Essays on Modeling, Hedging and Pricing of Insurance and Financial Products

Thèse

Denis-Alexandre Trottier

Doctorat en sciences de l'administration - finance et assurance
Philosophiæ doctor (Ph. D.)

Québec, Canada

Essays on Modeling, Hedging and Pricing of Insurance and Financial Products

Thèse

Denis-Alexandre Trottier

Sous la direction de:

Van Son Lai, directeur de recherche

Résumé

Cette thèse est composée de trois articles abordant différentes problématiques en relation avec la modélisation, la couverture et la tarification des risques d'assurance et financiers.

"A general class of distortion operators for pricing contingent claims with applications to CAT bonds" est un projet présentant une méthode générale pour dériver des opérateurs de distorsion compatibles avec la valorisation sans arbitrage. Ce travail offre également une nouvelle classe simple d'opérateurs de distorsion afin d'expliquer les primes observées dans le marché des obligations catastrophes.

"Local hedging of variable annuities in the presence of basis risk" est un travail dans lequel une méthode de couverture des rentes variables en présence de risque de base est développée. La méthode de couverture proposée bénéficie d'une exposition plus élevée au risque de marché et d'une diversification temporelle du risque pour obtenir un rendement excédentaire et faciliter l'accumulation de capital.

"Option pricing under regime-switching models : Novel approaches removing path-dependence" est un projet dans lequel diverses mesures neutres au risque sont construites pour les modèles à changement de régime de manière à générer des processus de prix d'option qui ne présentent pas de dépendance au chemin, en plus de satisfaire d'autres propriétés jugées intuitives et souhaitables.

Abstract

This thesis is composed of three papers addressing different issues in relation to the modeling, hedging and pricing of insurance and financial risks.

"A general class of distortion operators for pricing contingent claims with applications to CAT bonds" is a project presenting a general method for deriving probability distortion operators consistent with arbitrage-free pricing. This work also offers a simple novel class of distortions operators for explaining catastrophe (CAT) bond spreads.

"Local hedging of variable annuities in the presence of basis risk" is a work in which a method to hedge variable annuities in the presence of basis risk is developed. The proposed hedging scheme benefits from a higher exposure to equity risk and from time diversification of risk to earn excess return and facilitate the accumulation of capital.

"Option pricing under regime-switching models: Novel approaches removing path-dependence" is a project in which various risk-neutral measures for hidden regime-switching models are constructed in such a way that they generate option price processes which do not exhibit path-dependence in addition to satisfy other properties deemed intuitive and desirable.

Table des matières

Liste des tableaux

Liste des figures

Remerciements

Je tiens tout d'abord à remercier Prof. Van Son Lai pour m'avoir accompagné à travers mes études, non seulement en tant que superviseur, mais également en tant que mentor, me guidant dans la sélection des meilleures ressources pour effectuer mes recherches et améliorer davantage mes compétences en ingénierie financière. Van Son fut un superviseur de recherche très généreux et engagé, ainsi qu'une excellente référence.

Je tiens également à remercier mes co-auteurs, Prof. Frédéric Godin et Emmanuel Hamel pour m'avoir invité à participer à des recherches très stimulantes. Prof. Frédéric Godin a contribué à la complétion de mes projets de recherches et m'a permis de bénéficier de sa grande expérience et de ses judicieux conseils.

Je remercie également les membres de mon comité de thèse, Prof. Marie-Claude Beaulieu, Prof. Issouf Soumaré et Prof. Bruno Rémillard pour avoir pris le temps de s'intéresser à mes travaux de recherche et pour leurs recommandations qui ont contribué à les améliorer.

Je suis extrêmement reconnaissant envers la Bourse de Montréal, l'Autorité des marchés financiers, le Conseil de recherches en sciences naturelles et en génie du Canada, le Fonds de recherche du Québec - Société et culture, la Banque Nationale et le Centre de recherche sur les risques, les enjeux économiques et les politiques publiques. Je me sens profondément privilégié pour ce financement qui m'a permis de me concentrer pleinement sur mes recherches.

Finalement, je remercie ma famille et ma copine, Anne-Marie, pour leur incroyable support.

Avant-propos

Mon co-auteur pour l'article intitulé *"A general class of distortion operators for pricing contingent claims with applications to CAT bonds"* est Prof. Van Son Lai. J'ai proposé l'idée sous-jacente à l'article et ai effectué l'ensemble des travaux de recherche et de rédaction, avec des apports de Prof. Van Son Lai pour compléter la bibliographie, se positionner par rapport à la littérature, renforcer la structure de l'article et polir la version finale.

L'article intitulé *"Local hedging of variable annuities in the presence of basis risk"* a été accepté pour publication dans *ASTIN Bulletin*. La version acceptée pour publication est celle présentée dans cette thèse. J'ai le rang de premier auteur dans cet article et j'ai pour co-auteurs Prof. Frédéric Godin et Emmanuel Hamel. Ces derniers m'ont invité au début de l'été 2016 à participer à ce projet de recherche financé par l'Autorité des marchés financiers. La problématique générale étudiée dans cet article (couverture de fonds distincts en présence de risque de base) fut proposée par Emmanuel Hamel et Prof. Frédéric Godin. La stratégie de recherche, i.e., étudier une version simplifiée du modèle afin d'établir les résultats fondamentaux, a été proposée par moi-même. Je suis responsable pour la majorité des résultats analytiques et numériques présentés dans l'article. Une exception est l'équation (2.3.12) proposée par Prof. Frédéric Godin, qui est une version améliorée de l'approximation que j'avais initialement proposée. La rédaction de l'article a principalement été effectuée par moi-même et Prof. Frédéric Godin, avec certains apports techniques de Emmanuel Hamel.

L'article *"Option pricing under regime-switching models : Novel approaches removing path-dependence"* a été soumis à *Finance and Stochastics* le 25 novembre 2017. Prof. Van Son Lai et Prof. Frédéric Godin sont mes co-auteurs. L'idée sous-jacente à cet article est née de discussions entre ce-dernier et moi-même. Prof. Frédéric Godin a initialement proposé la solution à la Section 3.4.2 et j'ai ensuite dérivé la généralisation à la Section 3.4. Je suis également responsable pour les solutions de la Section 3.5. La rédaction de l'article a principalement été effectuée par moi-même et Prof. Frédéric Godin, en plus de grandement bénéficier des apports de Prof. Van Son Lai pour le polissage et le positionnement de la version finale.

Introduction

Les différents travaux présentés dans cette thèse sont issus de deux lignées de recherche : les obligations catastrophes et les fonds distincts.

Les obligations catastrophes furent inventées par les (ré)assureurs comme moyen de transférer certains risques de catastrophes naturelles vers les marchés de capitaux. Ces instruments offrent aux investisseurs une exposition au risque de désastre naturel sous la forme familière d'une obligation dont les coupons et le principal sont à risque, i.e., contingents à l'absence de désastres naturels qui sont tels que spécifiés par le contrat. La littérature théorique sur la valorisation d'obligations catastrophes est relativement développée et sophistiquée, principalement basée sur la valorisation neutre au risque en marché incomplet pour des processus stochastiques avec sauts ou autres formes de discontinuités. Il en est autrement pour la recherche empirique, qui se base plutôt sur des approches de régression ad hoc sans fondement théorique. Selon moi, l'obstacle principal à l'application empirique des modèles de valorisation théoriques issus de la littérature financière est que ceux-ci sont généralement formulés en termes d'information indisponible publiquement pour les obligations catastrophes, e.g., le seuil de déclenchement. Le premier projet présenté dans cette thèse adresse ce problème en reformulant les modèles de valorisation neutre au risque sous le formalisme des opérateurs de distorsion qui, eux, peuvent être appliqués à partir des données disponibles publiquement pour les obligations catastrophes. Plus précisément, ce travail présente une méthodologie générale permettant de construire des opérateurs de distorsion compatibles avec les modèles de valorisation neutre au risque retrouvés dans la littérature financière. De tels résultats engendrent plusieurs opportunités de recherches empiriques portant sur les obligations catastrophes. Nous utilisons ces résultats pour dériver de nouveaux opérateurs de distorsion servant à expliquer les primes de risques observées empiriquement dans ces marchés.

Le second projet de recherche porte sur la couverture des risques associés aux fonds distincts en présence de risque de base. Il s'agit d'une problématique complexe étant donné la grande quantité de risques impliqués, e.g., risques d'équité, de taux d'intérêt, de mortalité, de base et d'abandon. De plus, la couverture des options à longue maturité, telles que celles incluses au sein des polices de fonds distincts, est peu étudiée étant donné que les options à court terme sont plus typiques dans l'industrie financière. Nous avons donc développé une approche locale de couverture s'appliquant aux polices de fonds distincts lorsqu'il y a risque de base qui surpasse et généralise les méthodes de références dans la littérature et utilisées en pratique.

Indirectement, cette dernière recherche nous a également poussés à réfléchir à certains problèmes en lien avec la valorisation de produits dérivés en contexte de marché à changement de régime. Nous avons réalisé que l'approche habituelle pour construire la mesure neutre au risque impliquait des résultats contre-intuitifs lorsque appliquée à un sous-jacent dont la dynamique est à changement de régime, i.e., on montre qu'il y a alors dépendance au chemin dans les prix d'options vanilles. C'est ainsi que nous avons entrepris le dernier projet de recherche présenté dans cette thèse. Ce projet développe de nouvelles mesures neutres au risque intuitives pouvant incorporer de manière simple l'aversion au risque de régime et qui n'entraînent pas de tels effets secondaires de dépendance au chemin.

Cette thèse est structurée de la façon suivante. Le chapitre 1 présente la méthodologie générale pour dériver des opérateurs de distorsion compatibles avec la valorisation neutre au risque et présente les applications en lien avec les obligations catastrophes. Le chapitre 2 présente notre méthode de couverture des rentes variables en présence de risque de base. Le chapitre 3 présente nos nouvelles mesures neutres au risque pour les modèles à changement de régime, qui n'entraînent pas les effets secondaires de dépendance au chemin.

Chapter 1

A general class of distortion operators for pricing contingent claims with applications to CAT bonds

Résumé

Wang (2000) proposa un opérateur de distorsion permettant de récupérer les formules Black-Scholes de tarification d'options. Godin et al. (2012) généralisent cette approche de distorsion à la tarification sans arbitrage à une extension du modèle Black-Scholes basée sur la distribution normale inverse gaussienne pour la sous-classe de mesures martingales correctrices de moyenne. Nous généralisons ces travaux en offrant une classe d'opérateurs de distorsion compatible de façon plus générale avec la valorisation neutre au risque, ce qui ajoute de la flexibilité pour choisir le modèle actif/passif et la mesure neutre au risque. Nous dérivons ensuite plusieurs nouveaux opérateurs de distorsion améliorés permettant de valoriser les risques financiers et d'assurance. Enfin, nous présentons une nouvelle classe de distorsions pour évaluer les obligations catastrophes et offrons une validation empirique.

Abstract

Wang (2000) proposes a distortion operator that recuperates the Black-Scholes option pricing formulas. Godin et al. (2012) extend this distortion-based arbitrage-free pricing approach to a Normal Inverse Gaussian Black-Scholes world for the mean-correcting subclass of risk-neutral measures. We generalize this line of work by offering a class of distortion operators that is compatible with risk-neutral valuation more broadly, adding flexibility to the choices of the asset/liability model and the risk-neutral measure underlying the distortion. We then derive several new and improved distortion operators that can be used to price both financial and insurance risks. Finally, we present a novel class of distortions to price catastrophe bonds and provide an empirical validation.

Keywords : Distortion operator, Wang transform, Distortion risk measure, Arbitrage-free pricing, Insurance pricing, Contingent claim pricing, Pricing of CAT bonds.

1.1 Introduction

Wang (2000) proposes a probability distortion operator $g_\alpha(u) = \Phi(\Phi^{-1}(u) + \alpha)$, $u \in [0, 1]$, to price both financial and insurance risks, where Φ is the standard normal cumulative distribution function. In particular, Wang shows that this transform can recover the classical Black-Scholes option pricing formulas. Hamada and Sherris (2003) and Pelsser (2008), among others, study further the applicability of the Wang transform for contingent claims pricing. These authors find the Wang transform consistent with arbitrage-free pricing when the underlying asset follows a geometric Brownian motion, but inadequate under non-Gaussian assumptions. Addressing this limitation, by way of an exponential Normal Inverse Gaussian (NIG) Lévy motion, Godin et al. (2012) propose a distortion operator that can recuperate the arbitrage-free prices under the mean-correcting equivalent martingale measure.

In this paper, we develop a framework for deriving distortion operators that are compatible with risk-neutral valuation under more general assumptions for the underlying asset model and the risk-neutral measure. Using our methodology, we produce new distortions that can recuperate pricing functionals of popular financial and insurance models well beyond the Wang transform and its non-Gaussian extensions. We also present empirical applications of our approach in the characterization of catastrophe (CAT) bond spreads.

Our paper is related to the literature that studies the connection between the Wang transform and other forms of risk pricing. Conditions under which the Bühlmann (1980)'s pricing equilibrium yields the Wang transform are derived by Wang (2003). Multivariate extensions of the Wang transform based on similar lines of reasoning can be found in Kijima (2006), Wang (2007), and Kijima and Muromachi (2008). The equivalence between the Wang transform and the Esscher-Girsanov change of measure proposed by Goovaerts and Laeven (2008) is demonstrated in a static setting by Labuschagne and Offwood (2010). Relative to this literature, our paper is the first to describe the general connection between distortion operators and other pricing principles. The connections between the Wang transform, the Black-Scholes model, the Esscher-Girsanov change of measure, and the Bühlmann's equilibrium are recovered as special cases of our analyses.

Specifically, our contributions are as follows. First, we present the general expression of the distortion operator that recovers risk-neutral pricing functionals and we derive the conditions of applicability. Second, we characterize the change of probability measure applied by our distortion operator, and we show how the connections found in the literature between the Wang transform and other pricing frameworks can be viewed as manifestations of this more fundamental result. Third, we derive new distortion operators that are suitable for financial and insurance risk pricing. Our first distortion extends the NIG distortion of Godin et al. (2012) to the Esscher equivalent martingale measure. The second distortion recuperates the equilibrium prices in Kou (2002)'s jump-diffusion model. The third recovers the size-biased premium principle using the generalized beta of the second kind (GB2) distribution for a loss variable. The fourth distortion operator recuperates the Esscher premium principle for a gamma distributed loss variable. Finally, we propose distortion operators that can depict CAT bond market spreads and show their usefulness in an empirical analysis. Our results provide an interesting evidence that jump-diffusion models are appropriate for pricing CAT bonds, but that investors are averse to natural disasters.

As pointed out by Hamada and Sherris (2003), Pelsser (2008), and Godin et al. (2012), the normality assumption underlying the Wang transform poses significant limitations for practical applications. Our set of new distortion operators extends Godin et al. (2012) in the following manner. We offer a general class of distortion operators that is compatible with risk-neutral valuation, yielding flexibility in the selection of the asset/liability model and the risk-neutral measure. This opens a wide range of possible applications for future research. For instance, our distortion operators could be used to produce new *distortion risk measures*, which are quantile-based measures used by practitioners in finance and insurance, e.g., Dowd and Blake (2006). One advantage for the risk measures based on our proposed distortion operators is that they enable the incorporation of risk-aversion and other considerations that are embedded into the choice of an equivalent martingale measure, as well as risk distribution features (e.g., skewness and kurtosis). In different directions, the previous literature has much attempted to incorporate these features into risk measures (e.g., Bali and Theodossiou, 2008; Gzyl and Mayoral, 2008).

The paper proceeds as follows. Section 1.2 provides some background on distortion operators and other forms of risk pricing. Section 1.3 presents the general expression of our distortion operator and characterizes its applicability to arbitrage-free pricing. Section 1.4 derives new distortion operators consistent with popular financial non-Gaussian option pricing models. Section 1.5 performs the same exercise for insurance pricing models. Section 1.6 presents empirical applications to the pricing of catastrophe bonds. Section 1.7 concludes and discusses other potential applications.

1.2 Background on risk pricing

In this section, we briefly review popular financial and insurance pricing principles.

1.2.1 Arbitrage-free pricing

Let us consider a continuous-time economy where time t takes value within $[0, T]$. This economy stochastic behaviour is characterized by a probability space $(\Omega, \mathcal{F}_T, \mathbb{P})$ equipped with a filtration

$\{\mathcal{F}_t\}_{t\in[0,T]}$ satisfying the usual conditions.

A central result in the theory of asset pricing by arbitrage, the so-called first fundamental theorem of asset pricing, is the equivalence between the absence of (quasi-)arbitrage opportunities and the existence of an equivalent martingale measure $\mathbb{Q}$. See Delbaen and Schachermayer (1994) for the history of this theorem which goes back to the seminal papers of Harrison and Kreps (1979), and Harrison and Pliska (1981). Such a measure is often called a *risk-neutral measure* because the arbitrage-free price process $\{S_t\}_{t\in[0,T]}$ of any traded asset and derivative must satisfy

$$S_t = B_t\,\mathbb{E}^{\mathbb{Q}}\left[\frac{S_s}{B_s}\middle|\mathcal{F}_t\right], \qquad \forall t \le s \le T, \tag{1.2.1}$$

where $\{B_t\}_{t\in[0,T]}$ is the risk-free asset price process.

(Re)insurance contracts (Delbaen and Haezendonck, 1989; Sondermann, 1991) and insurance-linked securities (Vaugirard, 2003) employ arbitrage-free pricing techniques initially developed for financial derivatives. Such frameworks are often characterized by an underlying insurance or catastrophe loss process whose dynamics is punctuated by random jumps, e.g., a jump-diffusion process. It is well-known that with such (non-locally bounded) processes, [1] the market is incomplete, and therefore there exists an infinite number of risk-neutral measures. In this case, the choice of risk-neutral measure is determined by the market, i.e., by the equilibrium resulting from supply and demand which are in turn determined by aggregate risk-aversion, liquidity needs, and other factors. In the realm of incomplete markets, popular modelling assumptions are the minimal martingale measure (Föllmer and Schweizer, 1991), the Esscher martingale measure (Gerber and Shiu, 1994), the variance optimal martingale measure (Schweizer, 1995), the mean-correcting martingale measure, and equilibrium-based martingale measures.

Equilibrium-based martingale measures

General equilibrium models can be viewed as a subset of the arbitrage-free pricing framework in the sense that they are possible approaches for selecting the risk-neutral measure. Indeed, there are no arbitrage opportunities in a rational expectation equilibrium.

For example, consider a continuous-time extension of Lucas (1978)'s model. We assume there is a representative agent possessing endowments who maximizes the agents aggregate utility. Let $U(t, c_t)$ be the aggregate utility at time t for the consumption process $\{c_t\}_{t\in[0,T]}$. Under mild conditions, one can show that this setup produces a pricing kernel that depends only on the aggregate endowment process, denoted by $\{\delta_t\}_{t\in[0,T]}$, such that the price process $\{S_t\}_{t\in[0,T]}$ of

1. The version of the first fundamental theorem of asset pricing for non-locally bounded processes states that the condition of no free lunch with vanishing risk is equivalent to the existence of an equivalent sigma-martingale measure (Delbaen and Schachermayer, 1998). A semi-martingale X is a sigma-martingale if there exists a martingale M and an M-integrable predictable process ϕ such that $X_t = \int_0^t \phi_u dM_u$.

any traded asset and derivative must satisfy the following condition in equilibrium :

$$S_t = \mathbb{E}^{\mathbb{P}}\left[\frac{U_c(s,\delta_s)}{U_c(t,\delta_t)}\,S_s\middle|\mathcal{F}_t\right], \qquad \forall t \leq s \leq T, \tag{1.2.2}$$

where $\mathbb{P}$ is the physical measure, and $U_c \equiv \frac{\partial U}{\partial c}$. This equilibrium condition can be written in the form of equation (1.2.1) by noting that the pricing kernel can be used to define a Radon-Nikodym derivative.

1.2.2 Actuarial premium calculation principles

A prominent problem in actuarial science is to derive premium calculation principles (PCPs) that satisfy a number of desirable properties (see, e.g., Laeven and Goovaerts, 2008). We present below two of such popular approaches. [2]

Distortion operators

The history of distortion operators goes back to Yaari (1987)'s dual theory of choice under risk, in which attitudes toward risks are characterized by a distortion function rather than by an expected utility function. Distortion operators also stem from the axiomatic approach of Wang et al. (1997) to characterize insurance prices.

Let X be a random variable distributed with survival function $\bar{F}_{\mathbb{P}}(x) \equiv \mathbb{P}(X > x)$ under the physical measure $\mathbb{P}$. We introduce a distortion operator g which is an increasing and differentiable function such that $g(0) = 0$, $g(1) = 1$, and $g(u) \in [0,1]$ for all $u \in [0,1]$. It defines a change of measure such that X is distributed with survival function $\bar{F}_g(x) \equiv g(\bar{F}_{\mathbb{P}}(x))$ under the new probability measure.

The price of X is obtained via the expected value under the distorted probability measure. One can show that this expected value has the following Choquet integral representation :

$$H[X;g] \equiv \int_{-\infty}^{0} \left[g\big(\bar{F}_{\mathbb{P}}(x)\big) - 1\right]dx + \int_{0}^{\infty} g\big(\bar{F}_{\mathbb{P}}(x)\big)dx, \qquad \bar{F}_{\mathbb{P}}(x) \equiv \mathbb{P}(X > x). \tag{1.2.3}$$

This Choquet integral exhibits monotonicity, translation invariance, positive homogeneity, and is sub-additive if g is concave (Denneberg, 1994). Hence, the functional $H[\cdot;g]$ defines a distortion-based risk measure that is coherent in the sense of Artzner et al. (1999) if g is concave. [3]

2. There is a plethora of other insurance pricing principles, the reader is referred to Laeven and Goovaerts (2008) and Ai and Brockett (2008) for thorough accounts of these.

3. A more comprehensive treatise on distortion-based risk measures can be found in Dowd and Blake (2006).

Actuarial weighted pricing principles

Furman and Zitikis (2008) propose a broad class of PCPs based on weighted loss distributions. This class can also be viewed as a subclass of the loss function approach (see Remark 1 of Heilmann (1989)). Under this approach, the price of a risk $X \geq 0$ is given by

$$\Pi[X; w] = \frac{\mathbb{E}^{\mathbb{P}}[w(X)X]}{\mathbb{E}^{\mathbb{P}}[w(X)]}, \tag{1.2.4}$$

where $w(x) \geq 0$ for all $x \geq 0$. Thus, the change of probability measure whose Radon-Nikodym derivative is $w(x)/\mathbb{E}^{\mathbb{P}}[w(X)]$ characterizes this pricing principle. Several popular PCPs are contained within the weighted family. For instance, the Esscher principle is obtained with $w(x) = e^{ax}$ and the size-biased premium principle with $w(x) = x^a$, where $a \geq 0$ in both cases. [4]

1.3 A general framework for distortion-based risk-neutral valuation

This section contains our main theoretical results. First, we present the general definition of our distortion operator. Then, we characterize the change of probability measure it applies, and the conditions under which it can be used to compute arbitrage-free prices of derivatives.

1.3.1 A new class of distortion operators

We define below our distortion operator by Definition 1.3.1. This general expression can be reduced to a simpler form with fewer parameters using Proposition 1.3.1 whose proof is in Appendix 1.A.1. This can be crucial to circumvent the parameter identification issues that could otherwise arise in calibration.

Definition 1.3.1. Let X be a continuous random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{Q}$ be a probability measure equivalent to $\mathbb{P}$ on $\mathcal{F}$. Let $\bar{F}_{\mathbb{P}}(x) \equiv \mathbb{P}(X > x)$ and $\bar{F}_{\mathbb{Q}}(x) \equiv \mathbb{Q}(X > x)$. We define the following distortion operator :

$$g_X^{\mathbb{Q},\mathbb{P}}(u) \equiv \bar{F}_{\mathbb{Q}}\big(\bar{F}_{\mathbb{P}}^{-1}(u)\big), \qquad u \in [0, 1], \tag{1.3.1}$$

where $\bar{F}_{\mathbb{P}}^{-1}$ is the inverse of $\bar{F}_{\mathbb{P}}$ with the convention $\bar{F}_{\mathbb{P}}^{-1}(0) = +\infty$ and $\bar{F}_{\mathbb{P}}^{-1}(1) = -\infty$.

Proposition 1.3.1. *Let X be a continuous random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{Q}$ be a probability measure equivalent to $\mathbb{P}$ on $\mathcal{F}$. For any continuous and increasing function h,*

$$g_{h(X)}^{\mathbb{Q},\mathbb{P}}(u) = g_X^{\mathbb{Q},\mathbb{P}}(u), \qquad \forall u \in [0, 1]. \tag{1.3.2}$$

4. We refer the reader to Table 1 of Furman and Zitikis (2009) for additional examples.

Remark 1.3.1 (Wang transform)**.** Suppose that X is a standard normal $N(0,1)$ random variable under the measure $\mathbb{P}$, and that it is shifted to a $N(\theta,1)$ distribution under the measure $\mathbb{Q}$. In other words, we have $\mathbb{P}(X > x) = 1 - \Phi(x)$, and $\mathbb{Q}(X > x) = 1 - \Phi(x-\theta)$. Let h be any continuous and increasing function. By Proposition 1.3.1, $g^{\mathbb{Q},\mathbb{P}}_{h(X)}(u) = g^{\mathbb{Q},\mathbb{P}}_{X}(u) = \Phi\big(\Phi^{-1}(u)+\theta\big)$.

1.3.2 Change of measure performed by the distortion operator

The change of probability measure performed by the distortion operator $g^{\mathbb{Q},\mathbb{P}}_{X}$ in Definition 1.3.1 is characterized below by Theorem 1.3.1 whose proof is in Appendix 1.A.2. As stated in Corollary 1.3.1 (proven in Appendix 1.A.3), a key feature of the distortion $g^{\mathbb{Q},\mathbb{P}}_{X}$ is that it changes the probability measure from $\mathbb{P}$ to $\mathbb{Q}$ when applied on any random variable of the form $h(X)$, where h is any continuous and increasing function.

Definition 1.3.2. Let Z be a continuous random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{Q}$ be a probability measure equivalent to $\mathbb{P}$ on $\mathcal{F}$. Define $q_{\mathbb{P}}$ as the probability density function (PDF) of Z under $\mathbb{P}$, and define $q_{\mathbb{Q}}$ as its PDF under $\mathbb{Q}$. We define the likelihood ratio

$$\xi^{\mathbb{Q},\mathbb{P}}_{Z}(z) \equiv \frac{q_{\mathbb{Q}}(z)}{q_{\mathbb{P}}(z)}. \tag{1.3.3}$$

Theorem 1.3.1. *Let X and Z be continuous random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{Q}$ be a probability measure equivalent to $\mathbb{P}$ on $\mathcal{F}$. Denote the survival functions of X and Z under $\mathbb{P}$ by $\bar{F}_{\mathbb{P}}(x) \equiv \mathbb{P}(X > x)$ and $\bar{Q}_{\mathbb{P}}(z) \equiv \mathbb{P}(Z > z)$, and let the PDF of X under $\mathbb{P}$ be denoted by $f_{\mathbb{P}}$. The PDF of X under the probability measure distorted by $g^{\mathbb{Q},\mathbb{P}}_{Z}$ is given by*

$$f_{g^{\mathbb{Q},\mathbb{P}}_{Z}}(x) = f_{\mathbb{P}}(x)\, \xi^{\mathbb{Q},\mathbb{P}}_{Z}\big(\bar{Q}^{-1}_{\mathbb{P}}(\bar{F}_{\mathbb{P}}(x))\big), \tag{1.3.4}$$

where $\bar{Q}^{-1}_{\mathbb{P}}$ is the inverse of $\bar{Q}_{\mathbb{P}}$.

Corollary 1.3.1. *If $X = h(Z)$, where h is continuous and increasing, then the distorted distribution of X coincides with its distribution under $\mathbb{Q}$: $f_{g^{\mathbb{Q},\mathbb{P}}_{Z}}(x) = f_{\mathbb{Q}}(x)$, for all x in the support.*

1.3.3 Connections between the Wang transform and other pricing frameworks

There is a literature studying the connections between the Wang transform, the Esscher-Girsanov change of measure, and the Bühlmann general equilibrium model. We now show that these results can be recovered by virtue of Theorem 1.3.1. We refer to Wang (2003) and Labuschagne and Offwood (2010) for the original proofs. [5]

5. See Kijima (2006), Wang (2007), and Kijima and Muromachi (2008) for closely related works.

First, let's see how the Wang transform can be related to the Esscher-Girsanov change of measure. Suppose that Z is a $N(0,1)$ random variable under $\mathbb{P}$, and that its distribution is shifted to a $N(\theta,1)$ under $\mathbb{Q}$. As stated in Remark 1.3.1, the distortion operator is the Wang transform $g_Z^{\mathbb{Q},\mathbb{P}}(u) = \Phi\big(\Phi^{-1}(u)+\theta\big)$. Moreover, one can readily show that the Radon-Nikodym derivative is $\xi_Z^{\mathbb{Q},\mathbb{P}}(x) = e^{\theta x-\theta^2/2}$. Therefore, applying Theorem 1.3.1 with $\bar{Q}_{\mathbb{P}}(x) = 1-\Phi(x)$ gives us

$$f_{g_Z^{\mathbb{Q},\mathbb{P}}}(x) = e^{\theta\Phi^{-1}(F_{\mathbb{P}}(x))-\theta^2/2} f_{\mathbb{P}}(x), \tag{1.3.5}$$

where $F_{\mathbb{P}}(x) \equiv 1-\bar{F}_{\mathbb{P}}(x)$. When $\theta = h\nu$, the Definition 3.1 Esscher-Girsanov change of measure recovers the one in Labuschagne and Offwood (2010).

Next, let's see how the Wang transform can yield the Bühlmann (1980)'s pricing equilibrium. The key assumption in Wang (2003) is that X and Z and co-monotone in the sense that under $\mathbb{P}$ they can be expressed as $X = F_{\mathbb{P}}^{-1}(U)$ and $Z = \Phi^{-1}(U)$, where U is uniformly distributed between 0 and 1. Assuming this, using the expression $Z = \Phi^{-1}(F_{\mathbb{P}}(X))$ in (1.3.5) yields

$$f_{g_Z^{\mathbb{Q},\mathbb{P}}}(x) = \mathbb{E}^{\mathbb{P}}\left[e^{\theta Z-\theta^2/2}\middle|\, X = x\right] f_{\mathbb{P}}(x) = \frac{\mathbb{E}^{\mathbb{P}}\left[e^{\theta Z}\middle| X = x\right]}{\mathbb{E}^{\mathbb{P}}\left[e^{\theta Z}\right]} f_{\mathbb{P}}(x). \tag{1.3.6}$$

1.3.4 Risk-neutral pricing using distortion operators

Let X be a random variable representing a financial or insurance risk. For example, X could be the price of a traded asset (as it is for the case of financial derivatives), but it could also be something else like the number of heating degree days (say a weather derivative) or the value of a catastrophe loss index (as for insurance-linked securities). Our goal is to price derivatives on X using a distortion operator. The proposition below whose proof is in Appendix 1.A.4 gives the general solution to this problem.

Proposition 1.3.2. *Let X be a random variable on $(\Omega,\mathcal{F},\mathbb{P})$, and let $\mathbb{Q}$ be a measure equivalent to $\mathbb{P}$ on $\mathcal{F}$. For any continuous and increasing function h, we have*

$$H\left[h(X); g_X^{\mathbb{Q},\mathbb{P}}\right] = \mathbb{E}^{\mathbb{Q}}[h(X)]. \tag{1.3.7}$$

In particular, Proposition 1.3.2 also holds for any equivalent martingale measure $\mathbb{Q}$, in which case the arbitrage-free price of a derivative with terminal payoff $h(X)$ is given by the discounted value of $H\left[h(X); g_X^{\mathbb{Q},\mathbb{P}}\right]$. Less general versions of this result can be found in the literature. For instance, the proof when X is the terminal value of a geometric Brownian motion (see Remark 1.3.2) can be found in Hamada and Sherris (2003). The proof when X is the terminal value of an exponential NIG Lévy motion and $\mathbb{Q}$ is the mean-correcting martingale measure is in Godin et al. (2012).

Remark 1.3.2 (Black-Scholes model)**.** It follows from Proposition 1.3.2 that the distortion operator that can recuperate the arbitrage-free prices under the Black-Scholes model is $g_{S_T}^{\mathbb{Q},\mathbb{P}}$, where S_T is the terminal value of a geometric Brownian motion, with constant drift μ and volatility σ under the physical measure $\mathbb{P}$, and with drift r and volatility σ under the risk-neutral measure $\mathbb{Q}$, where r is the risk-free rate. Following Remark 1.3.1, one can check that this distortion reduces to the Wang transform $g_{S_T}^{\mathbb{Q},\mathbb{P}}(u) = \Phi\big(\Phi^{-1}(u) + \varphi\big)$ where $\varphi = \big(\frac{r-\mu}{\sigma}\big)\sqrt{T}$, with T denoting the maturity.

1.4 Distortion operators for financial models

Next, we derive distortion operators compatible with arbitrage-free pricing under non-Gaussian extensions of the Black-Scholes model. Empirically, it turns out that such extensions are needed to reproduce well-documented phenomena, such as the “volatility smiles" observed in option markets, and the fact that asset returns exhibit heavier-skewed tails than the normal distribution underpinning the geometric Brownian motion.

We consider a continuous-time market, with time $t \in [0, T]$, containing a liquid asset growing at the risk-free rate r and a (possibly non-traded) underlying risky process $\{S_t\}_{t\in[0,T]}$ defined on a probability space $(\Omega, \mathcal{F}_T, \mathbb{P})$ equipped with a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ satisfying the usual assumptions. In Section 1.4.1, we model the underlying asset with an infinite activity Lévy process, and in Section 1.4.2 with a jump-diffusion process. Both approaches, now standard in the literature, have been widely applied in finance (see, e.g., Schoutens, 2003; Cont and Tankov, 2004).

1.4.1 Normal Inverse Gaussian distortion based on the Esscher martingale measure

Godin et al. (2012) propose a distortion operator that can recuperate the arbitrage-free prices for a non-Gaussian extension of the Black-Scholes model based on the Normal Inverse Gaussian (NIG) distribution and the mean-correcting martingale measure. Their distortion operator is proposed in the form of an educated guess in their Definition 2. Here, we take a different approach. We start directly from the general expression of the distortion operator (Definition 1.3.1) and simplify it using Proposition 1.3.1. The latter approach has the advantage of not requiring an *ansatz* for the correct form of the distortion operator. To make this exercise rewarding, we extend the work of Godin et al. (2012) to the Esscher martingale measure, which provides a new NIG-based distortion operator that benefits from the very same advantages.

Dynamics under the physical measure

Under the physical measure, the underlying asset price follows an exponential NIG Lévy process with parameters $(\alpha, \beta, \delta, \mu)$. Its terminal value takes the form (see, e.g., Schoutens, 2003)

$$S_T = S_0 e^{X_T}, \qquad X_T \sim \mathrm{NIG}(\alpha, \beta, \delta T, \mu T), \tag{1.4.1}$$

where $\mathrm{NIG}(\alpha, \beta, \delta, \mu)$ is the Normal Inverse Gaussian distribution as defined in Godin et al. (2012). The distribution is a generalization of the normal distribution that allows for skewness and excess kurtosis. The cumulative distribution function (CDF) and survival function of $\mathrm{NIG}(\alpha, \beta, \delta, \mu)$ are respectively denoted by $\mathbf{NIG}(x; \alpha, \beta, \delta, \mu)$ and $\overline{\mathbf{NIG}}(x; \alpha, \beta, \delta, \mu)$, $x \in \mathbb{R}$. We state below some useful remarks about this distribution.

Remark 1.4.1. If $X \sim \mathrm{NIG}(\alpha, \beta, \delta, \mu)$ we have that $Y = aX + b$, for $a > 0$ and $b \in \mathbb{R}$, is such that $Y \sim \mathrm{NIG}(\alpha/a, \beta/a, a\delta, a\mu + b)$.

Remark 1.4.2. The NIG distribution possesses the symmetry

$$\overline{\mathbf{NIG}}(x; \alpha, \beta, \delta, 0) = \mathbf{NIG}(-x; \alpha, -\beta, \delta, 0).$$

Dynamics under the risk-neutral measure

It turns out that the market described above is incomplete, and therefore there exists an infinite number of equivalent sigma-martingale measures (Eberlein and Jacod, 1997). We choose the Esscher martingale measure as it is a popular choice (e.g., Gerber and Shiu, 1994). Under this measure, the process S is an exponential NIG Lévy process with parameters $(\alpha, \beta + \theta, \delta, \mu)$.[6] Under the Esscher martingale measure $\mathbb{Q}$, the terminal value of the underlying is such that

$$S_T = S_0 e^{X_T^{\mathbb{Q}}}, \qquad X_T^{\mathbb{Q}} \sim \mathrm{NIG}\big(\alpha, \beta + \theta, \delta T, \mu T\big). \tag{1.4.2}$$

Derivation of the distortion operator

Using Remark 1.4.1, we can express (1.4.1) and (1.4.2) more compactly as

$$S_T = h(Z) \equiv S_0 \exp\Big\{\mu T + \sqrt{\delta/\alpha}\, Z\Big\}, \tag{1.4.3}$$

where the random variable Z is such that

$$Z \sim \begin{cases} \mathrm{NIG}\big(\sqrt{\alpha\delta}, \beta\sqrt{\delta/\alpha}, T\sqrt{\alpha\delta}, 0\big), & \text{under } \mathbb{P}, \\ \mathrm{NIG}\big(\sqrt{\alpha\delta}, (\beta+\theta)\sqrt{\delta/\alpha}, T\sqrt{\alpha\delta}, 0\big), & \text{under } \mathbb{Q}. \end{cases} \tag{1.4.4}$$

6. Moreover, if S is the price process of a traded asset, then θ must be determined so that the discounted price process $\{S_t e^{-rt}\}_{t\in[0,T]}$ is a martingale under $\mathbb{Q}$ (see Schoutens, 2003, p. 79).

To obtain the expression of the distortion operator, we start from Definition 1.3.1 and simplify it using Proposition 1.3.1 and the symmetry property of Remark 1.4.2 :

$$g_{S_T}^{\mathbb{Q},\mathbb{P}}(u) = g_{h(Z)}^{\mathbb{Q},\mathbb{P}}(u) = g_Z^{\mathbb{Q},\mathbb{P}}(u) = \Phi_{\mathbf{NIG}}^{\mathbb{Q}}\big(\Phi_{\mathbf{NIG}}^{-1}(u)\big), \tag{1.4.5}$$

where the following definitions are used :

$$\Phi_{\mathbf{NIG}}(x) \equiv \mathbf{NIG}\big(x; \xi, \zeta, T\xi, 0\big), \qquad \Phi_{\mathbf{NIG}}^{\mathbb{Q}}(x) \equiv \mathbf{NIG}\big(x; \xi, \zeta^{\mathbb{Q}}, T\xi, 0\big),$$

with $\xi \equiv \sqrt{\alpha\delta}$, $\zeta \equiv -\beta\sqrt{\delta/\alpha}$, $\zeta^{\mathbb{Q}} \equiv -(\beta+\theta)\sqrt{\delta/\alpha}$, and where $\Phi_{\mathbf{NIG}}^{-1}$ is the inverse of $\Phi_{\mathbf{NIG}}$.

By Proposition 1.3.2, it follows that this distortion recovers the arbitrage-free prices under the Esscher martingale measure. In fact, Black-Scholes-type formulas can be recuperated through (1.3.7) in a similar fashion as in Hamada and Sherris (2003). Note that the difference between our NIG distortion and the one proposed by Godin et al. (2012) lies in the choice of the equivalent martingale measure, the later being based on the mean-correcting measure.[7] Since our new Esscher-based NIG distortion exhibits the same improvements over the Wang transform as achieved by Godin et al. (2012), we refer to that work for a thorough account of these improvements.

1.4.2 A distortion operator based on Kou (2002)'s jump-diffusion model

Let us consider now the Kou (2002)'s jump-diffusion option pricing model. This model is based on the equilibrium framework of Section 1.2.1, where the utility function of the representative agent is assumed to be of the form $U(t,c) = e^{-\kappa t}\frac{c^\alpha}{\alpha}$ for $\alpha \in (0,1]$, and $U(t,c) = e^{-\kappa t}\ln c$ for $\alpha = 0$, with κ being a subjective discount factor. Kou's model offers an attractive tradeoff between reality and tractability. It is able to reproduce the leptokurtic feature of the return distribution and the "volatility smile" observed in option markets, yet is simple enough to produce analytical formulas for call/put options, interest rate derivatives, and a variety of path-dependent options. We first describe Kou's model, and then derive its associated distortion.

The endowment process is modelled by the following jump-diffusion process under $\mathbb{P}$:

$$\frac{d\delta_t}{\delta_{t^-}} = \mu_1 dt + \sigma_1 dW_t^{(1)} + d\left[\sum_{i=1}^{N_t}(V_i - 1)\right], \tag{1.4.6}$$

where $\mu_1 \in \mathbb{R}$ and $\sigma_1 > 0$ are constants, $\{W_t^{(1)}\}_{t\in[0,T]}$ is a Wiener process, $\{N_t\}_{t\in[0,T]}$ is a standard Poisson process with intensity $\lambda > 0$, $\{V_i\}_{i\geq 1}$ is a sequence of i.i.d. non-negative random variables, and all sources of randomness are independent.

7. Godin et al. (2012) choose the mean-correcting martingale measure rather than the Esscher martingale measure. Under this measure, S is an exponential NIG Lévy process with parameters $(\alpha, \beta, \delta, \mu+\theta)$. One can check that applying our approach indeed yields the NIG distortion proposed in their Definition 2 : $\Phi_{\mathbf{NIG}}(\Phi_{\mathbf{NIG}}^{-1}(u) + \theta T\sqrt{\alpha/\delta})$.

The underlying follows the jump-diffusion process

$$\frac{dS_t}{S_{t^-}} = \mu dt + \sigma dW_t + d\left[\sum_{i=1}^{N_t}\left(V_i^{\beta} - 1\right)\right], \tag{1.4.7}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, and $\beta \in \mathbb{R}$ are constants. The new Wiener process $\{W_t\}_{t\in[0,T]}$ has constant correlation $\rho \in [-1, 1]$ to $W^{(1)}$ and is independent from the other sources of randomness.

Dynamics under the physical measure

The distribution of the log-size jumps is modelled by a Asymmetric Double Exponential distribution ADE(η_1, η_2, p). The PDF of this distribution is

$$\mathbf{ade}(x; \eta_1, \eta_2, p) = p\eta_1 e^{-\eta_1 x}\, \mathbb{1}_{x\geq 0} + (1-p)\eta_2 e^{\eta_2 x}\, \mathbb{1}_{x<0}, \qquad x \in \mathbb{R}. \tag{1.4.8}$$

The parameter domain is $\eta_1 > 0$, $\eta_2 > 0$, $p \in [0, 1]$. As stated in Remark 1.4.3, the ADE distribution is closed under scaling.

Remark 1.4.3. If $X \sim \text{ADE}(\eta_1, \eta_2, p)$ then we have that $Y \equiv aX$ is such that

$$Y \sim \begin{cases} \text{ADE}(\eta_1/a, \eta_2/a, p), & a \geq 0, \\ \text{ADE}(\eta_2/|a|, \eta_1/|a|, 1-p), & a < 0. \end{cases}$$

Solving the stochastic differential equation (1.4.7) yields

$$S_T = S_0 \exp\left\{ \left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W_T + \beta \sum_{i=1}^{N_T} Y_i \right\}, \tag{1.4.9}$$

with $\{Y_i \equiv \ln V_i\}_{i\geq 1} \overset{\text{i.i.d.}}{\sim} \text{ADE}(\eta_1, \eta_2, p)$, $W_T \sim N(0, T)$, and $N_T \sim \text{Poisson}(\lambda T)$.

Dynamics under the risk-neutral measure

Theorem 1 of Kou (2002) describes the dynamics under the risk-neutral measure and the conditions for the existence of this measure. From this theorem, it can be shown that under $\mathbb{Q}$ the terminal value of the underlying is such that

$$S_T = S_0 \exp\left\{ \left(\mu - \frac{\sigma^2}{2} - \rho\sigma\sigma_1(1-\alpha)\right)T + \sigma W_T^{\mathbb{Q}} + \beta \sum_{i=1}^{N_T} Y_i \right\}, \tag{1.4.10}$$

with $\{Y_i\}_{i\geq 1} \overset{\text{i.i.d.}}{\sim} \text{ADE}\left(\eta_1^{\mathbb{Q}}, \eta_2^{\mathbb{Q}}, p^{\mathbb{Q}}\right)$, $W_T^{\mathbb{Q}} \sim N(0, T)$, and $N_T \sim \text{Poisson}(\lambda^{\mathbb{Q}} T)$, where

$$\eta_1^{\mathbb{Q}} = \eta_1 - \alpha + 1, \qquad \eta_2^{\mathbb{Q}} = \eta_2 + \alpha - 1, \qquad p^{\mathbb{Q}} = \frac{p}{\zeta}\frac{\eta_1}{\eta_1 - \alpha + 1}, \qquad \lambda^{\mathbb{Q}} = \zeta\lambda, \tag{1.4.11}$$

with $\zeta \equiv \frac{p\eta_1}{\eta_1 - \alpha + 1} + \frac{(1-p)\eta_2}{\eta_2 + \alpha - 1}$.

Derivation of the distortion operator

The survival function defined below will be useful in defining the distortion operator. We refer to Appendix B of Kou (2002) for results that ease its numerical implementation.

Definition 1.4.1. Let the following random variables be independent : $Z \sim N(\mu, \sigma^2)$ is a normal variable, $P \sim \text{Poisson}(\lambda)$ is a Poisson variable, and $\{Y_i\}_{i\geq 1} \overset{\text{i.i.d.}}{\sim} \text{ADE}(\eta_1, \eta_2, p)$. The CDF of the *Normal Compound Poisson distribution NCP*$(\lambda, \eta_1, \eta_2, p, \sigma, \mu)$ is defined as

$$\mathbf{NCP}\big(x; \lambda, \eta_1, \eta_2, p, \sigma, \mu\big) \equiv \Pr\left(Z + \sum_{i=1}^{P} Y_i \leq x\right), \qquad x \in \mathbb{R}. \tag{1.4.12}$$

The survival function is denoted by $\overline{\mathbf{NCP}}\big(x; \lambda, \eta_1, \eta_2, p, \sigma, \mu\big) = 1 - \mathbf{NCP}\big(x; \lambda, \eta_1, \eta_2, p, \sigma, \mu\big)$.

Using Remark 1.4.3, it is a straightforward exercise to show that the NCP distribution is closed under affine transformations, as stated below in Remark 1.4.4. Moreover, the CDF and survival function of the NCP distribution are related by the symmetry property stated in Remark 1.4.5.

Remark 1.4.4. If $X \sim \text{NCP}(\lambda, \eta_1, \eta_2, p, \sigma, \mu)$ then we have that $Y = aX + b$ is such that

$$Y \sim \begin{cases} \text{NCP}\big(\lambda, \eta_1/a, \eta_2/a, p, a\sigma, a\mu + b\big), & a \geq 0, \\ \text{NCP}\big(\lambda, \eta_2/|a|, \eta_1/|a|, 1-p, |a|\sigma, a\mu + b\big), & a < 0. \end{cases}$$

Remark 1.4.5. The NCP distribution possesses the following symmetry :

$$\overline{\mathbf{NCP}}(x; \lambda, \eta_1, \eta_2, p, \sigma, \mu) = \mathbf{NCP}(-x; \lambda, \eta_2, \eta_1, 1-p, \sigma, -\mu).$$

We can make use of Remark 1.4.4 to express (1.4.9) and (1.4.10) more compactly as

$$S_T = h(X) \equiv S_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}X\right\}, \tag{1.4.13}$$

where the random variable X is such that

$$X \sim \begin{cases} \text{NCP}\big(\lambda T, \tilde{\eta}_1, \tilde{\eta}_2, \tilde{p}, 1, 0\big), & \text{under } \mathbb{P}, \\ \text{NCP}\big(\lambda^{\mathbb{Q}} T, \tilde{\eta}_1^{\mathbb{Q}}, \tilde{\eta}_2^{\mathbb{Q}}, \tilde{p}^{\mathbb{Q}}, 1, -\rho\sigma_1(1-\alpha)\sqrt{T}\big), & \text{under } \mathbb{Q}, \end{cases} \tag{1.4.14}$$

where the tilde parameters are defined in terms of $\nu \equiv \sigma\sqrt{T}/|\beta|$ as follows :

$$\begin{array}{lllllll} (\text{if } \beta \geq 0) & \tilde{\eta}_1 = \eta_1\nu, & \tilde{\eta}_1^{\mathbb{Q}} = \eta_1^{\mathbb{Q}}\nu, & \tilde{\eta}_2 = \eta_2\nu, & \tilde{\eta}_2^{\mathbb{Q}} = \eta_2^{\mathbb{Q}}\nu, & \tilde{p} = p, & \tilde{p}^{\mathbb{Q}} = p^{\mathbb{Q}}, \\ \quad\text{or} & & & & & & \\ (\text{if } \beta < 0) & \tilde{\eta}_1 = \eta_2\nu, & \tilde{\eta}_1^{\mathbb{Q}} = \eta_2^{\mathbb{Q}}\nu, & \tilde{\eta}_2 = \eta_1\nu, & \tilde{\eta}_2^{\mathbb{Q}} = \eta_1^{\mathbb{Q}}\nu, & \tilde{p} = 1-p, & \tilde{p}^{\mathbb{Q}} = 1-p^{\mathbb{Q}}. \end{array} \tag{1.4.15}$$

To obtain the expression of the distortion operator, we start from Definition 1.3.1 and simplify it by using Proposition 1.3.1 and the symmetry property of Remark 1.4.5 :

$$g_{S_T}^{\mathbb{Q},\mathbb{P}}(u) = g_{h(X)}^{\mathbb{Q},\mathbb{P}}(u) = g_X^{\mathbb{Q},\mathbb{P}}(u) = \Phi_{\mathbf{NCP}}^{\mathbb{Q}}\Big(\Phi_{\mathbf{NCP}}^{-1}(u) - \rho\sigma_1(1-\alpha)\sqrt{T}\Big), \tag{1.4.16}$$

where the following definitions are used :

$$\Phi_{\mathbf{NCP}}(x) \equiv \mathbf{NCP}\big(x; \lambda T, \tilde{\eta}_2, \tilde{\eta}_1, 1-\tilde{p}, 1, 0\big), \qquad \Phi^{\mathbb{Q}}_{\mathbf{NCP}}(x) \equiv \mathbf{NCP}\big(x; \lambda^{\mathbb{Q}} T, \tilde{\eta}^{\mathbb{Q}}_2, \tilde{\eta}^{\mathbb{Q}}_1, 1-\tilde{p}^{\mathbb{Q}}, 1, 0\big),$$

with $\Phi^{-1}_{\mathbf{NCP}}$ defined as the inverse of $\Phi_{\mathbf{NCP}}$.

From Proposition 1.3.2, it follows that the distortion operator (1.4.16) recovers the pricing equilibrium described above. For instance, through (1.3.7), the distortion can recuperate the European call formula (20) in Kou (2002). This distortion operator produces a risk-neutralized distribution that can be viewed as capturing premiums for both sources of risks, i.e., jumps and Brownian motion.

1.5 Distortion operators for insurance models

We now derive distortion operators for two insurance pricing models. The first is based on the size-biased pricing principle for the generalized beta of the second kind distribution, and the second is based on the Esscher principle for the gamma distribution. In both cases, the underlying risk is an insurance claim represented by a positive random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$.

We adopt the conventions used in Klugman et al. (2012) for the incomplete beta function and the incomplete gamma function. The incomplete beta function is defined as

$$\beta(x; \tau, \alpha) \equiv \frac{\Gamma(\tau+\alpha)}{\Gamma(\tau)\Gamma(\alpha)} \int_0^x t^{\tau-1}(1-t)^{\alpha-1} dt, \qquad \tau > 0, \quad \alpha > 0, \quad x \in (0,1), \tag{1.5.1}$$

and the (lower) incomplete gamma function is defined as

$$\Gamma(x; \alpha) \equiv \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt, \qquad \alpha > 0, \quad x > 0, \tag{1.5.2}$$

where $\Gamma(\alpha) \equiv \int_0^\infty t^{\alpha-1} e^{-t} dt$ is the gamma function.

1.5.1 A distortion based on the generalized beta of the second kind distribution

The generalized beta of the second kind (GB2) distribution, sometimes called the transformed beta distribution, is a member of the celebrated Pearson system and was first proposed as a model of the size-of-loss distribution in actuarial sciences by Venter (1983). This large family of heavy-tailed distributions contains the Burr, generalized Pareto, generalized gamma, log-t, and other commonly used distributions. It provides a fairly flexible form that can be used to model highly skewed loss distributions such as those typically observed in non-life insurance (Cummins et al., 1990). Here, we use the GB2 distribution to model the risk X. A thorough account of other size-of-loss distributions can be found in Kleiber and Kotz (2003) and in Klugman et al. (2012).

The GB2 distribution has heavy tails, and therefore only a few of the moments exist (Kleiber, 1997). This implies that some risk-neutral measures, such as the Esscher measure, may exist only for a certain range of the shape parameters. Among the possible risk-neutral measures, the size-biased subclass is an interesting choice as it preserves the shape of the GB2 family, yielding a simple Wang-like distortion operator. We underline that our approach also applies to non shape-preserving changes of measure, although it may require tedious algebra to simplify the form of the distortion function.

Distribution under the physical measure

Under the physical measure $\mathbb{P}$, the risk X follows a GB2$(\alpha, \theta, \gamma, \tau)$ distribution as defined in Klugman et al. (2012). The PDF of this distribution is

$$\mathbf{gb2}(x; \alpha, \theta, \gamma, \tau) = \frac{\Gamma(\alpha+\tau)}{\Gamma(\alpha)\Gamma(\tau)} \frac{\gamma (x/\theta)^{\gamma\tau}}{x[1+(x/\theta)^{\gamma}]^{\alpha+\tau}}, \qquad x \geq 0, \tag{1.5.3}$$

and its CDF is

$$\mathbf{GB2}(x; \alpha, \theta, \gamma, \tau) = \beta\biggl(\frac{(x/\theta)^{\gamma}}{1+(x/\theta)^{\gamma}}; \tau, \alpha\biggr), \qquad x \geq 0. \tag{1.5.4}$$

The parameter domain of this distribution is $\alpha > 0$, $\theta > 0$, $\gamma > 0$, $\tau > 0$. The parameter controlling the scale is θ. The other parameters control the tail behaviour and the shape in general. It is interesting to note that the log-normal distribution is a limiting case of the GB2 distribution, as stated in the following remark.

Remark 1.5.1 (Log-normal limit)**.** A log-normal distribution with PDF $\mathbf{logN}(x; \mu, \sigma) \equiv \frac{\phi\left(\frac{\ln x - \mu}{\sigma}\right)}{\sigma x}$, $x \geq 0$, is obtained as a limiting case of the GB2$(\alpha, \theta, \gamma, \tau)$ distribution for $\alpha \to \infty$, $\gamma \to 0$, $\theta = (\alpha\gamma^2\sigma^2)^{1/\gamma}$ and $\tau = (\gamma\mu + 1)/(\sigma^2\gamma^2)$ (from McDonald, 1987).

The survival function under $\mathbb{P}$ of the risk X is thus given by

$$\bar{F}_{\mathbb{P}}(x) \equiv \mathbb{P}(X > x) = 1 - \beta\biggl(\frac{(x/\theta)^{\gamma}}{1+(x/\theta)^{\gamma}}; \tau, \alpha\biggr), \qquad x \geq 0. \tag{1.5.5}$$

Distribution under the risk-neutral measure

Under the size-biased risk-neutral measure $\mathbb{Q}$, the PDF of X is given by

$$\frac{x^a}{\mathbb{E}^{\mathbb{P}}[X^a]} \mathbf{gb2}(x; \alpha, \theta, \gamma, \tau) \propto \mathbf{gb2}(x; \tilde{\alpha}, \theta, \gamma, \tilde{\tau}), \tag{1.5.6}$$

where

$$\tilde{\alpha} \equiv \alpha - a/\gamma, \qquad \tilde{\tau} \equiv \tau + a/\gamma. \tag{1.5.7}$$

By normalization, the right-hand side of (1.5.6) implies that $X \sim \text{GB2}(\tilde{\alpha}, \theta, \gamma, \tilde{\tau})$ under $\mathbb{Q}$. Note that the requirement $-\gamma\tau < a < \alpha\gamma$ is needed for the existence of this measure.

The survival function under $\mathbb{Q}$ of X is therefore given by

$$\bar{F}_{\mathbb{Q}}(x) \equiv \mathbb{Q}(X > x) = 1 - \beta\left(\frac{(x/\theta)^{\gamma}}{1 + (x/\theta)^{\gamma}}; \tilde{\tau}, \tilde{\alpha}\right), \qquad x \geq 0. \tag{1.5.8}$$

Derivation of the distortion operator

To obtain the distortion operator, the first step is to use (1.5.8) in Definition 1.3.1 to obtain

$$g_X^{\mathbb{Q},\mathbb{P}}(u) \equiv \bar{F}_{\mathbb{Q}}\big(\bar{F}_{\mathbb{P}}^{-1}(u)\big) = 1 - \beta\left(\frac{\big(\bar{F}_{\mathbb{P}}^{-1}(u)/\theta\big)^{\gamma}}{1 + \big(\bar{F}_{\mathbb{P}}^{-1}(u)/\theta\big)^{\gamma}}; \tilde{\tau}, \tilde{\alpha}\right). \tag{1.5.9}$$

Next, we use $x = \bar{F}_{\mathbb{P}}^{-1}(u)$ in (1.5.5) to obtain the following relation :

$$u = 1 - \beta\left(\frac{\big(\bar{F}_{\mathbb{P}}^{-1}(u)/\theta\big)^{\gamma}}{1 + \big(\bar{F}_{\mathbb{P}}^{-1}(u)/\theta\big)^{\gamma}}; \tau, \alpha\right) \quad \Rightarrow \quad \frac{\big(\bar{F}_{\mathbb{P}}^{-1}(u)/\theta\big)^{\gamma}}{1 + \big(\bar{F}_{\mathbb{P}}^{-1}(u)/\theta\big)^{\gamma}} = \beta^{-1}\big(1 - u; \tau, \alpha\big),$$

where the expression of the right-hand side is obtained by rearranging the terms and taking the inverse incomplete beta function β^{-1}. Using this expression in (1.5.9) gives us the GB2 distortion :

$$g_X^{\mathbb{Q},\mathbb{P}}(u) = 1 - \beta\big(\beta^{-1}(1 - u; \tau, \alpha); \tilde{\tau}, \tilde{\alpha}\big), \tag{1.5.10}$$

where $\tilde{\tau} \equiv \tau + a/\gamma$ and $\tilde{\alpha} \equiv \alpha - a/\gamma$.

An interesting special case of the GB2 distortion is obtained for $\tau = \alpha = 1$, for which it is a straightforward exercice to prove that we obtain $g_X^{\mathbb{Q},\mathbb{P}}(u) = 1 - \beta\big(1 - u; 1 + a/\gamma, 1 - a/\gamma\big)$. This distortion is a variant of the beta transform in Wirch and Hardy (1999). In this paper, the authors argue that the corresponding distortion risk measure has advantages over the expected shortfall because it utilizes the whole distribution rather than focusing only the tail.

An even more interesting property of the GB2 distortion is that it reduces to the Wang transform $g_X^{\mathbb{Q},\mathbb{P}}(u) = \Phi\big(\Phi^{-1}(u) + a\sigma\big)$ in the log-normal limiting case stated in Remark 1.5.1.[8] The GB2 distortion can therefore be seen as a very flexible generalization of the Wang transform that can account for heavier tails. It is a well-known fact that correctly modelling the tail behaviour is crucially important for premium calculation and risk measurement. Indeed, the Wang transform must be modified to capture the heavy tail feature.

8. Let X be a log-normal variable distributed with density $\mathbf{logN}(x; \mu, \sigma) \equiv \phi\big(\frac{\ln x - \mu}{\sigma}\big)/(\sigma x)$, $x \geq 0$. It can be shown that $\frac{x^a}{\mathbb{E}^{\mathbb{P}}[X^a]}\mathbf{logN}(x; \mu, \sigma) = \mathbf{logN}(x; \mu + a\sigma^2, \sigma)$. The distortion operator, obtained in a similar fashion as in Section 1.5.1, turns out to be the Wang transform $g_X^{\mathbb{Q},\mathbb{P}}(u) = \Phi\big(\Phi^{-1}(u) + a\sigma\big)$.

1.5.2 An extension of the proportional hazards transform

We now turn to the Esscher pricing principle. Because this principle relies on an exponential weight function, the risk-neutral measure does not exist for the heavy-tailed distributions in the GB2 family. It is however applicable to the gamma subfamily, which includes the exponential and chi-squared distributions. Here, we suppose that the insurance claim X is gamma distributed, and we derive the distortion that recuperates the Esscher principle. Interestingly, it turns out that the proportional hazards (PH) transform (Wang, 1995) is a special case of this distortion.

Distribution under the physical measure

Under the physical measure $\mathbb{P}$, the risk X follows a Gamma(α, θ) distribution. The parameter domain of this distribution is $\alpha > 0$, $\theta > 0$. The PDF and CDF are given by

$$\mathbf{gamma}(x; \alpha, \theta) = \frac{(x/\theta)^{\alpha} e^{-x/\theta}}{x\Gamma(\alpha)}, \qquad \mathbf{Gamma}(x; \alpha, \theta) = \Gamma\big(x/\theta; \alpha\big), \qquad x \geq 0. \tag{1.5.11}$$

The survival function of X, under $\mathbb{P}$, is thus $\bar{F}_{\mathbb{P}}(x) = 1 - \Gamma\big(x/\theta; \alpha\big)$, $x \geq 0$.

Distribution under the risk-neutral measure

Under the Esscher risk-neutral measure $\mathbb{Q}$, the PDF of X is given by

$$\frac{e^{ax}}{\mathbb{E}^{\mathbb{P}}[e^{aX}]} \mathbf{gamma}(x; \alpha, \theta) \propto \mathbf{gamma}(x; \alpha, \tilde{\theta}), \qquad \tilde{\theta} \equiv \frac{\theta}{1 - a\theta}. \tag{1.5.12}$$

By normalization, the right-hand side indeed implies that $X \sim$ Gamma$(\alpha, \tilde{\theta})$ under $\mathbb{Q}$. Note that $a < 1/\theta$ is required to ensure the existence of this measure.

The survival function of X, under $\mathbb{Q}$, is thus $\bar{F}_{\mathbb{Q}}(x) = 1 - \Gamma\big(x/\tilde{\theta}; \alpha\big)$, $x \geq 0$.

Derivation of the distortion operator

The derivation of the distortion operator follows the same steps as in Section 1.5.1. We obtain

$$g_X^{\mathbb{Q},\mathbb{P}}(u) = 1 - \Gamma\big(\Gamma^{-1}(1 - u; \alpha)[1 - a\theta]; \alpha\big), \tag{1.5.13}$$

where Γ^{-1} is the inverse of the incomplete gamma function Γ of (1.5.2).

Note that for the special case $\alpha = 1$ (i.e., the exponential distribution) we obtain $g_X^{\mathbb{Q},\mathbb{P}}(u) = u^{1-a\theta}$, which is the so-called PH transform (Wang, 1995). Our gamma-based distortion can therefore be seen as a natural generalization of the PH transform obtained from a more flexible distribution. Moreover, our analysis clearly deepens and extends the discussions in Wang (1996) regarding the connections between risk-neutral valuation and the PH transform.

1.6 Empirical applications to CAT bonds

Catastrophe (CAT) bonds are insurance-linked securities devised by insurers and reinsurers to shift natural disaster risks to the capital markets. [9] Several authors have proposed to price CAT bonds using the theory of contingent claim pricing for jump-diffusion processes (e.g., Vaugirard, 2003). These CAT bond pricing models usually involve extensive information that are not publicly disclosed (e.g., the trigger level or strike price). [10] It ensues that these types of models have yet to be tested empirically. In fact, most empirical studies on CAT bond spreads have instead relied on pure regression models (e.g., Bodoff and Gan, 2012; Braun, 2015). Meanwhile, explaining CAT bond spreads using distortion operators has been suggested by several works ; calibration of the Wang transform to CAT bond spreads is carried out, for instance, in Wang (2004) and in Galeotti et al. (2013). In this section, we employ our approach to perform further empirical tests on such models as it enables us to reformulate these in terms of the information available to the econometrician (e.g., the expected loss, the probability of first loss, the probability of last loss). Extending this stream of empirical studies, we make several interesting findings, in particular, it turns out that incorporating CAT risk and CAT risk-aversion into the distortion is important for explaining CAT bond spreads. This is established by calibrating a jump-diffusion distortion operator to market spreads. By way of a mixture of normal and exponential distributions, we also offer a simplified approximation of the latter distortion.

1.6.1 Setup

Let the payout to the ceding (re)insurer be contingent upon a risk X breaching a pre-agreed attachment point b, in which case the collateral is liquidated to reimburse the sponsor up to the par amount p paid by the investor at the issue date. If there is no triggering event during the term of the CAT bond, which is typically between one and four years, the principal is returned to the investor plus a coupon payment with spread S above the risk-free rate. Under the distortion premium principle, the spread is given by (from Galeotti et al., 2013)

$$\mathrm{S} = \frac{1}{p}\int_b^{b+p} g\big(\bar{F}_{\mathbb{P}}(x)\big)dx, \qquad \bar{F}_{\mathbb{P}}(x) \equiv \mathbb{P}(X > x). \tag{1.6.1}$$

The integral can be approximated using the trapezoidal rule :

$$\mathrm{S} \approx \big(g(\mathrm{PFL}) + g(\mathrm{PLL})\big)/2, \qquad \mathrm{PFL} \equiv \bar{F}_{\mathbb{P}}(b), \quad \mathrm{PLL} \equiv \bar{F}_{\mathbb{P}}(b+p). \tag{1.6.2}$$

9. CAT bond transactions usually involve a Special-Purpose Vehicle (SPV), located in a tax-efficient jurisdiction, that sells catastrophe protection to a ceding (re)insurer in the form of a reinsurance contract. The SPV then effectively transfers its risk exposure by issuing CAT bond tranches to capital market investors. In order to offer a virtually pure exposure to the natural disaster risk, the proceeds of the issuance are invested by the SPV in highly rated short-term assets that are held in a collateral account.

10. Available databases for CAT bond transactions include Artemis.bm's deal directory, and the reports published by Aon Benfield, Swiss Re, Plenum, and Lane Financial LLC.

If either of the probability of first loss PFL or the probability of last loss PLL is not available, the integral can be approximated using the rectangle method :

$$\mathrm{S} \approx g(\mathrm{EL}), \qquad \mathrm{EL} \equiv \frac{1}{p}\int_b^{b+p} \bar{F}_{\mathbb{P}}(x)dx. \tag{1.6.3}$$

A data set containing 508 CAT bond tranches issued between 1997 and 2016 has been made available to us by Artemis.bm. Roughly half of the CAT bonds in the sample cover the United States, approximatively one quarter are multi-territory, and the rest cover Europe, Japan or other areas. About half of the CAT tranches are multi-peril, and others are mainly wind-specific or earthquake-specific. For each of the 508 transactions, the available information includes the per annum spread S and expected loss EL. The probability of first loss PFL and the probability of last loss PLL are however available only for 284 transactions. Calibration is therefore carried out under the rectangle method (1.6.3), by minimizing the sum of squared pricing errors. Fitting adequacy is investigated using nonparametric regressions to detect systematic pricing errors ; see Azzalini et al. (1989), Zheng (1996), and Li and Racine (2007) for works on nonparametric specification testing.

1.6.2 Pricing CAT bonds using a jump-diffusion distortion

An interesting property of our distortion operators is that they inherit the features of the original root model they are derived from. For instance, the jump-diffusion distortion (1.4.16) can incorporate jump risk, making it an interesting candidate for pricing CAT bonds. Several frameworks based on jump-diffusion processes have in fact been proposed to price such insurance-linked securities (e.g., Vaugirard, 2003). In this section, instead of using the Wang transform as in Wang (2004) and Galeotti et al. (2013), we calibrate our jump-diffusion distortion to CAT bond spreads. Furthermore, we investigate whether jump risk is priced by the market.

Suppose the jumps are used to model natural disasters. Because such events can only lower the aggregate endowment, we are interested in the special case $p = 0$ and $\beta < 0$ of the model presented in Section 1.4.2, where the jump-diffusion process S is used here to represent a catastrophe loss index whose dynamics is affected by positive jumps. For this special case, it is a straightforward exercise to show that the distortion operator (1.4.16) can be simplified and rewritten as follows :

$$g_{\varphi,\lambda,\eta,\nu}(u) \equiv \Upsilon_{\lambda\nu,\frac{\eta}{\nu}}\left(\Upsilon^{-1}_{\lambda,\eta}(u) + \varphi\right), \tag{1.6.4}$$

where $\lambda > 0$, $\eta > 0$, $\nu \geq 1$, $\varphi \in \mathbb{R}$, and $\Upsilon^{-1}_{\lambda,\eta}$ is the inverse of the function defined below.[11]

Definition 1.6.1. Let $\{Y_i\}_{i\geq 1}$ be a sequence of i.i.d. exponential random variables with rate $\eta > 0$, P be a Poisson random variable with rate $\lambda > 0$, and Z be a standard normal random

11. Note that the parameters λ, η, ν, and φ are defined differently than in preceding sections ; the latter definition corresponds to the simplified econometric specification.

variable. We define the following CDF :

$$\Upsilon_{\lambda,\eta}(x) \equiv \Pr\left(Z - \sum_{i=1}^{P} Y_i \leq x\right), \qquad x \in \mathbb{R}. \tag{1.6.5}$$

We calibrate the distortion depicted by (1.6.4) using our CAT bond tranches data set. It turns out that there is an infinite set of solutions that yield the same fitted curve. For example, the solution $(\hat{\varphi}, \hat{\lambda}, \hat{\eta}, \hat{\nu}) = (0.02, 0.20, 4.50, 2.92)$ provides the same distortion as (0.11, 0.23, 2.00, 1.87). We are also interested in determining whether jump risk is priced by the market. We seek an answer to this research question by calibrating the distortion operator under the constraint $\nu = 1$, which states that the jump amplitude and frequency are unchanged under the risk-neutral measure, i.e., that jump risk is not priced by the market. The calibration results are exhibited in Figure 1.1, where we can see that assuming idiosyncratic (i.e., unpriced) jump risk leads to a mis-specified model. This provides, for the first time, interesting evidence that jump-diffusion models are appropriate for pricing CAT bonds, but that investors are risk-averse to natural disasters. We defer to future work for delving further into this question.

1.6.3 Results for other distortions

One potential disadvantage of the jump-diffusion distortion is the use of sophisticated analytical results on the Hh special function from mathematical physics in order to compute it efficiently (see, e.g., Kou, 2002, Appendix B). Hence, it would be interesting to find a readily implementable distortion that also fits well to CAT bond spreads.

Results for calibrating the GB2 distortion (1.5.10) to CAT bond market spreads are exhibited in Figure 1.2. Graph 1.2(b) reveals systematic pricing errors, exacerbated at small values of the expected loss EL. Unreported tests available upon request from the authors show that other distortion operators suffer from similar issues ; such as the two NIG distortions and also Wang (2004)'s distortion based on the Student-t distribution. Hence, such distortions are less suitable for explaining CAT bond spreads. This might appear surprising as the GB2 family is known for its flexibility in modeling insurance losses (Cummins et al., 1990). On the other hand, the theoretical literature on CAT bonds and other insurance-linked securities advocate jump-diffusion processes and compound Poisson processes to model catastrophe loss processes.[12] In the previous sections, we have presented empirical evidence supporting the latter models. The inadequacy of the above distortion functions for explaining CAT bond spreads may therefore be attributed to their inability in capturing and pricing catastrophe risks, in contrast with the jump-diffusion distortion which performs well. Next, we present a general framework addressing this issue.

12. See Vaugirard (2003), Nowak and Romaniuk (2013), Ma and Ma (2013), Perrakis and Boloorforoosh (2013) and Lai et al. (2014) for a non-exhaustive list of such works.

1.6.4 A simple class of distortion operators to price CAT bonds

This section offers a simple class of Esscher-type distortions based on a mixture of distributions to characterize catastrophe risks. It is shown such distortions can provide an accurate representation of CAT bond spreads while being straightforward to implement in practice.

General framework

Let X be a continuous random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ representing a risk. A probability mixture distribution having the following representation is considered :

$$X = e^{\mu + \sigma Z_I}, \tag{1.6.6}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are constants, I is a discrete random variable such that

$$\mathbb{P}(I = i) = p_i, \qquad i \in \{1, \dots, m\}, \qquad \sum_{i=1}^{m} p_i = 1, \tag{1.6.7}$$

and the $Z_1, \dots, Z_m$ are independent random variables with PDFs denoted by $f_1, \dots, f_m$. Hence,

$$\mathbb{P}(Z_I \leq z) = \sum_{i=1}^{m} p_i F_i(z), \qquad F_i(z) \equiv \int_{-\infty}^{z} f_i(x) dx, \qquad z \in \mathbb{R}, \tag{1.6.8}$$

as I is presumed independent from the $Z_1, \dots, Z_m$. Here, $m \geq 1$ is a given constant denoting the number of components in the probability mixture. Such a model can account for a hidden multi-state risk structure and is therefore deemed appropriate for catastrophe-linked risks.

The following notation will be used for the moment generating functions :

$$\zeta_i^{(t)} \equiv \mathbb{E}^{\mathbb{P}}\left[e^{tZ_i}\right], \qquad i \in \{1, \dots, m\}, \quad t \in \mathbb{R}. \tag{1.6.9}$$

The Esscher risk-neutral measure $\mathbb{Q}$ is considered as it is a common choice in the literature (e.g., Gerber and Shiu, 1994). The Radon-Nikodym derivative associated to this measure is

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{X^a}{\mathbb{E}^{\mathbb{P}}[X^a]} = \frac{e^{a\sigma Z_I}}{\mathbb{E}^{\mathbb{P}}[e^{a\sigma Z_I}]}, \tag{1.6.10}$$

where $a \in \mathbb{R}$ is a constant. The distribution of Z_I under $\mathbb{Q}$ is stated in the following proposition proven in Appendix 1.A.5.

Proposition 1.6.1. *We have*

$$\mathbb{Q}(Z_I \leq z) = \sum_{i=1}^{m} \tilde{p}_i \tilde{F}_i(z), \qquad z \in \mathbb{R}, \tag{1.6.11}$$

where, for all $i \in \{1, \dots, m\}$,

$$\tilde{p}_i \equiv \frac{p_i \zeta_i^{(a\sigma)}}{\sum_{j=1}^{m} p_j \zeta_j^{(a\sigma)}}, \qquad \tilde{F}_i(z) \equiv \int_{-\infty}^{z} \tilde{f}_i(x) dx, \qquad \tilde{f}_i(x) \equiv \frac{e^{a\sigma x}}{\zeta_i^{(a\sigma)}} f_i(x). \tag{1.6.12}$$

The distortion operator g that applies the Esscher change of measure for the risk X, i.e., such that $g\big(\mathbb{P}(X > x)\big) = \mathbb{Q}(X > x)$ for all $x \in \mathbb{R}$, is given in the proposition below whose proof is a direct application of Corollary 1.3.1.

Proposition 1.6.2. *The distortion that performs the Esscher change of measure for X is*

$$g_{Z_I}^{\mathbb{Q},\mathbb{P}}(u) \equiv \bar{F}_{\mathbb{Q}}\big(\bar{F}_{\mathbb{P}}^{-1}(u)\big), \qquad u \in [0,1], \tag{1.6.13}$$

where $\bar{F}_{\mathbb{Q}}(z) \equiv 1 - \mathbb{Q}(Z_I \leq z)$, and $\bar{F}_{\mathbb{P}}^{-1}$ is the inverse of the function $\bar{F}_{\mathbb{P}}(z) \equiv 1 - \mathbb{P}(Z_I \leq z)$.

Recuperating distortions from previous literature

The above framework is very general as it encompasses several distortion operators encountered in the previous literature. Some examples are given below for the case $m = 1$; detailed proofs are available from the authors upon request.

1. The PH transform (from Wang, 1995) $g_{Z_1}^{\mathbb{Q},\mathbb{P}}(u) = u^{1-a\sigma}$ is obtained if Z_1 is a standard exponential variable under $\mathbb{P}$.
2. The Wang transform $g_{Z_1}^{\mathbb{Q},\mathbb{P}}(u) = \Phi\big(\Phi^{-1}(u) + a\sigma\big)$ is obtained if $Z_1 \sim N(0,1)$ under $\mathbb{P}$.
3. The GB2 distortion $g_{Z_1}^{\mathbb{Q},\mathbb{P}}(u) = 1 - \beta\big(\beta^{-1}(1-u;\tau,\alpha);\tau + a\sigma, \alpha - a\sigma\big)$ is obtained if e^{Z_1} follows a GB2$(\alpha, 1, 1, \tau)$ distribution (as defined in Klugman et al., 2012) under $\mathbb{P}$.

The multi-state extensions of the above distortions are also directly derived from the above general framework.

A normal-exponential mixture distortion

To approximate the jump-diffusion distortion (1.6.4), we now consider a special case. The mixture consists of $m = 2$ states such that $Z_1 \sim N(0,1)$ and $Z_2 \sim \text{Exp}(\eta)$ under $\mathbb{P}$, where $\text{Exp}(\eta)$ is the exponential distribution with inverse scale parameter $\eta > 0$.

It follows that

$$\mathbb{P}(Z_I \leq z) = p_1 \Phi(z) + (1-p_1)(1 - e^{-\eta z})\mathbb{1}_{\{z \geq 0\}}, \qquad z \in \mathbb{R}. \tag{1.6.14}$$

As shown in Appendix 1.A.6, provided that $\eta > a\sigma$,

$$\begin{aligned} \mathbb{Q}(Z_I \leq z) &= \tilde{p}_1 \Phi(z - a\sigma) + (1-\tilde{p}_1)\big(1 - e^{-(\eta - a\sigma)z}\big)\mathbb{1}_{\{z \geq 0\}}, \qquad z \in \mathbb{R}, \\ \tilde{p}_1 &= \frac{p_1 e^{\frac{1}{2}(a\sigma)^2}}{p_1 e^{\frac{1}{2}(a\sigma)^2} + (1-p_1)\frac{\eta}{\eta - a\sigma}}. \end{aligned} \tag{1.6.15}$$

The corresponding distortion function is given by Proposition 1.6.2. It is characterized by the parameters (p_1, η, ξ) where $\xi \equiv a\sigma$, with domain $p_1 \in [0,1]$, $\eta > 0$, $\xi < \eta$. The computation of this distortion is quite straightforward as it involves only basic functions. Moreover, it provides a more accurate explanation of CAT bond spreads as illustrated in Figure 1.2. In particular, residuals in Figure 1.2(c) do not display systematic pricing errors in contrast with the previous distortions. Hence, our mixed distribution approach is successful in providing a more accurate description of observed CAT bond spreads while being simple to implement in practice.

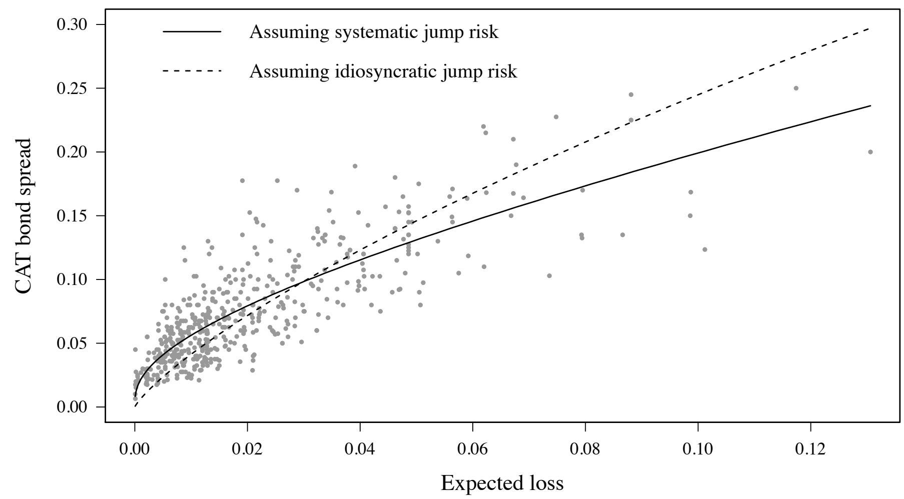
1.7 Conclusion

We propose a general class of probability distortion operators consistent with arbitrage-free pricing. Previous attempts in this direction are the Wang (2000) transform and the NIG distortion of Godin et al. (2012). To illustrate our approach, we derive several new distortions that improve upon the original Wang transform in a similar fashion as the more recent NIG distortion, while generalizing the latter by bringing flexibility in choosing the equivalent martingale measure and the underlying distribution. In fact, our framework makes it a straightforward exercise to derive new distortion operators, for instance from standard non-Gaussian financial theory (see, e.g., Schoutens, 2003). Our research also provides new twists to existing works that investigate the connections between distortion risk measures and other forms of risk pricing (e.g., Wang, 2003; Labuschagne and Offwood, 2010). We effectively characterize the change of measure performed by our distortion operators and show that this offers a deeper understanding of such connections.

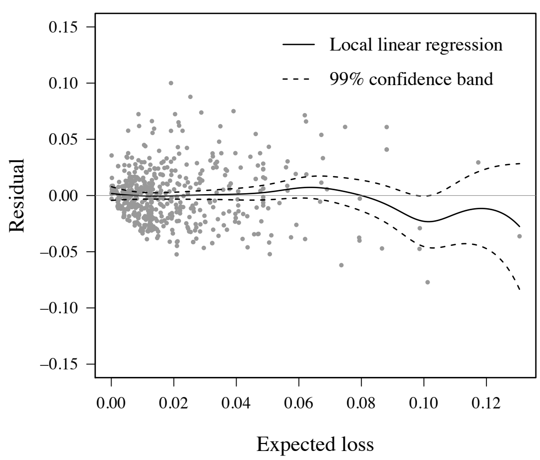
An important area of research opened by our work is the testing of catastrophe (CAT) bond pricing models from publicly available transaction information. Following the lead of Wang (2004) and Galeotti et al. (2013), we provide a explanatory empirical study which indicates that an exponential jump-diffusion distortion is adequate for explaining CAT bond spreads, but only if we allow the distortion to incorporate risk-aversion to natural disasters. Also, a general yet simple class of probability distortion operators based on the Esscher change of measure for multi-state structured risks is proposed to price CAT risks. This new class of distortions provides an accurate depiction of observed CAT bond spreads while being straightforward to implement in practice.

Another potential application of our distortion operator is to produce, by way of equation (1.2.3), distortion-based risk measures that can be used for a wide range of applications, including capital allocation and optimal reinsurance (see Dowd and Blake, 2006, for an account on the applications of distortion risk measures). We also contribute to the discussion in the literature regarding the connection between risk measures, heavy-tailed skewed distributions, and risk pricing. Indeed, our new distortion operator expressed by equation (1.3.1) is directly defined in terms of the physical distribution of the underlying risk and the choice of the equivalent martingale measure. As such, it provides a parametric approach to produce risk measures that incorporate stylized features of financial (or insurance) risk distributions (e.g., skewness and kurtosis) as well as risk-aversion and other determinants of market prices of risks. In fact, there have been several attempts to incorporate these features into risk measures (e.g., Bali and Theodossiou, 2008; Gzyl and Mayoral, 2008). In particular, it is well-known that accounting for the heavy tail feature of the loss distributions observed in non-life insurance is critically important for premium calculation and risk measurement. Our new distortion operator based on the generalized beta of the second kind distribution is a flexible generalization of the Wang transform that can capture the heavy tail feature.

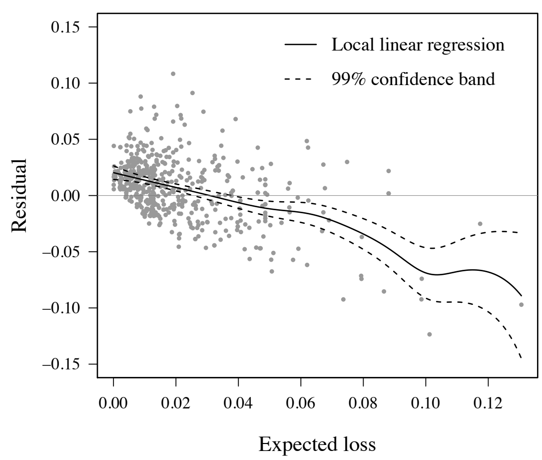
Regarding extensions, it has been shown that the Wang transform can be used to obtain the Black-Scholes prices of exotic options (Labuschagne and Offwood, 2013). The Wang transform has also been extended to a multivariate setting (see, e.g., Kijima, 2006; Wang, 2007). It seems reasonable to presume that our framework can be extended in similar directions. We leave these questions open for future research.



(a) Illustration of the calibrated distortion operators

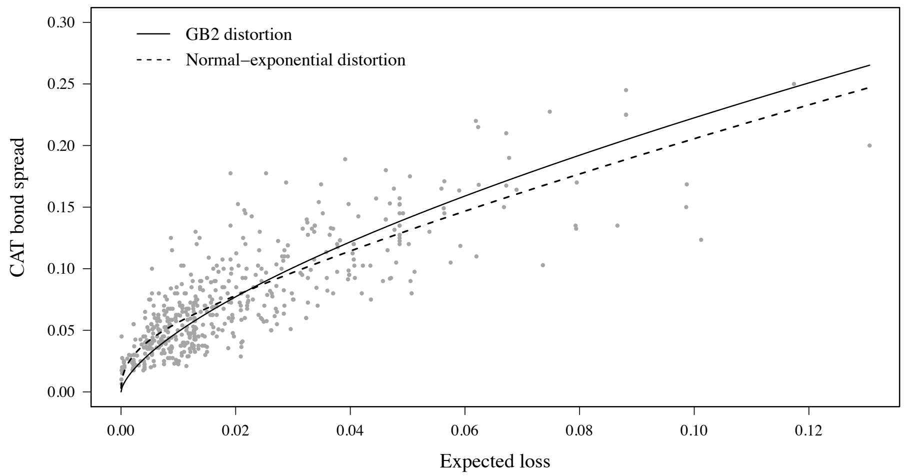


(b) Residuals : Systematic jump risk ($\nu \geq 1$)

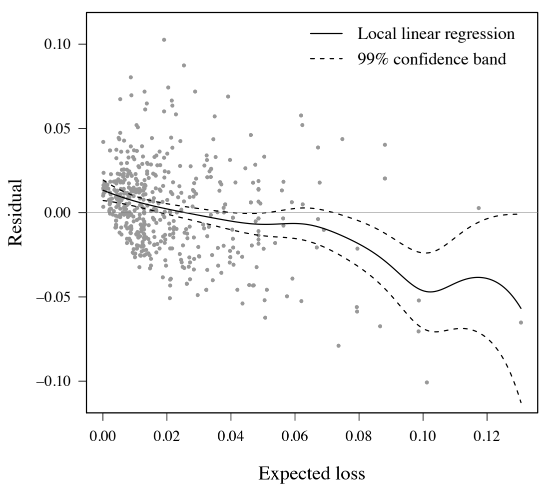


(c) Residuals : Idiosyncratic jump risk ($\nu = 1$)

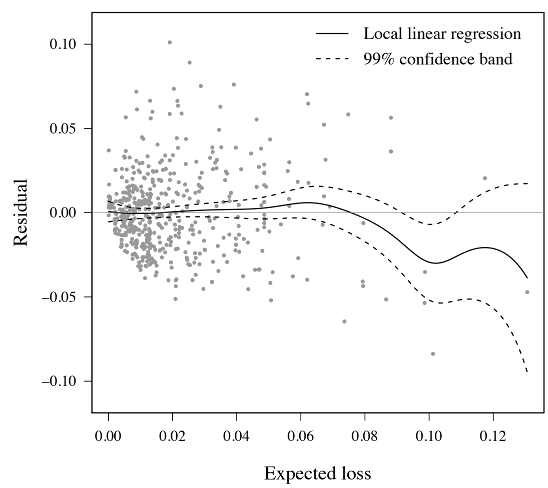
FIGURE 1.1 – Calibration results for our Artemis.bm data set consisting of 508 CAT bond tranches issued between 1997 and 2016. The calibration is done by minimizing the sum of squared pricing errors. Graph **(a)** illustrates the empirical data and the calibrated distortion operator $g_{\varphi,\lambda,\eta,\nu}$ of (1.6.4). The dashed line is obtained under the constraint $\nu = 1$ (idiosyncratic jump risk), and the solid line is obtained without this constraint (i.e., with $\nu \geq 1$). The residuals are plotted against the expected loss in Graph **(b)** for the unconstrained case, and in Graph **(c)** for the constrained case. A local linear regression of the residuals against the expected loss is carried out to detect systematic pricing errors.



(a) Illustration of the calibrated distortion operators



(b) Residuals : GB2 distortion



(c) Residuals : Normal-exponential distortion

Figure 1.2 – Calibration results for our Artemis.bm data set consisting of 508 CAT bond tranches issued between 1997 and 2016. The calibration is done by minimizing the sum of squared pricing errors. Graph **(a)** illustrates the empirical data and the calibrated distortion operators. The solid line corresponds to the GB2 distortion operator of (1.5.10), and the dashed line corresponds to the normal-exponential mixture distortion proposed in Section 1.6.4 of this paper. The residuals are plotted against the expected loss in Graph **(b)** for the GB2 distortion, and in Graph **(c)** for the normal-exponential distortion. A local linear regression of the residuals against the expected loss is carried out to detect systematic pricing errors.

Appendix

1.A Proofs

1.A.1 Proof of Proposition 1.3.1

Let the survival function of X under $\mathbb{P}$ and $\mathbb{Q}$ be denoted by $\bar{F}_{\mathbb{P}}(x) \equiv \mathbb{P}(X > x)$ and $\bar{F}_{\mathbb{Q}}(x) \equiv \mathbb{Q}(X > x)$. Similarly for $h(X)$, define $\bar{Q}_{\mathbb{P}}(x) \equiv \mathbb{P}(h(X) > x)$ and $\bar{Q}_{\mathbb{Q}}(x) \equiv \mathbb{Q}(h(X) > x)$. Since h is continuous and increasing, we have $\bar{Q}_{\mathbb{Q}}(x) = \bar{F}_{\mathbb{Q}}\big(h^{-1}(x)\big)$ and $\bar{Q}_{\mathbb{P}}^{-1}(u) = h\big(\bar{F}_{\mathbb{P}}^{-1}(u)\big)$. Using these last two identities in Definition 1.3.1 gives us

$$g_{h(X)}^{\mathbb{Q},\mathbb{P}}(u) \equiv \bar{Q}_{\mathbb{Q}}\big(\bar{Q}_{\mathbb{P}}^{-1}(u)\big) = \bar{F}_{\mathbb{Q}}\big(\bar{F}_{\mathbb{P}}^{-1}(u)\big) \equiv g_X^{\mathbb{Q},\mathbb{P}}(u).$$

1.A.2 Proof of Theorem 1.3.1

X follows a survival distribution function $\bar{F}_{g_Z^{\mathbb{Q},\mathbb{P}}}(x) = g_Z^{\mathbb{Q},\mathbb{P}}\big(\bar{F}_{\mathbb{P}}(x)\big)$ under the distorted measure. Using the chain rule to take the derivative of this equation with respect to x gives us

$$f_{g_Z^{\mathbb{Q},\mathbb{P}}}(x) = \dot{g}_Z^{\mathbb{Q},\mathbb{P}}\big(\bar{F}_{\mathbb{P}}(x)\big) f_{\mathbb{P}}(x), \tag{1.A.1}$$

where $\dot{g}_Z^{\mathbb{Q},\mathbb{P}}(u) \equiv \frac{d}{du} g_Z^{\mathbb{Q},\mathbb{P}}(u)$.

By Definition 1.3.1, we have $g_Z^{\mathbb{Q},\mathbb{P}}(u) \equiv \bar{Q}_{\mathbb{Q}}\big(\bar{Q}_{\mathbb{P}}^{-1}(u)\big)$. Taking the derivative with respect to u yields

$$\dot{g}_Z^{\mathbb{Q},\mathbb{P}}(u) = -q_{\mathbb{Q}}\big(\bar{Q}_{\mathbb{P}}^{-1}(u)\big) \frac{d\bar{Q}_{\mathbb{P}}^{-1}(u)}{du}, \tag{1.A.2}$$

where $q_{\mathbb{Q}}$ is the PDF of Z under $\mathbb{Q}$. Next, we note that

$$1 = \frac{d}{du} u = \frac{d}{du} \bar{Q}_{\mathbb{P}}\big(\bar{Q}_{\mathbb{P}}^{-1}(u)\big) = -q_{\mathbb{P}}\big(\bar{Q}_{\mathbb{P}}^{-1}(u)\big) \frac{d\bar{Q}_{\mathbb{P}}^{-1}(u)}{du} \quad \Rightarrow \quad \frac{d\bar{Q}_{\mathbb{P}}^{-1}(u)}{du} = \frac{-1}{q_{\mathbb{P}}\big(\bar{Q}_{\mathbb{P}}^{-1}(u)\big)}.$$

Using this last equality in (1.A.2) yields

$$\dot{g}_Z^{\mathbb{Q},\mathbb{P}}(u) = \frac{q_{\mathbb{Q}}\big(\bar{Q}_{\mathbb{P}}^{-1}(u)\big)}{q_{\mathbb{P}}\big(\bar{Q}_{\mathbb{P}}^{-1}(u)\big)} \equiv \xi_Z^{\mathbb{Q},\mathbb{P}}\big(\bar{Q}_{\mathbb{P}}^{-1}(u)\big), \tag{1.A.3}$$

where we have used Definition 1.3.2. Using (1.A.3) in (1.A.1) concludes the proof.

1.A.3 Proof of Corollary 1.3.1

Because $X = h(Z)$, where h is continuous and increasing, we have $\bar{F}_{\mathbb{P}}(x) = \bar{Q}_{\mathbb{P}}(h^{-1}(x))$. Then, by Theorem 1.3.1 it follows that $f_{g_Z^{\mathbb{Q},\mathbb{P}}}(x) = f_{\mathbb{P}}(x)\xi_Z^{\mathbb{Q},\mathbb{P}}(h^{-1}(x))$. Using standard results on the transformation of random variables, one can show that $f_{\mathbb{Q}}(x) = f_{\mathbb{P}}(x)\xi_Z^{\mathbb{Q},\mathbb{P}}(h^{-1}(x))$. [13]

1.A.4 Proof of Proposition 1.3.2

Because h is a continuous and increasing function, it follows from Corollary 1.3.1 that

$$g_X^{\mathbb{Q},\mathbb{P}}(\mathbb{P}(h(X) > x)) = \mathbb{Q}(h(X) > x), \qquad \forall x \in \mathbb{R}.$$

Using this in the definition (1.2.3) of the functional H gives us

$$H\Big[h(X); g_X^{\mathbb{Q},\mathbb{P}}\Big] = \int_{-\infty}^{0} \big[\mathbb{Q}(h(X) > x) - 1\big]dx + \int_0^{\infty} \mathbb{Q}(h(X) > x)dx = \mathbb{E}^{\mathbb{Q}}[h(X)],$$

where the last equality is a well-known identity.

1.A.5 Proof of Proposition 1.6.1

From (1.6.7),

$$\mathbb{E}^{\mathbb{P}}\left[e^{a\sigma Z_I}\right] = \sum_{i=1}^{m} \mathbb{P}(I = i)\mathbb{E}^{\mathbb{P}}\left[e^{a\sigma Z_I}\middle| I = i\right] = \sum_{i=1}^{m} p_i \mathbb{E}^{\mathbb{P}}\left[e^{a\sigma Z_i}\right] = \sum_{i=1}^{m} p_i \zeta_i^{(a\sigma)},$$

where definition (1.6.9) is used. Using this in (1.6.10) yields $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{e^{a\sigma Z_I}}{\sum_{i=1}^m p_i \zeta_i^{(a\sigma)}}$. Hence,

$$\begin{aligned}
\mathbb{Q}(Z_I \le z) \equiv \mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{Z_I \le z\}}\right] \equiv \mathbb{E}^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\mathbb{1}_{\{Z_I \le z\}}\right] &= \frac{\mathbb{E}^{\mathbb{P}}\left[e^{a\sigma Z_I}\mathbb{1}_{\{Z_I \le z\}}\right]}{\sum_{j=1}^m p_j \zeta_j^{(a\sigma)}}, \\
&= \sum_{i=1}^{m} \frac{p_i \mathbb{E}^{\mathbb{P}}\left[e^{a\sigma Z_i}\mathbb{1}_{\{Z_i \le z\}}\right]}{\sum_{j=1}^m p_j \zeta_j^{(a\sigma)}}, \\
&= \sum_{i=1}^{m} \frac{p_i \zeta_i^{(a\sigma)}}{\sum_{j=1}^m p_j \zeta_j^{(a\sigma)}} \mathbb{E}^{\mathbb{P}}\left[\frac{e^{a\sigma Z_i}}{\zeta_i^{(a\sigma)}}\mathbb{1}_{\{Z_i \le z\}}\right],
\end{aligned}$$

from which one can indeed conclude (1.6.11).

13. Define $q_{\mathbb{P}}$ as the PDF of Z under $\mathbb{P}$. Since h is continuous and increasing, it follows that the PDF of $X = h(Z)$ is given by $f_{\mathbb{P}}(x) = q_{\mathbb{P}}(h^{-1}(x))/h'(h^{-1}(x))$, where h' is the derivative of h. Similarly, under $\mathbb{Q}$ we have $f_{\mathbb{Q}}(x) = q_{\mathbb{Q}}(h^{-1}(x))/h'(h^{-1}(x))$. Therefore : $f_{\mathbb{Q}}(x) = f_{\mathbb{P}}(x)\xi_Z^{\mathbb{Q},\mathbb{P}}(h^{-1}(x))$, where $\xi_Z^{\mathbb{Q},\mathbb{P}}(z) \equiv q_{\mathbb{Q}}(z)/q_{\mathbb{P}}(z)$.

1.A.6 Proof of Eq. (1.6.15)

Under $\mathbb{P}$, $Z_1 \sim N(0,1)$ and $Z_2 \sim \text{Exp}(\eta)$. Hence, the PDFs of Z_1 and Z_2 are respectively

$$f_1(x) = \phi(x) \equiv \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \qquad f_2(x) = \eta e^{-\eta x} \mathbb{1}_{\{x\geq 0\}}, \qquad x \in \mathbb{R}.$$

Furthermore,

$$\zeta_1^{(a\sigma)} \equiv \mathbb{E}^{\mathbb{P}}\big[e^{a\sigma Z_1}\big] = e^{\frac{1}{2}(a\sigma)^2}, \qquad \zeta_2^{(a\sigma)} \equiv \mathbb{E}^{\mathbb{P}}\big[e^{a\sigma Z_2}\big] = \frac{\eta}{\eta - a\sigma},$$

where the latter holds if $\eta > a\sigma$. It follows that

$$\tilde{f}_1(x) \equiv \frac{e^{a\sigma x}}{\zeta_1^{(a\sigma)}} f_1(x) = \phi(x - a\sigma), \qquad \tilde{f}_2(x) \equiv \frac{e^{a\sigma x}}{\zeta_2^{(a\sigma)}} f_2(x) = (\eta - a\sigma) e^{-(\eta - a\sigma)x} \mathbb{1}_{\{x\geq 0\}}, \qquad x \in \mathbb{R}.$$

Then, applying Proposition 1.6.1 yields (1.6.15).

References

Ai, J. and Brockett, P. L. (2008). Insurance pricing/nonlife. In Melnick., E. L. and Everitt, B. S., editors, *Encyclopedia of Quantitative Risk Analysis and Assessment*, volume 2. John Wiley & Sons, Ltd.

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3) :203–228.

Azzalini, A., Bowman, A. W., and Härdle, W. (1989). On the use of nonparametric regression for model checking. *Biometrika*, 76(1) :1–11.

Bali, T. G. and Theodossiou, P. (2008). Risk management performance of alternative distribution functions. *The Journal of Risk and Insurance*, 75(2) :411–437.

Bodoff, N. M. and Gan, Y. (2012). An analysis of the market price of Cat bonds. *Variance*, 6(2) :161–177.

Braun, A. (2015). Pricing in the primary market for Cat bonds : New empirical evidence. *The Journal of Risk and Insurance*. Article first published online : DOI : 10.1111/jori.12067.

Bühlmann, H. (1980). An economic premium principle. *ASTIN Bulletin*, 11 :52–60.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC.

Cummins, J. D., Dionne, G., McDonald, J. B., and Pritchett, B. M. (1990). Applications of the GB2 family of distributions in modeling insurance loss processes. *Insurance : Mathematics and Economics*, 9(4) :257–272.

Delbaen, F. and Haezendonck, J. (1989). A martingale approach to premium calculation principles in an arbitrage free market. *Insurance : Mathematics and Economics*, 8(4) :269–277.

Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(1) :463–520.

Delbaen, F. and Schachermayer, W. (1998). The fundamental theorem of asset pricing for unbounded stochastic processes. *Mathematische Annalen*, 312(2) :215–250.

Denneberg, D. (1994). *Non-Additive Measure and Integral*. Springer.

Dowd, K. and Blake, D. (2006). After VaR : The theory, estimation, and insurance applications of quantile-based risk measures. *The Journal of Risk and Insurance*, 73(2) :193–229.

Eberlein, E. and Jacod, J. (1997). On the range of options prices. *Finance and Stochastics*, 1(2) :131–140.

Föllmer, H. and Schweizer, M. (1991). Hedging of contingent claims under incomplete information. In Davis, M. H. A. and Elliot, R. J., editors, *Applied Stochastic Analysis*, pages 389–414. Gordon and Breach.

Furman, E. and Zitikis, R. (2008). Weighted premium calculation principles. *Insurance : Mathematics and Economics*, 42(1) :459–465.

Furman, E. and Zitikis, R. (2009). Weighted pricing functionals with applications to insurance : An overview. *North American Actuarial Journal*, 13(4) :483–496.

Galeotti, M., Gürtler, M., and Winkelvos, C. (2013). Accuracy of premium calculation models for CAT bonds - An empirical analysis. *The Journal of Risk and Insurance*, 80(2) :401–421.

Gerber, H. U. and Shiu, E. S. W. (1994). Option pricing by Esscher transforms. *Transactions of the Society of Actuaries*, 46 :99–191.

Godin, F., Mayoral, S., and Morales, M. (2012). Contingent claim pricing using a normal inverse gaussian probability distortion operator. *The Journal of Risk and Insurance*, 79(3) :841–866.

Goovaerts, M. J. and Laeven, R. J. A. (2008). Actuarial risk measures for financial derivative pricing. *Insurance : Mathematics and Economics*, 42(2) :540–547.

Gzyl, H. and Mayoral, S. (2008). Determination of risk pricing measures from market prices of risk. *Insurance : Mathematics and Economics*, 43(3) :437–443.

Hamada, M. and Sherris, M. (2003). Contingent claim pricing using probability distortion operators : Methods from insurance risk pricing and their relationship to financial theory. *Applied Mathematical Finance*, 10 :19–47.

Harrison, J. M. and Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3) :381–408.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3) :215–260.

Heilmann, W. R. (1989). Decision theoretic foundations of credibility theory. *Insurance : Mathematics and Economics*, 8(1) :77–95.

Kijima, M. (2006). A multivariate extension of equilibrium pricing transforms : The multivariate Esscher and Wang transforms for pricing financial and insurance risks. *ASTIN Bulletin*, 36(1) :269–283.

Kijima, M. and Muromachi, Y. (2008). An extension of the Wang transform derived from Buhlmann's economic premium principle for insurance risk. *Insurance : Mathematics and Economics*, 42(3) :887–896.

Kleiber, C. (1997). The existence of population inequality measures. *Economics Letters*, 57(1) :39–44.

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons, Ltd.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models : From Data to Decisions*. Wiley, 4th edition.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48(8) :1086–1101.

Labuschagne, C. C. A. and Offwood, T. M. (2010). A note on the connection between the Esscher-Girsanov transform and the Wang transform. *Insurance : Mathematics and Economics*, 47(3) :385–390.

Labuschagne, C. C. A. and Offwood, T. M. (2013). Pricing exotic options using the Wang transform. *North American Journal of Economics and Finance*, 25 :139–150.

Laeven, R. J. A. and Goovaerts, M. J. (2008). Premium calculation and insurance pricing. In Melnick., E. L. and Everitt, B. S., editors, *Encyclopedia of Quantitative Risk Analysis and Assessment*, volume 3. John Wiley & Sons, Ltd.

Lai, V. S., Parcollet, M., and Lamond, B. F. (2014). The valuation of catastrophe bonds with exposure to currency exchange risk. *International Review of Financial Analysis*, 33(2) :243–252.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics : Theory and practice*. Princeton University Press.

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica*, 46(6) :1429–1445.

Ma, Z. G. and Ma, C. Q. (2013). Pricing catastrophe risk bonds : A mixed approximation method. *Insurance : Mathematics and Economics*, 52(2) :243–254.

McDonald, J. B. (1987). Model selection : Some generalized distributions. *Communications in Statistics - Theory and Methods*, 16(4) :1049–1074.

Nowak, P. and Romaniuk, M. (2013). Pricing and simulations of catastrophe bonds. *Insurance : Mathematics and Economics*, 52(1) :18–28.

Pelsser, A. (2008). On the applicability of the Wang transform for pricing financial risks. *ASTIN Bulletin*, 38(1) :171–181.

Perrakis, S. and Boloorforoosh, A. (2013). Valuing catastrophe derivatives under limited diversification : A stochastic dominance approach. *Journal of Banking & Finance*, 37(8) :3157–3168.

Schoutens, W. (2003). *Lévy Processes in Finance : Pricing Financial Derivatives*. Wiley.

Schweizer, M. (1995). On the minimal martingale measure and the Föllmer-Schweizer decomposition. *Stochastic Analysis and Applications*, 13(5) :573–599.

Sondermann, D. (1991). Reinsurance in arbitrage-free markets. *Insurance : Mathematics and Economics*, 10(3) :191–202.

Vaugirard, V. E. (2003). Pricing catastrophe bonds by an arbitrage approach. *The Quarterly Review of Economics and Finance*, 43(1) :119–132.

Venter, G. (1983). Transformed beta and gamma distributions and aggregate losses. *Proceedings of the Casualty Actuarial Society*, 70 :156–193.

Wang, S. S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance : Mathematics and Economics*, 17(1) :43–54.

Wang, S. S. (1996). Ordering of risks under PH-transforms. *Insurance : Mathematics and Economics*, 18(2) :109–114.

Wang, S. S. (2000). A class of distortion operators for pricing financial and insurance risks. *The Journal of Risk and Insurance*, 67(1) :15–36.

Wang, S. S. (2003). Equilibrium pricing transforms : New results using Buhlmann's 1980 economic model. *ASTIN Bulletin*, 33(1) :57–73.

Wang, S. S. (2004). Cat bond pricing using probability transforms. *Geneva Papers : Etudes et Dossiers*, 278 :19–29.

Wang, S. S. (2007). Normalized exponential tilting : Pricing and measuring multivariate risks. *North American Actuarial Journal*, 11(3) :89–99.

Wang, S. S., Young, V. R., and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance : Mathematics and Economics*, 21(2) :173–183.

Wirch, J. L. and Hardy, M. R. (1999). A synthesis of risk measures for capital adequacy. *Insurance : Mathematics and Economics*, 25(3) :337–347.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, 55(1) :95–115.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75(2) :263–289.

Chapter 2

Local hedging of variable annuities in the presence of basis risk

Résumé

Une méthode de couverture des rentes variables en présence de risque de base est développée. Un modèle à changement de régime est considéré pour la dynamique des actifs du marché. L'approche est basée sur une optimisation locale du risque et est donc très flexible. Le critère d'optimisation locale est lui-même optimisé pour minimiser les exigences de fonds propres associées aux rentes variables, ces dernières étant quantifiées par la mesure de risque CVaR. Par rapport aux benchmarks, notre méthode réussit à réduire simultanément les exigences de fonds propres et à augmenter la rentabilité. En effet, le schéma de couverture locale proposé bénéficie d'une exposition plus élevée au risque de marché et d'une diversification temporelle du risque pour gagner un rendement excédentaire et faciliter l'accumulation de capital. Une version robuste des stratégies de couverture couvrant le risque de modèle et l'incertitude des paramètres est également offerte.

Abstract

A method to hedge variable annuities in the presence of basis risk is developed. A regime-switching model is considered for the dynamics of market assets. The approach is based on a local optimization of risk and is therefore very tractable and flexible. The local optimization criterion is itself optimized to minimize capital requirements associated with the variable annuity policy, the latter being quantified by the CVaR risk metric. In comparison to benchmarks, our method is successful in simultaneously reducing capital requirements and increasing profitability. Indeed the proposed local hedging scheme benefits from a higher exposure to equity risk and from time diversification of risk to earn excess return and facilitate the accumulation of capital. A robust version of the hedging strategies addressing model risk and parameter uncertainty is also provided.

Keywords : Basis Risk, Hedging, Segregated Funds, Variable Annuities, Risk Measures, Risk Management, Regime-Switching Models.

2.1 Introduction

Variable annuity policies issued by life insurance companies are hybrid contracts involving both savings and insurance features. Indeed, such contracts allow the policyholder investing its account value in a mutual fund and obtaining variable returns tied to equity market performance. Moreover, those policies also offer guarantees taking various possible forms : a minimal rate of return on investments, a minimal benefit amount upon the death of the policyholder, etc.

The hedging of variable annuity guarantees by insurers presents multiple specific challenges since such products involve several features which are not present for vanilla options : mortality risk, lapse risk, periodic management charges to policyholder, fancy guarantee structures (e.g., ratchet features or GMWBs [1]), long maturities, and basis risk. Basis risk stems from the fact that insurers apply in practice a cross-hedge with liquid futures to mitigate risk associated with variable annuity liabilities due to the inconvenience or impossibility of shorting shares of the underlying mutual fund. Basis risk therefore refers to the imperfect correlation between returns of funds underlying variable annuity guarantees and returns of futures used to perform the hedge.

The current work aims at studying hedging schemes applicable to variable annuities by placing a special emphasis on the presence of basis risk, which is known to have a substantial impact on hedging residual risk in practice. As indicated in Zhang (2010), during the 2008 financial crisis, basis risk was one the most important sources of losses among insurers which implemented a dynamic hedging schemes to hedge guarantees associated with variable annuities. Although this risk has a material impact on hedging efficiency, basis risk has not received extensive attention within the literature in the context of variable annuities. An exception to this is the work of Ankirchner et al. (2014) who study the impact of variable annuities product design (proportional versus fixed charges) on the magnitude of basis risk and liquidity risk faced by the insurer. Basis risk in the context of option hedging is also studied in Zhang et al. (2017) who provide an analytical solution to a global mean-variance dynamic hedging problem under a bivariate Itô diffusion framework.

The current paper provides with three main contributions. First, a tractable and efficient hedging scheme making use of futures contracts is designed to hedge equity risk related to variable annuities issued by an insurer in the presence of basis risk. The optimization of the hedge is done through a local criterion which is itself optimized to minimize capital requirements associated with a given policy. Using a local criterion provides with sufficient tractability to consider realistic regime-switching asset price dynamics. The latter model is sufficiently realistic

1. A Guaranteed Minimum Withdrawal Benefit (GMWB) is a guarantee attached to a variable annuity which provides to the policyholder the right to withdraw from its policy a minimal amount each month until his initial investment is recouped.

to replicate stylized facts of financial markets (see Augustyniak and Boudreault, 2012, who show that regime-switching models can reproduce the thick left tail of financial returns), but also sufficiently parsimonious to retain tractability. Regime-switching models in the context of variable annuities hedging were considered by Wang and Yin (2012) and Qian et al. (2011) who applied respectively quantile hedging and local-risk minimization to perform the hedge. However, their framework does not include basis risk which is a desirable addition provided by the current paper. A key observation stemming from the simulation experiments illustrated in the current paper is that the omission of basis risk leads to severe risk under-estimation.

Our work mainly focuses on equity risk, which is the only source of uncertainty in the developed hedging scheme. The mutual fund underlying the variable annuity is assumed to be fully invested in equity, and therefore the hedge is performed with an equity futures. Extensions to our model handling stochastic interest rates, dynamic lapses and stochastic mortality improvement will be developed in upcoming papers from the authors. For instance, the inclusion of mortality related hedging instruments such as longevity bonds within the hedging scheme would be very relevant. Including additional sources of risk within our framework would have the effect of increasing capital requirements and potentially reducing the proportion of risk generated by basis risk since the latter would be diluted among other sources of risk. Additionally, the inclusion of dynamic lapses might increase the fair fees level and the magnitude of tail losses ; policyholders are expected to act in an adversarial manner and keep the policy in-force in scenarios where markets go down and large losses are incurred. The impact of embedding stochastic mortality and lapses within hedging schemes is investigated in Gaillardetz et al. (2012), Kling et al. (2014) and Boudreault and Augustyniak (2015) among others.

The second contribution of the current paper is the benchmarking of our method against approaches commonly found in the literature. Our local approach is shown in numerical experiments to greatly outperform the local minimal variance strategy simultaneously in terms of profitability and capital requirements, which is a significant contribution. In absence of model risk, the optimal local hedging approach even outperforms the no-hedging approach in terms of profitability. This is surprising due to the traditional premise stipulating that reducing risk through hedging comes at the expense of lower expected returns. However, this last finding is shown to be sensitive to model risk and parameter uncertainty ; when considering a robust version of the hedging strategy which addresses model risk, the no-hedging is much riskier but slightly more profitable in average than the optimal mean-variance hedge. Our approach benefits from time diversification of risk as it increases its local exposure to equity risk to generate higher expected returns and facilitate the accumulation of capital through time. Many commonly used approaches such as delta hedging attempt minimizing the exposure to equity risk and thus are unable to benefit from time diversification of risk. The sub-optimality of delta hedging in terms of optimization of global risk (as measured for instance by capital requirements) is well documented, see for instance Brandt (2003), Godin

(2016) and Augustyniak et al. (2016).

The third contribution relates to the numerical implementation of the proposed hedging methodology. Approximations based on Taylor expansions are applied on guarantee values to achieve a dimension reduction which substantially increases computational speed and convenience. This approach leads to a representation of the optimal hedging strategy which shows our method is a generalization of Greek-based methods implemented in practice such as delta hedging.

The current paper is divided as follows. Section 2.2 details the mathematical representation of cash flows involved in a dynamic hedging scheme for variable annuity guarantees. Local hedging optimization criteria are discussed in this section. Section 2.3 presents the regime-switching market model and outlines the application of the local hedging methodology to this particular market. Simulation based numerical experiments are presented in Section 2.4. Section 2.5 illustrates the implementation of a robust version of the hedging strategies addressing model risk and parameter uncertainty. Section 2.6 concludes.

2.2 Variable annuities hedging mechanics

This section outlines the mathematical model representing cash flows involved during the hedging of variable annuity guarantees by insurers.

2.2.1 Cash flows to the insurer

Consider a discrete set of monthly time steps $\mathcal{T} = \{0, \ldots, T\}$ and a probability space $(\Omega, \mathcal{F}_T, \mathbb{P})$ equipped with a filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t\in\mathcal{T}}$. The insurance company issues variable annuities to policyholders at time $t = 0$ and hedges risk pertaining to these contracts. Without loss of generality, a Guaranteed Minimal Maturity Benefit (GMMB) policy is considered.

Cash flows

The policy account is invested in a mutual fund whose value is a $\mathcal{F}$-adapted process denoted by $F = \{F_t\}_{t\in\mathcal{T}}$. Fees are periodically charged to policyholders and withdrawn from the policy account. Fees apply to all policyholders alive at the beginning of the period, but are only charged at the end of the period. Although fees are sometimes charged at the beginning of the period in practice, the impact of this fees timing assumption is very limited. Indeed, the discount factor applying to a single monthly time period is very close to 1. The total fee rate charged to policyholders is denoted by ω_{tot}. For a policyholder active at time $t-1$, the fees charged at time t are given by $\omega_{tot} A_{t-1} \frac{F_t}{F_{t-1}}$, where A_{t-1} is the post-fee policy account value at time $t-1$. The dynamics of $A = \{A_t\}_{t\in\mathcal{T}}$ is thus given by

$$A_{t+1} = A_t(1-\omega_{tot})\frac{F_{t+1}}{F_t}, \qquad t \in \{0,\dots,T-1\}. \tag{2.2.1}$$

$$\Rightarrow A_t = F_t\frac{A_0}{F_0}(1-\omega_{tot})^t, \qquad t \in \mathcal{T}. \tag{2.2.2}$$

A policyholder who dies during the period $(t, t+1]$ receives the amount A_{t+1} at time $t+1$. Moreover, at any time before maturity, the policyholder has the possibility to withdraw its investment from the variable annuity, which is called a lapse. In practice, insurers can diversify away a large proportion of idiosyncratic mortality and lapse risks by insuring a large number of policyholders. In the current framework, it is assumed that these risks can be fully diversified in this manner and are thus deterministic. The constant lapse rate on each period is denoted by b. Furthermore, ${}_tp_x$ is defined as the probability that a policyholder aged x months at time 0 survives t months. The proportion of policies that are still active at time t is thus given by

$$\ell_t = (1-b)^t\,{}_tp_x, \qquad t \in \mathcal{T}. \tag{2.2.3}$$

Surrender charges are not considered for simplicity, and therefore a policyholder who lapses during the period $(t, t+1]$ receives the amount A_{t+1} at time $t+1$. Such a simplification is acceptable in our framework due to the constant lapse rate assumption. However, extensions embedding dynamic lapses would definitely require the inclusion of surrender charges since such charges have a significant impact on policyholder surrender incentives as shown for instance by MacKay et al. (2017).

For a GMMB contract, in the case of a policyholder lapse or death, the benefit provided to the policyholder is fully funded by its policy account ; the insurer does not incur an outflow in this case. However, due to the guarantee, the insurer needs to pay the difference between the benefit and the policy account value at maturity if the policyholder is alive and the policy remains in-force until maturity. The benefit in excess of the account value paid to a GMMB policyholder whose policy is still active at the maturity T is given by $\max(0, K - A_T)$, where the guaranteed amount K is considered to be deterministic for simplicity.

Only a portion of fees collected from policyholders is allocated to the hedging portfolio for the guarantee, as the rest is allocated to profits and expenses. The fee rate which relates to fees allocated to hedging is denoted by ω_{opt}. Hence, if the policyholder is active at time $t-1$, the amount $\omega_{opt}A_{t-1}\frac{F_t}{F_{t-1}} = \frac{\omega_{opt}}{1-\omega_{tot}}A_t$ is received at time t and used by the insurance company for hedging. Therefore, the net cash outflow for the insurer at time t is given by

$$CF_t = -\frac{\omega_{opt}}{1-\omega_{tot}}A_t\ell_{t-1} + \mathbb{1}_{\{t=T\}}\max(0, K - A_T)\ell_T, \qquad t \in \{1,\dots,T\}, \tag{2.2.4}$$

with $CF_0 = 0$ as no immediate cash flows are involved at time $t = 0$. Defining

$$\tilde{K} \equiv \frac{KF_0}{A_0(1-\omega_{tot})^T}, \qquad \gamma_t \equiv \frac{A_0}{F_0}(1-\omega_{tot})^t\ell_t, \qquad t \in \mathcal{T}, \tag{2.2.5}$$

the relationship (2.2.2) can be used to express the cash flows directly in terms of the fund's value :

$$CF_t = -\omega_{opt}\gamma_{t-1}F_t + \mathbb{1}_{\{t=T\}}\gamma_T \max\bigl(0, \tilde{K} - F_T\bigr), \qquad t \in \{1, \dots, T\}. \tag{2.2.6}$$

Pricing

Since variable annuity option liabilities are not openly traded in markets, their fair value must be modeled. A possibility would be to use a value that is endogenous to the hedging strategy, such as the value that would allow optimizing the hedge according to some predefined criterion. Such price could be obtained through the quadratic global hedging approach ; see Rémillard and Rubenthaler (2013) for a general framework and Rémillard et al. (2017) for the Gaussian regime-switching specialization. However, prices are not endogenized to the hedging strategy in practice since insurance companies already have a pricing model which they use to determine the appropriate level of management expenses (MER) charged to clients.

The fair value of liabilities is obtained through risk-neutral valuation using an equivalent martingale measure $\mathbb{Q}$. We consider a risk-free asset whose price at time t is denoted $B_t = e^{rt}$ where r is the periodic risk-free rate. Since the discount price process $\{F_t/B_t\}_{t\in\mathcal{T}}$ is a martingale under $\mathbb{Q}$, it is straightforward to show that the time-t GMMB contract value is

$$\Pi_t = B_t\,\mathbb{E}^{\mathbb{Q}}\left[\sum_{j=t+1}^{T}\frac{CF_j}{B_j}\middle|\mathcal{F}_t\right] = -\omega_{opt}F_t\sum_{j=t+1}^{T}\gamma_{j-1} + \mathbb{1}_{\{t<T\}}\gamma_T G_t, \qquad t \in \mathcal{T}, \tag{2.2.7}$$

where

$$G_t \equiv B_t\,\mathbb{E}^{\mathbb{Q}}\left[\frac{\max\bigl(0, \tilde{K} - F_T\bigr)}{B_T}\middle|\mathcal{F}_t\right]. \tag{2.2.8}$$

Under this convention, the price Π_t excludes the cash flow CF_t from the pricing. In particular, $\Pi_T = 0$. The fee rate ω_{opt} is assumed to be a fair fee rate, i.e., the amount of fees which leads to a null initial value for the guarantee. Setting $\Pi_0 = 0$ in (2.2.7) leads to $\omega_{opt} = \frac{\gamma_T G_0}{F_0\sum_{j=1}^{T}\gamma_{j-1}}$.

Note that if the profitability provided by the fair fee rate ω_{opt} is deemed inadequate, the insurer might decide to adjust the total fee rate ω_{tot} provided that it remains competitive. Thus ω_{tot} and ω_{opt} would be related through a complex non-linear optimization in this case. However, we do not investigate such mechanisms in the current work since the pricing parameter ω_{tot} is assumed to be given.

Hedging

To mitigate risks embedded in guarantees provided by variables annuities, insurers perform a cross hedge based on a different hedging asset S, which creates basis risk since the assets F

and S are not perfectly correlated. The insurer sets up a hedging portfolio taking positions in two assets : the risk-free asset $B = \{B_t\}_{t\in\mathcal{T}}$ and a risky equity futures contract $S = \{S_t\}_{t\in\mathcal{T}}$, where S_t denotes the futures price at time t. The use of futures as hedging instruments is consistent with insurers practices, see for instance Chopra et al. (2009), iA Financial Group (2016) and Manulife Financial Corporation (2016). It is justified by the impossibility of taking short positions on many index funds and by the high liquidity of index futures. The number of long positions within the hedging portfolio during the time interval $(t, t+1]$ are respectively denoted by $\theta^{(B)}_{t+1}$ and $\theta^{(S)}_{t+1}$, with the convention $\theta^{(B)}_0 = \theta^{(S)}_0 = 0$. The insurer performs periodic injections or withdrawals of liquidities from the hedging portfolio at each time step. The injection at time t is denoted by I_t (negative amounts correspond to withdrawals). Define V^θ_{t-} and V^θ_{t+} as the value of the hedging portfolio respectively before and after the injection I_t and the cash flow CF_t at time t. This leads to

$$I_t = V^\theta_{t+} - V^\theta_{t-} + CF_t, \qquad t \in \mathcal{T}. \tag{2.2.9}$$

To ensure the hedging portfolio value tracks the guarantee value, at all steps $t \in \mathcal{T}$ the cash flow injection (or withdrawal) I_t is performed such that the post-injection portfolio value is equal to the guarantee value :

$$V^\theta_{t+} = \Pi_t, \qquad t \in \mathcal{T}. \tag{2.2.10}$$

In particular, the time 0 injection is nil, $I_0 = 0$, because $V^\theta_{0-} \equiv 0$, $CF_0 = 0$ and $\Pi_0 = 0$.

After the post-injection portfolio value $V^\theta_{t+} = \Pi_t$ is observed, the hedging portfolio is rebalanced. A new futures position $\theta^{(S)}_{t+1}$ is decided based on the selected hedging strategy, and then the whole portfolio value is invested at the risk-free rate ; entering positions on futures does not involve immediate cash flows besides margin requirements, and liquidities deposited inside the margin are assumed to accrue at the risk-free rate :

$$\theta^{(B)}_{t+1} = \frac{V^\theta_{t+}}{B_t} = \frac{\Pi_t}{B_t}. \tag{2.2.11}$$

The new pre-injection portfolio value at time $t+1$ is then obtained by summing the amount accrued at risk-free rate and profits/losses from futures positions :

$$V^\theta_{(t+1)-} = \theta^{(B)}_{t+1} B_{t+1} + \theta^{(S)}_{t+1}(S_{t+1} - S_t). \tag{2.2.12}$$

The following proposition proven in Appendix 2.A.1 gives an explicit expression for injections.

Proposition 2.2.1. *For $t \in \{0, \ldots, T-1\}$, define $\delta F_t \equiv F_{t+1} - F_t$, $\delta S_t \equiv S_{t+1} - S_t$, and $\delta G_t = G_{t+1} - G_t$. Cash flow injections are given by*

$$I_{t+1} = \Pi_t(1 - e^r) - \delta F_t \, \omega_{opt} \sum_{j=t+1}^{T} \gamma_{j-1} + \gamma_T \delta G_t - \theta^{(S)}_{t+1} \delta S_t. \tag{2.2.13}$$

Remark 2.2.1. The no-hedging (or unhedged) injection I^{NH}_{t+1} is defined as the injection value that would occur with $\theta^{(S)}_{t+1} = 0$. It follows from Proposition 2.2.1 that $I_{t+1} = I^{\text{NH}}_{t+1} - \theta^{(S)}_{t+1}\delta S_t$.

2.2.2 Capital requirements

Insurers must hold reserves and capital to meet future variable annuity guarantee liabilities. In Canada, insurers issuing segregated fund policies (which are the Canadian equivalent of variable annuities) are required by the Office of the Superintendent of Financial Institutions (OSFI) to hold a Total Gross Capital Required (TGCR) at time $t = 0$ represented by [2]

$$\text{TGCR} = \text{CVaR}^{\mathbb{P}}_{0.95}\left[\sum_{t=1}^{T} e^{-rt} I_t\right]. \tag{2.2.14}$$

The CVaR risk measure is defined rigorously in Rockafellar and Uryasev (2002). For a continuous random variable X, the $\text{CVaR}_\alpha[X]$ can be interpreted as the average of the worse $100(1-\alpha)\%$ scenarios. Note that U.S. recommendations for determining capital requirements for variable annuities are also based on the CVaR risk measure. [3]

2.2.3 Selection of the hedging strategy

The current paper's approach for the selection of the hedging strategy $\theta^{(S)} = \{\theta_t^{(S)}\}_{t\in\mathcal{T}}$ is to use a local criterion based on risk measures to optimize the risk. An $\mathcal{F}_t$ risk measure is a mapping $\mathcal{R}_t : \mathcal{X}_{\mathcal{F}_T} \to \mathcal{X}_{\mathcal{F}_t}$, where $\mathcal{X}_{\mathcal{G}}$ is the set of $\mathcal{G}$-measurable random variables for some sigma-algebra $\mathcal{G}$. The number of futures positions in the hedging portfolio is chosen by minimizing the risk related to the next cash injection :

$$\theta_{t+1}^{(S)*} = \underset{\theta_{t+1}^{(S)}}{\arg\min}\, \mathcal{R}_t(I_{t+1}), \qquad t \in \{0, \dots, T-1\}, \tag{2.2.15}$$

for a given dynamic risk measure $\{\mathcal{R}_t\}_{t=0}^{T-1}$, where $\mathcal{R}_t$ is an $\mathcal{F}_t$ risk measure for each t. Local procedures in hedging were pioneered by Ederington (1979) who uses $\mathcal{R}_t(\bullet) = \text{Var}^{\mathbb{P}}[\bullet|\mathcal{F}_t]$. However, an important drawback associated with the variance is that it focuses purely on risk in general, penalizing upside risk and failing to incorporate expected costs in the tradeoff. Rockafellar and Uryasev (2000) use the CVaR measure to optimize their hedge which allows reducing the magnitude and frequency of extreme losses. A plethora of other classes of risk measures that were developed in the literature could also be considered, for instance coherent risk measures (Artzner et al., 1999), deviation measures (Rockafellar et al., 2002) and distortion measures (Wang, 2000).

2. See Section 11 of the instruction guide by the OSFI on the "Use of Internal Models for Determining Required Capital for Segregated Fund Risks" which instructs using $\text{CVaR}^{\mathbb{P}}_{0.95}$ to determine the TGCR for segregated funds. See also Chapter 6 of the "Capital Adequacy Guideline" by the *Autorité des Marchés Financiers* for the Province of Québec.

3. See page 11 of the report "The Application of C-3 Phase II and Actuarial Guideline XLIII" by the American Academy of Actuaries which recommends using the $\text{CVaR}^{\mathbb{P}}_{0.90}$ to determine the TGCR for variable annuities.

2.3 A model involving regime-switching equity risk

This section presents the regime-switching price dynamics model for risky assets S and F.

2.3.1 Market model

Over a long period of time, markets go through various periods of either prosperity or turbulence. To model such dynamics, regime-switching processes have become very popular in the actuarial literature, see for instance Hardy (2001) and Hardy (2003). In these models, the overall state of the market is represented by the state of a Markov chain. Asset return distributions are then presumed to be a function of the current market state.

A regime process $h = \{h_t\}_{t\in\mathcal{T}}$ characterizes the state of the market, where $h_t \in \{1, 2\}$ can only take two possible values. Regimes h_t are latent variables, i.e., they are not directly observable. The following regime-switching dynamics is assumed for the risky asset prices :

$$R^{(F)}_{t+1} \equiv \log\left(\frac{F_{t+1}}{F_t}\right) = \mu^{(F)}_{h_t} + \sigma^{(F)}_{h_t} z^{(F)}_{t+1}, \qquad R^{(S)}_{t+1} \equiv \log\left(\frac{S_{t+1}}{S_t}\right) = \mu^{(S)}_{h_t} + \sigma^{(S)}_{h_t} z^{(S)}_{t+1},$$
$$z_{t+1} \equiv \begin{bmatrix} z^{(F)}_{t+1} \\ z^{(S)}_{t+1} \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{h_t} \\ \rho_{h_t} & 1 \end{bmatrix}\right), \tag{2.3.1}$$

where $N_2(\mu, \Sigma)$ is the bivariate normal distribution with mean vector μ and covariance matrix Σ, $z = \{z_t\}_{t\in\mathcal{T}}$ is a strong standardized Gaussian bivariate white noise, and the remaining parameters are constants to be estimated. Under this model, asset prices evolve according to

$$F_{t+1} = F_t e^{\mu^{(F)}_{h_t} + \sigma^{(F)}_{h_t} z^{(F)}_{t+1}}, \qquad S_{t+1} = S_t e^{\mu^{(S)}_{h_t} + \sigma^{(S)}_{h_t} z^{(S)}_{t+1}}. \tag{2.3.2}$$

The market information at time t is $\mathcal{F}_t \equiv \sigma\big(S_u, F_u : u = 0, \ldots, t\big)$. Following the lines of François et al. (2014), a full information filtration $\mathcal{G} \equiv \{\mathcal{G}_t\}_{t\in\mathcal{T}}$, where $\mathcal{G}_t \equiv \mathcal{F}_t \vee \sigma\big(h_u : u = 0, \ldots, t\big)$ is introduced. The regime process h is assumed to have the Markov property with respect to $\mathcal{G}$, i.e., for some transition matrix

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} \\ P_{2,1} & P_{2,2} \end{bmatrix}, \tag{2.3.3}$$

with $P_{1,2} = 1 - P_{1,1}$ and $P_{2,2} = 1 - P_{2,1}$, the following relationship holds for $i, j \in \{1, 2\}$:

$$\mathbb{P}\big(h_{t+1} = j \big| h_t = i, \{h_u\}^{t-1}_{u=0}, \{(F_u, S_u)\}^{t}_{u=0}\big) = \mathbb{P}\big(h_{t+1} = j \big| h_t = i\big) = P_{i,j}. \tag{2.3.4}$$

As in François et al. (2014), the regime mass functions given partial information $\mathcal{F}_t$ are defined as follows :

$$\eta^{\mathbb{P}}_{i,t} \equiv \mathbb{P}(h_t = i|\mathcal{F}_t), \qquad i \in \{1, 2\}, \tag{2.3.5}$$

and can be computed recursively through

$$\eta^{\mathbb{P}}_{i,t+1} = \frac{\sum_{j=1}^{2} \phi_{\mu_j,\Sigma_j}\left(R^{(F)}_{t+1}, R^{(S)}_{t+1}\right) \eta^{\mathbb{P}}_{j,t} P_{j,i}}{\sum_{j=1}^{2} \phi_{\mu_j,\Sigma_j}\left(R^{(F)}_{t+1}, R^{(S)}_{t+1}\right) \eta^{\mathbb{P}}_{j,t}}, \qquad i \in \{1,2\}, \tag{2.3.6}$$

where ϕ_{μ_j,Σ_j} denotes the bivariate Gaussian probability density function with mean vector and covariance matrix given by

$$\mu_j \equiv \begin{bmatrix} \mu^{(F)}_j \\ \mu^{(S)}_j \end{bmatrix}, \qquad \Sigma_j \equiv \begin{bmatrix} \left(\sigma^{(F)}_j\right)^2 & \rho_j \sigma^{(F)}_j \sigma^{(S)}_j \\ \rho_j \sigma^{(F)}_j \sigma^{(S)}_j & \left(\sigma^{(S)}_j\right)^2 \end{bmatrix}. \tag{2.3.7}$$

The following result deriving from (2.3.2) is useful to develop offline schemes for computing the local hedging strategies.

Proposition 2.3.1. *The $\mathcal{F}_t$-conditional distribution of $\left(\frac{F_{t+1}}{F_t}, \frac{S_{t+1}}{S_t}\right)$ depends only on $\eta^{\mathbb{P}}_{1,t}$ under the physical measure $\mathbb{P}$.*

Sojourn time conditional probabilities are quite useful to obtain analytical option pricing formulas (see, e.g., Hardy, 2001). For $t \in \{0, \ldots, T-1\}$, $\tau \in \{0, \ldots, T-t\}$, and $i \in \{1,2\}$, we define the following :

$$H_{t,\tau,i} \equiv \mathbb{P}\big(Y_t = \tau \big| h_t = i\big), \qquad Y_t \equiv \sum_{u=t}^{T-1} \mathbb{1}_{\{h_u = 1\}}, \tag{2.3.8}$$

where Y_t is the sojourn time in regime 1 between time t and time T.

A recursive algorithm is available for computing them. Initialize with $H_{T-1,1,1} = H_{T-1,0,2} = 1$ and $H_{T-1,j,1} = H_{T-1,k,2} = 0$ for all $j \neq 1$ and all $k \neq 0$. Starting at $t = T-2$, the following backward induction formulas from Hardy (2001) are used :

$$H_{t,k,1} = \sum_{i=1}^{2} P_{1,i}\, H_{t+1,k-1,i}, \qquad H_{t,k,2} = \sum_{i=1}^{2} P_{2,i}\, H_{t+1,k,i}.$$

2.3.2 Valuing the guarantee

The above market model is incomplete, and therefore there exists an infinite number of risk-neutral measures. For analytical tractability, the chosen $\mathbb{Q}$ is such that the dynamics remain a regime-switching model of the same form and with the same transition matrix, but with the drift parameters $\left[\mu^{(F)}_j, \mu^{(S)}_j\right]$ replaced by $\left[r - \frac{1}{2}\left(\sigma^{(F)}_j\right)^2, -\frac{1}{2}\left(\sigma^{(S)}_j\right)^2\right]$ for $j \in \{1,2\}$. Note that the risk-free rate r does not appear in the risk-neutral drift of S since it is a futures contract. The usual Girsanov-type change of measures could be applied to show the existence of such a measure, see for instance Elliott et al. (2005) for analogous work in continuous-time.

An analytical formula for the price G_t, see (2.2.8), of the option embedded within the guarantee can be obtained following the lines of Hardy (2001). However, the option price in the latter paper depends on the regime currently prevailing in the economy. In the current work, regimes are unobservable and the option price is therefore a weighted average of the prices associated with each regime where weights are the respective risk-neutral probabilities of currently being in each regime, namely $\eta_{i,t}^{\mathbb{Q}} \equiv \mathbb{Q}\big(h_t = i|\mathcal{F}_t\big)$.

The option price G_t is represented by a function $g\big(t, F_t, \eta_{1,t}^{\mathbb{Q}}\big)$, which for $t < T$ is given by

$$g(t,F,\eta) = \sum_{\tau=0}^{T-t}\Big(\eta H_{t,\tau,1} + (1-\eta)H_{t,\tau,2}\Big)\Big(\tilde{K}e^{-r(T-t)}\Phi\big(-d_2(\tau)\big) - F\Phi\big(-d_1(\tau)\big)\Big), \quad (2.3.9)$$

where Φ is the standard normal cumulative distribution function, and

$$\begin{aligned} d_1(\tau) &\equiv \frac{\log\big(F/\tilde{K}\big) + (T-t)r + \frac{1}{2}\left[\tau\Big(\sigma_1^{(F)}\Big)^2 + (T-t-\tau)\Big(\sigma_2^{(F)}\Big)^2\right]}{\left[\tau\Big(\sigma_1^{(F)}\Big)^2 + (T-t-\tau)\Big(\sigma_2^{(F)}\Big)^2\right]^{1/2}}, \\ d_2(\tau) &\equiv d_1(\tau) - \left[\tau\Big(\sigma_1^{(F)}\Big)^2 + (T-t-\tau)\Big(\sigma_2^{(F)}\Big)^2\right]^{1/2}. \end{aligned} \quad (2.3.10)$$

The Delta is given by

$$\frac{\partial g}{\partial F}(t,F,\eta) = -\sum_{\tau=0}^{T-t}\big(\eta H_{t,\tau,1} + (1-\eta)H_{t,\tau,2}\big)\Phi\big(-d_1(\tau)\big), \qquad t < T, \quad (2.3.11)$$

and $\frac{\partial g}{\partial F}(T,F,\eta) = -\mathbb{1}_{\{\tilde{K}>F\}}$, i.e., for $t = T$.

2.3.3 Taylor expansion on injections

The cash flow injections (or withdrawals) involved in the monthly rebalancing of the hedging portfolio are characterized in Proposition 2.2.1, where the expression of the injection I_{t+1} at time $t+1$ involves the monthly change in value of the guarantee : $\delta G_t \equiv G_{t+1} - G_t$. In order to simplify the solution to the hedging optimization problem, a Taylor approximation of δG_t can be applied using the option Greeks. The pricing function $g(t,F,\eta)$ given in (2.3.9) is only defined for discrete values of t, and thus the expansion cannot be centered on the time t as it would require a well-defined time sensitivity. Our solution to this issue is based on the following delta-type approximation :

$$\begin{aligned} \delta G_t &= \underbrace{g\big(t+1, F_{t+1}, \eta_{1,t+1}^{\mathbb{Q}}\big) - g\big(t+1, F_t, \eta_{1,t}^{\mathbb{Q}}\big)}_{\approx \delta F_t \frac{\partial g}{\partial F}\left(t+1, F_t, \eta_{1,t}^{\mathbb{Q}}\right)} + g\big(t+1, F_t, \eta_{1,t}^{\mathbb{Q}}\big) - g\big(t, F_t, \eta_{1,t}^{\mathbb{Q}}\big), \\ \Rightarrow \delta G_t &\approx \delta F_t \frac{\partial g}{\partial F}\big(t+1, F_t, \eta_{1,t}^{\mathbb{Q}}\big) + g\big(t+1, F_t, \eta_{1,t}^{\mathbb{Q}}\big) - g\big(t, F_t, \eta_{1,t}^{\mathbb{Q}}\big), \end{aligned} \quad (2.3.12)$$

where $\delta F_t = F_{t+1} - F_t$. Note that the impact of the variation $\delta\eta^{\mathbb{Q}}_{1,t} = \eta^{\mathbb{Q}}_{1,t+1} - \eta^{\mathbb{Q}}_{1,t}$ and of higher order variations, e.g., $(\delta F_t)^2$, have been omitted from the approximation. Although our hedging methodology can be extended to incorporate such corrections, unreported tests, available from the authors upon request, showed that their impact is not material.

Next, define the following Greek letters for $t \in \{0, \dots, T-1\}$:

$$\begin{aligned}\Theta_t &\equiv \Pi_t(1 - e^r) + \gamma_T\Big[g\big(t+1, F_t, \eta^{\mathbb{Q}}_{1,t}\big) - g\big(t, F_t, \eta^{\mathbb{Q}}_{1,t}\big)\Big],\\ \Delta_t &\equiv -\omega_{opt}\sum_{j=t+1}^{T}\gamma_{j-1} + \gamma_T\frac{\partial g}{\partial F}\big(t+1, F_t, \eta^{\mathbb{Q}}_{1,t}\big).\end{aligned} \tag{2.3.13}$$

Using the approximation (2.3.12) in the injection formula of Proposition 2.2.1 provides us with a delta-type approximation of the injection : $I_{t+1} \approx \tilde{I}_{t+1}$, where

$$\tilde{I}_{t+1} \equiv \Theta_t + \Delta_t \delta F_t - \theta^{(S)}_{t+1}\delta S_t. \tag{2.3.14}$$

The Greek Θ_t represents the value of the injection at time $t+1$ if F and S are unchanged from time t to time $t+1$. The Greek Δ_t measures the sensitivity of the injection to F.

In the current work, the injection approximation $\tilde{I}_{t+1}$ defined in (2.3.14) is used to tackle a simplified version of the hedging problem (2.2.15) that is more tractable and more easily solved. Since this approximation embeds the bulk of the risk related to the injection I_{t+1}, such a formulation is deemed a very good approximation of the hedging problem. Moreover, it provides a generalization of the delta-based approach to hedging and, as such, it is therefore in line with industry practices.

2.3.4 Local hedging based on the mean-variance risk measure

Here, a local hedging strategy based on the mean-variance family of risk measures is considered :

$$\theta^{(S)*}_{t+1} = \underset{\theta^{(S)}_{t+1}}{\arg\min}\Big\{\mathrm{Var}^{\mathbb{P}}\big[\tilde{I}_{t+1}\big|\mathcal{F}_t\big] + 2\lambda\mathbb{E}^{\mathbb{P}}\big[\tilde{I}_{t+1}\big|\mathcal{F}_t\big]\Big\}, \qquad t \in \{0, \dots, T-1\}, \tag{2.3.15}$$

where $\tilde{I}_{t+1}$ is the injection approximation, see (2.3.14), and $\lambda \geq 0$ is a chosen constant quantifying the mean-variance tradeoff. The mean-variance risk measure offers a flexible parametrization of the risk-return tradeoff while benefiting from convenient analytical properties. The solution to the above minimization problem is given in Proposition 2.3.2. The proof of this proposition is omitted since it is straightforward. The explicit formulas for the variance, covariance, and expectation involved in this proposition are provided in Appendix 2.A.2.

Proposition 2.3.2. *The mean-variance hedging strategy* (2.3.15) *is given by*

$$\theta^{(S)*}_{t+1} = \Delta_t\frac{\mathrm{Cov}^{\mathbb{P}}[F_{t+1}, S_{t+1}|\mathcal{F}_t]}{\mathrm{Var}^{\mathbb{P}}[S_{t+1}|\mathcal{F}_t]} + \lambda\frac{\mathbb{E}^{\mathbb{P}}[S_{t+1}|\mathcal{F}_t] - S_t}{\mathrm{Var}^{\mathbb{P}}[S_{t+1}|\mathcal{F}_t]}. \tag{2.3.16}$$

The strategy obtained under the choice $\lambda = 0$ is referred to as the *minimal variance strategy*. Interestingly, in the absence of basis risk the minimal variance strategy coincides with the usual form of delta hedging. The mean-variance strategy can be thought of as a generalization of delta hedging which can account for basis risk in addition of bringing expected costs into the tradeoff.

Properties of the mean-variance strategy

The proposition below characterizes the upper bound to the reduction of local risk, as measured by the variance, that is attainable in the presence of basis risk. It states, for instance, that if the correlation between returns of S and F is 90%, then the proportion of the standard deviation that can be eliminated is $1 - \sqrt{1 - 0.9^2} \approx 56.4\%$. The proof of Proposition 2.3.3 is obtained by combining (2.3.16) and (2.3.14).

Proposition 2.3.3. *Let $\tilde{I}_{t+1}^{\text{NH}}$ and $\tilde{I}_{t+1}^{\text{MV}}$ be the injection approximation* (2.3.14) *respectively for $\theta_{t+1}^{(S)} = 0$ (no hedging) and for $\theta_{t+1}^{(S)}$ given by* (2.3.16) *with $\lambda = 0$ (minimal variance). Then*

$$\frac{\text{Var}^{\mathbb{P}}\left[\tilde{I}_{t+1}^{\text{MV}} \middle| \mathcal{F}_t\right]}{\text{Var}^{\mathbb{P}}\left[\tilde{I}_{t+1}^{\text{NH}} \middle| \mathcal{F}_t\right]} = 1 - \text{Corr}^{\mathbb{P}}\left[\delta F_t, \delta S_t \middle| \mathcal{F}_t\right]^2. \tag{2.3.17}$$

The following remark relates the injection under the minimal variance strategy to the injection under the more general mean-variance strategy. This relation will prove to be useful when analyzing the simulation results. To prove it, one simply has to use Proposition 2.3.2 in the injection approximation formula (2.3.14).

Remark 2.3.1. Let $\tilde{I}_{t+1}$ and $\tilde{I}_{t+1}^{\text{MV}}$ be the injection approximation (2.3.14) respectively under the mean-variance strategy for some $\lambda \geq 0$ and the minimal variance strategy ($\lambda = 0$). Then

$$\tilde{I}_{t+1} = \tilde{I}_{t+1}^{\text{MV}} - \lambda \frac{\mathbb{E}^{\mathbb{P}}[S_{t+1}|\mathcal{F}_t] - S_t}{\text{Var}^{\mathbb{P}}[S_{t+1}|\mathcal{F}_t]} \delta S_t.$$

Optimizing the mean-variance tradeoff

For the mean-variance family of risk measures, the free parameter to be optimized is λ, which characterizes the risk-reward tradeoff. The objective is thus to find the value of that parameter which minimizes the capital (2.2.14) that the insurer is required to hold at time $t = 0$. We solve

$$\lambda^* \equiv \arg\min_{\lambda} \text{CVaR}_{0.95}^{\mathbb{P}} \left[\sum_{t=1}^{T} e^{-rt} I_t \right], \tag{2.3.18}$$

over hedging strategies of the form outlined in Proposition 2.3.2. Note that this optimization problem is based on the exact injections I_t rather than on their approximations $\tilde{I}_t$; this is because there are no numerical incentives not to rely on an exact formulation here.

A simple Monte Carlo simulation procedure is described to compute the value of λ^*. Consider a hedging strategy of the from

$$\theta_{t+1}^{(S)} = \alpha_t + \lambda\beta_t, \tag{2.3.19}$$

where $\lambda \in \mathbb{R}$, and (α_t, β_t) are some given $\mathcal{F}_t$-measurable random variables. Note that the mean-variance hedging strategy outlined in Proposition 2.3.2 has this form. For such strategies, it follows from Remark 2.2.1 that the injection value at time $t+1$ can be written as

$$I_{t+1} = I_{t+1}^{\text{NH}} - \alpha_t \delta S_t - \lambda \beta_t \delta S_t,$$

where I_{t+1}^{NH} is the injection value that would occur in the absence of hedging. We therefore have

$$\sum_{t=1}^{T} e^{-rt} I_t = \xi_1 - \lambda \xi_2, \tag{2.3.20}$$

where

$$\xi_1 \equiv \sum_{t=1}^{T} e^{-rt}[I_t^{\text{NH}} - \alpha_{t-1}\delta S_{t-1}], \qquad \xi_2 \equiv \sum_{t=1}^{T} e^{-rt}\beta_{t-1}\delta S_{t-1}. \tag{2.3.21}$$

Since ξ_1 and ξ_2 do not depend on λ, a single Monte Carlo simulation of the random variables (ξ_1, ξ_2) is required to optimize λ through

$$\lambda^* = \arg\min_{\lambda} \text{CVaR}_{0.95}^{\mathbb{P}}\left[\xi_1 - \lambda\xi_2\right]. \tag{2.3.22}$$

We refer to Rockafellar and Uryasev (2000) for a study on the numerical optimization of CVaR risk measures.

2.3.5 Local hedging based on a general class of risk measures

The current section characterizes the solution of the local hedging problem for a very large class of risk measures. A hedging strategy which minimizes capital requirements is also presented.

A general class of dynamic risk measures

A general class of risk-measures called the $\mathcal{F}_t$-reducible risk measures is introduced.

Definition 2.3.1. $\mathcal{R}_t$ is a $\mathcal{F}_t$*-reducible risk measure* if it satisfies the following :

1. $\mathcal{R}_t$ is a law-invariant $\mathcal{F}_t$ risk measure.
2. There exists a real function f_1 such that $\mathcal{R}_t(Y_t + X) = f_1(Y_t) + \mathcal{R}_t(X)$ for any risk X and any $\mathcal{F}_t$-measurable random variable Y_t.
3. There exists a nonnegative real function f_2 such that $\mathcal{R}_t(Y_t X) = f_2(Y_t)\mathcal{R}_t(X)$ for any admissible risk X and any $\mathcal{F}_t$-measurable random variable $Y_t \geq 0$ a.s.

The $\mathcal{F}_t$-reducible class of risk measures is quite large. In particular, it includes the $\mathcal{F}_t$-conditional counterparts of all coherent risk measures in the sense of Artzner et al. (1999), such as the Conditional Value-at-Risk, $\text{CVaR}_\alpha[\bullet|\mathcal{F}_t]$. It is however more general as it also includes the $\mathcal{F}_t$-conditional variance, $\text{Var}[\bullet|\mathcal{F}_t]$, and the $\mathcal{F}_t$-conditional Value-at-Risk, $\text{VaR}_\alpha[\bullet|\mathcal{F}_t]$. Furthermore, as shown in the proposition below, this class includes a large family of risk-reward tradeoffs, e.g., $\text{Std}[\bullet|\mathcal{F}_t] + \lambda\mathbb{E}^{\mathbb{P}}[\bullet|\mathcal{F}_t]$, $\text{VaR}_\alpha[\bullet|\mathcal{F}_t] + \lambda\mathbb{E}^{\mathbb{P}}[\bullet|\mathcal{F}_t]$, and $\text{CVaR}_\alpha[\bullet|\mathcal{F}_t] + \lambda\mathbb{E}^{\mathbb{P}}[\bullet|\mathcal{F}_t]$.

Proposition 2.3.4. *Let $\mathcal{R}_t$ be a $\mathcal{F}_t$-reducible risk measure such that $\mathcal{R}_t(Y_t + Z_t X) = f_1(Y_t) + Z_t\mathcal{R}_t(X)$ for all admissible risk X and all $\mathcal{F}_t$-measurable random variables (Y_t, Z_t) where $Z_t \geq 0$ a.s. Then, for any constant $\lambda \in \mathbb{R}$, the risk-measure $\mathcal{M}_t(\bullet) \equiv \mathcal{R}_t(\bullet) + \lambda\mathbb{E}^{\mathbb{P}}[\bullet|\mathcal{F}_t]$ is also $\mathcal{F}_t$-reducible.*

The proof of the above statement is rather straightforward and therefore omitted.

For the purpose of a dynamic hedging strategy, sequences of risk measures that satisfy the temporal law-invariance property of Definition 2.3.2 are considered.

Definition 2.3.2. Let $\{\mathcal{R}_t\}_{t=0}^{T-1}$ be a sequence where $\mathcal{R}_t$ is a $\mathcal{F}_t$ risk measure for each t. This sequence is said to have the *temporal law-invariance* property if the following condition is satisfied : for all $t_1, t_2 \geq 0$ and all random variables X_1, X_2, if the conditional distribution of X_1 given $\mathcal{F}_{t_1}$ is the same that the conditional distribution of X_2 given $\mathcal{F}_{t_2}$, then $\mathcal{R}_{t_1}(X_1) = \mathcal{R}_{t_2}(X_2)$.

Offline calculation of the hedging strategies

Here, a local hedging strategy of the following form is considered :

$$\theta_{t+1}^{(S)*} = \underset{\theta_{t+1}^{(S)}}{\arg\min}\ \mathcal{R}_t\big(\tilde{I}_{t+1}\big), \qquad t \in \{0, \dots, T-1\}, \tag{2.3.23}$$

where $\tilde{I}_{t+1}$ is the injection approximation, see (2.3.14), and $\{\mathcal{R}_t\}_{t=0}^{T-1}$ a dynamic risk measure such that $\mathcal{R}_t$ is $\mathcal{F}_t$-reducible for each t.

For temporal law-invariant sequences of reducible risk measures, it turns out that efficient pre-calculation of such hedging strategies is made possible by a trick that reduces the dimension of the associated optimization problem. This result is presented in Theorem 2.3.1 whose proof is in Appendix 2.A.2.

Theorem 2.3.1. *For any $\mathcal{F}_t$-reducible risk measure $\mathcal{R}_t$, the hedging strategy (2.3.23) is*

$$\theta_{t+1}^{(S)*} = \psi_{t+1}^* \frac{\Delta_t F_t}{S_t}, \qquad \psi_{t+1}^* \equiv \underset{\psi}{\arg\min}\ \mathcal{R}_t\Big(\psi\tfrac{\delta S_t}{S_t} - \tfrac{\delta F_t}{F_t}\Big). \tag{2.3.24}$$

Moreover, the solution can be expressed as $\psi_{t+1}^* = \Psi_{\mathcal{R}_t}\big(\eta_{1,t}^{\mathbb{P}}\big)$ for some function $\Psi_{\mathcal{R}_t} : [0,1] \to \mathbb{R}$. Furthermore, for a temporal law-invariant sequence $\{\mathcal{R}_t\}_{t=0}^{T-1}$, one has $\Psi_{\mathcal{R}_0} = \cdots = \Psi_{\mathcal{R}_{T-1}}$.

Theorem 2.3.1 has several consequences worthy of noticing. First, note from (2.3.14) that the $\mathcal{F}_t$-conditional distribution of $\tilde{I}_{t+1}$ depends on the state variables $(\Theta_t, \Delta_t, F_t, S_t, \eta^{\mathbb{P}}_{1,t})$ and the control parameter $\theta^{(S)}_{t+1}$. Hence, it might appear that the solution to the hedging problem is a function of five state variables, making it difficult to use an offline approach to efficiently pre-calculate the solution over a grid of points and then use interpolation methods, especially because the state variables F_t an S_t can cover a large range when the time index t is high, which is the case for variable annuities having a maturity of several years. However, due to Theorem 2.3.1, it is actually possible to reduce the dimensionality of the hedging problem to the single state variable $\eta^{\mathbb{P}}_{1,t} \in [0, 1]$. The solution can therefore be pre-calculated offline for a small grid of points over the domain [0,1] to build a continuous function using linear interpolation, reducing computational time by several orders of magnitude.

Finding the strategy that minimizes capital requirements

The objective of this section is to show how to design a hedging strategy that attains capital requirements at least as low as the best-performing hedging strategy based on a reducible risk measure. This strategy is referred to as the the *minimal TGCR strategy.*

For the class of dynamic reducible risk measures considered in Theorem 2.3.1, the main result is that each hedging strategy has the form $\theta^{(S)*}_{t+1} = \Psi(\eta^{\mathbb{P}}_{1,t})\frac{\Delta_t F_t}{S_t}$ for some function $\Psi : [0, 1] \to \mathbb{R}$. Rather than optimizing over the choice of the risk measure, the approach we propose is to directly optimize over the latter function to minimize the capital (2.2.14) that the insurer is required to hold at time $t = 0$ to meet future variable annuity guarantee liabilities. The optimization problem is

$$\Psi^* \equiv \underset{\Psi:[0,1]\to\mathbb{R}}{\arg\min}\ \mathrm{CVaR}^{\mathbb{P}}_{0.95}\left[\sum_{t=1}^{T} e^{-rt} I_t\right], \tag{2.3.25}$$

where $\theta^{(S)}_t = \Psi(\eta^{\mathbb{P}}_{1,t-1})\frac{\Delta_{t-1}F_{t-1}}{S_{t-1}}$. Note that this optimization problem is based on the exact injections I_t rather than on their approximations $\tilde{I}_t$. This is because there are no reasons not to rely on the exact formulation here, in contrast with the local hedging optimization problem for which the delta approximation leads to a tremendous reduction in computational time.

Remark 2.3.2. It is not guaranteed that there exists a risk measure which corresponds to the optimal function Ψ^*. In fact, such a risk measure does not even need to exist.

To solve the above optimization problem, a parametric approximation of the solution is considered by employing a polynomial of degree n :

$$\Psi(\eta) \equiv \sum_{i=0}^{n} a_i \eta^i, \tag{2.3.26}$$

where $\{a_i\}_{i=0}^n$ are constants to be determined. For the strategy $\theta_{t+1}^{(S)} = \Psi(\eta_{1,t}^{\mathbb{P}})\frac{\Delta_t F_t}{S_t}$, it follows from Remark 2.2.1 that the injection value at time $t+1$ can be expressed as

$$I_{t+1} = I_{t+1}^{\text{NH}} - \frac{\Delta_t F_t}{S_t}\delta S_t \sum_{i=0}^{n} a_i \big(\eta_{1,t}^{\mathbb{P}}\big)^i,$$

where I_{t+1}^{NH} is the injection value that would occur if no hedging were performed, i.e., with $\theta_{t+1}^{(S)} = 0$. We thus have

$$\sum_{t=1}^{T} e^{-rt} I_t = \xi_1 - \sum_{i=0}^{n} a_i \xi_{2,i}, \tag{2.3.27}$$

where

$$\xi_1 \equiv \sum_{t=1}^{T} e^{-rt} I_t^{\text{NH}}, \qquad \xi_{2,i} \equiv \sum_{t=1}^{T} e^{-rt} \Delta_{t-1} F_{t-1} \frac{\delta S_{t-1}}{S_{t-1}} \big(\eta_{1,t-1}^{\mathbb{P}}\big)^i. \tag{2.3.28}$$

The solution can therefore be formulated as

$$(a_0^*, \ldots, a_n^*) \equiv \underset{(a_0,\ldots,a_n)}{\arg\min}\ \text{CVaR}_{0.95}^{\mathbb{P}}\left[\xi_1 - \sum_{i=0}^{n} a_i \xi_{2,i}\right]. \tag{2.3.29}$$

This problem can be solved numerically by first simulating a sufficiently large sample of the random variables $(\xi_1, \xi_{2,0}, \ldots, \xi_{2,n})$ to estimate the CVaR and minimize it using standard algorithms.

To choose the degree n, one can fix some pre-determined value $n_{\max}$ and simulate $(\xi_1, \xi_{2,0}, \ldots, \xi_{2,n_{\max}})$. This simulated sample suffices to test all degrees below $n_{\max}$. In practice, n is chosen as the smallest value for which the choice $n+1$ yields no further improvements.

2.4 Simulation experiments

The numerical simulations presented here form the basis of several analyses that allow making important findings about the properties of the hedging strategies and of the optimal mitigation of risks embedded in variable annuities under the presence of basis risk.

2.4.1 Setup

It is assumed that F is the Great-West Life Canadian Equity (GWLIM) BEL fund. The vast majority of the fund wealth (roughly 90%) is invested in Canadian Equity (the remainder being cash investments). The constant risk-free rate assumption is reasonable in this context ; if the mutual fund invested in fixed income, including interest shocks would have been necessary to impact the fluctuation of the fund value. Since the mutual fund is invested in Canadian equity, this justifies using a Canadian equity index to perform the hedge. Thus, S is presumed to be the futures prices of the TSX 60 index.

Estimation results

The bivariate regime-switching model (2.3.1) is estimated using maximum likelihood. The log-likelihood function is computed through Hamilton (1989)'s filter and maximized using standard global optimization routines. For the first 66 months, the estimation methodology is based on the marginal likelihood function of the Great-West fund as the TSX 60 futures index was launched later (in September, 1999). For all subsequent months the joint density of both time series is considered. Maximum likelihood estimation results are presented in Table 2.1, where it can be seen that the first regime is characterized by positive expected returns and low volatility (bull market), whereas the second regime describes a state of higher volatility and negative expected returns (bear market). It is interesting to note that the correlation of the mutual fund and index futures returns is high in each regime ; it is $\rho_1 = 94.39\%$ in the bull market regime, which is slightly higher than in the bear market regime where it is $\rho_2 = 90.68\%$. Nevertheless, basis risk is not nil. The higher correlation in bull markets is surprising since all asset values are usually expected to depreciate simultaneously during financial crises. A possible explanation could be that, as stated in Robidoux (2015), the tactical (short-term) asset allocation within the mutual fund could significantly differ from the strategic (long-term) asset allocation target during specific market circumstances, for instance a flight to quality during a crisis, and thus the correlation structure could be altered in this situation. This observation entails that insurers should not rely on the expectation that basis risk will dampen during stress periods when hedging is most needed.

Table 2.1 – Maximum likelihood estimation results for the bivariate lognormal two-state regime-switching model of (2.3.1). The first component is Great-West Life Canadian Equity (GWLIM) BEL fund and the second is the TSX 60 index futures.

$\mu_1^{(j)}$	$\sigma_1^{(j)}$	$\mu_2^{(j)}$	$\sigma_2^{(j)}$
Great-West Life Canadian Equity (GWLIM) BEL ($j = F$)			
0.0084 (0.0024)	0.0330 (0.0019)	-0.0080 (0.0104)	0.0734 (0.0081)
TSX 60 index futures ($j = S$)			
0.0085 (0.0026)	0.0348 (0.0022)	-0.0134 (0.0126)	0.0858 (0.0097)
Correlations		*Transition matrix*	
ρ_1	ρ_2	$P_{1,1}$	$P_{2,1}$
0.9439 (0.0090)	0.9068 (0.0269)	0.9767 (0.0137)	0.0850 (0.0527)

Note : Standard errors are given in parentheses.

Baseline parameters

The baseline parameters of the simulation study are presented in Table 2.1 and Table 2.2. Policyholders aged 55 years at time $t = 0$ purchasing an at-the-money GMMB variable annuity with maturity of $T = 120$ months (10 years) are considered. The survival probabilities are obtained following the methodology recommended by the Canadian Institute of Actuaries ; base mortality rates are obtained from table CPM2014, see CIA (2014), and mortality improvements are projected with their proposed rates, see Appendix C of CIA (2010). Monthly mortality rates are obtained from annual rates by assuming that the force of mortality is constant within a given year. The parameters related to asset dynamics are taken from the maximum likelihood estimation results of Table 2.1. Other parameters are deemed representative of real-life practice. For instance, the utilized lapse rate whose annualized value is roughly 4% is consistent with lapse rates presented in Ledlie et al. (2008) which range between 2% and 6%.

Performance measures

The various local hedging strategies are benchmarked in terms of capital requirements. As explained in Section 2.2.2, the Total Gross Capital Required (TGCR) that must be held at time $t = 0$ by insurance companies in Canada can be modeled by the $\text{CVaR}_{0.95}$ of the discounted sum of injections. Moreover, the $\text{CVaR}_{0.80}$ is recommended to determine reserves :

$$\text{TGCR} = \text{CVaR}^{\mathbb{P}}_{0.95}\left[\sum_{t=1}^{T} e^{-rt} I_t\right], \qquad \text{Reserve} = \text{CVaR}^{\mathbb{P}}_{0.80}\left[\sum_{t=1}^{T} e^{-rt} I_t\right].$$

The main performance metric used in this work is the TGCR as defined above. In particular, hedging strategies that can be optimized to minimize capital requirements (see Sections 2.3.4 and 2.3.5) are implemented under this definition. Note however that such approaches can be generalized to any capital measurement criterion that is based on the discounted sum of injections.

Table 2.2 – Baseline parameters in monthly frequency.

Maturity (in months)	T	120
Survival probability	${}_tp_{660}$	Projected CPM2014
Lapse rate	b	0.34%
Total fee rate	ω_{tot}	0.29%
Risk-free rate	r	0.25%
GMMB guarantee	K	100
Initial value of F	F_0	100
Initial value of S	S_0	100

In the U.S., according to AAA (2011), the $\text{CVaR}_{0.90}^{\mathbb{P}}$ and the $\text{CVaR}_{0.70}^{\mathbb{P}}$ are recommended to quantify capital requirements. These values are therefore also presented. Moreover, the $\text{CVaR}_{0.99}^{\mathbb{P}}$ is given to detect potential flaws of the hedging strategies in terms of heavy tail risk.

Hedging strategies

This section presents the hedging strategies considered in the simulation experiment of Table 2.3, which will be analyzed in the next section.

The “mean-variance" strategy entails the minimization of a tradeoff between the variance and the expected value of the next cash injection. It is based on the local hedging problem of (2.3.15), the solution to which is given by Proposition 2.3.2. The value of the tradeoff parameter λ minimizing the TGCR is determined with the approach of Section 2.3.4 and is marked by a star ($\star$). The minimal variance strategy corresponds to the special case $\lambda = 0$.

The “minimal $\text{VaR}_{\alpha}^{\mathbb{P}}$" and the “minimal $\text{CVaR}_{\alpha}^{\mathbb{P}}$" strategies entails minimizing the risk concerning the next injection, based on the risk measures they are named after. These strategies solve the hedging problem of (2.3.23) where the risk measure $\mathcal{R}_t(\bullet)$ is $\text{VaR}_{\alpha}^{\mathbb{P}}[\bullet|\mathcal{F}_t]$ and $\text{CVaR}_{\alpha}^{\mathbb{P}}[\bullet|\mathcal{F}_t]$, respectively. Their solutions are calculated offline using Theorem 2.3.1. For both strategies, the parameter $\alpha \in [0, 1]$ controls the tradeoff between risk and cost minimization ; higher values of α encourage a pure risk reduction. The chosen values of α in the table are restricted to the range within which these strategies are well-behaved ; lower values of α can lead to a non-finite number of futures positions (unbounded solutions) and are therefore avoided.

The “minimal TGCR" strategy refers to the one described in Section 2.3.5. Contrary to the other hedging strategies, this one does not involve minimizing the risk of the next injection. Instead, it is based on the TGCR minimization problem of (2.3.25) under the polynomial model (2.3.26), which yields the numerical optimization problem (2.3.29). A polynomial of degree $n = 8$ is deemed satisfactory for the example currently considered ; in unreported tests, performance results (i.e., see Table 2.3) obtained using polynomials of degree higher than 8 are virtually identical to those obtained with the degree 8 polynomial.

For comparison, the results when no hedging is used are also given : $\theta_t^{(S)} = 0$ for all $t \in \mathcal{T}$, which is equivalent to the absence of a hedging portfolio. Furthermore, each hedging strategy is implemented in an hypothetical ideal case in which there is no basis risk. This is done by supposing that there exist futures contracts on the underlying mutual fund F. Although this is not the case in real life, such strategies are nevertheless implemented for the sake of our numerical study which aims at quantifying the impact of basis risk.

2.4.2 Results

The results for the setup outlined in the preceding section are presented in Table 2.3. These are obtained from 50,000 Monte Carlo simulation runs of the hedging strategies.

First, let's analyse the case of hedging under no basis risk. It can be seen that the strategies that are best-performing in terms of TGCR are those characterized by a small standard deviation, i.e., those entailing pure dispersion minimization. In particular, the minimal variance strategy ($\lambda = 0$) virtually coincides with minimal VaR and minimal CVaR strategies. Note that under the assumption of absence of basis risk, the minimal variance hedging strategy collapses to the delta-hedging approach widely used in the industry. Moreover, hedging strategies designed to minimize the TGCR do not substantially further improve the results obtained with the minimal variance strategy. For instance, the optimal mean-variance ($\lambda = 1.5$) and the minimal TGCR strategies both attain a TGCR of 3.7, which is quite close to the value of 4.7 obtained under the minimal variance strategy. These strategies constitute an important improvement over the no-hedging strategy, for which the TGCR is 25.3. Compared to the minimal variance strategy, mean-variance strategies with a higher value of the tradeoff parameter λ lead to higher TGCR values. As discussed below Proposition 2.3.2, the minimal variance strategy actually corresponds to standard delta hedging as there is no basis risk. Hence, the above results show that delta hedging is quite efficient in the absence of basis risk.

The results for the case of hedging under basis risk are richer and more subtle. Minimal variance, minimal VaR, and minimal CVaR strategies do not coincide with each other as they did in the absence of basis risk. Moreover, the strategies which perform the best in terms of TGCR are not necessarily those which attain a small standard deviation ; the minimal variance strategy in fact yields the worst TGCR. The minimal TGCR strategy and the optimal mean-variance ($\lambda = 7$) respectively attain a TGCR of 8.7 and 8.6. Note that the TGCR of the mean-variance strategy can be lower than the one pertaining to the minimal TGCR strategy as the mean-variance risk measure is not included in the family of reducible risk measures. This result shows that the optimal mean-variance strategy performs at least as well as the best strategy based on reducible risk measures. This is a surprising result as the latter class of hedging strategies is very large and encompasses other forms of risk-return tradeoffs such as the mean-standard deviation, mean-VaR and mean-CVaR. Moreover the optimal mean-variance strategy leads to a value of the $\text{CVaR}^{\mathbb{P}}_{0.99}$ that is smaller than for the minimal variance strategy (17.2 vs. 18.7). The results therefore show no evidence that reducing capital requirements comes at the expense of higher tail risk for levels beyond 95%. Nevertheless, care should be applied when interpreting this result as it is not impossible that risk could be increased in the far tail at levels higher than 99%. It is also interesting to note that the optimal mean-variance strategy leads to both a smaller CVaR and a better expected value ; only the variance is worse compared to the minimal variance hedging strategy. Indeed, the CVaR is also a form of risk-reward tradeoff, and therefore it is possible that it can be lowered simultaneously with the expected value even if this entails a higher variance. The variance increase observed when applying the optimal mean-variance hedge instead of the minimal variance hedge is caused by an increase in upside risk for the insurer, which is a desirable feature.

Another consideration that is worthy of emphasizing is the large difference in capital requirements obtained when comparing the respective cases where basis risk is absent or present. For instance, the TGCR of the minimal variance strategy is 4.7 in the absence of basis risk versus 14.8 when basis risk is considered, which is a 215% increase. Even for an optimized hedging strategy, capital requirements jump from 3.7 to 8.6 (optimal mean-variance strategy) or 8.7 (minimal TGCR strategy) when basis risk is incorporated. Indeed, Proposition 2.3.3 shows that even a small amount of basis risk can lead to a substantial loss of hedging performance. Omitting basis risk in hedging schemes performance assessments could therefore lead to severe risk under-estimation. Results of this nature could have been obtained through a poor hedging instrument choice ; basis risk will be very important if the mutual fund behaves very differently than the hedging asset. However, the statistics presented in Table 2.1 (i.e., the correlation coefficients) indicate that this is not the case here. Such high correlations make it difficult to believe that the insurer could significantly improve upon the presented hedge by choosing different hedging instruments.

The above discussion can be summarized briefly. In the absence of basis risk, conditional variance minimization coincides with delta hedging and is very efficient at reducing the TGCR. In contrast, under basis risk, the strategies attaining the lowest TGCR values are those which put some weight on expected cost minimization. These results ask for further analysis to shed light on the mechanics and determinants of the optimal hedging strategy under basis risk.

Unveiling the risk mitigation mechanics

With the exception of the mean-variance strategy, the hedging strategies in Table 2.3 can be expressed as $\theta^{(S)}_{t+1} = \Psi(\eta^{\mathbb{P}}_{1,t})\frac{\Delta_t F_t}{S_t}$, and are therefore fully characterized by their function Ψ. These are illustrated in Figure 2.1 and allow for a straightforward interpretation of the risk mitigation mechanics. For the minimal variance strategy, the function Ψ is roughly constant and positive, implying the use of short positions only (because $\Delta_t \leq 0$ a.s.). The minimal $\text{VaR}^{\mathbb{P}}_\alpha$ and minimal $\text{CVaR}^{\mathbb{P}}_\alpha$ strategies also rely almost exclusively on short positions. In contrast, the minimal TGCR strategy uses long positions when the conditional probability of the bull market regime is above a certain threshold. In other words, this strategy uses the futures as an investment vehicle, as opposed to a hedging instrument, in bull market time periods. This behavior is also found in the mean-variance strategy, as shown in Figure 2.2 for a simulated trajectory of the hedging portfolio composition. These observations highlight the fact that there are actually two ways for the insurer to meet futures variable annuity liabilities :

1. The insurer can use the futures contract as a hedging asset, which entails shorting it to cover the long position in the underlying mutual fund. Doing this reduces the conditional variance of the cash flow injections as shown in Proposition 2.3.3. This strategy is however costly in bull markets because it involves shorting an asset whose price grows on average. Moreover, hedging should intuitively be less needed in bull market because the underlying mutual fund also grows on average.
2. The insurer can also invest money in capital markets through long positions in the TSX 60 futures contract. Such risky investments benefit from time diversification of risk, making them smarter choices than the risk-free asset if the time horizon is sufficiently long. Time diversification refers to the imperfect correlation between all remaining futures log-returns until maturity ; investing through a futures over a long horizon therefore reduces risk associated with the latter position when compared to a short-term investment.

Note that other investment vehicles than long equity futures positions could have been considered. Such positions could be replaced by long positions in general investment portfolios targeting long-term growth. Low-volatility funds and risk-managed funds could for instance be considered as these are designed to provide decent returns with low downside risk, which are in line with the use of investment in the current hedging framework. Such extensions are left as further work.

Time diversification

This section explains why a large (small) value of the mean-variance tradeoff parameter λ is optimal when basis risk is present (absent). The starting point is Remark 2.3.1, which states that the injection under the mean-variance strategy with parameter λ can be expressed as

$$\tilde{I}_{t+1} = \tilde{I}^{\text{MV}}_{t+1} + \lambda J_{t+1}, \qquad J_{t+1} \equiv -\frac{\mathbb{E}^{\mathbb{P}}[\delta S_t | \mathcal{F}_t]}{\text{Var}^{\mathbb{P}}[S_{t+1} | \mathcal{F}_t]} \delta S_t, \tag{2.4.1}$$

where $\tilde{I}^{\text{MV}}_{t+1}$ is the injection under the minimal variance strategy. Hence, the mean-variance injection can be represented as a departure "$+\lambda J_{t+1}$" from the minimal variance injection. The above definition implies that $\mathbb{E}^{\mathbb{P}}[J_{t+1}] \leq 0$, so this departure indeed implies a reduction in the expected injection value. Note that it can also be expressed as

$$J_{t+1} \equiv -\frac{\mathbb{E}^{\mathbb{P}}[\delta S_t / S_t | \mathcal{F}_t]}{\text{Var}^{\mathbb{P}}[S_{t+1} / S_t | \mathcal{F}_t]} \frac{\delta S_t}{S_t},$$

which shows that the $\mathcal{F}_t$-conditional distribution of J_{t+1} depends only on $\eta^{\mathbb{P}}_{1,t}$; this is because the $\mathcal{F}_t$-conditional distribution of $\frac{\delta S_t}{S_t}$ depends only on $\eta^{\mathbb{P}}_{1,t}$, as it can be seen from (2.3.2). Furthermore, one can show from the above equation that

$$\mathbb{P}(J_{t+1} > 0 | \mathcal{F}_t) = \sum_{i=1}^{2} \eta^{\mathbb{P}}_{i,t} \left[\Phi\left(\frac{\mu_i^{(S)}}{\sigma_i^{(S)}} \right) \mathbb{1}_{\left\{ \mu_i^{(S)} + \frac{1}{2}\left(\sigma_i^{(S)}\right)^2 < 0 \right\}} + \Phi\left(-\frac{\mu_i^{(S)}}{\sigma_i^{(S)}} \right) \mathbb{1}_{\left\{ \mu_i^{(S)} + \frac{1}{2}\left(\sigma_i^{(S)}\right)^2 > 0 \right\}} \right].$$

In particular, with parameters presented in Table 2.1,

$$\mathbb{P}(J_{t+1} > 0|\mathcal{F}_t) \in [40.0\%, 45.7\%],$$

which means that the departure from the minimal variance strategy entails at least a 40% conditional probability of ending up with a higher injection value. A myopic risk manager could be temped to stick to the minimal variance strategy to reduce the injections volatility. Potential advantages of a mean-variance tradeoff are however revealed when the entire time horizon is considered ; under the baseline parameters of Tables 2.1 and 2.2, one can show through Monte Carlo simulations that

$$\mathbb{P}\left(\sum_{t=1}^{T} J_t e^{-rt} > 0\right) = 2.8\%, \tag{2.4.2}$$

which is because downside risk is reduced through time diversification.

A more formal discussion can be drawn up based on what insurers are interested in : discounted sum of injections. From (2.4.1),

$$\sum_{t=1}^{T} \tilde{I}_t e^{-rt} = \sum_{t=1}^{T} \tilde{I}_t^{\text{MV}} e^{-rt} + \lambda \sum_{t=1}^{T} J_t e^{-rt}. \tag{2.4.3}$$

This provides the last piece required to explain the simulation results :

– Suppose the insurer performing the hedge is not confronted with basis risk ; futures on the fund F are available for hedging. The conditional variance of the minimal variance injection approximation $\tilde{I}_{t+1}^{\text{MV}}$ is completely eliminated, as shown by Proposition 2.3.3. Moreover, this is done at no net cost because the price of risk is the same for the underlying fund and the hedging instrument. This explains why the first sum on the right-hand side of (2.4.3) is distributed around zero with a very small dispersion, and why a near-zero TGCR is thus obtained for the choice $\lambda = 0$ (see Table 2.3). This makes it optimal to use such a small value of λ, as the TGCR would be increased by the second sum despite the fact that its downside risk is reduced through time diversification.

– Suppose the insurer is confronted with basis risk ; futures S are used to hedge the GMMB contract with the underlying fund F. The minimal variance injection still contains an important portion of the no-hedging conditional standard deviation. In fact, Proposition 2.3.3 shows that for a correlation of around 90%, only about 56% of the no-hedging standard deviation is eliminated by the minimal variance strategy. The first sum on the right-hand side of (2.4.3) is therefore not distributed with a very small dispersion as it would be in the absence of basis risk. Hence, it can be optimal to use a large value of λ as the second sum can now help in reducing the CVaR through time diversification, as shown by (2.4.2).

The above explanation can be summarized as follows : the minimal variance strategy is less efficient under basis risk, as shown by Proposition 2.3.3, and moreover it neglects the risk reduction offered by time diversification through the long maturity of variable annuities. This explains why a large value of the mean-variance tradeoff parameter λ is optimal when there is basis risk.

2.5 Robustification against drift uncertainty

A surprising observation from Table 2.3 is that the optimal mean-variance hedge is more profitable in average than the no-hedging strategy (see the expected error of -22.8 vs -6.5). This is at odds with the traditional premise that hedging reduces risk at the expense of a lower average return. The lower expected return under the absence of hedging results from the optimal hedging strategy which promotes an aggressive investment into the index futures in order to take advantage of the excess equity growth rate of the asset (under $\mathbb{P}$) over the risk-free rate that is used to discount the capital injections, at least when the market is very likely to be in the bull market regime. This is illustrated by a simulated hedging portfolio composition path plotted in Figure 2.2. The use of long futures positions within the hedging portfolio during bull markets raises several concerns. The hedging strategy exhibits regime-timing behavior which could be deemed undesirable, both from the point of view of risk management and of the regulator. Such a strategy relies on the theoretical ability to forecast the conditional drift through the regime mass function η. In practice, model risk and parameter uncertainty could be very detrimental to the performance of the presented hedging scheme due to the inability of efficiently forecasting drifts, which is a notoriously difficult problem. For instance, the standard errors of the bear market drifts $\mu_2^{(S)}$ and $\mu_2^{(F)}$ displayed in Table 2.1 are very large. Moreover, the statistical uncertainty related to the regime transition probability $P_{2,1}$ in the bear market regime is substantial as indicated by its large standard error provided in Table 2.1. Hence, it is important to investigate whether risk-reward tradeoff based strategies remain useful in such a context of drift uncertainty.

The current section presents a robust version of the risk-reward tradeoff strategies developed in the previous section which does not rely on the ability to accurately forecast the time-variation of the drifts. Although there exists multiple approaches to embed model risk and estimation risk into the hedging strategy, a thorough investigation of other schemes are left as further work. In the robust version, during the optimization of the hedging strategies (see (2.3.16)-(2.3.18) for the mean-variance approach and (2.3.25) for the minimal TGCR approach), the physical measure $\mathbb{P}$ is replaced by a new probability measure $\mathbb{Z}$ under which the drift is time-invariant. More precisely, the constrained probability measure $\mathbb{Z}$ is such that the asset dynamics remains the same lognormal two-state regime-switching model than under $\mathbb{P}$, but with the following drift parameters : $\tilde{\mu}_1^{(S)} = \tilde{\mu}_2^{(S)} = \bar{\mu}^{(S)}$ and $\tilde{\mu}_1^{(F)} = \tilde{\mu}_2^{(F)} = \bar{\mu}^{(F)}$, where

$$\bar{\mu}^{(S)} \equiv \pi_1 \, \mu_1^{(S)} + (1-\pi_1)\mu_2^{(S)} = 0.0037, \qquad \bar{\mu}^{(F)} \equiv \pi_1 \, \mu_1^{(F)} + (1-\pi_1)\mu_2^{(F)} = 0.0048 \quad (2.5.1)$$

are the stationary expected returns, and $\pi_1 = 0.78$ is the stationary probability associated with regime 1.

Calculating the hedging strategies under $\mathbb{Z}$ entails that such strategies do not have the ability to forecast expected returns, in line with drift uncertainty concerns as discussed above. Such strategies are referred to as the *drift-constrained* counterparts. Note that the ability to forecast volatility is preserved. Indeed, there is a large literature documenting the ability of various frameworks to forecast volatility under model/parameter uncertainty (see, e.g., Ardia et al., 2017). The robustification procedure therefore focuses on drift uncertainty.

A realization of the drift-constrained mean-variance strategy is illustrated in Figure 2.3. The reducible drift-constrained strategies are illustrated in Figure 2.4. The drift-constrained strategies almost exclusively use net short positions in the futures, in contrast with their unconstrained counterparts. They represent an under-hedge compared to the minimal variance strategy, i.e., they rely on a long temporal horizon to use temporal diversification so as to save on the cost of shorting an asset whose value grows on average. Such strategies therefore address the concern of model risk since they rely on a more conservative investment component than the unconstrained strategies.

The performance of the robust strategies was assessed by simulating the hedging process by applying the drift-constrained strategies optimized under $\mathbb{Z}$ over underlying asset paths simulated with the physical data-generation measure $\mathbb{P}$. The hedging simulation results obtained are presented in Table 2.4. The optimal value of λ is smaller for the drift-constrained mean-variance strategy than for the unconstrained version optimized under $\mathbb{P}$ (4.4 vs. 7). The robust version of the strategy is therefore less aggressive and puts less weight on the mean component than its unconstrained counterpart. Nevertheless, even when considering the robust version of the hedging strategy, departing from the minimal variance strategy still provides the opportunity to both increase profitability through a lower injections mean and reduce capital requirements through a lower TGCR. The robust version of the hedging strategy should therefore be very attractive to practitioners concerned with model risk who wish to increase expected returns without increasing capital requirements. However, for the robust strategy, the expected hedging error is higher than for the no-hedging strategy (-4.7 vs -6.5). This result is therefore consistent with the traditional premise stipulating that hedging reduces capital requirements at the expense of lesser profitability.

2.6 Conclusion

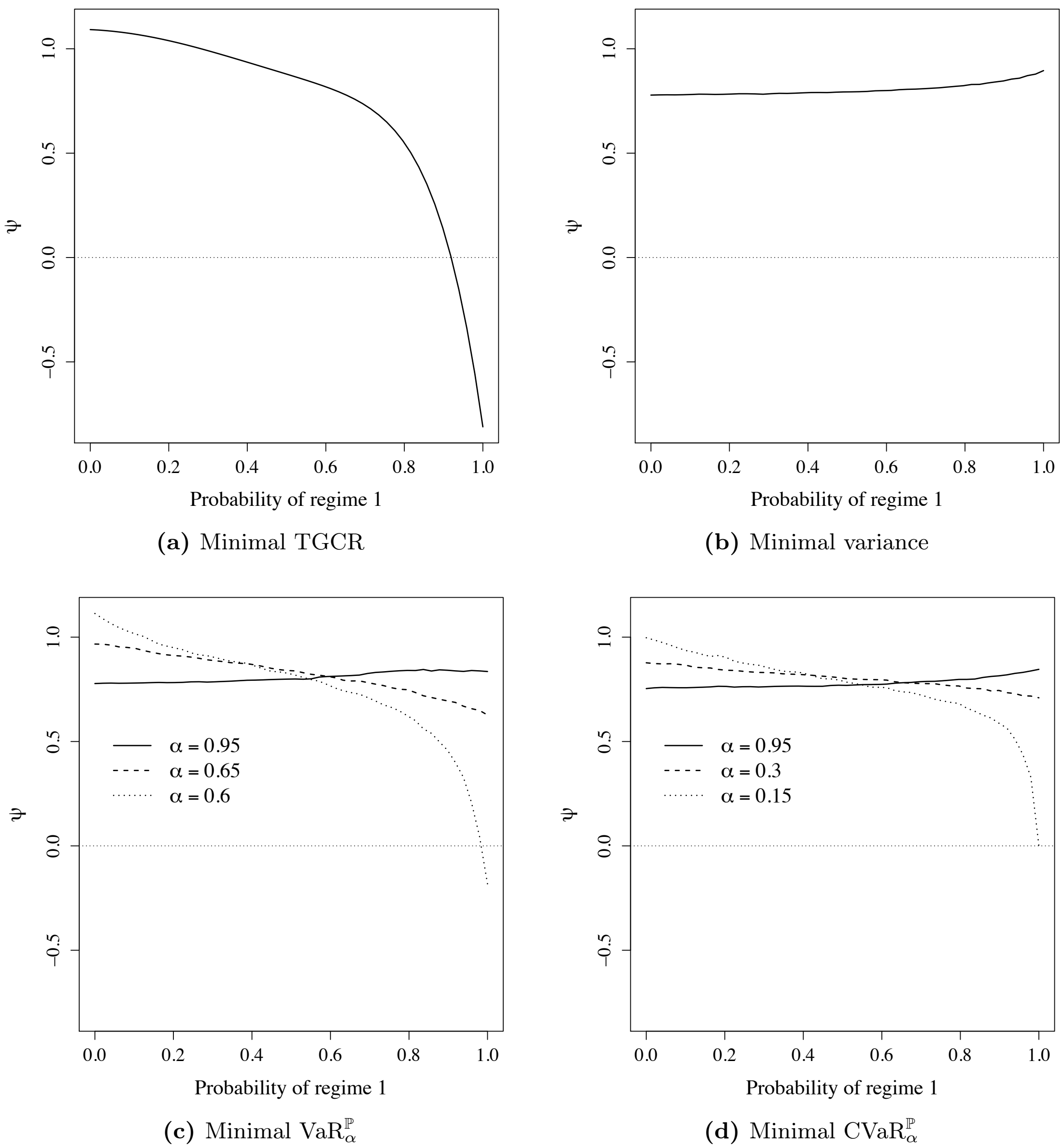
An efficient and tractable methodology is developed for insurers hedging equity risk related to guarantees associated with variable annuity polices in the presence of basis risk. Although the optimization criterion is local, a global flavor is given to the hedge as the local risk measure is optimized to minimize capital requirements. This enables the local hedging approach incorporating time diversification of risk into its design. Taylor expansions on liquidity injections are applied to simplify the implementation of the methodology. Such approximations lead to a family of solutions to the hedging problem encompassing multiple hedging approaches found in the literature such as delta hedging and minimal local variance hedges. Multivariate regime-switching models are considered for the joint dynamics of the guarantee underlying asset and the index futures.

Within simulation experiments, our method is compared to benchmarks drawn from the literature such as minimal variance hedging. The outperformance of our method versus benchmarks in terms of both capital requirements reduction and expected return enhancement is explained by the time diversification of risk associated to additional exposure to equity risk resulting from the optimal local criterion ; an additional exposure to equity risk proves beneficial in the long-term since a higher expected return facilitates the accumulation of capital and since the impact of downturns is dampened over a long horizon. The omission of basis risk within hedging performance assessment simulations is shown to lead to a severe under-estimation of risk. For instance, in the simulation experiment presented the inclusion of basis risk within the hedging scheme leads to a drastic increase of 215% of the Total Gross Capital Required (TGCR) associated with a minimal variance strategy.

Table 2.3 – Comparison of various hedging strategies for the simulation study of Section 2.4.1. The statistics are for the discounted sum of injections : $\sum_{t=1}^{T} e^{-rt} I_t$.

	Mean	Std.Dev.	$\text{CVaR}^{\mathbb{P}}_{0.70}$	$\text{CVaR}^{\mathbb{P}}_{0.80}$	$\text{CVaR}^{\mathbb{P}}_{0.90}$	$\text{CVaR}^{\mathbb{P}}_{0.95}$	$\text{CVaR}^{\mathbb{P}}_{0.99}$
No hedging							
	-6.5	11.7	8.7	14.0	20.7	25.3	32.2
Hedging under basis risk							
Mean-variance							
$\lambda = 0$	2.9	4.9	8.9	10.5	12.8	14.8	18.7
$\lambda = 2$	-4.4	6.7	3.8	5.7	8.5	10.8	15.1
$\lambda = 5$	-15.4	11.6	-1.7	1.2	5.4	8.9	15.5
$\lambda = 7 \star$	-22.8	15.3	-4.9	-1.2	4.1	8.6	17.2
$\lambda = 10$	-33.8	21.0	-9.3	-4.4	2.8	8.9	20.5
Minimal TGCR							
$n = 8$	-15.1	10.0	-3.2	-0.1	4.6	8.7	16.9
Minimal $\text{VaR}^{\mathbb{P}}_{\alpha}$							
$\alpha = 0.95$	2.4	4.9	8.4	9.9	12.3	14.3	18.4
$\alpha = 0.60$	-7.3	6.8	0.8	3.0	6.3	9.2	15.2
Minimal $\text{CVaR}^{\mathbb{P}}_{\alpha}$							
$\alpha = 0.95$	2.4	5.0	4.3	6.3	9.2	11.7	16.0
$\alpha = 0.15$	-3.7	5.6	3.1	4.9	7.7	10.2	15.1
Hedging without basis risk							
Mean-variance							
$\lambda = 0$	0.3	1.7	2.4	2.9	3.8	4.7	6.6
$\lambda = 1.5 \star$	-2.5	5.9	0.9	1.6	2.7	3.7	6.0
$\lambda = 5$	-9.2	7.4	-0.7	1.0	3.5	5.6	9.6
$\lambda = 7$	-13.1	10.1	-1.5	0.9	4.2	7.1	12.5
$\lambda = 10$	-18.9	14.1	-2.5	0.7	5.5	9.5	17.1
Minimal TGCR							
$n = 8$	-1.7	2.3	1.2	1.8	2.8	3.7	5.7
Minimal $\text{VaR}^{\mathbb{P}}_{\alpha}$							
$\alpha = 0.95$	0.3	1.7	2.4	2.9	3.8	4.7	6.6
$\alpha = 0.60$	0.3	1.7	2.4	2.9	3.8	4.7	6.6
Minimal $\text{CVaR}^{\mathbb{P}}_{\alpha}$							
$\alpha = 0.95$	0.3	1.7	2.4	2.9	3.8	4.7	6.6
$\alpha = 0.15$	0.3	1.7	2.4	2.9	3.8	4.7	6.6

Notes : The "mean-variance" strategy is characterized by Proposition 2.3.2. The value of λ minimizing capital requirements is marked by a star ($\star$) and is determined with the approach of Section 2.3.4. The "minimal TGCR" strategy refers to the one described in Section 2.3.5. The "minimal $\text{VaR}^{\mathbb{P}}_{\alpha}$" and the "minimal $\text{CVaR}^{\mathbb{P}}_{\alpha}$" strategies are characterized by Theorem 2.3.1 where the risk measure $\mathcal{R}_t(\bullet)$ is $\text{VaR}^{\mathbb{P}}_{\alpha}[\bullet|\mathcal{F}_t]$ and $\text{CVaR}^{\mathbb{P}}_{\alpha}[\bullet|\mathcal{F}_t]$, respectively. The "Hedging without basis risk" pannel refers to a ficticious case where the futures underlying asset is the mutual fund whereas the "Hedging under basis risk" pannel integrates imperfect correlation between the futures underlying asset and the mutual fund.



(a) Minimal TGCR

(b) Minimal variance

(c) Minimal $\mathrm{VaR}_\alpha^{\mathbb{P}}$

(d) Minimal $\mathrm{CVaR}_\alpha^{\mathbb{P}}$

FIGURE 2.1 – Representation of hedging strategies for the simulation study of Section 2.4.1. The "minimal TGCR" strategy refers to the one described in Section 2.3.5. The "minimal variance", "minimal $\mathrm{VaR}_\alpha^{\mathbb{P}}$", and "minimal $\mathrm{CVaR}_\alpha^{\mathbb{P}}$" strategies are characterized by Theorem 2.3.1 where the risk measure $\mathcal{R}_t(\bullet)$ is respectively $\mathrm{Var}^{\mathbb{P}}[\bullet|\mathcal{F}_t]$, $\mathrm{VaR}_\alpha^{\mathbb{P}}[\bullet|\mathcal{F}_t]$, and $\mathrm{CVaR}_\alpha^{\mathbb{P}}[\bullet|\mathcal{F}_t]$. Each strategy can be expressed as $\theta_{t+1}^{(S)} = \Psi(\eta_{1,t}^{\mathbb{P}})\frac{\Delta_t F_t}{S_t}$ where the functions Ψ are those illustrated. Curves above the zero level corresponds to short positions (because $\Delta_t \leq 0$ a.s.).

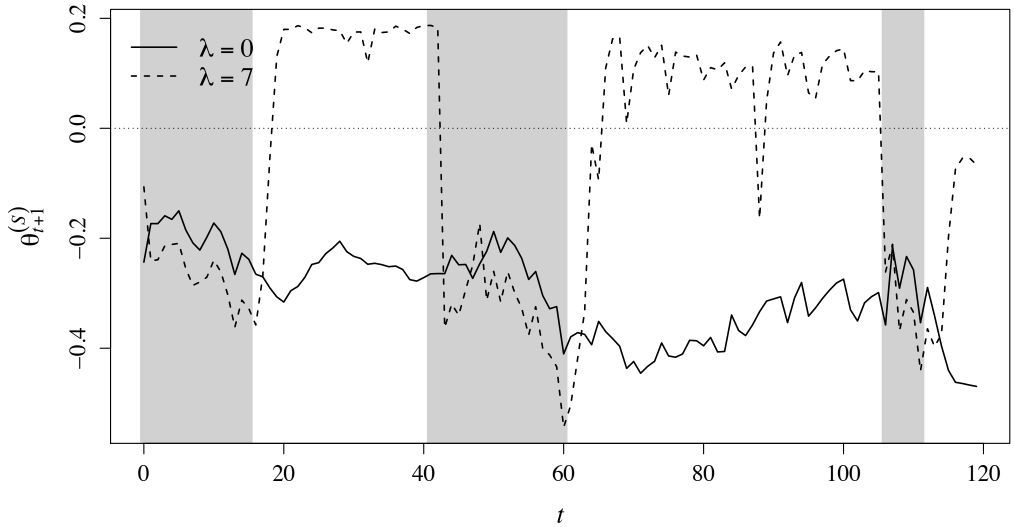


Figure 2.2 – Example of mean-variance hedging simulation in the setup of Section 2.4.1. The number of long positions in the futures is illustrated as a function of time (in months). Shaded areas correspond to bear market regime periods. One can see that the mean-variance strategy with $\lambda = 7$ combines both long and short positions. In contrast, the minimal variance strategy ($\lambda = 0$) relies solely on short positions.

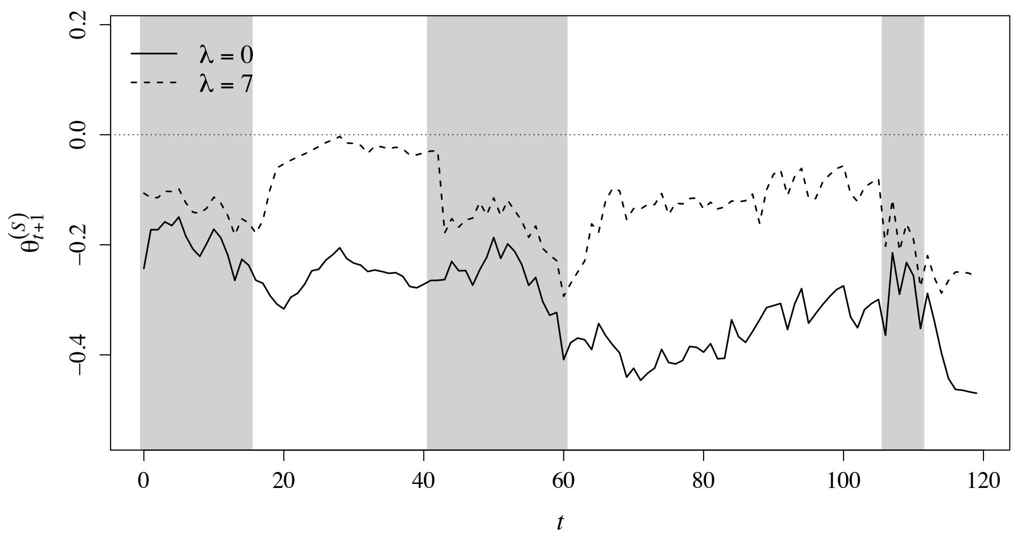
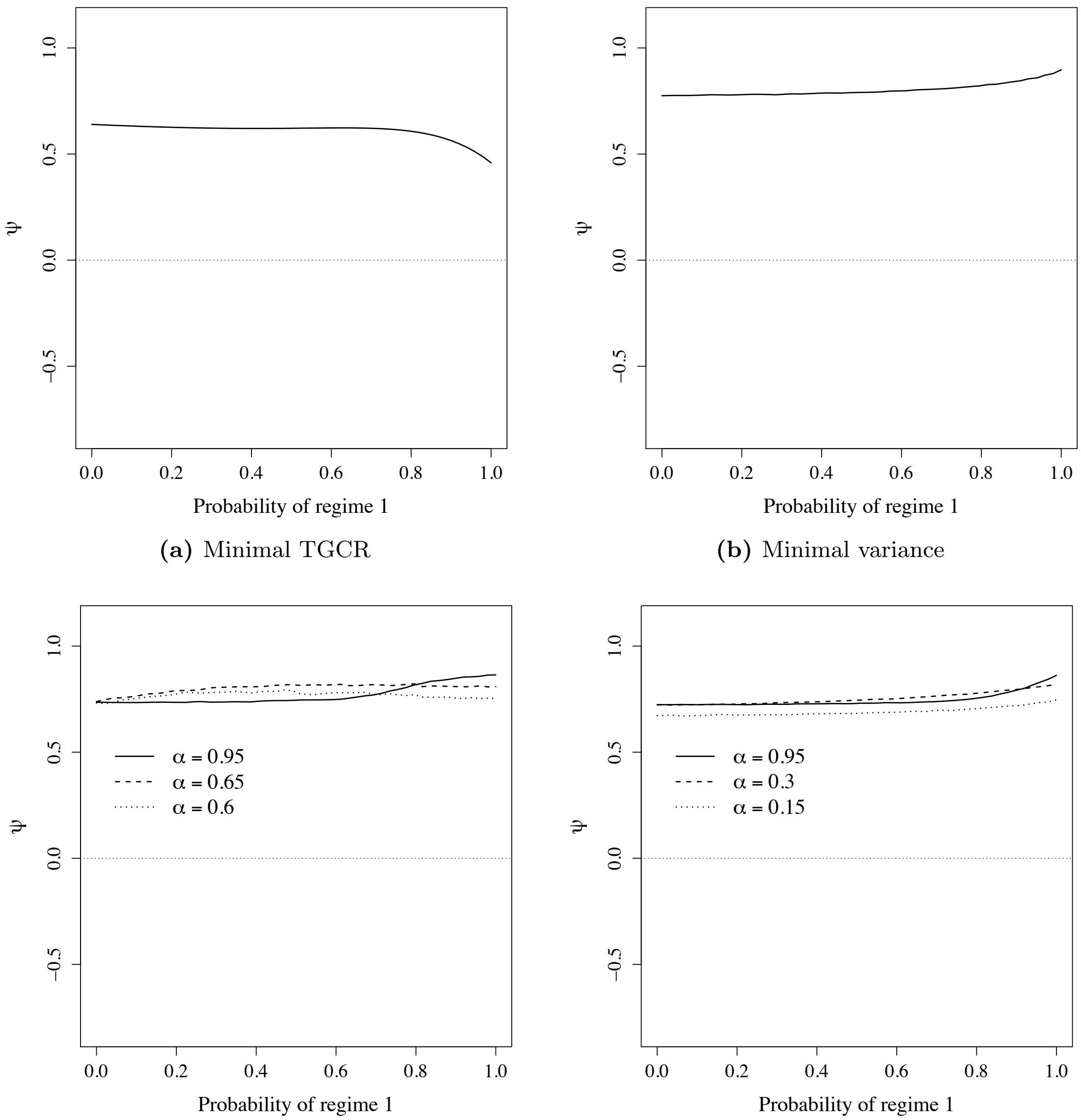


Figure 2.3 – Example of drift-constrained (see Section 2.5) mean-variance hedging simulation in the setup of Section 2.4.1. The number of positions in the futures is illustrated as a function of time (in months). Shaded areas correspond to bear market regime periods. One can see that the constrained mean-variance strategy represents an under-hedge compared to minimal variance hedging. Both strategies rely only on short positions, in contrast with Figure 2.2.



(a) Minimal TGCR **(b)** Minimal variance

(c) Minimal $\text{VaR}_\alpha^{\mathbb{Z}}$ **(d)** Minimal $\text{CVaR}_\alpha^{\mathbb{Z}}$

Figure 2.4 – Representation of the drift-constrained hedging strategies for the simulation study of Section 2.4.1. The strategies are all calculated under the measure $\mathbb{Z}$ described in Section 2.5. The "minimal TGCR" strategy refers to the one described in Section 2.3.5, but where the $\text{CVaR}_{0.95}^{\mathbb{P}}$ risk measure is replaced by $\text{CVaR}_{0.95}^{\mathbb{Z}}$. The "minimal variance", "minimal $\text{VaR}_\alpha^{\mathbb{Z}}$", and "minimal $\text{CVaR}_\alpha^{\mathbb{Z}}$" strategies are characterized by Theorem 2.3.1 where the risk measure $\mathcal{R}_t(\bullet)$ is respectively $\text{Var}^{\mathbb{Z}}[\bullet|\mathcal{F}_t]$, $\text{VaR}_\alpha^{\mathbb{Z}}[\bullet|\mathcal{F}_t]$, and $\text{CVaR}_\alpha^{\mathbb{Z}}[\bullet|\mathcal{F}_t]$. Each strategy can be expressed as $\theta_{t+1}^{(S)} = \Psi(\eta_{1,t}^{\mathbb{Z}})\frac{\Delta_t F_t}{S_t}$ where the functions Ψ are those illustrated. We note that only short positions are used for the drift-constrained strategies, in contrast with their basic counterparts (see Figure 2.1).

Table 2.4 – Comparison of drift-constrained hedging strategies (see Section 2.5) for the simulation study of Section 2.4.1. The statistics are for the discounted sum of injections : $\sum_{t=1}^{T} e^{-rt} I_t$.

	Mean	Std.Dev.	$\text{CVaR}^{\mathbb{P}}_{0.70}$	$\text{CVaR}^{\mathbb{P}}_{0.80}$	$\text{CVaR}^{\mathbb{P}}_{0.90}$	$\text{CVaR}^{\mathbb{P}}_{0.95}$	$\text{CVaR}^{\mathbb{P}}_{0.99}$
No hedging							
	-6.5	11.7	8.7	14.0	20.7	25.3	32.2
Hedging under basis risk							
Mean-variance							
$\lambda = 0$	3.0	5.0	9.0	10.6	12.9	15.0	19.0
$\lambda = 2$	-0.5	5.9	7.0	8.8	11.2	13.2	17.1
$\lambda = 4.4 \star$	-4.7	8.0	5.3	7.5	10.3	12.6	16.8
$\lambda = 7$	-9.2	10.7	3.9	6.6	10.3	13.2	18.7
$\lambda = 10$	-14.5	14.1	2.6	6.1	10.8	14.7	22.2
Minimal TGCR							
$n = 8$	-1.6	5.7	5.7	7.8	10.6	12.9	16.4
Minimal $\text{VaR}^{\mathbb{Z}}_{\alpha}$							
$\alpha = 0.95$	2.9	5.1	9.1	10.7	13.1	15.1	19.2
$\alpha = 0.60$	1.1	5.0	7.2	8.9	11.3	13.4	17.4
Minimal $\text{CVaR}^{\mathbb{Z}}_{\alpha}$							
$\alpha = 0.95$	2.4	5.1	8.7	10.3	12.7	14.7	18.6
$\alpha = 0.15$	1.3	5.2	7.9	9.6	12.1	14.2	18.0
Hedging without basis risk							
Mean-variance							
$\lambda = 0$	0.3	1.7	2.4	2.9	3.8	4.7	6.6
$\lambda = 1.3 \star$	-0.8	2.3	2.1	2.7	3.5	4.3	6.2
$\lambda = 4.4$	-3.4	4.5	2.1	3.2	4.6	5.8	8.1
$\lambda = 7$	-5.6	6.6	2.4	3.9	6.1	7.9	11.5
$\lambda = 10$	-8.2	9.1	2.7	4.9	7.9	10.6	15.8
Minimal TGCR							
$n = 8$	-0.8	2.3	2.2	2.9	3.7	4.4	5.8
Minimal $\text{VaR}^{\mathbb{Z}}_{\alpha}$							
$\alpha = 0.95$	0.3	1.7	2.4	2.9	3.8	4.7	6.6
$\alpha = 0.60$	0.3	1.7	2.4	2.9	3.8	4.7	6.6
Minimal $\text{CVaR}^{\mathbb{Z}}_{\alpha}$							
$\alpha = 0.95$	0.3	1.7	2.4	2.9	3.8	4.7	6.6
$\alpha = 0.15$	0.3	1.7	2.4	2.9	3.8	4.7	6.6

Notes : The strategies are all calculated under the measure $\mathbb{Z}$ described in Section 2.5. The "mean-variance" strategy is characterized by Proposition 2.3.2 where the measure $\mathbb{P}$ is replaced by $\mathbb{Z}$. The value of λ minimizing capital requirements is marked by a star ($\star$) and is determined with the approach of Section 2.3.4, with $\mathbb{P}$ replaced by $\mathbb{Z}$. The "minimal TGCR" strategy refers to the one described in Section 2.3.5, again with $\mathbb{P}$ replaced by $\mathbb{Z}$. The "minimal $\text{VaR}^{\mathbb{Z}}_{\alpha}$" and the "minimal $\text{CVaR}^{\mathbb{Z}}_{\alpha}$" strategies are characterized by Theorem 2.3.1 where the risk measure $\mathcal{R}_t(\bullet)$ is $\text{VaR}^{\mathbb{Z}}_{\alpha}[\bullet|\mathcal{F}_t]$ and $\text{CVaR}^{\mathbb{Z}}_{\alpha}[\bullet|\mathcal{F}_t]$, respectively. The "Hedging without basis risk" pannel refers to a ficticious case where the futures underlying asset is the mutual fund whereas the "Hedging under basis risk" pannel integrates imperfect correlation between the futures underlying asset and the mutual fund.

Appendix

2.A Proofs

2.A.1 Proof of Proposition 2.2.1

Define $\delta F_t = F_{t+1} - F_t$. From the GMMB pricing formula (2.2.7), for $t \in \{0, \ldots, T-1\}$,

$$\Pi_{t+1} - \Pi_t = -\omega_{opt}\delta F_t \sum_{j=t+1}^{T} \gamma_{j-1} + \omega_{opt}\gamma_t F_{t+1} + \mathbb{1}_{\{t+1<T\}}\gamma_T G_{t+1} - \mathbb{1}_{\{t<T\}}\gamma_T G_t. \tag{2.A.1}$$

Rewriting the injection formula (2.2.9) for I_{t+1} using (2.2.6), (2.2.10) and (2.2.12) yields

$$\begin{aligned} I_{t+1} &= V^{\theta}_{(t+1)+} + CF_{t+1} - V^{\theta}_{(t+1)-}, \\ &= \Pi_{t+1} - \omega_{opt}\gamma_t F_{t+1} + \mathbb{1}_{\{t+1=T\}}\gamma_T G_T - \theta^{(B)}_{t+1} B_{t+1} - \theta^{(S)}_{t+1}\delta S_t, \end{aligned}$$

where $G_T \equiv \max\big(0, \tilde{K} - F_T\big)$, and $\delta S_t = S_{t+1} - S_t$. Next, substitute Π_{t+1} and $\theta^{(B)}_{t+1}$ by the expressions prescribed by (2.A.1) and (2.2.11), respectively. This gives

$$\begin{aligned} I_{t+1} &= \Pi_t - \omega_{opt}\delta F_t \sum_{j=t+1}^{T} \gamma_{j-1} + \omega_{opt}\gamma_t F_{t+1} + \mathbb{1}_{\{t+1<T\}}\gamma_T G_{t+1} - \mathbb{1}_{\{t<T\}}\gamma_T G_t \\ &\quad -\omega_{opt}\gamma_t F_{t+1} + \mathbb{1}_{\{t+1=T\}}\gamma_T G_T - \frac{\Pi_t}{B_t} B_{t+1} - \theta^{(S)}_{t+1}\delta S_t, \\ &= \Pi_t(1 - e^r) - \omega_{opt}\delta F_t \sum_{j=t+1}^{T} \gamma_{j-1} - \theta^{(S)}_{t+1}\delta S_t + \gamma_T\big(\mathbb{1}_{\{t+1<T\}} G_{t+1} - \mathbb{1}_{\{t<T\}} G_t + \mathbb{1}_{\{t+1=T\}} G_T\big), \end{aligned}$$

where $\frac{B_{t+1}}{B_t} = e^r$ was used. At last, the identities

$$\mathbb{1}_{\{t+1=T\}} G_T = \mathbb{1}_{\{t+1=T\}} G_{t+1}, \qquad \mathbb{1}_{\{t+1<T\}} + \mathbb{1}_{\{t+1=T\}} = \mathbb{1}_{\{t+1\leq T\}} = \mathbb{1}_{\{t<T\}},$$

are used to obtain

$$I_{t+1} = \Pi_t(1 - e^r) - \omega_{opt}\delta F_t \sum_{j=t+1}^{T} \gamma_{j-1} - \theta^{(S)}_{t+1}\delta S_t + \gamma_T \mathbb{1}_{\{t<T\}}\delta G_t, \tag{2.A.2}$$

where $\delta G_t = G_{t+1} - G_t$. This concludes the proof.

2.A.2 Proof of Theorem 2.3.1

First, note that the approximation of the injection, see (2.3.14), can be written as

$$\tilde{I}_{t+1} = \Theta_t - \Delta_t F_t \Big[\psi_{t+1} \tfrac{\delta S_t}{S_t} - \tfrac{\delta F_t}{F_t}\Big], \qquad \psi_{t+1} \equiv \frac{\theta^{(S)}_{t+1} S_t}{\Delta_t F_t}. \tag{2.A.3}$$

Because the embedded option is a put, $-\Delta_t > 0$ almost surely. Hence, for any $\mathcal{F}_t$-reducible risk measure $\mathcal{R}_t$, there exists a real function f_1 and a nonnegative function f_2 such that

$$\mathcal{R}_t\big(\tilde{I}_{t+1}\big) = f_1(\Theta_t) + f_2(-\Delta_t F_t)\mathcal{R}_t\Big(\psi_{t+1} \tfrac{\delta S_t}{S_t} - \tfrac{\delta F_t}{F_t}\Big).$$

It follows that

$$\psi^*_{t+1} \equiv \underset{\psi_{t+1}}{\arg\min}\, \mathcal{R}_t\big(\tilde{I}_{t+1}\big) = \underset{\psi_{t+1}}{\arg\min}\, \mathcal{R}_t\Big(\psi_{t+1} \tfrac{\delta S_t}{S_t} - \tfrac{\delta F_t}{F_t}\Big), \tag{2.A.4}$$

from which (2.3.24) can indeed be concluded.

As stated in Proposition 2.3.1, the $\mathcal{F}_t$-conditional distribution of $\big(\frac{F_{t+1}}{F_t}, \frac{S_{t+1}}{S_t}\big)$ under $\mathbb{P}$ depends only on the first regime conditional likelihood, $\eta^{\mathbb{P}}_{1,t}$. It follows that the solution to the minimization problem (2.A.4) can be expressed as a function of the form $\psi^*_{t+1} = \Psi_{\mathcal{R}_t}\big(\eta^{\mathbb{P}}_{1,t}\big)$. Finally, from the temporal law-invariance property of the sequence $\{\mathcal{R}_t\}_{t=0}^{T-1}$, it follows that $\Psi_{\mathcal{R}_0} = \cdots = \Psi_{\mathcal{R}_{T-1}}$.

Explicit formulas

Deriving the following formulas under the regime-switching model of Section 2.3.1 is a straightforward exercice that involves conditioning on the state of the regime and using the moment generating function of the normal distribution :

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}[S_{t+1}|\mathcal{F}_t] &= S_t \sum_{j=1}^{2} \eta^{\mathbb{P}}_{j,t} e^{\mu^{(S)}_j + \frac{1}{2}\left(\sigma^{(S)}_j\right)^2}. \\
\mathrm{Var}^{\mathbb{P}}[S_{t+1}|\mathcal{F}_t] &= S_t^2 \left(\sum_{j=1}^{2} \eta^{\mathbb{P}}_{j,t} e^{2\mu^{(S)}_j + 2\left(\sigma^{(S)}_j\right)^2} - \left[\sum_{j=1}^{2} \eta^{\mathbb{P}}_{j,t} e^{\mu^{(S)}_j + \frac{1}{2}\left(\sigma^{(S)}_j\right)^2} \right]^2 \right). \qquad (2.A.5) \\
\mathrm{Cov}^{\mathbb{P}}[F_{t+1}, S_{t+1}|\mathcal{F}_t] &= S_t F_t \Bigg\{ \sum_{j=1}^{2} \eta^{\mathbb{P}}_{j,t} \left[e^{\mu^{(S)}_j + \mu^{(F)}_j + \frac{1}{2}\left(\sigma^{(S)}_j\right)^2 + \rho_j \sigma^{(S)}_j \sigma^{(F)}_j + \frac{1}{2}\left(\sigma^{(F)}_j\right)^2} \right] \qquad (2.A.6) \\
&\quad - \left(\sum_{j=1}^{2} \eta^{\mathbb{P}}_{j,t} e^{\mu^{(S)}_j + \frac{1}{2}\left(\sigma^{(S)}_j\right)^2} \right) \left(\sum_{j=1}^{2} \eta^{\mathbb{P}}_{j,t} e^{\mu^{(F)}_j + \frac{1}{2}\left(\sigma^{(F)}_j\right)^2} \right) \Bigg\}.
\end{aligned}$$

References

AAA (2011). The Application of C-3 Phase II and Actuarial Guideline XLIII. Variable Annuity Practice Note Work Group, American Academy of Actuaries.

Ankirchner, S., Schneider, J. C., and Schweizer, N. (2014). Cross-hedging minimum return guarantees : Basis and liquidity risks. *Journal of Economic Dynamics and Control*, 41 :93–109.

Ardia, D., Kolly, J., and Trottier, D.-A. (2017). The impact of parameter and model uncertainty on market risk predictions from GARCH-type models. *Journal of Forecasting*, 36(7) :808–823.

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3) :203–228.

Augustyniak, M. and 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(2) :183–206.

Augustyniak, M., Godin, F., and Simard, C. (2016). Assessing the effectiveness of local and global quadratic hedging under GARCH models. *Quantitative Finance*. Forthcoming.

Boudreault, M. and Augustyniak, M. (2015). On the importance of hedging dynamic lapses in variable annuities. *Risk and Rewards*, 66 :12–16.

Brandt, M. W. (2003). Hedging demands in hedging contingent claims. *The Review of Economics and Statistics*, 85(1) :119–140.

Chopra, D., Erzan, O., de Gantès, G., Grepin, L., and Slawner, C. (2009). Responding to the variable annuity crisis. *Mckinsey Working Papers on Risk*.

CIA (2010). Mortality Improvement Research Paper. Document 210065, Canadian Institute of Actuaries.

CIA (2014). Final Report: Canadian Pensioners Mortality. Document 214013, Canadian Institute of Actuaries.

Ederington, L. H. (1979). The hedging performance of the new futures markets. *The Journal of Finance*, 34(1) :157–170.

Elliott, R. J., Chan, L., and Siu, T. K. (2005). Option pricing and esscher transform under regime switching. *Annals of Finance*, 1(4) :423–432.

François, P., Gauthier, G., and Godin, F. (2014). Optimal hedging when the underlying asset follows a regime-switching Markov process. *European Journal of Operational Research*, 237(1) :312–322.

Gaillardetz, P., Li, H. Y., and MacKay, A. (2012). Equity-linked products : Evaluation of the dynamic hedging errors under stochastic mortality. *European Actuarial Journal*, 2(2) :243–258.

Godin, F. (2016). Minimizing CVaR in global dynamic hedging with transaction costs. *Quantitative Finance*, 16(3) :461–475.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2) :357–384.

Hardy, M. R. (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2) :41–53.

Hardy, M. R. (2003). *Investment Guarantees : Modeling and Risk Management for Equity-Linked Life Insurance*. John Wiley & Sons, Ltd.

iA Financial Group (2016). 2015 Annual report.

Kling, A., Ruez, F., and Ruß, J. (2014). The impact of policyholder behavior on pricing, hedging, and hedge efficiency of withdrawal benefit guarantees in variable annuities. *European Actuarial Journal*, 4(2) :281–314.

Ledlie, M. C., Corry, D. P., Finkelstein, G. S., Ritchie, A. J., Su, K., and Wilson, D. C. E. (2008). Variable annuities. *British Actuarial Journal*, 14(02) :327–389.

MacKay, A., Augustyniak, M., Bernard, C., and Hardy, M. R. (2017). Risk management of policyholder behavior in equity-linked life insurance. *Journal of Risk and Insurance*, 84(2) :661–690.

Manulife Financial Corporation (2016). 2015 Annual report.

Qian, L., Yang, H., and Wang, R. (2011). Locally risk-minimizing hedging strategies for unit-linked life insurance contracts under a regime switching lévy model. *Frontiers of Mathematics in China*, 6(6) :1185–1202.

Rémillard, B., Hocquard, A., Lamarre, H., and Papageorgiou, N. (2017). Option pricing and hedging for discrete time regime-switching models. *Modern Economy*, 8 :1005–1032.

Rémillard, B. and Rubenthaler, S. (2013). Optimal hedging in discrete time. *Quantitative Finance*, 13(6) :819–825.

Robidoux, B. (2015). Introduction to predictive modeling of fund manager behavior for variable annuities riders. *Predictive Analytics and Futurism*, 12 :24–26.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2 :21–42.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7) :1443–1471.

Rockafellar, R. T., Uryasev, S. P., and Zabarankin, M. (2002). Deviation measures in risk analysis and optimization. Working Paper, University of Florida.

Wang, S. S. (2000). A class of distortion operators for pricing financial and insurance risks. *The Journal of Risk and Insurance*, 67(1) :15–36.

Wang, Y. and Yin, G. (2012). Quantile hedging for guaranteed minimum death benefits with regime switching. *Stochastic Analysis and Applications*, 30(5) :799–826.

Zhang, F. (2010). Integrating robust risk management into pricing : New thinking for VA writers. *Risk and Rewards*, 55 :34–36.

Zhang, J., Tan, K. S., and Weng, C. (2017). Optimal hedging with basis risk under mean–variance criterion. *Insurance : Mathematics and Economics*, 75 :1–15.

Chapter 3

Option pricing under regime-switching models : Novel approaches removing path-dependence

Résumé

Une approche connue pour la tarification des options dans le cadre de modèles à changement de régime consiste à adapter le principe de Girsanov. Une façon d'incorporer l'incertitude de régime consiste alors à calculer les probabilités des régimes sous cette mesure de probabilité neutre au risque. Cet article montre qu'une telle approche, bien que naturelle, engendre des problèmes de dépendance au chemin dans les prix d'options vanilles. Nous argumentons que cette propriété est contre-intuitive et indésirable. Ce travail développe des mesures neutres au risque intuitives pouvant incorporer de manière simple l'aversion au risque de régime et qui n'entraînent pas de tels effets secondaires de dépendance au chemin. Des schémas numériques basés sur la programmation dynamique ainsi que des méthodes de simulations Monte-Carlo pour calculer les prix des options sont présentés pour ces nouvelles mesures neutres au risque.

Abstract

A well-known approach for the pricing of options under regime-switching models is to use the extended Girsanov principle to obtain risk-neutrality. One way to handle regime uncertainty consists in using regime probabilities that are filtered under this risk-neutral measure to compute risk-neutral expected payoffs. The current paper shows that this natural approach creates path-dependence issues within option price dynamics. Indeed, since the underlying asset price can be embedded in a Markov process under the physical measure even when regimes are unobservable, such path-dependence behavior of vanilla option prices is puzzling and may entail non-trivial theoretical features (e.g., time non-separable preferences) in a way that is difficult to characterize. This work develops novel and intuitive risk-neutral measures that can incorporate regime risk-aversion in a simple fashion and which do not lead to such path-dependence side effects. Numerical schemes either based on dynamic programming or Monte-Carlo simulations to compute option prices under the novel risk-neutral dynamics are presented.

Keywords : Option pricing, Regime-switching models, Hidden Markov models, Esscher transform, Path-dependence.

3.1 Introduction

Since their introduction in the economics literature by Hamilton (1989), regime-switching models have received extensive attention in the context of derivatives pricing. This can be explained by the ability of regime-switching models to reproduce stylized facts of financial log-returns such as fat tails, volatility clusters and momentum, see for instance Ang and Timmermann (2012). In particular, regime-switching models are used to price long-dated options such as those embedded in variable annuities, see Hardy (2003). Regime-switching models are sensible choices in such circumstances since the underlying asset of a long-dated option might go through multiple business cycles or varying financial conditions throughout the life of the option. Moreover, regime-switching dynamics allow recovering volatility smiles exhibited by empirical option prices, see Ishijima and Kihara (2005) and Yao et al. (2006).

The usual route to obtain a risk-neutral measure in the context of regime-switching models is to use the extended Girsanov principle (also sometimes referred to as the mean-correcting transform or the regime-switching Esscher transform) which preserves the model specification and shifts the drift to the risk-free rate, see for instance Hardy (2001) and Buffington and Elliott (2002b). Elliott et al. (2005) provide a justification for using the latter transform by showing that it leads to the minimal entropy martingale measure. In previous works, the Girsanov transform is often applied under the assumption of observable regimes. Failing to recognize that latent variables are unobserved can lead to systematic bias in option prices, see Bégin and Gauthier (2017). To handle regime latency, the typical approach found for instance in Liew and Siu (2010) is to compute the filtered risk-neutral distribution of the hidden regimes to obtain weights for derivatives prices associated with each regime which lead to a price in the context of regime uncertainty.

The current paper shows that combining the usual Girsanov transform with the risk-neutral filter in the context of regime-switching models provides price dynamics exhibiting path-dependence even though the underlying asset price can be embedded in a Markov process under the physical measure. Such a feature points toward non-trivial theoretical implications such as time non-separable preferences as in Garcia et al. (2003). Moreover, the interpretation for the time evolution mechanism of risk-neutral regime probabilities in terms of the underlying asset price movements is not very clear. Modeling option prices from a dynamic perspective rather than a static one is very important since such dynamic models are embedded into dynamic hedging performance assessment models, see for instance Trottier et al. (2017).

In the current paper, alternative risk-neutral measures which possess intuitive properties and remove the path-dependence feature are developed. A first approach is a modified version of the regime-switching Girsanov transform that leads to the construction of a wide class of risk-neutral measures by engineering a dynamic transition matrix so as to yield option prices exhibiting the Markov property. Such risk-neutral measures are obtained by the successive alteration of transition probabilities and of the underlying asset drift. A second approach explores two different families of martingale measures whose Radon-Nikodym derivatives are measurable given the partial observable information. For the latter measures, option prices exhibit the Markov property, and furthermore the conditional distribution of the past (unobservable) regime trajectory given the asset full trajectory set is left unaltered. The latter property is consistent with the interpretation of a risk-neutral measure as a representation of aggregate risk-aversion and other determinants of equilibrium prices ; these factors should not distort past risk distributions given the full asset trajectory. Under all of our introduced martingale measures, option prices can be calculated simply either through a dynamic program or a Monte-Carlo simulation.

Several other interesting papers from the regime-switching option pricing literature should be mentioned. Classical regime-switching dynamics were expanded by incorporating jumps, see Naik (1993), Elliott et al. (2007) and Elliott and Siu (2013), or GARCH feedback effects (Duan et al., 2002). European options are priced in a Gaussian regime-switching setting using quadratic global hedging in Rémillard et al. (2017). Multiple types of derivatives were priced such as American options (Buffington and Elliott, 2002a), perpetual American options (Zhang and Guo, 2004), barrier options (Jobert and Rogers, 2006; Ranjbar and Seifi, 2015), and other exotic options such as Asian and lookback options (Boyle and Draviam, 2007). The incorporation to the market of an additional asset providing payoffs at regime switches which allows completing the market is investigated in Guo (2001) and Fuh et al. (2012). The partial differential equations approach to price derivatives in regime-switching markets is presented in Mamon and Rodrigo (2005). Di Masi et al. (1995) investigate mean-variance hedging in the presence of regimes. Various numerical schemes were developed to price options in the regime-switching context, such as trees (Bollen, 1998; Yuen and Yang, 2009), and the fast Fourier transform (Liu et al., 2006). Finally, alternative approaches to pricing such as equilibrium and stochastic games are considered in Garcia et al. (2003) and Shen and Siu (2013).

The paper continues as follows. Section 3.2 introduces the regime-switching market. Section 3.3 illustrates the use of the mean-correcting transform to price options under regime uncertainty. The non-Markov behavior of option prices under such a transform is discussed. Section 3.4 introduces a wide class of risk-neutral measures based on the successive alteration of transition probabilities and of the underlying asset drift. Section 3.5 explores two different families of martingale measures whose Radon-Nikodym derivatives are measurable given the asset trajectory. Section 3.6 concludes.

3.2 Regime-switching market

This section introduces the regime-switching market model. We adopt the shorthand notation $x_{1:n} \equiv (x_1, \dots, x_n)$, and denote the conditional PDF of random variables X given Y by $f_{X|Y}$.

3.2.1 Regime-switching model

Consider a discrete time space $\mathcal{T} = \{0, \dots, T\}$ and a probability space $(\Omega, \mathcal{F}_T, \mathbb{P})$. Define a regime process $h = \{h_t\}_{t=0}^{T-1}$ and an innovation process $z^{\mathbb{P}} = \{z_t^{\mathbb{P}}\}_{t=1}^{T}$ which are independent under $\mathbb{P}$. The process $z^{\mathbb{P}}$ is a strong standardized Gaussian white noise. Possible values for regimes are $h_t(\omega) \in \{1, \dots, H\}$ for all $\omega \in \Omega$, where H is a positive integer. A risk-free asset is introduced and its price is given by $B_t = e^{rt}$ with r being the constant risk-free rate. A risky asset price process is defined by

$$S_t = S_0 \exp\left(\sum_{j=1}^{t} \epsilon_j\right), \qquad t \in \mathcal{T}, \tag{3.2.1}$$

where the asset log-returns are given by

$$\epsilon_{t+1} = \mu_{h_t} + \sigma_{h_t} z_{t+1}^{\mathbb{P}}, \qquad t \in \{0, \dots, T-1\}, \tag{3.2.2}$$

for some constants μ_i and σ_i, $i \in \{1, \dots, H\}$.

The filtrations $\mathcal{G} = \{\mathcal{G}_t\}_{t=0}^{T}$, $\mathcal{H} = \{\mathcal{H}_t\}_{t=0}^{T}$ and $\mathcal{F} = \{\mathcal{F}_t\}_{t=0}^{T}$ are defined as

$$\mathcal{G}_t = \sigma(S_0, \dots, S_t), \qquad \mathcal{H}_t = \sigma(h_0, \dots, h_t), \qquad \mathcal{F}_t = \mathcal{G}_t \vee \mathcal{H}_t. \tag{3.2.3}$$

$\mathcal{G}$ and $\mathcal{H}$ are sub-σ-algebras of $\mathcal{F}$. The filtration $\mathcal{G}$ is referred to as the partial information whereas $\mathcal{F}$ is called the full information. In practice, investors only have access to information $\mathcal{G}_t$ at time t as regimes are hidden variables.

A standard assumption in the literature is to assume the regime process h is a Markov chain. We therefore assume that for all $j \in \{1, \dots, H\}$,

$$\mathbb{P}[h_{t+1} = j | \mathcal{G}_{t+1} \vee \mathcal{H}_t] = P_{h_t, j}, \tag{3.2.4}$$

where $P_{k,j}$ represents the probability of a transition $k \to j$ of the Markov chain h. This implies

$$\mathbb{P}[h_{t+1} = j | \mathcal{F}_t] = P_{h_t, j}.$$

This framework is known as a regime-switching (RS) model. The joint mixed PDF of $\left(\epsilon_{1:T}, h_{0:T-1}\right)$ under such model is (proof in Appendix 3.A.1)

$$f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1}) = f^{\mathbb{P}}_{h_0}(h_0) \prod_{t=2}^{T} P_{h_{t-2},h_{t-1}} \prod_{t=1}^{T} \phi^{\mathbb{P}}_{h_{t-1}}(\epsilon_t), \tag{3.2.5}$$

where we have introduced the functions $\phi^{\mathbb{P}}_i$, $i \in \{1, \dots, H\}$, which are defined as

$$\phi^{\mathbb{P}}_i(x) \equiv \frac{1}{\sigma_i}\phi\left(\frac{x-\mu_i}{\sigma_i}\right), \qquad x \in \mathbb{R}, \tag{3.2.6}$$

with $\phi(z) \equiv \frac{e^{-z^2/2}}{\sqrt{2\pi}}$ denoting the standard normal PDF. Hence, $\phi^{\mathbb{P}}_i$ is the Gaussian density with mean μ_i and variance σ_i^2.

3.2.2 Regime mass function

Following François et al. (2014), we introduce $\eta^{\mathbb{P}} = \{\eta^{\mathbb{P}}_t\}_{t=0}^{T}$ where $\eta^{\mathbb{P}}_t = \left(\eta^{\mathbb{P}}_{t,1}, \dots, \eta^{\mathbb{P}}_{t,H}\right)$ is defined as the regime mass function process, or filtered density, with respect to the partial information :

$$\eta^{\mathbb{P}}_{t,j} \equiv \mathbb{P}[h_t = j | \mathcal{G}_t], \qquad j \in \{1, \dots, H\}. \tag{3.2.7}$$

The random vector $\eta^{\mathbb{P}}_t = \left(\eta^{\mathbb{P}}_{t,1}, \dots, \eta^{\mathbb{P}}_{t,H}\right)$ determines what are the probabilities at time t that the regime process is currently in each respective possible regime given the observable information.

François et al. (2014) show that the process $\eta^{\mathbb{P}}$ can be computed through the following recursion :

$$\eta^{\mathbb{P}}_{t+1,i} = \frac{\sum_{j=1}^{H} P_{j,i}\, \phi^{\mathbb{P}}_j(\epsilon_{t+1})\eta^{\mathbb{P}}_{t,j}}{\sum_{\ell=1}^{H}\sum_{j=1}^{H} P_{j,\ell}\, \phi^{\mathbb{P}}_j(\epsilon_{t+1})\eta^{\mathbb{P}}_{t,j}} = \frac{\sum_{j=1}^{H} P_{j,i}\, \phi^{\mathbb{P}}_j(\epsilon_{t+1})\eta^{\mathbb{P}}_{t,j}}{\sum_{j=1}^{H} \phi^{\mathbb{P}}_j(\epsilon_{t+1})\eta^{\mathbb{P}}_{t,j}}, \qquad i \in \{1, \dots, H\}. \tag{3.2.8}$$

A direct consequence of this relation is the following proposition.

Proposition 3.2.1 (François et al. 2014)**.** *The joint process $\left\{\left(S_t, \eta^{\mathbb{P}}_t\right)\right\}_{t=0}^{T}$ has the Markov property with respect to the filtration $\mathcal{G}$ under the physical measure $\mathbb{P}$.*

The conditional density of the stock price process under $\mathbb{P}$ is

$$f^{\mathbb{P}}_{S_{t+1}|S_{0:t}}(s|S_{0:t}) = \sum_{k=1}^{H} \eta^{\mathbb{P}}_{t,k} \frac{1}{s\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{\left[\log(s/S_t) - \mu_k\right]^2}{2\sigma_k^2}\right), \qquad s \geq 0, \tag{3.2.9}$$

which is a mixture of log-normal distributions with mixing weights $\eta^{\mathbb{P}}_t$.

3.3 The RS mean-correcting martingale measure

This section illustrates the traditional approach to option pricing based on a regime-switching version of the mean-correcting martingale measure as in Hardy (2001) and Elliott et al. (2005). This procedure is shown to entail complicated counterintuitive theoretical features such as non-Markovian option price dynamics even though the underlying asset price process can be embedded in a Markov process under the physical measure.

3.3.1 Constructing the RS mean-correcting martingale measure

Consider a European-type contingent claim whose payoff at time T is $\Psi(S_T)$, for some non-negative real function Ψ. For instance, a call option has a payoff $\Psi(S_T) = \max(S_T - K, 0)$ where $K \geq 0$ is the strike price. The problem considered in the current paper is to identify a suitable price process $\Pi = \{\Pi_t\}_{t=0}^T$ for the contingent claim, where Π_t represents the contingent claim price at time t. Since regimes are unobservable, only prices Π_t that are $\mathcal{G}_t$-measurable are considered as prices cannot depend on information that is unavailable to investors. This approach is different from the one of Hardy (2001) where the option price depends on the currently prevailing regime.

Define $\mathcal{Q}$ as the set of all probability measures $\mathbb{Q}$ that are equivalent to $\mathbb{P}$ and such that the discounted price process $\{e^{-rt}S_t\}_{t=0}^T$ is a $\mathcal{G}$-martingale under the measure $\mathbb{Q}$. Such probability measures are referred to as martingale measures. A well known result from option pricing theory (see, e.g., Harrison and Kreps, 1979, for a proof) is that the set of all pricing processes which do not generate arbitrage opportunities is characterized by

$$\left\{\Pi^{\mathbb{Q}} = \left\{e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[\Psi(S_T)|\mathcal{G}_t]\right\}_{t=0}^T : \mathbb{Q} \in \mathcal{Q}\right\}.$$

Because the market is incomplete under the regime-switching framework, an infinite number of martingale measures exist and solutions to the option pricing problem are thus not unique.

A common approach is to select a particular martingale measure under which the asset price dynamics remains in the same class of models. This approach is followed for instance by Hardy (2001) who considers a martingale measure under which the risky asset price returns are still a Gaussian regime-switching process with transition probabilities $P_{i,j}$, but where the drift in each regime μ_i is replaced by $r - \frac{1}{2}\sigma_i^2$. Such a martingale measure can be constructed using a regime-switching mean-correcting change of measure following the lines of Elliott et al. (2005) who perform a similar exercise in a continuous-time framework. Replacing μ_i by $r - \frac{1}{2}\sigma_i^2$ in (3.2.6), the joint mixed PDF of returns and regimes under such a risk-neutral measure $\mathbb{Q}$ is

$$f^{\mathbb{Q}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1}) = f^{\mathbb{P}}_{h_0}(h_0)\prod_{t=2}^{T} P_{h_{t-2},h_{t-1}}\prod_{t=1}^{T}\phi^{\mathbb{Q}}_{h_{t-1}}(\epsilon_t), \tag{3.3.1}$$

where the functions $\phi_i^{\mathbb{Q}}$, $i \in \{1, \dots, H\}$, are defined as

$$\phi_i^{\mathbb{Q}}(x) \equiv \frac{1}{\sigma_i}\phi\left(\frac{x - r + \frac{1}{2}\sigma_i^2}{\sigma_i}\right), \qquad x \in \mathbb{R}. \tag{3.3.2}$$

An assumption implicit to (3.3.1) is that the distribution of the initial regime h_0 is left untouched by the change of measure i.e. $f_{h_0}^{\mathbb{P}} = f_{h_0}^{\mathbb{Q}}$.

The following result (proven in the Online Appendix 3.D.1) shows how to create a new probability measure under which the underlying asset price and regimes dynamics matches the desired one.

Proposition 3.3.1. *Consider any joint mixed PDF for $(\epsilon_{1:T}, h_{0:T-1})$ denoted by $f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}$. Then the measure defined by $\mathbb{Z}[A] \equiv \mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_A \frac{d\mathbb{Z}}{d\mathbb{P}}\right]$, for all $A \in \mathcal{F}_T$, where*

$$\frac{d\mathbb{Z}}{d\mathbb{P}} \equiv \frac{f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})}{f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})}, \tag{3.3.3}$$

is a probability measure. $\mathbb{Z}$ is equivalent to $\mathbb{P}$ if and only if $f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})$ is strictly positive almost surely. Furthermore, the joint mixed PDF of $(\epsilon_{1:T}, h_{0:T-1})$ under $\mathbb{Z}$ is $f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}$.

By Theorem 3.3.1, we thus consider the measure $\mathbb{Q}$ generated by the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{f^{\mathbb{Q}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})}{f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})}, \tag{3.3.4}$$

where $f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}$ and $f^{\mathbb{Q}}_{\epsilon_{1:T},h_{0:T-1}}$ are defined as before; see (3.2.5) and (3.3.1). Simplifying yields (see Online Appendix 3.D.3)

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \prod_{t=1}^{T} \xi_t, \qquad \xi_t = e^{z_t^{\mathbb{P}}\lambda_t - \frac{1}{2}\lambda_t^2}, \tag{3.3.5}$$

where

$$\lambda_t \equiv -\frac{\mu_{h_{t-1}} - r + \frac{1}{2}\sigma^2_{h_{t-1}}}{\sigma_{h_{t-1}}}. \tag{3.3.6}$$

From (3.2.2), defining $z_t^{\mathbb{Q}} \equiv z_t^{\mathbb{P}} - \lambda_t$ yields

$$\epsilon_{t+1} = r - \frac{1}{2}\sigma^2_{h_t} + \sigma_{h_t} z^{\mathbb{Q}}_{t+1}.$$

By Theorem 3.3.1, the joint PDF of $(\epsilon_{1:T}, h_{0:T-1})$ under $\mathbb{Q}$ is $f^{\mathbb{Q}}_{\epsilon_{1:T},h_{0:T-1}}$. Furthermore,

- $\{z_t^{\mathbb{Q}}\}_{t=1}^{T}$ are independent standard Gaussian random variables under $\mathbb{Q}$,
- $\{z_t^{\mathbb{Q}}\}_{t=1}^{T}$ and $\{h_t\}_{t=0}^{T-1}$ are independent processes under $\mathbb{Q}$,
- $\mathbb{Q}[h_{t+1} = j | \mathcal{G}_{t+1} \vee \mathcal{H}_t] = \mathbb{Q}[h_{t+1} = j | \mathcal{F}_t] = P_{h_t,j}$.

3.3.2 Contingent claim pricing

The joint process $\{(S_t, h_t)\}_{t=0}^T$ possesses the Markov property under $\mathbb{Q}$ with respect to the filtration $\mathcal{F}$. The contingent claim price is thus given by

$$\begin{aligned}
\Pi_t^{\mathbb{Q}} &= \mathbb{E}^{\mathbb{Q}}\big[e^{-r(T-t)}\Psi(S_T)\big|\mathcal{G}_t\big], \\
&= \mathbb{E}^{\mathbb{Q}}\Big[\mathbb{E}^{\mathbb{Q}}\big[e^{-r(T-t)}\Psi(S_T)\big|\mathcal{F}_t\big]\Big|\mathcal{G}_t\Big], \\
&= \mathbb{E}^{\mathbb{Q}}\big[g_t(S_t, h_t)\big|\mathcal{G}_t\big], \quad \text{by the Markov property of } \{(S_t, h_t)\}_{t=0}^T, \\
&= \sum_{k=1}^{H} \eta_{t,k}^{\mathbb{Q}}\, g_t(S_t, k), \qquad (3.3.7)
\end{aligned}$$

where $\eta_{t,j}^{\mathbb{Q}} \equiv \mathbb{Q}\,[h_t = j|\mathcal{G}_t]$, and with g_t, $t \in \{0, \dots, T\}$, being real functions characterized by the following dynamic programming scheme starting with $g_T(s,k) = \Psi(s)$:

$$g_t(s,k) = \sum_{\ell=1}^{H} P_{k,\ell} \int_{-\infty}^{\infty} g_{t+1}\Big(se^{r-\sigma_k^2/2+\sigma_k z}, \ell\Big) \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \qquad t \in \{0, \dots, T-1\}.$$

For European options, i.e., for $\Psi(s) = \max(s-K, 0)$, Hardy (2001) provides an explicit expression for g_t in the two regimes case.

The formula (3.3.7) illustrates the path-dependence feature generated by the RS mean-correcting transform. At time t, for an investor, $(S_t, \eta_t^{\mathbb{P}})$ completely characterizes the likelihood of every possible future scenarios under the physical measure $\mathbb{P}$ due to the Markov property of $(S, \eta^{\mathbb{P}})$ with respect to the partial information $\mathcal{G}$. Indeed, $f^{\mathbb{P}}_{S_{t+1:T}|\mathcal{G}_t} = f^{\mathbb{P}}_{S_{t+1:T}|S_t,\eta_t^{\mathbb{P}}}$. It would be intuitive to expect that the option price at time t would be measurable with respect to $\sigma(S_t, \eta_t^{\mathbb{P}})$. This is however not the case with the RS mean-correcting transform as $\Pi_t^{\mathbb{Q}}$ is a function of $\eta_t^{\mathbb{Q}}$ which is not $\sigma(S_t, \eta_t^{\mathbb{P}})$-measurable in general since it depends on the whole path $S_0, \dots, S_t$. The option price $\Pi_t^{\mathbb{Q}}$ therefore exhibits path-dependence (non-Markovian behavior) although the underlying asset payoff can be expressed as a function of the last observation of the $\mathcal{G}$-Markov process $(S, \eta^{\mathbb{P}})$ under $\mathbb{P}$. This leads us to question the appropriateness of the RS mean-correcting transform applied to regime-switching models when regimes are latent ; it creates path-dependence in option prices when it would be reasonable to expect these to exhibit the Markov property. The Online Appendix 3.B further illustrates the path-dependence feature in a simplified setting.

3.3.3 Stochastic discount factor representation

The path-dependence feature can be visualized through a *stochastic discount factor* (SDF) representation. As shown in the Online Appendix 3.D.2, prices obey the following relationship :

$$\Pi_t^{\mathbb{Q}} = \mathbb{E}^{\mathbb{P}}\left[\Pi_{t+1}^{\mathbb{Q}} m_{t+1}^{\mathbb{Q}} \middle| \mathcal{G}_t\right], \qquad m_{t+1}^{\mathbb{Q}} = e^{-r} \frac{\sum_{i=1}^{H} \eta_{t,i}^{\mathbb{Q}} \, \phi_i^{\mathbb{Q}}(\epsilon_{t+1})}{\sum_{i=1}^{H} \eta_{t,i}^{\mathbb{P}} \, \phi_i^{\mathbb{P}}(\epsilon_{t+1})}. \tag{3.3.8}$$

Therefore, the SDF $m_t^{\mathbb{Q}}$ is not $\sigma(\epsilon_t, \eta_{t-1}^{\mathbb{P}})$-measurable. Pricing under $\mathbb{Q}$ in fact entails weighing prices at time $t+1$ based on the risk-neutral filtered regime probabilities $\eta_t^{\mathbb{Q}}$, and thus in a path-dependent fashion. This could point to complicated theoretical implications such as time non-separable preferences as in Garcia et al. (2003).

3.4 A new family of RS mean-correcting martingale measures

This section shows how the concept of regime-switching mean-correcting change of measure can be adapted to yield a $\sigma(S_t, \eta_t^{\mathbb{P}})$-measurable time-$t$ option price. The key takeaway is that the statistical properties of the regime process must be altered in suitable ways, i.e., so as to remove non-Markovian effects.

3.4.1 General construction of an alternative martingale measure

The joint mixed PDF of $(\epsilon_{1:T}, h_{0:T-1})$ under any probability measure $\mathbb{M}$ can be expressed as

$$\begin{aligned} f^{\mathbb{M}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1}) &= f^{\mathbb{M}}_{h_0}(h_0) f^{\mathbb{M}}_{\epsilon_1|h_0}(\epsilon_1|h_0) \times \\ &\prod_{t=2}^{T} f^{\mathbb{M}}_{h_{t-1}|h_{0:t-2},\epsilon_{1:t-1}}\left(h_{t-1}|h_{0:t-2}, \epsilon_{1:t-1}\right) f^{\mathbb{M}}_{\epsilon_t|h_{0:t-1},\epsilon_{1:t-1}}(\epsilon_t|h_{0:t-1}, \epsilon_{1:t-1}). \end{aligned} \tag{3.4.1}$$

To obtain the martingale property, we apply a RS mean correction, i.e., we impose that conditionally on the current regime h_{t-1}, the distribution of the log-return ϵ_t is still Gaussian with a variance equal to the physical one and a mean of $r - \frac{1}{2}\sigma^2_{h_{t-1}}$. Therefore,

$$f^{\mathbb{M}}_{\epsilon_t|h_{0:t-1},\epsilon_{1:t-1}} = \phi^{\mathbb{Q}}_{h_{t-1}}, \qquad t \geq 1, \tag{3.4.2}$$

where $\phi_i^{\mathbb{Q}}$, $i \in \{1, \ldots, H\}$, is defined as before ; see (3.3.2).

Alterations on transition probabilities of the regime process are applied to remove non-Makovian effects on option prices. Consider a multivariate process $\psi = \{\psi_t\}_{t=1}^{T-1}$ where $\psi_t = \left[\psi_t^{(i,j)}\right]_{i,j=1}^{H}$ is a $\mathcal{G}_t$-measurable $H \times H$ random matrix for all $t \in \{0, \ldots, T-1\}$. Transition probabilities of the following form are assumed under $\mathbb{M}$:

$$f^{\mathbb{M}}_{h_{t-1}|h_{0:t-2},\epsilon_{1:t-1}}(h_{t-1}|h_{0:t-2}, \epsilon_{1:t-1}) = P_{h_{t-2},h_{t-1}} \psi_{t-1}^{(h_{t-2},h_{t-1})}, \qquad t \geq 2. \tag{3.4.3}$$

This imposes that for all $t \in \{1, \ldots, T-1\}$ and all $i, j \in \{1, \ldots, H\}$,

$$\psi_t^{(i,j)} > 0 \text{ almost surely}, \qquad \text{and} \qquad \sum_{j=1}^{H} P_{i,j} \psi_t^{(i,j)} = 1 \text{ almost surely}, \tag{3.4.4}$$

to ensure positiveness and normalization. Note also that the initial mass function of the first regime can be modified from $f_{h_0}^{\mathbb{P}}(h_0)$ to $f_{h_0}^{\mathbb{M}}(h_0)$ during the passage from $\mathbb{P}$ to $\mathbb{M}$.

By Theorem 3.3.1, such a measure $\mathbb{M}$ is constructed by the Radon-Nikodym derivative

$$\frac{d\mathbb{M}}{d\mathbb{P}} = \frac{f_{\epsilon_{1:T}, h_{0:T-1}}^{\mathbb{M}}(\epsilon_{1:T}, h_{0:T-1})}{f_{\epsilon_{1:T}, h_{0:T-1}}^{\mathbb{P}}(\epsilon_{1:T}, h_{0:T-1})} = \frac{f_{h_0}^{\mathbb{M}}(h_0)}{f_{h_0}^{\mathbb{P}}(h_0)} \prod_{t=2}^{T} \psi_{t-1}^{(h_{t-2}, h_{t-1})} \prod_{t=1}^{T} \xi_t, \tag{3.4.5}$$

where ξ_t is defined as in (3.3.5).

As shown in Appendix 3.A.2, the risk-neutral mass function of regimes is given by

$$\eta_{t+1,i}^{\mathbb{M}} \equiv \mathbb{M}[h_{t+1} = i | \mathcal{G}_{t+1}] = \frac{\sum_{j=1}^{H} P_{j,i} \psi_{t+1}^{(j,i)} \phi_j^{\mathbb{Q}}(\epsilon_{t+1}) \eta_{t,j}^{\mathbb{M}}}{\sum_{j=1}^{H} \phi_j^{\mathbb{Q}}(\epsilon_{t+1}) \eta_{t,j}^{\mathbb{M}}}, \qquad t \in \{0, \ldots T-1\}, \tag{3.4.6}$$

with $\eta_{0,i}^{\mathbb{M}} = f_{h_0}^{\mathbb{M}}(i)$.

Using (3.4.2) and (3.4.6), it is straightforward to show that

$$f_{\epsilon_{t+1}|\epsilon_{1:t}}^{\mathbb{M}}(\epsilon_{t+1}|\epsilon_{1:t}) = \sum_{i=1}^{H} \eta_{t,i}^{\mathbb{M}} \phi_i^{\mathbb{Q}}(\epsilon_{t+1}). \tag{3.4.7}$$

Hence, provided that $\eta_t^{\mathbb{M}}$ is $\sigma(\eta_t^{\mathbb{P}})$-measurable for all $t \geq 0$, we have that the $\mathcal{G}_t$-conditional distribution of the log-return ϵ_{t+1} under $\mathbb{M}$ depends exclusively on $\eta_t^{\mathbb{P}}$. Furthermore, $\eta_{t+1}^{\mathbb{P}}$ is a function of $(\epsilon_{t+1}, \eta_t^{\mathbb{P}})$, as shown by (3.2.8). Applying this reasoning recursively, it follows that the $\mathcal{G}_t$-conditional distribution of $\epsilon_{t+1:T}$ under $\mathbb{M}$ depends only on $\eta_t^{\mathbb{P}}$. This leads to the following result :

Proposition 3.4.1. *The joint process $\{(S_t, \eta_t^{\mathbb{P}})\}_{t=0}^{T}$ has the Markov property with respect to the filtration $\mathcal{G}$ under the probability measure $\mathbb{M}$ if $\eta_t^{\mathbb{M}}$ is $\sigma(\eta_t^{\mathbb{P}})$-measurable for all $t \geq 0$.*

Under the conditions stated in the above proposition, it follows that the option price

$$\Pi_t^{\mathbb{M}} = \mathbb{E}^{\mathbb{M}}\left[e^{-r(T-t)} \Psi(S_T) \middle| \mathcal{G}_t\right]$$

is $\sigma(S_t, \eta_t^{\mathbb{P}})$-measurable by the Markov property. A simple way of designing a probability measure $\mathbb{M}$ satisfying such conditions is provided next.

3.4.2 A simple construction of an alternative martingale measure

A special case is obtained by specifying the measure $\mathbb{M}$ through the conditions

$$f_{h_0}^{\mathbb{M}} = f_{h_0}^{\mathbb{P}}, \quad \text{and} \quad \psi_t^{(j,i)} = \frac{\eta_{t,i}^{\mathbb{P}}}{P_{j,i}} \quad \text{almost surely}, \qquad i,j \in \{1,\dots,H\}. \tag{3.4.8}$$

Using (3.4.8) in (3.4.6) yields

$$\eta_t^{\mathbb{M}} = \eta_t^{\mathbb{P}} \quad \text{almost surely}. \tag{3.4.9}$$

The condition from Proposition 3.4.1 requiring $\eta_t^{\mathbb{M}}$ to be $\sigma\big(\eta_t^{\mathbb{P}}\big)$-measurable for all $t \geq 0$ is thus trivially satisfied. As stated in Remark 3.4.1, it turns out that the martingale measure $\mathbb{M}$ obtained in this fashion has an interesting interpretation.

Remark 3.4.1. The martingale measure $\mathbb{M}$ obtained with (3.4.8) can be understood as a sequence of two consecutive changes of measure : one from the physical measure $\mathbb{P}$ to an equivalent measure $\tilde{\mathbb{P}}$ under which the statistical properties of returns are preserved, and another from $\tilde{\mathbb{P}}$ to $\mathbb{M}$ which induces the martingale property through a RS mean correction.

Indeed, assume $\tilde{\mathbb{P}}$ is a probability measure such that for all $t \in \{1,\dots,T\}$ and all $j \in \{1,\dots,H\}$,

$$\tilde{\mathbb{P}}[h_0 = j] = f_{h_0}^{\mathbb{P}}(j), \tag{3.4.10}$$

$$\tilde{\mathbb{P}}[h_t = j | \mathcal{G}_t \vee \mathcal{H}_{t-1}] = \eta_{t,j}^{\mathbb{P}}, \tag{3.4.11}$$

$$f^{\tilde{\mathbb{P}}}_{\epsilon_t | h_{0:t-1}, \epsilon_{1:t-1}} = f^{\mathbb{P}}_{\epsilon_t | h_{0:t-1}, \epsilon_{1:t-1}} = \phi^{\mathbb{P}}_{h_{t-1}}. \tag{3.4.12}$$

In other words, when passing from $\mathbb{P}$ to $\tilde{\mathbb{P}}$, only the transition probabilities are shifted, from P_{h_{t-1},h_t} to $\eta^{\mathbb{P}}_{t,h_t}$. For such a measure $\tilde{\mathbb{P}}$, it can be shown (see Appendix 3.A.3 for a proof) that

$$f^{\tilde{\mathbb{P}}}_{\epsilon_{t+1}|\mathcal{G}_t} = f^{\mathbb{P}}_{\epsilon_{t+1}|\mathcal{G}_t}. \tag{3.4.13}$$

This implies the joint distribution of log-returns is identical under both $\mathbb{P}$ and $\tilde{\mathbb{P}}$, and thus the change of measure from $\mathbb{P}$ to $\tilde{\mathbb{P}}$ preserves the statistical properties of the underlying asset S. Because regime-switching model adequacy and goodness-of-fit statistical tests are characterized by the distribution of the underlying process, there is no reason why $\mathbb{P}$ might be preferred to $\tilde{\mathbb{P}}$ when a regime-switching model is deemed appropriate for the price dynamics of some asset ; both have the same joint distribution. Thus, $\tilde{\mathbb{P}}$ could even be viewed as the physical measure.

Next, let's see how the change of measure can be decomposed. As shown in Appendix 3.A.4, the joint mixed PDF of $(\epsilon_{1:T}, h_{0:T-1})$ under $\tilde{\mathbb{P}}$ is

$$f^{\tilde{\mathbb{P}}}_{\epsilon_{1:T}, h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1}) = f^{\mathbb{P}}_{h_0}(h_0) \prod_{t=2}^{T} \eta^{\mathbb{P}}_{t-1,h_{t-1}} \prod_{t=1}^{T} \phi^{\mathbb{P}}_{h_{t-1}}(\epsilon_t). \tag{3.4.14}$$

This implies the following representation of $\mathbb{M}$:

$$\frac{d\mathbb{M}}{d\mathbb{P}} \equiv \frac{d\mathbb{M}}{d\tilde{\mathbb{P}}} \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}, \tag{3.4.15}$$

where

$$\frac{d\mathbb{M}}{d\tilde{\mathbb{P}}} = \frac{f^{\mathbb{M}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})}{f^{\tilde{\mathbb{P}}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})} = \prod_{t=1}^{T} \xi_t, \tag{3.4.16}$$

with ξ_t defined as in (3.3.5), and

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \frac{f^{\tilde{\mathbb{P}}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})}{f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})} = \prod_{t=2}^{T} \frac{\eta^{\mathbb{P}}_{t-1,h_{t-1}}}{P_{h_{t-2},h_{t-1}}}. \tag{3.4.17}$$

Therefore, $\mathbb{M}$ can be constructed by applying a regular Girsanov-type change of drift through (3.4.16) to a measure $\tilde{\mathbb{P}}$ under which the risky asset has the same statistical properties as under the physical measure $\mathbb{P}$. This confirms the statement in Remark 3.4.1. In summary, the regime-switching mean-correcting change of measure can be used to yield Markovian option prices, but it must be applied on $\tilde{\mathbb{P}}$, rather than $\mathbb{P}$.

3.4.3 Incorporating regime uncertainty aversion

The condition (3.4.9) implies that regime uncertainty risk is unpriced as the conditional distribution of the hidden regime h_t is left untouched by the passage from $\mathbb{P}$ to $\mathbb{M}$. The current section illustrates a generalization of the previous method which can incorporate regime uncertainty aversion through a so-called *conversion function*. Such a function relates $\eta^{\mathbb{M}}$ to $\eta^{\mathbb{P}}$ by applying a distortion to the regime mass function process.

Definition 3.4.1. Consider functions $\zeta_k : [0,1]^H \to [0,1]$, $k \in \{1,\ldots,H\}$, having the property

$$\sum_{k=1}^{H} \zeta_k(\eta_1,\ldots,\eta_H) = 1, \qquad \text{for all } (\eta_1,\ldots,\eta_H) \in [0,1]^H \text{ such that } \sum_{i=1}^{H} \eta_i = 1.$$

The function $\zeta = (\zeta_1,\ldots,\zeta_H)$ is referred to as a *conversion function*.

The $\psi_t^{(j,i)}$ from (3.4.3) characterizing the martingale measure $\mathbb{M}$ are determined to enforce the chosen conversion :

$$\eta^{\mathbb{M}}_{t,k} = \zeta_k\big(\eta^{\mathbb{P}}_t\big) \text{ almost surely for all } t \text{ and all } k. \tag{3.4.18}$$

By Proposition 3.4.1, path-dependence problems are purged when such a measure $\mathbb{M}$ is used as a martingale measure for pricing. From (3.4.4) and (3.4.6), the above condition involves using $\psi_t^{(i,j)}$ that are solutions of the following linear system of equations, for all $t \geq 1$:

$$\begin{aligned} \frac{\sum_{j=1}^{H} P_{j,i}\psi_t^{(j,i)}\phi_j^{\mathbb{Q}}(\epsilon_t)\zeta_j\big(\eta^{\mathbb{P}}_{t-1}\big)}{\sum_{j=1}^{H}\phi_j^{\mathbb{Q}}(\epsilon_t)\zeta_j\big(\eta^{\mathbb{P}}_{t-1}\big)} &= \zeta_i\big(\eta^{\mathbb{P}}_t\big), \qquad i \in \{1,\ldots,H\}, \\ \sum_{j=1}^{H} P_{i,j}\psi_t^{(i,j)} &= 1, \qquad i \in \{1,\ldots,H\}. \end{aligned} \tag{3.4.19}$$

The solutions are characterized in the proposition below whose proof is in Appendix 3.A.5.

Proposition 3.4.2. *The system of equations* (3.4.19) *admits an infinite number of solutions. The trivial solution is*

$$\psi_t^{(j,i)} = \frac{\zeta_i\big(\eta_t^{\mathbb{P}}\big)}{P_{j,i}}, \qquad i,j \in \{1,\ldots,H\}. \tag{3.4.20}$$

A non-trivial solution to the system (3.4.19) is presented in the Online Appendix 3.C.

Examples of conversion functions could include for instance :

- The identity conversion function :

$$\zeta_k(\eta_1,\ldots,\eta_H) = \eta_k, \tag{3.4.21}$$

- The softmax function : for some real constants a_i, b_i, with $i \in \{1,\ldots,H\}$,

$$\zeta_k(\eta_1,\ldots,\eta_H) = \frac{\exp(a_k + b_k\eta_k)}{\sum_{i=1}^{H}\exp(a_i + b_i\eta_i)}. \tag{3.4.22}$$

The identity conversion function case described in Section 3.4.2 would reflect risk-neutrality with respect to regime uncertainty, whereas the softmax function could reflect risk-aversion to regime uncertainty. Values for parameters (a_k, b_k) of the softmax function could be obtained through calibration using market option prices.

3.4.4 Price computation algorithms

Using martingale measures $\mathbb{M}$ described in the current section, options can be priced by means either of Monte-Carlo simulations or a dynamic programming approach. Both methods are outlined below.

Monte-Carlo simulations

A fairly simple recipe to simulate log-returns ϵ_t within a Monte-Carlo simulation under the measure $\mathbb{M}$ is given : at each $t = 0,\ldots,T-1$,

1. Calculate $\eta_t^{\mathbb{P}}$ from (3.2.8),
2. Calculate $\eta_{t,i}^{\mathbb{M}} = \zeta_i\big(\eta_t^{\mathbb{P}}\big)$, for $i \in \{1,\ldots,H\}$,
3. Draw ϵ_{t+1} from the Gaussian mixture (3.4.7).

Dynamic program

Dynamic programming can be used to price simple contingent claims. By Proposition 3.4.1, the option price is $\sigma\big(S_t, \eta_t^{\mathbb{P}}\big)$-measurable since (3.4.18). Hence

$$\Pi_t^{\mathbb{M}} = \mathbb{E}^{\mathbb{M}}\big[e^{-r(T-t)}\Psi(S_T)\big|\mathcal{G}_t\big] = \pi_t^{\mathbb{M}}\big(S_t, \eta_t^{\mathbb{P}}\big),$$

for some real functions $\pi_0^{\mathbb{M}}, \ldots, \pi_T^{\mathbb{M}}$.

The functions $\pi_0^{\mathbb{M}}, \ldots, \pi_T^{\mathbb{M}}$ can be computed through a simple dynamic program provided by Proposition 3.4.3 which is proven in Appendix 3.A.6.

Proposition 3.4.3. *For $i \in \{1, \ldots, H\}$ and $t \in \{0, \ldots, T-1\}$, define the functions*

$$\chi_{t+1,i}(\eta, \epsilon) \equiv \frac{\sum_{j=1}^{H} P_{j,i}\, \phi_j^{\mathbb{P}}(\epsilon)\eta_j}{\sum_{j=1}^{H} \phi_j^{\mathbb{P}}(\epsilon)\eta_j} \tag{3.4.23}$$

and

$$\chi_{t+1}\big(\eta_t^{\mathbb{P}}, \epsilon_{t+1}\big) = \Big(\chi_{t+1,1}\big(\eta_t^{\mathbb{P}}, \epsilon_{t+1}\big), \ldots, \chi_{t+1,H}\big(\eta_t^{\mathbb{P}}, \epsilon_{t+1}\big)\Big). \tag{3.4.24}$$

Then, for any $t \in \{0, \ldots, T-1\}$ and any possible value of S_t and $\eta_t^{\mathbb{P}}$:

$$\pi_t^{\mathbb{M}}\big(S_t, \eta_t^{\mathbb{P}}\big) = e^{-r} \sum_{k=1}^{H} \zeta_k\big(\eta_t^{\mathbb{P}}\big) \int_{-\infty}^{\infty} \pi_{t+1}^{\mathbb{M}}\Big(S_t e^{r-\sigma_k^2/2+\sigma_k z}, \chi_{t+1}\Big(\eta_t^{\mathbb{P}}, r - \sigma_k^2/2 + \sigma_k z\Big)\Big) \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz, \tag{3.4.25}$$

with $\pi_T^{\mathbb{M}}\big(S_T, \eta_T^{\mathbb{P}}\big) = \Psi(S_T)$ where Ψ is the payoff function.

Moreover, the dimension of the pricing functional can be reduced by one as stated below.

Remark 3.4.2. Because $\sum_{k=1}^{H} \eta_{t,k}^{\mathbb{P}} = 1$ almost surely (since they represent probabilities of a sample space partition), the function $\chi_{t+1,i}(\eta, \epsilon)$ only needs to be computed at points where $\eta_1 + \ldots + \eta_H = 1$. Because of this, we can drop $\eta_{t,H}^{\mathbb{P}}$ from the state variables since it is a known quantity when $\eta_{t,1}^{\mathbb{P}}, \ldots, \eta_{t,H-1}^{\mathbb{P}}$ are given. This reduces the dimension of the pricing functional by one since it is possible to write $\pi_t^{\mathbb{M}}\big(S_t, \eta_t^{\mathbb{P}}\big) = \bar{g}_t\big(S_t, \eta_{t,1}^{\mathbb{P}}, \ldots, \eta_{t,H-1}^{\mathbb{P}}\big)$ for some function $\bar{g}_t$, $t \in \mathcal{T}$.

3.5 Martingale measures based on $\mathcal{G}_T$-measurable transforms

There are still some conceptual issues for the approach presented in the previous section. In particular, because the Radon-Nikodym derivative $\frac{d\mathbb{M}}{d\mathbb{P}}$ is not $\mathcal{G}_T$-measurable, there exists events $A \in \mathcal{F}_T$ such that

$$\mathbb{M}[A|\mathcal{G}_T] \neq \mathbb{P}[A|\mathcal{G}_T]. \tag{3.5.1}$$

This means that such risk-neutral measures can alter the likelihood of past regimes given the full asset trajectory. For instance, the most probable regime trajectory could differ significantly under $\mathbb{M}$ (compared to under $\mathbb{P}$). This property might seem counter-intuitive. Indeed, a risk-neutral measure reflects risk-aversion and other considerations that affect equilibrium prices; as such it might be desirable not to alter the posterior regime distribution when there is no asset risk left, i.e., given $S_{0:T}$.

This section illustrates the construction of martingale measures which leave the $\mathcal{G}_T$-conditional distribution of past regimes unaffected by the change of measure. A first approach relies on the adaptation of the well-known Esscher transform to the latent regimes framework. A second approach, based on a regime-mixture approach, combines features of the Esscher transform and of martingale measures constructed in Section 3.4.

3.5.1 A conditional version of the Esscher transform

The Esscher transform is a popular concept in finance and insurance, and it is therefore relevant to investigate whether it can be adapted to regime-switching models so as to provide a natural solution to path-dependence issues. The Esscher transform presented hereby is a particular case of the general pricing approach under heteroskedasticity of Christoffersen et al. (2009).

The conditional Esscher risk-neutral measure $\widehat{\mathbb{Q}}$ is defined by the Radon-Nikodym derivative

$$\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}} = \prod_{t=1}^{T} \widehat{\xi}_t, \qquad \widehat{\xi}_t \equiv e^{-\theta_{t-1}} \left(\frac{S_t}{S_{t-1}} \right)^{\alpha_{t-1}}, \tag{3.5.2}$$

where $\{\theta_t\}_{t=0}^T$ and $\{\alpha_t\}_{t=0}^T$ are $\mathcal{G}$-adapted processes to be defined. As shown in Appendix 3.A.7, the following condition, which is assumed to hold, ensures that $\widehat{\mathbb{Q}}$ is a probability measure :

$$\theta_t = \log \left(\sum_{k=1}^{H} \eta_{t,k}^{\mathbb{P}} \exp\left(\alpha_t \mu_k + \frac{1}{2}\alpha_t^2 \sigma_k^2 \right) \right). \tag{3.5.3}$$

Moreover, assuming this condition holds, as shown in Appendix 3.A.8, the following condition is necessary and sufficient to ensure that $\widehat{\mathbb{Q}}$ is a risk-neutral measure :

$$\sum_{k=1}^{H} \eta_{t,k}^{\mathbb{P}} \exp\left(\alpha_t \mu_k + \frac{1}{2}\alpha_t^2 \sigma_k^2 \right) \left[1 - \exp\left(\mu_k + \alpha_t \sigma_k^2 + \frac{1}{2}\sigma_k^2 - r \right) \right] = 0. \tag{3.5.4}$$

A solution to this equation always exists since the left hand side tends to minus infinity as $\alpha_t \to \infty$ and to infinity as $\alpha_t \to -\infty$, on top of being a continuous function of α_t. Equation (3.5.4) can be solved numerically to determine α_t ; the solution is a function of $\eta_t^{\mathbb{P}}$, and therefore (θ_t, α_t) is a function of $\eta_t^{\mathbb{P}}$.

Appendix 3.A.9 shows that the distribution of returns under the measure $\widehat{\mathbb{Q}}$ is characterized by

$$\widehat{\mathbb{Q}}[\epsilon_{t+1} \leq x | \mathcal{G}_t] = \sum_{i=1}^{H} \hat{\eta}_{t,i}^{\mathbb{P}} \Phi\left(\frac{x - \mu_i - \alpha_t \sigma_i^2}{\sigma_i} \right), \qquad x \in \mathbb{R}, \tag{3.5.5}$$

where Φ is the standard Gaussian cumulative distribution function, and

$$\hat{\eta}_{t,i}^{\mathbb{P}} = \frac{\eta_{t,i}^{\mathbb{P}} \exp\left(\alpha_t \mu_i + \frac{1}{2}\alpha_t^2 \sigma_i^2 \right)}{\sum_{k=1}^{H} \eta_{t,k}^{\mathbb{P}} \exp\left(\alpha_t \mu_k + \frac{1}{2}\alpha_t^2 \sigma_k^2 \right)}. \tag{3.5.6}$$

The log-returns $\mathcal{G}_t$-conditional distribution under $\widehat{\mathbb{Q}}$ is therefore still a Gaussian mixture with modified mixing weights $\hat{\eta}_t^{\mathbb{P}}$ and means shifted from μ_i to $\mu_i - \alpha_t \sigma_i^2$ for each regime $i \in \{1, \ldots, H\}$. Note that the passage from $\eta_t^{\mathbb{P}}$ to $\hat{\eta}_t^{\mathbb{P}}$ is an instance of a conversion function since α_t is a function of $\eta_t^{\mathbb{P}}$ as shown by (3.5.4).

Equations (3.5.5)-(3.5.6) indicate the $\widehat{\mathbb{Q}}$ distribution of the log-return ϵ_{t+1} given $\mathcal{G}_t$ depends exclusively on $\eta_t^{\mathbb{P}}$ since α_t and $\hat{\eta}_t^{\mathbb{P}}$ are functions of $\eta_t^{\mathbb{P}}$. Furthermore, $\eta_{t+1}^{\mathbb{P}}$ is a function of $(\epsilon_{t+1}, \eta_t^{\mathbb{P}})$; see (3.2.8). Applying this reasoning recursively, it follows that the $\mathcal{G}_t$-conditional distribution of $\epsilon_{t+1:T}$ under $\widehat{\mathbb{Q}}$ depends only on $\eta_t^{\mathbb{P}}$. This leads to the following result :

Proposition 3.5.1. *The joint process* $\{(S_t, \eta_t^{\mathbb{P}})\}_{t=0}^{T}$ *has the Markov property with respect to the filtration* $\mathcal{G}$ *under the probability measure* $\widehat{\mathbb{Q}}$.

This result entails that the option price at time t is $\sigma(S_t, \eta_t^{\mathbb{P}})$-measurable as desired. Other desirable theoretical properties satisfied by this measure are outlined in the remark below.

Remark 3.5.1. The risk-neutral measure $\widehat{\mathbb{Q}}$ displays the following desirable properties :

- The option price $\Pi_t^{\widehat{\mathbb{Q}}} = \mathbb{E}^{\widehat{\mathbb{Q}}}\big[e^{-r(T-t)}\Psi(S_T)\big|\mathcal{G}_t\big]$ is $\sigma(S_t, \eta_t^{\mathbb{P}})$-measurable.
- $\widehat{\xi}_t$ is $\mathcal{G}_t$-measurable for all $t \in \mathcal{T}$ and therefore $\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}} \in \mathcal{G}_T$. Thus, the $\mathcal{G}_T$-conditional distribution of past risks is unaffected by the change of measure : $\widehat{\mathbb{Q}}[A|\mathcal{G}_T] = \mathbb{P}[A|\mathcal{G}_T]$, $\forall A \in \mathcal{F}_T$.
- If the martingale property is already satisfied under $\mathbb{P}$, i.e., $\phi_i^{\mathbb{Q}} = \phi_i^{\mathbb{P}}$ for all $i \in \{1, \ldots, H\}$, then there is no change of measure, i.e., $\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}} = 1$ almost surely. [1]
- In the single-regime case ($H = 1$), $\widehat{\mathbb{Q}}$ reduces to the usual Esscher martingale measure $\mathbb{Q}$.

Option pricing schemes

A simple recipe is available to simulate log-returns under the measure $\widehat{\mathbb{Q}}$ within a Monte-Carlo simulation : at each $t = 0, \ldots, T-1$,

1. Calculate $\eta_{t,i}^{\mathbb{P}}$, $i \in \{1, \ldots, H\}$, from (3.2.8),
2. Solve numerically for α_t in (3.5.4),
3. Calculate $\hat{\eta}_{t,i}^{\mathbb{P}}$, $i \in \{1, \ldots, H\}$, from (3.5.6),
4. Draw ϵ_{t+1} from the Gaussian mixture (3.5.5).

Note that the second and third steps can be pre-calculated.

Simple contingent claims can also be priced by dynamic programming. Since the time-t option price is $\sigma(S_t, \eta_t^{\mathbb{P}})$-measurable, it follows that for all $t \in \mathcal{T}$ there exists a function $\pi_t^{\widehat{\mathbb{Q}}}$ such that

$$\Pi_t^{\widehat{\mathbb{Q}}} \equiv \mathbb{E}^{\widehat{\mathbb{Q}}}\big[e^{-r(T-t)}\Psi(S_T)\big|\mathcal{G}_t\big] = \pi_t^{\widehat{\mathbb{Q}}}(S_t, \eta_t^{\mathbb{P}}).$$

1. This is because we then have $\alpha_t = \theta_t = 0$ almost surely for all t.

The dynamic program that enables the recursive computation of the functions $\pi_t^{\widehat{\mathbb{Q}}}$ can be derived following the steps outlined in Section 3.4.4 :

$$\pi_t^{\widehat{\mathbb{Q}}}(S_t, \eta_t^{\mathbb{P}}) = e^{-r} \sum_{k=1}^{H} \hat{\eta}_{t,k}^{\mathbb{P}} \int_{-\infty}^{\infty} \pi_{t+1}^{\widehat{\mathbb{Q}}}\Big(S_t e^{\mu_k - \alpha_t \sigma_k^2 + \sigma_k z}, \chi_{t+1}\Big(\eta_t^{\mathbb{P}}, \mu_k - \alpha_t \sigma_k^2 + \sigma_k z\Big)\Big) \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz, \tag{3.5.7}$$

with $\pi_T^{\widehat{\mathbb{Q}}}(S_T, \eta_T^{\mathbb{P}}) = \Psi(S_T)$ where Ψ is the payoff function, $\hat{\eta}_t^{\mathbb{P}}$ is defined as a function of $\eta_t^{\mathbb{P}}$ through (3.5.6), and χ_{t+1} is defined by (3.4.24).

3.5.2 A regime-mixture transform

We present now a new family of martingale measures based on a regime-mixture approach. A measure from this new family is denoted by $\bar{\mathbb{Q}}$. Similarly to the conditional Esscher transform $\widehat{\mathbb{Q}}$ from Section 3.5.1, the Radon-Nikodym derivative characterizing the new regime-mixture martingale measure $\bar{\mathbb{Q}}$ is $\mathcal{G}_T$-measurable. This implies the $\mathcal{G}_T$-conditional distribution of regimes $h_{0:T-1}$ is left untouched by the change of measure, which can be deemed a desirable property as previously discussed. Moreover, as for RS mean-correcting measures $\mathbb{M}$, the risk-neutral one-period conditional distribution of asset log-returns is a mixture of Gaussian distribution whose mean is the risk-free rate minus the usual convexity correction. The regime-mixture approach therefore combines features of the two families of martingale measures previously considered, namely the new version of the RS mean-correcting measure $\mathbb{M}$ and the conditional Esscher transform $\widehat{\mathbb{Q}}$. We first explain how this measure can be derived.

The PDF of a trajectory under a probability measure $\bar{\mathbb{Q}}$ can be expressed as (see Appendix 3.A.10)

$$f^{\bar{\mathbb{Q}}}_{\epsilon_{1:T}, h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1}) = f^{\bar{\mathbb{Q}}}_{h_{0:T-1}|\mathcal{G}_T}(h_{0:T-1}|\mathcal{G}_T) \prod_{t=1}^{T} f^{\bar{\mathbb{Q}}}_{\epsilon_t|\mathcal{G}_{t-1}}(\epsilon_t|\mathcal{G}_{t-1}). \tag{3.5.8}$$

In comparison, the PDF under $\mathbb{P}$ is given by (see Appendix 3.A.11)

$$f^{\mathbb{P}}_{\epsilon_{1:T}, h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1}) = f^{\mathbb{P}}_{h_{0:T-1}|\mathcal{G}_T}(h_{0:T-1}|\mathcal{G}_T) \prod_{t=1}^{T} \sum_{i=1}^{H} \eta^{\mathbb{P}}_{t-1,i}\, \phi_i^{\mathbb{P}}(\epsilon_t). \tag{3.5.9}$$

The regime-mixture Esscher martingale measure $\bar{\mathbb{Q}}$ is constructed by enforcing

$$f^{\bar{\mathbb{Q}}}_{h_{0:T-1}|\mathcal{G}_T}(h_{0:T-1}|\mathcal{G}_T) = f^{\mathbb{P}}_{h_{0:T-1}|\mathcal{G}_T}(h_{0:T-1}|\mathcal{G}_T), \tag{3.5.10}$$

$$f^{\bar{\mathbb{Q}}}_{\epsilon_t|\mathcal{G}_{t-1}}(\epsilon_t|\mathcal{G}_{t-1}) = \sum_{i=1}^{H} \zeta_i(\eta^{\mathbb{P}}_{t-1}) \phi_i^{\mathbb{Q}}(\epsilon_t), \qquad \forall t \in \{1, \ldots, T\}, \tag{3.5.11}$$

where ζ is the conversion function, and $\phi_i^{\mathbb{Q}}$, $i \in \{1, \ldots, H\}$, is defined as before ; see (3.3.2). The property (3.5.10) states that the $\mathcal{G}_T$-conditional distribution of the regime trajectory is unaltered under $\bar{\mathbb{Q}}$. This is an intuitive feature as previously discussed. The property (3.5.11) states the $\mathcal{G}_{t-1}$-conditional distribution of the log-return ϵ_t under $\bar{\mathbb{Q}}$ is a Gaussian mixture with mixing weights given by the vector $\zeta(\eta^{\mathbb{P}}_{t-1})$, and means shifted from μ_i to $r - \frac{1}{2}\sigma_i^2$ for each regime $i \in \{1, \ldots, H\}$. The purpose of the latter condition is to ensure the martingale property is satisfied, and that regime risk is priced according to the chosen conversion function.

As shown in Appendix 3.A.12 the Radon-Nikodym derivative is

$$\frac{d\bar{\mathbb{Q}}}{d\mathbb{P}} = \prod_{t=1}^{T} \bar{\xi}_t, \qquad \bar{\xi}_t \equiv \frac{\sum_{i=1}^{H} \zeta_i(\eta^{\mathbb{P}}_{t-1})\phi^{\mathbb{Q}}_i(\epsilon_t)}{\sum_{i=1}^{H} \eta^{\mathbb{P}}_{t-1,i}\, \phi^{\mathbb{P}}_i(\epsilon_t)}. \tag{3.5.12}$$

Appendix 3.A.13 shows that the distribution of returns under this measure is characterized by

$$\bar{\mathbb{Q}}[\epsilon_{t+1} \leq x | \mathcal{G}_t] = \sum_{i=1}^{H} \zeta_i(\eta^{\mathbb{P}}_t)\Phi\left(\frac{x - r + \frac{1}{2}\sigma_i^2}{\sigma_i}\right), \qquad x \in \mathbb{R}. \tag{3.5.13}$$

Hence, for any $s = 0, \ldots, T-t-1$, the $\mathcal{G}_{t+s}$-conditional distribution of ϵ_{t+1+s} under $\bar{\mathbb{Q}}$ depends only on $\eta^{\mathbb{P}}_{t+s}$. Furthermore, by (3.2.8), $\eta^{\mathbb{P}}_{t+s}$ is a function of $\left(\epsilon_{t+s}, \eta^{\mathbb{P}}_{t+s-1}\right)$. The above reasoning, applied recursively, implies that the $\mathcal{G}_t$-conditional distribution of $\epsilon_{t+1:T}$ under $\bar{\mathbb{Q}}$ depends only on $\eta^{\mathbb{P}}_t$. The next proposition then follows.

Proposition 3.5.2. *The joint process $\left\{(S_t, \eta^{\mathbb{P}}_t)\right\}_{t=0}^{T}$ has the Markov property with respect to the filtration $\mathcal{G}$ under the probability measure $\bar{\mathbb{Q}}$.*

This property entails that the option price $\Pi^{\bar{\mathbb{Q}}}_t = \mathbb{E}^{\bar{\mathbb{Q}}}\left[e^{-r(T-t)}\Psi(S_T)\middle|\mathcal{G}_t\right]$ is $\sigma\left(S_t, \eta^{\mathbb{P}}_t\right)$-measurable. Furthermore, the other properties stated in Remark 3.5.1 also hold for $\bar{\mathbb{Q}}$. Finally, since the underlying asset price joint distribution are identical under $\mathbb{M}$ and $\bar{\mathbb{Q}}$, the pricing algorithms are identical to those given in Section 3.4.4.

3.6 Conclusion

The current work shows that the usual approach to construct martingale measures in a regime-switching framework based on the correction of the drift for each respective regime (i.e., regime-switching mean correction) leads to path-dependence even for vanilla options. More precisely, even if the joint process $(S, \eta^{\mathbb{P}})$ comprising the underlying asset price and the regime mass function given observable information has the Markov property, vanilla derivatives prices at time t would not be a function strictly of the current value of the latter process, i.e., of $(S_t, \eta^{\mathbb{P}}_t)$. The construction of multiple martingale measures possessing intuitive properties and removing the path-dependence feature is illustrated in the current paper.

Our first approach is a modified version of the above concept of RS mean-correcting martingale measure ; it also relies on RS mean correction to obtain the martingale property, but with the inclusion of transition probability transforms so as to recuperate the Markov property of option prices. This yields a very wide class of new martingale measures removing the path-dependence. This class includes an interesting special case which can be represented as the successive application of two changes of measures : a first one which allows retaining the exact same underlying asset statistical properties from the physical measure, and then a change of drift on each regime. Obtained generalizations allow for the pricing of regime uncertainty through conversion functions which distort the hidden regime distribution given the currently observed information.

A second approach developed is based on changes of measures whose Radon-Nikodym derivatives are $\sigma(S_0, \ldots, S_T)$-measurable, implying that they do not impact the conditional distribution of the regime hidden trajectory given the full asset trajectory. This approach embeds as a particular case the well-known Esscher transform.

Simple pricing procedures for contingent claims under the developed martingale measures based either on dynamic programming or Monte-Carlo simulations are also provided.

Appendix

3.A Proofs

3.A.1 Proof of Eq. (3.2.5)

$$\begin{aligned}
f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T},h_{0:T-1}) &= f^{\mathbb{P}}_{\epsilon_1,h_0}(\epsilon_1,h_0)\prod_{t=2}^{T} f^{\mathbb{P}}_{\epsilon_t,h_{t-1}|\epsilon_{1:t-1},h_{0:t-2}}(\epsilon_t,h_{t-1}|\epsilon_{1:t-1},h_{0:t-2}),\\
&= f^{\mathbb{P}}_{h_0}(h_0) f^{\mathbb{P}}_{\epsilon_1|h_0}(\epsilon_1|h_0)\times\\
&\quad \prod_{t=2}^{T} f^{\mathbb{P}}_{\epsilon_t|\epsilon_{1:t-1},h_{0:t-1}}(\epsilon_t|\epsilon_{1:t-1},h_{0:t-1}) f^{\mathbb{P}}_{h_{t-1}|\epsilon_{1:t-1},h_{0:t-2}}(h_{t-1}|\epsilon_{1:t-1},h_{0:t-2}),\\
&= f^{\mathbb{P}}_{h_0}(h_0)\prod_{t=2}^{T} P_{h_{t-2},h_{t-1}}\prod_{t=1}^{T}\frac{1}{\sigma_{h_{t-1}}}\phi\left(\frac{\epsilon_t-\mu_{h_{t-1}}}{\sigma_{h_{t-1}}}\right),
\end{aligned}$$

where the last equality follows from (3.2.2) and (3.2.4). Using definition (3.2.6) concludes the proof.

3.A.2 Proof of Eq. (3.4.6)

$$\begin{aligned}
\eta^{\mathbb{M}}_{t+1,i} &= \mathbb{M}\left[h_{t+1}=i|\mathcal{G}_{t+1}\right],\\
&= \sum_{j=1}^{H}\mathbb{M}\left[h_{t+1}=i|\mathcal{G}_{t+1},h_t=j\right]\mathbb{M}\left[h_t=j|\mathcal{G}_{t+1}\right],\\
&= \sum_{j=1}^{H} P_{j,i}\psi^{(j,i)}_{t+1}\frac{f^{\mathbb{M}}_{h_t,\epsilon_{t+1}|\epsilon_{1:t}}(j,\epsilon_{t+1}|\epsilon_{1:t})}{f^{\mathbb{M}}_{\epsilon_{t+1}|\epsilon_{1:t}}(\epsilon_{t+1}|\epsilon_{1:t})}, \qquad \text{from (3.4.3)},\\
&= \sum_{j=1}^{H} P_{j,i}\psi^{(j,i)}_{t+1}\frac{f^{\mathbb{M}}_{h_t|\epsilon_{1:t}}(j|\epsilon_{1:t}) f^{\mathbb{M}}_{\epsilon_{t+1}|h_t,\epsilon_{1:t}}(\epsilon_{t+1}|j,\epsilon_{1:t})}{\sum_{k=1}^{H} f^{\mathbb{M}}_{h_t|\epsilon_{1:t}}(k|\epsilon_{1:t}) f^{\mathbb{M}}_{\epsilon_{t+1}|h_t,\epsilon_{1:t}}(\epsilon_{t+1}|k,\epsilon_{1:t})},\\
&= \sum_{j=1}^{H} P_{j,i}\psi^{(j,i)}_{t+1}\frac{\eta^{\mathbb{M}}_{t,j}\phi^{\mathbb{Q}}_j(\epsilon_{t+1})}{\sum_{k=1}^{H}\eta^{\mathbb{M}}_{t,k}\phi^{\mathbb{Q}}_k(\epsilon_{t+1})}, \qquad \text{from (3.4.2)}.
\end{aligned}$$

3.A.3 Proof of Eq. (3.4.13)

$$f^{\tilde{\mathbb{P}}}_{\epsilon_{t+1}|\mathcal{G}_t}(x|\mathcal{G}_t) = \sum_{k=1}^{H} f^{\tilde{\mathbb{P}}}_{\epsilon_{t+1},h_t|\mathcal{G}_t}(x,k|\mathcal{G}_t) = \sum_{k=1}^{H} \tilde{\mathbb{P}}[h_t = k|\mathcal{G}_t] f^{\tilde{\mathbb{P}}}_{\epsilon_{t+1}|h_t,\mathcal{G}_t}(x|k,\mathcal{G}_t). \tag{3.A.1}$$

Moreover,

$$\tilde{\mathbb{P}}[h_t = k|\mathcal{G}_t] = \mathbb{E}^{\tilde{\mathbb{P}}}[\mathbb{1}_{\{h_t=k\}}|\mathcal{G}_t] = \mathbb{E}^{\tilde{\mathbb{P}}}\Big[\mathbb{E}^{\tilde{\mathbb{P}}}[\mathbb{1}_{\{h_t=k\}}|\mathcal{G}_t,\mathcal{H}_{t-1}]\Big|\mathcal{G}_t\Big] = \mathbb{E}^{\tilde{\mathbb{P}}}\Big[\underbrace{\tilde{\mathbb{P}}[h_t = k|\mathcal{G}_t,\mathcal{H}_{t-1}]}_{=\eta^{\mathbb{P}}_{t,k},\text{ by (3.4.11)}}\Big|\mathcal{G}_t\Big] = \eta^{\mathbb{P}}_{t,k}.$$

Similarly, it can be shown using (3.4.12) that

$$f^{\tilde{\mathbb{P}}}_{\epsilon_{t+1}|h_t,\mathcal{G}_t}(x|k,\mathcal{G}_t) = \phi^{\mathbb{P}}_k(x).$$

Using the above relations in (3.A.1) yields

$$f^{\tilde{\mathbb{P}}}_{\epsilon_{t+1}|\mathcal{G}_t}(x|\mathcal{G}_t) = \sum_{k=1}^{H} \eta^{\mathbb{P}}_{t,k}\phi^{\mathbb{P}}_k(x) = f^{\mathbb{P}}_{\epsilon_{t+1}|\mathcal{G}_t}(x|\mathcal{G}_t),$$

where the last equality is straightforward to prove. Hence, $f^{\tilde{\mathbb{P}}}_{\epsilon_{t+1}|\mathcal{G}_t} = f^{\mathbb{P}}_{\epsilon_{t+1}|\mathcal{G}_t}$.

3.A.4 Proof of Eq. (3.4.14)

$$\begin{aligned}
f^{\tilde{\mathbb{P}}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T},h_{0:T-1}) &= f^{\tilde{\mathbb{P}}}_{\epsilon_1,h_0}(\epsilon_1,h_0)\prod_{t=2}^{T} f^{\tilde{\mathbb{P}}}_{\epsilon_t,h_{t-1}|\epsilon_{1:t-1},h_{0:t-2}}(\epsilon_t,h_{t-1}|\epsilon_{1:t-1},h_{0:t-2}),\\
&= f^{\mathbb{P}}_{h_0}(h_0) f^{\tilde{\mathbb{P}}}_{\epsilon_1|h_0}(\epsilon_1|h_0)\times\\
&\quad \prod_{t=2}^{T} f^{\tilde{\mathbb{P}}}_{\epsilon_t|\epsilon_{1:t-1},h_{0:t-1}}(\epsilon_t|\epsilon_{1:t-1},h_{0:t-1}) f^{\tilde{\mathbb{P}}}_{h_{t-1}|\epsilon_{1:t-1},h_{0:t-2}}(h_{t-1}|\epsilon_{1:t-1},h_{0:t-2}),\\
&= f^{\mathbb{P}}_{h_0}(h_0)\prod_{t=2}^{T}\eta^{\mathbb{P}}_{t-1,h_{t-1}}\prod_{t=1}^{T}\phi^{\mathbb{P}}_{h_{t-1}}(\epsilon_t), \qquad \text{from (3.4.11) and (3.4.12).}
\end{aligned}$$

3.A.5 Proof of Proposition 3.4.2

The system (3.4.19) is equivalent to

$$\sum_{j=1}^{H}\tilde{\psi}^{(j,i)}_t\kappa_{t,j} = 0 \qquad \text{and} \qquad \sum_{j=1}^{H}\tilde{\psi}^{(i,j)}_t = 0, \qquad i\in\{1,\dots,H\},$$

where we have defined

$$\tilde{\psi}^{(j,i)}_t \equiv P_{j,i}\psi^{(j,i)}_t - \zeta_i\big(\eta^{\mathbb{P}}_t\big), \qquad \kappa_{t,j} \equiv \phi^{\mathbb{Q}}_j(\epsilon_t)\zeta_j\big(\eta^{\mathbb{P}}_{t-1}\big).$$

Indeed, the trivial solution is, for all $i, j \in \{1, \dots, H\}$,

$$\tilde{\psi}_t^{(j,i)} = 0 \qquad \Rightarrow \qquad \psi_t^{(j,i)} = \frac{\zeta_i\big(\eta_t^{\mathbb{P}}\big)}{P_{j,i}}. \tag{3.A.2}$$

The system has H^2 unknown values and $2H$ equations. If $H > 2$, the existence of a solution implies that an infinite number of solutions exist. Even if $H = 2$, we can show there exists an infinite number of solutions.

Indeed, the system can be written as follows for $H = 2$,

$$\underbrace{\begin{bmatrix} \kappa_{t,1} & 0 & \kappa_{t,2} & 0 \\ 0 & \kappa_{t,1} & 0 & \kappa_{t,2} \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}}_{\equiv \mathcal{M}} \begin{bmatrix} \tilde{\psi}_t^{(1,1)} \\ \tilde{\psi}_t^{(1,2)} \\ \tilde{\psi}_t^{(2,1)} \\ \tilde{\psi}_t^{(2,2)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Since $\det \mathcal{M} = \kappa_{t,1}\kappa_{t,2} - \kappa_{t,1}\kappa_{t,2} = 0$, an infinity of solutions exist by the properties of homogeneous linear systems.

3.A.6 Proof of Proposition 3.4.3

First,

$$\begin{aligned} \pi_t^{\mathbb{M}}\big(S_t, \eta_t^{\mathbb{P}}\big) &= \mathbb{E}^{\mathbb{M}}\Big[e^{-r(T-t)}\Psi(S_T)\Big|\mathcal{G}_t\Big], \\ &= \mathbb{E}^{\mathbb{M}}\Big[e^{-r}\mathbb{E}^{\mathbb{M}}\Big[e^{-r(T-(t+1))}\Psi(S_T)\Big|\mathcal{G}_{t+1}\Big]\Big|\mathcal{G}_t\Big], \\ &= e^{-r}\mathbb{E}^{\mathbb{M}}\Big[\pi_{t+1}^{\mathbb{M}}\big(S_{t+1}, \eta_{t+1}^{\mathbb{P}}\big)\Big|\mathcal{G}_t\Big], \\ &= e^{-r}\sum_{k=1}^{H}\mathbb{M}[h_t = k|\mathcal{G}_t]\,\mathbb{E}^{\mathbb{M}}\Big[\pi_{t+1}^{\mathbb{M}}\big(S_{t+1}, \eta_{t+1}^{\mathbb{P}}\big)\Big|\mathcal{G}_t, h_t = k\Big], \\ &= e^{-r}\sum_{k=1}^{H}\zeta_k\big(\eta_t^{\mathbb{P}}\big)\mathbb{E}^{\mathbb{M}}\Big[\pi_{t+1}^{\mathbb{M}}\big(S_{t+1}, \eta_{t+1}^{\mathbb{P}}\big)\Big|S_t, \eta_t^{\mathbb{P}}, h_t = k\Big], \qquad \text{by (3.4.18)} \end{aligned} \tag{3.A.3}$$

Moreover, from (3.2.8), the definition (3.4.23) implies that

$$\eta_{t+1,i}^{\mathbb{P}} = \chi_{t+1,i}\big(\eta_t^{\mathbb{P}}, \epsilon_{t+1}\big).$$

and thus

$$\eta_{t+1}^{\mathbb{P}} = \chi_{t+1}\big(\eta_t^{\mathbb{P}}, \epsilon_{t+1}\big). \tag{3.A.4}$$

This means

$$
\begin{aligned}
&\mathbb{E}^{\mathbb{M}}\left[\pi^{\mathbb{M}}_{t+1}\big(S_{t+1},\eta^{\mathbb{P}}_{t+1}\big)\Big|S_t,\eta^{\mathbb{P}}_t,h_t=k\right]\\
&=\mathbb{E}^{\mathbb{M}}\left[\pi^{\mathbb{M}}_{t+1}\big(S_te^{\epsilon_{t+1}},\chi_{t+1}(\eta^{\mathbb{P}}_t,\epsilon_{t+1})\big)\Big|S_t,\eta^{\mathbb{P}}_t,h_t=k\right],\\
&=\mathbb{E}^{\mathbb{M}}\left[\pi^{\mathbb{M}}_{t+1}\left(S_te^{r-\sigma_k^2/2+\sigma_k z^{\mathbb{M}}_{t+1}},\chi_{t+1}\left(\eta^{\mathbb{P}}_t,r-\sigma_k^2/2+\sigma_k z^{\mathbb{M}}_{t+1}\right)\right)\Big|S_t,\eta^{\mathbb{P}}_t,h_t=k\right],\\
&=\int_{-\infty}^{\infty}\pi^{\mathbb{M}}_{t+1}\left(S_te^{r-\sigma_k^2/2+\sigma_k z},\chi_{t+1}\left(\eta^{\mathbb{P}}_t,r-\sigma_k^2/2+\sigma_k z\right)\right)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz. \qquad (3.A.5)
\end{aligned}
$$

Combining (3.A.3) and (3.A.5) yields the recursive formula (3.4.25) to obtain the option price $\Pi^{\mathbb{M}}_t=\pi^{\mathbb{M}}_t(S_t,\eta^{\mathbb{P}}_t)$ from $\pi^{\mathbb{M}}_{t+1}$.

3.A.7 Proof of Eq. (3.5.3)

To ensure $\widehat{\mathbb{Q}}$ represents a change of probability measure, the following condition which guarantees that $\mathbb{E}^{\mathbb{P}}\left[\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}}\right]=1$ is assumed to hold for all $t\geq 0$:

$$
\begin{aligned}
1 &= \mathbb{E}^{\mathbb{P}}\left[\widehat{\xi}_{t+1}\Big|\mathcal{G}_t\right], \qquad (3.A.6)\\
&= e^{-\theta_t}\mathbb{E}^{\mathbb{P}}\left[\left(\frac{S_{t+1}}{S_t}\right)^{\alpha_t}\Big|\mathcal{G}_t\right],\\
&= e^{-\theta_t}\mathbb{E}^{\mathbb{P}}\left[\exp\left(\alpha_t\mu_{h_t}+\alpha_t\sigma_{h_t}z^{\mathbb{P}}_{t+1}\right)\Big|\mathcal{G}_t\right],\\
&= e^{-\theta_t}\sum_{k=1}^{H}\eta^{\mathbb{P}}_{t,k}\exp\left(\alpha_t\mu_k+\frac{1}{2}\alpha_t^2\sigma_k^2\right),\\
\Rightarrow \theta_t &= \log\left(\sum_{k=1}^{H}\eta^{\mathbb{P}}_{t,k}\exp\left(\alpha_t\mu_k+\frac{1}{2}\alpha_t^2\sigma_k^2\right)\right). \qquad (3.A.7)
\end{aligned}
$$

Next, let's prove that $\mathbb{E}^{\mathbb{P}}\left[\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}}\right]=1$. The following property will be useful :

$$
\widehat{\xi}_s \text{ is } \mathcal{G}_t\text{-measurable}, \qquad \forall s\leq t. \qquad (3.A.8)
$$

It thus follows that for all $t\geq 1$,

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\prod_{s=t}^{T}\widehat{\xi}_s\Big|\mathcal{G}_{t-1}\right] &= \mathbb{E}^{\mathbb{P}}\left[\prod_{s=t}^{T-1}\widehat{\xi}_s\underbrace{\mathbb{E}^{\mathbb{P}}\left[\widehat{\xi}_T\Big|\mathcal{G}_{T-1}\right]}_{=1,\text{ by (3.A.6)}}\Big|\mathcal{G}_{t-1}\right], \qquad \text{by (3.A.8)},\\
&\ \vdots \quad \text{(applying recursively)}\\
&= 1. \qquad (3.A.9)
\end{aligned}
$$

In particular, for $t=1$ the above statement is equivalent to $\mathbb{E}^{\mathbb{P}}\left[\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}}\right]=1$.

3.A.8 Proof of Eq. (3.5.4)

To ensure $\widehat{\mathbb{Q}}$ is a martingale measure, the following risk-neutral condition must hold :

$$\begin{aligned}
e^r &= \mathbb{E}^{\widehat{\mathbb{Q}}}\left[\frac{S_{t+1}}{S_t}\middle|\mathcal{G}_t\right], \\
&= \frac{\mathbb{E}^{\mathbb{P}}\left[\frac{S_{t+1}}{S_t}\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}}\middle|\mathcal{G}_t\right]}{\mathbb{E}^{\mathbb{P}}\left[\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}}\middle|\mathcal{G}_t\right]}, \\
&= \frac{\prod_{n=1}^{t}\widehat{\xi}_n\,\mathbb{E}^{\mathbb{P}}\left[\frac{S_{t+1}}{S_t}\prod_{n=t+1}^{T}\widehat{\xi}_n\middle|\mathcal{G}_t\right]}{\prod_{n=1}^{t}\widehat{\xi}_n\,\underbrace{\mathbb{E}^{\mathbb{P}}\left[\prod_{n=t+1}^{T}\widehat{\xi}_n\middle|\mathcal{G}_t\right]}_{=1,\text{ by (3.A.9)}}}, \qquad \text{by (3.A.8)}, \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}\left[\frac{S_{t+1}}{S_t}\prod_{n=t+1}^{T}\widehat{\xi}_n\middle|\mathcal{G}_{t+1}\right]\middle|\mathcal{G}_t\right], \\
&= \mathbb{E}^{\mathbb{P}}\left[\frac{S_{t+1}}{S_t}\widehat{\xi}_{t+1}\underbrace{\mathbb{E}^{\mathbb{P}}\left[\prod_{n=t+2}^{T}\widehat{\xi}_n\middle|\mathcal{G}_{t+1}\right]}_{=1,\text{ by (3.A.9)}}\middle|\mathcal{G}_t\right], \qquad \text{by (3.A.8)}, \\
&= \mathbb{E}^{\mathbb{P}}\left[e^{-\theta_t}\left(\frac{S_{t+1}}{S_t}\right)^{\alpha_t+1}\middle|\mathcal{G}_t\right], \\
&= e^{-\theta_t}\mathbb{E}^{\mathbb{P}}\left[\exp\left((\alpha_t+1)\mu_{h_t}+(\alpha_t+1)\sigma_{h_t}z_{t+1}^{\mathbb{P}}\right)\middle|\mathcal{G}_t\right], \\
&= e^{-\theta_t}\sum_{k=1}^{H}\eta_{t,k}^{\mathbb{P}}\exp\left((\alpha_t+1)\mu_k+\frac{1}{2}(\alpha_t+1)^2\sigma_k^2\right). \qquad (3.A.10)
\end{aligned}$$

Combining (3.5.3) and (3.A.10) yields

$$\begin{aligned}
&\sum_{k=1}^{H}\eta_{t,k}^{\mathbb{P}}\exp\left(\alpha_t\mu_k+\frac{1}{2}\alpha_t^2\sigma_k^2\right) = \sum_{k=1}^{H}\eta_{t,k}^{\mathbb{P}}\exp\left((\alpha_t+1)\mu_k+\frac{1}{2}(\alpha_t+1)^2\sigma_k^2-r\right), \\
\Rightarrow\quad &\sum_{k=1}^{H}\eta_{t,k}^{\mathbb{P}}\exp\left(\alpha_t\mu_k+\frac{1}{2}\alpha_t^2\sigma_k^2\right)\left[1-\exp\left(\mu_k+\alpha_t\sigma_k^2+\frac{1}{2}\sigma_k^2-r\right)\right]=0.
\end{aligned}$$

3.A.9 Proof of Eq. (3.5.5)

$$\begin{aligned}
\widehat{\mathbb{Q}}\left[\epsilon_{t+1} \leq x | \mathcal{G}_t\right] &= \frac{\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}}\middle|\mathcal{G}_t\right]}{\mathbb{E}^{\mathbb{P}}\left[\frac{d\widehat{\mathbb{Q}}}{d\mathbb{P}}\middle|\mathcal{G}_t\right]}, \\
&= \frac{\prod_{n=1}^{t}\widehat{\xi}_n\,\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}\prod_{n=t+1}^{T}\widehat{\xi}_n\middle|\mathcal{G}_t\right]}{\prod_{n=1}^{t}\widehat{\xi}_n\,\underbrace{\mathbb{E}^{\mathbb{P}}\left[\prod_{n=t+1}^{T}\widehat{\xi}_n\middle|\mathcal{G}_t\right]}_{=1,\text{ by (3.A.9)}}}, \qquad \text{by (3.A.8)}, \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}\widehat{\xi}_{t+1}\underbrace{\mathbb{E}^{\mathbb{P}}\left[\prod_{n=t+2}^{T}\widehat{\xi}_n\middle|\mathcal{G}_{t+1}\right]}_{=1,\text{ by (3.A.9)}}\middle|\mathcal{G}_t\right], \qquad \text{by (3.A.8)}, \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}e^{-\theta_t+\alpha_t\epsilon_{t+1}}\middle|\mathcal{G}_t\right], \\
&= e^{-\theta_t}\sum_{i=1}^{H}\eta_{t,i}^{\mathbb{P}}\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\mu_i+\sigma_i z_{t+1}^{\mathbb{P}}\leq x\}}e^{\alpha_t\mu_i+\alpha_t\sigma_i z_{t+1}^{\mathbb{P}}}\middle|\mathcal{G}_t, h_t = i\right]. \qquad (3.A.11)
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\mu_i+\sigma_i z_{t+1}^{\mathbb{P}}\leq x\}}e^{\alpha_t\mu_i+\alpha_t\sigma_i z_{t+1}^{\mathbb{P}}}\middle|\mathcal{G}_t, h_t = i\right] &= \int_{-\infty}^{(x-\mu_i)/\sigma_i} e^{\alpha_t\mu_i+\alpha_t\sigma_i z}\phi(z)dz, \\
&= \int_{-\infty}^{(x-\mu_i)/\sigma_i} e^{\alpha_t\mu_i+\alpha_t\sigma_i z}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz, \\
&= \int_{-\infty}^{(x-\mu_i)/\sigma_i} e^{\alpha_t\mu_i+\alpha_t^2\sigma_i^2/2}\frac{1}{\sqrt{2\pi}}e^{-(z-\alpha_t\sigma_i)^2/2}dz, \\
&= e^{\alpha_t\mu_i+\alpha_t^2\sigma_i^2/2}\Phi\left(\frac{x-\mu_i}{\sigma_i}-\alpha_t\sigma_i\right). \qquad (3.A.12)
\end{aligned}$$

Plugging (3.A.7) and (3.A.12) in (3.A.11), we obtain

$$\begin{aligned}
\widehat{\mathbb{Q}}\left[\epsilon_{t+1} \leq x | \mathcal{G}_t\right] &= e^{-\theta_t}\sum_{i=1}^{H}\eta_{t,i}^{\mathbb{P}}e^{\alpha_t\mu_i+\alpha_t^2\sigma_i^2/2}\Phi\left(\frac{x-\mu_i-\alpha_t\sigma_i^2}{\sigma_i}\right), \\
&= \sum_{i=1}^{H}\frac{\eta_{t,i}^{\mathbb{P}}e^{\alpha_t\mu_i+\alpha_t^2\sigma_i^2/2}}{\sum_{k=1}^{H}\eta_{t,k}^{\mathbb{P}}e^{\alpha_t\mu_k+\alpha_t^2\sigma_k^2/2}}\Phi\left(\frac{x-\mu_i-\alpha_t\sigma_i^2}{\sigma_i}\right), \\
&= \sum_{i=1}^{H}\hat{\eta}_{t,i}^{\mathbb{P}}\Phi\left(\frac{x-\mu_i-\alpha_t\sigma_i^2}{\sigma_i}\right).
\end{aligned}$$

3.A.10 Proof of Eq. (3.5.8)

The PDF of a trajectory $(\epsilon_{1:T}, h_{0:T-1})$ under a generic probability measure $\bar{\mathbb{Q}}$ can be expressed as

$$f^{\bar{\mathbb{Q}}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1}) = f^{\bar{\mathbb{Q}}}_{\epsilon_{1:T}}(\epsilon_{1:T}) f^{\bar{\mathbb{Q}}}_{h_{0:T-1}|\mathcal{G}_T}(h_{0:T-1}|\mathcal{G}_T), \tag{3.A.13}$$

since $\mathcal{G}_T \equiv \sigma(\epsilon_{1:T})$. Moreover,

$$\begin{aligned} f^{\bar{\mathbb{Q}}}_{\epsilon_{1:T}}(\epsilon_{1:T}) &= f^{\bar{\mathbb{Q}}}_{\epsilon_{1:T-1}}(\epsilon_{1:T-1}) f^{\bar{\mathbb{Q}}}_{\epsilon_T|\mathcal{G}_{T-1}}(\epsilon_T|\mathcal{G}_{T-1}), \\ &\vdots \quad \text{(applying recursively)} \\ &= \prod_{t=1}^{T} f^{\bar{\mathbb{Q}}}_{\epsilon_t|\mathcal{G}_{t-1}}(\epsilon_t|\mathcal{G}_{t-1}). \end{aligned} \tag{3.A.14}$$

Combining (3.A.13) and (3.A.14) yields (3.5.8).

3.A.11 Proof of Eq. (3.5.9)

The expression (3.5.8) also holds for $\mathbb{P}$, i.e.,

$$f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1}) = f^{\mathbb{P}}_{h_{0:T-1}|\mathcal{G}_T}(h_{0:T-1}|\mathcal{G}_T) \prod_{t=1}^{T} f^{\mathbb{P}}_{\epsilon_t|\mathcal{G}_{t-1}}(\epsilon_t|\mathcal{G}_{t-1}). \tag{3.A.15}$$

Plugging the following concludes the proof :

$$f^{\mathbb{P}}_{\epsilon_t|\mathcal{G}_{t-1}}(\epsilon_t|\mathcal{G}_{t-1}) = \sum_{i=1}^{H} \underbrace{\mathbb{P}[h_{t-1} = i|\mathcal{G}_{t-1}]}_{\equiv \eta^{\mathbb{P}}_{t-1,i}} \underbrace{f^{\mathbb{P}}_{\epsilon_t|\mathcal{G}_{t-1},h_{t-1}}(\epsilon_t|\mathcal{G}_{t-1}, i)}_{= \phi^{\mathbb{P}}_i(\epsilon_t)}. \tag{3.A.16}$$

3.A.12 Proof of Eq. (3.5.12)

The Radon-Nikodym derivative is (from Theorem 3.3.1)

$$\frac{d\bar{\mathbb{Q}}}{d\mathbb{P}} \equiv \frac{f^{\bar{\mathbb{Q}}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})}{f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(\epsilon_{1:T}, h_{0:T-1})}. \tag{3.A.17}$$

Plugging Equation (3.5.8), (3.5.9), (3.5.10) and (3.5.11) yields (3.5.12).

3.A.13 Proof of Eq. (3.5.13)

The following property will be useful :

$$\bar{\xi}_s \text{ is } \mathcal{G}_t\text{-measurable}, \qquad \forall s \leq t. \tag{3.A.18}$$

Also, note that for all $t \geq 1$

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}}[\bar{\xi}_t|\mathcal{G}_{t-1}] &= \int_{-\infty}^{\infty} \left\{ \frac{\sum_{i=1}^{H} \zeta_i(\eta_{t-1}^{\mathbb{P}}) \phi_i^{\mathbb{Q}}(x)}{\sum_{i=1}^{H} \eta_{t-1,i}^{\mathbb{P}} \, \phi_i^{\mathbb{P}}(x)} f_{\epsilon_t|\mathcal{G}_{t-1}}^{\mathbb{P}}(x|\mathcal{G}_{t-1}) \right\} dx, \\
&= \int_{-\infty}^{\infty} \left\{ \frac{\sum_{i=1}^{H} \zeta_i(\eta_{t-1}^{\mathbb{P}}) \phi_i^{\mathbb{Q}}(x)}{\sum_{i=1}^{H} \eta_{t-1,i}^{\mathbb{P}} \, \phi_i^{\mathbb{P}}(x)} \sum_{i=1}^{H} \eta_{t-1,i}^{\mathbb{P}} \, \phi_i^{\mathbb{P}}(x) \right\} dx, \\
&= \sum_{i=1}^{H} \zeta_i(\eta_{t-1}^{\mathbb{P}}) \underbrace{\left[\int_{-\infty}^{\infty} \phi_i^{\mathbb{Q}}(x) dx \right]}_{=1}, \\
&= 1. \qquad (3.A.19)
\end{aligned}
$$

Furthermore, for all $t \geq 1$

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\prod_{s=t}^{T} \bar{\xi}_s \middle| \mathcal{G}_{t-1} \right] &= \mathbb{E}^{\mathbb{P}}\left[\prod_{s=t}^{T-1} \bar{\xi}_s \underbrace{\mathbb{E}^{\mathbb{P}}\left[\bar{\xi}_T \middle| \mathcal{G}_{T-1} \right]}_{=1, \text{ by } (3.A.19)} \middle| \mathcal{G}_{t-1} \right], \qquad \text{by } (3.A.18), \\
&\vdots \quad \text{(applying recursively)} \\
&= 1. \qquad (3.A.20)
\end{aligned}
$$

We are now ready to carry out the main proof :

$$
\begin{aligned}
\bar{\mathbb{Q}}[\epsilon_{t+1} \leq x|\mathcal{G}_t] &= \mathbb{E}^{\bar{\mathbb{Q}}}\big[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}\big|\mathcal{G}_t\big], \\
&\equiv \frac{\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}\frac{d\bar{\mathbb{Q}}}{d\mathbb{P}}\middle|\mathcal{G}_t\right]}{\mathbb{E}^{\mathbb{P}}\left[\frac{d\bar{\mathbb{Q}}}{d\mathbb{P}}\middle|\mathcal{G}_t\right]}, \\
&= \frac{\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}\prod_{s=1}^{T}\bar{\xi}_s\middle|\mathcal{G}_t\right]}{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{T}\bar{\xi}_s\middle|\mathcal{G}_t\right]}, \\
&= \frac{\prod_{s=1}^{t}\bar{\xi}_s\,\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}\prod_{s=t+1}^{T}\bar{\xi}_s\middle|\mathcal{G}_t\right]}{\prod_{s=1}^{t}\bar{\xi}_s\,\underbrace{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=t+1}^{T}\bar{\xi}_s\middle|\mathcal{G}_t\right]}_{=1,\text{ by (3.A.20)}}}, \qquad \text{by (3.A.18)}, \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}\bar{\xi}_{t+1}\,\underbrace{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=t+2}^{T}\bar{\xi}_s\middle|\mathcal{G}_{t+1}\right]}_{=1,\text{ by (3.A.20)}}\middle|\mathcal{G}_t\right], \qquad \text{by (3.A.18)}, \\
&= \mathbb{E}^{\mathbb{P}}\big[\mathbb{1}_{\{\epsilon_{t+1}\leq x\}}\bar{\xi}_{t+1}\big|\mathcal{G}_t\big], \\
&= \int_{-\infty}^{x}\left\{\frac{\sum_{i=1}^{H}\zeta_i(\eta_t^{\mathbb{P}})\phi_i^{\mathbb{Q}}(y)}{\sum_{i=1}^{H}\eta_{t,i}^{\mathbb{P}}\,\phi_i^{\mathbb{P}}(y)}f^{\mathbb{P}}_{\epsilon_{t+1}|\mathcal{G}_t}(y|\mathcal{G}_t)\right\}dy, \\
&= \int_{-\infty}^{x}\left\{\frac{\sum_{i=1}^{H}\zeta_i(\eta_t^{\mathbb{P}})\phi_i^{\mathbb{Q}}(y)}{\sum_{i=1}^{H}\eta_{t,i}^{\mathbb{P}}\,\phi_i^{\mathbb{P}}(y)}\sum_{i=1}^{H}\eta_{t,i}^{\mathbb{P}}\,\phi_i^{\mathbb{P}}(y)\right\}dy, \\
&= \sum_{i=1}^{H}\zeta_i(\eta_t^{\mathbb{P}})\left[\int_{-\infty}^{x}\phi_i^{\mathbb{Q}}(y)dy\right], \\
&= \sum_{i=1}^{H}\zeta_i(\eta_t^{\mathbb{P}})\Phi\left(\frac{x-r+\frac{1}{2}\sigma_i^2}{\sigma_i}\right).
\end{aligned}
$$

References

Ang, A. and Timmermann, A. (2012). Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1) :313–337.

Bégin, J.-F. and Gauthier, G. (2017). Price bias and common practice in option pricing. Working paper. Simon Fraser University.

Bollen, N. P. (1998). Valuing options in regime-switching models. *The Journal of Derivatives*, 6(1) :38–49.

Boyle, P. and Draviam, T. (2007). Pricing exotic options under regime switching. *Insurance : Mathematics and Economics*, 40(2) :267–282.

Buffington, J. and Elliott, R. J. (2002a). American options with regime switching. *International Journal of Theoretical and Applied Finance*, 5(05) :497–514.

Buffington, J. and Elliott, R. J. (2002b). Regime switching and European options. In *Stochastic Theory and Control*, pages 73–82. Springer.

Christoffersen, P., Elkamhi, R., Feunou, B., and Jacobs, K. (2009). Option valuation with conditional heteroskedasticity and nonnormality. *The Review of Financial Studies*, 23(5) :2139–2183.

Di Masi, G. B., Kabanov, Y. M., and Runggaldier, W. J. (1995). Mean-variance hedging of options on stocks with Markov volatilities. *Theory of Probability & Its Applications*, 39(1) :172–182.

Duan, J.-C., Popova, I., Ritchken, P., et al. (2002). Option pricing under regime switching. *Quantitative Finance*, 2(116-132) :209.

Elliott, R. J., Chan, L., and Siu, T. K. (2005). Option pricing and Esscher transform under regime switching. *Annals of Finance*, 1(4) :423–432.

Elliott, R. J. and Siu, T. K. (2013). Option pricing and filtering with hidden Markov-modulated pure-jump processes. *Applied Mathematical Finance*, 20(1) :1–25.

Elliott, R. J., Siu, T. K., Chan, L., and Lau, J. W. (2007). Pricing options under a generalized Markov-modulated jump-diffusion model. *Stochastic Analysis and Applications*, 25(4) :821–843.

François, P., Gauthier, G., and Godin, F. (2014). Optimal hedging when the underlying asset follows a regime-switching Markov process. *European Journal of Operational Research*, 237(1) :312–322.

Fuh, C.-D., Ho, K. W. R., Inchi, H., and Wang, R.-H. (2012). Option pricing with Markov switching. *Journal of Data Science*, 10(3) :483–509.

Garcia, R., Luger, R., and Renault, E. (2003). Empirical assessment of an intertemporal option pricing model with latent variables. *Journal of Econometrics*, 116(1) :49–83.

Guo, X. (2001). Information and option pricings. *Quantitative Finance*, 1(1) :38–44.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2) :357–384.

Hardy, M. R. (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2) :41–53.

Hardy, M. R. (2003). *Investment Guarantees : Modeling and Risk Management for Equity-Linked Life Insurance*. John Wiley & Sons, Ltd.

Harrison, J. M. and Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, 20(3) :381–408.

Ishijima, H. and Kihara, T. (2005). Option pricing with hidden Markov models. Institute of Finance, Waseda University.

Jobert, A. and Rogers, L. C. (2006). Option pricing with Markov-modulated dynamics. *SIAM Journal on Control and Optimization*, 44(6) :2063–2078.

Lamberton, D. and Lapeyre, B. (2007). *Introduction to stochastic calculus applied to finance*. CRC press.

Liew, C. C. and Siu, T. K. (2010). A hidden Markov regime-switching model for option valuation. *Insurance : Mathematics and Economics*, 47(3) :374–384.

Liu, R., Zhang, Q., and Yin, G. (2006). Option pricing in a regime-switching model using the fast Fourier transform. *International Journal of Stochastic Analysis*, 2006.

Mamon, R. S. and Rodrigo, M. R. (2005). Explicit solutions to European options in a regime-switching economy. *Operations Research Letters*, 33(6) :581–586.

Naik, V. (1993). Option valuation and hedging strategies with jumps in the volatility of asset returns. *The Journal of Finance*, 48(5) :1969–1984.

Ranjbar, H. R. and Seifi, A. (2015). A path-independent method for barrier option pricing in hidden Markov models. *Physica A : Statistical Mechanics and its Applications*, 440 :1–8.

Rémillard, B., Hocquard, A., Lamarre, H., and Papageorgiou, N. (2017). Option pricing and hedging for discrete time regime-switching models. *Modern Economy*, 8 :1005–1032.

Shen, Y. and Siu, T. K. (2013). Stochastic differential game, Esscher transform and general equilibrium under a Markovian regime-switching Lévy model. *Insurance : Mathematics and Economics*, 53(3) :757–768.

Trottier, D.-A., Godin, F., and Hamel, E. (2017). Local hedging of variable annuities in the presence of basis risk. Working paper. Available at SSRN : https ://ssrn.com/abstract=2997118.

Yao, D. D., Zhang, Q., and Zhou, X. Y. (2006). A regime-switching model for European options. *Stochastic processes, optimization, and control theory : applications in financial engineering, queueing networks, and manufacturing systems*, pages 281–300.

Yuen, F. L. and Yang, H. (2009). Option pricing in a jump-diffusion model with regime switching. *ASTIN Bulletin : The Journal of the IAA*, 39(2) :515–539.

Zhang, Q. and Guo, X. (2004). Closed-form solutions for perpetual American put options with regime switching. *SIAM Journal on Applied Mathematics*, 64(6) :2034–2049.

Online Appendix (not part of the paper)

3.B A simple motivational example

This section presents a simple motivational example in which a Markovian underlying asset price process leads to a path dependent option price process for a vanilla call option.

Consider a three-period trinomial tree model with $\Omega = \big\{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{u, m, d\},\ i = 1, 2, 3\big\}$. Here, $\omega = (\omega_1, \omega_2, \omega_3)$ represents a trajectory, and u, m, d respectively stand for “up", “middle" and “down" to denote the underlying asset stock price movements. Consider an economy with a null risk-free rate for simplicity and with a stock price process $\{S_t\}_{t=0}^{3}$ evolving according to

$$S_{t+1}(\omega) = \begin{cases} S_t(\omega) + 1, & \text{if } \omega_{t+1} = u, \\ S_t(\omega), & \text{if } \omega_{t+1} = m, \\ S_t(\omega) - 1, & \text{if } \omega_{t+1} = d, \end{cases}$$

with $S_0(\omega) = 5$ for all $\omega \in \Omega$. We consider a physical probability measure $\mathbb{P}$ under which the stock price can go up, down or middle with the same probability $1/3$ at every node. This leads to $\mathbb{P}(\omega) = (1/3)^3 = 1/27$ for any $\omega \in \Omega$. The information on the market is characterized by the filtration $\mathcal{G} = \{\mathcal{G}_t\}_{t=0}^{3}$ where $\mathcal{G}_t = \sigma(S_0, \ldots, S_t)$ is the sigma-algebra generated by previous and contemporaneous stock prices.

Consider an at-the-money vanilla European call option maturing at time 3 whose payoff is $\Psi(S_3) = \max(0, S_3 - 5)$. Let $\mathcal{Q}$ be the set of all martingale measures on the space.[2] From the first fundamental theorem of financial mathematics (see, e.g., Lamberton and Lapeyre, 2007), the set of all possible option price processes $\Pi = \{\Pi_t\}_{t=0}^{3}$ that do not generate arbitrage opportunities is

$$\Big\{\Pi^{\mathbb{Q}} = \big\{\mathbb{E}^{\mathbb{Q}}[\Psi(S_3)|\mathcal{G}_t]\big\}_{t=0}^{3} : \mathbb{Q} \in \mathcal{Q}\Big\}.$$

2. $\mathbb{Q}$ is a martingale measure if $\mathbb{Q}$ is equivalent to $\mathbb{P}$ and if S is a $\mathbb{Q}$-martingale.

Since S is a $\mathbb{P}$-martingale, $\mathbb{P} \in \mathcal{Q}$. Therefore, since S is a Markov process under $\mathbb{P}$, using $\mathbb{Q} = \mathbb{P}$ as the pricing measure leads to a Markovian price process. However, it is possible to construct a martingale measure under which S loses the Markov property. Indeed consider $\mathbb{Z}$ defined through

$$\mathbb{Z}[\omega] = \begin{cases} 1/54, & \text{if } \omega = (m,m,u) \text{ or } \omega = (m,m,d), \\ 2/27, & \text{if } \omega = (m,m,m), \\ 1/27, & \text{otherwise .} \end{cases}$$

This measure changes the probability of an upward or a downward movement to 1/6 instead of 1/3 at the last time step if the two first stock movements were $\omega_1 = \omega_2 = m$. Straightforward calculations show that S is a $\mathbb{Z}$-martingale. However, S is not a Markov process under $\mathbb{Z}$ since $\mathbb{Z}\left[S_3 = 6 | S_2 = 5, S_1 = 5\right] = \frac{1/54}{1/9} = 1/6$ is not equal to $\mathbb{Z}\left[S_3 = 6 | S_2 = 5, S_1 = 6\right] = \frac{1/27}{1/9} = 1/3$. The non-Markovian dynamics of S under $\mathbb{Z}$ leads to a non-Markovian option price under $\mathbb{P}$ if the pricing measure is $\mathbb{Z}$. For example, let $\Pi_t^{\mathbb{Z}} \equiv \mathbb{E}^{\mathbb{Z}}\left[\Psi(S_3) | \mathcal{G}_t\right]$, then for any $\omega_3 \in \{u, d, m\}$,

$$\Pi_2^{\mathbb{Z}}(u, d, \omega_3) = (6-5) \times 1/3 = 1/3, \qquad \Pi_2^{\mathbb{Z}}(m, m, \omega_3) = (6-5) \times 1/6 = 1/6.$$

Hence, $\Pi_2^{\mathbb{Z}}(u, d, \omega_3) \neq \Pi_2^{\mathbb{Z}}(m, m, \omega_3)$ and this even if $S_2(u, d, \omega_3) = S_2(m, m, \omega_3) = 5$.

Although $\left\{\Pi_t^{\mathbb{Z}}\right\}_{t=0}^{3}$ is an arbitrage-free price process consistent with arbitrage pricing theory, it can be argued that it entails unintuitive properties. When an investor observes that the stock price at time 2 is $S_2 = 5$, he has a complete knowledge of the likelihood of all possible scenarios until maturity ($\omega_3 = u, d$ or m) under the physical measure $\mathbb{P}$; previous information such as S_1 does not alter the likelihood of each scenario. When $S_2 = 5$ at time 2, intuitively the investor should be indifferent to the value of S_1 when determining the option price. In other words, non-Markovian option prices under $\mathbb{P}$ when the underlying asset has the Markov property under $\mathbb{P}$ could be deemed counterintuitive. Thus, it is legitimate to search for martingale measures that do not exhibit such properties.

3.C Non-trivial RS mean-correcting measure without path dependence

This appendix presents the construction of another martingale measure $\mathbb{M}$ satisfying (3.4.19), which is not the trivial solution to the latter system of equations. Indeed, among the infinite set of martingale measures $\mathbb{M}$ satisfying (3.4.19), some are arguably more intuitive than others. Consider for instance the identity conversion case, see Section 3.4.2. Property (3.4.11) implies that the regime process h is not a Markov process under $\mathbb{M}$, which is a strange property. Furthermore, this is the case even if the discounted price process $\{e^{-rt}S_t\}_{t=0}^{T}$ is already a

$\mathbb{P}$-martingale, i.e., when $\lambda_t = 0$ almost surely. Indeed, using (3.4.8) and $\lambda_t = 0$ in (3.4.5) yields

$$\frac{d\mathbb{M}}{d\mathbb{P}} = \frac{f^{\mathbb{M}}_{h_0}(h_0)}{f^{\mathbb{P}}_{h_0}(h_0)} \prod_{t=2}^{T} \psi_{t-1}^{(h_{t-2},h_{t-1})} = \frac{f^{\mathbb{M}}_{h_0}(h_0)}{f^{\mathbb{P}}_{h_0}(h_0)} \prod_{t=2}^{T} \frac{\eta^{\mathbb{P}}_{t,j}}{P_{j,i}},$$

which means that a change of measure is still applied even though the martingale property is already satisfied under $\mathbb{P}$.

However, as stated in Proposition 3.4.2, there is an infinite number of other solutions. Among all the solutions, a specific one stated in the proposition below allows minimizing the disparity between the transition probabilities under $\mathbb{P}$ and $\mathbb{M}$, the latter being given by (3.4.3). As stated in the latter proposition, the choice (3.C.2) yields $\frac{d\mathbb{M}}{d\mathbb{P}} = \frac{f^{\mathbb{M}}_{h_0}(h_0)}{f^{\mathbb{P}}_{h_0}(h_0)}$ if the discounted underlying price is a $\mathbb{P}$-martingale. The proof of the following proposition is in the Online Appendix 3.D.4.

Proposition 3.C.1. *The minimization problem*

$$\min_{\left\{\psi_t^{(j,i)}\right\}_{i,j=1}^{H}} \sum_{i,j=1}^{H} \left(P_{j,i}\psi_t^{(j,i)} - P_{j,i}\right)^2 \tag{3.C.1}$$

subject to the constraints (3.4.19) *admits the solution*

$$\psi_t^{(j,i)} = 1 + \frac{\phi_j^{\mathbb{Q}}(\epsilon_t)\zeta_j(\eta^{\mathbb{P}}_{t-1})}{P_{j,i}} \frac{\sum_{\ell=1}^{H} \phi_\ell^{\mathbb{Q}}(\epsilon_t)\zeta_\ell(\eta^{\mathbb{P}}_{t-1}) \left(\zeta_i(\eta^{\mathbb{P}}_t) - P_{\ell,i}\right)}{\sum_{\ell=1}^{H} \left(\phi_\ell^{\mathbb{Q}}(\epsilon_t)\zeta_\ell(\eta^{\mathbb{P}}_{t-1})\right)^2}. \tag{3.C.2}$$

When $\phi_\ell^{\mathbb{Q}} = \phi_\ell^{\mathbb{P}}$, i.e., when $\lambda_t = 0$ almost surely, and when the identity conversion is considered, i.e., $\zeta_i(\eta^{\mathbb{P}}_t) = \eta^{\mathbb{P}}_{t,i}$ almost surely for $i = 1, \ldots, H$ and $t = 0, \ldots, T-1$, the expression (3.C.2) *simplifies to $\psi_t^{(j,i)} = 1$.*

3.D Additional proofs

3.D.1 Proof of Proposition 3.3.1

Define $\mathbb{H} = \{1, \ldots, H\}$, then

$$\mathbb{E}^{\mathbb{P}}\left[\frac{d\mathbb{Z}}{d\mathbb{P}}\right] \equiv \sum_{y\in\mathbb{H}^T} \int_{x\in\mathbb{R}^T} \frac{f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}(x,y)}{f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(x,y)} f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(x,y)dx = \sum_{y\in\mathbb{H}^T} \int_{x\in\mathbb{R}^T} f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}(x,y)dx = 1.$$

Since $f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}$ is non-negative this implies $\mathbb{Z}[A] \in [0,1]$ for all $A \in \mathcal{F}$. Clearly, $\mathbb{Z}$ is also sigma-additive and therefore is a probability measure. Furthermore, since $f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}$ is non-negative, showing the equivalence of $\mathbb{Z}$ and $\mathbb{P}$ is directly obtained from the properties of expectation.

Finally, for all Borel sets $E \times F \subseteq \mathbb{R}^T \times \mathbb{H}^T$,

$$\begin{aligned}
\mathbb{Z}[(\epsilon_{1:T}, h_{0:T-1}) \in E \times F] &= \sum_{y \in F} \int_{x \in E} \frac{f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}(x,y)}{f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(x,y)} f^{\mathbb{P}}_{\epsilon_{1:T},h_{0:T-1}}(x,y)dx, \\
&= \sum_{y \in F} \int_{x \in E} f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}(x,y)dx,
\end{aligned}$$

and thus $f^{\mathbb{Z}}_{\epsilon_{1:T},h_{0:T-1}}$ is the joint mixed PDF of $\epsilon_{1:T}, h_{0:T-1}$ under the probability measure $\mathbb{Z}$.

3.D.2 Proof of Eq. (3.3.8)

The following property will be useful :

$$\xi_s \text{ is } \mathcal{F}_t\text{-measurable}, \qquad \forall s \leq t. \tag{3.D.1}$$

Also, note that for all $t \geq 1$

$$\mathbb{E}^{\mathbb{P}}[\xi_t | \mathcal{F}_{t-1}] = 1. \tag{3.D.2}$$

Furthermore, for all $t \geq 1$

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\prod_{s=t}^{T} \xi_s \middle| \mathcal{F}_{t-1}\right] &= \mathbb{E}^{\mathbb{P}}\left[\prod_{s=t}^{T-1} \xi_s \underbrace{\mathbb{E}^{\mathbb{P}}\left[\xi_T \middle| \mathcal{F}_{T-1}\right]}_{=1, \text{ by (3.D.2)}} \middle| \mathcal{F}_{t-1}\right], \qquad \text{by (3.D.1)}, \\
&\vdots \quad \text{(applying recursively)} \\
&= 1.
\end{aligned} \tag{3.D.3}$$

Hence, the relation between $\eta^{\mathbb{Q}}_t$ and $\eta^{\mathbb{P}}_t$ is

$$\begin{aligned}
\eta^{\mathbb{Q}}_{t,i} \equiv \mathbb{Q}(h_t = i | \mathcal{G}_t) &= \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{h_t = i\}} | \mathcal{G}_t], \\
&= \frac{\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{h_t=i\}} \prod_{s=1}^{T} \xi_s \middle| \mathcal{G}_t\right]}{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{T} \xi_s \middle| \mathcal{G}_t\right]}, \\
&= \frac{\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{h_t=i\}} \prod_{s=1}^{t} \xi_s \overbrace{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=t+1}^{T} \xi_s \middle| \mathcal{G}_t, \mathcal{H}_{t-1}, h_t\right]}^{=1, \text{ by (3.D.3)}} \middle| \mathcal{G}_t\right]}{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t} \xi_s \underbrace{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=t+1}^{T} \xi_s \middle| \mathcal{F}_t\right]}_{=1, \text{ by (3.D.3)}} \middle| \mathcal{G}_t\right]}, \quad \text{by (3.D.1)}, \\
&= \frac{\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{h_t=i\}} \mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t} \xi_s \middle| \mathcal{G}_t, h_t\right] \middle| \mathcal{G}_t\right]}{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t} \xi_s \middle| \mathcal{G}_t\right]}, \\
&= \eta^{\mathbb{P}}_{t,i} \frac{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t} \xi_s \middle| \mathcal{G}_t, h_t = i\right]}{\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t} \xi_s \middle| \mathcal{G}_t\right]}.
\end{aligned} \tag{3.D.4}$$

Following similar steps, one can show that

$$\Pi_t^{\mathbb{Q}} = \mathbb{E}^{\mathbb{Q}}\big[e^{-r}\Pi_{t+1}^{\mathbb{Q}}\big|\mathcal{G}_t\big] = \mathbb{E}^{\mathbb{P}}\big[\Pi_{t+1}^{\mathbb{Q}} m_{t+1}^{\mathbb{Q}}\big|\mathcal{G}_t\big], \tag{3.D.5}$$

where the SDF is

$$m_{t+1}^{\mathbb{Q}} \equiv e^{-r}\frac{\mathbb{E}^{\mathbb{P}}\Big[\prod_{s=1}^{t+1}\xi_s\Big|\mathcal{G}_{t+1}\Big]}{\mathbb{E}^{\mathbb{P}}\Big[\prod_{s=1}^{t}\xi_s\Big|\mathcal{G}_t\Big]}. \tag{3.D.6}$$

The above expression can be simplified. First, since ξ_{t+1} is $\mathcal{G}_{t+1}\vee\sigma(h_t)$-measurable,

$$\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t+1}\xi_s\middle|\mathcal{G}_{t+1}\right] = \mathbb{E}^{\mathbb{P}}\left[\xi_{t+1}\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t}\xi_s\middle|\mathcal{G}_{t+1},h_t\right]\middle|\mathcal{G}_{t+1}\right].$$

Furthermore, note that

$$\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t}\xi_s\middle|\mathcal{G}_{t+1},h_t\right] = \mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t}\xi_s\middle|\mathcal{G}_{t},h_t\right].$$

Hence,

$$\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t+1}\xi_s\middle|\mathcal{G}_{t+1}\right] = \mathbb{E}^{\mathbb{P}}\left[\xi_{t+1}\mathbb{E}^{\mathbb{P}}\left[\prod_{s=1}^{t}\xi_s\middle|\mathcal{G}_{t},h_t\right]\middle|\mathcal{G}_{t+1}\right]. \tag{3.D.7}$$

Using the above expression in (3.D.6) yields

$$m_{t+1}^{\mathbb{Q}} = e^{-r}\mathbb{E}^{\mathbb{P}}\left[\xi_{t+1}\frac{\mathbb{E}^{\mathbb{P}}\big[\prod_{s=1}^{t}\xi_s\big|\mathcal{G}_t,h_t\big]}{\mathbb{E}^{\mathbb{P}}\big[\prod_{s=1}^{t}\xi_s\big|\mathcal{G}_t\big]}\middle|\mathcal{G}_{t+1}\right] = e^{-r}\mathbb{E}^{\mathbb{P}}\left[\xi_{t+1}\frac{\eta_{t,h_t}^{\mathbb{Q}}}{\eta_{t,h_t}^{\mathbb{P}}}\middle|\mathcal{G}_{t+1}\right], \quad \text{by (3.D.4).} \tag{3.D.8}$$

One last step is required to further simplify this expression. Note that, see (3.D.11),

$$\xi_{t+1} = \frac{\phi_{h_t}^{\mathbb{Q}}(\epsilon_{t+1})}{\phi_{h_t}^{\mathbb{P}}(\epsilon_{t+1})},$$

and also, using Bayes' theorem, that

$$\mathbb{P}(h_t = i|\mathcal{G}_{t+1}) = \frac{\phi_i^{\mathbb{P}}(\epsilon_{t+1})\eta_{t,i}^{\mathbb{P}}}{\sum_{j=1}^{H}\phi_j^{\mathbb{P}}(\epsilon_{t+1})\eta_{t,j}^{\mathbb{P}}}.$$

Using the above expressions in (3.D.8) yields

$$m_{t+1}^{\mathbb{Q}} = e^{-r}\frac{\sum_{i=1}^{H}\eta_{t,i}^{\mathbb{Q}}\phi_i^{\mathbb{Q}}(\epsilon_{t+1})}{\sum_{i=1}^{H}\eta_{t,i}^{\mathbb{P}}\phi_i^{\mathbb{P}}(\epsilon_{t+1})}, \tag{3.D.9}$$

which concludes the proof.

3.D.3 Proof of Eq. (3.3.5)

Using (3.2.5) and (3.3.1) in (3.3.4) yields

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \prod_{t=1}^{T} \xi_t, \tag{3.D.10}$$

where

$$\xi_t \equiv \frac{\phi^{\mathbb{Q}}_{h_{t-1}}(\epsilon_t)}{\phi^{\mathbb{P}}_{h_{t-1}}(\epsilon_t)}. \tag{3.D.11}$$

The definitions (3.2.6) and (3.3.2) can then be used to simplify :

$$\begin{aligned}
\xi_t &= \frac{\phi\left(\frac{\epsilon_t - r + \frac{1}{2}\sigma^2_{h_{t-1}}}{\sigma_{h_{t-1}}}\right)}{\phi\left(\frac{\epsilon_t - \mu_{h_{t-1}}}{\sigma_{h_{t-1}}}\right)}, \\
&= \exp\left[-\frac{1}{2\sigma^2_{h_{t-1}}}\left(\left(\epsilon_t - r + \sigma^2_{h_{t-1}}/2\right)^2 - \left(\epsilon_t - \mu_{h_{t-1}}\right)^2\right)\right], \\
&= \exp\left[-\frac{1}{2\sigma^2_{h_{t-1}}}\left(\left(\epsilon_t - \mu_{h_{t-1}} + \mu_{h_{t-1}} - r + \sigma^2_{h_{t-1}}/2\right)^2 - \left(\epsilon_t - \mu_{h_{t-1}}\right)^2\right)\right], \\
&= \exp\left[-\frac{1}{2\sigma^2_{h_{t-1}}}\left(2\left(\epsilon_t - \mu_{h_{t-1}}\right)\left(\mu_{h_{t-1}} - r + \sigma^2_{h_{t-1}}/2\right) + \left(\mu_{h_{t-1}} - r + \sigma^2_{h_{t-1}}/2\right)^2\right)\right].
\end{aligned}$$

Defining the Girsanov kernel $\lambda_t = -\frac{\mu_{h_{t-1}} - r + \frac{1}{2}\sigma^2_{h_{t-1}}}{\sigma_{h_{t-1}}}$,

$$\xi_t = \exp\left[\frac{(\epsilon_t - \mu_{h_{t-1}})\lambda_t}{\sigma_{h_{t-1}}} - \frac{1}{2}\lambda_t^2\right] = e^{z^{\mathbb{P}}_t \lambda_t - \frac{1}{2}\lambda_t^2}. \tag{3.D.12}$$

This concludes the proof.

3.D.4 Proof of Proposition 3.C.1

The Lagrangian of the optimization problem (3.C.1) under constraints (3.4.19) is

$$\mathcal{L} = \sum_{i,j=1}^{H}\left(P_{i,j}\psi_t^{(i,j)} - P_{i,j}\right)^2 - \sum_{i=1}^{H}\nu_i \sum_{k=1}^{H}\kappa_{t,k}\left(\psi_t^{(k,i)}P_{k,i} - \zeta_i(\eta_t^{\mathbb{P}})\right) - \sum_{k=1}^{H}\tau_k\left(\sum_{i=1}^{H}\psi_t^{(k,i)}P_{k,i} - 1\right),$$

where $\{\nu_i\}_{i=1}^H$ and $\{\tau_k\}_{k=1}^H$ are the Lagrange multipliers and $\kappa_{t,k} \equiv \phi_k^{\mathbb{Q}}(\epsilon_t)\zeta_k(\eta^{\mathbb{P}}_{t-1})$.

The first order conditions are

$$\frac{\partial \mathcal{L}}{\partial \psi_t^{(k,i)}} = 2P_{k,i}\left(P_{k,i}\psi_t^{(k,i)} - P_{k,i}\right) - \nu_i\kappa_{t,k}P_{k,i} - \tau_k P_{k,i} = 0, \qquad i,k \in \{1,\dots,H\}, \tag{3.D.13}$$

$$\frac{\partial \mathcal{L}}{\partial \nu_i} = \sum_{k=1}^{H}\kappa_{t,k}\left(\psi_t^{(k,i)}P_{k,i} - \zeta_i(\eta_t^{\mathbb{P}})\right) = 0, \qquad i \in \{1,\dots,H\}, \tag{3.D.14}$$

$$\frac{\partial \mathcal{L}}{\partial \tau_k} = \sum_{i=1}^{H}\psi_t^{(k,i)}P_{k,i} - 1 = 0, \qquad k \in \{1,\dots,H\}. \tag{3.D.15}$$

Dividing (3.D.13) by $P_{k,i}$, summing across all values of $i = 1, \ldots, H$ and using (3.D.15) yields

$$-\kappa_{t,k}\sum_{i=1}^{H}\nu_i - H\tau_k = 0, \qquad k \in \{1,\ldots,H\},$$

and thus

$$\tau_k = -\frac{\kappa_{t,k}}{H}\sum_{i=1}^{H}\nu_i, \qquad k \in \{1,\ldots,H\}. \tag{3.D.16}$$

Moreover, multiplying (3.D.13) by $\frac{\kappa_{t,k}}{P_{k,i}}$ and summing across all values of k and using (3.D.14) yields

$$\begin{aligned} 0 &= 2\sum_{k=1}^{H}\kappa_{t,k}\left(P_{k,i}\psi_t^{(k,i)} - P_{k,i}\right) - \nu_i\sum_{k=1}^{H}\kappa_{t,k}^2 - \sum_{k=1}^{H}\tau_k\kappa_{t,k}, \\ &= 2\sum_{k=1}^{H}\kappa_{t,k}\left(\zeta_i(\eta_t^{\mathbb{P}}) - P_{k,i}\right) - \nu_i\sum_{k=1}^{H}\kappa_{t,k}^2 - \sum_{k=1}^{H}\tau_k\kappa_{t,k}, \qquad i \in \{1,\ldots,H\}. \end{aligned}$$

Using (3.D.16) in the latter expression implies, for $i \in \{1,\ldots,H\}$, that

$$\begin{aligned} \nu_i\sum_{k=1}^{H}\kappa_{t,k}^2 - \sum_{k=1}^{H}\frac{\kappa_{t,k}^2}{H}\sum_{i=1}^{H}\nu_i &= 2\sum_{k=1}^{H}\kappa_{t,k}\left(\zeta_i(\eta_t^{\mathbb{P}}) - P_{k,i}\right), \\ \Rightarrow \sum_{k=1}^{H}\kappa_{t,k}^2\left(\nu_i - \frac{1}{H}\sum_{\ell=1}^{H}\nu_\ell\right) &= 2\sum_{k=1}^{H}\kappa_{t,k}\left(\zeta_i(\eta_t^{\mathbb{P}}) - P_{k,i}\right), \\ \Rightarrow \nu_i - \frac{1}{H}\sum_{\ell=1}^{H}\nu_\ell &= \frac{2\sum_{k=1}^{H}\kappa_{t,k}\left(\zeta_i(\eta_t^{\mathbb{P}}) - P_{k,i}\right)}{\sum_{k=1}^{H}\kappa_{t,k}^2}. \end{aligned} \tag{3.D.17}$$

Furthermore, using (3.D.16) in (3.D.13) leads to

$$\begin{aligned} 2P_{k,i}\left(P_{k,i}\psi_t^{(k,i)} - P_{k,i}\right) &= \nu_i\kappa_{t,k}P_{k,i} - \frac{\kappa_{t,k}}{H}\sum_{\ell=1}^{H}\nu_\ell P_{k,i}, \\ &= \kappa_{t,k}P_{k,i}\left(\nu_i - \frac{1}{H}\sum_{\ell=1}^{H}\nu_\ell\right), \qquad i,k \in \{1,\ldots,H\}. \end{aligned}$$

Using (3.D.17) then yields

$$\begin{aligned} 2P_{k,i}\left(P_{k,i}\psi_t^{(k,i)} - P_{k,i}\right) &= 2\kappa_{t,k}P_{k,i}\frac{\sum_{\ell=1}^{H}\kappa_{t,\ell}\left(\zeta_i(\eta_t^{\mathbb{P}}) - P_{\ell,i}\right)}{\sum_{\ell=1}^{H}\kappa_{t,\ell}^2}, \qquad i,k \in \{1,\ldots,H\}. \\ \Rightarrow \psi_t^{(k,i)} &= 1 + \frac{\kappa_{t,k}}{P_{k,i}}\frac{\sum_{\ell=1}^{H}\kappa_{t,\ell}\left(\zeta_i(\eta_t^{\mathbb{P}}) - P_{\ell,i}\right)}{\sum_{\ell=1}^{H}\kappa_{t,\ell}^2}, \qquad i,k \in \{1,\ldots,H\}. \end{aligned}$$

At last, we use the definition $\kappa_{t,k} \equiv \phi_k^{\mathbb{Q}}(\epsilon_t)\zeta_k(\eta_{t-1}^{\mathbb{P}})$ to obtain

$$\psi_t^{(k,i)} = 1 + \frac{\phi_k^{\mathbb{Q}}(\epsilon_t)\zeta_k(\eta_{t-1}^{\mathbb{P}})}{P_{k,i}}\frac{\sum_{\ell=1}^{H}\phi_\ell^{\mathbb{Q}}(\epsilon_t)\zeta_\ell(\eta_{t-1}^{\mathbb{P}})\left(\zeta_i(\eta_t^{\mathbb{P}}) - P_{\ell,i}\right)}{\sum_{\ell=1}^{H}\left(\phi_\ell^{\mathbb{Q}}(\epsilon_t)\zeta_\ell(\eta_{t-1}^{\mathbb{P}})\right)^2}, \qquad i,k \in \{1,\ldots,H\}, \tag{3.D.18}$$

which concludes the proof of (3.C.2).

Finally, when $\phi_\ell^{\mathbb{Q}} = \phi_\ell^{\mathbb{P}}$ and if the identity conversion is considered, i.e., $\zeta_i(\eta_t^{\mathbb{P}}) = \eta_{t,i}^{\mathbb{P}}$ almost surely for $i \in \{1, \ldots, H\}$ and $t \in \{0, \ldots, T-1\}$, it follows that

$$\begin{aligned}
\sum_{\ell=1}^{H} \phi_\ell^{\mathbb{Q}}(\epsilon_t)\zeta_\ell(\eta_{t-1}^{\mathbb{P}}) \left(\zeta_i(\eta_t^{\mathbb{P}}) - P_{\ell,i}\right) &= \sum_{\ell=1}^{H} \phi_\ell^{\mathbb{P}}(\epsilon_t)\eta_{t-1,\ell}^{\mathbb{P}} \left(\eta_{t,i}^{\mathbb{P}} - P_{\ell,i}\right) \\
&= \eta_{t,i}^{\mathbb{P}} \sum_{\ell=1}^{H} \phi_\ell^{\mathbb{P}}(\epsilon_t)\eta_{t-1,\ell}^{\mathbb{P}} - \sum_{\ell=1}^{H} \phi_\ell^{\mathbb{P}}(\epsilon_t)\eta_{t-1,\ell}^{\mathbb{P}} P_{\ell,i}, \\
&= 0, \quad \text{from (3.2.8)}.
\end{aligned}$$

The expression (3.D.18) therefore implies that $\psi_t^{(k,i)} = 1$, $i, k \in \{1, \ldots, H\}$, under such conditions.

Conclusion

Cette thèse est constituée de trois projets de recherche portant sur la modélisation, la couverture et la tarification des risques en finance et assurance.

Le chapitre 1 propose une classe générale d'opérateurs de distorsion compatibles avec la valorisation par absence d'arbitrage. Cette méthodologie générale est utilisée dans l'article afin de dériver plusieurs nouvelles distorsions qui améliorent et généralisent la transformation de Wang. Une étude empirique visant à expliquer les primes de risque observées sur le marché des obligations catastrophes est ensuite effectuée. En outre, une nouvelle classe générale et simple d'opérateurs de distorsion basée sur le changement de mesure de Esscher pour les risques avec une structure multi-états est proposée à cette fin. L'étude empirique révèle qu'un opérateur de distorsion basé sur un modèle de mouvement Brownien avec sauts exponentiels est adéquat pour expliquer les primes de risque d'obligations catastrophes, mais seulement si on permet à la distorsion d'intégrer l'aversion au risque de catastrophes naturelles. Cela remet en question l'hypothèse populaire dans la littérature théorique selon laquelle les catastrophes naturelles sont traitées comme un facteur de risque idiosyncratique.

Le chapitre 2 développe une méthodologie efficace et souple pour couvrir le risque lié aux polices de rentes variables en présence de risque de base. L'approche est basée sur une couverture locale du risque, avec une saveur de couverture globale qui consiste à optimiser la mesure de risque locale afin de minimiser les fonds propres requis. Cela permet notamment d'incorporer la diversification temporelle du risque dans la conception de la stratégie de couverture, ce qui permet à notre méthode de surpasser les benchmarks (e.g., couverture à variance minimale) dans le cadre d'expériences de simulations. La sur-performance de notre méthode peut plus précisément être expliquée par le fait que la minimisation locale du risque est sous-optimale en présence du risque de base qui réduit substantiellement la capacité à éliminer le risque ; le coût implicite engendré par la position courte dans un actif dont la valeur augmente sur un long horizon est alors moins rentable en terme de réduction du risque, et de plus une telle approche néglige la diversification temporelle comme source naturelle de réduction du risque.

Le chapitre 3 montre que l'approche habituelle pour construire la mesure neutre au risque (i.e., changement du paramètre de dérive pour obtenir la propriété martingale) conduit à des prix d'options avec dépendance au chemin lorsque appliquée en contexte de sous-jacent avec dynamique à changement de régime. La construction de plusieurs mesures neutres au risque possédant des propriétés intuitives et supprimant la dépendance au chemin est illustrée. Les généralisations obtenues permettent la tarification de l'aversion au risque de régime au moyen de fonctions de conversion. Des procédures simples de tarification pour les titres contingents sous ces nouvelles mesures neutres au risque, fondées soit sur la programmation dynamique, soit sur des simulations Monte-Carlo, sont également fournies.