf of the variable X at the point x conditional on $\mathcal{G}_{T-1}$. The integral in (4.52) can be computed exactly. Recall from (4.10) that $\mathcal{L}_T=\mu_T^{\mathcal{L}}+\sigma_T^{\mathcal{L}}\epsilon_T^{(\mathcal{L})}$ where

$$\mu_T^{\mathcal{L}} := g(T)+\gamma(\mathcal{L}_{T-1}-g(T-1)) \qquad (4.53)$$

$$\sigma_T^{\mathcal{L}} := \sqrt{v(T)}. \qquad (4.54)$$

Therefore,

$$\begin{aligned}
\mathcal{L}_T(F_{T,T}^{(m)}-F_{t_0,T})-V_T &= (\mu_T^{\mathcal{L}}+\sigma_T^{\mathcal{L}}\epsilon_T^{(\mathcal{L})})(F_{T,T}^{(m)}-F_{t_0,T})-V_T \qquad (4.55) \\
&= \epsilon_T^{(\mathcal{L})}\sigma_T^{\mathcal{L}}(F_{T,T}^{(m)}-F_{t_0,T})+\mu_T^{\mathcal{L}}(F_{T,T}^{(m)}-F_{t_0,T})-V_T \qquad (4.56)
\end{aligned}$$

Denote

$$K := \frac{V_T - \mu_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T})}{\sigma_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T})}.$$

This yields

$$\chi(\epsilon_T^{(\mathcal{L})}) := \mathbb{I}_{\{\mathcal{L}_T(F_{T,T}-F_{t_0,T})>V_T\}} = \begin{cases} \mathbb{I}_{\{\epsilon_T^{(\mathcal{L})}>K\}}, & \text{if } F_{T,T}^{(m)} > F_{t_0,T} \\ \mathbb{I}_{\{\epsilon_T^{(\mathcal{L})}<K\}}, & \text{if } F_{T,T}^{(m)} < F_{t_0,T} \\ \mathbb{I}_{\{V_T<0\}}, & \text{if } F_{T,T}^{(m)} = F_{t_0,T} \end{cases} \tag{4.57}$$

Furthermore,

$$\begin{aligned} (\mathcal{L}_T(F_{T,T}^{(m)} - F_{t_0,T}) - V_T)^2 \quad = \quad & \left(\epsilon_T^{(\mathcal{L})}\right)^2 \left(\sigma_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T})\right)^2 \\ & + \epsilon_T^{(\mathcal{L})} 2\sigma_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T}) \left(\mu_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T}) - V_T\right) \\ & + \left(\mu_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T}) - V_T\right)^2 \end{aligned} \tag{4.58}$$

Plugging (4.57) and (4.58) into (4.52),

$$\begin{aligned} & \int_{\bar{L}} (\bar{L}(F_{T,T}^{(m)} - F_{t_0,T}) - V_T)^2 \mathbb{I}_{\{\bar{L}(F_{T,T}^{(m)} - F_{t_0,T})>V_T\}} f_{\mathcal{L}_T}(\bar{L}) d\bar{L} \\ & \quad = \left(\sigma_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T})\right)^2 \int_{\bar{\epsilon}\in\mathbb{R}} \bar{\epsilon}^2 \chi(\bar{\epsilon}) f_{\epsilon_T^{(\mathcal{L})}}(\bar{\epsilon}) d\bar{\epsilon} \\ & \quad +2\sigma_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T}) \left(\mu_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T}) - V_T\right) \int_{\bar{\epsilon}\in\mathbb{R}} \bar{\epsilon} \chi(\bar{\epsilon}) f_{\epsilon_T^{(\mathcal{L})}}(\bar{\epsilon}) d\bar{\epsilon} \\ & \quad + \left(\mu_T^{\mathcal{L}}(F_{T,T}^{(m)} - F_{t_0,T}) - V_T\right)^2 \int_{\bar{\epsilon}\in\mathbb{R}} \chi(\bar{\epsilon}) f_{\epsilon_T^{(\mathcal{L})}}(\bar{\epsilon}) d\bar{\epsilon} \end{aligned} \tag{4.59}$$

The three integrals in the right side of (4.59) can be computed analytically by Theorem 4.10.5 since $\epsilon_T^{(\mathcal{L})}$ has the standard Gaussian distribution.

Theorem 4.10.5 *Let Y be a standard Gaussian random variable, ϕ be its pdf and Φ be its cdf. Let a be a constant. Then,*

$$\mathbb{E}\left[Y\mathbb{I}_{\{Y>a\}}\right] \quad = \quad \phi(a), \tag{4.60}$$

$$\mathbb{E}\left[Y^2\mathbb{I}_{\{Y>a\}}\right] \quad = \quad 1 - \Phi(a) + a\phi(a). \tag{4.61}$$

4.10.12 Linear-quadratic penalty

The quadratic penalty is symmetric and therefore it sanctions gains and losses equally. To alleviate this problem, an incentive to obtain gains could be incorporated in the penalty.

Such an alternative is to consider the linear-quadratic penalty used in Ni et al. (2012) that offers a trade-off between upside and downside risk :

$$g(x) = x + \zeta x^2 \tag{4.62}$$

for some constant ζ characterizing the risk aversion of the hedger. This penalty is asymmetric and permits to penalize losses and gains unequally. Another advantage of this penalty is its flexibility : it permits to calibrate the risk aversion of the hedger by varying the ζ parameter. However large gains will still be penalized in this case, which is not a desirable property.

Computing the solution

The next theorem implies that the linear-quadratic penalty can be seen as an extension of the quadratic one :

Theorem 4.10.6 *Define $g(x) := x^2$ and $\tilde{g} := x + \zeta x^2$. Then, the two following problems yield the same solution :*

$$\begin{aligned} &\underset{(\theta_{t_0+1},\ldots,\theta_T)\in\Theta}{\arg\min} \ \mathbb{E}\left[\tilde{g}(\Psi_T - V_T)|\mathcal{G}_{t_0}\right] \\ &\underset{(\theta_{t_0+1},\ldots,\theta_T)\in\Theta}{\arg\min} \ \mathbb{E}\left[g(\tilde{\Psi}_T - V_T)|\mathcal{G}_{t_0}\right] \end{aligned} \tag{4.63}$$

where $\tilde{\Psi}_T := \Psi_T + \frac{1}{2\zeta}$.

Proof of Theorem 4.10.6 : $x + \zeta x^2 = \zeta\left(x + \frac{1}{2\zeta}\right)^2 - \frac{1}{4\zeta}$. $\square$

Thus, using the linear-quadratic penalty is equivalent to using the quadratic one where the target Ψ_T is shifted by some constant. This therefore allows using the semi-explicit formulas of Schweizer (1995) to solve Equation (4.63). Some criterion should be used to select an appropriate value for ζ.

From (4.39), the two factors that must be computed to identify the optimal portfolio solving problem (4.63) are "α" and "b" factors. Let $\tilde{b}_t$ be the "b" coefficient solving problem (4.63) and b_t be the one solving problem (4.8). Similarly, let $\tilde{\alpha}_t$ be the "α" coefficient solving problem (4.63) and α_t be the one solving problem (4.8). Since the "b" does not depend on the payoff Ψ_T, Theorem 4.10.6 implies that $\tilde{b}_t = b_t$. Furthermore, α_t and $\tilde{\alpha}_t$ can be linked through a simple relationship. From results of Section 4.10.9,

$$\begin{aligned}
\tilde{\alpha}_t &:= A_t^{-1}\mathbb{E}\left[B_T^{-1}\tilde{\Psi}_T\Delta_t P_{t+1}|\mathcal{G}_{t-1}\right] \\
&= A_t^{-1}\mathbb{E}\left[B_T^{-1}\Psi_T\Delta_t P_{t+1}|\mathcal{G}_{t-1}\right] + A_t^{-1}\mathbb{E}\left[B_T^{-1}\frac{1}{2\zeta}\Delta_t P_{t+1}|\mathcal{G}_{t-1}\right] \\
&= \alpha_t + B_T^{-1}\frac{1}{2\zeta}b_t \qquad (4.64)
\end{aligned}$$

From (4.39), the solution to problem (4.63) is

$$\begin{aligned}
\theta_t &= \tilde{\alpha}_t - B_{t-1}^{-1}V_{t-1}\tilde{b}_t \qquad (4.65)\\
&= \alpha_t - \left(B_T^{-1}\frac{1}{2\zeta} + B_{t-1}^{-1}V_{t-1}\right)b_t \quad \text{(from (4.64))} \qquad (4.66)\\
&= \alpha_t - B_{t-1}^{-1}\tilde{V}_{t-1}b_t \qquad (4.67)
\end{aligned}$$

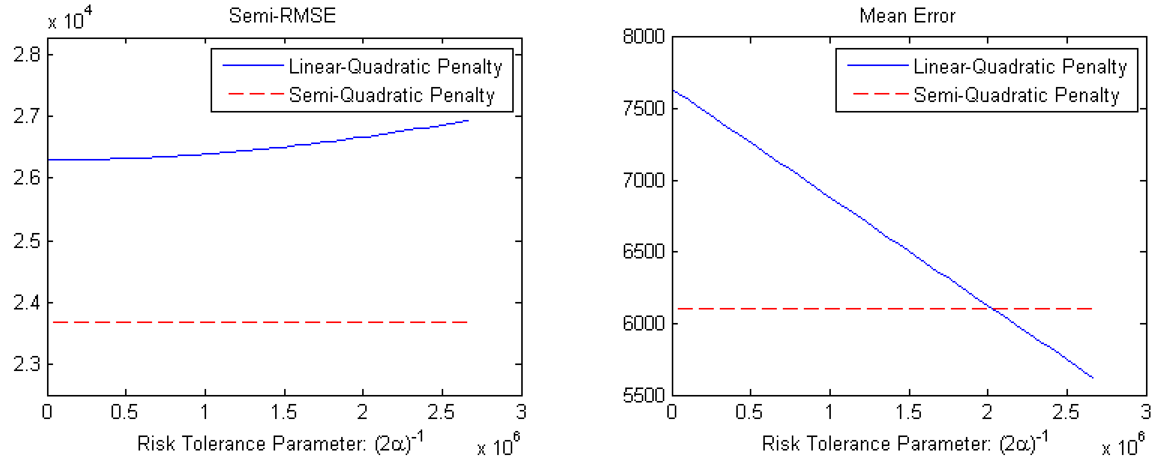
where $\tilde{V}_{t-1} = V_{t-1} + \frac{B_T^{-1}}{B_{t-1}^{-1}}\frac{1}{2\zeta}$. The solution to problem (4.63) can therefore be obtained simply by using the α_t and b_t solving (4.8). Thus, by solving (4.8), one simultaneously solves (4.63) for all ζ, avoiding additional calculations.

4.10.13 Numerical results for the linear-quadratic penalty

The in-sample experiment with real data of Section 4.4.2 is repeated with the linear-quadratic benchmark characterized by the problem (4.63). Instead of fixing the risk-aversion parameter ζ, various choices are attempted and the impact of this parameter is observed. Define Tol $:= \frac{1}{2\zeta}$ to be the risk tolerance. The case Tol $= 0$ is the quadratic penalty case given by problem (4.8) with $G(x) = x^2$. The mean and semi-RMSE of hedging errors are computed for the various choices of ζ and compared to the semi-quadratic hedging method SQDGH. The results are found in Figure 4.9.

The risk-reward trade-off offered by varying ζ is clearly illustrated in Figure 4.9 ; as the risk tolerance increases, the mean error decreases and the semi-RMSE increases. However, the semi-RMSE is lower for the semi-quadratic SQDGH than for the linear-quadratic hedging with any value of ζ. The former method is therefore more efficient in reducing risk than its counterpart. However, if the LSE hedger desires to achieve some risk-reward trade-off instead of only minimizing risk, the linear-quadratic penalty could be considered. Those observations indicate that the linear semi-quadratic penalty $g(x) = x + \zeta x^2 \mathbb{I}_{\{x>0\}}$ could also be used to achieve a risk-reward trade-off.

FIGURE 4.9 – Linear-quadratic hedging in the in-sample experiment



Comparing the linear-quadratic hedging and the semi-quadratic hedging SQDGH in the in-sample hedging back-test.

4.10.14 Further extensions

The models used in the current paper could be refined by means of several following extensions. A first possibility would be to re-balance the hedging portfolio daily instead of weekly to see it the hedging performance is improved. Models for the daily load, spot and futures price would be required in that case. Also, as indicated in Lucia and Torro, hydropower reservoir level can affect the relationship between the spot and futures price of electricty. Using reservoir levels as an additional explanatory variable could improve the spot and futures price predictive distribution. Weather related variables could also be incorporated in the model to improve the prediction distribution of the load variable. Concerning the futures price model, a time-varying mean in returns, which would represent a time varying risk premium as identified in Lucia & Torro (2011), could be implemented.

Additional risks faced by LSE could also be considered. The risk that a LSE pays a spot price that differs from the system price, due to grid congestions, might be considered. This risk can be at least partially hedged through contracts for difference (CfD) available on NASDAQ OMX. Moreover, LSE can bear a currency risk if it uses another currency than Euros to report its profit since the futures contracts are quoted in Euros. Currency exchange rates movements could be considered.

Finally, the short term horizon of hedges described in this paper could be increased to implement monthly, quarterly and yearly hedges with forward contracts.

References

[1] Abdel-Aal, R.E. (2008). Univariate modeling and forecasting of monthly energy demand time series using abductive and neural networks. *Computers & Industrial Engineering.* **(54)**. pp. 903-917.

[2] Abdel-Aal, R.E., Al-Garni, A.Z. (1997). Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis. *Energy.* **(22)**. pp. 1059-1069.

[3] Barakat, E.H., Al-Qasem, J.M. (1998). Methodology for weekly load forecasting. *IEEE Transactions on Power Systems.* **(13)**. pp. 1548-1555.

[4] Bessembinder, H., Lemmon, M.L. (2002). Equilibrium pricing and optimal hedging in electricity forward markets. *Journal of Finance.* **(57)**. pp. 1347-1382.

[5] Benth, F.E., Benth, J.S., Koekebakker, S. (2008). *Stochastic Modelling of Electricity and Related Markets.* World Scientific, Singapore.

[6] Bertsekas, D.P. (1995). *Dynamic Programming and Optimal Control.* Athena Scientific, Belmont, MA.

[7] Byström, H.N.E. (2003). The hedging performance of electricity futures on the Nordic power exchange. *Applied Economics.* **(35)**. p. 1-11.

[8] Coulon, M., Powell, W.B., Sircar, R. (2012). A model for hedging load and price risk in the Texas electricity market. Working paper, Princeton University, NJ.

[9] Deng, S.J., Oren, S.S. (2006). Electricity derivatives and risk management. *Energy.* **(31)**. pp. 940-953.

[10] Deng, S.J., Xu, L. (2009). Mean-risk efficient portfolio analysis of demand response and supply resources. *Energy* **(34)**. pp. 1523-1529.

[11] Ederington, L.H. (1979). The hedging performance of the new futures markets. *Journal of Finance.* **(34)**. pp. 157-170.

[12] Eydeland, A., Wolyniec, K. (2003). *Energy and Power Risk Management : New Developments in Modeling, Pricing, and Hedging*. Wiley, Hoboken, NJ

[13] Fleten, S.E., Brathen, E., Nissen-Meyer, S.E. (2010). Evaluation of static hedging strategies for hydropower producers in the Nordic market. *Journal of Energy Markets*. **(3)**. pp. 1-28.

[14] Föllmer, H., Leukert, P. (1999). Quantile hedging. *Finance and Stochastics* **(3)**. pp. 251-273.

[15] Föllmer, H., Leukert, P. (2000). Efficient hedging : cost versus shortfall risk. *Finance and Stochastics*. **(4)**. pp. 117-146.

[16] François, P., Gauthier, G., Godin, F. (2014). Optimal hedging when the underlying asset follows a regime-switching Markov process. *European Journal of Operational Research*. **(237)**. pp. 312-322.

[17] Genest, C., Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*. **(44)**. pp. 1096-1127.

[18] González-Romera, E., Jaramillo-Morán, M.A., Carmona-Fernández, D. 2008. Monthly electric energy demand forecasting with neural networks and Fourier series. *Energy Conversion and Management*. **(49)**. pp. 3135-3142.

[19] Hahn, H., Meyer-Nieberg, S., Pickl, S. (2009). Electric load forecasting methods : Tools for decision making. *European Journal of Operational Research*. **(199)**. pp. 902-907.

[20] Huovila, S. 2003. Short-term forecasting of power demand in the Nord Pool market. Master thesis. Lappeenranta University of Technology. Lappeenranta. Finland.

[21] Johnsen, T.A., Olsen, O.J. (2011). Regulated and unregulated Nordic retail prices. *Energy Policy*. **(39)**. pp. 3337-3345.

[22] Koekebakker, S., Ollmar, F. (2005). Forward curve dynamics in the Nordic electricity market. *Managerial Finance*. **(31)**. pp. 73-94.

[23] Kleindorfer, P. R., Li, L. (2005). Multi-period VaR-constrained portfolio optimization with applications to the electric power sector. *Energy* **(26)**. pp. 1-26.

[24] Madaleno, M., Pinho, C. (2008). The hedging effectiveness of electricity futures. Working paper. University of Aveiro. Portugal.

[25] Liu, S.D., Jian, J.B., Wang, Y.Y. (2010). Optimal dynamic hedging of electricity futures based on copula-GARCH models. *Industrial Engineering and Engineering Management (IEEM) Conference Proceedings*. Macau. China. pp. 2498-2502.

[26] Lucia, J.J., Torro, H. (2011). On the risk premium in Nordic electricity futures prices. *International Review of Economics & Finance*. **(20)**. pp. 750-763.

[27] Ni, J., Chu, L.K., Wu, F., Sculli, D., Shi, Y. (2012). A multi-stage financial hedging approach for the procurement of manufacturing materials. *European Journal of Operational Research*. **(221)**. pp. 424-431.

[28] Nordic Energy Regulators (NordREG). (2010). The Nordic financial electricity market. Report 8/2010, November 2010.

[29] Oum, Y., Oren, S.S., Deng, S.J. (2006). Hedging quantity risks with standard power options in a competitive wholesale electricity market. *Naval Research Logistics*. **(53)**. pp. 697-712.

[30] Oum, Y., Oren, S.S. (2009). VaR constrained hedging of fixed price load-following obligations in competitive electricity markets. *Risk and Decision Analysis* **(1)**. pp. 43-56.

[31] Oum, Y., Oren, S.S. (2010). Optimal static hedging of volumetric risk in a competitive wholesale electricity market. *Decision Analysis* **(7)**. pp. 107-122.

[32] Rémillard, B. 2013. *Statistical Methods for Financial Engineering*. CRC Press. Boca Raton. FL.

[33] Rémillard, B., Langlois, H., Hocquard, A., Papageorgiou, N. (2010). Optimal hedging of American options in discrete time. R. A. Carmona, P. Del Moral, P. Hu, N. Oudjane, eds. *Numerical Methods in Finance*. Springer. Berlin. Germany. pp. 145-170.

[34] Rémillard, B., Rubenthaler, S. (2013). Optimal hedging in discrete time. *Quantitative Finance*. **(13)**. pp. 819-825.

[35] Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*. **(23)**. pp. 470-472.

[36] Schweizer, M. (1995). Variance-optimal hedging in discrete time. *Mathematics of Operation Research*. **(20)**. pp. 1-32.

[37] State of California. (2004). Order Instituting Rulemaking to Establish Policies and Cost Recovery Mechanisms for Generation Procurement and Renewable Resource

Development Interim Opinion. Public Utility Commission Decision No. 04-01-050. http ://docs.cpuc.ca.gov/PUBLISHED/FINAL_DECISION/33625.htm

[38] Stoft, S., Belden, T., Goldman, C., Pickle., S. (1998). A primer on electricity futures and other derivatives. Technical Report. Laurence Berkeley National Laboratory. University of California. Berkeley.

[39] Torro, H. (2011). Assessing the influence of spot price predictability on electricity futures hedging. *Journal of Risk*. **(13)**. pp. 31-61.

[40] Von der Fehr, N.H.M., Hansen, P.V. (2010). Electricity retailing in Norway. *Energy*. **(31)**. pp. 25-45.

[41] Wagner, M., Skantze, P., Ilic, M. (2003). Hedging optimization algorithms for deregulated electricity markets. *Proceedings of the 12th conference on intelligent systems application to power systems*. Lemnos. Greece.

[42] Weron, R. (2006). *Modeling and Forecasting Electricity Loads and Prices*. Wiley. Chichester. England.

[43] Woo, C.K., Karimov, R., Horowitz, I. (2004). Managing electricity procurement cost and risk by a local distribution company. *Energy Policy*. **(32)**. pp. 635-645.

[44] Zanotti, G., Gabbi, G., Geranio, M. (2010). Hedging with futures : Efficacy of GARCH correlation models to European electricity markets. *Journal of International Financial Markets, Institutions and Money*. **(20)**. pp. 135-148.

Chapitre 5

Conclusion générale

La présente thèse illustre le développement et l'implémentation de la méthode de couverture globale de risques dans différents marchés incomplets, dont les marchés à changement de régime, les marchés où des frais de transaction sont encourus et les marchés de l'électricité. Tout d'abord, les problèmes de couverture sont définis rigoureusement dans chacun des articles. Ensuite, des algorithmes théoriques et numériques sont proposés afin de résoudre ceux-ci. Dans certain cas, comme par exemple dans l'article traitant du marché de l'électricité, des modèles sont bâtis pour représenter la dynamique des variables d'état impliquées dans le problème. Des simulations numériques ou des expériences avec des données réelles ont été réalisées dans chacun des articles pour évaluer l'efficacité des méthodes qui ont été proposées.

La performance de la méthode de couverture globale se compare avantageusement à d'autres méthodes utilisées dans la littérature, dont le *delta-hedging*, les méthodes de couverture locale et les stratégies de couverture statique ; les estimations des mesures de risques appliquées aux erreurs de couverture obtenues lors de simulations numériques sont moindres pour les méthodes de couverture globale que pour les autres. On peut donc considérer que celles-ci sont plus efficaces pour réduire les risques associés à l'usage de produit dérivés.

Le principal obstacle pratique à l'implémentation d'algorithmes de couverture globale des risques est leur lourdeur numérique. Il sera donc nécessaire d'identifier des méthodes numériques plus efficaces afin de résoudre plus rapidement les problèmes d'optimisation stochastique, dont l'équation de Bellman de la programmation dynamique. Les méthodes d'interpolation spectrale ou de simulation et régression sont de bons candidats potentiels pour l'accomplissement d'une telle tâche. Les outils d'apprentissage automatique (*machine learning*) pourraient aussi être à considérer. Le développement de méthodes numériques plus efficaces pourrait aussi permettre d'ajouter des dimensions aux problèmes de couverture et ainsi de considérer par exemple une volatilité stochastique, un taux d'intérêt sans-risque stochastique, de l'auto-corrélation dans les innovations ou même l'ajout d'autres actifs financiers dans le portefeuille de couverture.

Étant donné que les procédures de couverture globale ne sont pas encore répandues dans la littérature, il serait aussi souhaitable d'effectuer plusieurs tests avec des données réelles de différents marchés pour vérifier si l'efficacité de la méthode par rapport aux autres méthodes persiste. Il serait aussi nécessaire d'évaluer la robustesse de la méthode par rapport au risque de modèle i.e. lorsque le modèle présumé par l'algorithme de couverture diffère de la dynamique réelle du sous-jacent.

Bref, plusieurs travaux de recherche sont probablement encore nécessaires avant de voir les méthodes de couverture globale se répandre à travers la littérature et dans l'industrie financière. Cependant, les résultats obtenus dans la présente thèse par rapport à cette méthode de couverture sont très prometteurs, et cette dernière pourrait devenir un puissant outil de gestion des risque si elle est bien développée.