Thèse

présentée pour obtenir le grade de docteur de l'Université de Strasbourg
Spécialité MATHÉMATIQUES APPLIQUÉES

Djibril Gueye

Some Contributions to Financial Risk Management

Soutenue le 09 Juillet 2021
devant la commission d'examen

Etienne Chevalier, examinateur
Areski Cousin, directeur
Abdou Ka Diongue, rapporteur
Armelle Guillou, examinatrice
Monique Jeanblanc, examinatrice
Ying Jiao, rapportrice

IRMA
Institut de Recherche
Mathématique Avancée

cnrs

Université
de Strasbourg

Université de Strasbourg

Institut de Recherche Mathématique Avancée (IRMA)

Some contributions to Financial Risk Management

Auteur:

Djibril Gueye

Composition du jury:

Etienne Chevalier
Areski Cousin
Abdou Ka Diongue
Armelle Guillou
Monique Jeanblanc
Ying Jiao

Thèse soutenue le 09 Juillet 2021

Remerciements

Je commence par exprimer ma gratitude envers mon directeur de thèse Areski Cousin pour m'avoir soutenu pendant toutes ces trois années et demie sans ménager aucun effort pour la réussite de ce travail. Je te remercie pour ta disponibilité et pour ta gentillesse. Merci encore une fois pour avoir contribué à ma future carrière professionnelle. Je t'en suis très reconnaissant.

Je voudrais exprimer ma profonde reconnaissance à une personne éminente, Monique Jeanblanc, qui m'a tant soutenu sur tous les plans. Toi qui m'as fait aimer la théorie de grossissement de filtrations ainsi que celle des processus stochastiques. J'ai tant appris de toi. Tu m'as tellement apporté. Pour ton soutien, ton attention envers ma personne, pour ta compréhension ; merci pour tout Monique!

J'aimerais ensuite remercier Ying Jiao et Abdou Ka Diongue pour avoir accepté de rapporter cette thèse en consacrant leur temps et leurs énergies à la lecture attentive de ce manuscrit. Je tiens également à exprimer toute ma reconnaissance à Armelle Guillou et Etienne Chevalier pour avoir accepté de faire partie de mon jury de thèse. Je les remercie également pour avoir été membres de mon premier comité de suivi ainsi que Ying Jiao qui a été membre de mon second comité de suivi.

J'adresse également un grand merci à toute l'équipe de calcul stochastique de l'IRMA. Tout particulièrement à Xaolin pour ces suggestions et remarques dans ce manuscrit, à Jean Berard, Nicolas Juillet et Vincent Vignon pour leurs conseils et surtout leur ouverture. Merci aussi aux autres membres de l'IRMA que j'ai côtoyés : Samuel Maistre, Michaël Gutnic, Laurent Gardes, Souhil Chakar, Sylvain P. Blanc,...
Je remercie aussi l'ensemble des enseignants de l'UFR de Sciences Appliquées et Technologies de l'Université Gaston Berger.

Mes remerciements vont également à l'endroit des doctorants et anciens doctorants de l'IRMA que j'ai rencontrés et avec qui j'ai échangé : Frédéric Valet, Claire, Camille, Guillaume, Nico, Phillipe Meyer, Lorenzo, Martin, Xavier, Phillipe Ricka, Lukas, Luca, Laura, Alexander, Marie Chion, Archia, Victoria, Antoine, Basile, Raoul, Clarence,...
Un grand merci à mes co-bureaux, Pierre-Alexandre, Yohann, Francisco et Valdo pour ces années de partage et d'entraide.

Merci à tous mes amis et frères africains résidents en France : Elhadji Ngom, Amady, Hamady, Youssoupha, Malick, Momo Sylla, Mamadou Sène, Ismaël, Arlindo, Momo Sambou, Mor, Prince, Omar Kata,...

Contents

List of Figures

Á mes parents.

Á mon fils Ousmane Gueye et sa maman Chantal.

Introduction détaillée en français

Cette thèse s'articule autour de deux parties différentes qui contribuent à la gestion quantitative des risques financiers : la modélisation du risque de crédit et la quantification de l'incertitude liée à l'estimation des risques financiers. La première partie traite de la modélisation des temps de défaut en risque de crédit, tandis que la seconde porte sur la construction de surfaces de volatilité à l'aide du krigeage.

Partie 1: Modèlisation du temps de défaut

Dans la littérature du risque de crédit, la modélisation du temps de défaut repose essentiellement sur deux approches fondamentales, à savoir l'approche structurelle et l'approche à forme réduite. La principale différence entre ces deux approches est le choix de la filtration. Dans l'approche structurelle (voir Merton, 1974), l'information est modélisée par une filtration $\mathbb{F}$ pour laquelle le temps de défaut est un temps d'arrêt qui peut être prévisible ou totalement inaccessible (dans ce cas, il est nécessaire de calculer son compensateur). Tandis que, dans l'approche à forme réduite (voir Duffie, Schroder, and Skiadas, 1996; Lando, 1995), on commence par une filtration de référence $\mathbb{F}$ dans laquelle le temps de défaut τ n'est pas un temps d'arrêt et on la grossit progressivement avec τ pour obtenir une filtration $\mathbb{G}$. La théorie de grossissement de filtration remonte aux années 70-80 suite aux travaux d'Itô (1978), Barlow (1978), Jeulin and Yor (1978), Yor (1978), Jacod (1985) et s'applique dans plusieurs domaines notamment en finance et en assurance.

Très souvent, la construction des modèles de défaut dans l'approche à forme réduite conduit à des temps de défauts qui évitent les temps d'arrêt de la filtration de référence $\mathbb{F}$. Ce qui fait que ces modèles deviennent inadaptés quand il s'agit de modéliser certains produits financiers soumis au défaut en présence de chocs économiques, lorsque le temps de défaut τ peut être égal, avec une probabilité strictement positive, à l'un des instants d'événement de chocs qui correspondent à des temps d'arrêt de la filtration de référence. Ceci conduit certains auteurs à s'intéresser récemment au cas où cette hypothèse est omise, en utilisant des modèles dits hybrides. Par exemple, Gehmlich and Schmidt (2018) et Fontana and Schmidt (2018) n'utilisent pas le grossissement de filtration et considèrent un modèle où le temps de défaut est un temps d'arrêt par rapport à une filtration $\mathbb{G}$ qui n'évite pas les $\mathbb{G}$-temps d'arrêt prévisibles. Dans leur travail, ils supposent l'existence du compensateur du défaut qui peut être décomposé en une partie absolument continue et un saut pur (prévisible). Mais ils ne font aucune construction du temps d'arrêt associé.
Dans cet ordre d'idées, Jiao and Li (2018) ont élaboré un modèle selon lequel le temps de défaut peut coïncider avec des instants de chocs économiques prévisibles. Ils s'intéressent au compensateur du temps de défaut en utilisant l'approche de densité généralisée et montrent que le processus de taux d'intensité n'existe pas dans leur cadre. En outre, leur modèle permet de capter les sauts de prix des obligations zéro-coupon soumises au défaut ainsi que ceux des spreads correspondants.

Un des objectifs de cette thèse est d'étendre le modèle de Jiao and Li (2018) dans le cas où les instants de chocs ne sont pas prévisibles. Malgré son attractivité dans la modélisation des produits financiers soumis au défaut en présence de chocs économiques prévisibles, le modèle de Jiao and Li (2018) est difficile à mettre en œuvre lorsque les chocs ne sont pas prévisibles (i.e., ces chocs apparaissent par surprise) et ne permettent pas toujours de capter les sauts de prix des obligations zéro-coupon.

Pour modéliser les produits financiers soumis au défaut en présence de chocs économiques qui ne sont pas prévisibles, nous proposons un modèle de Cox généralisé qui étend celui de Lando (1998) dans lequel le temps de défaut τ est le premier instant où un processus K croissant adapté à une filtration donnée (la filtration de référence) $\mathbb{F}$, absolument continue par rapport à la mesure de Lebesgue passe par une barrière stochastique Θ indépendante de $\mathbb{F}$. Il s'ensuit que ce temps de défaut évite tous les temps d'arrêt dans la filtration de référence $\mathbb{F}$. Nous assouplissons l'hypothèse selon laquelle le processus croissant K est absolument continu. Ici, nous travaillons dans un cadre plus général, ce processus K étant adapté, croissant et continu à droite avec des limites à gauche ou continu à gauche avec des limites à droite. Cela nous conduit à un temps aléatoire qui n'évite pas les temps d'arrêt de $\mathbb{F}$. Dans les deux cas, nous nous intéressons aux caractéristiques du temps de défaut comme la martingale de survie conditionnelle, la supermartingale Azéma, le compensateur du temps de défaut ainsi que son processus de réduction prévisible (présenté dans la Section 1.2) et la décomposition multiplicative de la supermartingale d'Azéma. Nous étudions également l'existence de densités conditionnelles dans le sens de Jacod (1985) que nous dénotons (CL), dans le sens de Jiao and Li (2015), dénoté par (GD) et l'hypothèse étendue de Jacod introduite dans Li and Rutkowski (2014).

On peut trouver une première tentative de généralisation dans Bélanger, Shreve, and Wong (2004), où le processus croissant est prévisible et continu à droite.

Contrairement au modèle de Jiao and Li (2018), dans le modèle généralisé de Cox, le processus de taux d'intensité peut exister.

Structure des chapitres de la partie 1

Cette première partie de la thèse est structurée comme suit.

Dans le chapitre 1, nous commençons par rappeler des faits bien connus sur la théorie des processus stochastiques et modèles de temps de défaut que nous utiliserons tout au long de cette partie, avant de traiter l'exemple du cas gaussien en densité conditionnelle. Nous définissons toutes les caractéristiques du temps de défaut, particulièrement la réduction prévisible de son compensateur. Nous mettons l'accent ensuite sur l'hypothèse d'existence de densité conditionnelle de Jacod que l'on notera (CL) en étudiant l'exemple explicite de densité gaussienne de El Karoui et al. (2014). Ensuite nous montrons, en utilisant divers techniques de calcul stochastique, que cet exemple de densité conditionnelle gaussienne est unique dans une large classe.

Dans le chapitre 2, nous généralisons le modèle de Jiao and Li (2018) qui est développé dans un contexte où le temps de défaut peut coïncider avec certains temps d'arrêt prévisibles $(\tau_i)_i$ de $\mathbb{F}$ qui sont des temps de chocs exogènes prévisibles où ces chocs sont modélisés par un processus de Poisson N non homogène. Par conséquant l'hypothèse de Jacod n'est plus satisfaite. Notre généralisation consiste à relâcher l'hypothèse que les temps de chocs sont prévisibles et à remplacer le processus de Poisson par un processus plus général noté X. Nous étudions les caractéristiques du temps de défaut.

Dans ce cadre, nous considérons $(\Omega, \mathcal{G}, \mathbb{F}, \mathbb{P})$ un espace probabilisé filtré. Nous notons par Ψ la fonction croissante vérifiant $P(X_u \leq 1) = e^{-\Psi(u)}$ (avec $\Psi(0) = 0$ et $\Psi(\infty) = \infty$) et par $\mathbb{F}^X$ la filtration naturelle de X. Soit $(\tau_i)_i$ une suite strictement croissante de temps d'arrêt de $\mathbb{F}$ telle que $\lim_{i \to \infty} \tau_i = \infty$ et $\tau_0 = 0$. Le temps de défaut τ est défini comme suit :

$$\tau = \theta \wedge \xi$$

où θ est donné par

$$\theta = \tau_i \quad \text{on} \quad \{X_{\tau_{i-1}} \leq 1 < X_{\tau_i}\}, \text{ for } i \geq 1$$

et

$$\xi := \inf\{t \geq 0 : \Gamma_t \geq \Theta\}$$

avec Γ un processus croissant $\mathbb{F}$-adapté (tel que $\Gamma_0 = 0$, $\Gamma_t < \infty$ pour tout t, et $\Gamma_\infty = \infty$) et Θ une variable aléatoire (supposée être de loi exponentielle de paramètre 1) indépendante de $\mathbb{F}$ et de $\mathbb{F}^X$.

Dans ce modèle, le temps de défaut τ n'évite pas les temps d'arrêt de la filtration de référence $\mathbb{F}$ et pour tout $t \geq 0$, on obtient

$$\mathbb{P}(\tau = \tau_i | \mathcal{F}_t) = \mathbb{E}\left[e^{-\Gamma_{\tau_i}} \left(e^{-\Psi(\tau_{i-1})} - e^{-\Psi(\tau_i)} \right) | \mathcal{F}_t \right], \ \forall i \geq 1.$$

En particulier,

$$\mathbb{1}_{\{\tau_i \leq t\}} \mathbb{P}(\tau = \tau_i | \mathcal{F}_t) = e^{-\Gamma_{\tau_i}} \left(e^{-\Psi(\tau_{i-1})} - e^{-\Psi(\tau_i)} \right) \mathbb{1}_{\{\tau_i \leq t\}}, \ \forall i \geq 1.$$

En notant par $\mathbb{G}$ le grossissement progressif de $\mathbb{F}$ par τ, on obtient la propriété d'immersion entre $\mathbb{F}$ et $\mathbb{G}$. Ce qui implique que la supermartingale d'Azéma Z est décroissante.
Comme dans Jiao and Li (2018), nous calculons également la loi de survie conditionelle de τ qui est donnée pour tout $t \in \mathbb{R}^+$ par

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = \mathbb{E}\left[\exp\left(-\sum_{i=1}^\infty \mathbb{1}_{\{\tau_i \leq u\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})] - \Gamma_u \right) | \mathcal{F}_t \right] \text{ for } u \in \mathbb{R}^+.$$

En particulier, on a pour $t \geq u$,

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = \exp\left(-\sum_{i=1}^\infty \mathbb{1}_{\{\tau_i \leq u\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})] - \Gamma_u \right).$$

Ainsi la supermatingale d'Azéma Z s'exprime comme

$$Z_t = \exp\left(-\sum_{i=1}^\infty \mathbb{1}_{\{\tau_i \leq t\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})] - \Gamma_t \right).$$

Dans le cas où les temps d'arrêt $(\tau_i)_i$ de $\mathbb{F}$ sont prévisibles, la réduction $\mathbb{F}$-prévisible du compensateur Λ de τ est

$$\Lambda_t = \Gamma_t + \sum_{i=1}^\infty \mathbb{1}_{\{\tau_i \leq t\}} (1 - \exp(-[\Psi(\tau_i) - \Psi(\tau_{i-1})])), \text{ for all } t \geq 0$$

ce qui montre, comme dans Jiao and Li (2018), que le taux d'intensité du temps de défaut n'existe pas. Cependant, si certains des temps d'arrêt $(\tau_i)_i$ de $\mathbb{F}$ ne sont pas prévisibles, on suppose connaître pour tout i, le $\mathbb{F}$-compensateur de τ_i, noté J^i. Ainsi, la réduction $\mathbb{F}$-prévisible du compensateur de τ vérifie

$$d\Lambda_t = \frac{1}{Z_{t-}} dA_t^p = d\Gamma_t + \frac{d\zeta_t}{Q_{t-}}$$

avec Q la supermartingale

$$Q_t := \exp\left(-\sum_{i=1}^\infty \mathbb{1}_{\{\tau_i \leq t\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})] \right)$$

et ζ le processus croissant $\mathbb{F}$-prévisible défini par

$$\zeta_t := \sum_{i=0}^\infty \int_0^t \left(e^{-\Psi(\tau_i)} - e^{-\Psi(s)} \right) dJ_s^{i+1}.$$

Nous montrons que, dans le cas particulier où les temps de chocs sont les temps de sauts d'un processus de Poisson homogène N (d'ntensité λ) engendrant la filtration $\mathbb{F}$, la réduction prévisible du compensateur Λ est continue et que le processus de taux d'intensité existe et est égale, pour tout $t \geq 0$ à

$$\gamma_t + \frac{\lambda \sum_{i=0}^{\infty} \mathbb{1}_{[\tau_i, \tau_{i+1}]}(t)\big(e^{-\Psi(\tau_i)} - e^{-\Psi(t)}\big)}{Q_{t-}}.$$

Dans le chapitre 3, nous incorporons le modèle étendu de Jiao et Li étudié dans le chapitre 2 dans un cadre beaucoup plus général de modèle de Cox généralisé. Nous travaillons toujours dans l'espace $(\Omega, \mathcal{G}, \mathbb{F}, \mathbb{P}, \mathbb{G})$ où $\mathbb{G}$ est le grossissement progressif de $\mathbb{F}$ par τ. Le modèle de Cox (voir Lando, 1998) est le modèle le plus largement utilisé au marché financier où le temps de défaut (supposé être fini) est modélisé comme suit :

$$\tau = \inf\{t \geq 0 : K_t \geq \Theta\}$$

avec le processus K qui est supposé être continu croissant $\mathbb{F}$-adapté tel que $K_0 = 0$ et Θ une v.a indépendante de $\mathbb{F}$ (de loi exponentielle de paramètre 1).

Dans ce modèle, le temps de défaut τ évite les temps d'arrêt de $\mathbb{F}$ et les hypothèses d'intensité et de Jacod sont satisfaites. Nous proposons d'étendre ce modèle afin que le temps de défaut τ puisse coïncider avec des temps d'arrêt de $\mathbb{F}$. Notre généralisation consiste à relâcher l'hypothèse de continuité sur le processus K. Nous supposons que le processus K, croissant $\mathbb{F}$-adapté, est continue à droite avec des limites à gauche (ou continu à gauche avec des limites à doite).

Dans un tel cadre, les temps de sauts de K (notés $(\tau_i)_i$) sont les $\mathbb{F}$-temps d'arrêt que τ n'évite pas et on a

$$\mathbb{P}(\tau = \tau_i | \mathcal{F}_t) = \mathbb{E}[e^{-K_{\tau_i-}}(1 - e^{-\Delta K_{\tau_i}})|\mathcal{F}_t].$$

L'étude de la réduction prévisible du compensateur de τ nécessite le résultat suivant (qui calcule la projection duale $\mathbb{F}$-prévisible de τ).

Lemme : Soit I la $\mathbb{F}$-sous-martingale càdlàg $I_t = \sum_{s \leq t}(1 - e^{-\Delta K_s})$ et A^I la partie prévisible de la décomposition de Doob-Meyer de I. Alors, la projection duale $\mathbb{F}$-pévisible de τ vérifie

$$dA_t^p = e^{-K_{t-}}(dK_t^c + dA_t^I),$$

où K^c est la partie continue de K.
Par conséquent, la réduction prévisible du compensateur de τ est $\Lambda = K^c + A^I$.

Nous étudions également toutes les autres caractéristiques du temps de défaut, particulièrement la décomposition multiplicative de la super-martingale d'Azéma Z qui est donnée par :

$$\mathcal{E}(Y)_t \mathcal{E}(-\Lambda)_t = \mathcal{E}(Y)_t e^{-\Lambda_t} \prod_{s \leq t}(1 - \Delta A_s^I)e^{\Delta A_s^I}$$

où $\mathcal{E}(-\Lambda)$ est la partie $\mathbb{F}$-prévisible décroissante et $\mathcal{E}(Y)$ la partie $\mathbb{F}$-locale martingale, avec $Y_t = \int_0^t \frac{1}{Z_{s-}(1 - \Delta A_s^I)} dM_s$, où M étant la partie martingale de la décomposition de Doob-Meyer de Z et Λ la réduction $\mathbb{F}$-prévisible du compensateur de τ.
Ce résultat est établi en toute généralité dans Jeanblanc and Li (2020), où les auteurs "devinent" cette forme et la vérifient. Cependant, il semble intéressant de le retrouver sans deviner cette décomposition, et nous présentons une preuve dans ce manuscrit.

Pour illustrer notre construction, nous donnons de nombreux exemples. Pour cela, Nous nous intéresserons à des processus croissant càdlàg partant de zéro qui nous permettent de construire le temps de défaut τ et ensuite d'expliciter toutes ses caractéristiques. Un exemple est la construction suivante où K est un shot noise.

Exemple du shot-noise :

Soit $\mathbb{F}$ une filtration, $(\tau_i)_i$ une suite croissante de $\mathbb{F}$-temps d'arrêt, et $(\gamma_i)_i$ une suite de variables aléatoires avec $\gamma_i \in \mathcal{F}_{\tau_i}$. Soit μ la mesure aléatoire du processus marqué $(\tau_i, \gamma_i)_i$ et ν son compensateur. On pose $\tilde{\mu} = \mu - \nu$. On définit le bruit quantique suivant (Shot-noise en anglais) :

$$K_t = \sum_{i=0}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}} G(t - \tau_i, \gamma_i) = \int_0^t \int_{\mathbb{R}} G(t-s, x) \mu(ds, dx)$$

où la fonction $G : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^+$ vérifie

$$G(t, x) = G(0, x) + \int_0^t g(s, x) ds \quad \forall\, t \geq 0,\ x \in \mathbb{R}$$

avec g une fonction borélienne positive de $\mathbb{R}^+ \times \mathbb{R}$, telle que G est croissante dans la direction du temps. Dans cet exemple, nous montrons d'abord que K est une $\mathbb{F}$-sous-martingale et sa décomposition de Doob-Meyer est $K_t = A_t^K + M_t^K$ avec A^K sa partie prévisible (et croissante) donnée par

$$A_t^K = \int_{u=0}^t \left(\int_{s=0}^u \int_{\mathbb{R}} g(u-s, x) \mu(ds, dx) \right) du + \int_{s=0}^t \int_{\mathbb{R}} G(0, x) \nu(ds, dx)$$

et M^K sa partie martingale qui a pour expression

$$M_t^K = \int_0^t \int_{\mathbb{R}} G(0, x) \tilde{\mu}(ds, dx).$$

Le processus de réduction prévisible du compensateur de τ est :

$$\Lambda_t = \int_{u=0}^t \left(\int_{s=0}^u \int_{\mathbb{R}} g(u-s, x) e^{-G(u-s, x)} \mu(ds, dx) \right) du + \int_0^t \int_{\mathbb{R}} (1 - e^{-G(0, x)}) \nu(ds, dx) \,.$$

Si ν est déterministe, alors pour tout $u \geq t$, la martingale de survie conditionnelle de τ s'exprime

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = c(u) L_t(u)$$

où $c(u)$ est une constante explicite et $L(u)$ la martingale

$$L_t(u) = \exp\left(-\int_0^t \int_{\mathbb{R}} G(u-s,\ x) \mu(ds, dx) - \int_0^t \int_{\mathbb{R}} (e^{-G(u-s,\ x)} - 1) \nu(ds, dx) \right).$$

Dans le cas où le processus croissant K est continu à gauche (càglàd), nous modifions un peu la définition de τ dans le but d'obtenir un temps aléatoire avec de jolies propriétés. Ainsi, τ est défini comme suit :

$$\tau := \inf\{t \,:\, K_t > \Theta\}.$$

Cette définition nous permet d'être capable de caractériser l'ensemble $\{K_t > \Theta\}$ et d'en déduire par la suite que $\{K_t \leq \Theta\} = \{\tau \geq t\}$, ce qui implique que $\widetilde{Z}_t = e^{-K_t}$ et donc $Z = e^{-K_+}$. Par conséquent, ce cadre peut être réduit au cas précédent sur la continuité à droite en utilisant le processus κ, vérifiant $\kappa_t = K_t+$ pour tout $t \geq 0$, à la place de K.

Dans le chapitre 4, nous étudions certaines applications du modèle généralisé de Cox dans le pricing des produits soumis au défaut. Nous donnons l'expression des prix des obligations zéro-coupon ainsi que leurs

dynamiques à travers un cadre général. Nous nous intéresserons également aux impacts des sauts de K sur les prix avant défaut des obligations zéro-coupon soumis au défaut.

Partie 2: Le krigeage pour la construction de surfaces de volatilité

Les surfaces de volatilité constituent des outils importants en gestion des risques. Elles constituent des composantes élémentaires pour la tarification et la couverture des produits dérivés financiers. Elles permettent par exemple d'estimer, à partir du prix d'options liquides, la valeur des produits financiers dont les caractéristiques sont non standards et dont le prix n'est pas observé sur le marché. La volatilité dépend des caractéristiques contractuelles de l'option (nature de l'actif sous-jacent, maturité, prix d'exercice) et n'est observée que pour un nombre limité de caractéristiques standard (couple maturité - prix d'exercice) et peut nécessiter la surface entière de prix ou le prix de certaines caractéristiques non observées pour certaines applications. En outre, la construction de ces surfaces est habituellement faite en respectant le principe d'absence d'opportunité d'arbitrage, i.e., les prix des options obtenus à partir de ces surfaces ne permettent pas de réaliser d'arbitrage financier. La construction d'une surface de prix sans arbitrage peut être transformée en un problème d'apprentissage fonctionnel avec des contraintes de forme et éventuellement des observations brutées.

Une fois la surface de prix interpolée, on peut utiliser la formule d'inversion de Black et Scholes (voir Black and Scholes, 1973) pour obtenir la surface de volatilité implicite. Pour obtenir la surface de volatilité locale, nous pouvons utiliser une approximation de différence finie de la formule de Dupire (voir Dupire, 1994). En outre, il est possible de construire la surface de volatilité locale en utilisant les volatilités implicites Il s'agit d'extraire la surface de volatilité locale de la volatilité implicite en utilisant la formule (1.10) de Gatheral (2011). Ces approches qui utilisent les prix observés (ou la volatilité implicite pour construire la volatilité locale) sont connues sous le nom d'approches indirectes.

Au-delà de ces approches directes, il existe certaines qui paramétrent directement la volatilité non observée à travers une fonction dans laquelle ses paramètres peuvent être estimés en minimisant une fonction objective appropriée obtenue pour calibrer cette fonction aux données de marché. Il s'agit notamment de la méthode SVI qui paramétrent le smile de volatilité implicite et son extension (SSVI) pour la surface de volatilité implicite (voir Gatheral and Jacquier, 2014; Gatheral, 2004.

Les techniques utilisées habituellement dans les approches indirectes reposent sur des splines contraintes (voir par exemple Fengler, 2009; Homescu, 2011; Laurini, 2011; Wang, Yin, and Qi, 2004, etc.).

Récemment, des techniques de réseaux de neuronnes ont été adoptées pour améliorer ces modèles habituels en enregistrant de meilleures performances de surface de prix sans arbitrage. Le travail de Dugas et al. (2009) est le point de départ de l'intégration de contraintes sans arbitrage dans les réseaux de neurones pour l'apprentissage des prix des options. Dugas et al. (2009) ont utilisé les contraintes dites hard à travers une architecture spéciale. Cette approche est différente de celle appelée approche par contraintes soft qui consiste à pénaliser certaines variables, dans la fonction objective, qui ne vérifient pas les contraintes (voir, par exemple, Itkin, 2019). La principale difficulté de l'approche de Dugas et al. (2009) s'appuie sur le fait, qu'en pratique, l'incorporation de contraintes hard dans les réseaux neuronaux réduit leur flexibilité et présente une transformation hautement non linéaire pour le calcul de prix des options à chaque itération. Selon Ackerer, Tagasovska, and Vatter (2019), les contraintes hard de Dugas sur les prix font perdre trop de précisions. Ainsi, pour la construction de la surface de volatilité locale, ces derniers ont proposé d'utiliser directement la surface de volatilité implicite au lieu de la surface de prix en appliquant l'approche soft. Chataigner, Crépey, and Dixon (2020) ont proposé des architectures de réseaux de neurones simples pour les prix des options sans arbitrage et pour la volatilité implicite. Ils montrent que les contraintes hard réduisent la puissance du réseau et que les contraintes soft fournissent les meilleurs prix précis et les volatilités implicites.

L'objectif de cette partie est d'adapter des techniques de krigeage contrainte développées dans Maatouk and Bay (2017) afin de quantifier l'incertitude associée à la construction de surfaces de volatilité (telles que la volatilité implicite et la surface de volatilité locale) dans un cadre qui respecte les conditions d'absence d'opportunités d'arbitrage. Nous utilisons les approches indirectes qui consistent à apprendre les prix par le krigeage et ensuite utiliser la formule d'inversion pour obtenir la surface de volatilité implicite et en déduire la surface de volatilité locale. Il n'est pas nécessaire d'étendre l'approche de Maatouk and Bay (2017) pour des contraintes d'inégalité plus générales, comme cela a été fait dans López-Lopera et al. (2018), puisque nous nous limitons aux contraintes de convexité et de monotonie.

L'idée est de supposer que la fonction prix est une réalisation d'un champ gaussien tronqué, vérifiant certaines contraintes. Comme pour les courbes de taux (voir Cousin, Maatouk, and Rullière, 2016), les contraintes de compatibilité avec les observations de marché (compatibilité aux prix des options européennes) se traduisent sous forme d'une relation linéaire sur le processus Gaussien. L'hypothèse d'absence d'arbitrages est équivalente à l'existence d'une mesure martingale : les réalisations du processus de krigeage doivent correspondre à des lois marginales de martingales. D'après le Th. 2.6 de Beiglböck, Juillet, et al. (2016), il s'agit alors de construire une famille de distributions marginales à partir d'un processus gaussien bivarié, indéxé par la maturité et le prix d'exercice, qui est :

- croissante par rapport à la direction des maturités,

- convexe par rapport à la direction des strikes,

- solution d'une relation linéaire sur certaines maturités et strikes.

Le krigeage (ou la regression par processus de Gauss) est une technique de géostatistique initialement introduite pour estimer la densité de minerais dans une zone pré-définie de l'espace, étant donné quelques observations issues d'expériences de forage à des points particuliers du sol. Le principe du krigeage repose sur la détermination de la distribution conditionnelle d'un champ aléatoire gaussien connaissant les valeurs de ce champ à certains points de l'espace, appelés points d'expérience (voir par exemple, Matheron, 1963; Cressie, 1990; Krige and Magri, 1982). Le principal intérêt de cette approche est qu'elle permet d'estimer les valeurs de la variable de référence à d'autres points de l'espace en quantifiant l'incertitude associée à cette estimation. Par ailleurs, dans un contexte où les observations sont coûteuses à obtenir, le plan d'expérience (points du support où la variable d'intérêt est observée) peut être conçu de manière adaptative (voir Williams and Rasmussen, 2006). Cette technique a rapidement gagné en popularité suite aux travaux de Williams and Rasmussen (2006) et a été adaptée dans divers domaines tels que l'hydrologie, la météorologie ou l'épidémiologie, ect. Quelques travaux ont été récemment développés en sciences actuarielles.

Par exemple, Sousa, Esquível, and Gaspar (2012) ont utilisé le krigeage pour estimer les paramètres du modèle d'apprentissage lors de l'étalonnage du modèle de taux d'intérêt Vasicek sous la mesure risque neutre. Ils ont considéré un prior gaussien sur les prix des obligations zéro-coupon et ont estimé les paramètres du modèle en maximisant la log de vraissemblance des données d'entraînement compte tenu des paramètres. Ludkovski (2018) pour améliorer la méthode des moindres carrés Monte-Carlo en valorisation d'option de type américain, a utilisé le krigeage pour améliorer l'étape de la régression qui consiste à faire rapprocher la valeur espérée de celle de continuation.

De Spiegeleer et al. (2018) ont montré la rapidité du krigeage par rapport à la méthode de monte-carlo pour l'approximation des prix d'options et de la volatilité implicite. Toutefois, ils n'ont pas tenu en compte les conditions d'arbitrage dans leur approche. En modélisation du portefeuille de produits, Dixon and Crépey (2018) ont considéré une technique de krigeage avec multi-réponses pour calculer la Credit Valuation Adjustment (CVA). Ils ont montré que le krigeage permet une approximation rapide de la valeur du portefeuille.

Gonzalvez et al. (2019) ont exploré l'utilisation du krigeage en finance en montrant que ce dernier est un outil puissant pour ajuster la courbe de rendement des prix des actions. Cependant leur interpolation ne prend pas également en considération les conditions de non arbitrage du marché.

Une première tentative d'utilisation du krigeage pour la volatilité locale est effectuée par Tegnér and Roberts (2019) qui placent un prior gaussien directement sur la surface de volatilité locale (pour garantir la positivité de la volatilité locale, ils attribuent une fonction positive sur le prior). Une telle approche conduit à une fonction de perte des moindres carrés non linéaires, car elle implique la transformation non linéaire de la volatilité locale en prix d'option vanille correspondants. Une telle fonction de perte n'est pas notoirement favorable à la descente de gradient (stochastique ou non), de sorte que les auteurs recourent à une optimisation MCMC. De plus, la performance de leur GPs est mesuré par RMSE en échantillon et ne semble pas être évalué pour le surapprentissage. En outre, ils ne comparent pas leur approche à d'autres approches alternatives.

En fait, ajouter une telle contrainte de positivité ainsi que d'autres contraintes d'inégalité linéaire (telles que la monotonie, la convexité, les bornes) dans le krigeage n'est pas une tâche facile et constitue un grand défi. À cette fin, certains auteurs se sont récemment penchés sur cette question, connue sous le nom de krigeage contraint ou de régression par processus gaussien (GPs) contraints. Les cadres présentés dans Da Veiga and Marrel (2012), Golchi et al. (2015), Riihimäki and Vehtari (2010), and Wang and Berger (2016), entre autres, utilisent une subdivision de l'espace d'entrée. Cependant, ils se basent sur le fait que les contraintes ne sont satisfaites que sur des observations virtuelles et ne garantissent pas les contraintes dans l'ensemble du domaine. Par exemple, pour intégrer la contrainte de monotonie croissante dans le krigeage, Wang and Berger (2016) ont proposé de sélectionner les localisations avec de fortes probabilités d'avoir des dérivées négatives comme les localisations virtuelles dans lesquelles ils forcent les contraintes. L'approche d'Agrell (2019) est une extension de celle de Wang and Berger (2016) pour des contraintes multiples grâce à une méthode efficace de simulation du processus à postériori basé sur la dérivation de processus gaussien contraint à l'aide d'un opérateur linéaire.

Lorsque des contraintes d'inégalité sont ajoutées sur les quantités d'intérêt, le processus à postériori n'est plus gaussien et les contraintes sont habituellement infini-dimensionnelles. C'est ce qui a motivé Maatouk and Bay (2014) à utiliser l'approximation fini-dimensionnelle des processus gaussiens pour lesquels les contraintes d'inégalité sont faciles à vérifier. Ils ont développé un schéma appelé rejection autour du mode (RSM) pour la simulation de la distribution gaussienne multivariée tronquée nécessaire à l'estimation du processus a postériori. Leur approche garantit les contraintes dans le domaine entier. Cousin, Maatouk, and Rullière (2016) utilisent les techniques de krigeage contraint mises au point dans Maatouk and Bay (2014) pour construire des courbes de taux et des courbes CDS en respectant les conditions de non arbitrage et de contrôler des erreurs. López-Lopera et al. (2018) ont étendu le cadre de Maatouk and Bay (2014) pour des contraintes d'inégalité plus générales et ont proposé d'utiliser la méthode de Monte Carlo Hamiltonienne (HMC) de Pakman and Paninski (2014) qui est plus efficace que le rejet autour du mode de Maatouk and Bay (2014) pour la simulation du processus a posteriori.

Contrairement à l'interpolation par splines, et aux approches SSVI et réseau de neurones, le krigeage est une technique de regression permettant de quantifier l'incertitude dans l'estimation des variables d'intérêt. Il est possible par exemple d'obtenir des intervalles de confiance aux points de la surface où les observations sont indisponibles ou considérées comme peu fiables.

La principale contribution de cette partie de la thèse est de montrer que le krigeage est un outil approprié pour construire des prix d'option et pour quantifier l'incertitude dans l'estimation de la volatilité locale et implicite.

Structure des chapitres de la partie 2

Cette deuxième partie de la thèse est structurée en trois chapitres.

Le chapitre 5 est un chapitre introductif dans lequel on présente des outils du krigeage qui vont être utilisés dans les chapitres 6 et 7. Une attention particulière sera portée sur la simulation des coefficients gaussiens lorsque les contraintes de monotonie sont saturées et nous proposons une solution numérique à ce problème.

Le chapitre 6 porte sur la construction de la surface de volatilité implicite à l'aide du krigeage. Nous adoptons le krigeage contraint développé dans Maatouk and Bay (2014) pour construire des prix sans arbitrage. Nous comparons les performances du krigeage classique et celles du krigeage contraint dans ce contexte à l'aide d'une étude empirique utilisant les données Euro Stoxx 50 du 10 janvier 2019. L'approche par krigeage contraint fournit les meilleures performances que celle par krigeage classique. Par contre, Les mesures de précision des estimateurs des MAP montrent qu'à un certain pourcentage des données utilisées, le MAP estimé à partir du krigeage classique fonctionne aussi bien que celui estimé à partir du krigeage contraint.

Le chapitre 7 est consacré à la construction de la surface de volatilité locale en utilisant le krigeage. Tout d'abord, nous étendons la construction de surface de prix d'option sans arbitrage dans le cas des dividendes et des taux d'intérêt. Avec une telle surface de prix, nous modélisons la surface de volatilité locale en utilisant l'équation aux dérivées partielles de Dupire. La quantification de l'incertitude fournit des intervalles de confiance dans les modèles de volatilité locale. Nous comparons les performances de notre modèle à celles des approches SSVI et réseaux de neurones. Nous utilisons un réseaux de neurones avec des contraintes dites "soft" sur la formule de variance implicite pour dériver la surface de volatilité locale à partir des volatilités implicites de Black-Scholes (voir la flèche bleue de la Fig. 7.1).
L'ajustement du réseau de neurones est fait sur les donnés mid tandis que celui du krigeage s'est fait sur une réplication des données bid et ask. Cette propriété du krigeage permet d'obtenir une surface de prix qui respecte les contraintes des écarts de bid-ask. Autrement dit, la surface de prix obtenu se trouve entièrement entre les prix de bid et les prix ask.
Le krigeage offre de meilleures performances pour le pricing des oprions et est la seule approche qui permet de quantifier l'incertitude, ce qui est utile pour l'analyse des risques.

La lecture de ce manuscrit peut se faire par le guide du diagramme suivant :
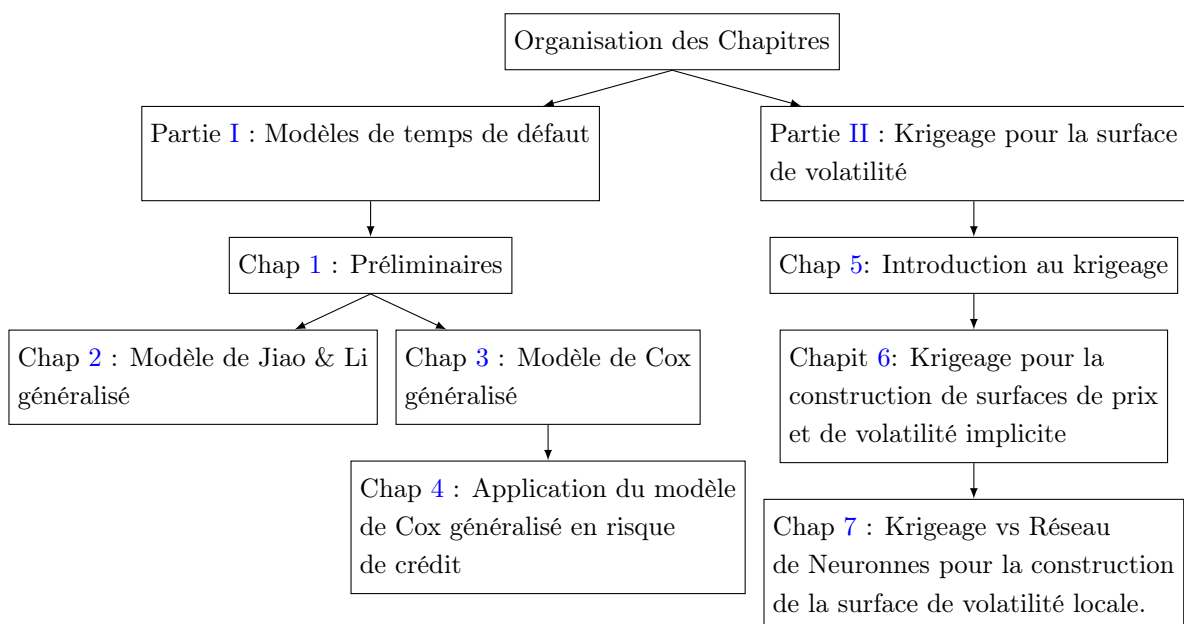
Introduction in english

This thesis treats two different parts which contribute in the modelisation and uncertainty quantification of financial risk. The first part deals with default time modeling in credit risks while the second one focuses on volatility surfaces construction using kriging.

Part 1: Models of default times

This part addresses issues related to default time modeling in credit risk. The two fundamental approaches for default models are the structural approach and the reduced form approach. The main difference between these two approaches is the choice of the filtration. In the structural approach (see Merton, 1974), the information is modeled by a filtration $\mathbb{F}$ in which the default time τ is a stopping time which can be predictable or totally inaccessible (in this case, one has to compute its compensator). Whereas, in the reduced form approach (see Duffie, Schroder, and Skiadas, 1996; Lando, 1995), one starts with a reference filtration $\mathbb{F}$ in which the default time τ is not a stopping time and one enlarges progressively $\mathbb{F}$ with τ to obtain $\mathbb{G}$.

The enlargement of filtration theory has been in operation since the 70's following the works of Itô (1978), Barlow (1978), Jacod (1985), Jeulin and Yor (1978) and Yor (1978), it can be applied in several areas, particularly in finance and insurance.

Usually, the construction of τ is such that τ avoids the $\mathbb{F}$-stopping times. However, recently some authors were interested in the case where this hypothesis is violated (see, e.g., Jiao and Li, 2015; Jiao and Li, 2018). This fact can be considered, for instance, from the perspective of the need of modeling some defaultable claims under presence of economic shocks, in which the default time could coincide with one of the economic shocks times, which are stopping times.

In Gehmlich and Schmidt (2018), as well as in Fontana and Schmidt (2018), the authors do not make use of enlargement of filtration and consider a model where the $\mathbb{G}$-stopping time τ does not avoid $\mathbb{G}$-predictable stopping times. In their work, they assume the existence of the compensator of the default time, which is decomposed into an sum of absolutely continuous part and a (predictable) pure-jump part, but they did not construct of the associated stopping time.

Jiao and Li (2018) developed a model under which the default time could coincide with some predictable shock times. They are interested in the compensator of the default time, by using the generalized density approach, they show that the intensity rate process does not exist in their setting. Moreover, their model allows one to capture the jumps in the prices of zero-coupon bonds as well as in the corresponding spreads.

One of the objectives of this work is to extend the model of Jiao and Li in the case where the shock times are not predictable. Despite its attractiveness in modeling defaultable claims under some predictable shocks, the model of Jiao and Li is difficult to implement if the shocks are not predictable and does not always allow to capture the jumps of the zero-coupon bond prices. Our extention allows one to model such claims.

To model defaultable claims under some economic shocks which are not predictable, we propose a generalized Cox model which extend the one of Lando (1998) in which the default time is the first time when an increasing process adapted to a given filtration, absolutely continuous with respect to Lebesgue's measure hits a certain level, which is a random variable independent of the given filtration. It follows that this random time avoids all stopping times in the reference filtration. We relax the assumption that the increasing process that hits the threshold level is absolutely continuous. Here we are working in a more general case, this process being adapted, increasing and continuous on right with limits on left or continuous on left with limits on right. This leads us to a random time which does not avoid $\mathbb{F}$-stopping times. In both cases, we are interested in the characteristics of the default time such as the conditional survival martingale, the Azéma supermartingale and the compensator of the default time, as well as its predictable reduction process (presented in Section 1.2) and the multiplicative decomposition of the Azéma supermartingale. We also study the existence of conditional densities in the sense of Jacod (1985), in the sense of Jiao and Li (2015) and the extended Jacod's hypothesis introduced in Li and Rutkowski (2014).

A first attempt to such a generalisation can be found in Bélanger, Shreve, and Wong (2004) where the increasing process is predictable and right continuous.

Unlike the model of Jiao and Li, under the generalized Cox model, the intensity rate process may exist.

Organization of chapters

In Chapter 1, we begin by introducing well known facts about stochastic calculus and models of default times that we will used throughout this part of the thesis. We then illustrate an Gaussian example of conditional density.
We define all the characteristics of the default time. In particular, we appoint the definition of what we call predictable reduction of the compensator of the default time, which is sometimes called $\mathbb{F}$-compensator, where $\mathbb{F}$ is the reference filtration. We show that the Gaussian conditional density example is unique in a wide class.

In Chapter 2, we generalize the model of Jiao and Li (2018), which is developed in the context where the default time can coincide with some predictable stopping times, the latter are some predictable exogenous shock times where the shocks are modeled by an inhomogeneous Poisson process N. Our generalisation consists in relaxing the assumption that the shock times are predictable and in replacing the Poisson process by a more general process X. We investigate the characteristics of the default time. We show that, in the particular case where the shock times are the jump times of a Poisson process N which generates the filtration $\mathbb{F}$, the predictable reduction of the compensator is continuous and the intensity rate process exists.

Chapter 3 focuses on the generalized Cox model. We first present the classical Cox model of Lando (1998) where the default time is the first time when an increasing process K adapted to a given filtration and absolutely continuous with respect to Lebesgue's measure hits a level, which is a random variable independent of the given filtration. We then introduce the general case where the increasing process K is right continuous with left limits (càdlàg). We investigate the characteristics of the default time and we give many examples of our construction. A special attention will be paid to the shot-noise processes. We show that the model of Jiao and Li is a special one of the generalized Cox model. We end this chapter by introducing the case where the increasing process K is left continuous with right (càglàd) with a slight modification of the definition of the default time. We conclude that this model can be reduced to the case where K is right continuous.

In Chapter 4, we investigate some applications of the generalized Cox model in credit risk. We give closed form expression for the prices of some claims and their dynamics through a general framework. We will also be interested in the impacts of the jumps of K on the pre-default prices of defaultable bonds.

Part 2: Kriging for volatility surfaces construction

Volatility surfaces are important building blocks of risk management systems designed for pricing and hedging of financial derivatives or optional guarantees under market-consistent asset models. For a given underlying asset, the option value depends on the characteristics of the contract, i.e., the date of maturity and the strike price for equity options. However, the market value of such options is typically available (or reliable) for a limited number of standard characteristics (maturity, strike price) whereas one may require the whole price surface or the price for some unobserved characteristics for some applications. In addition, the construction of these surfaces is usually made in accordance with the principle of no arbitrage opportunity, i.e, the constructed options prices must be exempted financial arbitrage. Constructing no-arbitrage price surface can be cast into a functional learning problem with shape constraints and possibly noisy observations.

Once the price surface is interpolated, we can use the inversion formula of Black and Scholes for obtaining the implied volatility surface. For getting the local volatility surface, we can use a finite difference approximation of Dupire's formula (see Dupire, 1994). Besides it is possible to construct the local volatility surface using the implied volatilities rather than prices. This consists in extracting the local volatility surface from the implied volatility one using the formula (1.10) of Gatheral (2011). These approaches that use the observed prices (or implied volatility for constructing local volatility) in the volatility surfaces construction are known as indirect.

Beyond these direct approches, there exist some ones which parametrize directly the unobserved volatility through a function in which its parameters can be estimated by minimizing a suitable objective function obtained for calibrating that function to market data. These include among others the stochastic volatility inspired (SVI) which parametrize the implied volatility smile and its extension (SSVI) for the implied volatility surface (see Gatheral and Jacquier, 2014; Gatheral, 2004).

The most known techniques in the context of volatility surface construction using indirect approaches have been based on constrained splines, see for example, Fengler (2009), Homescu (2011), Laurini (2011), and Wang, Yin, and Qi (2004) and so forth.

Some techniques in deep learning have recently been developed in order to fill the gap in the usual models by recording better performances of learning arbitrage-free price surface. The work of Dugas et al. (2009) is the starting point for incorporating no-arbitrage constraints in neural networks (NN) for learning option prices. Dugas et al. (2009) used the so-called hard constraints approach through a special neural network architecture. This approach is different to the one called soft constraints approach which consists in penalizing some variables, in the objective function, that do not verify the constraints (see, e.g., Itkin, 2019).

The main difficulty of the approach of Dugas et al. (2009) relies on the fact that incorporating inequality constraints in the neural networks reduces its flexibility in practice and presents highly non linear transformation for option pricing computation at each iteration as shown in Ackerer, Tagasovska, and Vatter (2019). For constructing the local volatility surface, Ackerer, Tagasovska, and Vatter (2019) proposed to directly learn the implied volatility surface instead of prices surfaces by using the soft constraint approach.

Chataigner, Crépey, and Dixon (2020) proposed simple neural network architectures for both arbitrage-free option prices and implied volatility. They show that hard constraints reduce the power of the network and that soft constraints provide best accurate prices and implied volatilities.

The main limit of these approaches lies on the fact that they do not allow to quantify uncertainty in the estimation of the variables of interest.

The aim of this part is to adopt the constrained kriging techniques developed in Maatouk and Bay (2017) for imposing hard constraints and quantifying the associated uncertainty in the context of volatility surfaces construction such as implied volatility and local volatility surfaces. We use the indirect approaches that consist in learning prices through kriging and then use the inversion formula in term of implied volatility surface and calibrate Dupire's formula in term of local volatility. There is no need to extend the approach of Maatouk and Bay (2017) for more general inequality constraints as it is done in López-Lopera et al. (2018), since we limit ourselves to the convexity and monotonicity constraints.

The idea is to assume that the distribution of the underlying process is a realization of a truncated Gaussian process, verifying a number of constraints. As for the rate curves (see Cousin, Maatouk, and Rullière (2016)), the compatibility constraints with the observations (price compatibility of European options) is translated into a linear relationship on the kriging process. The arbitrage-free hypothesis on option prices is equivalent to the existence of a martingale measure: the achievements of the kriging process must correspond to marginal distributions of martingales. According to Theorem 2.6 of Beiglböck, Juillet, et al. (2016), it is then necessary to build a family of marginal distributions from a bivariate Gaussian process which is conditionally:

- increasing with respect to the direction of maturities

- convex with respect to the strike prices

- solution of a linear relationship on certain maturities and the strike prices.

Therefore, the first two constraints correspond to the inequality constraints that must be taken into account by the kriging model. The third constraint makes the kriging model compatible with the observed European options with different maturities and exercise prices.

Kriging, known as Gaussian process (GPs) regression, is a spatial interpolation approach and its techniques have been developed in geostatistics for estimating the distribution of mineral resource in the ground given the relatively small set of boreholes (see, e.g., Matheron, 1963; Cressie, 1990; Krige and Magri, 1982). It gains popularity with the works of Williams and Rasmussen (2006) and has been adapted into various areas such as hydrology, meteorology, epidemiology and among others. Recent works in kriging have been developed in quantitative finance. For instance, Sousa, Esquível, and Gaspar (2012) used kriging for learning model parameters when calibrating the Vasicek interest rate model under the risk neutral measure. They made a Gaussian prior on the zero coupon bond log prices and learnt model parameters by maximizing the log likelihood of the training data given the parameters. Ludkovski (2018) improves the Monte Carlo Least square method for the valuation of Bermuda option by using kriging for the regression step which consists in approximating the expected value from continuation. De Spiegeleer et al. (2018) show the speed up of kriging comparing to monte carlo method for pricing options and approximating implied volatility. However they do not take into account the arbitrage-free conditions in their approximation. In derivative portfolio modeling, Dixon and Crépey (2018) used a multi-response kriging in portfolio valuation through computation of credit valuation adjustment (CVA). They show that kriging provides fast approximation of portfolio value. Gonzalvez et al. (2019) explore the use of kriging in finance and show that this latter is a powerful tool for fitting the yield curve. However their interpolation does not take in consideration the arbitrage-free conditions too.

A first attempt at using kriging for local volatililty is done by Tegnér and Roberts (2019) who place a Gaussian prior directly on the local volatility surface (to guarantee the positivity of the local volatility, they assign a positive function on the prior). Such an approach leads to a nonlinear least squares training loss function, as it involves the nonlinear transformation of the local volatility into the corresponding vanilla option prices. Such a loss function is not obviously amenable to gradient descent (stochastic or not), so the authors resort to a MCMC optimization. Moreover, the performance of their GPs is measured by in-sample RMSE and does not seem to be assessed for overfitting. Furthermore, they do not benchmark their approach

against alternatives.

Actually, adding that positivity constraint and several others (such as monotonicity, convexity, boundedness) in kriging is not an easy task and constitutes a big challenge. Several approaches exist. The frameworks presented in Da Veiga and Marrel (2012), Golchi et al. (2015), Riihimäki and Vehtari (2010), and Wang and Berger (2016), among many others, are based on the fact that the constraints are only satisfied on a virtual observation locations and does not guarantee the constraints in the entire domain. For instance, for incorporating increasing monotonicity constraint in kriging, Wang and Berger (2016) proposed to select the locations with large probabilities of having negative derivatives as the virtual locations set in which they enforce the constraints. The approach of Agrell (2019) extended that of Wang and Berger (2016) for multiple constraints through an efficient method for sampling the posterior process based on the derivation of the posterior of the constrained Gaussian process using a linear operator.

Adding some inequality constraints in the quantities of interest leads to the fact that the posterior process is no more Gaussian and these constraints are usually infinite-dimensional. This motivated Maatouk and Bay (2014) to use the finite dimensional approximation of Gaussian processes for which the the inequality constraints are easy to check. They developed a sampling scheme called Rejection Sampling from the Mode (RSM) for sampling from truncated multivariate Gaussian distribution which is needed for estimating the posterior process. Their approach guarantees the constraints in the whole domain. Cousin, Maatouk, and Rullière (2016) show the extensions of classical spline interpolation by constrained kriging techniques developed in Maatouk and Bay (2014) to ensure non-arbitrable and error-controlled yield-curve and CDS curve interpolation. López-Lopera et al. (2018) extended the framework of Maatouk and Bay (2014) for more general inequality constraints and proposed to use the Hamiltonian Monte Carlo (HMC) of Pakman and Paninski (2014) which is more efficient than the RSM of Maatouk and Bay (2014) for sampling from truncated multivariate Gaussian distribution.

Unlike spline interpolation and NN approaches, kriging makes it possible to quantify the uncertainty in the estimation of the variables of interest. It is possible for example, to obtain confidence intervals at the points of the surface where the observations are unavailable or considered unsuitable.

The main contributions of the part is to show that kriging is a suitable tool for constructing option prices and quantifying uncertainty in the presence of noisy observation, and for computing the associated local and implied volatility.

Organization of chapters

The outline of the part is organized as follows.

In Chapter 5, we present some techniques in classical and constrained kriging that we use in the following chapters. A particular attention will be paid in the sampling of Gaussian random coefficients when some monotonicity constraints are saturated and we will propose a solution of this problem.

Chapter 6 focuses on the construction of implied volatility surface using kriging. We adopt the constrained kriging developed in Maatouk and Bay (2014) for constructing no-arbitrage prices. We compare the performances of the classical and constrained kriging in this context through an empirical study using the Euro Stoxx 50 data of January 10, 2019. Constrained kriging provides, over the whole interpolation criteria, better performances than claissical kriging. Moreover, due to the shape constraints, the constrained kriging approach behave better in extrapolation than an classical kriging approach. However, when the percentage of the used data is too large the estimated MAP from classical kriging performs as well as the one estimated from the constrained kriging in interpolation. This is shown by studying the accuracy of the Maximum A

Posterior (MAP) estimators using the RMSE measure.

   Chapter 7 is devoted to the construction of local volatility surface using kriging. It makes several contributions to the literature. First, we extend the free-arbitrage option price surface construction in the case with dividends and interest rate. With such free-arbitrage extended price surface, we model the local volatility surface by using Dupire's partial differential equation. Uncertainty quantification provides confidence bounds in the local volatility models. We benchmark the performance of our model against the SSVI and the NN approaches. We use a soft-constrained NN local volatility implied variance formula to derive a local volatility surface from Black-Scholes implied volatilities (see blue arrow in Fig. 7.1).
The fitting of the NN is done with mid-quotes while the one of the kriging is done with a replication of the bid and ask quotes. This is a nice property of Kriging which allows to obtain a price surface that respects the bid-ask spread constraints. In other terms, the obtained surface lies between the bid and the ask prices. Kriging provides better performances for pricing and is the only approach that gives uncertainty quantification, which is useful for risk analysis.

Part I

Part 1: Models of default times

..

Chapter 1

Preliminaries

In this chapter, we recall some definitions and standard results which will be useful in the following chapters.

1.1 Stochastic processes

Let $(\Omega, \mathcal{G}, \mathbb{H}, \mathbb{P})$ be a filtered probability space with $\mathbb{H}$ a generic filtration which verifies the usual conditions (i.e., completed and right-continuous). For any process with limits on right and limits on left (làdlàg) Y, we denote its left jump by ΔY and its right jump by $\Delta^+ Y$ which are respectively given at time t by $\Delta Y_t = Y_t - Y_{t-}$ and $\Delta^+ Y_t = Y_{t+} - Y_t$, where Y_{t-} and Y_{t+} are its left and right limits at t. If Y is continuous on right with limit on left (càdlàg), one has $\Delta^+ Y_t = 0$ for any $t \geq 0$, if Y is continuous on left with limits on right (càglàd) $\Delta Y_t = 0$ for any $t \geq 0$, and if Y is continuous $\Delta Y_t = \Delta^+ Y_t = 0$, for any $t \geq 0$. For any càdlàg increasing (or decreasing)[1] process Y, we denote by Y^c [2] its continuous part, i.e., $Y_t^c = Y_t - \sum_{s \leq t} \Delta Y_s$.

The Lebesgue-Stieltjes integral of a bounded process V with respect to a càdlàg increasing process X is, for $s < t$, denoted $\int_s^t V_u dX_u := \int_{]s,t]} V_u dX_u$. Note that $\int_{]0,t]} dX_u = X_t - X_0$ and, with the convention $X_{0-} = 0$, that $\int_{[0,t]} dX_u = X_t - X_0 + \Delta X_0 = X_t$.

In what follows, we recall the notion of semimartingale. More details about this notion can be found, e.g., in Aksamit and Jeanblanc, 2017, subsection 1.2; He, Wang, and Yan, 2018, Chapter 8.

Definition 1 *An $\mathbb{H}$-semimartingale is an $\mathbb{H}$-adapted càdlàg process X with the following decomposition:*

$$X = X_0 + N + V$$

where N is an $\mathbb{H}$-local martingale with $N_0 = 0$ and V an $\mathbb{H}$-adapted càdlàg process with finite variation and $V_0 = 0$. If, in addition, V is $\mathbb{H}$-predictable, this decomposition is unique and in that case X is called a special semimartingale.

If X is a continuous $\mathbb{H}$-semimartingale, then X is a special semimartingale, with N and V being continuous. A càdlàg $\mathbb{H}$-supermartingale (resp. $\mathbb{H}$-submartingale) is a special semimartingale.

Definition 2 *For any bounded $\mathbb{H}$-predictable process H and any càdlàg $\mathbb{H}$-special semimartingale X with decomposition $X = X_0 + N + V$, we denote by $(H \cdot X)$ the stochastic integral of H w.r.t. X which is the $\mathbb{H}$-special semimartingale given by (see He, Wang, and Yan, 2018, Definition 9.13)*

$$(H \cdot X)_t := H_0 X_0 + \int_0^t H_s dX_s = H_0 X_0 + \int_0^t H_s dN_s + \int_0^t H_s dV_s, \ \forall t \geq 0.$$

Definition 3 *Let X^c and Y^c be the continuous martingale parts of two $\mathbb{H}$-semimartingales X and Y. The quadratic covariation process of X and Y is the $\mathbb{H}$-adapted process with finite variation denoted by $[X, Y]$*

---

[1] We say that a process Y is increasing (resp. decreasing) if for $0 \leq s \leq t$, $Y_t \geq Y_s$ (resp. $Y_t \leq Y_s$) a.s.

[2] This notation should be not confused with the notation X^c, which, for a semimartingale X denotes the continuous martingale part of X.

such that (see He, Wang, and Yan, 2018, Definition 8.2)

$$[X, Y]_t := X_0 Y_0 + \langle X^c, Y^c \rangle_t + \sum_{0 < s \leq t} \Delta X_s \Delta Y_s, \ \forall \ t \geq 0$$

where $\langle X^c, Y^c \rangle$ is the unique $\mathbb{H}$-predictable process with finite variation such that $X^c Y^c - \langle X^c, Y^c \rangle$ is an $\mathbb{H}$-local martingale.

The following result is known as Yoeurp's lemma. We recall it as given in Aksamit and Jeanblanc, 2017, Proposition 1.16.

Lemma 4 *(Yoeurp's lemma). The quadratic covariation $[X, V]$ of a càdlàg $\mathbb{H}$-semimartingale X and a finite variation $\mathbb{H}$-adapted càdlàg process V verifies*

$$[X, V]_t = X_0 V_0 + (\Delta X \cdot V)_t = X_0 V_0 + \int_0^t \Delta X_s dV_s, \ \forall \ t \geq 0$$

and the integration by parts formula can be written as

$$X_t V_t = X_0 V_0 + \int_0^t X_s dV_s + \int_0^t V_{s-} dX_s, \ \forall \ t \geq 0.$$

If in addition V is $\mathbb{H}$-predictable, then $[X, V]_t = X_0 V_0 + (\Delta V \cdot X)_t$ and

$$X_t V_t = X_0 V_0 + \int_0^t X_{s-} dV_s + \int_0^t V_s dX_s, \ \forall \ t \geq 0.$$

Definition 5 *(see He, Wang, and Yan, 2018, Definition 3.15)*
We denote by $\mathcal{O}(\mathbb{H})$ (resp. $\mathcal{P}(\mathbb{H})$) the $\mathbb{H}$-optional (resp. $\mathbb{H}$-predictable) σ-algebra on $\mathbb{R}^+ \times \Omega$ and by $\mathcal{B}(\mathbb{R}^+)$ the Borelien sets of $\mathbb{R}^+$. Let V be a locally integrable variation process càdlàg with $V_0 = 0$. There exists a unique $\mathbb{H}$-optional locally integrable variation process $V^{o,\mathbb{H}}$, called the $\mathbb{H}$-dual optional projection of V, such that $V_0^{o,\mathbb{H}} = 0$ and

$$\mathbb{E}\left[\int_{[0,\infty[} X_s dV_s\right] = \mathbb{E}\left[\int_{[0,\infty[} X_s dV_s^{o,\mathbb{H}}\right]$$

for any bounded $\mathbb{H}$-optional process X such that $\mathbb{E}\left[\int_{[0,\infty[} |X_s| d|V|_s\right] < \infty$.
There exists a unique $\mathbb{H}$-predictable locally integrable variation process $V^{p,\mathbb{H}}$, called the $\mathbb{H}$-dual predictable projection of V, such that $V_0^{p,\mathbb{H}} = 0$ and

$$\mathbb{E}\left[\int_{[0,\infty[} X_s dV_s\right] = \mathbb{E}\left[\int_{[0,\infty[} X_s dV_s^{p,\mathbb{H}}\right]$$

for any bounded $\mathbb{H}$-predictable process X such that $\mathbb{E}\left[\int_{[0,\infty[} |X_s| d|V|_s\right] < \infty$. If V is $\mathbb{H}$-adapted, then $V - V^{p,\mathbb{H}}$ is an $\mathbb{H}$-martingale and $V^{p,\mathbb{H}}$ is called the $\mathbb{H}$-compensator of V.

Let us note that these projections depend on the filtration that one refers to.

We recall that any càdlàg $\mathbb{H}$-supermartingale (resp. submartingale) X admits a unique Doob-Meyer decomposition, i.e., $X = M^X - A^X$ (resp. $X = M^X + A^X$) where M^X is an $\mathbb{H}$-martingale and A^X an increasing $\mathbb{H}$-predictable process with $A_0^X = 0$.
We recall also that any strictly positive bounded càdlàg $\mathbb{H}$-supermartingale Y admits a multiplicative decomposition of the form $Y = N e^{-H}$, where N is an $\mathbb{H}$-local martingale and H an increasing $\mathbb{H}$-predictable process (see, e.g., Aksamit and Jeanblanc, 2017, Proposition 1.32).

Definition 6 *A filtration $\mathbb{F}$ is said to be immersed in a filtration $\mathbb{H}$, with $\mathbb{F} \subset \mathbb{H}$ if any $\mathbb{F}$-martingale is an $\mathbb{H}$-martingale (see Jeanblanc, Yor, and Chesney, 2009, p.316; Brémaud and Yor (1978)). In the literature (see Brémaud and Yor, 1978), it is said that the H-hypothesis holds between $\mathbb{F}$ and $\mathbb{H}$.*

Note that the immersion property depends on the choice of the probability measure.

A trivial (and useful) example of immersion between two filtrations $\mathbb{H}$ and $\mathbb{F}$ is when $\mathbb{H} = \mathbb{F} \vee \tilde{\mathbb{F}}$ where $\tilde{\mathbb{F}}$ is independent of $\mathbb{F}$ under $\mathbb{P}$.

Lemma 7 *Let $\mathbb{H}^{(1)}, \mathbb{H}^{(2)}$ and $\mathbb{H}^{(3)}$ be three filtrations. If $\mathbb{H}^{(1)} \subset \mathbb{H}^{(3)}$ and Y is an $\mathbb{H}^{(3)}$-martingale which is $\mathbb{H}^{(1)}$-adapted, then Y is an $\mathbb{H}^{(1)}$-martingale. In particular, if $\mathbb{H}^{(1)}$ is immersed in $\mathbb{H}^{(3)}$ and $\mathbb{H}^{(1)} \subset \mathbb{H}^{(2)} \subset \mathbb{H}^{(3)}$, then $\mathbb{H}^{(1)}$ is immersed in $\mathbb{H}^{(2)}$.*

PROOF: Let Y be an $\mathbb{H}^{(1)}$-adapted $\mathbb{H}^{(3)}$-martingale. For any $0 \le s \le t$, one has by using the tower property

$$\mathbb{E}[Y_t | \mathcal{H}_s^{(1)}] = \mathbb{E}[\mathbb{E}[Y_t | \mathcal{H}_s^{(3)}] | \mathcal{H}_s^{(1)}] = \mathbb{E}[Y_s | \mathcal{H}_s^{(1)}] = Y_s$$

where we have used, in the second equality, the martingale property of Y with respect to $\mathbb{H}^{(3)}$ and, in the last equality, the fact that Y is $\mathbb{H}^{(1)}$-adapted. $\square$

Lemma 8 *Let $\mathbb{F} = (\mathcal{F}_t)_{t \ge 0}$, $\mathbb{H}^1 = (\mathcal{H}_t^1)_{t \ge 0}$ and $\mathbb{H}^2 = (\mathcal{H}_t^2)_{t \ge 0}$ be three filtrations such that $\mathbb{H}^1$ is independent of $\mathbb{H}^2$ and $\mathcal{F}_t = \mathcal{H}_t^1 \vee \mathcal{H}_t^2, \forall t \ge 0$. Then, for $T \ge t$, $\mathcal{H}_T^1$ and $\mathcal{H}_T^2$ are conditionally independent with respect to $\mathcal{F}_t$.*

PROOF: Let X, Y be bounded random variables with $X \in \mathcal{H}_T^1$ and $Y \in \mathcal{H}_T^2$. Then, from tower property

$$\mathbb{E}[XY | \mathcal{F}_t] = \mathbb{E}[\mathbb{E}[XY | \mathcal{H}_T^2 \vee \mathcal{H}_t^1] | \mathcal{F}_t]$$
$$= \mathbb{E}[Y \mathbb{E}[X | \mathcal{H}_T^2 \vee \mathcal{H}_t^1] | \mathcal{F}_t]$$

where the second equality follows from the fact that $Y \in \mathcal{H}_T^2 \subset \mathcal{H}_T^2 \vee \mathcal{H}_t^1$. Since $\mathbb{H}^1$ is independent of $\mathbb{H}^2$, $X \in \mathcal{H}_T^1$ is independent of $\mathcal{H}_T^2$ and $\mathbb{E}[X | \mathcal{H}_t^1 \vee \mathcal{H}_T^2] = \mathbb{E}[X | \mathcal{H}_t^1]$. Hence,

$$\mathbb{E}[XY | \mathcal{F}_t] = \mathbb{E}[Y \mathbb{E}[X | \mathcal{H}_t^1] | \mathcal{F}_t]$$
$$= \mathbb{E}[X | \mathcal{H}_t^1] \mathbb{E}[Y | \mathcal{F}_t]$$
$$= \mathbb{E}[X | \mathcal{F}_t] \mathbb{E}[Y | \mathcal{F}_t].$$

The second equality is due to the fact that $\mathbb{E}[X | \mathcal{H}_t^1]$ is $\mathcal{F}_t$-measurable and the last one comes from the independence of $\mathbb{H}^1$ and $\mathbb{H}^2$ which allows to write, for $X \in \mathcal{H}_T^1$, $\mathbb{E}[X | \mathcal{H}_t^1] = \mathbb{E}[X | \mathcal{H}_t^1 \vee \mathcal{H}_t^2] = \mathbb{E}[X | \mathcal{F}_t]$. $\square$

We recall that the graph of a finite random time τ (i.e., a non-negative random variable) is the subset $[\![\tau]\!]$ of $\Omega \times \mathbb{R}^+$ defined as

$$[\![\tau]\!] = \{(\omega, t) : \tau(\omega) = t\}.$$

Definition 9 *A random time τ avoids all $\mathbb{H}$-stopping times if $\mathbb{P}(\tau = \zeta < \infty) = 0$, for any $\mathbb{H}$-stopping time ξ.*

Definition 10 *(see He, Wang, and Yan, 2018, Section 3, Chapter 3)*
An $\mathbb{H}$-stopping time ϑ is said to be $\mathbb{H}$-predictable if there exists an increasing sequence of $\mathbb{H}$-stopping times $(\vartheta_i)_{i \ge 1}$ converging to ϑ such that $\vartheta_i < \vartheta$ on the set $\{\vartheta_i > 0\}$, for all i. If ϑ is $\mathbb{H}$-predictable, $(\mathbb{1}_{\{\vartheta \le t\}}, t \ge 0)$ is a predictable process.

An $\mathbb{H}$-stopping ϑ is said to be accessible if $[\![\vartheta]\!] \subset \cup_i [\![\vartheta_i]\!]$ where $(\vartheta_i)_{i \geq 1}$ are $\mathbb{H}$-predictable stopping times.
An $\mathbb{H}$-stopping time ϑ is said to be totally inaccessible if it avoids all $\mathbb{H}$-predictable stopping times (i.e., $\mathbb{P}(\vartheta = \xi < \infty) = 0$ for any $\mathbb{H}$-predictable stopping time ξ).

1.2 Default time

We consider a probability space $(\Omega, \mathcal{G}, \mathbb{P})$ and τ a random time that we assume to be finite (except for some specific examples), defined on $(\Omega, \mathcal{G})$. We introduce the right-continuous increasing default process $A_t = \mathbb{1}_{\{\tau \leq t\}}$ associated with τ and we denote by $\mathbb{A} = (\mathcal{A}_t)_{t \geq 0}$ the filtration (completed and right-continuous) generated by this default process. We recall that, for any process X, one has $\int_u^t X_s dA_s = \int_{]u,t]} X_s dA_s = X_\tau \mathbb{1}_{\{u < \tau \leq t\}}$. For a given filtration $\mathbb{H}$ on Ω, we introduce the $\mathbb{H}$-dual predictable projection of the default process A, i.e., $A^{p,\mathbb{H}}$. By abuse of language, we shall sometimes say that $A^{p,\mathbb{H}}$ is the $\mathbb{H}$-dual predictable projection of τ.

If τ is an $\mathbb{H}$-stopping time, the compensator of τ is by definition the unique $\mathbb{H}$-predictable increasing process $J^{\mathbb{H}}$ such that $J_0^{\mathbb{H}} = 0$ and $A_t - J_t^{\mathbb{H}}$ is an $\mathbb{H}$-martingale (see Jeanblanc, Yor, and Chesney, 2009, p.265). Note that $J_t^{\mathbb{H}} = J_{t \wedge \tau}^{\mathbb{H}}$. This compensator $J^{\mathbb{H}}$ of τ is nothing else than $A^{p,\mathbb{H}}$. This property extends as follows:

Lemma 11 *For any $\mathbb{H}$-predictable bounded process H, the process*

$$H_\tau \mathbb{1}_{\{\tau \leq t\}} - \int_0^{t \wedge \tau} H_s dA_s^{p,\mathbb{H}}$$

is an $\mathbb{H}$-martingale.

PROOF: This result follows from the fact that

$$H_\tau \mathbb{1}_{\{\tau \leq t\}} = \int_0^t H_s dA_s = (H \cdot A)_t,$$

and, for H being $\mathbb{H}$-predictable, the $\mathbb{H}$-dual predictable projection of $H \cdot A$ is $H \cdot A^{p,\mathbb{H}}$ (see He, Wang, and Yan, 2018, p. 148, Theorem 5.23). $\qquad\square$

We now work on a filtered probability space $(\Omega, \mathcal{G}, \mathbb{F}, \mathbb{P})$ on which a random time τ is defined. We call $\mathbb{F}$ the reference filtration. We consider the links between the default time and the reference filtration. We denote by Z the Azéma supermartingale (see Azéma, 1972; Jeanblanc, Yor, and Chesney, 2009, Subsection 5.9.4; Nikeghbali, 2006) associated with τ, which satisfies $Z_t := \mathbb{P}(\tau > t | \mathcal{F}_t)$. Note that $Z_t > 0$ on $\{\tau > t\}$ and $Z_{t-} > 0$ on $\{\tau \geq t\}$ (see Aksamit and Jeanblanc, 2017, Lemma 2.14). Then, $A^{p,\mathbb{F}}$, the $\mathbb{F}$-dual predictable projection of A, is also the predictable part in the Doob-Meyer decomposition of $Z_t := m_t - A_t^{p,\mathbb{F}}$ where m is an $\mathbb{F}$-martingale (see Aksamit and Jeanblanc, 2017, subsection 2.2, page 33).

We introduce the so-called second Azéma's supermartingale $\tilde{Z}$ associated to τ, which satisfies $\tilde{Z}_t := \mathbb{P}(\tau \geq t | \mathcal{F}_t) = Z_t - \Delta A_t^{o,\mathbb{F}}$ (see Aksamit and Jeanblanc, 2017, Proposition 1.46). Note that the Azéma supermartingale Z is right-continuous with left limits while the supermartingale $\tilde{Z}$ is a process with right and left limits.

Definition 12 *Let $\mathbb{G} = (\mathcal{G}_t)_{t \geq 0}$ be the progressive enlargement of $\mathbb{F}$ with τ, i.e., $\mathbb{G} = \mathbb{F} \vee \mathbb{A}$, which means that $\mathcal{G}_t = \cap_{\epsilon > 0} \mathcal{G}_{t+\epsilon}^0$, with $\mathcal{G}_s^0 = \mathcal{F}_s \vee \mathcal{A}_s$ for every $s \geq 0$ (see, e.g., Jeulin and Yor, 1978; Yor, 1978).*
The filtration $\mathbb{G}$ is the smallest filtration satisfying the usual hypotheses containing $\mathbb{F}$ and turning out τ into a stopping time.

Definition 13 *The $\mathbb{F}$-predictable reduction of the compensator of τ*
The process Λ given by

$$\Lambda_t = \int_0^t \mathbb{1}_{\{Z_{s-} > 0\}} \frac{dA_s^{p,\mathbb{F}}}{Z_{s-}} \tag{1.2.1}$$

is $\mathbb{F}$-predictable and increasing, and, denoting by Λ^τ the process Λ stopped at time τ,

$$A_t - \Lambda_{t \wedge \tau} = A_t - \Lambda_t^\tau = A_t - \int_0^{t \wedge \tau} \frac{dA_s^{p,\mathbb{F}}}{Z_{s-}}$$

is a $\mathbb{G}$-martingale (see Aksamit and Jeanblanc, 2017, Proposition 2.15).
The process Λ^τ is the $\mathbb{G}$-compensator of the default process A (we shall also say compensator of τ) and we call Λ the $\mathbb{F}$-predictable reduction of the $\mathbb{G}$-compensator of τ.
If Λ is absolutely continuous with respect to the Lebesgue measure, i.e., $\Lambda_t = \int_0^t \lambda_s ds$, then its derivative λ, which is a non-negative $\mathbb{F}$-predictable process, is called the $\mathbb{F}$-intensity rate.

We recall that the default time τ avoids all $\mathbb{F}$-stopping times (resp. all $\mathbb{F}$-predictable stopping times) if and only if $A^{o,\mathbb{F}}$ (resp. $A^{p,\mathbb{F}}$) is continuous (see Aksamit and Jeanblanc, 2017, Proposition 1.43). It can be proved that the jump times of $A^{o,\mathbb{F}}$ are $\mathbb{F}$-stopping times not avoided by τ.
The assumption that the default time τ avoids all $\mathbb{F}$-stopping times is widely made in the literature of progressive enlargement of filtration (see, e.g., in Nikeghbali, 2006).

In the case of progressive enlargement of filtration, the immersion property is easily characterized:

Lemma 14 *(see Aksamit and Jeanblanc, 2017, Lemma 3.8)*
The filtration $\mathbb{F}$ is immersed in $\mathbb{G}$ if and only if

$$\mathbb{P}(\tau > t | \mathcal{F}_t) = \mathbb{P}(\tau > t | \mathcal{F}_\infty), \forall t \geq 0.$$

We now give some definitions relative to the conditional law of τ

Definition 15 *A family of stochastic processes $(M_t(u))_{u,t \in \mathbb{R}^+}$ is called a $(\mathbb{P}, \mathbb{F})$-martingale survival process if it satisfies (see El Karoui et al., 2014, Section 2):*

- *for every $u \in \mathbb{R}^+$, for every $t \in \mathbb{R}^+$, the random variable $M_t(u)$ is valued in $[0,1]$,*

- *for every $u \in \mathbb{R}^+$, the process $(M_t(u))_{t \in \mathbb{R}^+}$ is a $(\mathbb{P}, \mathbb{F})$-martingale,*

- *for every $t \in \mathbb{R}^+$, the family $M_t(u), u \in \mathbb{R}^+$ is decreasing with respect to u.*

In a default setting, we will be interested with the conditional survival process of the default time τ, i.e., $M_t(u) = \mathbb{P}(\tau > u | \mathcal{F}_t)$, which is a $(\mathbb{P}, \mathbb{F})$-martingale survival process. It is shown in Jeanblanc and Song (2011a) that to any $(\mathbb{P}, \mathbb{F})$-martingale survival process M, one can associate, on an extended probability space, a random time τ such that M is the conditional survival process of τ.

Definition 16 *A family of non-negative $\mathcal{O}(\mathbb{F}) \bigotimes \mathcal{B}(\mathbb{R}^+)$-measurable functions $(\omega, t, u) \to p_t(\omega, u)$ is called a $(\mathbb{P}, \mathbb{F})$-conditional density process if (see El Karoui et al., 2014, Condition (A))*

- *for every $u \in \mathbb{R}^+, (p_t(u))_{t \in \mathbb{R}^+}$ is a non-negative $(\mathbb{P}, \mathbb{F})$-martingale,*

- *for every $t \in \mathbb{R}^+, \int_0^\infty p_t(u)\eta(du) = 1$ a.s, where η is a probability law on $\mathbb{R}^+$.*

If M is a $(\mathbb{P}, \mathbb{F})$-martingale survival process satisfying $M_t(0) = 1$ for all $t \geq 0$ and $M_t(u) = \int_u^\infty p_t(\theta)\eta(d\theta)$ a.s, with η is a probability law on $\mathbb{R}^+$, then p is a conditional density process.
The processes $Z, \widetilde{Z}$, the $\mathbb{F}$-dual predictable and optional projections $A^{p,\mathbb{F}}, A^{o,\mathbb{F}}$ of τ, the compensator process Λ^τ as well as its $\mathbb{F}$-predictable reduction Λ and the conditional survival process $(M_t(u))_{u,t \in \mathbb{R}^+}$ are called the characteristics of the default time τ (see Jeanblanc and Li, 2020 for more details).
In what follows, $\mathbb{F}$ being the reference filtration, we shall simply denote by A^p and A^o the $\mathbb{F}$-dual predictable and optional projections of A.

Lemma 17 ***Key Lemma 1*** *(see, e.g., Jeanblanc, Yor, and Chesney, 2009, Lemma 7.4.1.1)*
For any integrable random variable Y, one has

$$\mathbb{1}_{\{\tau>t\}}\mathbb{E}[Y|\mathcal{G}_t] = \mathbb{1}_{\{\tau>t\}}\frac{\mathbb{E}[Y\mathbb{1}_{\{\tau>t\}}|\mathcal{F}_t]}{Z_t}. \tag{1.2.2}$$

If furthermore $Y \in \mathcal{F}_T$,

$$\mathbb{E}[Y\mathbb{1}_{\{\tau>T\}}|\mathcal{G}_t] = \mathbb{1}_{\{\tau>t\}}\frac{\mathbb{E}[YZ_T|\mathcal{F}_t]}{Z_t}. \tag{1.2.3}$$

Lemma 18 ***Key Lemma 2*** *(see, e.g., Jeanblanc, Yor, and Chesney, 2009, Lemma 7.4.1.2)*
For any bounded $\mathbb{F}$-predictable process R,

$$\mathbb{E}[R_\tau\mathbb{1}_{\{\tau\leq T\}}|\mathcal{G}_t] = R_\tau\mathbb{1}_{\{\tau\leq t\}} - \mathbb{1}_{\{\tau>t\}}\frac{\mathbb{E}[\int_t^T R_u dZ_u|\mathcal{F}_t]}{Z_t} = R_\tau\mathbb{1}_{\{\tau\leq t\}} + \mathbb{1}_{\{\tau>t\}}\frac{\mathbb{E}[\int_t^T R_u dA_u^p|\mathcal{F}_t]}{Z_t}. \tag{1.2.4}$$

1.3 Conditional density approach

In this Section, we investigate the conditional density approach for credit risk modeling. We first introduce the density hypothesis within the meaning of Jacod (1985, Condition (A)), before tackling the so-called extended density hypothesis of Li and Rutkowski (2014, Proposition 2.5) and the so-called generalized density hypothesis of Jiao and Li (2015). We also investigate some examples of conditional density processes.
In order to highlight the density approach rule in default modeling we first recall the general framework to valuate a defaultable claim.

1.3.1 General framework of valuation of defaultable claims

This paragraph is based on Bielecki and Rutkowski (2002) and more details related to this topic can be found in Bielecki, Jeanblanc, and Rutkowski (2009); Jeanblanc, Yor, and Chesney, 2009, Chapter 7; Aksamit and Jeanblanc, 2017, Chapter 2.
The valuation formula of a defaultable claim with default payment (or recovery) h relies on some characteristics of the default time. Following Jeanblanc and Li (2020), if the recovery process h is $\mathbb{F}$-predictable, then the pre-default value is obtained using the $\mathbb{F}$-predictable projection A^p of the default time or equivalently by the Azéma supermartingale Z, while when h is $\mathbb{F}$-optional, the pre-default value of the claim is determined the $\mathbb{F}$-dual optional projection A^o. In what follows, we recall the pricing formula by considering predictable recovery. The case with optional recovery can be obtained in the same manner by letting A^o to play the rule of A^p in the first case.

Consider a defaultable claim which consists in a single payment of a positive value V_T (where V_T is an $\mathcal{F}_T$-measurable integrable random variable) at maturity T when default does not occur before T and a payment of the recovery h (with h a bounded $\mathbb{F}$-predictable process) evaluated at τ when default occurs before maturity T. When the payment of the recovery is made at hit, the discounted dividend process ζ at time t has the following expression

$$\zeta := e^{-\int_t^T r_s ds}V_T\mathbb{1}_{\{\tau>T\}} + e^{-\int_t^\tau r_s ds}h_\tau\mathbb{1}_{\{t<\tau\leq T\}},$$

with r a non-negative $\mathbb{F}$-adapted process which is the interest rate. By assuming that $\mathbb{P}$ is the pricing measure, the price of the defaultable claim is given by a direct application of Lemma 17 and 18 (see, e.g., Bielecki and Rutkowski, 2002, proposition 8.2.1; Bielecki, Jeanblanc, and Rutkowski, 2009, Lemma 7.4.1.2)

$$\mathbb{E}[\zeta|\mathcal{G}_t] = \frac{\mathbb{E}\left[e^{-\int_t^T r_s ds}V_T Z_T + \int_t^T e^{-\int_t^u r_s ds}h_u dA_u^{p,\mathbb{F}}|\mathcal{F}_t\right]}{Z_t}\mathbb{1}_{\{\tau>t\}}. \tag{1.3.1}$$

If the actualized payoff at time t is of the form $\zeta := e^{-\int_t^T r_s ds} f(V_T, \tau) \mathbb{1}_{\{\tau < T\}}$, where f is a bounded function from $\mathbb{R} \times \mathbb{R}^+$ into $\mathbb{R}$, then the Azéma supermartingale Z does not allow to compute the quantity $\mathbb{E}[\zeta | \mathcal{G}_t]$ on the set $\{\tau > t\}$. In such a case, the $\mathbb{F}$-conditional density process $p(u)$ of the default time τ, if it exists (see below the definition) provides the needed tool.

Definition 19 *The default time τ satisfies the density hypothesis (CL) - which we mean classical density hypothesis - Jacod, 1985, Condition (A), if there exists a conditional density process $(p_t(u))_{t, u \in \mathbb{R}^+}$ (we call it (CL)-conditional density) such that for any bounded Borel function h*

$$\mathbb{E}[h(\tau) | \mathcal{F}_t] = \int_0^\infty h(u) p_t(u) \eta(du), \forall t \geq 0, \ a.s$$

where η is a probability law on $\mathbb{R}^+$.

The existence of a (CL) conditional density is a strong hypothesis, not always satisfied, whereas the martingale survival process of the default time always exist.

If this hypothesis is satisfied, one can always choose $\eta = \beta$, where β is the law of τ (see Jacod, 1985, Proposition 1.5). Furthermore, if β is no atomic, τ avoids $\mathbb{F}$-stopping times (see El Karoui, Jeanblanc, and Jiao, 2010, Corollary 2.2). Nevertheless, if β has an atom at t^*, then $\mathbb{P}(\tau = t^*) > 0$ and the constant stopping time t^* is not avoided by τ.

Under (CL) $A_t^p = \int_{[0,t]} p_{u-}(u) \eta(du)$ and $A_t^o = \int_{[0,t]} p_u(u) \eta(du)$ (see Aksamit and Jeanblanc, 2017, Corollary 5.27). Furthermore, $\mathbb{F}$ is immersed in $\mathbb{G}$ if and only if $p_t(u) = p_u(u)$, for η-a.e. u satisfying $u \leq t$. (see El Karoui, Jeanblanc, and Jiao, 2010, subsection 3.2)

If the conditional density exists, then

$$M_t(u) := \mathbb{P}(\tau > u | \mathcal{F}_t) = \int_u^\infty p_t(\theta) \eta(d\theta), a.s \tag{1.3.2}$$

is the family of martingale survival process associated with τ.

The price of the defaultable claim $\zeta := f(V_T, \tau) \mathbb{1}_{\{\tau < T\}}$ paid at time T is given by (see El Karoui, Jeanblanc, and Jiao, 2010, Theorem 3.1)

$$\mathbb{E}[\zeta | \mathcal{G}_t] = \mathbb{1}_{\{\tau > t\}} \frac{\mathbb{E}\left[\int_t^T e^{-\int_t^T r_s ds} f(V_T, u) p_T(u) \eta(du) | \mathcal{F}_t\right]}{Z_t}. \tag{1.3.3}$$

This formula (1.3.3) implies the fact that knowing the conditional law of τ with respect to the reference filtration provides more support in the pricing derivatives.

Li and Rutkowski, 2014, Definition 2.5 introduced the extended density hypothesis denoted by (ED) in order to overcome a non-degeneracy condition of the $\mathbb{F}$-conditional law of τ. The following definition focusses on this extended density approach.

Definition 20 *The default time τ satisfies the extended density hypothesis (ED) (see Li and Rutkowski, 2014), if, for any u, there exists a family of non-negative martingales $(m_t(u)_{t \geq u})$ and an $\mathbb{F}$-adapted, increasing process D, with $D_0 = 0$ such that for all $0 \leq u \leq t$,*

$$\mathbb{P}(\tau \leq u | \mathcal{F}_t) = \int_0^u m_t(s) dD_s, \forall t \geq 0.$$

Under the classical density approach (CL), if the law of τ has no atoms, τ is a totally inaccessible $\mathbb{G}$-stopping time which avoids all $\mathbb{F}$-stopping times and this does not allow to model some credit risks where the default time can coincide with some $\mathbb{F}$-predictable stopping times. That motivated Jiao and Li to introduce the generalized density hypothesis that we denote (GD) which postulates that the part of the random time

τ which avoids a sequence of $\mathbb{F}$-stopping times, admits an $\mathbb{F}$-conditional density (see Jiao and Li, 2015, Assumption 2.4).

Definition 21 *The default time τ satisfies the generalized density hypothesis (GD) (see Jiao and Li, 2015) if there exists a non-atomic non-negative σ-finite Borel measure υ on $\mathbb{R}^+$, a strictly increasing sequence of $\mathbb{F}$-stopping times $(\tau_i)_i$ and a family $(\alpha_t(u))_{t,u\in\mathbb{R}^+}$ of non-negative $\mathbb{F}$-martingales (we call it (GD)-conditional density) such that for any bounded Borel function h ,*

$$\mathbb{E}\left[h(\tau)\prod_{i\geq 1}\mathbb{1}_{\{\tau\neq\tau_i\}}|\mathcal{F}_t\right] = \int_{\mathbb{R}^+} h(u)\alpha_t(u)\upsilon(du), \ \forall t\geq 0, \ \ a.s.$$

Proposition 22 *For any bounded Borel function h, if the Generalized density hypothesis is satisfied,*

$$\mathbb{E}[h(\tau)|\mathcal{F}_t] = \int_0^\infty h(u)\alpha_t(u)\upsilon(du) + \sum_{i=1}^\infty \mathbb{E}[h(\tau_i)p_{t\vee\tau_i}^i|\mathcal{F}_t].$$

PROOF:

$$\mathbb{E}[h(\tau)|\mathcal{F}_t] = \mathbb{E}[h(\tau)\prod_{i=1}^\infty \mathbb{1}_{\{\tau\neq\tau_i\}}|\mathcal{F}_t] + \sum_{i=1}^\infty \mathbb{E}[h(\tau)\mathbb{1}_{\{\tau=\tau_i\}}|\mathcal{F}_t].$$

Using the tower property, we obtain

$$\mathbb{E}[h(\tau)\mathbb{1}_{\{\tau=\tau_i\}}|\mathcal{F}_t] = \mathbb{E}\left[h(\tau_i)\mathbb{P}(\tau=\tau_i|\mathcal{F}_{\tau_i\vee t})|\mathcal{F}_t\right] = \mathbb{E}\left[h(\tau_i)p_{t\vee\tau_i}^i|\mathcal{F}_t\right].$$

$\square$

Comments :

- If τ avoids all $\mathbb{F}$-stopping times, the existence of a (GD)-conditional density is equivalent to the existence of a (CL)-conditional density.

- If the (CL)-conditional density exists then the extended density hypothesis (ED) exists. The (ED) may exist while the (CL) and the (GD)-conditional density do not exist.

1.3.2 Gaussian example

1.3.2.1 El Karoui et al. (2014) example

In this subsection, we recall the example of El Karoui et al. (2014) (see also Zargari, 2011, Chapter 2, subsection 2.2).

Following El Karoui et al. (2014), starting from a square integrable deterministic function g on $\mathbb{R}^+$ and introducing the martingale μ as $\mu_t = \int_0^t g(s)dB_s$, where B is a Brownian motion, one can construct a random time $\tau = l^{-1}(\mu_\infty)$, where l is a strictly increasing function, differentiable, from $\mathbb{R}^+$ to $\mathbb{R}$. By denoting $\mathbb{F}$ the Brownian filtration generated by B, it is shown in El Karoui et al. (2014) that, for any $t, \theta \geq 0$,

$$M_t(u) := \mathbb{P}(\tau > u|\mathcal{F}_t) = \Phi\left(\frac{\mu_t - l(u)}{\nu(t)}\right), \tag{1.3.4}$$

where $\nu(t) := \int_t^\infty g^2(s)ds$ is assumed to be strictly positive, and Φ is the standard Gaussian cumulative function. The family $(M_t(u))_{t,u\geq 0}$ is a family of martingales, obviously valued in $[0,1]$ and decreasing w.r.t. u with dynamics

$$dM_t(u) = \varphi\left(\frac{\mu_t - l(u)}{\nu(t)}\right)\frac{g(t)}{\nu(t)}dB_t$$

where φ is the standard Gaussian density function, or setting $U_t(u) = \frac{\mu_t - l(u)}{\nu(t)}$

$$dM_t(u) = M_t(u)\frac{\varphi(U_t(u))}{\Phi(U_t(u))}\frac{g(t)}{\nu(t)}dB_t \tag{1.3.5}$$

which allows us to write $M(u)$ as the exponential martingale of $\int_0^{\cdot} \frac{\varphi(U_t(u))}{\Phi(U_t(u))}\frac{g(t)}{\nu(t)}dB_t$.

The density of τ is obtained by differentiating $M_0(u)$ w.r.t. u, i.e.,

$$\rho(u) = \frac{1}{\sqrt{2\pi\nu(0)}}l'(u)\exp\left\{-\frac{1}{2}\left(\frac{\mu_0 - l(u)}{\nu(0)}\right)^2\right\}$$

where $\nu(0) := \int_0^\infty g^2(s)ds$. The conditional density process obtained by differentiating $M_t(u)$ w.r.t. u, (and taking η as the law of τ) is

$$p_t(u) = \frac{1}{\rho(u)\sqrt{2\pi\nu(t)}}l'(u)\exp\left\{-\frac{1}{2}\left(\frac{\mu_t - l(u)}{\nu(t)}\right)^2\right\}$$

and its dynamics is given by

$$dp_t(u) = -p_t(u)\frac{\mu_t - l(u)}{\nu^2(t)}g(t)dB_t.$$

1.3.2.2   Starting from Φ

In the previous subsubsection, we have recalled a family of conditional laws constructed from the standard Gaussian cumulative function. One wonders whether it is possible to obtain, from this distribution function, additional conditional laws. In this subsubsection, we show that the previous example is unique in a wide class. As such, we look for a diffusion X satisfying the following differential stochastic equation

$$dX_t = a(X_t, t)dt + \sigma(X_t, t)dB_t, \ X_0 = \theta$$

where B is a standard Brownian motion, and two deterministic functions f and h from $\mathbb{R}^+$ to $\mathbb{R}$ such that

$$Z_t(u) := \Phi(X_t - f(t)h(u))$$

defines a martingale survival process (see Definition 15).

The decreasing property w.r.t. the parameter u is obtained if h is an increasing strictly positive function and f is a strictly positive function. The initial (resp. terminal) condition on $Z_t(\cdot)$ imposes that $Z_0(0) = 1$ and $Z_0(\infty) = 0$, hence we restrict our attention to the case $h(0) = -\infty$ and $h(\infty) = \infty$.

We now determine some conditions on the coefficients of the diffusion X such that $Z(u)$ is a martingale for any u. As such, by denoting $U_t := X_t - f(t)h(u)$, the Itô formula shows that

$$dZ_t(u) = \left\{\Phi'(U_t)(-f'(t)h(u) + a(X_t, t)) + \frac{1}{2}\Phi''(U_t)\sigma^2(X_t, t)\right\}dt + \Phi'(U_t)\sigma(X_t, t)dB_t \tag{1.3.6}$$

and by using the fact that $\Phi''(x) = -x\Phi'(x)$, it follows that for all $u \geq 0$, $Z(u)$ is a local martingale if and only if

$$\Phi'(U_t)\left\{-f'(t)h(u) + a(X_t, t) - \frac{1}{2}U_t\sigma^2(X_t, t)\right\} = 0. \tag{1.3.7}$$

Note that Z being bounded will be a true martingale. This implies that for all $u, t \geq 0$

$$h(u)\left\{-f'(t) + \frac{1}{2}f(t)\sigma^2(X_t, t)\right\} + a(X_t, t) - \frac{1}{2}X_t\sigma^2(X_t, t) = 0.$$

Knowing that this last equality holds for all $u \geq 0$, one obtains for any $t \geq 0$

$$a(X_t, t) - \frac{1}{2} X_t \sigma^2(X_t, t) = 0 \tag{1.3.8}$$

$$- f'(t) + \frac{1}{2} f(t) \sigma^2(X_t, t) = 0 \ . \tag{1.3.9}$$

By assuming that the diffusion X is not degenerated (i.e., X is not constant), the equation (1.3.9) establishes that $\sigma^2(x, t)$ does not depend on x. Hence one denotes it by $\sigma^2(t)$. The equations (1.3.8) and (1.3.9) imply

$$a(x, t) - \frac{1}{2} x \sigma^2(t) = 0 \tag{1.3.10}$$

$$- f'(t) + \frac{1}{2} f(t) \sigma^2(t) = 0 \tag{1.3.11}$$

for all x which belongs to the support of X_t. Therefore, it follows that

$$\sigma^2(t) = 2 \frac{f'(t)}{f(t)}$$

$$a(x, t) = \frac{1}{2} x \sigma^2(t) = x \frac{f'(t)}{f(t)}$$

and

$$dX_t = X_t \frac{f'(t)}{f(t)} dt + \sqrt{2 \frac{f'(t)}{f(t)}} dB_t,$$

for f such that $\frac{f'}{f} \geq 0$ for all $t \geq 0$, hence for f such that $f' \geq 0$ since $f > 0$.
This line of thinking proves the following result:

Proposition 23 *Let h be an increasing deterministic function from $\mathbb{R}^+$ to $\mathbb{R}$ and f a strictly positive and differentiable increasing function. The set of diffusions X, such that*

$$Z_t(u) := \Phi \left(X_t - f(t) h(u) \right) \tag{1.3.12}$$

defines a martingale survival process, is the one of the form

$$dX_t = X_t \frac{f'(t)}{f(t)} dt + \sqrt{2 \frac{f'(t)}{f(t)}} dB_t, \ X_0 = \theta. \tag{1.3.13}$$

One denotes by $\mathcal{D}$ the set of such diffusions X.

1.3.2.3 The Gaussian conditional density construction using the class $\mathcal{D}$

We now investigate this class of diffusions $\mathcal{D}$ in the perspective of the result obtained in the previous subsubsection which stipulated that

$$\Phi \left(\frac{\mu_t - l(u)}{\nu(t)} \right),$$

where $\nu(t) := \int_t^\infty g^2(s) ds$ and $\mu_t = \int_0^t g(s) dB_s$, is a martingale survival process. We show that this result is driven by a class of diffusions denoted $\mathcal{C}$, which is contained in the class $\mathcal{D}$. As such, we take $X_t = \frac{\mu_t}{\nu(t)}$, $f(t) = \frac{1}{\nu(t)}$ and $h = l$, since for any $u \geq 0$, $\Phi(X_t - \frac{1}{\nu(t)} h(u))$ is a martingale, the process X belongs to the

class $\mathcal{D}$ with $X_0 = 0$ and satisfies

$$
\begin{aligned}
dX_t =& - X_t \frac{\nu'(t)}{\nu(t)} dt + \frac{g(t)}{\nu(t)} dB_t \\
=& - X_t \frac{\nu'(t)}{\nu(t)} dt + \sqrt{2(-\frac{\nu'(t)}{\nu(t)})} dB_t \\
=& X_t \frac{f'(t)}{f(t)} dt + \sqrt{2\frac{f'(t)}{f(t)}} dB_t.
\end{aligned}
$$

By consequence, the class $\mathcal{D}$ is a little bit larger than $\mathcal{C}$ due to the choice of the initial value of X (it is rather obvious that, w.l.g., one can choose $f(0) = 1$).

Consider now an element X of the class $\mathcal{D}$ (i.e., the dynamics of X is of the form (1.3.13)) with initial value θ. These dynamics are well known as a linear stochastic differential equation and admit the following solution

$$
X_t(\theta) = \exp[\int_0^s \frac{f'(u)}{f(u)} du] \left(\theta + \int_0^t \sqrt{2\frac{f'(s)}{f(s)}} \exp[-\int_0^s \frac{f'(u)}{f(u)} du] dB_s \right).
$$

Since $f(0) = 1$, one has $\exp[\int_0^s \frac{f'(u)}{f(u)} du] = f(t)$. Therefore, by denoting $\mu_t := \int_0^t g(s) dB_s$ where $g(t) := \sqrt{2\frac{f'(t)}{f(t)}} \frac{1}{f(t)}$ and $\nu(t) := \frac{1}{f(t)}$ (which implies, by simple computation, that $\nu^2(t) = \int_t^\infty g(s) ds$), one has from $\mathbb{P}(\tau > u | \mathcal{F}_t) = \Phi(X_t - f(t)h(u))$, that $\mathbb{P}(\tau > u) = \Phi(\theta - h(u))$, which can be greater or smaller that the value obtained in the case $\theta = 0$.

1.3.2.4   Dynamics of $Z(u)$

We have shown that for any martingale X of the form (1.3.12) (i.e., X belongs to the class $\mathcal{D}$), the family $Z_t(u) := \Phi(X_t - f(t)h(u))$, with h an increasing deterministic function from $\mathbb{R}^+$ to $\mathbb{R}$ and f a strictly positive function, defines a martingale survival process. Using (1.3.6) and (1.3.7), the dynamics of $Z(u)$ is of the form

$$
dZ_t(u) = \varphi(X_t - f(t)h(u)) \sqrt{2\frac{f'(t)}{f(t)}} dB_t \tag{1.3.14}
$$

where φ is the standard Gaussian density function.

By setting $Y_t = X_t - f(t)h(t)$, the dynamics of the supermartingale Z (which is given by $Z_t = \Phi(X_t - f(t)h(t))$) has the following form, which is its Doob-Meyer decomposition,

$$
dZ_t = -\varphi(Y_t)h'(t)f(t)dt + \varphi(Y_t)\sqrt{2\frac{f'(t)}{f(t)}} dB_t, \tag{1.3.15}
$$

where we have applied Itô's formula to $Z = \Phi(Y)$ and used (1.3.7).
We consider the multiplicative decomposition of Z, which is of the form $Z = Ne^{-H}$, where N is a positive local martingale and H an increasing predictable process. We have, by integration by parts, $dZ_t = -Z_t dH_t + e^{-H_t} dN_t$, hence, by uniqueness of the Doob-Meyer decomposition, we obtain

$$
dH_t = \frac{\varphi(Y_t)}{\Phi(Y_t)} h'(t)f(t)dt, \qquad dN_t = \frac{\varphi(Y_t)}{\Phi(Y_t)} N_t \sqrt{2\frac{f'(t)}{f(t)}} dB_t.
$$

Note that the quantity $\frac{\varphi(Y_t)}{\Phi(Y_t)} h'(t)f(t)$ is the intensity rate of τ.

1.3.3   Random times with supermartingale valued in $[0,1]$

Here we apply the result first introduced by Gapeev et al. (2010) and later by Jeanblanc and Song (2011a) and Jeanblanc and Song (2011b) to the Gaussian case.

It consists in starting with a given supermartingale $\widehat{Z}$ valued in $[0,1[$ with the multiplicative decomposition Ne^{-H}, where N is a non-negative local martingale and $dH_t = \lambda_t dt$, and associate to it a martingale survival process $(\widehat{M}_t(u))_{t,u\in\mathbb{R}^+}$, hence a random time which admits that martingale survival probability.

The process $(\widehat{M}_t(u))_{t,u\in\mathbb{R}^+}$, given by

$$\widehat{M}_t(u) = \begin{cases} 1 - (1 - \widehat{Z}_t)\exp\left(-\int_u^t \frac{\widehat{Z}_s}{1-\widehat{Z}_s}\lambda_s ds\right), & \text{for } u \leq t \\ \mathbb{E}[\widehat{Z}_u | \mathcal{F}_t], & \text{for } u > t, \end{cases}$$

is a survival martingale process and its dynamics for u fixed is given for, $t \geq u$, by

$$d\widehat{M}_t(u) = \frac{1 - \widehat{M}_t(u)}{1 - \widehat{Z}_t}e^{-H_t}dN_t. \tag{1.3.16}$$

Futhermore, if there exists ρ such that $\hat{M}_0(u) = \mathbb{E}[\widehat{Z}_u] = \int_u^\infty \rho(\theta)d\theta$, then ρ is the density function of τ, and the family $(\hat{p}_t(u))_{t,u\in\mathbb{R}^+}$, given by

$$\hat{p}_t(u) = \begin{cases} (1 - \widehat{M}_t(u))\frac{\widehat{Z}_u}{1-\widehat{Z}_u}\frac{\lambda_u}{\rho(u)}, & \text{for } u \leq t \\ \mathbb{E}[\hat{p}_u(u)|\mathcal{F}_t], & \text{for } u > t, \end{cases}$$

is a conditional density process associated with $M(u)$.

This result is first introduced by Gapeev et al. (2010) and later by Jeanblanc and Song (2011a) and Jeanblanc and Song (2011b).

We can apply this methodology to the Gaussian case with $\widehat{Z}_t = \Phi(Y_t)$ and $Y_t = \frac{\mu_t - l(t)}{\nu(t)}$.

We then obtain, for $u \leq t$,

$$\begin{aligned} d\widehat{M}_t(u) &= \frac{1 - \widehat{M}_t(u)}{1 - \Phi(Y_t)}e^{-H_t}\frac{\varphi(Y_t)}{\Phi(Y_t)}N_t\sqrt{2\frac{f'(t)}{f(t)}}dB_t \\ &= (1 - \widehat{M}_t(u))\frac{\varphi(Y_t)}{\Phi(-Y_t)}\sqrt{2\frac{f'(t)}{f(t)}}dB_t, \end{aligned}$$

where we have used the fact that $\Phi(-Y_t) = 1 - \Phi(Y_t)$. Note that $\widehat{M}(u)$ is different from $M(u)$ defined in the Gaussian case in (1.3.5).

Chapter 2

Generalization of the model of Jiao and Li

In the sovereign risk modeling, Jiao and Li (2018), by considering a reference filtration $\mathbb{F}$, propose the following form of the default time:

$$\tau = \theta \wedge \xi$$

where

$$\theta = \tau_i \quad \text{on} \quad \{N_{\tau_{i-1}} = 0\} \cap \{N_{\tau_i} = 1\}, \text{ for } i \in \{1, ..., n\}$$

with N an inhomogeneous Poisson process independent of $\mathbb{F}$, with deterministic intensity rate function $t \to \lambda^N(t)$ (i.e., a non-negative function λ^N such that $N_t - \int_0^t \lambda^N(s)ds$ is an $\mathbb{F}^N$-martingale, where $\mathbb{F}^N$ is the filtration generated by N), $(\tau_i)_{i \in \{1, ..., n\}}$ an increasing sequence of $\mathbb{F}$-predictable stopping times and

$$\xi := \inf\{t \geq 0 : \Gamma_t \geq \Theta\}$$

with Γ an increasing $\mathbb{F}$-adapted continuous process with $\Gamma_0 = 0$ and Θ a random variable, independent of $\mathbb{F}$ and $\mathbb{F}^N$, with a unit exponential law.

The authors are interested in some characteristics of the default time such as its compensator and its $\mathbb{F}$-conditional laws. They show that the intensity rate of τ does not exist in their setting.

Our goal is to generalize this model. We modify the role of the Poisson process N by using an increasing càdlàg process X such that $X_0 = 0$, $X_\infty = \infty$, independent of $\mathbb{F}$ and relax the assumption of the predictability of the stopping times $(\tau_i)_{i \geq 1}$. We investigate the characteristics of the default time. In particular, we compute the compensator of the default time which is more general than the one obtained by Jiao and Li (2018). As such, we consider $(\Omega, \mathcal{G}, \mathbb{F}, \mathbb{P})$ a filtered probability space. One denotes by Ψ the increasing deterministic function with $\Psi(0) = 0$ and $\Psi(\infty) = \infty$ such that $P(X_u \leq 1) = e^{-\Psi(u)}$. One denotes by $\mathbb{F}^X$ the natural filtration of X. Let $(\tau_i)_i$ be a strictly increasing sequence of finite $\mathbb{F}$-stopping times such that $\lim_{i \to \infty} \tau_i = \infty$ and set $\tau_0 = 0$. We define a random time [1] θ by

$$\theta = \tau_i \quad \text{on} \quad \{X_{\tau_{i-1}} \leq 1 < X_{\tau_i}\}, \text{ for } i \geq 1.$$

Note that $\theta < \infty$, a.s. since $\mathbb{P}(X_\infty \leq 1) = 0$.

We also define:

$$\xi := \inf\{t \geq 0 : \Gamma_t \geq \Theta\}$$

where Γ is an increasing $\mathbb{F}$-adapted continuous process with $\Gamma_0 = 0$, $\Gamma_t < \infty$ for all t, and $\Gamma_\infty = \infty$ and Θ a random variable, independent of $\mathbb{F}$ and $\mathbb{F}^X$, with a unit exponential law.

---

[1] The constant 1 in the definition can be changed in any constant, or even in any random variable independent of $(X, \mathbb{F})$ and of the random variable Θ.

We define
$$\tau = \theta \wedge \xi.$$

By construction, τ does not avoid the $\mathbb{F}$-stopping times $(\tau_i)_i$. By denoting $\mathbb{G}$ the smallest filtration containing $\mathbb{F}$ under which τ is a stopping time, one obtains the immersion property between $\mathbb{F}$ and $\mathbb{G}$. Indeed, knowing that $\mathbb{G} \subset \mathbb{F} \vee \sigma(\Theta) \vee \mathbb{F}^X$ and that $\mathbb{F}$ is immersed in $\mathbb{F} \vee \sigma(\Theta) \vee \mathbb{F}^X$ (by the independence of $\mathbb{F}$ from Θ and X) implies that $\mathbb{F}$ is immersed in $\mathbb{G}$ (see Lemma 7).

In what follows, we first compute the $\mathbb{F}$-conditional probability that the default time coincides with one of the $\mathbb{F}$-stopping times τ_i, before computing the $\mathbb{F}$-conditional survival process of τ. We mimic the proofs of Jiao and Li.

We shall frequently use the equality

$$\sum_{i \geq 1} \mathbb{1}_{\{\tau_i > t \geq \tau_{i-1}\}} e^{-U(\tau_{i-1})} = \exp\left(-\sum_{i \geq 1} \mathbb{1}_{\{\tau_i > t \geq \tau_{i-1}\}} U(\tau_{i-1}) \right), \tag{2.0.1}$$

for any function $U : \mathbb{R}^+ \to \mathbb{R}$. This is trivial since, for any i, on the set $\{\tau_i > t \geq \tau_{i-1}\}$ the left and right-hand sides are both equal to $e^{-U(\tau_{i-1})}$.

2.1 The $\mathbb{F}$ -conditional distributions of τ

Lemma 24 *For all $t \geq 0$, one has*

$$\mathbb{P}(\tau = \tau_i | \mathcal{F}_t) = \mathbb{E}\left[e^{-\Gamma_{\tau_i}} \left(e^{-\Psi(\tau_{i-1})} - e^{-\Psi(\tau_i)} \right) | \mathcal{F}_t \right], \ \forall i \geq 1. \tag{2.1.1}$$

In particular,

$$\mathbb{1}_{\{\tau_i \leq t\}} \mathbb{P}(\tau = \tau_i | \mathcal{F}_t) = e^{-\Gamma_{\tau_i}} \left(e^{-\Psi(\tau_{i-1})} - e^{-\Psi(\tau_i)} \right) \mathbb{1}_{\{\tau_i \leq t\}}, \ \forall i \geq 1. \tag{2.1.2}$$

PROOF: From the definition of τ, one has the following equality

$$\{\tau = \tau_i\} = \{\xi > \tau_i, X_{\tau_{i-1}} \leq 1 < X_{\tau_i}\}.$$

Hence, by using the definition of ξ as well as the fact that $(\tau_i)_i$ are $\mathbb{F}$-stopping times, hence are $\mathcal{F}_\infty$-measurable random variables and that the random variable X_{τ_i} is, for any i, $\mathcal{F}_\infty \vee \mathcal{F}_\infty^X$ measurable, one obtains

$$
\begin{aligned}
\mathbb{P}(\xi > \tau_i, X_{\tau_{i-1}} \leq 1 < X_{\tau_i} | \mathcal{F}_\infty) &= \mathbb{E}\left[\mathbb{P}(\xi > \tau_i | \mathcal{F}_\infty \vee \mathcal{F}_\infty^X) \mathbb{1}_{\{X_{\tau_{i-1}} \leq 1 < X_{\tau_i}\}} | \mathcal{F}_\infty \right] \\
&= \mathbb{E}\left[\mathbb{P}(\Gamma_{\tau_i} < \Theta | \mathcal{F}_\infty \vee \mathcal{F}_\infty^X) \mathbb{1}_{\{X_{\tau_{i-1}} \leq 1 < X_{\tau_i}\}} | \mathcal{F}_\infty \right]
\end{aligned}
$$

where the first equality requires the tower property.

Since Θ and X are mutually independent and are independent of $\mathcal{F}_\infty$, then using the fact that $\Gamma_{\tau_i} \in \mathcal{F}_\infty$ leads to

$$\mathbb{P}(\Gamma_{\tau_i} < \Theta | \mathcal{F}_\infty \vee \mathcal{F}_\infty^X) = \mathbb{P}(\Gamma_{\tau_i} < \Theta | \mathcal{F}_\infty) = e^{-\Gamma_{\tau_i}}. \tag{2.1.3}$$

Therefore, it follows

$$
\begin{aligned}
\mathbb{P}(\xi > \tau_i, X_{\tau_{i-1}} \leq 1 < X_{\tau_i} | \mathcal{F}_\infty) &= e^{-\Gamma_{\tau_i}} \mathbb{P}(X_{\tau_{i-1}} \leq 1 < X_{\tau_i} | \mathcal{F}_\infty) \\
&= e^{-\Gamma_{\tau_i}} \{ \mathbb{P}(X_{\tau_{i-1}} \leq 1 | \mathcal{F}_\infty) - \mathbb{P}(X_{\tau_i} \leq 1 | \mathcal{F}_\infty) \} \\
&= e^{-\Gamma_{\tau_i}} \left\{ e^{-\Psi(\tau_{i-1})} - e^{-\Psi(\tau_i)} \right\}. \tag{2.1.4}
\end{aligned}
$$

The last equality is due to the fact that the random variables τ_i are $\mathcal{F}_\infty$-measurable and the process X is independent of $\mathcal{F}_\infty$.

Therefore

$$\mathbb{P}(\tau = \tau_i | \mathcal{F}_t) = \mathbb{E}[e^{-\Gamma_{\tau_i}} \left\{ e^{-\Psi(\tau_{i-1})} - e^{-\Psi(\tau_i)} \right\} | \mathcal{F}_t].$$

Since the set $\{\tau_i \leq t\}$ is $\mathcal{F}_t$-measurable, by multiplying (2.1.4) by $\mathbb{1}_{\{\tau_i \leq t\}}$, one deduces that

$$\mathbb{1}_{\{\tau_i \leq t\}} \mathbb{P}(\tau = \tau_i | \mathcal{F}_t) = e^{-\Gamma_{\tau_i}} \left\{ e^{-\Psi(\tau_{i-1})} - e^{-\Psi(\tau_i)} \right\} \mathbb{1}_{\{\tau_i \leq t\}},$$

where, in the last equality, we have used the fact that, on the set $\{\tau_i \leq t\}$, the random variables Γ_{τ_i}, $\Psi(\tau_i)$ and $\Psi(\tau_{i-1})$ are $\mathcal{F}_t$-measurable.

$\square$

Lemma 25 *The conditional survival process of τ is given by, for $t \in \mathbb{R}^+$*

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = \mathbb{E}\left[\exp\left(-\sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq u\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})] - \Gamma_u \right) | \mathcal{F}_t \right] \ for \ u \ \in \mathbb{R}^+. \tag{2.1.5}$$

In particular, for $t \geq u$,

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = \exp\left(-\sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq u\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})] - \Gamma_u \right). \tag{2.1.6}$$

Then, the Azéma supermartingale Z has the following expression

$$Z_t = \exp\left(-\sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})] - \Gamma_t \right). \tag{2.1.7}$$

PROOF: For all $t, u \in \mathbb{R}^+$,

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = \sum_{i=1}^{\infty} \mathbb{P}(\theta > u, \xi > u, \theta = \tau_i | \mathcal{F}_t)$$

$$= \sum_{i=1}^{\infty} \mathbb{P}(\tau_i > u, \xi > u, X_{\tau_{i-1}} \leq 1 < X_{\tau_i} | \mathcal{F}_t).$$

This implies that

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = \sum_{i=1}^{\infty} \mathbb{P}(\tau_i > u, \xi > u, X_{\tau_{i-1}} \leq 1 | \mathcal{F}_t) - \sum_{i=1}^{\infty} \mathbb{P}(\tau_i > u, \xi > u, X_{\tau_i} \leq 1 | \mathcal{F}_t).$$

By using the fact that $\tau_0 = 0$ which implies that the set $\{\tau_0 > u\}$ is empty, one has

$$\mathbb{P}(\tau > u|\mathcal{F}_t) = \sum_{i=1}^{\infty} \mathbb{P}(\tau_i > u, \xi > u, X_{\tau_{i-1}} \le 1|\mathcal{F}_t) - \sum_{i=0}^{\infty} \mathbb{P}(\tau_i > u, \xi > u, X_{\tau_i} \le 1|\mathcal{F}_t)$$

$$= \sum_{i=1}^{\infty} \mathbb{P}(\tau_i > u, \xi > u, X_{\tau_{i-1}} \le 1|\mathcal{F}_t) - \sum_{k=1}^{\infty} \mathbb{P}(\tau_{k-1} > u, \xi > u, X_{\tau_{k-1}} \le 1|\mathcal{F}_t)$$

$$= \sum_{i=1}^{\infty} \mathbb{P}(\tau_i > u, \xi > u, X_{\tau_{i-1}} \le 1|\mathcal{F}_t) - \sum_{i=1}^{\infty} \mathbb{P}(\tau_{i-1} > u, \xi > u, X_{\tau_{i-1}} \le 1|\mathcal{F}_t).$$

Therefore, it follows from the trivial equality $\mathbb{1}_{\{\tau_i > u\}} - \mathbb{1}_{\{\tau_{i-1} > u\}} = \mathbb{1}_{\{\tau_i > u \ge \tau_{i-1}\}}$ that

$$\mathbb{P}(\tau > u|\mathcal{F}_t) = \sum_{i=1}^{\infty} \mathbb{P}(\tau_i > u \ge \tau_{i-1}, \xi > u, X_{\tau_{i-1}} \le 1|\mathcal{F}_t)$$

$$= \sum_{i=1}^{\infty} \mathbb{E}[\mathbb{1}_{\{\tau_i > u \ge \tau_{i-1}\}} \mathbb{P}(\xi > u|\mathcal{F}_{\infty} \vee \mathcal{F}_{\infty}^X) \mathbb{1}_{\{X_{\tau_{i-1}} \le 1\}}|\mathcal{F}_t]$$

$$= \sum_{i=1}^{\infty} \mathbb{E}[\mathbb{1}_{\{\tau_i > u \ge \tau_{i-1}\}} e^{-\Gamma_u} \mathbb{1}_{\{X_{\tau_{i-1}} \le 1\}}|\mathcal{F}_t],$$

where we have used, in the last equality, the fact that Θ is independent of $\mathbb{F} \vee \mathbb{F}^X$. This implies

$$\mathbb{P}(\tau > u|\mathcal{F}_t) = \sum_{i=1}^{\infty} \mathbb{E}[\mathbb{1}_{\{\tau_i > u \ge \tau_{i-1}\}} e^{-\Gamma_u} \mathbb{P}(X_{\tau_{i-1}} \le 1|\mathcal{F}_{\infty})|\mathcal{F}_t]$$

$$= \sum_{i=1}^{\infty} \mathbb{E}[\mathbb{1}_{\{\tau_i > u \ge \tau_{i-1}\}} e^{-\Gamma_u} e^{-\Psi(\tau_{i-1})}|\mathcal{F}_t]$$

$$= \mathbb{E}\left[\exp\left(-\sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i > u \ge \tau_{i-1}\}} \Psi(\tau_{i-1})\right) e^{-\Gamma_u}|\mathcal{F}_t\right]$$

$$= \mathbb{E}\left[\exp\left(-\sum_{i=1}^{\infty} (\mathbb{1}_{\{\tau_{i-1} \le u\}} - \mathbb{1}_{\{\tau_i \le u\}}) \Psi(\tau_{i-1})\right) e^{-\Gamma_u}|\mathcal{F}_t\right]$$

$$= \mathbb{E}\left[\exp\left(-\sum_{i=0}^{\infty} \mathbb{1}_{\{\tau_i \le u\}} \Psi(\tau_i) + \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \le u\}} \Psi(\tau_{i-1})\right) e^{-\Gamma_u}|\mathcal{F}_t\right].$$

Since $\Psi(\tau_0) = 0$, one has

$$\sum_{i=0}^{\infty} \mathbb{1}_{\{\tau_i \le u\}} \Psi(\tau_i) = \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \le u\}} \Psi(\tau_i).$$

Hence, it follows

$$\mathbb{P}(\tau > u|\mathcal{F}_t) = \mathbb{E}\left[\exp\left(-\sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \le u\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})]\right) e^{-\Gamma_u}|\mathcal{F}_t\right]$$

$$= \mathbb{E}\left[\exp\left(-\sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \le u\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})] - \Gamma_u\right)|\mathcal{F}_t\right].$$

If $t \ge u$, the random variables $e^{-\Gamma_u}$, $\mathbb{1}_{\{\tau_i \le u\}}$ and $\mathbb{1}_{\{\tau_i \le u\}}[\Psi(\tau_i) - \Psi(\tau_{i-1})]$ are $\mathcal{F}_t$-measurable. Hence, one has

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = \exp\left(-\sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq u\}}[\Psi(\tau_i) - \Psi(\tau_{i-1})] - \Gamma_u\right).$$

In particular, one checks that

$$\mathbb{P}(\tau > u | \mathcal{F}_u) = \mathbb{P}(\tau > u | \mathcal{F}_\infty),$$

which also follows from the immersion property. Note that Z is decreasing, which stems from the immersion property (it can also be seen from (2.1.7)). $\qquad\square$

2.2   The compensator of the default time τ

Now, we focus on the compensator of the default time τ. We first consider the case where the stopping times $(\tau_i)_i$ are predictable and then consider the general case.

2.2.1   Case where the stopping times $(\tau_i)_i$ are predictable

If the stopping times $(\tau_i)_i$ are $\mathbb{F}$-predictable, Z is $\mathbb{F}$-predictable and the $\mathbb{F}$-dual predictable projection of $A_t = \mathbb{1}_{\{\tau \leq t\}}$ is $1 - Z$. Hence, it is clear that $dA_t^p = -dZ_t$ and then using (1.2.1), the $\mathbb{F}$-predictable reduction Λ of the compensator of τ verifies (note that $Z_- > 0$):

$$d\Lambda_t = -\frac{dZ_t}{Z_{t-}}. \tag{2.2.1}$$

Since the Azéma supermartingale Z has the explicit form $Z_t = e^{-Y_t}$ where Y is the increasing process

$$Y_t = \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}}[\Psi(\tau_i) - \Psi(\tau_{i-1})] + \Gamma_t,$$

by applying Itô's formula for semimartingale (see Jeanblanc, Yor, and Chesney, 2009, Subsection 9.4.1) to e^{-Y}, we obtain

$$dZ_t = -Z_{t-}dY_t + Z_{t-}(e^{-\Delta Y_t} - 1) + Z_{t-}\Delta Y_t,$$

and by plugging it in (2.2.1), one obtains the following expression of the $\mathbb{F}$-predictable reduction of the compensator of τ

$$d\Lambda_t = dY_t - (e^{-\Delta Y_t} - 1) - \Delta Y_t = dY_t^c - (e^{-\Delta Y_t} - 1).$$

Hence, it follows that, since $\Lambda_0 = 0$

$$\Lambda_t = \Gamma_t + \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}}(1 - \exp(-[\Psi(\tau_i) - \Psi(\tau_{i-1})])), \text{ for all } t \geq 0. \tag{2.2.2}$$

2.2.2   Case where some of the stopping times $(\tau_i)_i$ are not predictable

If some stopping times $(\tau_i)_i$ are not predictable (which can occur only if the filtration $\mathbb{F}$ is not continuous [2]), one assumes to know, for any i, the $\mathbb{F}$-compensator of τ_i, denoted by J^i which is the $\mathbb{F}$-dual predictable

---

[2] A filtration $\mathbb{F}$ is continuous if all $\mathbb{F}$-martingales are continuous. In that case, every $\mathbb{F}$-stopping time is predictable (see, e.g., Nikeghbali, 2006, Theorem 4.11).

projection of the $\mathbb{F}$-adapted process $A^i = \mathbb{1}_{\{\tau_i \leq \cdot\}}$ (note that if τ_k is predictable, then $J^k = A^k$). In other terms, we consider the $\mathbb{F}$-predictable increasing process J^i, with $J_0^i = 0$, such that

$$(\mathbb{1}_{\{\tau_i \leq t\}} - J_{t \wedge \tau_i}^i, t \geq 0)$$

is an $\mathbb{F}$-martingale. In what follows, we give the $\mathbb{F}$-dual predictable projection process A^p as well as the $\mathbb{F}$-predictable reduction of the compensator of τ. We first introduce the following Lemma.

Lemma 26 *Consider, for $i \geq 1$, the two $\mathbb{F}$-adapted increasing processes $Q_t^{1,i} := \mathbb{1}_{\{\tau_i \leq t\}} e^{-\Psi(\tau_i)}$ and $Q_t^{2,i} :=$ $\mathbb{1}_{\{\tau_{i+1} \leq t\}} e^{-\Psi(\tau_i)}$. Then the processes defined as*

$$Q_t^{1,i} - \int_0^t e^{-\Psi(s)} dJ_s^i \tag{2.2.3}$$

and

$$Q_t^{2,i} - \int_0^t \mathbb{1}_{\{\tau_i < s\}} e^{-\Psi(\tau_i)} dJ_s^{i+1} \tag{2.2.4}$$

are $\mathbb{F}$-martingales.

PROOF: By applying Lemma 11 for the process A^i, the predictable process H being the function $H_t = e^{-\Psi(t)}$, it follows that $Q_t^{1,i} - \int_0^t e^{-\Psi(s)} dJ_s^i$ is an $\mathbb{F}$-martingale.
The process $Q^{2,i}$ which is $Q_t^{2,i} = \int_0^t e^{-\Psi(\tau_i)} dA_s^{i+1}$ can be written as

$$Q_t^{2,i} = \int_0^t e^{-\Psi(\tau_i)} \mathbb{1}_{\{\tau_i < s\}} dA_s^{i+1}.$$

Indeed, due to the strictly increasing property of the τ_i,

$$
\begin{aligned}
\int_0^t e^{-\Psi(\tau_i)} \mathbb{1}_{\{\tau_i < s\}} dA_s^{i+1} &= e^{-\Psi(\tau_i)} \int_0^t \mathbb{1}_{\{\tau_i < s\}} dA_s^{i+1} \\
&= e^{-\Psi(\tau_i)} \mathbb{1}_{\{\tau_i < \tau_{i+1}\}} \mathbb{1}_{\{\tau_{i+1} \leq t\}} = e^{-\Psi(\tau_i)} \mathbb{1}_{\{\tau_{i+1} \leq t\}} = Q_t^{2,i}.
\end{aligned}
$$

By setting $Y_s = \mathbb{1}_{\{\tau_i < s\}} e^{-\Psi(\tau_i)}$, which is left-continuous and adapted hence predictable, and by applying Lemma 11, the process

$$Y_{\tau_{i+1}} \mathbb{1}_{\{\tau_{i+1} \leq t\}} - \int_0^{t \wedge \tau_{i+1}} Y_s dJ_s^{i+1}$$

is an $\mathbb{F}$-martingale, or equivalently, the process

$$
\begin{aligned}
\mathbb{1}_{\{\tau_i < \tau_{i+1}\}} e^{-\Psi(\tau_i)} \mathbb{1}_{\{\tau_{i+1} \leq t\}} - \int_0^{t \wedge \tau_{i+1}} Y_s dJ_s^{i+1} &= Q_t^{2,i} - \int_0^{t \wedge \tau_{i+1}} Y_s dJ_s^{i+1} \\
&= Q_t^{2,i} - \int_0^t \mathbb{1}_{\{\tau_i < s\}} e^{-\Psi(\tau_i)} dJ_s^{i+1}
\end{aligned}
$$

is an $\mathbb{F}$-martingale. $\qquad\square$

Lemma 27 *The $\mathbb{F}$-dual predictable projection A^p of the default time τ verifies:*

$$dA_t^p = Z_t d\Gamma_t + e^{-\Gamma_t} d\zeta_t, \tag{2.2.5}$$

where ζ is the predictable increasing process defined as

$$\zeta_t := \sum_{i=0}^{\infty} \int_0^t \left(e^{-\Psi(\tau_i)} - e^{-\Psi(s)} \right) dJ_s^{i+1}. \tag{2.2.6}$$

PROOF: The process

$$Q_t := \exp\left(-\sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}}[\Psi(\tau_i) - \Psi(\tau_{i-1})]\right) \tag{2.2.7}$$

is decreasing, hence is a supermartingale. Using (2.0.1), we get

$$Q_t = \sum_{i=0}^{\infty} \mathbb{1}_{\{\tau_i \leq t < \tau_{i+1}\}} e^{-\Psi(\tau_i)} = \sum_{i=0}^{\infty} \left(\mathbb{1}_{\{\tau_i \leq t\}} e^{-\Psi(\tau_i)} - \mathbb{1}_{\{\tau_{i+1} \leq t\}} e^{-\Psi(\tau_i)}\right) = \sum_{i=0}^{\infty} \left(Q_t^{1,i} - Q_t^{2,i}\right)$$

where the processes $Q^{1,i}$ and $Q^{2,i}$ are the ones defined in Lemma 26. Therefore, from Lemma 26, it admits the decomposition $Q = m^Q - \zeta$, where m^Q is a martingale and ζ the predictable process given as

$$\zeta_t = \sum_{i=0}^{\infty} \left(\int_0^t \mathbb{1}_{\{\tau_i < s\}} e^{-\psi(\tau_i)} dJ_s^{i+1} - \int_0^t e^{-\psi(s)} dJ_s^i\right).$$

Since $M^{i+1} = A^{i+1} - J^{i+1}$ is a martingale with $M_0^{i+1} = 0$ and $A_{\tau_i}^{i+1} = 0$, we obtain

$$\mathbb{E}[J_{\tau_i}^{i+1}] = \mathbb{E}[A_{\tau_i}^{i+1} - M_{\tau_i}^{i+1}] = -\mathbb{E}[M_{\tau_i}^{i+1}] = 0$$

which implies that $J_{\tau_i}^{i+1} = 0$ and then due to the increasing property of J^{i+1}, $J_t^{i+1} = 0$ on $\{t \leq \tau_i\}$. This shows that the support of J^{i+1} is $[\tau_i, \tau_{i+1}]$.

Hence, it follows that $\int_0^t \mathbb{1}_{\{\tau_i < s\}} e^{-\Psi(\tau_i)} dJ_s^{i+1} = \int_0^t e^{-\Psi(\tau_i)} dJ_s^{i+1}$, and then

$$\zeta_t = \sum_{i=0}^{\infty} \int_0^t e^{-\psi(\tau_i)} dJ_s^{i+1} - \sum_{i=1}^{\infty} \int_0^t e^{-\psi(s)} dJ_s^i = \sum_{i=0}^{\infty} \int_0^t \left(e^{-\psi(\tau_i)} - e^{-\psi(s)}\right) dJ_s^{i+1}$$

where we have used the fact that $J^0 = 0$.

Due to the form of the support of J^{i+1}, we have $\zeta_t = \sum_{i=0}^{\infty} \int_0^t \mathbb{1}_{\{s < \tau_i\}} \left(e^{-\psi(\tau_i)} - e^{-\psi(s)}\right) dJ_s^{i+1}$, hence, since $e^{-\psi(\tau_i)} - e^{-\psi(s)} \geq 0$ for $s > \tau_i$, the process ζ is increasing.

By integration by parts of the multiplicative decomposition of the supermartingale Z, which is (from (2.1.7))

$$Z_t = e^{-\Gamma_t} Q_t = e^{-\Gamma_t}(m_t^Q - \zeta_t),$$

and by applying Yoeurp's lemma, we get

$$\begin{aligned} dZ_t = & -e^{-\Gamma_t}(m_t^Q - \zeta_t)d\Gamma_t + e^{-\Gamma_t}(dm_t^Q - d\zeta_t) \\ = & e^{-\Gamma_t} dm_t^Q - e^{-\Gamma_t}(m_t^Q - \zeta_t)d\Gamma_t - e^{-\Gamma_t} d\zeta_t \\ = & e^{-\Gamma_t} dm_t^Q - (Z_t d\Gamma_t + e^{-\Gamma_t} d\zeta_t) \end{aligned}$$

which leads to the Doob-Meyer decomposition of Z. Hence the $\mathbb{F}$-dual predictable projection A^p of τ verifies $dA_t^p = Z_t d\Gamma_t + e^{-\Gamma_t} d\zeta_t$. $\qquad \square$

Finally, using 1.2.1 and the fact that Γ is continuous, the $\mathbb{F}$-predictable reduction Λ of the compensator of τ is given by

$$d\Lambda_t = \frac{1}{Z_{t-}} dA_t^p = d\Gamma_t + \frac{d\zeta_t}{Q_{t-}}. \tag{2.2.8}$$

Remark 28 *Note that if all the stopping times $(\tau_i)_{i \geq 1}$ are predictable, then the supermartingale Q is predictable and $Q_t = 1 - \zeta_t$, hence $dQ_t = -d\zeta_t$. Therefore, by noting that $\frac{dQ_t}{Q_{t-}} = e^{-\Delta K_t} - 1$, (2.2.8) becomes*

$$d\Lambda_t = d\Gamma_t - \frac{dQ_t}{Q_{t-}} = d\Gamma_t + 1 - e^{-\Delta K_t}$$

and which is the dynamic form of (2.2.2).

As an example, assume that the $(\tau_i)_{i \geq 1}$ are the jump times of a Poisson process N with constant intensity λ which generates the filtration $\mathbb{F}$. Then

$$dJ_t^{i+1} = \lambda \mathbb{1}_{[\tau_i, \tau_{i+1}]}(t)dt.$$

Hence from Lemma 27, one has

$$\zeta_t = \int_0^t \left(\lambda \sum_{i=0}^{\infty} \mathbb{1}_{[\tau_i, \tau_{i+1}]}(s) \left(e^{-\Psi(\tau_i)} - e^{-\Psi(s)} \right) \right) ds$$

which is continuous. Therefore, if $\Gamma_t = \int_0^t \gamma_s ds$, with γ a non-negative $\mathbb{F}$-adapted process, one has

$$d\Lambda_t = \left(\gamma_t + \frac{\lambda \sum_{i=0}^{\infty} \mathbb{1}_{[\tau_i, \tau_{i+1}]}(t) \left(e^{-\Psi(\tau_i)} - e^{-\Psi(t)} \right)}{Q_{t-}} \right) dt \tag{2.2.9}$$

where Q is given in (2.2.7). This implies that the intensity rate of the default time exists and is given at time t by

$$\gamma_t + \frac{\lambda \sum_{i=0}^{\infty} \mathbb{1}_{[\tau_i, \tau_{i+1}]}(t) \left(e^{-\Psi(\tau_i)} - e^{-\Psi(t)} \right)}{Q_{t-}}.$$

Chapter 3

Generalized Cox model

This chapter is based on the working paper Gueye, Jeanblanc, and Li (2019).

3.1 Introduction and motivations

In the works of default time modeling through Cox framework (see Lando, 1998), one usually starts with a reference filtration $\mathbb{F}$ and an $\mathbb{F}$-adapted process Λ (absolutely continuous with respect to Lebesgue's measure and increasing). Then one defines the default time as the first time this process hits a stochastic barrier which is assumed to be independent of $\mathbb{F}$. Hence one establishes that the process Λ is the $\mathbb{F}$-predictable reduction of the compensator of τ, and shows that it can be calibrated, for instance from the credit default swap spread, in order to compute the prices of products subject to default risk. In this setting, the default time avoids $\mathbb{F}$-stopping times and this does not allow to cover a wide range of modeling of these products, for example in the case which induces some economic shocks. The model of Bélanger, Shreve, and Wong (2004) generalizes that of Lando (1998), but covers only the setting where the shock times are predictable. The model of Gehmlich and Schmidt (2018), which is also the one of Fontana and Schmidt (2018), is not a model of enlargement of filtration and presents an interesting case where the compensator of τ admits predictable jumps. In some sense, it is a structural model in which the default time is supposed to have nice properties, and there is no construction of such a default time. We shall extend the construction given in Bélanger, Shreve, and Wong (2004) to a more general framework, constructing a default time which does not avoid a given sequence of $\mathbb{F}$-stopping times.

We start from a right-continuous with left limits (or a left-continuous with right limits) increasing process K, adapted with respect to a given reference filtration $\mathbb{F}$, not necessarily predictable and which jumps at some given $\mathbb{F}$-stopping times and we define the default time τ as in Cox's paradigm, i.e., τ is the first hitting time of this process K of a stochastic barrier independent of the reference filtration. As before, $\mathbb{G}$ is the filtration $\mathbb{F}$ progressively enlarged with τ. In both cases, we compute the characteristics of the default time, i.e., the $\mathbb{F}$-predictable reduction of the compensator of τ, the Azéma supermartingale and its multiplicative decomposition and the conditional survival process. We also study existence of conditional densities (i.e., the (CL), (ED) and (GD) conditions).

In what follows, we first present our model before illustrating our construction through many examples.

We consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, endowed with a filtration $\mathbb{F}$ satisfying usual hypotheses and a random time τ. We consider an increasing $\mathbb{F}$-adapted process K with $K_0 = 0$ and Θ a unit exponential r.v. independent of $\mathbb{F}$. Then, if K is right-continuous with left limits (càdlàg), we define the random time τ as

$$\tau = \inf\{t \geq 0 : K_t \geq \Theta\},$$

and, if K is left-continuous with right limits (càglàd), one modifies a little bit the definition by setting (we shall see why later)

$$\tau = \inf\{t \geq 0 : K_t > \Theta\}.$$

In both cases, immersion holds between $\mathbb{F}$ and $\mathbb{G}$, since $\mathbb{F}$ is immersed in $\mathbb{F} \vee \sigma(\Theta)$ and $\mathbb{F} \subset \mathbb{G} \subset \mathbb{F} \vee \sigma(\Theta)$. This property implies that $A^o = 1 - Z$ (see Aksamit and Li, 2016, Th. 6), therefore, since $\tilde{Z} = Z - \Delta A^o$, then $\tilde{Z} = Z_-$, where as in the previous chapters A^o is the $\mathbb{F}$-optional projection of the process $A = \mathbb{1}_{\{\tau \leq \cdot\}}$ and $Z, \tilde{Z}$ the Azéma supermartingales.

3.2 Case where K is continuous

When the process K is continuous, we define $\tau = \inf\{t : K_t \geq \Theta\}$ and due to $\{K_t < \Theta\} = \{\tau > t\}$, we obtain immediately that $Z = e^{-K} = 1 - A^o = 1 - A^p$. From the fact that A^o is continuous, we obtain $\Delta A^o = 0$, hence $\tilde{Z} = Z$. The Doob-Meyer decomposition of Z is $Z_t = 1 - (1 - e^{-K_t})$, for all $t \geq 0$. In that setting, A^o being continuous, τ avoids $\mathbb{F}$-stopping times (see Aksamit and Jeanblanc, 2017, Proposition 1.43). The survival probability of τ is $\mathbb{P}(\tau > u) = \mathbb{E}[e^{-K_u}]$, so that for any bounded Borel function h, $\mathbb{E}[h(\tau)] = \mathbb{E}\big[\int_{\mathbb{R}^+} h(u)e^{-K_u}dK_u\big]$. The conditional survival process is

$$
\begin{aligned}
\mathbb{P}(\tau > u|\mathcal{F}_t) &= Z_u, \quad \text{for } u < t \\
&= \mathbb{E}[Z_u|\mathcal{F}_t], \quad \text{for } t \leq u,
\end{aligned}
$$

where the first equality is due to immersion property. If K is absolutely continuous w.r.t. Lebesgue's measure, i.e., $K_t = \int_0^t k_u du$, then, τ admits a density $f(u) = \mathbb{E}[k_u e^{-K_u}]$ and a (CL)-conditional density (with $\eta(du) = f(u)du$) given by

$$
\begin{aligned}
p_t(u) &= \frac{k_u e^{-K_u}}{\mathbb{E}[k_u e^{-K_u}]}, \quad \text{for } u < t, \\
&= \frac{\mathbb{E}[k_u e^{-K_u}|\mathcal{F}_t]}{\mathbb{E}[k_u e^{-K_u}]}, \quad \text{for } t \leq u,
\end{aligned}
$$

if $\mathbb{E}[k_u e^{-K_u}] > 0$ and $p_t(u) = 0$ if $\mathbb{E}[k_u e^{-K_u}] = 0$.

In general, the (CL) conditional density may fail to exist. For example, let $K_t = -\ln(1 - L_{t \wedge 1}) + \mathbb{1}_{\{t > 1\}}(t - 1)$ where L is the local time at level 0 of a standard Brownian motion. Then $\mathbb{P}(\tau > t) = \mathbb{E}[e^{-K_t}] = \mathbb{E}[1 - L_{t \wedge 1}]e^{-(t-1)^+}$. The Tanaka formula (see Øksendal, 2003, Chap 4, p59), i.e.,

$$|W_t| = \int_0^t \text{sgn}(W_s)dW_s + L_t,$$

where sgn is the sign function, i.e.,

$$
\text{sgn}(x) = \begin{cases} +1, & \text{if } x \geq 0 \\ -1, & \text{if } x < 0 \end{cases}
$$

implies (the stochastic integral being a martingale) that $\mathbb{E}[L_t] = \mathbb{E}[|W_t|] = \frac{\sqrt{2t}}{\sqrt{\pi}}$. Therefore,

$$
\mathbb{P}(\tau > t) = \begin{cases} 1 - \frac{\sqrt{2t}}{\sqrt{\pi}}, & \text{if } t \leq 1 \\ (1 - \frac{\sqrt{2}}{\sqrt{\pi}})e^{-t+1}, & \text{if } t > 1 \, . \end{cases}
$$

We deduce that τ has a density f w.r.t. Lebesgue's measure, where

$$f(t) = \begin{cases} \frac{1}{\sqrt{2t\pi}}, & \text{if } t \leq 1 \\ (1 - \frac{\sqrt{2}}{\sqrt{\pi}})e^{-t+1}, & \text{if } t > 1 . \end{cases}$$

Therefore, if the classical density hypothesis is satisfied, then by choosing $\eta(ds) = f(s)ds$, one would have, for $u < t$

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = Z_u = \int_u^\infty p_t(s)f(s)ds$$

and Z would be absolutely continuous w.r.t. Lebesgue's measure, which is not the case.

3.3   Case where K is càdlàg

We denote by $(\tau_i)_i$ the jump times of K.

Lemma 29 *For all $t \geq 0$, the following equality*

$$\{K_t \geq \Theta\} = \{\tau \leq t\} \tag{3.3.1}$$

holds, equivalently $\{K_t < \Theta\} = \{t < \tau\}$.

PROOF: In the proof, we proceed for ω fixed. The set $\{K_t \geq \Theta\}$ is either of the form $[a, \infty[$ or of the form $]a, \infty[$. If it is of the form $[a, \infty[$ it would mean that $K_a \geq \Theta$ and by definition of τ, $\tau = a$. If it is of the form $]a, \infty[$, it means that $K_{a+\varepsilon} \geq \Theta$, for any $\varepsilon > 0$, hence by right-continuity of K, letting ε go to 0, it follows that $K_a \geq \Theta$ and $a \in \{t \geq 0 : K_t \geq \Theta\}$ which is contradictory. $\square$

3.3.1   Computation of the dual predictable projection of τ

Lemma 30 *Let I be the càdlàg $\mathbb{F}$-submartingale $I_t = \sum_{s \leq t}(1 - e^{-\Delta K_s})$ and A^I be its predictable part in its Doob-Meyer decomposition. Then the $\mathbb{F}$-dual predictable projection of τ satisfies $dA_t^p = e^{-K_{t-}}(dK_t^c + dA_t^I)$, where K^c is the continuous part of K and the $\mathbb{F}$-predictable reduction of the compensator of τ is*

$$\Lambda = K^c + A^I . \tag{3.3.2}$$

PROOF: By using the equality (3.3.1), the fact that K is $\mathbb{F}$-adapted and the independence between Θ and $\mathbb{F}$, one has from the exponential law of Θ

$$Z_t := \mathbb{P}(\tau > t | \mathcal{F}_t) = \mathbb{P}(K_t < \Theta | \mathcal{F}_t) = e^{-K_t} .$$

Then, by Itô's formula, it follows that

$$\begin{aligned} dZ_t &= -e^{-K_{t-}}dK_t + e^{-K_{t-}}(e^{-\Delta K_t} - 1) + e^{-K_{t-}}\Delta K_t \\ &= -e^{-K_{t-}}dK_t^c + e^{-K_{t-}}(e^{-\Delta K_t} - 1) \end{aligned} \tag{3.3.3}$$

where K^c is the continuous part of K, i.e., $K_t^c = K_t - \sum_{s \leq t}\Delta K_s$.
The process K^c being increasing and continuous is a submartingale with Doob-Meyer's decomposition with no martingale part, i.e., $K_t^c = 0 + K_t^c$. The process I being increasing (indeed $1 - e^{-\Delta K_s} \geq 0$) is a

submartingale and admits a Doob-Meyer decomposition $I = M^I + A^I$. Finally

$$
\begin{aligned}
dZ_t &= -e^{-K_{t-}} dK_t^c - e^{-K_{t-}} dI_t \\
&= -e^{-K_{t-}}(dK_t^c + dM_t^I + dA_t^I) \\
&= -e^{-K_{t-}} dM_t^I - e^{-K_{t-}}(dK_t^c + dA_t^I).
\end{aligned}
\tag{3.3.4}
$$

Therefore

$$
dA_t^p = e^{-K_{t-}}(dK_t^c + dA_t^I)
\tag{3.3.5}
$$

and $A_t - \Lambda_{t \wedge \tau}$ is a $\mathbb{G}$-martingale where

$$
d\Lambda_t = \frac{1}{Z_{t-}} dA_t^p = dK_t^c + dA_t^I
$$

or $\Lambda_t = K_t^c + A_t^I$.

<div align="right">□</div>

3.3.2   The conditional laws of the default time τ

Proposition 31 *The conditional probability that the default occurs at time τ_i is given by*

$$
p_t^i := \mathbb{P}(\tau = \tau_i | \mathcal{F}_t) = \mathbb{E}[e^{-K_{\tau_i}-}(1 - e^{-\Delta K_{\tau_i}}) | \mathcal{F}_t], \ \forall t \geq 0.
\tag{3.3.6}
$$

PROOF: From (3.3.1), one has $\{s < \tau \leq t\} = \{K_s < \Theta \leq K_t\}$ and for any $\mathbb{F}$-stopping time ϑ, one obtains, denoting $\vartheta(\epsilon) = (\vartheta - \epsilon) \vee 0$, that $\{\vartheta(\epsilon) < \tau \leq \vartheta\} = \{K_{\vartheta(\epsilon)} < \Theta \leq K_\vartheta\}$ and

$$
\mathbb{P}(\vartheta(\epsilon) < \tau \leq \vartheta | \mathcal{F}_t) = \mathbb{E}[e^{-K_{\vartheta(\epsilon)}} - e^{-K_\vartheta} | \mathcal{F}_t].
$$

Passing to the limit when ϵ goes to 0, one obtains

$$
\mathbb{P}(\tau = \vartheta | \mathcal{F}_t) = \mathbb{E}[e^{-K_{\vartheta-}} - e^{-K_\vartheta} | \mathcal{F}_t].
$$

In particular,

$$
p_t^i := \mathbb{P}(\tau = \tau_i | \mathcal{F}_t) = \mathbb{E}[e^{-K_{\tau_i}-}(1 - e^{-\Delta K_{\tau_i}}) | \mathcal{F}_t].
$$

<div align="right">□</div>

Note that on the set $\{\tau_i \leq t\}$, one has $p_t^i = e^{-K_{\tau_i}-}(1 - e^{-\Delta K_{\tau_i}})$ and the random variable p_t^i does not depend on t, as it must be, from immersion property (see El Karoui, Jeanblanc, and Jiao, 2010, Subsection 3.2.).

Proposition 32 *If the continuous part of K is absolutely continuous, i.e., $K_t^c = \int_0^t k_s ds$, the generalized density hypothesis (GD) is satisfied, and $\alpha_t(u) = \mathbb{E}[k_u e^{-K_u} | \mathcal{F}_t]$ together with υ the Lebesgue's measure.*

PROOF: For any bounded Borel function h, the process

$$
X_t = h(t) \prod_i \mathbb{1}_{\{t \neq \tau_i\}} = h(t) \prod_i (\mathbb{1}_{\{t < \tau_i\}} + \mathbb{1}_{\{t > \tau_i\}})
$$

is $\mathbb{F}$-optional (Indeed, $\mathbb{1}_{\{\tau_i > \cdot\}} = 1 - \mathbb{1}_{\{\tau_i \leq \cdot\}}$ is right-continuous and $\mathbb{F}$-adapted, hence $\mathbb{F}$-optional and $\mathbb{1}_{\{\tau_i < \cdot\}}$ is left-continuous and $\mathbb{F}$-adapted hence $\mathbb{F}$-predictable therefore $\mathbb{F}$-optional).   Hence,

$$
\begin{aligned}
\mathbb{E}[h(\tau) \prod_i \mathbb{1}_{\{\tau \neq \tau_i\}} | \mathcal{F}_t] &= \mathbb{E}[X_\tau | \mathcal{F}_t] = \mathbb{E}[\int_0^\infty X_s dA_s^o | \mathcal{F}_t] = -\mathbb{E}[\int_0^\infty X_s dZ_s^c | \mathcal{F}_t] \\
&= \mathbb{E}[\int_0^\infty h(s) k_s e^{-K_s} ds | \mathcal{F}_t] = \int_0^\infty h(s) \mathbb{E}[k_s e^{-K_s} | \mathcal{F}_t] ds \,.
\end{aligned}
$$

The second equality is due to Aksamit and Jeanblanc, 2017, Corollary 2.10, the third equality is due to the fact that X vanishes on the discontinuities of Z, so that $\int_0^\infty X_s dZ_s = \int_0^\infty X_s dZ_s^c$, then we use the fact that, from (3.3.3) $Z_t^c = 1 - \int_0^t e^{-K_{s-}} dK_s^c = 1 - \int_0^t e^{-K_s} dK_s^c = 1 - \int_0^t e^{-K_s} k_s ds$. $\qquad \square$

Note that, in the case where the $(\tau_i)_i$ are ordered, the same kind of computation leads to

$$
\mathbb{E}\left[h(\tau) \mathbb{1}_{\{\tau_i < \tau < \tau_{i+1}\}} | \mathcal{F}_t \right] = \mathbb{E}\left[\int_0^\infty \mathbb{1}_{\{\tau_i < s < \tau_{i+1}\}} h(s) k_s e^{-K_s} ds | \mathcal{F}_t \right] \,.
$$

That is to consider $X_t = h(t) \mathbb{1}_{\{\tau_i < t < \tau_{i+1}\}} = h(t)(\mathbb{1}_{\{\tau_i < t\}} - \mathbb{1}_{\{\tau_{i+1} \leq t\}})$ which is $\mathbb{F}$-optional and then since X vanishes on the discontinuities of Z, it follows that

$$
\mathbb{E}[h(\tau) \mathbb{1}_{\{\tau_i < \tau < \tau_{i+1}\}} | \mathcal{F}_t] = \mathbb{E}[X_\tau | \mathcal{F}_t] = -\mathbb{E}[\int_0^\infty X_s dZ_s^c | \mathcal{F}_t] \,.
$$

 The classical density hypothesis (CL) may not be satisfied if the continuous part of K is not absolutely continuous (see, for example, the given counterexample in the case where K is continuous).

The extended density hypothesis (ED) is satisfied: it suffices to note that for $u < t$, $\mathbb{P}(\tau \leq u | \mathcal{F}_t) = 1 - e^{-K_u} = \int_0^u dD_s$ with $D_s = 1 - e^{-K_s}$ and $m_t \equiv 1$.

3.3.3   The multiplicative decomposition of Z

Lemma 33 *The multiplicative decomposition of Z is*[1]

$$
\mathcal{E}(Y)_t \mathcal{E}(-\Lambda)_t = \mathcal{E}(Y)_t e^{-\Lambda_t} \prod_{s \leq t} (1 - \Delta A_s^I) e^{\Delta A_s^I}
$$

where $\mathcal{E}(-\Lambda)$ is the decreasing $\mathbb{F}$-predictable part and $\mathcal{E}(Y)$ is the $\mathbb{F}$-local martingale part, with $Y_t = \int_0^t \dfrac{1}{Z_{s-}(1 - \Delta A_s^I)} dM_s$, where M being the martingale part of Z in its Doob-Meyer decomposition and Λ being the $\mathbb{F}$-predictable reduction of the compensator of τ given in (3.3.2).

 PROOF: This result is established in full generality in Jeanblanc and Li (2020), where the authors "guess" this form and check it. However, it seems interesting to recover that result without guessing this decomposition, and we present a proof. By using the multiplicative decomposition of Z as $Z = Ne^{-H}$, where H is a predictable càdlàg increasing process and N a local martingale and by applying the integration by parts formula using Yoeurp's lemma, we obtain

$$
dZ_t = e^{-H_t} dN_t + N_{t-} de^{-H_t} \,.
$$

---

[1]We recall that the Doléans-Dade exponential of a càdlàg semimartingale X is

$$
\mathcal{E}(X)_t = e^{X_t - \frac{1}{2}\langle X^c, X^c \rangle_t} \prod_{0 < s \leq t} (1 + \Delta X_s) e^{-\Delta X_s} \,,
$$

where X^c is the continuous martingale part of X.

By identifying this decomposition of Z and the DM decomposition of Z given in (3.3.4), one gets

$$e^{-H}dN = dM, \quad N_-de^{-H} = -Z_-(dK^c + dA^I),$$

hence $dH^c + 1 - e^{-\Delta H} = dK^c + dA^I$ which leads to

$$dH^c = dK^c + dA^{I,c} \text{ and } 1 - e^{-\Delta H} = \Delta A^I.$$

Finally, using the fact that $\Lambda = K^c + A^I = K^c + A^{I,c} + \sum_{s \le \cdot} \Delta A_s^I$, we obtain

$$e^{-H_t} = e^{-H_t^c} e^{-\sum_{s \le t} \Delta H_s} = e^{-\Lambda_t} \prod_{s \le t} (1 - \Delta A_s^I) e^{\Delta A_s^I}. \tag{3.3.7}$$

The equality $dN = e^H dM$ can be written as $dN = N_- Y dM$ with $Y = \frac{e^H}{N_-} = \frac{e^{\Delta H}}{Z_-} = \frac{1}{Z_-(1-\Delta A^I)}$, hence, using the fact that $\Lambda = K^c + A^I = K^c + A^{I,c} + \sum_{s \le \cdot} \Delta A_s^I = H^c + \sum_{s \le \cdot} \Delta A_s^I$, the proof is done. $\qquad \square$

One can check that $1 - \Delta A^I \ge 0$. This follows from (3.3.7). Another proof can be given from $\Delta A^I = {}^p(\Delta I) = {}^p(1 - e^{-\Delta K}) = 1 - {}^p(e^{-\Delta K}) \le 1$, where the first equality comes from Aksamit and Jeanblanc, 2017, Proposition 1.36 b. It follows that

$$\Delta \Lambda \le 1. \tag{3.3.8}$$

One can also check that $1 + \Delta N > 0$. Indeed,

$$\Delta N = -e^{-(K_- + H)} \Delta M^I = -e^{-(K_- + H)}(1 - e^{-\Delta K} - \Delta A^I) > -1.$$

3.3.4 Particular case: K is càdlàg $\mathbb{F}$-predictable

When the process K is càdlàg $\mathbb{F}$-predictable, then $Z = 1 - A^p = 1 - A^o$. Furthermore, only the jump times of K, which are all predictable, are $\mathbb{F}$-stopping times not avoided by τ. The same model was presented in Bélanger, Shreve, and Wong (2004), in a slightly more general case when Θ is not a unit exponential random variable [2] (but is still independent of $\mathbb{F}$). We assume that

$$K_t = \int_0^t k_s ds + \sum_{i \ge 1} \mathbb{1}_{\{\tau_i \le t\}} \theta_i$$

where $(\tau_i)_i$ is a strictly increasing sequence of $\mathbb{F}$-predictable stopping times, k is an $\mathbb{F}$-adapted non-negative process with $\int_0^t k_s ds < \infty$, $\forall t \ge 0$ and $(\theta_i)_i$ a sequence of non-negative random variables with $\theta_i \in \mathcal{F}_{\tau_i-}$ such that $\sum_{i=1}^\infty \theta_i < \infty$. Note that $I = \sum_i \mathbb{1}_{\{\tau_i \le \cdot\}}(1 - e^{-\theta_i})$ is $\mathbb{F}$-predictable and then $A^I = I$ and $M^I = 0$.

3.3.5 Examples

In this subsection, we give some examples when the process K is càdlàg.

3.3.5.1 Brownian filtration

We construct a simple example where K is càdlàg (but not continuous) and predictable. Let $\mathbb{F}$ be a Brownian filtration, $(T_n, n \ge 1)$ an increasing sequence of $\mathbb{F}$-stopping times (e.g., $T_n = \inf\{t : W_t = n\}$) and define $K_t = \int_0^t k_s ds + \sum_n \mathbb{1}_{\{T_n \le t\}} \theta_n$ where $(\theta_n)_{n \ge 1}$ is a sequence of non-negative random variables with $\theta_n \in$

---

[2] There is no matter if Θ is not an exponential random variable, since one can always reduce attention to the case with exponential one by modifying K, thanks to the inverse transform sampling theorem, at least if Θ is a random variable with strictly monotone cumulative distribution function. The case where Θ is not independent of $\mathbb{F}$ can be studied using the $\mathbb{F}$-conditional law of Θ (see El Karoui et al., 2014).

$\mathcal{F}_{T_n} = \mathcal{F}_{T_n-}$, i.e., $\theta_n = c + \int_0^{T_n} \psi_s^{(n)} dW_s$ for an $\mathbb{F}$-predictable process $\psi^{(n)}$ (such that $\theta_n \geq 0$).
In this case, $A^I = I = \sum_n \mathbb{1}_{\{T_n \leq \cdot\}}(1 - e^{-c - \int_0^{T_n} \psi_s^{(n)} dW_s})$ and $M^I = 0$.

3.3.5.2   Subordinator processes

Before going further, we recall the compensation formula for Lévy processes as described in Jeanblanc, Yor, and Chesney, 2009, Proposition 11.2.2.3.

Proposition 34 *Let X be a Lévy process with Lévy measure denoted by ν and let f be a non-negative function such that $f(0) = 0$. Then,*

$$\mathbb{E}\left(\sum_{0 < s \leq t} f(\Delta X_s)\right) = t \int_{\mathbb{R}} f(x)\nu(dx).$$

Furthermore, if $\int_{\mathbb{R}} f(x)\nu(dx) < \infty$, then $\sum_{0 < s \leq t} f(\Delta X_s)$ is a Lévy process and the process

$$\sum_{0 < s \leq t} f(\Delta X_s) - t \int_{\mathbb{R}} f(x)\nu(dx)$$

is a martingale.

3.3.5.2.1   Case where K is a subordinator   We consider the case where K is a pure subordinator, i.e., a càdlàg Lévy process without drift with increasing paths (see, e.g., Jeanblanc, Yor, and Chesney, 2009, Subsection 11.6; Cont and Tankov, 2004, Subsection 4.2.2). We denote by ν its Lévy measure, which is such that

$$\int_0^\infty (1 \wedge x)\nu(dx) < \infty . \tag{3.3.9}$$

We consider $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ to be the Laplace exponent of K, i.e., ψ verifies

$$\mathbb{E}[e^{-uK_t}] = e^{-t\psi(u)}, \ t, u \geq 0$$

and has the following expression

$$\psi(u) = \int_0^\infty (1 - e^{-ux})\nu(dx), \ u \geq 0. \tag{3.3.10}$$

Note that the restriction (3.3.9) implies that ψ is finite for all $u \geq 0$.
The natural filtration of K is denoted by $\mathbb{F}$. Our goal is to use the result of Lemma 30 in order to make explicit the characteristics of the default time τ. For this purpose, since the continuous part of the process K is null, it suffices to compute the compensator A^I of the process $I_t := \sum_{s \leq t}(1 - e^{-\Delta K_s})$.

By considering the non-negative function $f(x) = 1 - e^{-x}, \ \forall x \geq 0$, one can apply Proposition 34 which implies that

$$\mathbb{E}\left(\sum_{0 < s \leq t} f(\Delta K_s)\right) = \mathbb{E}\left(\sum_{s \leq t}(1 - e^{-\Delta K_s})\right) = t \int_0^\infty (1 - e^{-x})\nu(dx).$$

Since $\int_0^\infty (1 - e^{-x})\nu(dx) = \psi(1) < \infty$, it follows that the process I is a subordinator and

$$I_t - t \int_0^\infty (1 - e^{-x})\nu(dx) = I_t - t\psi(1) \tag{3.3.11}$$

is an $\mathbb{F}$-martingale.

By consequence, the $\mathbb{F}$-predictable reduction of the compensator of τ is given by

$$\Lambda_t = t\psi(1). \tag{3.3.12}$$

For any u, the process

$$n_t(u) = e^{-uK_t + t\psi(u)} \tag{3.3.13}$$

is an $\mathbb{F}$-martingale with $\mathbb{E}[n_t(u)] = 1$ (see, e.g., Cont and Tankov, 2004, Proposition 3.17). In particular, the process n, defined as

$$n_t := n_t(1) = e^{-K_t + t\psi(1)} \tag{3.3.14}$$

is an $\mathbb{F}$-martingale.

Therefore, the survival probability of τ is given by

$$P(\tau > t) = \mathbb{E}[Z_t] = \mathbb{E}[e^{-K_t}] = \mathbb{E}[n_t e^{-t\psi(1)}] = e^{-t\psi(1)}$$

and its density is $f(t) = \psi(1)e^{-t\psi(1)}$, i.e., τ has an exponential law.

For $\theta \geq t$, one has

$$\mathbb{P}(\tau > \theta | \mathcal{F}_t) = \mathbb{E}[Z_\theta | \mathcal{F}_t] = e^{-K_t}\mathbb{E}[e^{-K_{\theta-t}}] = e^{-K_t}e^{-\psi(1)(\theta-t)} = n_t e^{-\theta\psi(1)} = n_t \int_\theta^\infty f(s)ds$$

so that for $\theta \geq t$, $\mathbb{P}(\tau \in d\theta | \mathcal{F}_t) = n_t f(\theta)d\theta$.

Since Z_u is not absolutely continuous w.r.t. Lebesgue's measure, the (CL)-conditional density does not exist in this model.

One can check the existence of the (GD)-conditional density by using Proposition 32 which implies that $\alpha(u) = 0$ because K has no continuous part ($k = 0$). This can be confirmed by the fact that $\sum_i \mathbb{P}(\tau = \tau_i) = 1$ which implies that, for any bounded Borel function h, one has $\mathbb{E}\left[h(\tau)\prod_i \mathbb{1}_{\{\tau_i \neq \tau\}}\right] = 0$.

Remark 35 *These results about subordinators cover a wide range of examples because of the large families of these kind of Lévy processes. This means that for any subordinator process that represents the process K in the generalized Cox model, it suffices to know its Lévy measure in order to obtain all the characteristics of the default time τ. In what follows, we present some examples of these processes.*

Example 36 *Compound Poisson process*

We consider K a compound Poisson process with non-negative jumps (which is a subordinator), i.e., $K_t = \sum_{n=1}^{N_t} Y_n$, where N is a Poisson process with jump times $(T_n, n \geq 1)$ and intensity λ and $(Y_n, n \geq 1)$ non-negative random variables, i.i.d. and independent of N. Hence, the Lévy measure ν of K is given by $\nu(dy) = \lambda F(dy)$, where F is the cumulative distribution function of Y_1, and its Laplace exponent ψ is $\psi(u) = \int_0^\infty (1 - e^{-uy})\nu(dy) = \lambda \int_0^\infty (1 - e^{-uy})F(dy) = \lambda(1 - \mathbb{E}[e^{-uY_1}])$. Thus, from (3.3.12), the intensity rate of τ is

$$\psi(1) = \lambda(1 - \mathbb{E}[e^{-Y_1}]).$$

In the particular case where Y_1 has exponential distribution with parameter γ, then the $\mathbb{F}$-predictable reduction of the compensator of τ is $t\psi(1)$, where

$$\psi(1) = \frac{\lambda}{1 + \gamma}.$$

Example 37 *Gamma process*

Let K be the Gamma process denoted by $(G_t(\alpha, \mu), t \geq 0)$ which is a subordinator with Lévy measure

$$\nu(dx) = \frac{\alpha}{x}e^{-\mu x}\mathbb{1}_{\{x > 0\}}dx$$

with $\alpha, \mu > 0$. Hence, $\psi(1) = \int_0^\infty (1 - e^{-x}) \frac{\alpha}{x} e^{-\mu x} dx$ and the $\mathbb{F}$-predictable reduction of the compensator of τ is

$$t \int_0^\infty (1 - e^{-x}) \frac{\alpha}{x} e^{-\mu x} dx.$$

Example 38   *Tempered stable subordinator*
In the case where K is a tempered stable subordinator, its Lévy measure ν is given by (see Cont and Tankov, 2004, Subsection 4.2.2)

$$\nu(dx) = \frac{\gamma e^{-\lambda x}}{x^{\beta+1}} \mathbb{1}_{\{x>0\}} dx,$$

where $\gamma > 0, \lambda > 0, 0 \le \beta < 1$. Therefore, the $\mathbb{F}$-predictable reduction of the compensator of τ is then given by $t \int_0^\infty (1 - e^{-x}) \frac{\gamma e^{-\lambda x}}{x^{\beta+1}} dx$.

3.3.5.2.2   Case where K is a subordinator plus an absolutely continuous part   We set

$$K_t = \int_0^t k_s ds + X_t$$

where k is a non-negative process adapted to the filtration $\mathbb{F}^W$ generated by a Brownian motion W and X a pure subordinator process independent of $\mathbb{F}^W$. We denote by $\mathbb{F}^X$ the natural filtration of X and by ν its Lévy measure. We define the reference filtration $\mathbb{F}$ as $\mathbb{F} = \mathbb{F}^W \vee \mathbb{F}^X$.

The multiplicative decomposition of $Z = e^{-K}$ is given by the following equality

$$Z_t = n_t e^{-\int_0^t k_s ds - t\psi(1)} \tag{3.3.15}$$

where n is the $\mathbb{F}^X$-martingale (hence an $\mathbb{F}$-martingale) defined in (3.3.14) and ψ is the Laplace transform of X defined in (3.3.10).

By integration by parts, one obtains the Doob-Meyer decomposition of Z:

$$dZ_t = e^{-\int_0^t k_s ds - t\psi(1)} dn_t - (k_t + \psi(1))Z_t dt . \tag{3.3.16}$$

Hence, the $\mathbb{F}$-dual optional projection A^o of τ, which is given by $A^o = 1 - Z$, verifies

$$dA_t^o = -e^{-\int_0^t k_s ds - t\psi(1)} dn_t + (k_t + \psi(1))Z_t dt \tag{3.3.17}$$

and the $\mathbb{F}$-predictable part of this Doob-Meyer decomposition of Z which is the $\mathbb{F}$-dual predictable projection of τ, A^p satisfies

$$dA_t^p = (k_t + \psi(1))Z_t dt. \tag{3.3.18}$$

Thus, Λ, the $\mathbb{F}$-predictable reduction of the compensator of τ is

$$d\Lambda_t = (k_t + \psi(1))dt. \tag{3.3.19}$$

This result can also be obtained by using Lemma 30. It suffices to note that $I_t = \sum_i (1 - e^{-\Delta X_{\tau_i}}) \mathbb{1}_{\{\tau_i \le t\}}$ is a submartingale with $\mathbb{F}$-predictable part $A_t^I = \psi(1)t$.

The survival probability of τ is

$$\mathbb{P}(\tau > t) = \mathbb{E}[Z_t] = \mathbb{E}\left[n_t e^{-\int_0^t k_s ds - t\psi(1)} \right] = e^{-t\psi(1)} \mathbb{E}\left[e^{-\int_0^t k_s ds} \right]. \tag{3.3.20}$$

This is due to the fact that the martingale n is independent of $e^{-\int_0^t k_s ds}$ and $\mathbb{E}[n_t] = 1$.

The default time admits a (GD) conditional density where the martingale family α is given by $\alpha_t(u) = $

$\mathbb{E}[k_u e^{-K_u}|\mathcal{F}_t]$, for all $u, t \geq 0$ and the conditional survival process of τ is then given by for $u \geq t$,

$$\mathbb{P}(\tau \geq u|\mathcal{F}_t) = \int_u^\infty \alpha_t(s)ds + \sum_i \mathbb{1}_{\{\tau_i \geq t\}} p_{\tau_i \vee t}^i,$$

where $p_t^i = \mathbb{P}(\tau = \tau_i|\mathcal{F}_t)$ are given from (3.3.6) by

$$p_t^i = \mathbb{E}\left[e^{-\int_0^{\tau_i} k_s ds} e^{-X_{\tau_{i-1}}}(1 - e^{-\Delta X_{\tau_i}})|\mathcal{F}_t\right].$$

3.3.5.3 Marked point process

If an increasing sequence $(\tau_i)_i$ of $\mathbb{F}$-stopping times and a sequence of non-negative random variables variables $(\theta_i \in \mathcal{F}_{\tau_i}, i \geq 1)$ are given, as well as a non-negative $\mathbb{F}$-adapted process k, one can construct an increasing càdlàg process K as $K_t = \int_0^t k_s ds + \sum_{i \geq 1} \mathbb{1}_{\{\tau_i \leq t\}} \theta_i$. This framework covers many cases of generalized Cox model, i.e., all those where the sequence of jump times of K is increasing and the continuous part is absolutely continuous w.r.t. the Lebesgue measure.

The associated random time τ does not avoid the stopping random times $(\tau_i)_{i \geq 1}$ and we set $\tau_0 = 0$. We have $I_t = \sum_{i \geq 1}(1 - e^{-\theta_i})\mathbb{1}_{\{\tau_i \leq t\}} = \sum_{i \geq 1} \gamma_i \mathbb{1}_{\{\tau_i \leq t\}} = M_t^I + A_t^I$. We consider the marked point process $(\gamma_i, \tau_i)_i$ with jump measure μ defined as, for any Borel set A

$$\mu(\omega, [0, t], A) = \sum_{i \geq 1} \mathbb{1}_{\{\tau_i(\omega) \leq t\}}(\omega)\mathbb{1}_{\{\gamma_i(\omega) \in A\}}$$

and its compensator ν given by

$$\nu(dt, dx) = \sum_{i \geq 0} \mathbb{1}_{\{\tau_i < t \leq \tau_{i+1}\}} \frac{\mathbb{P}(\tau_{i+1} \in dt, \gamma_{i+1} \in dx|\mathcal{F}_{\tau_i})}{\mathbb{P}(\tau_{i+1} > t|\mathcal{F}_{\tau_i})}$$

(see Last and Brandt, 1995a, Section 1.10). Then

$$I_t = \int_0^t \int_{\mathbb{R}^+} x\mu(ds, dx) = M_t^I + A_t^I$$

where M is the martingale $\int_0^t \int_{\mathbb{R}^+} x(\mu(ds, dx) - \nu(ds, dx))$ and $A_t^I = \int_0^t \int_{\mathbb{R}^+} x\nu(ds, dx)$. Furthermore, $dA_t^p = e^{-K_{t-}}(dK_t^c + dA_t^I)$.

3.3.5.4 A particular case of Marked point process

In the particular case of point process, i.e., $K_t = \sum_{i \geq 1} \mathbb{1}_{\{\tau_i \leq t\}}$ for an increasing sequence of $(\tau_i)_i$, denoting by $\mathbb{F}$ the natural filtration of K, one obtains $dA_t^p = e^{K_{t-}} dA_t^I$ where $I = (1 - e^{-1})K$ so that $A^I = (1 - e^{-1})\sum_{i \geq 1} J^i$, where, for any i, the process J^i is the $\mathbb{F}$-compensator of $A^i = \mathbb{1}_{\{\tau_i \leq \cdot\}}$.

In the case where all the J^i are continuous, the process $N = Z \exp \Lambda$ is a martingale, where $\Lambda_t = \int_0^t \frac{dA_s^p}{Z_{s-}} = A_t^I$. Hence, $e^{-K+(1-e^{-1})\sum_{i \geq 1} J^i}$ is a martingale.

One can recover the result of Giesecke and Zhu, 2013, Proposition 3.1 that, setting $\psi(u) = 1 - e^{-u}$, the process $e^{-uK + \psi(u)\sum_{i \geq 1} J_t^i}$ is a martingale by considering the increasing process $\widehat{K} = uK$.

Note that, if all the τ_i are predictable, K is predictable hence $\sum_{i \geq 1} J_t^i = K$. Furthermore, as noticed in Subsection 3.3.4 $dA^p = dI = -dZ$. One has also $A^I = I$, so that $\Lambda = (1 - e^{-1})K$.

3.3.5.5 Shot-noise processes

Shot-noise processes have been introduced by Campbell (1909a), Campbell (1909b) in the context of studying discontinuous phenomena and light emission and became popular in the whole physical domain with

Schottky (1918), which earned him the nomination Schottky-noise. A wide variety of studies of these processes have been developed in the actuarial sciences issues. These include the one of Schmidt (2014) for modeling the aggregated losses process in catastrophe reinsurance. Dassios and Jang (2003), in pricing of catastrophe reinsurance and derivatives, used shot noise process for modeling the intensity catastrophe event times. Mikosch (2009) applied these processes in non life insurance. Recent works have been done in credit risk; these include Scherer, Schmid, and Schmidt (2012) in portfolio default modeling where the authors used the first-passage times of the common clock across independent exponentially distributed threshold levels but they have not been interested in the default times characteristics. There is also the work of Herbertsson, Jang, and Schmidt (2011) in pricing basket default swaps.

Definition 39 *Let $\mathbb{F}$ be a given filtration, $(\tau_i)_i$ be a strictly increasing sequence of $\mathbb{F}$-stopping times with $\tau_0 = 0$, and (γ_i) a sequence of random variables with $\gamma_i \in \mathcal{F}_{\tau_i}$. We consider μ the random jumps measure of the marked point process $(\tau_i, \gamma_i)_i$ defined as $\mu(\omega, [0, t], A) = \sum_i \mathbb{1}_{\{\tau_i(\omega) \le t\}} \mathbb{1}_{\{\gamma_i(\omega) \in A\}}$ for $A \in \mathcal{B}(\mathbb{R})$ and ν its compensator. We denote by $\tilde{\mu}$ the compensated random measure $\tilde{\mu} = \mu - \nu$. We define the shot-noise process $X = (X_t)_{t \ge 0}$ (see Scherer, Schmid, and Schmidt, 2012) as*

$$X_t = \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \le t\}} G(t - \tau_i, \gamma_i) = \int_0^t \int_{\mathbb{R}} G(t - s, x) \mu(ds, dx), \forall t \ge 0 \tag{3.3.21}$$

where the function $G : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^+$, called shots' decay patterns, satisfies

$$G(t, x) = G(0, x) + \int_0^t g(s, x) ds \quad \forall\, t \ge 0,\ x \in \mathbb{R} \tag{3.3.22}$$

with g a non-negative Borel function on $\mathbb{R}^+ \times \mathbb{R}$, so that G is increasing with respect to its first variable. We assume that

$$\int_0^T \int_{\mathbb{R}} (g(s, x))^2 \nu(ds, dx) < \infty, \ \forall\, T,\ a.s \tag{3.3.23}$$

and there exists a non negative function φ such that

$$|g(s, x)| \le \varphi(x), \ \forall\, (s, x) \ with \ \int_0^T \int_{\mathbb{R}} \varphi(x) \nu(ds, dx) < \infty, \ \forall\, T \ a.s. \tag{3.3.24}$$

Note that $\Delta X_{\tau_i} = G(0, \gamma_i)$ for any $i \ge 1$ and the continuous part of X is given by $X_t^c = \sum_{i \ge 1} \mathbb{1}_{\{\tau_i \le t\}} [G(t - \tau_i, \gamma_i) - G(0, \gamma_i)]$.

The configurations of the function G lead to some different specifications of the shot-noise process X. For instance, when G has the exponential structure, i.e., $G(t, x) = x e^{\alpha t} \mathbb{1}_{\{x \ge 0\}}$, for $t \ge 0$ with $\alpha > 0$, the process X is Markovian (see Scherer, Schmid, and Schmidt, 2012, Example 2.1). Other varieties of the function G can be found in Scherer, Schmid, and Schmidt (2012) and Schmidt (2014). The simple case $G(t, x) = x$ leads to a compound Poisson process, if the $(\tau_i)_i$ are the jumps of a Poisson process N and $(\gamma_i)_i$ are i.i.d. independent of N.

We aim to find the characteristics of the default time τ when K is the shot-noise process defined in (3.3.21) (i.e., when $K = X$). In the next lemma, we give the decomposition of the increasing $\mathbb{F}$-adapted process (hence the $\mathbb{F}$-submartingale) X which will be used later. Our result is similar to the one of Schmidt, 2014, Lemma 2 and we have an explicit form for the martingale part.

Lemma 40 *The shot-noise process X is an $\mathbb{F}$-submartingale and admits the following Doob-Meyer decomposition $X_t = M_t^X + A_t^X$ where the increasing $\mathbb{F}$-predictable part A^X is*

$$A_t^X = \int_{u=0}^t \left(\int_{s=0}^u \int_{\mathbb{R}} g(u-s,x)\mu(ds,dx) \right) du + \int_{s=0}^t \int_{\mathbb{R}} G(0,x)\nu(ds,dx)$$

and the $\mathbb{F}$-martingale part M^X has the following form:

$$M_t^X = \int_0^t \int_{\mathbb{R}} G(0,x)\tilde{\mu}(ds,dx) \, .$$

PROOF: We extend the definition of G to $\mathbb{R} \times \mathbb{R}$ by setting $G(u,x) = G(0,x)$, for $u < 0$. Introducing $Y_t(a) = \int_0^t \int_{\mathbb{R}} G(a-s,x)\tilde{\mu}(ds,dx)$ for any $a \in \mathbb{R}$, one has

$$X_t = Y_t(t) + \int_0^t \int_{\mathbb{R}} G(t-s,x)\nu(ds,dx).$$

We apply the Itô-Ventzell formula as developed in Øksendal and Zhang, 2007, Theorem 3.1 to the process $Y_t(a)$ with parameter a, where forward integral in Øksendal and Zhang (2007) is usual integral in our setting, due to the fact that we integrate only deterministic functions with respect to compensated random measure. The first derivative of $Y_t(a)$ with respect to the parameter a is (see Metivier, 1982) due to (3.3.24),

$$Y_t'(a) = \int_0^t \int_{\mathbb{R}} g(a-s,x)\tilde{\mu}(ds,dx).$$

Using Theorem 3.1 of Øksendal and Zhang (2007), it follows that

$$Y_t(t) = \int_{u=0}^t Y_u'(u)du + \int_{s=0}^t \int_{\mathbb{R}} G(0,x)\tilde{\mu}(ds,dx).$$

Note that $\int_{u=0}^t Y_u'(u)du$ is a continuous bounded variation adapted process, hence is a predictable process. Therefore,

$$\begin{aligned} X_t &= \int_{u=0}^t \left(\int_{s=0}^u \int_{\mathbb{R}} g(u-s,x)\tilde{\mu}(ds,dx) \right) du + \int_0^t \int_{\mathbb{R}} G(0,x)\tilde{\mu}(ds,dx) + \int_0^t \int_{\mathbb{R}} G(t-s,x)\nu(ds,dx) \\ &= M_t^X + \int_0^t \int_{\mathbb{R}} G(t-s,x)\nu(ds,dx) + \int_{u=0}^t \left(\int_{s=0}^u \int_{\mathbb{R}} g(u-s,x)\tilde{\mu}(ds,dx) \right) du \end{aligned}$$

where M^X is the local martingale $\int_0^t \int_{\mathbb{R}} G(0,x)\tilde{\mu}(ds,dx)$. Using (3.3.22), one has

$$\int_0^t \int_{\mathbb{R}} G(t-s,x)\nu(ds,dx) = \int_{s=0}^t \int_{\mathbb{R}} \int_{u=0}^{t-s} g(u,x)du \, \nu(ds,dx) + \int_0^t \int_{\mathbb{R}} G(0,x)\nu(ds,dx).$$

By making the change of variable $y = u + s$, one obtains

$$\int_{u=0}^{t-s} g(u,x)du = \int_{y=s}^t g(y-s,x)dy \, ,$$

hence by using stochastic Fubini's theorem (see Protter, 2005, Theorem 65) valid under (3.3.23), one gets

$$\begin{aligned} \int_0^t \int_{\mathbb{R}} G(t-s,x)\nu(ds,dx) &= \int_{s=0}^t \int_{\mathbb{R}} \int_{u=s}^t g(u-s,x)du \, \nu(ds,dx) + \int_0^t \int_{\mathbb{R}} G(0,x)\nu(ds,dx) \\ &= \int_{u=0}^t \left(\int_{s=0}^u \int_{\mathbb{R}} g(u-s,x) \, \nu(ds,dx) \right) du + \int_0^t \int_{\mathbb{R}} G(0,x)\nu(ds,dx). \end{aligned}$$

Therefore,

$$X_t = M_t^X + \int_{u=0}^t \left(\int_{s=0}^u \int_{\mathbb{R}} g(u-s,x)\, \mu(ds,dx) \right) du + \int_0^t \int_{\mathbb{R}} G(0,x)\nu(ds,dx) = M_t^X + A_t^X$$

where $A_t^X = \int_{u=0}^t \left(\int_{s=0}^u \int_{\mathbb{R}} g(u-s,x)\, \mu(ds,dx) \right) du + \int_0^t \int_{\mathbb{R}} G(0,x)\nu(ds,dx).$ $\qquad\qquad\square$

Now, we are ready to compute the $\mathbb{F}$-predictable reduction of the compensator of the default time τ. This is given in the next Lemma.

Lemma 41 *If the process K is of the form* (3.3.21), *then the $\mathbb{F}$-predictable reduction of the compensator of τ has the following expression:*

$$\Lambda_t = \int_{u=0}^t \left(\int_{\mathbb{R}} \int_{s=0}^u g(u-s,x)\mu(ds,dx) \right) du + \int_0^t \int_{\mathbb{R}} (1 - e^{-G(0,x)})\nu(ds,dx)\,. \qquad (3.3.25)$$

PROOF: The predictable part of the submartingale

$$I_t = \sum_i \mathbb{1}_{\{\tau_i \leq t\}}(1 - e^{-\Delta K_{\tau_i}}) = \int_0^t \int_{\mathbb{R}} (1 - e^{-G(0,x)})\mu(ds,dx)$$

is $A_t^I = \int_0^t \int_{\mathbb{R}} (1 - e^{-G(0,x)})\nu(ds,dx)$. In addition, the continuous part K^c of K is given by

$$\begin{aligned}
K_t^c &= \sum_{i \geq 1} \mathbb{1}_{\{\tau_i \leq t\}}[G(t-\tau_i,\gamma_i) - G(0,\gamma_i)] = \int_{s=0}^t \int_{\mathbb{R}} \left(\int_{u=0}^{t-s} g(u,x)du \right) \mu(ds,dx) \\
&= \int_{u=0}^t \left(\int_{\mathbb{R}} \int_{s=0}^u g(u-s,x)\mu(ds,dx) \right) du.
\end{aligned}$$

Therefore, from Lemma 3.3.1, the $\mathbb{F}$-predictable reduction of the compensator of τ is $K^c + A^I$. $\qquad\square$

Remark 42 *Note that the $\mathbb{F}$-predictable reduction of the compensator of τ can also be recovered by using* (1.2.1). *Indeed, by Itô's formula for semimartingales, one has*

$$\begin{aligned}
dZ_t &= -e^{-K_{t-}}dK_t + e^{-K_{t-}}(e^{-\Delta K_t} - 1) + e^{K_{t-}}\Delta K_t \\
&= e^{-K_{t-}}\big(-dM_t^K - dA_t^K + e^{-\Delta K_t} - 1 + \Delta K_t \big).
\end{aligned}$$

Hence by denoting by m the martingale part of the Doob Meyer decomposition of Z, it follows that

$$dZ_t = Z_{t-}\left(-dM_t^K - dA_t^K + \int_{\mathbb{R}} (e^{-G(0,x)} - 1 + G(0,x))\mu(dt,dx) \right) = dm_t - dA_t^p,$$

where $dA_t^p = Z_{t-}\left(dA_t^K + \int_{\mathbb{R}}(1 - e^{-G(0,x)} - G(0,x))\nu(dt,dx) \right)$. Hence by using (1.2.1), *one obtains*

$$d\Lambda_t = dA_t^K + \int_{\mathbb{R}} \left(1 - e^{-G(0,x)} - G(0,x) \right) \nu(dt,dx).$$

The result follows by replacing A^K by its expression, which is

$$A_t^K = \int_{u=0}^t \left(\int_{s=0}^u \int_{\mathbb{R}} g(u-s,x)\, \mu(ds,dx) \right) du + \int_0^t \int_{\mathbb{R}} G(0,x)\nu(ds,dx).$$

The following result gives the $\mathbb{F}$-conditional law of τ in the particular case where ν is deterministic.

Proposition 43 *If ν is deterministic, then for any $u \geq t$, one has*

$$\mathbb{P}(\tau > u | \mathcal{F}_t) = c(u) L_t(u) \tag{3.3.26}$$

with $c(u) = \exp\left(\int_0^u \int_{\mathbb{R}} (e^{-G(u-s,x)} - 1)\nu(ds, dx)\right)$ and $L(u)$ is the martingale

$$L_t(u) = \exp\left(-\int_0^t \int_{\mathbb{R}} G(u-s, \ x)\mu(ds, dx) - \int_0^t \int_{\mathbb{R}} (e^{-G(u-s, \ x)} - 1)\nu(ds, dx) \right).$$

For $u < t$, one has $\mathbb{P}(\tau > u | \mathcal{F}_t) = c(u) L_u(u)$ thanks to the immersion property.

PROOF: Let $V_t(u) = -\int_0^t \int_{\mathbb{R}} G(u-s, \ x)\mu(ds, dx) - \int_0^t \int_{\mathbb{R}} (e^{-G(u-s, \ x)} - 1)\nu(ds, dx)$ for $u \geq t \geq 0$. By applying Itô's formula for semimartingales to the process $L_t(u) = e^{V_t(u)}$, one has (where we denote, for u fixed, L_t and V_t instead of $L_t(u)$ and $V_t(u)$),

$$dL_t = L_{t-} \ (dV_t + e^{\Delta V_t} - 1 - \Delta V_t). \tag{3.3.27}$$

By writing $e^{\Delta V_t} - 1 - \Delta V_t = \int_{\mathbb{R}} \left(e^{-G(u-t,x)} - 1 + G(u-t,x) \right) \mu(dt, dx)$, one obtains

$$dL_t = L_{t-} \ \int_{\mathbb{R}} (e^{-G(u-t,x)} - 1)\tilde{\mu}(dt, dx) \tag{3.3.28}$$

which implies that L is a local martingale. By (3.3.28), it is a true martingale.

If ν is deterministic, due to the fact that $-K_u = V_u - \int_0^u \int_{\mathbb{R}} (1 - e^{-G(u-s,x)})\nu(ds, dx)$, one has for $u > t$

$$\mathbb{E}[e^{-K_u} | \mathcal{F}_t] = c(u) L_t$$

where

$$c(u) = \mathbb{E}[e^{-K_u}] = \exp\left(-\int_0^u \int_{\mathbb{R}} (1 - e^{-G(u-s,x)})\nu(ds, dx) \right). \tag{3.3.29}$$

in particular, we recover the special semimartingale feature of X proved in Altmann, Schmidt, and Stute, 2008, Proposition 2.1.

$\square$

Remark 44 *Note from (3.3.27) that $\Delta L_t(u) = e^{V_{t-}(u)}(e^{\Delta V_t(u)} - 1)$ for $0 \leq t < u$.*

3.4   Case where K is càglàd

We present the case where the increasing process K is left-continuous with right limits (càglàd) which implies its predictability. Under this setting, we define the random time

$$\tau = \inf\{t \ : \ K_t > \Theta\} \tag{3.4.1}$$

to be able to characterize the set $\{K_t > \Theta\}$.

Lemma 45 *If K is càglàd, then defining τ as above (3.4.1), for all $t \geq 0$, the following equality*

$$\{K_t \leq \Theta\} = \{\tau \geq t\}$$

holds.

PROOF: As in Lemma 29, we proceed for ω fixed. The set $\{K_t > \Theta\}$ is either of the form $[a, \infty[$ or of the form $]a, \infty[$ and $a = \tau$. If it is of the form $[a, \infty[$ it would mean that $a = \tau$ is in the set $\{K_t > \Theta\}$, hence $K_\tau > \Theta$. By left-continuity of K, one should have $K_{\tau-\varepsilon} > \Theta$ for ε small enough, hence a contradiction. It follows that $\{K_t > \Theta\} =]\tau, \infty] = \{t > \tau\}$. $\square$

Proposition 46 *If K is càglàd, and $\tau = \inf\{t : K_t > \Theta\}$, one has $Z = e^{-K_+}$, $d\Lambda = e^{-\Delta^+ K}(dK^c + dA^C)$ where A^C is the predictable part of the supermartingale $C_t = \sum_{s \leq t} \left(e^{-\Delta^+ K_s} - 1\right)$.*

PROOF: Let $\kappa_t := K_{t+} = K_t + \Delta^+ K_t$ where $\Delta^+ K_t = K_{t_+} - K_t$. Let $J_t = \sum_{s \leq t} \Delta^+ K_s$, so that $\kappa_t = K_t^c + J_t$, where K^c is the continuous part of K, defined as $K_t^c = K_t - \sum_{s \leq t} \Delta^+ K_s$.

From Lemma 45, one has $\widetilde{Z}_t = \mathbb{P}(\tau \geq t | \mathcal{F}_t) = \mathbb{P}(\Theta \geq K_t | \mathcal{F}_t) = e^{-K_t}$ and $Z_t = e^{-\kappa_t}$ (as $\widetilde{Z}_+ = Z_+ = Z$ (see Aksamit and Jeanblanc, 2017, Subsection 1.3.4)), one has

$$\begin{aligned} dZ_t &= -e^{-\kappa_{t-}} d\kappa_t + (e^{-\kappa_t} - e^{-\kappa_{t-}} + e^{-\kappa_{t-}} \Delta \kappa_t) \\ &= -e^{-\kappa_{t-}} dK_t^c + e^{-\kappa_{t-}}(e^{-\Delta \kappa_t} - 1). \end{aligned}$$

The decreasing càdlàg process C defined as $C_t := \sum_{s \leq t} e^{-\Delta \kappa_s} - 1 = \sum_{s \leq t} e^{-\Delta^+ K_s} - 1$ admits a Doob-Meyer decomposition $C = M^C - A^C$ and

$$\begin{aligned} dZ_t &= e^{-\kappa_{t-}} dM_t^C - e^{-\kappa_{t-}}(dK_t^c + dA_t^C) \\ &= e^{-K_t} dM_t^C - e^{-K_t}(dK_t^c + dA_t^C) \end{aligned}$$

and $dA_t^p = e^{-K_t}(dK_t^c + dA_t^C)$, $d\Lambda_t = e^{\Delta^+ K_t}(dK_t^c + dA_t^C)$. $\square$

For $u < t$, the equality

$$\mathbb{P}(\tau \leq u | \mathcal{F}_t) = 1 - e^{-\kappa_u} = \int_0^u dD_s \ ,$$

where $D_s = 1 - e^{-\kappa_s}$, proves that the extended LR-conditional density exists.

Chapter 4

Application of the generalized Cox model in credit risk modeling

In this chapter, as in the previous ones, we are working on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, endowed with a filtration $\mathbb{F}$ where a generalized Cox time (see Chapter 3) τ is defined and we consider $\mathbb{G}$ the progressive enlarged filtration of $\mathbb{F}$ with τ. As in the previous chapters, we shall simply denote by A^p and A^o the $\mathbb{F}$-dual predictable and optional projections of the process $A = \mathbb{1}_{\{\tau \leq \cdot\}}$.

The aim of this chapter is to highlight the importance of the generalized Cox framework. As such, we present some useful case studies in credit risk modeling through this model. We will be interested to the behavior of the prices at the jump times of the generalized Cox process K used in the definition of τ (4.1.2).

4.1 Framework

We consider a defaultable bond (in what follows, we shall only write a DB) with maturity T and payoff given by

$$\zeta := \mathbb{1}_{\{\tau > T\}} + h_\tau \mathbb{1}_{\{\tau \leq T\}},$$

where the recovery h is a bounded process which is either $\mathbb{F}$-predictable or $\mathbb{F}$-optional and which we suppose to be paid at the default time τ if $\tau \leq T$ whereas the unit payoff is done at time T if no default has occured before. In what follows, we suppose that the probability $\mathbb{P}$ is the pricing measure and the interest rate is zero. By denoting by $P_t(T), t \leq T$ the price of the defaultable bond which is given, for any $t \geq 0$, as

$$P_t(T) = \mathbb{E}[\mathbb{1}_{\{\tau > T\}} + h_\tau \mathbb{1}_{\{t < \tau \leq T\}} | \mathcal{G}_t],$$

we define the pre-default price (see Elliott, Jeanblanc, and Yor, 2000) of the defaultable bond as the $\mathbb{F}$-adapted process $D(T)$ such that for any $t \geq 0$,

$$D_t(T)\mathbb{1}_{\{\tau > t\}} = P_t(T)\mathbb{1}_{\{\tau > t\}}.$$

We also consider a Credit Default Swap (we only write a CDS) of maturity T, with spread κ and $\mathbb{F}$-predictable recovery δ which is a contract between two parties, the protection buyer and the protection seller having their payoffs which depend on the default time τ of a reference entity. Under this contract, the seller is committed to protect the buyer against the default of the entity by paying her δ_τ, at the default time τ, if the default of the entity occurs before the maturity T. In exchange, the protection buyer pays a premium κ, at some predetermined dates, to the protection seller till $\tau \wedge T$. For simplicity we assume that the premium is paid continuously, i.e., κdt is paid during $[t, t+dt]$ if $t + dt \leq \tau \wedge T$.

From the buyer point of view, the price at time $t < \tau \wedge T$ of the CDS, denoted $CDS_t(\kappa, \delta, T)$, is the difference between the expected present value of the amount received by the protection buyer $D_t^{leg} = \mathbb{E}[\delta_\tau \mathbb{1}_{\{t < \tau \leq T\}} | \mathcal{G}_t]$

and the one received by the protection seller $P_t^{leg} = \mathbb{E}[\int_t^T \mathbb{1}_{\{u<\tau\}}\kappa du|\mathcal{G}_t]$, i.e.,

$$CDS_t(\kappa,\delta,T) = \mathbb{E}[\delta_\tau \mathbb{1}_{\{t<\tau\leq T\}} - \kappa(T\wedge\tau - t)^+|\mathcal{G}_t] \; . \tag{4.1.1}$$

By definition of the CDS contract, one has $CDS_0(\kappa,\delta,T) = 0$ which determines the value of κ.

We consider the default time τ to be a generalized Cox time, i.e.,

$$\tau = \inf\{t \geq 0 : K_t \geq \Theta\}, \tag{4.1.2}$$

where K is an increasing càdlàg $\mathbb{F}$-adapted process and Θ a unit exponential r.v. independent of $\mathbb{F}$.
To make explicit the pre-default prices of the DB and the two legs of the CDS defined above, one needs to specify the situation in which we are, i.e., whether or not the process K is subject to some economic shocks, and if these shocks are predictable or not.
In what follows, we consider some different cases of the economy for which we compute the characteristics of τ before plugging them into the pricing formula.

4.2   The model

We consider an increasing sequence of $\mathbb{F}$-stopping times $(\tau_i)_i$, with $\tau_0 = 0$ and we define the increasing càdlàg $\mathbb{F}$-adapted process K as

$$K_t = \int_0^t k_s ds + \sum_{i\geq 1} \mathbb{1}_{\{\tau_i\leq t\}}\Delta K_{\tau_i} \tag{4.2.1}$$

where k is a non-negative $\mathbb{F}$-adapted process and $\Delta K_t = K_t - K_{t-}$ is non-negative.
The financial intuition behind this model can be seen in the particular case where the claim (DB) is issued by an entity (E) (which may be, for example, a firm or a sovereign country) with a risky asset S, tradable in the financial market, assumed to be $\mathbb{F}^W$-adapted, where $\mathbb{F}^W$ is the filtration generated by a Brownian motion W, and $\mathbb{F}^W \subset \mathbb{F}$. The asset S may be the total asset value of the entity (E) when (E) is a firm or its solvency process when (E) is a sovereign country. The process k which is supposed to be $\mathbb{F}^W$-adapted, can be considered (as in Jiao and Li, 2018) to be a function of S, i.e., $k = \varphi(S)$, where φ is a decreasing non-negative function and the jump times $(\tau_i)_i$ of the process K as the times of some possible economic shocks which can affect the price of the claims. In this setting, the times of the shocks capture the exogenous shocks of the economy while the process k captures the possible losses effect in the asset S, a useful type of this process k being the so-called negative power process given as $\varphi(S) = \alpha S^{-p}$, with α and p some non-negative constants, used by Linetsky (2006) in the context of pricing equity derivatives. The case where the entity (E) is a sovereign country and $\mathbb{F}^W = \mathbb{F}$ is studied by Jiao and Li (2018).

4.2.1   Case where the economic shocks are predictable

We consider the case where the jumps times $(\tau_i)_i$ are $\mathbb{F}$-predictable stopping times.
An example of this modeling consists in considering a decreasing sequence of positive fixed barriers $(b_i)_i$ and defining each τ_i as the first time the (positive) risky asset S ($\mathbb{F}$-adapted and continuous), with $S_\infty = 0$, being below the barrier b_i, i.e., $\tau_i = \inf\{t \geq 0 : S_t \leq b_i\}$, with $S_0 > b_1$.
One can define the jumps size of the process K as follows:

$$K_{\tau_i} - K_{\tau_i-} = \theta_i, \tag{4.2.2}$$

where $(\theta_i)_{i \geq 1}$ is a sequence of non-negative random variables with $\theta_i \in \mathcal{F}_{\tau_i -}$. In this case, the process K given by

$$K_t = \int_0^t k_s ds + \sum_{i \geq 1} \mathbb{1}_{\{\tau_i \leq t\}} \theta_i \qquad (4.2.3)$$

is increasing right-continuous $\mathbb{F}$-predictable. This model which is the one of Bélanger, Shreve, and Wong (2004) is a particular case of the one described in 3.3. In the particular case where $\theta_i = \Psi(\tau_i) - \Psi(\tau_{i-1})$, with Ψ an increasing function, we recover the generalized Jiao and Li model in which the associated K is

$$K_t = \Gamma_t + \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}} [\Psi(\tau_i) - \Psi(\tau_{i-1})] \qquad (4.2.4)$$

with $\Gamma_t = \int_0^t k_s ds$.

Let us now investigate the pre-default price of the defaultable bond defined above under this setting (i.e., K predictable defined by (4.2.3)).

Since K is $\mathbb{F}$-predictable the $\mathbb{F}$-dual projections A^o and A^p are the same and are equal to $(1 - Z)$. This implies that the pricing of defaultable contingent claims can be done by using any of them. By identifying $K_t^c = \int_0^t k_s ds$ and $A_t^I = I_t = \sum_{i=1}^{\infty}(1 - e^{-\theta_i})\mathbb{1}_{\{\tau_i \leq t\}}$ (which is due to the fact that I is $\mathbb{F}$-predictable and increasing hence $M^I = 0$), one deduces, from Lemma 30, that the $\mathbb{F}$-predictable reduction Λ of the compensator of τ is given by:

$$\Lambda_t = \int_0^t k_s ds + \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}}(1 - e^{-\theta_i}). \qquad (4.2.5)$$

From Proposition 32, the generalized density hypothesis is satisfied with $\alpha_t(s) = \mathbb{E}[e^{-K_s} k_s | \mathcal{F}_t]$ and from Proposition 22 it follows

$$Z_t = 1 - \mathbb{P}(\tau \leq t | \mathcal{F}_t) = 1 - \int_0^{\infty} \mathbb{1}_{\{u \leq t\}} \alpha_t(u) du - \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}} p_{\tau_i}^i$$

$$= 1 - \int_0^t \alpha_u(u) du - \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}} p_t^i \qquad (4.2.6)$$

where we have used the fact that

$$\mathbb{1}_{\{\tau_i \leq t\}} p_t^i = \mathbb{1}_{\{\tau_i \leq t\}} \mathbb{E}[e^{-K_{\tau_i -}}(1 - e^{-\theta_i}) | \mathcal{F}_t] = \mathbb{1}_{\{\tau_i \leq t\}} e^{-K_{\tau_i -}}(1 - e^{-\theta_i}) = \mathbb{1}_{\{\tau_i \leq t\}} p_{\tau_i}^i$$

and the fact that $\alpha_t(u) = \alpha_u(u)$ on $\{t \geq u\}$.

Proposition 47 *The pre-default price at time t, denoted $D_t(T)$, of the defaultable bond has the following expression*

$$D_t(T) = \widetilde{D}_t(T) + \widehat{D}_t(T) \qquad (4.2.7)$$

where

$$\widetilde{D}_t(T) = \mathbb{E}\left[\exp\left(-\sum_{i=1}^{\infty} \mathbb{1}_{\{t < \tau_i \leq T\}} \theta_i - \int_t^T k_s ds\right) | \mathcal{F}_t\right] \qquad (4.2.8)$$

is the predefault price of the terminal payoff 1 and

$$\widehat{D}_t(T) = \frac{\mathbb{E}\left[\int_t^T h_u \alpha_T(u) du + \sum_{i=1}^{\infty} h_{\tau_i} p_{\tau_i}^i \mathbb{1}_{\{t < \tau_i \leq T\}} | \mathcal{F}_t\right]}{Z_t} \qquad (4.2.9)$$

is the predefault price of the recovery.

PROOF: from the general pricing formula (1.3.1), one has

$$D_t(T) = \mathbb{E}\left[\frac{Z_T}{Z_t}|\mathcal{F}_t\right] + \frac{\mathbb{E}\left[\int_t^T h_u dA_u^p|\mathcal{F}_t\right]}{Z_t} = \widetilde{D}_t(T) + \widehat{D}_t(T),$$

with

$$\widetilde{D}_t(T) = \mathbb{E}\left[\frac{Z_T}{Z_t}|\mathcal{F}_t\right] = \mathbb{E}\left[e^{K_t - K_T}|\mathcal{F}_t\right]$$

$$= \mathbb{E}\left[\exp\left(-\sum_{i=1}^{\infty}\mathbb{1}_{\{t < \tau_i \le T\}}\theta_i - \int_t^T k_s ds\right)|\mathcal{F}_t\right]$$

and

$$\widehat{D}_t(T) = \frac{\mathbb{E}\left[\int_t^T h_u dA_u^p|\mathcal{F}_t\right]}{Z_t}.$$

Since on the set $\{\tau_i \le t\}$, one has $p_t^i = p_{\tau_i}^i$, and one obtains $d(\mathbb{1}_{\{\tau_i \le t\}}p_t^i) = p_{\tau_i}^i dA_t^i$, where $A_t^i = \mathbb{1}_{\{\tau_i \le t\}}$. From the equality $A_t^p = 1 - Z_t$, by using the fact that $\alpha_t(u) = \alpha_u(u)$ on the set $\{t > u\}$, one has

$$dA_t^p = -dZ_t = \alpha_t(t)dt + \sum_{i=1}^{\infty} p_t^i dA_t^i$$

where we have used the equality (4.2.6). Therefore, it follows

$$\int_t^T h_u dA_u^p = \int_t^T h_u \alpha_u(u)du + \sum_{i=1}^{\infty}\int_t^T h_u p_u^i dA_u^i$$

$$= \int_t^T h_u \alpha_u(u)du + \sum_{i=1}^{\infty} h_{\tau_i} p_{\tau_i}^i \mathbb{1}_{\{t < \tau_i \le T\}}.$$

Since $\alpha_T(u) = \alpha_u(u)$ on $\{u \le T\}$, one obtains

$$\mathbb{E}[\int_t^T h_u dA_u^p|\mathcal{F}_t] = \mathbb{E}\left[\int_t^T h_u \alpha_T(u)du + \sum_{i=1}^{\infty} h_{\tau_i} p_{\tau_i}^i \mathbb{1}_{\{t < \tau_i \le T\}}|\mathcal{F}_t\right].$$

By consequence,

$$\widehat{D}_t(T) = \frac{\mathbb{E}\left[\int_t^T h_u \alpha_T(u)du + \sum_{i=1}^{\infty} h_{\tau_i} p_{\tau_i}^i \mathbb{1}_{\{t < \tau_i \le T\}}|\mathcal{F}_t\right]}{Z_t}.$$

$\square$

As we mentioned, we can construct the process K which jumps at the times $(\tau_i)_i$ and in which we may assume to know the laws of its jump sizes, its jump times and its continuous part as well as the joint law of its jumps sizes. This allows us to get the $\mathbb{F}$-predictable reduction of the compensator of τ given by (4.2.5) of the default time τ which also jumps at the times $(\tau_i)_i$ with known jumps size. This construction leads to the same model as the model of Jiao and Li (2018) in context of modeling sovereign risks by means of the setting in (4.2.4), in the sense that the two models have two identical Azéma supermartingales and both verify the immersion property. Therefore, one can use our model in the modeling sovereign risks.

The inverse problem   From (4.2.5), one has $1 - \Delta\Lambda_{\tau_i} = e^{-\theta_i} > 0$ and $K_t = \Lambda_t - \sum_{i=1}^{\infty}\mathbb{1}_{\{\tau_i \le t\}}(1 - e^{-\theta_i} - \theta_i)$. Hence $\theta_i = -\ln(1 - \Delta\Lambda_{\tau_i})$ and

$$K_t = \Lambda_t - \sum_{i=1}^{\infty}\mathbb{1}_{\{\tau_i \le t\}}\left[\Delta\Lambda_{\tau_i} - \ln(1 - \Delta\Lambda_{\tau_i})\right]. \tag{4.2.10}$$

Hence by giving a process Λ, increasing predictable continuous on right which jumps at some increasing predictable times $(\tau_i)_i$, with $\tau_0 = 0$, and such that its jump size $\Delta\Lambda_{\tau_i}$ at τ_i verifies $0 \leq \Delta\Lambda_{\tau_i} < 1$ (see (3.3.8)), one can construct a default time τ which admits Λ as $\mathbb{F}$-predictable reduction of the compensator of τ under the Cox framework by defining K as in (4.2.10).

4.2.2 Case where some economic shocks are not predictable

When the economic shocks are no more predictable, one can consider for example their times of occurrence as the jump times of a compound Poisson process or more generally a shot-noise process. Here, we give these two examples.

4.2.2.1 Example of the compound Poisson process

We consider the times of occurrence of the shocks $(\tau_i)_i$ as the jump times of a Poisson process N with intensity λ and we introduce a Brownian motion W independent of N with natural filtration denoted by $\mathbb{F}^W$. We set the following form of K:

$$K_t = \int_0^t k_s ds + X_t \tag{4.2.11}$$

where k is an $\mathbb{F}^W$-adapted non-negative process and $X_t = \sum_{i=1}^{N_t} Y_i$ is a compound Poisson process with $(Y_i, i \geq 1)$ non-negative random variables i.i.d, independent of N and independent of $\mathbb{F}^W$. We denote by $\mathbb{F}^X$ the filtration generated by X and define the reference filtration $\mathbb{F}$ as the one generated by X and W.
This is a particular case of 3.3.5.2.2, where $X_t := \sum_{i=1}^{N_t} Y_i$ admits the Laplace transform ψ defined, for all $u \in \mathbb{R}^+$, by $\psi(u) = \int_0^\infty (1 - e^{-uy}) F(dy)$ with F the cumulative distribution function of Y_1 and the $\mathbb{F}$-martingale n in the multiplicative decomposition of Z (3.3.15) is given by $n_t = \exp\left(-X_t + t\lambda\psi(1)\right)$.

The pricing formula We assume that the recovery process h is a bounded $\mathbb{F}$-predictable process, then by using (3.3.17), one has

$$\mathbb{E}\left[\int_t^T h_s dA_s^o | \mathcal{F}_t\right] = \mathbb{E}\left[-\int_t^T h_s dZ_s | \mathcal{F}_t\right] = \mathbb{E}\left[\int_t^T h_s dA_s^p | \mathcal{F}_t\right].$$

The last equality is due to the fact that the conditional expectation of the integral w.r.t the martingale part of Z vanishes. Thus one can either use A^o or A^p in the valuation of claims when dealing with predictable recovery.

Proposition 48 *The pre-default price $D_t(T)$ at time t of the defaultable bond is given by*

$$D_t(T) = \widetilde{D}_t(T) + \widehat{D}_t(T)$$

where the quantity $\widetilde{D}(T)$ is given by

$$\widetilde{D}_t(T) = e^{-(T-t)\lambda\psi(1)} \exp\left(\int_0^t k_s ds\right) \mathbb{E}\left[\exp\left(-\int_0^T k_s ds\right) | \mathcal{F}_t^W\right] \tag{4.2.12}$$

and $\widehat{D}_t(T)$ has the following expression

$$\widehat{D}_t(T) = \frac{1}{Z_t} \mathbb{E}\left[\int_t^T (k_u + \lambda\psi(1)) h_u Z_u du | \mathcal{F}_t\right]. \tag{4.2.13}$$

PROOF: The quantity $\widetilde{D}_t(T)$ is given by

$$
\begin{aligned}
\widetilde{D}_t(T) &= \mathbb{E}\left[e^{K_t - K_T}|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[\exp\left(-\int_t^T k_s ds - (X_T - X_t)\right)|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[\exp\left(-\int_t^T k_s ds\right)\mathbb{E}[\exp\left(-(X_T - X_t)\right)|\mathcal{F}_T^W \vee \mathcal{F}_t^X]|\mathcal{F}_t\right]
\end{aligned}
$$

where we have used the tower property in the last equality. The fact that the compound Poisson process X is independent of $\mathbb{F}^W$ and $\mathcal{F}_t^X \subset \mathcal{F}_t$ leads to

$$
\begin{aligned}
\widetilde{D}_t(T) &= \mathbb{E}\left[\exp\left(-\int_t^T k_s ds\right)\mathbb{E}[\exp\left(-(X_T - X_t)\right)|\mathcal{F}_t^X]|\mathcal{F}_t\right] \\
&= \mathbb{E}[\exp\left(-(X_T - X_t)\right)|\mathcal{F}_t^X]\mathbb{E}\left[\exp\left(-\int_t^T k_s ds\right)|\mathcal{F}_t\right].
\end{aligned}
$$

The compound Poisson process X being with independent and stationary increments, one has

$$
\widetilde{D}_t(T) = \mathbb{E}[\exp\left(-X_{T-t}\right)]\mathbb{E}\left[\exp\left(-\int_t^T k_s ds\right)|\mathcal{F}_t\right].
$$

Thus by using the fact that $\mathbb{E}[\exp\left(-X_{T-t}\right)] = \mathbb{E}[e^{-(T-t)\lambda\psi(1)}n_{T-t}] = e^{-(T-t)\lambda\psi(1)}$, with the martingale n given by $n_t = e^{-K_t + t\lambda\psi(1)}$, and the independence of k from $\mathbb{F}^X$, one has

$$
\widetilde{D}_t(T) = e^{-(T-t)\lambda\psi(1)}\mathbb{E}\left[\exp\left(-\int_t^T k_s ds\right)|\mathcal{F}_t^W\right].
$$

The computation of $\widehat{D}_t(T)$ follows by replacing $dA_t^p = (k_t + \lambda\psi(1))Z_t\,dt$ in

$$
\widehat{D}_t(T) = \frac{\mathbb{E}\left[\int_t^T h_u dA_u^p|\mathcal{F}_t\right]}{Z_t},
$$

by means

$$
\widehat{D}_t(T) = \frac{\mathbb{E}\left[\int_t^T h_u(k_u + \lambda\psi(1))Z_u\,du|\mathcal{F}_t\right]}{Z_t}
$$

□

Term structure of defaultable bond with zero recovery If the recovery is zero, i.e., $h = 0$, the pre-default price of the DB is given by

$$
D_t(T) = e^{-(T-t)\lambda\psi(1)}\mathbb{E}\left[\exp\left(-\int_t^T k_s ds\right)|\mathcal{F}_t^W\right]. \tag{4.2.14}
$$

Whenever k is an affine process, this leads to the following analytical solution

$$
D_t(T) = e^{-\lambda\psi(1)(T-t)}\bar{Q}_t(T),
$$

with $\bar{Q}_t(T) := \mathbb{E}\left[e^{-\int_t^T k_s ds}|\mathcal{F}_t\right] = e^{A_t(T) - B_t(T)k_t}$, where A and B are differentiable functions with $A_T(T) = 0$ and $B_T(T) = 0$ and verify generalized Riccati ODEs (see Duffie, Filipović, and Schachermayer, 2003).

The pre-default yield spread $S_t(T)$ of the defaultable bond, which is the difference between the yield to maturity at time t of the defaultable zero-coupon bond and the one of a risk-free zero-coupon bond, which have the same maturity T, on $\{\tau > t\}$, is given by

$$S_t(T) = -\frac{1}{T-t} \ln \; D_t(T) = \lambda\psi(1) - \frac{1}{T-t} \ln \; \bar{Q}_t(T) \; . \tag{4.2.15}$$

Furthermore, when the process k is affine, this leads to

$$S_t(T) = \lambda\psi(1) - \frac{1}{T-t} \left[A_t(T) - B_t(T)k_t \right] \; . \tag{4.2.16}$$

For instance, we can consider the case where the process k is a CIR square root diffusion process, i.e., k verifies

$$dk_t = \gamma(\theta - k_t)dt + \sigma\sqrt{k_t}dB_t, \qquad k_0 = x,$$

where γ, θ, and σ are positive parameters. Then the quantities $A_t(T)$ and $B_t(T)$ are given by (see Maghsoodi, 1996): $A_t(T) = 2\frac{\gamma\theta}{\sigma^2} \ln \left(\frac{2h \; e^{\frac{1}{2}(\gamma+h)(T-t)}}{h-\gamma+e^{h(T-t)}(h+\gamma)} \right)$ and $B_t(T) = \frac{-2(e^{h(T-t)}-1)}{h-\gamma+e^{h(T-t)}(h+\gamma)}$, where $h = \sqrt{\gamma^2 + 2\sigma^2}$.

Comments 49 *We note that, under this model (where the jumps of K constitute a compound Poisson process and the continuous part is adapted with respect to a Brownian motion, independent of the jump part), the pre-default price of a DB with null recovery is continuous. Note that the pre-default price of the defaultable bond is decreasing w.r.t. the intensity of the Poisson process N. Furthermore, one observes, for any $0 \le t \le T$, an additional positive value in the spread of the defaultable bond which equals to $\lambda\psi(1)$. This shows the impacts of the jumps of K (triggered by the economic shocks) on both the pre-default price and spread of the defaultable bond.*

4.2.2.2   Example of shot-noise processes

Let X a shot noise process defined as in 3.3.21 of Chapter 3 with ν deterministic.   One considers the filtration $\mathbb{F}^W$ generated by a Brownian motion W independent of X and $\mathbb{F}^X$ the filtration generated by the shot noise process X. We define the reference filtration $\mathbb{F}$ as the one generated by X and W.

We consider

$$K_t = \int_0^t k_s ds + X_t \tag{4.2.17}$$

where k is an $\mathbb{F}^W$-adapted non-negative process.

Proposition 50 *Under the above hypotheses, the pre-default price $D_t(T)$ of the defaultable bond with zero-recovery is given by*

$$D_t(T) = \exp\left(\int_t^T \int_{\mathbb{R}} (e^{-G(T-s,x)} - 1)\nu(ds, dx) - \int_0^t \int_{\mathbb{R}} [G(T-s,x) - G(t-s,x)]\mu(ds,dx) \right) \bar{Q}_t(T),$$
$$\tag{4.2.18}$$

with $\bar{Q}_t(T) := \mathbb{E}\left[e^{-\int_t^T k_s ds} | \mathcal{F}_t \right]$. Note that $Q(T)$ is continuous.

PROOF: As in the proof of Proposition 48, one easily obtains

$$D_t(T) = \mathbb{E}[e^{-X_T + X_t} | \mathcal{F}_t]\bar{Q}_t(T) = e^{X_t}\mathbb{E}[e^{-X_T} | \mathcal{F}_t]\bar{Q}_t(T).$$

The result follows by using Proposition 43 where one has

$$\mathbb{E}[e^{-X_T} | \mathcal{F}_t] = \exp\left(\int_t^T \int_{\mathbb{R}} (e^{-G(T-s,x)} - 1)\nu(ds,dx) - \int_0^t \int_{\mathbb{R}} G(T-s, \; x)\mu(ds,dx) \right).$$

□

Markovian shot-noise   In the particular case where X has the following form

$$X_t = \sum_{i=0}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}} \gamma_i e^{\alpha(t-\tau_i)} \tag{4.2.19}$$

with (γ_i) non-negative random variables and $\alpha > 0$, one obtains

$$D_t(T) = \exp\left(\int_t^T \int_{\mathbb{R}^+} (e^{-xe^{\alpha(T-s)}} - 1)\nu(ds, dx) - (e^{\alpha(T-t)} - 1)X_t\right) \bar{Q}_t(T) . \tag{4.2.20}$$

This is due to the fact that $G(T - s, x) := xe^{\alpha(T-s)} = e^{\alpha(T-t)}G(t - s, x)$. Furthermore, $D(T)$ has negative jump size at $(\tau_i)_i$ given by

$$\Delta D_{\tau_i}(T)\mathbb{1}_{\{\tau_i \leq T\}} = D_{\tau_i-}(T)\left[\exp\left(-\left(e^{\alpha(T-\tau_i)} - 1\right)\gamma_i\right) - 1\right]\mathbb{1}_{\{\tau_i \leq T\}}.$$

The pre-default bond yield, denoted by $S_t(T)$ is then given by

$$S_t(T) := -\frac{1}{T-t}\left(\int_t^T \int_{\mathbb{R}^+} (e^{-xe^{\alpha(T-s)}} - 1)\nu(ds, dx) - (e^{\alpha(T-t)} - 1)X_t + \ln(\bar{Q}_t(T))\right) \tag{4.2.21}$$

and has a positive jump's size at time τ_i given on $\{\tau_i < T\}$ by

$$\Delta S_{\tau_i}(T)\mathbb{1}_{\{\tau_i \leq T\}} = \frac{1}{T-\tau_i}(e^{\alpha(T-\tau_i)} - 1)\Delta X_{\tau_i}\mathbb{1}_{\{\tau_i \leq T\}} = \frac{1}{T-\tau_i}(e^{\alpha(T-\tau_i)} - 1)\gamma_i\mathbb{1}_{\{\tau_i \leq T\}}. \tag{4.2.22}$$

Hence the jumps of K imply some negative jumps of the pre-default price of the defaultable bond and some positive jumps of its pre-default yield spread both at the same jump times of K.

Remark 51 *Note, from (4.2.20), that if $(\gamma_i, \tau_i)_i$ is associated to an increasing compound Poisson process (hence $\nu(dt, dx) = \lambda F(dx)dt$, where F is the law of γ_1) and $k = 0$, between two jumps, i.e., on $\tau_i < t < \tau_{i+1}$, the process $D(T)$ defined as*

$$D_t(T) = \exp\left(\int_t^T \int_{\mathbb{R}^+} (e^{-xe^{\alpha(T-s)}} - 1)\lambda F(dx)ds - (e^{\alpha(T-t)} - 1)(X_{\tau_i} + \gamma_i e^{\alpha(t-\tau_i)})\right)$$

is increasing.
Indeed since $e^{-xe^{\alpha(T-t)}} - 1$ is non-positive, $\int_t^T \int_{\mathbb{R}^+} (e^{-xe^{\alpha(T-s)}} - 1)\lambda F(dx)ds$ is increasing in t and

$$(e^{\alpha(T-t)} - 1)(X_{\tau_i} + \gamma_i e^{\alpha(t-\tau_i)}) = (e^{\alpha(T-t)} - 1)X_{\tau_i} + \gamma_i e^{\alpha(T-\tau_i)} - \gamma_i e^{\alpha(t-\tau_i)}$$

and $(e^{\alpha(T-t)} - 1)X_{\tau_i}$ as well as $-\gamma_i e^{\alpha(t-\tau_i)}$ are decreasing with respect to t.

4.2.3   Valuation of the CDS under economic shocks

The same kind of computations as in the case of the defaultable bond allows to obtain the pre-default price of two legs of the CDS.

4.2.3.1 The economic shock times are predictable

When the default time of the reference entity follows a generalized Cox model with K defined as in (4.2.1), with the assumption that the jump times of K are predictable, the value of the default payment leg at time t is given by

$$D_t^{leg} = \mathbb{1}_{\{\tau > t\}} \mathbb{E}[\delta_\tau \mathbb{1}_{\{\tau \leq T\}} | \mathcal{G}_t] = \mathbb{1}_{\{\tau > t\}} \frac{\mathbb{E}\left[\int_t^T \delta_u \alpha_T(u) du + \sum_{i=0}^\infty \delta_{\tau_i} p_{\tau_i}^i \mathbb{1}_{\{t < \tau_i \leq T\}} | \mathcal{F}_t\right]}{Z_t}.$$

This computation is the same as in (4.2.9) of Proposition 47, where δ plays the rule of h. The price of the premium leg at time t is given, on $\{\tau > t\}$, by

$$P_t^{leg} = \mathbb{1}_{\{\tau > t\}} \mathbb{E}[\int_t^T \mathbb{1}_{\{u < \tau\}} \kappa du | \mathcal{G}_t] = \mathbb{1}_{\{\tau > t\}} \frac{\mathbb{E}\left[\kappa \int_t^T Z_u du | \mathcal{F}_t\right]}{Z_t}.$$

The spread κ of the CDS is the one which verifies $P_0^{leg} = D_0^{leg}$ and its expression is given by

$$\kappa = \frac{\mathbb{E}\left[\int_0^T \delta_u \alpha_T(u) du + \sum_{i=0}^\infty \delta_{\tau_i} p_{\tau_i}^i \mathbb{1}_{\{\tau_i \leq T\}}\right]}{\mathbb{E}\left[\int_0^T Z_u du\right]}.$$

4.2.3.2 CDS price under surprising economic shocks: example of the compound Poisson process

Whenever the process K is given by (4.2.11) the price of the default payment leg at time t is given by

$$D_t^{leg} = \mathbb{1}_{\{\tau > t\}} \frac{1}{Z_t} \mathbb{E}\left[\int_t^T (k_u + \lambda \psi(1)) \delta_u Z_u du | \mathcal{F}_t\right]$$

whereas the value of the premium leg at time t is given by

$$P_t^{leg} = \mathbb{1}_{\{\tau > t\}} \frac{\mathbb{E}\left[\kappa \int_t^T Z_u du | \mathcal{F}_t\right]}{Z_t} = \mathbb{1}_{\{\tau > t\}} \frac{\mathbb{E}\left[\kappa \int_t^T e^{-X_u} e^{-\int_0^u k_s ds} du | \mathcal{F}_t\right]}{Z_t}.$$

By using the Tower property one has

$$P_t^{leg} = \mathbb{1}_{\{\tau > t\}} \frac{\kappa \int_t^T \mathbb{E}\left[e^{-\int_0^u k_s ds} \mathbb{E}\left[e^{-X_u} | \mathcal{F}_u^W \vee \mathcal{F}_t^X\right] | \mathcal{F}_t\right] du}{Z_t}.$$

The fact that X is independent of $\mathbb{F}^W$ and $\mathcal{F}_t^X \subset \mathcal{F}_t$ implies

$$P_t^{leg} = \mathbb{1}_{\{\tau > t\}} \frac{\kappa \int_t^T \mathbb{E}\left[e^{-\int_0^u k_s ds} | \mathcal{F}_t\right] \mathbb{E}\left[e^{-X_u} | \mathcal{F}_t^X\right] du}{Z_t}$$

Using the independence property of W and $\mathbb{F}^X$, we obtain

$$\begin{aligned}
P_t^{leg} &= \mathbb{1}_{\{\tau > t\}} \frac{\kappa \int_t^T \mathbb{E}\left[e^{-X_u} | \mathcal{F}_t^X\right] \mathbb{E}\left[e^{-\int_0^u k_s ds} | \mathcal{F}_t^W\right] du}{Z_t} \\
&= \mathbb{1}_{\{\tau > t\}} \frac{\kappa \int_t^T n_t e^{-u\psi(1)} \mathbb{E}\left[e^{-\int_0^u k_s ds} | \mathcal{F}_t^W\right] du}{Z_t} \\
&= \mathbb{1}_{\{\tau > t\}} \frac{\kappa \int_t^T e^{-u\psi(1)} y_t(u) du}{e^{-t\psi(1)}}
\end{aligned}$$

where $y_t(u) = \mathbb{E}\left[e^{-\int_0^u k_s ds}|\mathcal{F}_t^W\right]$. Therefore, the spread κ can be computed from $P_0^{leg} = D_0^{leg}$ which leads to

$$\kappa = \frac{\mathbb{E}\left[\int_0^T (k_u + \lambda\psi(1))\delta_u Z_u du\right]}{\int_0^T e^{-u\psi(1)}y_0(u)du} \ . \tag{4.2.23}$$

4.2.4   General framework of dynamics of defaultable Bond

In this subsection, we will be interested in the dynamics of the pre-default prices of defaultable bond with zero recovery, in a more general case, when the default time follows a generalized Cox time under which the process K is càdlàg. One knows that for any maturity T, the pre-default price $D_t(T)$ at time $t < T$ has the following form

$$D_t(T) = \frac{\mathbb{E}[Z_T|\mathcal{F}_t]}{Z_t} =: Y_t(T)e^{K_t}$$

where $Y(T)$ is the $\mathbb{F}$-martingale defined by $Y_t(T) := \mathbb{E}[Z_T|\mathcal{F}_t]$.

When there is no ambiguity, we delete T in the notation (i.e., we denote for T fixed Y_t instead of $Y_t(T)$, for $t < T$).

By Itô's integration by parts, one has, for $0 \leq t < T$,

$$\begin{aligned}
dD_t(T) &= Y_{t-}de^{K_t} + e^{K_{t-}}dY_t + d[Y, e^K]_t \\
&= Y_{t-}e^{K_{t-}}\left(dK_t^c + dJ_t\right) + e^{K_{t-}}dY_t + d[Y, e^K]_t
\end{aligned}$$

where J is the increasing process (hence a submartingale with a Doob-Meyer decomposition) $J_t = \sum_{s \leq t}(e^{\Delta K_s} - 1) = M_t^J + A_t^J$. The covariation process $[Y, e^K]_t = \int_0^t \Delta Y_s de^{K_s}$ is a semimartingale, that we assume to be special with decomposition $[Y, e^K] = A^* + M^*$ where M^* is a local martingale and A^* a predictable process.

Therefore, the local martingale part $M^D(T)$ of $D(T)$ is given by

$$M_t^D(T) = \int_0^t Y_{s-}e^{K_{s-}}dM_s^J + \int_0^t e^{K_{s-}}dY_s + M_t^* \tag{4.2.24}$$

and its predictable bounded variation part $A^D(T)$ is given by

$$A_t^D(T) = \int_0^t Y_{s-}e^{K_{s-}}\left(dK_s^c + dA_s^J\right) + A_t^* \ . \tag{4.2.25}$$

4.2.4.1   Examples

In this subsection, we give some particular cases which lead to more explicit formulae. In addition, we analyse the impacts of the jumps times of the generalized Cox process K to the prices.

Example 52 *If $\mathbb{F}$ is a continuous filtration, Y is continuous, hence $[Y, e^K] = 0$. Furthermore, since optional processes are predictable[1] , J is predictable and $M^J = 0, A^J = J$. Hence,*

$$dD_t(T) = Y_t de^{K_t} + e^{K_{t-}}dY_t = Y_t e^{K_{t-}}\left(dK_t^c + dJ_t\right) + e^{K_{t-}}dY_t = Y_t e^{K_t}d\Lambda_t + e^{K_{t-}}dY_t \tag{4.2.26}$$

and the special semimartingale $D(T)$ has a (continuous) local martingale part

$$M_t^D(T) = \int_0^t e^{K_{s-}}dY_s, \ \ \forall \, 0 \leq t < T$$

---

[1]In a continuous filtration, the two σ-algebra $\mathcal{O}(\mathbb{H})$ and $\mathcal{P}(\mathbb{H})$ are equal (see (Jeanblanc, Yor, and Chesney, 2009, Page 512)).

and a predictable part

$$A_t^D(T) = \int_0^t Y_s e^{K_s-} dK_s^c + \int_0^t Y_s e^{K_s-} dA_s^J, \quad \forall\, 0 \leq t < T.$$

It is clear, from (4.2.26), that the predefault-price $D(T)$ has predictable positive jumps at the jumps times $(\tau_i)_i$ of K which are given by

$$\Delta D_{\tau_i}(T) \mathbb{1}_{\{\tau_i < T\}} = D_{\tau_i-}(T) \Delta A_{\tau_i}^J \mathbb{1}_{\{\tau_i < T\}} = D_{\tau_i-}(T)(e^{\Delta K_{\tau_i}} - 1)\mathbb{1}_{\{\tau_i < T\}}.$$

Example 53 *If K is predictable, $M^J \equiv 0$ and $A^J = J$, and by Yoeurp's lemma $[Y, e^K]_t = \int_0^t \Delta e^{K_s} dY_s$ which is the local martingale that we have denoted M^*. Then $e^{K_s-} dY_s + dM_s^* = (e^{K_s-} + \Delta e^{K_s})dY_s = e^{K_s} dY_s$ and the local martingale part of $D_t(T)$ is*

$$M_t^D(T) = \int_0^t e^{K_s} dY_s, \quad \forall\, 0 \leq t < T. \tag{4.2.27}$$

The predictable bounded variation part is

$$A_t^D(T) = \int_0^t Y_{s-} e^{K_s-} \big(dK_s^c + dJ_s\big), \quad \forall\, 0 \leq t < T. \tag{4.2.28}$$

Note that, for any $0 \leq t < T$,

$$e^{K_t-} \big(dK_t^c + dJ_t\big) = e^{K_t} dK_t^c + e^{K_t} dI_t = e^{K_t} d\Lambda_t$$

so that the predictable part is

$$A_t^D(T) = \int_0^t Y_{s-} e^{K_s} d\Lambda_s, \quad \forall\, 0 \leq t < T.$$

In this case, the predefault price $D(T)$ has jumps at the jump times of K with sizes given, for all $0 \leq t < T$, by

$$\Delta D_{\tau_i}(T) \mathbb{1}_{\{\tau_i < T\}} = D_{\tau_i-}(T)\big(e^{\Delta K_{\tau_i}} \frac{Y_{\tau_i}}{Y_{\tau_i-}} - 1\big)\mathbb{1}_{\{\tau_i < T\}}.$$

Example 54 *If $\mathbb{F}$ is generated by a Brownian motion W (hence a continuous filtration) and K predictable, using the fact that $Y(T)$ is a strictly positive $\mathbb{F}$-local martingale which can be written, thanks to the martingale representation theorem, as $dY_t(T) = \phi_t(T)Y_t(T)dW_t$, for an $\mathbb{F}$-predictable process $\phi(T)$, one obtains*

$$dD_t(T) = D_t(T)\left(d\Lambda_t + \phi_t(T)dW_t\right). \tag{4.2.29}$$

We recover the result of Bélanger, Shreve, and Wong, 2004, formula 5.3, Theorem 5.1 in the case of zero recovery.

Example 55 *We assume to be in the case of shot-noise model presented in Subsection 4.2.2.2, and ν is deterministic. From the results of Subsection 4.2.2.2, it follows that if $k = 0$ then*

$$D_t(T) = \exp\left(\int_t^T \int_{\mathbb{R}} (e^{-G(T-s,x)} - 1)\nu(ds, dx) - \int_0^t \int_{\mathbb{R}} [G(T-s,x) - G(t-s,x)]\mu(ds, dx)\right).$$

The dynamics of $D(T)$ can be deduced by Itô-Ventcell's formula. We prefer to make use of the general results presented in the first part of this section, with $Y_t = c(T)L_t(T)$, where $c(T) = \exp\left(\int_0^T \int_{\mathbb{R}} (e^{-G(T-s,x)} - 1)\nu(ds, dx)\right)$

and

$$L_t(T) = \exp\left(-\int_0^t \int_{\mathbb{R}} G(T-s, x)\mu(ds, dx) - \int_0^t \int_{\mathbb{R}} (e^{-G(T-s, x)} - 1)\nu(ds, dx)\right).$$

The jumps of Y and e^K occur at times τ_i. One has, from (3.3.28)

$$\begin{aligned}
\Delta Y_{\tau_i} &= Y_{\tau_i-}(e^{-G(T-\tau_i, \gamma_i)} - 1) \\
\Delta(e^K)_{\tau_i} &= e^{K_{\tau_i-}}(e^{G(\tau_i - \tau_i, \gamma_i)} - 1) = e^{K_{\tau_i-}}(e^{G(0, \gamma_i)} - 1).
\end{aligned}$$

Hence

$$\begin{aligned}
[Y, e^K]_t &= \sum_{s \leq t} \Delta Y_s \Delta(e^K)_s = \int_0^t Y_{s-} e^{K_{s-}} \int_{\mathbb{R}} (e^{-G(T-s, x)} - 1)(e^{G(0, x)} - 1)\mu(ds, dx) \\
&= M_t^* + \int_0^t Y_{s-} e^{K_{s-}} \int_{\mathbb{R}} (e^{-G(T-s, x)} - 1)(e^{G(0, x)} - 1)\nu(ds, dx). \qquad (4.2.30)
\end{aligned}$$

As in section 3.3.5.5 for the computation of A^I, we obtain

$$A_t^J = \int_{s=0}^t \int_{\mathbb{R}} (e^{G(0, x)} - 1)\nu(ds, dx), \forall t \geq 0,$$

Therefore, the local martingale part of $D(T)$ is

$$M_t^D(T) = \int_0^t Y_{s-} e^{K_{s-}} dM_s^J + \int_0^t e^{K_{s-}} dY_s + M_t^*$$

where

$$\begin{aligned}
M_t^* &= \int_0^t \int_{\mathbb{R}} Y_{s-} e^{K_{s-}} (e^{-G(T-s, x)} - 1)(e^{G(0, x)} - 1)\widetilde{\mu}(ds, dx), \forall t \geq 0 \\
M_t^J &= \int_0^t \int_{\mathbb{R}} (e^{G(0, x)} - 1)\widetilde{\mu}(ds, dx), \forall t \geq 0
\end{aligned}$$

so that the local martingale part reduces, after simple computation, to

$$M_t^D(T) = \int_0^t \int_{\mathbb{R}} Y_{s-} e^{K_{s-}} \left(e^{-(G(T-s, x) - G(0, x))} - 1\right)\widetilde{\mu}(ds, dx) \qquad (4.2.31)$$

and the predictable bounded variation part is

$$A_t^D(T) = \int_0^t \int_{\mathbb{R}} Y_{s-} e^{K_{s-}} e^{-G(T-s, x)} (e^{G(0, x)} - 1)\nu(ds, dx).$$

In this case the predefault price $D(T)$ admits negative jumps at the jump times $(\tau_i)_i$ of K which are given by

$$\Delta D_{\tau_i}(T)\mathbb{1}_{\{\tau_i < T\}} = D_{\tau_i-}(T)\left(e^{G(0, \gamma_i) - G(T-\tau_i, \gamma_i)} - 1\right)\mathbb{1}_{\{\tau_i < T\}} \qquad (4.2.32)$$

which can be directly obtained from (4.2.31).

Example 56 *In the case of the Compound Poisson process and a Brownian filtration as given subsection 4.2.2.1, one has $D_t(T)$ which is given by (4.2.14) that shows that $D(T)$ is a continuous process. Setting $y_t(T) = \mathbb{E}[\exp\left(-\int_0^T k_s ds\right)|\mathcal{F}_t^W]$ and $\eta_t(T) = e^{-(T-t)\lambda\psi(-1)}\exp\left(\int_0^t k_s ds\right)$, the dynamics of $D_t(T)$ is*

$$dD_t(T) = \eta_t(T)dy_t(T) + y_t(T)\eta_t(T)(\lambda\psi(-1) + k_t)dt$$

where $y(T)$ is a continuous $\mathbb{F}$-martingale.

As a remark, we can note that it is not difficult to obtain (up to K^c) the same form for the dynamics of $D(T)$ (with longer computation) using the previous equalities. Indeed, as noticed previously, the CCP correspond to the case when the shot noise is obtained from $G(t, x) = x$. Then, $Y_t = e^{-X_t - (T-t)\lambda\psi(-1)}\mathbb{E}[\exp\big(-\int_0^T k_s ds\big)|\mathcal{F}_t^W]$ and we obtain

$$[Y, e^K]_t = \sum_{s \leq t} \Delta Y_s \Delta(e^K)_s = M_t^* + \int_0^t Y_{s-}e^{K_{s-}}\int_{\mathbb{R}}(2 - e^x - e^{-x})\lambda F(dx)ds.$$

4.2.4.2 Illustrative example of the shot-noise case

Here, we present the simulation results of the generalized Cox model in the case of Markovian shot-noise introduced in 4.2.2.2. This choice is due to the fact that the Markovian shot-noise processes are preferable for the computational perspective thanks to their analytical tractability.

The simulation of the particular case (where $(\gamma_i)_i$ are i.i.d) of the process given in (4.2.19) is done by using the standard Algorithm 2.1 of Scherer, Schmid, and Schmidt (2012). In Figure 4.1, we illustrate one trajectory of K where, as in (4.2.19), $K_t = \sum_{i=0}^{\infty} \mathbb{1}_{\{\tau_i \leq t\}}\gamma_i e^{\alpha(t-\tau_i)}$, in between 0 and $T = 6$ years, with i.i.d exponential jumps $(\gamma_i)_i$ of parameter $\gamma = 5$, with parameter $\alpha = 0.5$ and where the jumps times τ_i are jumps of a Poisson process N independent of the $(\gamma_i)_i$ with intensity $\lambda = 0.4$. One can observe three jumps of the shot-noise at times $\tau_1 = 2.028$, $\tau_2 = 2.994$ and $\tau_3 = 4.254$ with respectively sizes $\gamma_1 = 0.045$, $\gamma_2 = 0.014$, and $\gamma_3 = 0.094$. Figure 4.2 shows the corresponding time-varying Bond prices computed using (4.2.20) where $k = 0$ (with $\nu(ds, dx) = \lambda F(dx)ds$, where F is the law of γ_1). Unsurprisingly, one observes three negative jumps of the prices at the same jump times of the shot-noise. In addition, the bond price is increasing between the jumps times as we have shown in Subsection 4.2.2.2.

We end this example by investigating the survival probability of the default time τ which can be computed by using (3.3.29). Figure 4.3 shows this survival probability for different values of the the decay rate α, and different values of the intensity λ of the Poisson process N. It is clear that the survival probability is decreasing with respect to the intensity λ of N. However the speed of the degrowth increases with respect to the decay rate α.
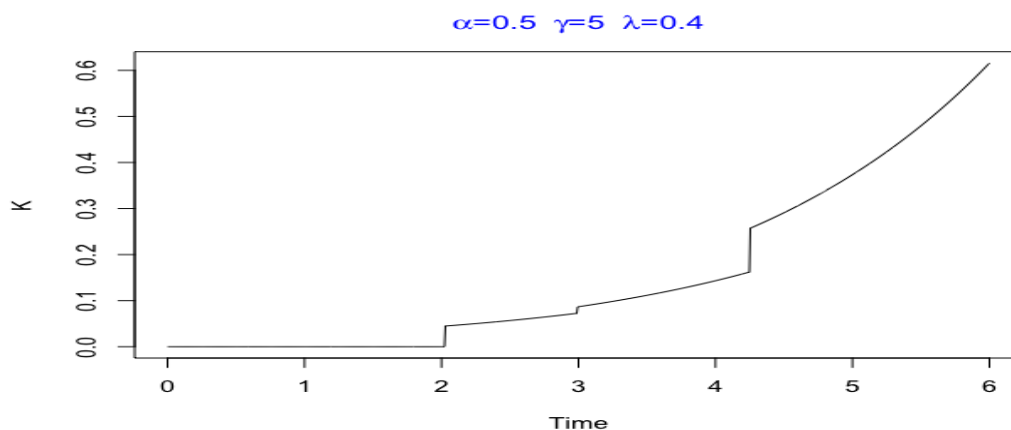

FIGURE 4.1: One path of K which is given as in the form (4.2.19)
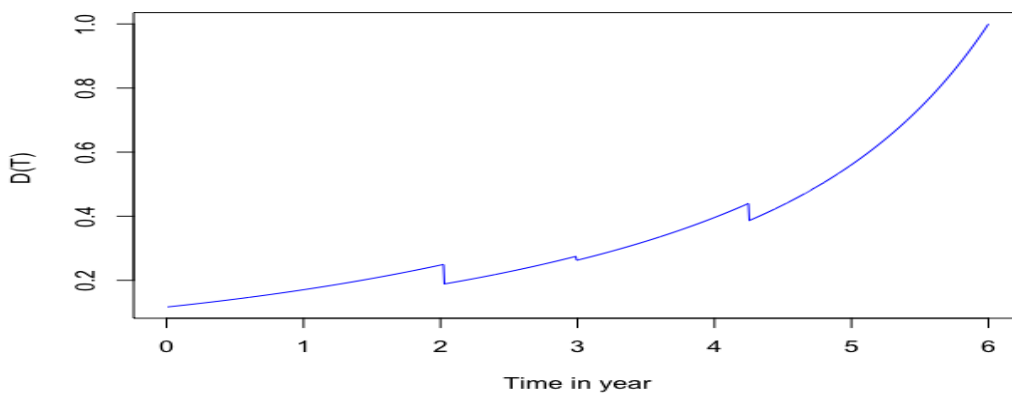
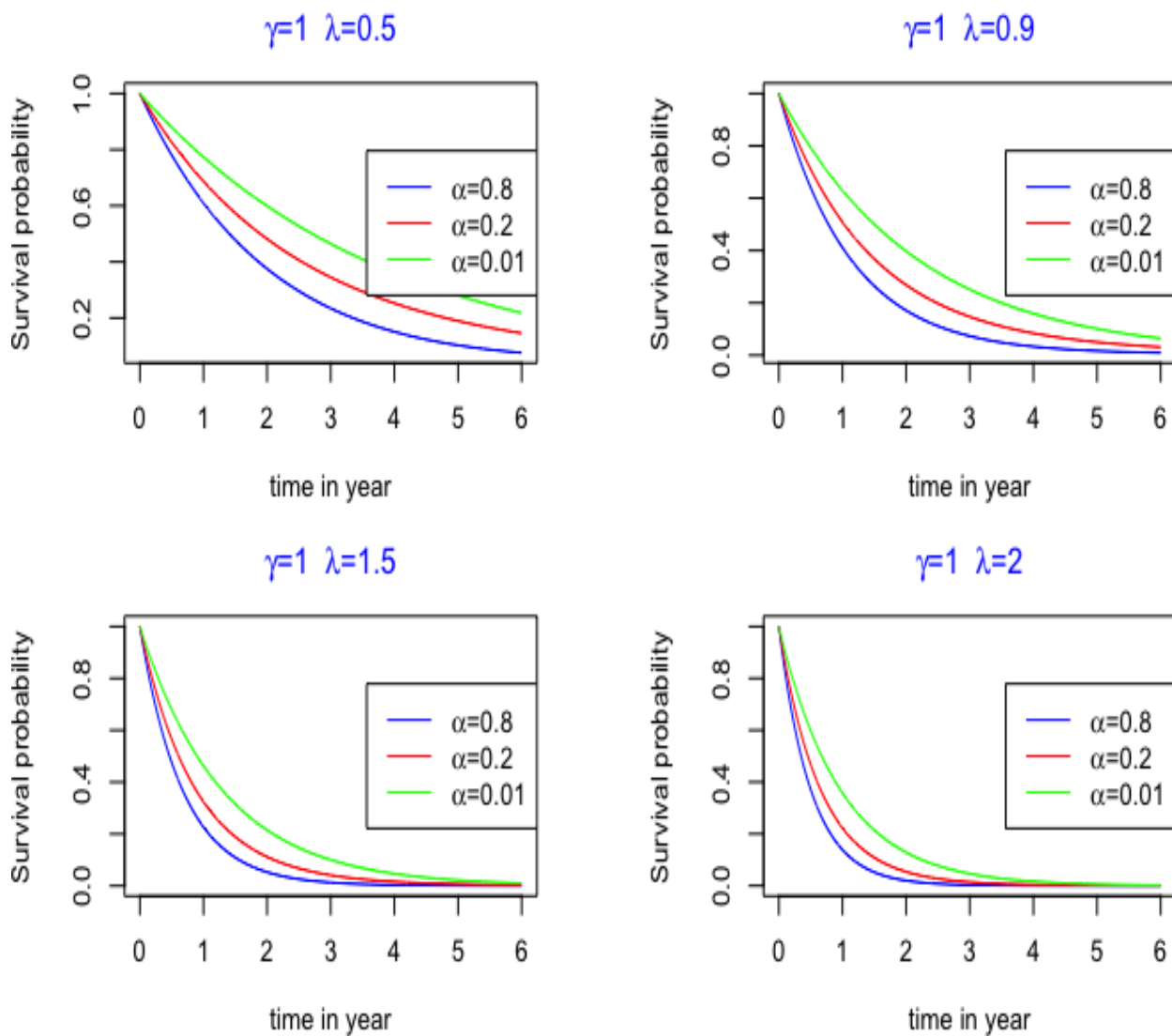FIGURE 4.2: Predefault bond price for the corresponding path of the Fig 4.1

FIGURE 4.3: Survival probability of τ for different values of α with different values of λ.

Part II

Kriging for volatility surfaces construction

Chapter 5

An introduction to kriging

In this chapter, we first, briefly recall some known facts about classical kriging before introducing the constrained kriging approach which extends the classical one in presence of some linear inequality constraints (such as monotonicity, bounded, convexity,...). More details about classical kriging can be found in Williams and Rasmussen (2006). Among the existing approaches for constrained kriging, we adopt the technique from Maatouk and Bay (2014) which is based on a finite dimensional approximation of the Gaussian processes prior. Note that this approach has been extended by López-Lopera et al. (2018) to more general linear constraints. However, in this thesis, we limit ourselves to convexity and monotonicity constraints that can be treated using the model of Maatouk and Bay (2014). A particular interest is focused on the simulation of the Gaussian random coefficients when some monotonicity constraints are saturated (see Subsection 5.2.4.2), a problem that is frequent in practice. Indeed this saturation leads to problems of efficient samples when using the improved rejection sampling (the RMS algorithm) developed by Maatouk and Bay (2016). As such, we consider an adapted basis function grid for allowing the efficiency of this sample method. We illustrate this problem through an example.

5.1 Classical kriging

Assume a function f of d-dimensional variables $\mathbf{x} \in \mathbb{R}^d$ to be known only at some input locations $\mathbf{x}_1, ..., \mathbf{x}_n$, with $\mathbf{x}_i \in \mathbb{R}^d$, for $i = 1, ..., n$. In order to estimate the function f in the entire input domain, one can use kriging (also known as Gaussian process regression) which consists in considering the unknown function f as a realization of a Gaussian process (GPs) $(Y(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d)$ and estimating the conditional process $Y|Y(\mathbf{x}_1) = f(\mathbf{x}_1), ..., Y(\mathbf{x}_n) = f(\mathbf{x}_n)$. The GPs $(Y(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d)$ is a collection of random variables, any finite number of which have (consistent) joint Gaussian distributions (see Williams and Rasmussen, 2006) and is characterized by its mean function

$$\mu : \mathbf{x} \in \mathbb{R}^d \longrightarrow \mathbb{E}(Y(\mathbf{x})) \in \mathbb{R}$$

and its covariance function (also called kernel)

$$\mathcal{K} : (\mathbf{x}, \mathbf{x}') \in \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \text{Cov}(Y(\mathbf{x}), Y(\mathbf{x}')) \in \mathbb{R}.$$

We consider the covariance function $\mathcal{K}$ to be defined such that Y has almost surely continuous and differentiable paths. For instance, in the empirical investigation, we consider the d-dimensional isotropic covariance kernel given, for any $\mathbf{x} = (x^1, ..., x^d)$ and $\mathbf{x}' = (x'^1, ..., x'^d)$ as

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{i=1}^{d} R_i(x^i - x'^i, \theta_{x^i})$$

where $\theta = (\theta_{x^1}, ..., \theta_{x^d}) \in \mathbb{R}^d$ and σ are respectively the length scale and the variance hyper-parameters of the kernel function $\mathcal{K}$ and the functions $(R_i)_i$ are kernel correlation functions. These hyper-parameters can be either specified (based on expert knowledge) or estimated by using some existing methods in the literature (see Subsection 5.1.2). Some usual kernel correlation functions are presented in Table 5.1, they are ranked by degree of regularity.

Note that changing the kernel $\mathcal{K}$ means changing the initial belief on f (i.e., the prior). Figure 5.1 shows how the choice of the kernel affects the paths of the GPs.

1D kriging kernel	$\mathcal{K}(x, x')$	Class
Gaussian	$\sigma^2 \exp\left(-\frac{(x-x')^2}{2\theta^2}\right)$	$\mathcal{C}^\infty$
Matérn 5/2	$\sigma^2 \left(1 + \frac{\sqrt{5}|x-x'|}{\theta} + \frac{5(x-x')^2}{3\theta^2}\right) \exp\left(-\frac{\sqrt{5}|x-x'|}{\theta}\right)$	$\mathcal{C}^2$
Matérn 3/2	$\sigma^2 \left(1 + \frac{\sqrt{3}|x-x'|}{\theta}\right) \exp\left(-\frac{\sqrt{3}|x-x'|}{\theta}\right)$	$\mathcal{C}^1$
Exponential	$\sigma^2 \exp\left(-\frac{|x-x'|}{\theta}\right)$	$\mathcal{C}^0$

TABLE 5.1: Some usual kernel correlation functions in kriging techniques.
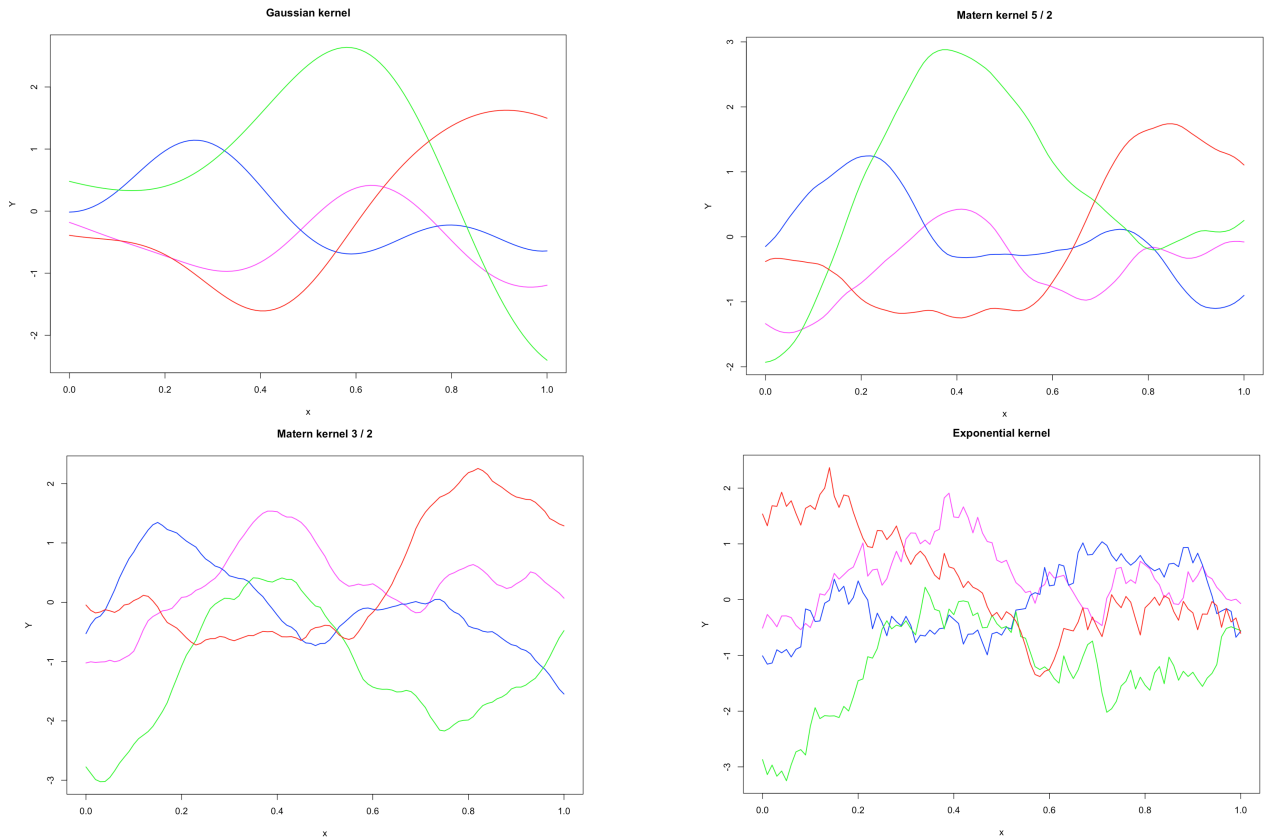
FIGURE 5.1: Sample paths of the Gaussian prior Y for different choices of kernel function with parameters $\sigma^2 = 1$ and $\theta = 0.2$.

By denoting $\boldsymbol{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top \in \mathbb{R}^{n \times d}$ a d-dimensional design points, $\boldsymbol{y} = [y_1, \ldots, y_n]^\top \in \mathbb{R}^n$, the observed values of f at these points (i.e, $y_i = f(\mathbf{x}_i)$, for $i = 1, ..., n$) and $Y(\boldsymbol{X}) = (Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_n))^\top$ the vector composed of Y at $\boldsymbol{X}$, the conditional process (called the posterior process) $Y|Y(\boldsymbol{X}) = \boldsymbol{y}$ is a GPs with mean function (see Williams and Rasmussen, 2006)

$$\eta(\mathbf{x}) = \mu(\mathbf{x}) + \boldsymbol{c}(\mathbf{x})^\top \boldsymbol{C}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}), \quad \mathbf{x} \in \mathbb{R}^d \tag{5.1.1}$$

and covariance function $\mathcal{K}^\star$ given by

$$\mathcal{K}^\star(\mathbf{x}, \mathbf{x}') = \mathcal{K}(\mathbf{x}, \mathbf{x}') - \boldsymbol{c}(\mathbf{x})^\top \boldsymbol{C}^{-1} \boldsymbol{c}(\mathbf{x}'), \quad \mathbf{x}, \mathbf{x}' \in \mathbb{R}^d \tag{5.1.2}$$

where $\boldsymbol{\mu} = \mu(\mathbf{X}) = [\mu(\mathbf{x}_1), \dots, \mu(\mathbf{x}_n)]^\top$, $\boldsymbol{C}$ is the covariance matrix of $Y(\boldsymbol{X})$ and $\boldsymbol{c}(\mathbf{x}) = [\mathcal{K}(\mathbf{x}, \mathbf{x}_1), \dots, \mathcal{K}(\mathbf{x}, \mathbf{x}_n)]^\top$.

Without consideration of any inequality constraint on the function f, kriging prediction and uncertainty quantification are made using the posterior process $Y|Y(\boldsymbol{X}) = \boldsymbol{y}$. This is known as classical kriging (or unconstrained kriging). In this case, the Best Linear Unbiased Estimator of $Y(x)$ is the posterior mean function (5.1.1) which is also called (unconstrained) kriging mean (see Jones, Schonlau, and Welch, 1998). We can then use the conditional covariance function (5.1.2) to obtain confidence bands arround the predicted function. The hyperparmeters of the kernel function $\mathcal{K}$ can be either specified or estimated using, for instance, the maximum likelihood estimator (MLE) (see Subsection 5.1.2).

5.1.1   Presence of noise :

The previous framework can be extended to the case when $y_i = f(\mathbf{x}_i) + \varepsilon_i$ for $i = 1, \dots n$, where ε_i is assumed to be a zero-mean Gaussian random variable with homoscedastic variance ς^2 and ε_i independent from Y. In this setting, the conditional process $Y|Y(\boldsymbol{X}) + \boldsymbol{\varepsilon} = \boldsymbol{y}$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top \in \mathbb{R}^n$ is a zero-mean Gaussian vector with covariance matrix $\varsigma^2 I_n$ (I_n being the identity matrix of dimension n), is still a GPs with mean function (see Williams and Rasmussen, 2006)

$$\eta(\mathbf{x}) = \mu(\mathbf{x}) + \boldsymbol{c}(\mathbf{x})^\top \left(\boldsymbol{C} + \varsigma^2 I_n\right)^{-1} (\boldsymbol{y} - \boldsymbol{\mu}), \quad \mathbf{x} \in \mathbb{R}^d$$

and covariance function $\mathcal{K}^\star$ given by

$$\mathcal{K}^\star(\mathbf{x}, \mathbf{x}') = \mathcal{K}(\mathbf{x}, \mathbf{x}') - \boldsymbol{c}(\mathbf{x})^\top \left(\boldsymbol{C} + \varsigma^2 I_n\right)^{-1} \boldsymbol{c}(\mathbf{x}'), \quad \mathbf{x}, \mathbf{x}' \in \mathbb{R}^d.$$

5.1.2   Hyper-parameters learning :

Hyper-parameters consist in the length scale θ and the variance parameter σ of the kernel function $\mathcal{K}$ as well as the noise parameter ς. The most commonly used methods for estimating them are cross validation (CV) (see, e.g., Bachoc, 2013; Zhang and Wang, 2010) and Maximum Likelihood Estimators (MLE) (see Jones, Schonlau, and Welch, 1998). We only present the MLE which will be used later (we shall say why in the section dealing with the constrained kriging).

By considering a Zero-mean Gaussian Y (i.e., $\mu = 0$) the MLE consists in maximizing the marginal log likelihood for Y with respect to the parameter $\lambda = [\theta, \sigma, \varsigma]^\top$. This log marginal likelihood has the following form (see Williams and Rasmussen, 2006, section 2.2):

$$\mathcal{L}(\lambda) = -\frac{1}{2}\boldsymbol{y}^\top (\boldsymbol{C} + \varsigma^2 I_n)^{-1} \boldsymbol{y} - \frac{1}{2}\log(\det(\boldsymbol{C} + \varsigma^2 I_n)) - \frac{n}{2}\log(2\pi) \tag{5.1.3}$$

where $\boldsymbol{y} = [Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)]^\top$ and $\boldsymbol{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$. Hence the vector of estimated parameters $\hat{\lambda}$ verifies :

$$\hat{\lambda} = \arg\max_\lambda \mathcal{L}(\lambda).$$

This can be solved by using, for example, the gradient ascent method.

5.2 Constrained kriging

In Machine Learning techniques, learning a function f of the data mapping when f is subjected to some inequality constraints is not an easy task most especially in Gaussian process regression. Suppose the unknown real function f is observed with some noises and such that f belongs to a set of inequality constraints $\mathcal{M}$, i.e.,

$$f : \mathcal{D} \to \mathbb{R}$$

such that

$$\begin{cases} f(\boldsymbol{X}) + \boldsymbol{\varepsilon} = \boldsymbol{y} \\ f \in \mathcal{M} \end{cases} \tag{5.2.1}$$

where $\mathcal{M}$ is a convex set of functions satisfying some shape properties.

For instance, $\mathcal{M}$ can be one of the following convex sets:

- $\mathcal{M}_0^d := \{ f \in \mathcal{C}(\mathbb{R}^d, \mathbb{R}) \mid y_{\min} \leq f(\mathbf{x}) \leq y_{\max}, \ \forall \mathbf{x} \in \mathcal{D} \}$

- $\mathcal{M}_1^1 := \{ f \in \mathcal{C}(\mathbb{R}, \mathbb{R}) \mid f \text{ is non-decreasing} \}$

- $\mathcal{M}_2^1 = \{ f \in \mathcal{C}(\mathbb{R}, \mathbb{R}) \mid f \text{ is convex} \}$

- $\mathcal{M}_{12}^2 = \{ f \in \mathcal{C}(\mathbb{R}^2, \mathbb{R}) \mid f \text{ is non-decreasing in } x \text{ and convex in } y \}$.

Because making a Gaussian prior in f leads to the fact that the posterior is no Gaussian anymore and the shape conditions are usually infinite-dimensional, we propose to use the finite dimensional approximation of the Gaussian prior on f (see Maatouk and Bay, 2014) for which the constraints can be imposed in the entire domain $\mathcal{D}$ with finite number of checks. In the next subsection, we present this approximation in one dimensional case. The two dimensional one can be directly obtained with little modifications of the one dimensional case and will be presented in Chapter 6.

5.2.1 Finite dimensional approximation of GPs in 1d

We consider the input domain $\mathcal{D}$ to be an interval in $\mathbb{R}$ of the form $\mathcal{D} = [\underline{x}, \overline{x}]$ of $\mathbb{R}$ and we make a Gaussian prior Y on f with covariance function $\mathcal{K}$ and (for simplicity) zero-mean. We discretize $\mathcal{D}$ on a regular subdivision $u_0 < \ldots < u_N$ with a constant mesh δ. For each u_i, we consider hat functions $\phi_i(x) := \max\left(1 - \frac{|x - u_i|}{\delta}, 0\right)$. The Gaussian process Y is approximated on $\mathcal{D}$ by the process Y^N given by (see Maatouk and Bay, 2014; Cousin, Maatouk, and Rullière, 2016)

$$Y^N(x) = \sum_{i=0}^{N} Y(u_i)\phi_i(x). \tag{5.2.2}$$

The interest of using piecewise linear basis function is the fact that Y^N is a piecewise and the inequality constraints are satisfied everywhere in the input domain whenever they are satisfied at the knots.

Let $\Phi(x)$ denote the vector of size $N + 1$ given by $\Phi(x) = (\phi_0(x), \ldots, \phi_N(x))^\top$ and let $\boldsymbol{\Phi}(\mathbf{X})$ denote the $n \times (N + 1)$ matrix of basis functions in which each row l corresponds to the vector $\Phi(x_l)$.

Proposition 57 *(see Maatouk and Bay, 2014)*
The following results hold.

- *The finite-dimensional process $Y^N(\cdot) = \sum_{i=0}^{N} Y(u_i)\phi_i(\cdot)$ uniformly converges to Y on $\mathcal{D}$ as $N \to \infty$, almost surely.*

- *$Y^N(x) = \boldsymbol{\Phi}(\mathbf{X})\boldsymbol{\xi}$ where $\boldsymbol{\xi} := (Y(u_0), \ldots, Y(u_N))^\top$ is a zero-mean Gaussian vector with covariance matrix Γ^N such that $\Gamma_{i,j}^N = \mathcal{K}(u_i, u_j)$, for $i, j = 1, \ldots, N$.*

Some shape-preserving conditions :

1) Y^N takes values on $[y_{\min}, y_{\max}]$ if and only if $y_{\min} \leq \xi_i \leq y_{\max}$, $i = 0, \ldots, N$

2) Y^N is non-decreasing on $\mathcal{D}$ if and only if $\xi_{i+1} \geq \xi_i$, $i = 0, \ldots, N - 1$

3) Y^N is convex on $\mathcal{D}$ if and only if $\xi_{i+2} - \xi_{i+1} \geq \xi_{i+1} - \xi_i$, $i = 0, \ldots, N - 2$.

In accordance with Proposition 57, kriging the unknown function f boils down to finding the conditional distribution of Y^N given

$$
\begin{cases}
Y^N(\mathbf{X}) + \boldsymbol{\varepsilon} = \boldsymbol{y} \\
Y^N \in \mathcal{M}
\end{cases} .
\tag{5.2.3}
$$

This is equivalent to finding the distribution of the truncated Gaussian vector $\boldsymbol{\xi} \sim \mathcal{N}(0, \Gamma^N)$ given that

$$
\begin{cases}
\boldsymbol{\Phi}(\mathbf{X}) \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} = \boldsymbol{y} \\
\boldsymbol{\xi} \in \mathcal{C}_{ineq}
\end{cases}
\tag{5.2.4}
$$

where $\mathcal{C}_{ineq}$ is a set of linear inequality constraints on $\boldsymbol{\xi}$ as given in 2) and 3) of Proposition 57.

5.2.2   The most probable response curve and measurement noises

The maximum a Posteriori (MAP) of the Gaussian process Y^N conditionally to (5.2.3) satisfies the inequality constraints on the entire domain of interest and corresponds to the most likely curve (see Cousin, Maatouk, and Rullière, 2016). Its expression is given by

$$
\boldsymbol{\nu}_{Y^N}(x) := \sum_{i=0}^{N} \boldsymbol{\nu}_{\boldsymbol{\xi}}^i \phi_i(x)
\tag{5.2.5}
$$

where $\boldsymbol{\nu}_{\boldsymbol{\xi}} = \left(\boldsymbol{\nu}_{\boldsymbol{\xi}}^0, \ldots \boldsymbol{\nu}_{\boldsymbol{\xi}}^N \right)^\top$ is the MAP of the truncated Gaussian vector $\boldsymbol{\xi}$ which is the solution of the following convex optimization problem (see, e.g., Cousin, Maatouk, and Rullière, 2016):

$$
\min_{\boldsymbol{\Phi} \cdot \boldsymbol{\vartheta} + \boldsymbol{\varepsilon} = \boldsymbol{y}, \, \boldsymbol{\vartheta} \in C_{ineq}} (\frac{1}{2} \boldsymbol{\vartheta}^\top (\Gamma^N)^{-1} \boldsymbol{\vartheta}),
\tag{5.2.6}
$$

where Γ^N is the covariance matrix of $\boldsymbol{\xi}$, where we have used the the shorthand $\boldsymbol{\Phi}(\mathbf{X}) = \boldsymbol{\Phi}$.
The solution of the optimization problem (5.2.6) can also be obtained by maximizing the he density function of the conditional Gaussian vector $\boldsymbol{\xi} | \boldsymbol{\xi} \in C_{ineq}$ restricted to C_{ineq} which leads to

$$
\boldsymbol{\nu}_{\boldsymbol{\xi}} \propto \arg \max_{\boldsymbol{\vartheta} \in C_{ineq}} \exp\{-\frac{1}{2} (\boldsymbol{\vartheta} - \mu_{\text{cond}})^\top \Sigma_{\text{cond}}^{-1} (\boldsymbol{\vartheta} - \mu_{\text{cond}})\}
$$

where μ_{cond} and Σ_{cond} are respectively defined by (5.2.9) and (5.2.10).
Therefore the MAP of the truncated Gaussian vector $\boldsymbol{\xi}$ is also given by

$$
\boldsymbol{\nu}_{\boldsymbol{\xi}} = \arg \min_{\boldsymbol{\vartheta} \in C_{ineq}} (\frac{1}{2} \boldsymbol{\vartheta}^\top \Sigma_{\text{cond}}^{-1} \boldsymbol{\vartheta} - \boldsymbol{\vartheta}^\top \Sigma_{\text{cond}}^{-1} \mu_{\text{cond}}).
\tag{5.2.7}
$$

Remark 58 *Note that in the case of noiseless (i.e., if $\boldsymbol{\varepsilon} = 0$), the mode does not depend on the parameter σ of the covariance function $\mathcal{K}$. Indeed, in this case, σ does not affect the the solution of (5.2.6) since σ^2 is a multiplicative constant in Γ^N. However, in the case with noisy data, this is no more true.*

5.2.3 Hyper-parameters learning in constrained kriging

There are two recent methods for estimating the hyper-parameters. The first one is the Adapted Cross-Validation (ACV) of Maatouk, Roustant, and Richet (2015) which readjust the classical Leave-One-Out mean square error criterion by using the MAP instead of the (unconstrained) kriging mean for estimating the length scale hyper-parameter θ of the covariance function $\mathcal{K}$. However, this method works only in the case of noiseless, because of the fact that the parameter σ does not depend on the MAP in the case without noise, as we mentioned in Remark 58.

The second one is the so-called constrained Maximum Likelihood Estimator (cMLE) of Bachoc, Lagnoux, López-Lopera, et al. (2019) which is developed for taking into account the inequality constraints in the estimation of the hyper-parameters. However, in their study for asymptotic properties of cMLE, Bachoc, Lagnoux, López-Lopera, et al. (2019) show that any consistency with unconstrained Gaussian process, is preserved when adding inequality constraints. This is because for large sample sizes the constraints slightly impact the log marginal likelihood. Hence the cMLE has the same asymptotic distribution as the MLE. In other terms, conditioning by the constraints (significantly increases the computational burden and) has a negligible impact on the MLE, unless perhaps the sample size is very small.

In this line of thinking, we limit ourselves to the MLE in this work. By using the finite dimensional approximation (5.2.2) of the Gaussian process Y, the log marginal likelihood defined in (5.1.3) is then given by

$$\log(p(Y^N|\gamma, \mathbf{X})) = -\frac{1}{2}\boldsymbol{y}^\top(\Phi\Gamma^N\Phi^\top + \varsigma^2 I_n^2)^{-1}\boldsymbol{y} - \frac{1}{2}\log(\det(\Phi\Gamma^N\Phi^\top + \varsigma^2 I_n^2)) - \frac{n}{2}\log(2\pi) \qquad (5.2.8)$$

and the vector of estimated parameters $\hat{\gamma}$ verifies :

$$\hat{\gamma} = \arg\max_\gamma \log(p(Y^N|\gamma, \mathbf{X})) \,.$$

5.2.4 Sampling finite-dimensional GP with shape constraints

In this section, we present the technique of simulating the coefficients $\boldsymbol{\xi}$. We present the problem of constraints saturation when using the improved rejection sampling technique of Maatouk and Bay (2016) and propose a numerical solution to that.

5.2.4.1 Sampling finite dimensional Gaussian processes under shape constraints

In view of (5.2.4), kriging the constrained function f consists in sampling $\boldsymbol{\Phi}(X) \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} = \boldsymbol{y}$ truncated on $\mathcal{C}_{ineq}$. The conditional distribution of $\boldsymbol{\xi}$ given $\boldsymbol{\Phi}(X) \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} = \boldsymbol{y}$ is multivariate Normal with $\mathcal{N}(\boldsymbol{\mu}_{cond}, \Sigma_{cond})$ (see Williams and Rasmussen, 2006) where

$$\mu_{\text{cond}} = \Gamma^N\Phi^\top(\Phi\Gamma^N\Phi^\top + \varsigma^2 I_n)^{-1}\boldsymbol{y} \qquad (5.2.9)$$

and

$$\Sigma_{\text{cond}} = \Gamma^N - \Gamma^N\Phi^\top(\Phi\Gamma^N\Phi^\top + \varsigma^2 I_n)^{-1}\Phi\Gamma^N. \qquad (5.2.10)$$

Hence we face the problem of sampling from a truncated multivariate Gaussian distribution, which can be done by using different methods among them the Rejection Sampling from the Mode (RSM) developed by Maatouk and Bay (2014). López-Lopera et al. (2018) have investigated the following methods : RSM (Maatouk and Bay, 2014), Gibbs Sampling (Gibbs) (Taylor and Benjamini, 2016), Exponential Tilting (ET) (Botev, 2017), Metropolis-Hasting (MH) (Robert, 2014) and Hamiltonian Monte Carlo (HMC) (Pakman and Paninski, 2014) and it turns out that the Exact Hamiltonian Monte Carlo is the more efficient in term of Effective sample size (ESS) (Thiébaux and Zwiers, 1984) and less expensive in term of CPU time.

An appropriate choice of the initial vector (which must verify the constraints) in the algorithm of Exact

Hamiltonian Monte Carlo for sampling $\boldsymbol{\xi}$, is the MAP of $\boldsymbol{\xi}$ (see López-Lopera et al., 2018). We will consider this sampling technique in Chapters 6 and 7.

5.2.4.2 The saturated monotonicity constraints problem when using the RMS algorithm

By saturated monotonicity constraints, we mean the region where outputs are close. This is characterized by the fact that the mode of the GPs lies in boundary of the constraints. That problem is frequent in practice and requires special attention when simulating the posterior process using the RMS algorithm.

The algorithm of improved rejection sampling from the MAP (RMS) of Maatouk and Bay (2016) is a generalization of the rejection technique of Von Neumann (1951) for simulating truncated multivariate Gaussian random variables. This method consists in simulating at first the conditional Gaussian vector $\boldsymbol{\xi}|\boldsymbol{\Phi}(X)\cdot\boldsymbol{\xi}+\boldsymbol{\varepsilon}=\boldsymbol{y}$ and then using an improved rejection sampling around the MAP of $\boldsymbol{\xi}$ (indeed this verifies the constraints in the whole domain) that selects only the random coefficients that belong to the convex set C_{ineq}.

Although this algorithm outperforms Von Neumann's method for the simulation of such random variables, saturation of monotonicity constraints raises problems for efficient samples, in particular for large dimension vector. While using this algorithm, a very poor acceptance rate in the regions of the input state space is reached, where constraints are saturated. Indeed, as long as the updated value is different to the MAP in the saturation region, the algorithm will not work. A possible solution consists inusing a non-homogeneous adapted grid in order to get an efficient sample. For this purpose, we use a class of the so-called *h-adaptivity* scheme in which the mesh connectivity changes accordingly to the level of constraint saturation (see Löhner and Baum, 1992; Mitchell, 1991). In particular, it boils down to reduce the number of basis functions in the regions of the input state space where constraints are saturated. This consists in

- fixing a level of saturation constraints,

- detecting the saturation regions,

- reducing the number of basis functions in these regions for getting an adapted grid

- and using the sample technique in this adapted grid.

Illustrative example   For illustrating the adaptive mesh refinement approach, we consider the curve construction of a non-negative and increasing function f which is only known at some inputs locations as given by Table 5.2. Without loss of generality, we assume that f is observed at x without noise.

x	f (x)
0.22	103.050
0.32	128.925
0.57	129.025
0.82	141.825
1.32	175.700
1.82	175.800
2.81	204.700

TABLE 5.2: Market data

We consider f as a sample of zero-mean Gaussian process Y with finite dimensional approximation given as in (5.2.2).

We consider the Matérn 5/2 kernel function with length scale parameter $\theta = 1$ and variance parameter $\sigma = 100$, we then choose $N = 100$ basis functions. The red curve in Figure 5.2 shows the Maximum a Posteriori of the constrained Gaussian Process which is the most likely curve. With a saturation level of order 1, we can observe two regions of constraints saturation which coincide with the respective input intervals $[0.32, 0.57]$ (with outputs 129.025 and 129.025), $[1.32, 1.82]$ (with outputs 175.800 and 175.700). In

order to get an efficient sample of the truncated Gaussian vector ξ we reduce the number of basis functions in these two saturation regions. By this mean, we construct 95% confidence interval of the 100 simulated paths from the constrained Gaussian using the RMS algorithm (dashed lines).
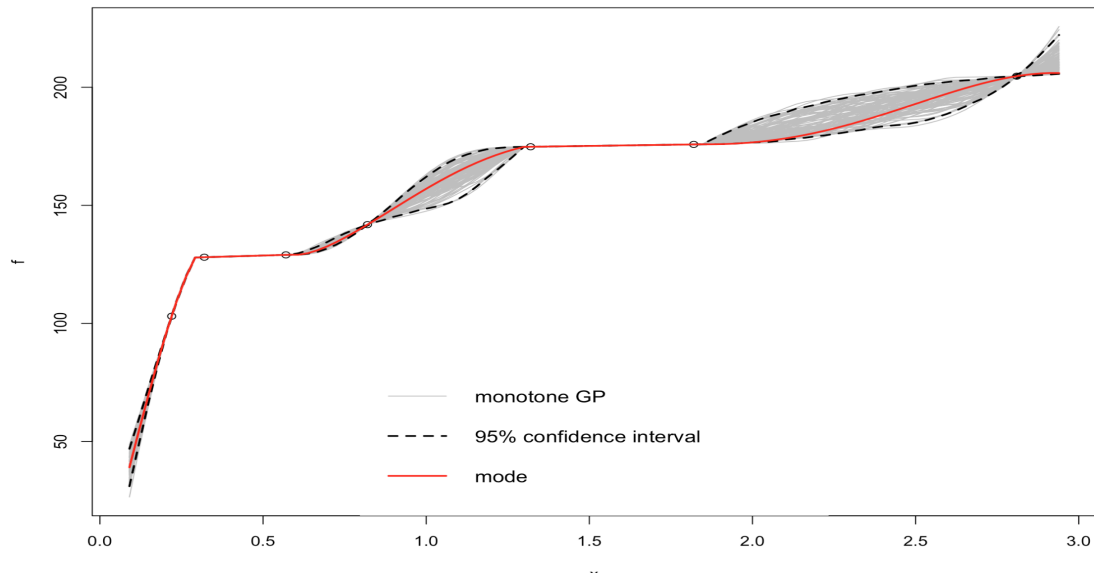

FIGURE 5.2: Illustration of learning f with two saturation regions.

The main advantage of this technique is the fact that it allows to get efficient sample and to reduce computational costs. But a caveat to take into account is the choice of the level of saturation constraints. It requires a careful reflection for solving this disadvantage and can be an open problem for improving the method. Fortunately, even if the Hamiltonian Monte Carlo method (HMC) could be time consuming when constraints are saturated, we demonstrate in the next chapter that this method is suitable for sampling the GP paths in the volatility surface construction. This is way we propose to use it in what follows.

In the next chapters, we will extend the kriging technique in two dimensions and we will show how to adapt this technique for the construction of option price and volatility surfaces which respect the no-arbitrage conditions.

Chapter 6

Kriging for implied volatility surface

This chapter is based on A. Cousin, D. Gueye, Kriging for implied volatility surface, working paper.

6.1 Introduction

Implied volatility surface is of crucial interest for risk management and exotic option pricing models. Its construction is usually carried out in accordance with the arbitrage-free principle. This condition leads to shape restrictions on the option prices such as monotonicity with respect to maturities and convexity with respect to strike prices. In this chapter, we propose a new arbitrage-free construction method that extends classical spline techniques by additionally allowing for quantification of uncertainty. The proposed method extends the constrained kriging techniques developed in Maatouk and Bay (2014) and Cousin, Maatouk, and Rullière (2016) to the context of volatility surface construction. Assuming a Gaussian process prior, the posterior price surface becomes a truncated Gaussian field given shape constraints and market observations. Prices of illiquid instruments can also be incorporated when considered as noisy observations. Starting from a suitable finite-dimensional approximation of the Gaussian process prior, the no-arbitrage condition on the entire input domain is characterized by a finite number of linear inequality constraints. We define the most likely response surface and the most-likely noise values as the solution of a quadratic optimization problem. We use Hamiltonian Monte Carlo (HMC) techniques to simulate the posterior truncated Gaussian surface and build pointwise confidence bands. The Gaussian process hyper-parameters are estimated using the maximum likelihood estimator (MLE). The method is illustrated on Euro Stoxx 50 option prices by building no-arbitrage implied volatility surfaces and their corresponding confidence bands.

This chapter is organized as follows. In section 6.2, we recall some well known facts concerning implied volatility. Section 6.3 emphasizes on constrained kriging and Section 6.4 is devoted to the numerical illustrations.

6.2 Option pricing in no-arbitrage models

The goal of this section is to recall the notion of implied volatility surface and explain the construction problem of such surface under the no-arbitrage assumptions. For this purpose, we discuss it through the standard model of Black-Scholes. More details can be founded in Jeanblanc, Yor, and Chesney (2009, subsection 2.3), Shreve (2004, subsection 4.5), Karatzas (1997, section 1.2).

6.2.1 Black-Scholes model

We consider an European option on a given asset S, with maturity T and strike price X. By denoting Φ its payoff function, the value of the option at time T is given, in term of Call option, by $\Phi(S_T) = (S_T - X)^+$, where $(S_T - X)^+ = \max(S_T - X; 0)$. While, in term of put option, its value at time T is $\Phi(S_T) = \max(X - S_T; 0)$.

In Black and Scholes (1973) model, the risky asset price $S = (S)_{t \geq 0}$ defined on the filtered probability space $(\Omega, \mathcal{G}, \mathbb{P}, \mathbb{F})$ is assumed to follow a Geometric Brownian Motion, i.e., its dynamics are of the form :

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \ \forall t \geq 0 \tag{6.2.1}$$

where $B = (B_t)_{t \geq 0}$ is a Brownian Motion defined on $(\Omega, \mathcal{G}, \mathbb{P})$, μ represents the drift of the returns of S, σ its instantaneous volatility and $S_0 > 0$ the initial stock price. The interest rate r is assumed to be constant and positive so that the discounted stock price $\hat{S}$ has the following dynamics under the probability $\mathbb{P}$:

$$d\hat{S}_t = (\mu - r)\hat{S}_t dt + \sigma \hat{S}_t dB_t, \ \forall t \geq 0 \ .$$

One can show from the Girsanov theorem that there exists a (unique) measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, given by $\mathbb{Q}|_{\mathcal{F}_t} = e^{-\theta B_t - \frac{1}{2}\theta^2 t}\mathbb{P}|_{\mathcal{F}_t}$, with $\theta = \frac{\mu - r}{\sigma}$ Jeanblanc, Yor, and Chesney (2009, Proposition 2.3.1.1) and under which $\hat{S}$ is an $\mathbb{F}$-martingale. The fundamental theorem of asset pricing provides that this existence of such a $\mathbb{Q}$ is equivalent to the absence of arbitrage opportunity of this market model. Therefore, under the arbitrage-free conditions, the value at time t, of the European option is the expected present value under the risk neutral measure $\mathbb{Q}$ of future payoffs, which has the following expression:

$$V_t(S_t, t) = \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}\Phi(S_T)|\mathcal{F}_t\right]. \tag{6.2.2}$$

The computation of the expression (6.2.2) leads to the famous Black-Scholes formula.
For instance, the value of a put option on S, denoted by P_t^{BS} is given by

$$P_t^{BS}(S_t, T, X, \sigma, r) = Xe^{-r(T-t)}N(-d_2) - S_t N(-d_1), \tag{6.2.3}$$

where

$$d_1 = \frac{\log(S_t/X) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}},$$
$$d_2 = d_1 - \sigma\sqrt{T-t}.$$

and N is the cumulative normal density function.

6.2.2 Implied volatility surface

The equality (6.2.3) links the European put option price to the underlying price S, the interest rate r, the volatility of the stock σ, the strike price X and the maturity T. All these parameters except the volatility σ may be directly observed in the market. Actually, the option price is a function of the volatility of the underlying asset and then, inverting it leads to the so called implied volatility, which is of crucial interest for risk management and exotic option pricing models. This means that for a given observed market option price P_t^{Market}, quoted at the date t, with corresponding maturity T and Strike X, the implied volatility is the solution value $\sigma_t^{IV}(T, X)$ of

$$P_t^{BS}(S_t, T, X, \sigma_t^{IV}(T, X), r) = P_t^{Market}. \tag{6.2.4}$$

Accordingly, our interpolation problem consists in computing at any time t, the whole surface $(T, X) \rightarrow P_t^{BS}(T, X)$ of option prices which allows us to derive the implied volatility surface $(T, X) \rightarrow \sigma_t^{IV}(T, X)$ by using the inversion problem. However, this construction must take into account some characteristics such as :

- **Incomplete information** : the option price surface is only known or can only be estimated for few input locations (the observed couples Strike-maturity).

- **Noisy measurement** : observed prices may not be fully reliable (ex : price of illiquid instruments).

- **Smoothness constraints** : the price surface should be differentiable (important for deriving the local volatility surface, see Chapter 7).

- **Shape constraints** : the price surface should not allow to generate arbitrage.

In this regard, we propose to use the kriging techniques for such construction. We discuss our methodology that requires the transition from classical kriging, which does not allow to obtain arbitrage-free option price surfaces, to the constrained kriging which does allow to get free-arbitrage option price surfaces. We present these techniques in the next section.

In what follows, we only consider Put option, Call option price can be recovered using the call-put parity. Before going further, we recall how no-arbitrage conditions translate into shape constraints on the put option price.

Proposition 59 *We place ourselves at a fixed date of evolution t_0 and we consider S_0 the value of the underlying at t_0.*
The put price surface $(T, X) \to P(T, X)$ is free of static arbitrage if and only if

(i) $X \to P(T, X)$ *is a convex function such that* $P(T, 0) = 0$ *and* $\frac{\partial P}{\partial X}(T, 0) = 0$, *for any* $T \geq 0$.

(ii) $T \to P(T, X)$ *is a non-decreasing function, for any* $X \geq 0$.

(iii) $\lim\limits_{X \to \infty} P(T, X) = X - S_0$.

(iv) $P(0, X) = (X - S_0)^+$.

PROOF: This follows by using the call-put parity formula in Roper (2010, Theorem 2.1).

6.3   Kriging for learning arbitrage-free put option price surfaces

In this suebsection, we present the classical kriging techniques which allow to construct option price surface by only incorporating the three first characteristics mentioned above (i.e, incomplete information, noisy measurement and smoothness constraints).

Given the input domain $\mathcal{D}$ in time and space, we aim at constructing, at a given quotation date, put price surface

$$P : \mathcal{D} \to \mathbb{R}^+$$
$$(T, X) \mapsto P(T, X)$$

satisfying arbitrage-free conditions given in Proposition 59 and given n noisy observations $\boldsymbol{y} = [y_1, ..., y_n]^\top$ of function P at input points $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_n]$ where $\mathbf{X}_i = (T_i, X_i)$, $i = 1, \ldots, n$. Then, this construction should be compatible with market fit condition

$$\boldsymbol{y} = P(\mathbf{X}) + \boldsymbol{\varepsilon} \tag{6.3.1}$$

where $P(\mathbf{X}) := [P(\mathbf{X}_1), ..., P(\mathbf{X}_n)]^\top$. The additive noise term $\boldsymbol{\varepsilon} = [\varepsilon_1, ..., \varepsilon_n]^T$ is assumed to be a zero-mean Gaussian vector, independent from P, and with an homoscedastic covariance matrix given as $\varsigma^2 I_n$, where I_n is the identity matrix of size n.

Comments 60 *As mentioned in Cousin, Maatouk, and Rullière (2016), this framework which takes account of the presence of error noise being quite considerable inasmuch as it allows to construct implied volatility surface in presence of illiquid options. Accordingly a best way to incorporate some noises in the response variable is to investigate the main sources of these noises. In Hentschel, 2003 three kinds of sources of measurement errors have been referred in the option prices such as the finite quote precision which is based on the tick price that represents the minimum increment between bid and ask prices of an asset in the trading system which orients the prices movement in a discrete setting. However the real market prices move in a continuous way and this discrete increment should be a source of noisy observations. In addition to the finite quote precision, Hentschel invoques the non-synchronous prices and the bid-ask spread which is the difference between the ask (which represents the supply for a particular asset) and the bid (the demand for asset). The higher the bid-ask spread of an option, the more liquid this option becomes. In Chapter 7, we have considered bid and ask prices of the same option as two replicate noisy observations at the same input location.*

Two approaches for choosing our observations and incorporating noises might be taken into account, the first one consists in considering the realizations of the Gaussian process to be the mid-point prices and assuming the noise variance term ς to be proportional to the magnitudes of the bid-ask spreads. The second one consists in considering both the bid and ask prices as two independent realizations of the Gaussian process and estimate ς by using an appropriate method of parameter estimation. This last approach allows to obtain a option price surface which lies between the free-arbitrage bid and ask surfaces and it will be further detailed in Chapter 7.

6.3.1   Classical GPs regression or kriging

We consider a zero-mean Gaussian process prior on the mapping $P = P(T,X)_{(T,X)\in\mathcal{D}}$ with covariance function (kernel function) $\mathcal{K}$. Then, the output vector $P(\mathbf{X})$ has a normal distribution with zero mean and covariance matrix $\boldsymbol{C}$ with components $\mathrm{cov}(P(T_i, X_i), P(T_j, X_j)) = \mathcal{K}((T_i, X_i), (T_j, X_j))$. We consider a 2-dimensional isotropic covariance kernel given as a tensor product, i.e., for $\mathbf{x} = (T, X)$ and $\mathbf{x}' = (T', X')$ two elements of $\mathcal{D}$,

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \sigma^2 R_T(T - T', \theta_T) R_X(X - X', \theta_X)$$

where $\theta = (\theta_T, \theta_X)$ and σ^2 correspond to the length scale and the variance hyper-parameters of the kernel function $\mathcal{K}$ and the functions R_T and R_X are kernel correlation functions.

It is well known that the conditional process $P \mid \boldsymbol{y} = P(\mathbf{X}) + \boldsymbol{\varepsilon}$ is Gaussian with mean function η and covariance function $\mathcal{K}^\star$ given respectively by (see Williams and Rasmussen, 2006):

$$\eta(\mathbf{x}) = \mathbf{c}(\mathbf{x})^\top (\boldsymbol{C} + \varsigma^2 I_n)^{-1} \boldsymbol{y}, \quad \mathbf{x} = (T, X) \in \mathcal{D} \tag{6.3.2}$$

$$\mathcal{K}^\star(\mathbf{x}, \mathbf{x}') = \mathcal{K}(\mathbf{x}, \mathbf{x}') - \mathbf{c}(\mathbf{x})^\top (\boldsymbol{C} + \varsigma^2 I_n)^{-1} \mathbf{c}(\mathbf{x}'), \quad \mathbf{x}, \mathbf{x}' \in \mathcal{D} \tag{6.3.3}$$

where $\mathbf{c}(\mathbf{x}) = [\mathcal{K}(\mathbf{x}, (T_1, X_1)), ..., \mathcal{K}(\mathbf{x}, (T_n, X_n))]^\top$.

Without considering arbitrage-free conditions as described in Proposition 59, estimation of the price function P under this framework is known as classical GPs regression or classical kriging. In this setting, prediction and uncertainty quantification is made using the conditional distribution $P \mid \boldsymbol{y} = P(\mathbf{X}) + \boldsymbol{\varepsilon}$. The Best Linear Unbiased Estimator (BLUE) of P is given as the kriging mean function (6.3.2). The conditional covariance function $\mathcal{K}^*$ can be used to obtain confidence bands around the predicted price surface. The hyper-parameters of the kernel function $\mathcal{K}$ as well as the variance of the noise can be estimated using the maximum likelihood estimator (MLE) (see, e.g., Bachoc, Lagnoux, and López-Lopera, 2018).

6.3.2   Imposing the no-arbitrage conditions

Conditionally to the market fit condition (6.3.1) and conditions (iii) and (iv) of Proposition 59, P is still Gaussian. However, conditionally to the inequality constraints (i.e, the monotonicity of the put price with respect to the maturities direction and its convexity with respect to the strike prices) by means (i) and (ii) of Proposition 59, the process P is no longer Gaussian and this issue obviously run across the difficulties of simulating the posterior process in the sense that the range of constraint check points is usually infinite-dimensional in the simulation. We adopt the solution of Cousin, Maatouk, and Rullière (2016) that consists in constructing a finite-dimensional approximation P^N of the Gaussian prior P for which the constraints can be checked in the entire domain $\mathcal{D}$ with a finite number of checks.

We first consider a discretized version of the input space $\mathcal{D}$ as a $N = (N_T + 1) \times (N_X + 1)$ regular grid with knots (u_i, v_j), $i = 1, ..., N_T$, $j = 1, ..., N_X$ with $u_i = i\delta_T$ and $v_j = j\delta_X$, where $\delta_T = \frac{1}{N_T}$ and $\delta_X = \frac{1}{N_X}$. For each knot (u_i, v_j), we introduce the hat basis function defined as the following tensor product

$$\phi_{i,j}(T, X) := \max(1 - \frac{|T - u_i|}{\delta_T}, 0) \max(1 - \frac{|X - v_j|}{\delta_X}, 0) \ .$$

Then, the process P is approximated on $\mathcal{D}$ by the process P^N given by

$$P^N(T, X) = \sum_{i=0}^{N_T} \sum_{j=0}^{N_X} P(u_i, v_j)\phi_{i,j}(T, X), \quad \text{for all } (T, X) \in \mathcal{D} \tag{6.3.4}$$

which is a piecewise linear interpolation of P at knots $(u_i, v_j)_{i,j}$. If we denote $\xi_{i,j} = P(u_i, v_j)$, for $i = 1, ..., N_T$, $j = 1, ..., N_X$, then $\boldsymbol{\xi} = [\xi_{0,0}, ..., \xi_{i,j}..., \xi_{N_T, N_X}]^\top$ is a zero-mean Gaussian vector with $N \times N$ covariance matrix Γ^N such that $\Gamma^h_{i_1, i_2} = \mathcal{K}((u_{i_1}, v_{j_1}), (u_{i_2}, v_{j_2}))$, for any two grid index pairs (i_1, j_1) and (i_2, j_2) corresponding to global indices ι_1 and ι_2 respectively and $\boldsymbol{\phi}(T, X)$ a vector of size N given by

$$\boldsymbol{\phi}(T, X) = [\phi_{0,0}(T, X), ..., \phi_{i,j}(T, X), ..., \phi_{N_T, N_X}(T, X)].$$

The equality (6.3.4) can be written in the following matrix form

$$P^N(T, X) = \boldsymbol{\phi}(T, X) \cdot \boldsymbol{\xi}$$

so that when denoting $\boldsymbol{\Phi}(\mathbf{X})$ the $n \times N$ matrix of basis function in which, each row l corresponds to the vector $\boldsymbol{\phi}(T_l, X_l)$, one has $P^N(\mathbf{X}) = \boldsymbol{\Phi}(\mathbf{X}) \cdot \boldsymbol{\xi}$, with $P^N(\mathbf{X}) := [P^N(\mathbf{X}_1), ..., P^N(\mathbf{X}_n)]^\top$.

In what follows we use the shorthand $\boldsymbol{\Phi}(\mathbf{X}) = \boldsymbol{\Phi}$.

Proposition 61 *(see Maatouk, 2017)*
The following statements hold.

- *The finite-dimensional process P^N uniformly converges to P on $\mathcal{D}$ as $N_X \to \infty$ and $N_T \to \infty$, almost surely,*

- *$P^N(T, X)$ is non-decreasing function of T if and only if $\xi_{i+1,j} \geq \xi_{i,j}$,*

- *$P^N(T, X)$ is a convex function of X if and only if $\xi_{i,j+2} - \xi_{i,j+1} \geq \xi_{i,j+1} - \xi_{i,j}$.*

Given the first statement of Proposition 61, by denoting $\mathcal{M}$ the convex set of inequality constraints, i.e., $\mathcal{M}$ is the set of 2-d continuous functions which are no-decreasing w.r.t. T and convex w.r.t. the X, our construction problem (we denote it $(\mathcal{P})$) consists in finding the conditional distribution of P^N given

$$\begin{cases} \boldsymbol{y} = P^N(\mathbf{X}) + \boldsymbol{\varepsilon} \\ P^N \in \mathcal{M}. \end{cases}$$

The last two statements of Proposition 61 justify the choice of the hat basis functions which is due to the fact that P^N satisfies the inequality constraints in the entire domain $\mathcal{D}$ when it satisfies these constraints at the knots (see Maatouk and Bay, 2014), i.e., $P^N \in \mathcal{M}$ if and only if $\boldsymbol{\xi} \in \mathcal{C}_{ineq}$ where $\mathcal{C}_{ineq}$ is a set of linear inequality constraints on $\boldsymbol{\xi}$ as given by the two last points of Proposition 61. In this line of thinking, our construction problem $(\mathcal{P})$ is equivalent to estimate $\boldsymbol{\xi}$ restricted to

$$\begin{cases} \boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} \\ \boldsymbol{\xi} \in \mathcal{C}_{ineq}. \end{cases}$$

6.3.3 Hyper-parameter learning

Hyper-parameters, namely the length scale θ and the variance parameter σ, of the kernel function $\mathcal{K}$ as well as the noise parameter ς can be either specified or estimated. By denoting λ the set of these parameters (i.e., $\lambda = [\theta, \sigma, \varsigma]^\top$), we propose to maximize the marginal log likelihood $\mathcal{L}(\lambda)$ for the process P^N w.r.t. λ for parameters learning.

Under the finite dimensional approximation, the marginal log likelihood can be expressed as (see, e.g., Williams and Rasmussen, 2006):

$$\mathcal{L}(\lambda) = -\frac{1}{2}\boldsymbol{y}^\top (\boldsymbol{\Phi}\Gamma^N \boldsymbol{\Phi}^\top + \varsigma^2 I_n)^{-1}\boldsymbol{y} - \frac{1}{2}\log(\det(\boldsymbol{\Phi}\Gamma^N \boldsymbol{\Phi}^\top + \varsigma^2 I_n)) - \frac{n}{2}\log(2\pi). \tag{6.3.5}$$

6.3.4 The most probable response surface and measurement noises

The MAP of P^N is given by

$$\mathbf{m}_{P^N}(T, X) := \sum_{i=0}^{N_X} \sum_{j=0}^{N_T} \boldsymbol{\nu}_{\boldsymbol{\xi}}^{i,j} \phi_{i,j}(T, X) \tag{6.3.6}$$

where $\boldsymbol{\nu}_{\boldsymbol{\xi}} = \left(\boldsymbol{\nu}_{\boldsymbol{\xi}}^{(0,0)}, \ldots \boldsymbol{\nu}_{\boldsymbol{\xi}}^{(i,j)}, \cdots, \boldsymbol{\nu}_{\boldsymbol{\xi}}^{(N_x, N_t)} \right)^\top$ is the MAP of the Gaussian coefficients $\boldsymbol{\xi}$. As we mentioned in Chapter 5, the MAP $\boldsymbol{\nu}_{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$ is paramount in the sampling of $\boldsymbol{\xi}$ since it satisfies the inequality constraints hence it can be considered as the initial vector when using the sampling algorithm of HMC (see López-Lopera et al., 2018). As explained in Section 5.2.2 of Chapter 5, this can be obtained directly by maximizing the density function of the conditional Gaussian vector $\boldsymbol{\xi}|\boldsymbol{\xi} \in C_{ineq}$ restricted to C_{ineq} which implies that

$$\boldsymbol{\nu}_{\boldsymbol{\xi}} = \arg\min_{\boldsymbol{\vartheta} \in C_{ineq}} (\boldsymbol{\vartheta}^\top \Sigma_{\text{cond}}^{-1} \boldsymbol{\vartheta} - \boldsymbol{\vartheta}^\top \Sigma_{\text{cond}}^{-1} \mu_{\text{cond}}) \tag{6.3.7}$$

where μ_{cond} and Σ_{cond} are respectively defined by (6.3.10) and (6.3.11). This is a quadratic optimization problem and it is equivalent to

$$\boldsymbol{\nu}_{\boldsymbol{\xi}} := \arg\min_{\boldsymbol{\Phi}\cdot\boldsymbol{\vartheta}+\boldsymbol{e}=\boldsymbol{Y}, \boldsymbol{\vartheta}\in\mathcal{C}_{ineq}} (\boldsymbol{\vartheta}^\top (\Gamma^N)^{-1} \boldsymbol{\vartheta}), \tag{6.3.8}$$

with Γ^N the covariance matrix of $\boldsymbol{\xi}$.

In order to identify the locations x of the largest noises and their values, we compute the joint MAP $(\boldsymbol{\nu}_{\boldsymbol{\xi}}, \boldsymbol{\nu}_{\boldsymbol{\varepsilon}})$ of the truncated gaussian vector $\boldsymbol{\xi}$ and the MAP of the Gaussian noise vector $\boldsymbol{\varepsilon}$. This can be defined as solution of

$$\max_{\boldsymbol{\vartheta}, \boldsymbol{e}} \mathbb{P} \left(\boldsymbol{\xi} \in [\boldsymbol{\vartheta}, \boldsymbol{\vartheta} + d\boldsymbol{\vartheta}], \boldsymbol{\varepsilon} \in [\boldsymbol{e}, \boldsymbol{e} + d\boldsymbol{e}] \mid \boldsymbol{\Phi} \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} = \tilde{\boldsymbol{v}}, \boldsymbol{\xi} \in \mathcal{C}_{ineq} \right).$$

As $(\boldsymbol{\xi}, \boldsymbol{\nu}_{\boldsymbol{\varepsilon}})$ is Gaussian centered with block-diagonal covariance matrix with blocks Γ^N and $\varsigma^2 I_n$, this implies that the mode $(\boldsymbol{\nu}_{\boldsymbol{\xi}}, \boldsymbol{\nu}_{\boldsymbol{\varepsilon}})$ is a solution to the following quadratic problem

$$\min_{\boldsymbol{\Phi}\cdot\boldsymbol{\vartheta}+\boldsymbol{e}=\tilde{\boldsymbol{v}}, \boldsymbol{\vartheta}\in\mathcal{C}_{ineq}} \left(\boldsymbol{\vartheta}^\top (\Gamma^N)^{-1} \boldsymbol{\vartheta} + \boldsymbol{e}^\top (\varsigma^2 I_n)^{-1} \boldsymbol{e} \right). \tag{6.3.9}$$

As a consequence, we define the most probable measurement noises to be $\boldsymbol{\nu_\varepsilon}$ and the most probable response surface $\mathbf{m}_{PN}$ given by $\mathbf{m}_{PN} := \boldsymbol{\Phi} \cdot \boldsymbol{\nu_\xi}$. Distance to the data can be an effect of arbitrage opportunities within the data and/or the misspecification / lack of expressiveness of the kernel.

6.3.5   Sampling finite-dimensional GPs with shape constraints

As we mentioned in Subsection 6.3.2, the construction of the put price surface consists in sampling $\boldsymbol{\xi}$ truncated on C_{ineq}. Known that the distribution of $\boldsymbol{\xi}$ given $\boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon}$ is multivariate Normal $\mathcal{N}(\boldsymbol{\mu}_{cond}, \Sigma_{cond})$ (see Williams and Rasmussen, 2006) where

$$\mu_{\mathrm{cond}} = \Gamma^N \boldsymbol{\Phi}^\top (\boldsymbol{\Phi}\Gamma^N\boldsymbol{\Phi}^\top + \varsigma^2 I_n)^{-1}\boldsymbol{y} \tag{6.3.10}$$

and

$$\Sigma_{\mathrm{cond}} = \Gamma^N \boldsymbol{\Phi}^\top (\boldsymbol{\Phi}\Gamma^N\boldsymbol{\Phi}^\top + \varsigma^2 I_n)^{-1}\boldsymbol{\Phi}\Gamma^N. \tag{6.3.11}$$

Hence we are face to a problem of sampling from truncated multivariate Gaussian distribution, which we do by Hamiltonian Monte Carlo (see López-Lopera et al., 2018), using the maximum a posteriori probability estimate (MAP) of $\boldsymbol{\xi}$ as initial vector (which must verify the constraints) in the algorithm, computed as explained in Chapter 5.

6.4   Numerical illustrations

The aim of this empirical investigation is to illustrate the construction methods introduced in the previous sections using real financial data. In particular, we construct the put prices surface from both classical and constrained kriging in such a way to use it for deriving the implied volatility surface from the inversion techniques. Our study turns out to be very interesting since the results would certainly be useful for no iliquid options pricing. It would also be possible through this study to understand the option prices consistency under bid ask spreads. In addition, a useful case study could be to investigate for instance, the liquidity measure which is one of the most important tools of a financial markets and characterizes the ability of market makers to execute trades at the determined market prices and for a large volume without affecting the stock price.

The illustrations are carried out on the Euro Stoxx 50. We present an approach which consists in constructing the put prices surface through the mid price by estimating the standard deviation of the noise by MLE.

Remark 62 *One can also construct the whole put prices surface through the mid price by defining the standard deviation of the noise as the difference between the ask and mid prices. Another approach consists in considering both the bid and ask put prices as independent responses of a zero-mean Gaussian process Y and constructing a put prices surface by estimating the attributed noise variance. As such, we can study the behavior of this surface relatively to the surfaces constructed from the bid put prices and the ask put prices. This last approach is used in Chapter 7.*

We observe at a particular market quotes t_0 as of January 10, 2019, a series of put option prices $f(T_i, X_i) = P(T_i, X_i)$ for different characteristics (T_i, X_i), $i = 1, \ldots, n = 1232$ with T_i the maturities in which, its range goes from January 18, 2019 to December 17, 2021 and X_i the strike prices which range from 250 to 4000. The spot price is equal to $S_{t_0} = 3070.24$. These observations are represented in Figure 6.1. Our goal is to construct the whole surface of the put option prices (using both classical and constrained kriging) that we use for computing the implied volatility surface by the inversion techniques.

To this end, we consider our input data which contains $n_T = 15$ maturities and $n_X = 88$ strike prices and which does not necessarily contain gridded data (as we may observe in Figure 6.1).

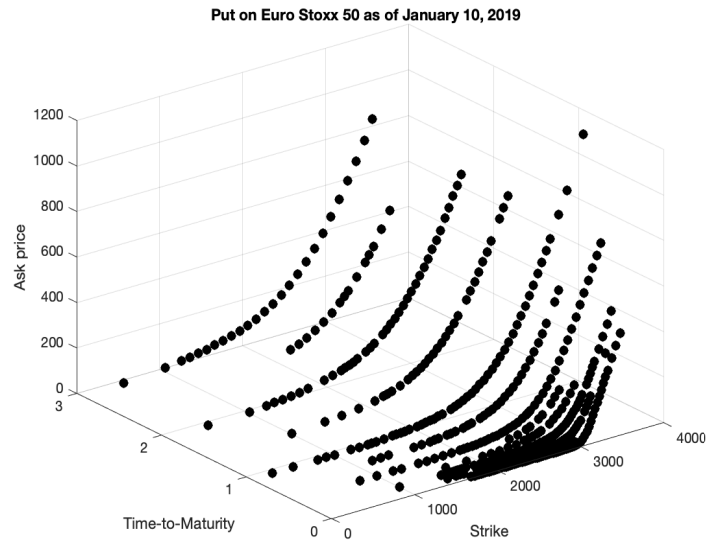**Put on Euro Stoxx 50 as of January 10, 2019**

FIGURE 6.1: Import input observed data

We randomly choose 5% of these data as training set. Thus the unknown function is evaluated at only 5% of the input data observed. After scaling the input space $\mathcal{D}$ into $[0, 1] \times [0, 1]$, we choose $N = 600$ basis functions given by $N = N_T \times N_X$ where $N_T = 30$ represents the number of nodes related to the time-to-maturities in $[0, 1]$ and $N_X = 20$ the number of nodes linked to the strike prices in $[0, 1]$. We consider a two dimensional Gaussian covariance function defined as

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp \left(-\frac{(X - X')^2}{\theta_X^2} - \frac{(T - T')^2}{\theta_T^2} \right)$$

where the components of the two vectors $\mathbf{x} = (T, X)$ and $\mathbf{x}' = (T', X')$ represent respectively the vectors of strike prices and time to maturities.
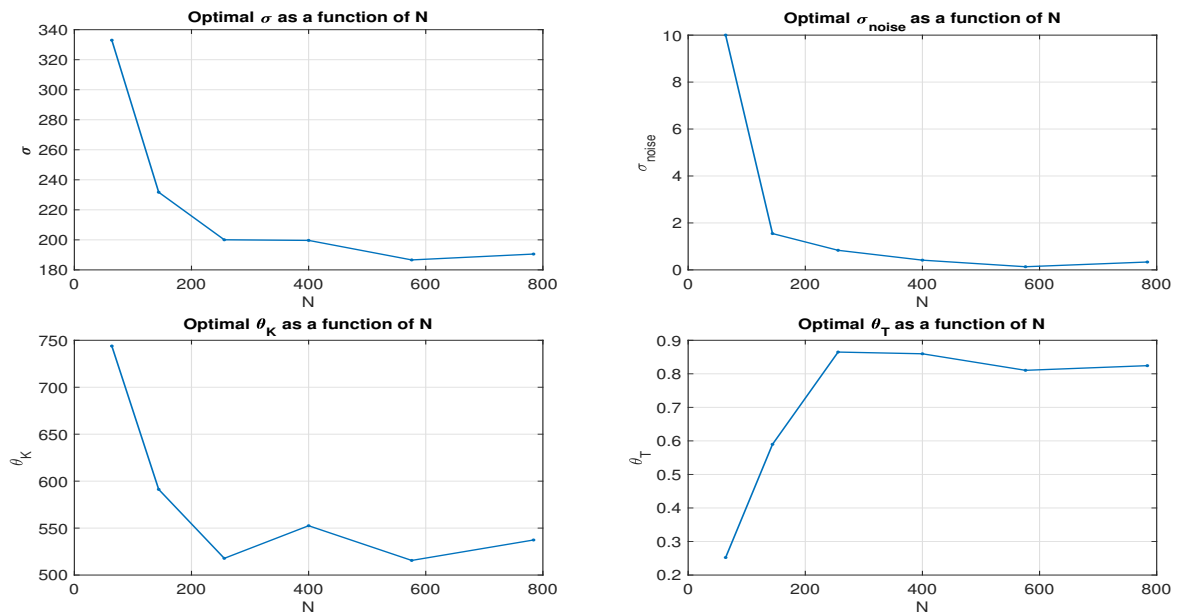
FIGURE 6.2: Convergence of optimal parameter as a function of N (number of basis functions).

The GP hyperparameters $\theta_X, \theta_T, \sigma$ and the standard deviation of the noise ς are estimated using the MLE. Figure 6.2 represents their stability relatively to the chosen number of basis functions. Here, we study the convergence of the parameters by increasing the number of basis function. It is clear that this convergence is achieved among from 250 basis functions. In short, parameters convergence is reached when the number of basis functions tends to infinity.

Then, as explained in Subsection 6.3.4, quadratic programming is used to find the most probable response surface and measurement noises, while enforcing the constraints, using the `interior-point-convex` iterative algorithm with a tolerance of 1×10^{-12}.

In Figure 6.3, we represent the most probable response surface (left) and measurement noises (right). It is clear that the surface of put prices respects the arbitrage-free conditions. The scattered noisy points show that low maturity options are associated with a significant error and therefore a high distance from the most probable non-arbitrable surface.
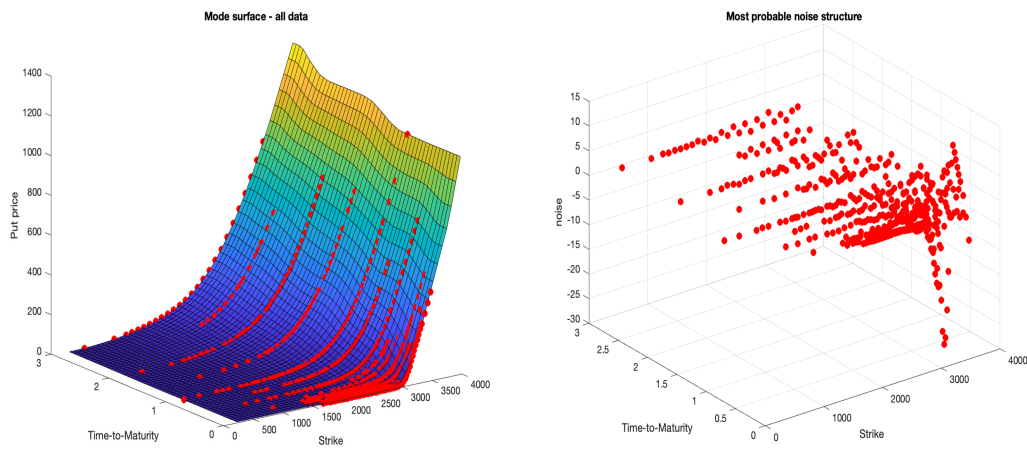

FIGURE 6.3: Most probable surface (left) vs most probable noise values (right).
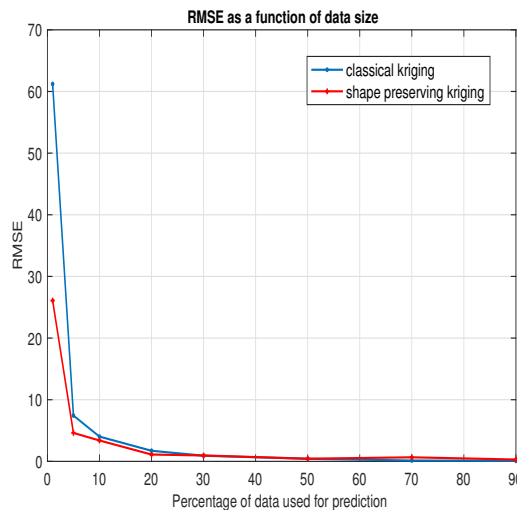

FIGURE 6.4: Mode estimator - prediction accuracy

Figure 6.4, illustrates the prediction accuracy of the mode estimator (MAP) which has been carried out by the following way:

- We first construct a series of randomly chosen data subsets with increasing number of points,

- We apply classical kriging and shape-preserving kriging on these subsets by computing the two mode estimators,

- For each data size, we compute the Root Mean Square Error (RMSE) with respect to the original data set.

The results show that for a small percent of data size (less than 30 %), the MAP estimated from the constrained kriging outperforms the one estimated from the unconstrained kriging. However, when more than 30% of data points are used for training, the two estimated MAP present a similar accuracy with their RMSE which tend to zero.

We now generate 5000 paths from both the classical and the constrained kriging using the Exact Hamiltonian Monte Carlo algorithm (see Pakman and Paninski, 2014). Then we compute the whole implied volatility surface from the inversion of the Black and Scholes formula as described in Subsection 6.2.1 with the spot price S_{t_0} and interest rate $r = 1\%$. Figure 6.5 presents comparison results on the MAP estimator between classical and constrained kriging. One can see that the put prices surface constructed from constrained kriging (in left) clearly verifies the no-arbitrage constraints while the one constructed using classical kriging does not fulfill these constraints. Also Figure 6.6 represents the obtained implied volatility surface which shows that the constrained kriging allows to obtain some surfaces which are more smooth than the ones obtained in classical kriging. In Figure 6.7 which shows 5% and 95% estimated quantiles of the fitted GPs, one can see a large confidence interval in the case of classical kriging where there is no observation while this confidence interval becomes more restricted in the constrained kriging, due to the knowledge of the no-arbitrage conditions.
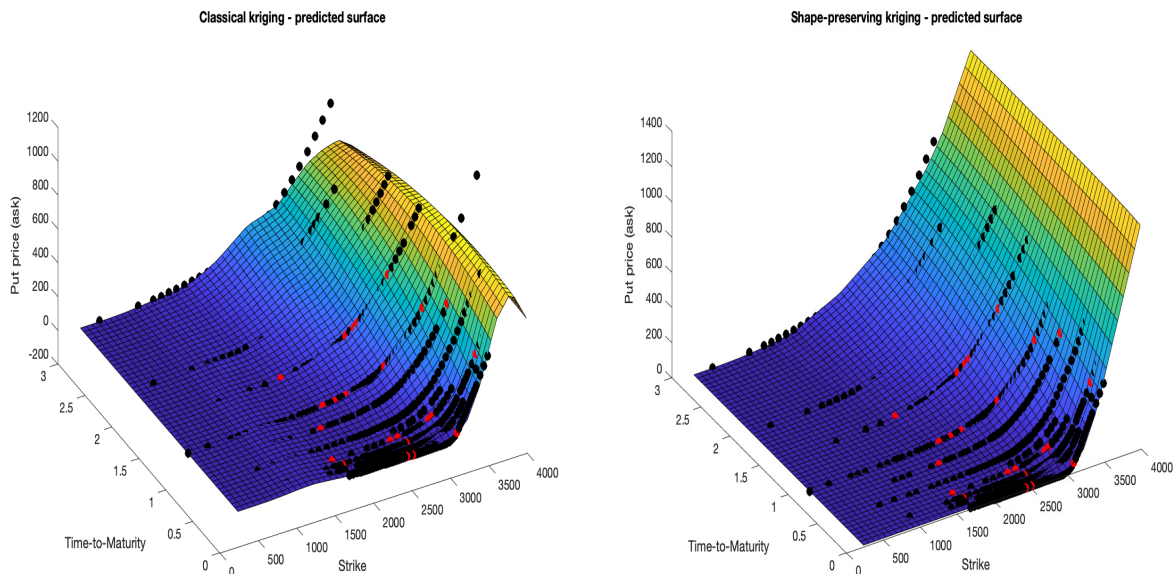

FIGURE 6.5: Put prices surface constructed from classical kriging (left) vs put prices surface constructed from constrained kriging (right). The red points represent the 5% of the data used for training.
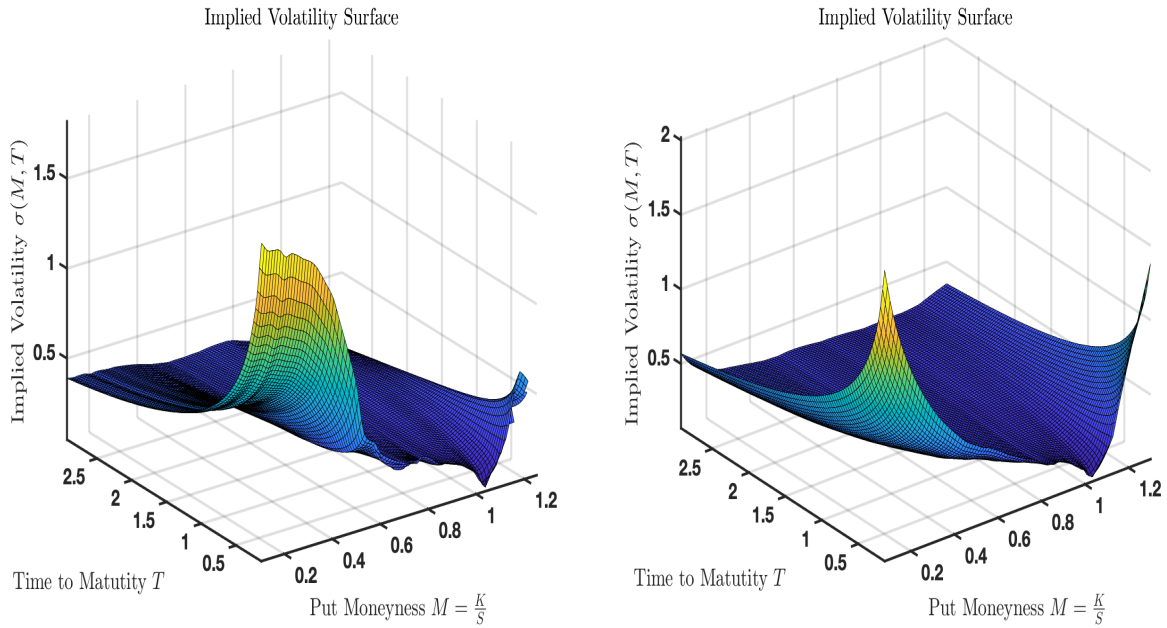
FIGURE 6.6: Implied volatility surface obtained from classical kriging (left) vs Implied volatility surface obtained from constrained kriging (right). Implied volatility surface obtained from constrained kriging is more smoother.
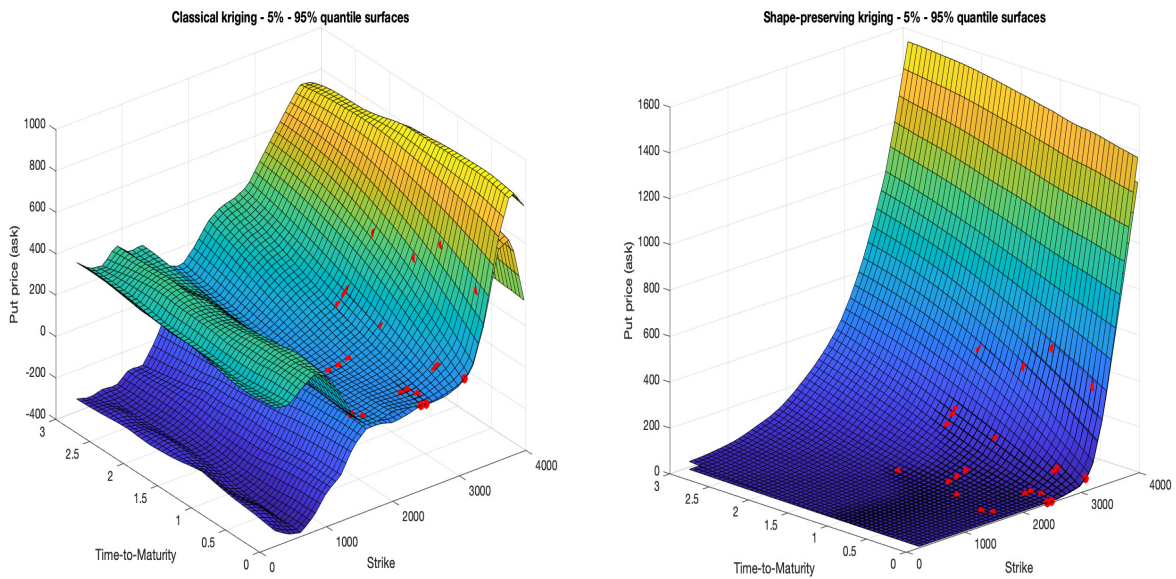
FIGURE 6.7: 5% and 95% estimated pointwise quantiles of the fitted GPs with classical kriging (left) vs constrained kriging (right).

In Figure 6.8, we present a 5% and 95% estimated pointwise quantiles of the constrained fitted Gaussian process with extrapolation in the time-to-maturities direction by adding two years. Unsurprisingly, this leads to an increase of the confidence interval due to the fact that no price is observed for maturities greater than two years.

FIGURE 6.8: 5% and 95% estimated pointwise quantiles of the constrained fitted GPs with
extrapolation in the time-to-maturities direction (adding 2 years).

6.5   Conclusion

In this chapter we have adopted the constrained kriging techniques in the aim of constructing implied
volatility surface. We have shown that constrained kriging techniques allow to estimate the option prices
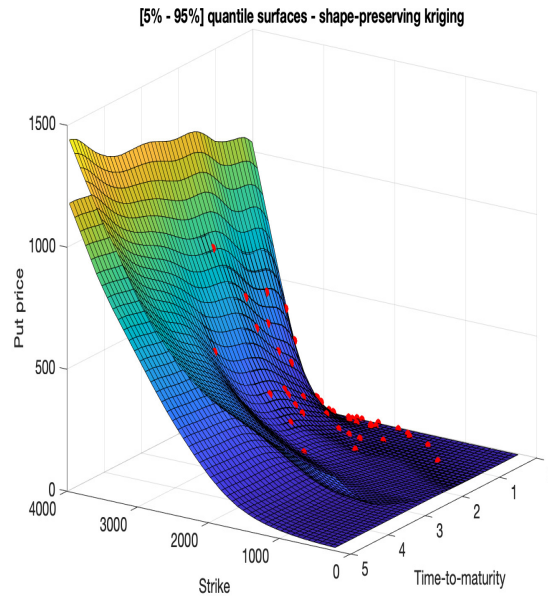in a market where only noisy prices are available. Our approach incorporates all the characteristics of
surface construction problem, such as incomplete information, indirect observation, noisy measurement and
shape constraints. The most likely measurement noises have been computed. We have shown, through a
comparison study, that constrained kriging is more adapted for implied volatility surface construction than
the classical one. In this study, we did not incorporate dividends and interest rates in the kriging methods.
This setting requires a small extension of the method. We do not also compare our method to the alternative
approaches such as the SSVI and the constrained neural network. In our next chapter, we explore those
issues in the SPX European puts data on 18th May 2019.

Chapter 7

Kriging vs Neural Network for local volatility construction

This chapter is a slightly modified version of the paper A. Cousin, S. Crepey, M.F. Dixon, and D. Gueye, Beyond Surrogate Modeling: Learning the Local Volatility Via Shape Constraints, under review, 2020.

7.1 Introduction

Local volatility surface is an important tool used by practitioners for pricing and hedging exotic options. Its construction can be done in two ways. As explained in Figure 7.1, one can derive it either from the Black-Scholes implied volatility surface by using Gatheral formula, or from the no-arbitrage continuous surface of option prices through Dupire formula.

Deep learning and kriging for option pricing have emerged as novel methodologies for fast computations with applications in calibration and computation of sensitivities. However, most of these approaches do not enforce any no-arbitrage conditions, and the subsequent local volatility surface is never considered. As shown by results of Chapter 6, constrained kriging is a suitable tool for constructing option prices and quatifying incertainty in the presence of noisy data. In this chapter, we propose a fully viable benchmarking approach for kriging, by building on Ackerer, Tagasovska, and Vatter (2019) using a neural network approximation of the local volatility implied variance formula (i.e. the Dupire formula restated in terms of implied variance) to stabilize the local volatility surface. This approach is in contrast to Chataigner, Crépey, and Dixon (2020) who consider the analogous approach of using the Dupire formula with price interpolation which is included in the numerical results section.

Regularization tends to reduce static arbitrage violation on the training set but does not exclude violation on the testing set. This is a by product of using stochastic gradient descent. Unlike interior point methods, which use barrier functions to avoid leaving the feasible set (but are not applicable to neural networks), stochastic gradient descent does not ensure saturation of the penalization. A manifestation of the limitations of the soft constraints approach is the instability of the local volatility surface which has a tendency to be quite irregular, especially away from training points.

As a focal point of the chapter, we emphasize that a single optimization of a GPs or neural network jointly yields a shape-constrained price or implied volatility estimation together with the local volatility surface. The extraction of a nonparametric representation of the local volatility surface, is not only intrinsically useful for exotic option pricing, but also serves as a stabilization approach for price or implied volatility estimation in a neural network (such stabilization approach is not needed for the GPs approach).

In summary, we shall (i) use a hard-constrained kriging Dupire formula to derive a local volatility surface from prices; and (ii) a soft-constrained NN local volatility implied variance formula to derive a local volatility surface from Black-Scholes implied volatilities (see blue arrow in Fig. 7.1). As we illustrate later in this chapter, such a local volatility surface shall in fact be jointly derived and, at the same time, further regularized.

The remainder of the chapter is outlined as follows. Section 7.2 describes the kriging approach to probabilistic modeling of local volatility with no-arbitrage constraints. Then in Section 7.3 we introduce the neural network approach to local volatility modeling from implied volatility. Details of our numerical experiment setup and results comparing the performance of kriging are given in Section 7.4. Finally, Section 7.5 concludes with further directions for research.
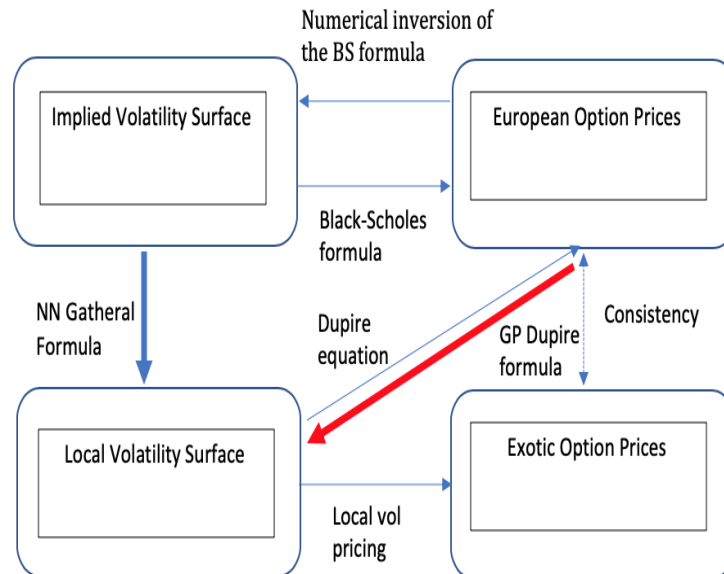
FIGURE 7.1: *Mathematical connections between option prices, implied, and local volatility, and the goal of this chapter, namely to either use the Dupire formula with Gaussian processes to jointly approximate the vanilla price and local volatility surfaces, or use the Gatheral formula with neural networks to jointly approximate the implied volatility and local volatility surfaces.*

7.2 Kriging the Local Volatility from Prices

This section is devoted to the construction of local volatility surface using kriging. Since, we use dividends and interest rates, it will be interesting to understand how to extend the kriging techniques presented in the previous chapters for no-arbitrage option price surface construction in presence of dividends and interest rates. As such, we first introduce this extension in the next subsection before tackling the calibration of Dupire formula in the following one.

7.2.1 No-arbitrage conditions reformulation in Kriging framework

Given a stock S we consider an European vanilla put option on S, with maturity T and strike price X. We consider the following risk neutral dynamics of S

$$dS_t = S_t[(r(t) - q(t))dt + \sigma(t, S_t)dB_t], \ \forall t \geq 0 \tag{7.2.1}$$

where $B = (B_t)_{t \geq 0}$ is a Brownian Motion, $r(t)$ represents a deterministic short interest rate term structure of the corresponding economy and $q(t)$ a deterministic continuous-dividend-yields on S and $\sigma(t, S_t)$ the instantaneous volatility function of S and t. We denote respectively by $D_t = \exp\left(-\int_0^t r(s)ds\right)$ and $H_t = \exp\left(-\int_0^t q(s)ds\right)$ the discount and dividend factors.

Without any restriction, we only pay particular attention to the put prices constraints since the the call prices constraints can be derived from the Call-Put parity. As such, let us denote by $P(T, X)$ and $C(T, X)$,

the prices at time $t = 0$ of, respectively, the European put and the European Call options on the underlying asset S with maturity T and strike price X. We recall the Call-Put parity which can be expressed as

$$C(T, X) - P(T, X) = S_0 H(T) - X D(T) \tag{7.2.2}$$

Once, we know $C(T, X) - P(T, X)$ for $T = T_1, ..., T_j$, with $j \in \mathbb{N}$, we can regress $C(T, X) - P(T, X)$ on the direction of X, for $X = X_1, ..., X_{n_j}$, with $n_j \in \mathbb{N}$ in order to obtain an approximation of $H(T_j)$ and $D(T_j)$. One further constructs $H(T)$ and $D(T)$, for any T by interpolation.

Definition 63 *Under the arbitrage-free conditions, the transformed process $M = (M_t)_{t \geq 0}$, given by*

$$M_t = \exp\left(-\int_0^t (r(s) - q(s)) ds\right) S_t,$$

is a martingale (see, e.g., Shreve, 2004, Subsection 5.5.1) and the price at time $t = 0$ of the put option is given by

$$P(T, X) = \mathbb{E}\left[\exp\left(-\int_0^T r(s) ds\right)(X - S_T)^+\right]. \tag{7.2.3}$$

The equality (7.2.3) can be expressed as

$$P(T, X) = H_T \mathbb{E}\left[\left(\frac{D_T}{H_T} X - M_T\right)^+\right] \tag{7.2.4}$$

with

$$M_T = \frac{D_T}{H_T} S_T.$$

By setting

$$k = \frac{D_T}{H_T} X, \tag{7.2.5}$$

one has, from (7.2.4),

$$P(T, X) = H_T \widehat{P}(T, k) \tag{7.2.6}$$

where

$$\widehat{P}(T, k) = \mathbb{E}\left[(k - M_T)^+\right]. \tag{7.2.7}$$

Proposition 64 *In presence of dividend and interest rate, a collection of Put prices $P(T, X)_{T \geq 0, X \geq 0}$ is arbitrage-free if and only if the process $\widehat{P}(T, k)_{T, k}$ given by (7.2.7) verifies the following conditions*

1. *$\widehat{P}(T, .)$ is a convexe function, for any $T \geq 0$,*

2. *$\widehat{P}(., k)$ is an non-decreasing function, for any $k \geq 0$,*

3. *$\widehat{P}(T, 0) = 0, \forall T \geq 0$,*

4. *$\frac{\partial}{\partial k} \widehat{P}(T, 0) = 0, \forall T \geq 0$,*

5. *$\lim_{k \to \infty} \frac{\partial}{\partial k} \widehat{P}(T, k) = k - M_0$,*

6. *$\widehat{P}(0, k) = (k - M_0)^+, \forall k \geq 0$.*

Accordingly, learning the put price function $P(\cdot, \cdot)$ is equivalent to learning the reduced price $\widehat{P}$ given the observations and the constraints.

7.2.2 GPs for calibration of Dupire formula

By considering the dynamics of S given in (7.2.1) with constant interest rate r and constant dividend-yields q, we obtain the Dupire (1994) formula

$$\frac{\sigma^2(T, X)}{2} = \frac{\partial_T P(T, X) + (r - q)X \, \partial_X P(T, X) + qP(T, X)}{X^2 \partial^2_{X^2} P(T, X)} \tag{7.2.8}$$

which establishes a relation between the put price and the volatility. The function $\sigma : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is known as local volatility and is consistent with no-arbitrage market prices for any options on S. Thus, the construction of $\sigma(\cdot, \cdot)$ given the market put prices $P(T_1, X_1), ..., P(T_n, X_n)$, observed for the finite number of pairs $(T_i, X_i)_{i=1,...,n}$ can be done by first constructing a no-arbitrage put price surface $P : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ which is compatible with these observed prices and then use the Dupire (1994) formula to get the local volatility surface.

As we show above, learning the put price surface P corresponds to leaning the reduced price $\widehat{P}$ and using (7.2.7). Fortunately, the Dupire formula can be rewritten in term of $\widehat{P}$. Indeed, one has from the (7.2.6) the following derivatives

$$\partial_T P(T, X) = \partial_T[H_T \widehat{P}(T, k)] = \partial_T H_T \, \widehat{P}(T, k) + H_T \, \partial_T \widehat{P}(T, k),$$

$$\partial_X P(T, X) = \partial_X[H_T \, \widehat{P}(T, k)] = H_T \, \partial_k \widehat{P}(T, k) \, \partial_X k = D_T \, \partial_k \widehat{P}(T, k),$$

$$\partial^2_{X^2} P(T, X) = \partial_K[D_T \, \partial_k \widehat{P}(T, k)] = D_T \partial^2_{k^2} \widehat{P}(T, k) \partial_K k = \frac{D_T^2}{H_T} \partial_k \widehat{P}(T, k).$$

Replacing these derivatives into (7.2.8) leads to

$$\frac{\sigma^2(T, X)}{2} = \frac{\partial_T \widehat{P}(T, k)}{k^2 \partial^2_{k^2} \widehat{P}(T, k)} := \mathrm{dup}(T, k). \tag{7.2.9}$$

which implies that constructing the local volatility surface can be done by constructing the reduced price surface and using Dupire's formula to get the whole local volatility surface.

Our construction consists in using the constrained kriging presented in Chapter 6 to learn the reduced price $\widehat{P}$, given in Proposition 64, which must be compatible to the market fit. In other terms, for a given quotation date, we construct reduced put price surfaces $(T, k) \to \widehat{P}(T, k)$ satisfying the arbitrage-free conditions in Proposition 64 from n noisy observations $y = [y_1, \ldots, y_n]^\top$ of function $\widehat{P}$ at input points $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]$. The input points $\mathbf{X}_i = (T_i, k_i)$, $i = 1, \ldots, n$ correspond to observed term-to-maturities and reduced option strikes. Then, the construction of the reduced price function should be compatible with the following market fit condition

$$y = \widehat{P}(\mathbf{X}) + \boldsymbol{\varepsilon}, \tag{7.2.10}$$

where $\widehat{P}(\mathbf{X}) := [\widehat{P}(\mathbf{X}_1), \ldots, \widehat{P}(\mathbf{X}_n)]^\top$ is the vector composed of reduced put prices at observation points $\mathbf{X}_1, \ldots, \mathbf{X}_n$. The additive noise term $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_n]^T$ is assumed to be a zero-mean Gaussian vector, independent from $\widehat{P}$, and with an homoscedastic covariance matrix given as $\sigma^2_{\mathrm{noise}} I_n$, where I_n is the identity matrix of dimension n.

7.2.3 The methodology of the construction

The construction of the local volatility consists first of all in kriging the reduced price $\widehat{P}$ which is compatible with (7.2.10) and which respects the conditions of Proposition 64 and then using the Dupire's formula.

7.2.3.1   kriging the reduced price

Kriging the reduced price $\widehat{P}$ can be done as explained in Section 6.3.2 of Chapter 6. By means, we consider a discretized version of the input space $\mathcal{D}$ as a $N = (N_T + 1) \times (N_k + 1)$ regular grid $\mathcal{D}^N := \{(u_i, v_j) \mid u_i = ih_T,\ v_j = jh_k,\ i = 0, \ldots, N_T,\ j = 0, \ldots, N_k\}$, where $h_T = \frac{1}{N_T}$ and $h_k = \frac{1}{N_k}$. For each knot (u_i, v_j), we introduce the hat basis function $\phi(T, k)$ over $\mathcal{D}$, as the following tensor product

$$\phi_{i,j}(T, k) := \max(1 - \frac{|T - u_i|}{h_T}, 0) \max(1 - \frac{|k - v_j|}{h_k}, 0) \,.$$

so that $supp(\phi_{ij}) = [k_{i-1}, k_{i+1}] \times [T_{j-1}, T_{j+1}]$.

Let $V = H^1(\mathcal{D})$ denote the space of $\widehat{P}$ and $V^N \subset V$ denote the finite dimensional linear subspace spanned by the N linearly independent basis functions so that $V^N = \text{span}\{\phi_{0,0}, \ldots, \phi_{N_T,N_k}\}$ with $dim(V^N) = N$.

The surface $\widehat{P} \in V$ is projected onto V^N:

$$\widehat{P}^N(T, k) = \sum_{i=0}^{N_T} \sum_{j=0}^{N_k} \widehat{P}(u_i, v_j) \phi_{i,j}(T, k), \quad \text{for any } (T, k) \in \mathcal{D}. \tag{7.2.11}$$

$\widehat{P}^N \in V^N$ is a bilinear quadrilateral finite element approximation of the values of $\widehat{P}$ at knots $(u_i, v_j)_{i,j}$. If we denote $\xi_{i,j} := \widehat{P}(u_i, v_j)$, for $i = 0, \ldots, N_T$, $j = 0, \ldots, N_k$, then $\boldsymbol{\xi} = [\xi_{0,0}, \ldots, \xi_{i,j}, \ldots, \xi_{N_T,N_k}]^\top$ is a zero-mean Gaussian vector with $N \times N$ covariance matrix Γ^N such that $\Gamma^h_{i_1,i_2} = \boldsymbol{\kappa}((u_{i_1}, v_{j_1}), (u_{i_2}, v_{j_2}))$, for any two grid index pairs (i_1, j_1) and (i_2, j_2) corresponding to global indices i_1 and i_2 respectively. Let $\boldsymbol{\phi}(T, k)$ denote the vector of size N given by

$$\boldsymbol{\phi}(T, k) = [\phi_{0,0}(T, k), \ldots, \phi_{i,j}(T, k), \ldots, \phi_{N_T,N_k}(T, k)].$$

The equality (7.2.11) can be written in the following matrix form

$$\widehat{P}^N(T, k) = \boldsymbol{\phi}(T, k) \cdot \boldsymbol{\xi}$$

so that when denoting $\boldsymbol{\Phi}(\mathbf{X})$ the $n \times N$ matrix of basis function in which, each row ℓ corresponds to the vector $\boldsymbol{\phi}(T_\ell, k_\ell)$, one has $\widehat{P}^N(\mathbf{X}) = \boldsymbol{\Phi}(\mathbf{X}) \cdot \boldsymbol{\xi}$, with $\widehat{P}^N(\mathbf{X}) := [\widehat{P}^N(\mathbf{X}_1), \ldots, \widehat{P}^N(\mathbf{X}_n)]^\top$.

The main steps of constructing the reduced price surface are

1. Find GPs hyperparameters by maximizing the marginal log-likelihood of the (made finite dimensional) put price surface $\widehat{P}^N$ with respect to a training dataset.

2. Find the MAP (mode of the posterior distribution) $\tilde{\boldsymbol{\xi}}$ of the (finite reduction of the) GPs put price surface $\boldsymbol{\xi}$ by quadratic minimization, using the hyperparameters found in step 1.

3. Use $\tilde{\boldsymbol{\xi}}$ to initialize a Hamilton MC sampler and sample realizations of $\boldsymbol{\xi}$ or, equivalently, $\widehat{P}^N$, from the posterior distribution of the GPs put price surface.

7.2.3.2   Turning into local volatility surface

Once the kriging of $\widehat{P}^N$ has been done as presented above, the realizations of the local volatility surface can be obtained by deriving their corresponding finite element approximations for each realization of $\widehat{P}^N$.

For this purpose, the finite dimensional approximation $\widehat{P}^N$ should be convex with respect to the strike direction. However, note that $\widehat{P}^N$ is not differentiable, because our basis functions are only of class $\mathcal{C}^0$. To address this issue, we may use quite a large number of nodes in strike direction in order to get the local volatility surface.

An other solution is to use a regularization technique which consists in formulating a weak form of the Dupire equation and construct the local volatility surface using this finite element method. This regularization can

be done as follows.

Recall from Eq. 7.2.9 that the Dupire formula in reduced variables (T, k) specifies the local half-variance surface $\text{dup} \in W = H^0(\mathcal{D})$. We construct a piecewise constant approximation of the local volatility. Note that the approach described in this section can be readily extended to smoother approximations, but requires more elaborate calculations. The finite dimensional linear solution subspace $W^N \subset W$ is spanned by N linearly independent basis functions so that $W^N = span\{\psi_{0,0}, \ldots, \psi_{N_T-1, N_k-1}\}$ with $dim(W^N) = N = N_T \times N_k$. The piecewise constant basis functions are given as $\psi_{i,j} := \mathbb{1}_{[u_i, u_{i+1}) \times [v_j, v_{j+1})}$ with $supp(\psi_{i,j}) = [u_i, u_{i+1}) \times [v_j, v_{j+1})$ so that

$$\text{dup}^h(T, k) = \sum_{i,j} \psi_{i,j}(T, k)\text{dup}_{i,j} = \boldsymbol{\psi} \cdot \mathbf{dup}.$$

If $\widehat{P} \in V \subset H^2(\mathcal{D})$, then under a change of variables $\kappa = ln\ k$ we could write the Dupire equation as:

$$\partial_{\kappa^2}^2 \widehat{P}(T, \kappa)\text{dup}(T, \kappa) = \partial_T \widehat{P}(T, \kappa). \tag{7.2.12}$$

Instead, we must take the weak form of the Dupire equation to solve for the local half variance by integrating against a test function $v^N \in V^N$. Substituting $\widehat{P}^N$ for $\widehat{P}$ and dup^h for dup we write:

$$\int_{\mathcal{D}} \partial_{\kappa^2}^2 \widehat{P}^N(T, \kappa)\text{dup}^h(T, \kappa)v^N d\kappa dT = \int_{\mathcal{D}} \partial_T \widehat{P}^N(T, \kappa)v^N d\kappa dT, \forall v^N \in V^N,$$

and after integration by parts to eliminate the second derivative on $\widehat{P}$,

$$- \int_{\mathcal{D}} \partial_\kappa \widehat{P}^N(T, \kappa)\partial_\kappa(\text{dup}^h(T, \kappa)v^N)d\kappa dT + \int_{T \in (0, T_{max}]} \partial_\kappa \widehat{P}^N(T, \kappa)\text{dup}^h(T, \kappa)v^N \mid_{\kappa_{min}}^{\kappa_{max}} dT$$

$$= \int_{\mathcal{D}} \partial_T \widehat{P}^N(T, \kappa)v^N d\kappa dT, \forall v^N \in V^N,$$

where the second integral on the left hand side is a maturity boundary integral. Adopting global indices $\imath, \jmath \in \{0, \ldots, N-1\}$, substituting the form of $\widehat{P}^N$ and choosing $v^N = \phi_\jmath$:

$$- \sum_\imath \xi_\imath \int_{T_{i-1}}^{T_{i+1}} \int_{\kappa_{j-1}}^{\kappa_{j+1}} \partial_\kappa \phi_\imath(T, \kappa)\partial_\kappa \left(\text{dup}^h(T, \kappa)\phi_\jmath(T, \kappa)\right) d\kappa dT +$$

$$\sum_\imath \xi_\imath \int_{T_{i-1}}^{T_{i+1}} \partial_\kappa \phi_\imath(T, \kappa)\text{dup}^h(T, \kappa)\phi_\jmath(T, \kappa) \mid_{\kappa_{min}}^{\kappa_{max}} dT$$

$$= \sum_\imath \xi_\imath \int_{T_{i-1}}^{T_{i+1}} \int_{\kappa_{j-1}}^{\kappa_{j+1}} \partial_T \phi_\imath(T, \kappa)\phi_\jmath(T, \kappa)d\kappa dT, \ \forall \imath, \jmath \in \{0, \ldots, N-1\}.$$

Each double integral is defined over a quadrant consisting of four elements. Since dup^h is only constant over each element in the quadrant, we can separate the double integral into integrals over each element to give the linear system

$$H(\mathbf{dup})\boldsymbol{\xi} = F\boldsymbol{\xi}, \tag{7.2.13}$$

where the stiffness matrix H and RHS matrix F has a nine-stencil and 6-stencil respectively and are given in the appendix 8. This form is inconvenient for finding $\mathbf{dup}$ and we can rearrange the computations to solve for $\mathbf{dup}$ to give

$$A(\boldsymbol{\xi})\mathbf{dup} = F\boldsymbol{\xi}, \tag{7.2.14}$$

where $A \in \mathbb{R}^{N \times N}$ is a quad-diagonal rectangular matrix (see appendix for further details) and $\mathbf{dup}$ is found by the pseudo-inverse of A. The matrix A can be viewed as a combination of time weighted local differencing over $\boldsymbol{\xi}$ w.r.t. κ and subsequent averaging of dup^h over the four elements in each quadrant. We further note

that the linear system in Eq. 7.2.14 provides the opportunity for further regularization techniques such as, for examples, preconditioners.

In order to constrain the derivatives to be greater than zero during fitting, it is necessary to evaluate the weak form of the second derivative of $\widehat{P}^N$ w.r.t. k. The weak form of the first derivative of $\widehat{P}^N$ w.r.t. T is equivalent to the pointwise derivative because $\widehat{P}^N \in V^N \subset H^1(\mathcal{D})$. Returning to grid indices (i,j) we see that

$$\partial_T \widehat{P}^N(T,k) = \sum_{ij} \phi_i'(T)\phi_j(k)\xi_{ij} = \frac{1}{h_k}(\xi_{i+1,j} - \xi_{i,j}), (T,k) \in [T_i, T_{i+1}] \times [k_j, k_{j+1}]$$

is just the 2-stencil corresponding to a forward difference over any element. Hence we may freely choose where to evaluate finite differences in each element. The weak second derivative

$$-\int_{k=k_{min}}^{k_{max}} \widehat{P}^N(T,k) v^N(T,k) dk = -\sum_i \xi_i \int_{k_{j-1}}^{k_{j+1}} \partial_k \phi_i(T,k)\partial_k \phi_j(T,k) dk = \sum_i H_{j,i} \xi_i,$$

can be written as $H\boldsymbol{\xi}$, where the stiffness matrix is a 3-stencil corresponding to a second order finite difference operator.

7.3   Neural networks implied volatility metamodeling

Our second goal is to use neural nets (NN) to construct a continuous Σ put surface $\Sigma : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$, interpolating Σ market quotes Σ_* up to some error term, both being stated in terms of a put option maturity T and log-(forward) moneyness $\kappa = \log(\frac{k}{S_0}) = \log\left(\frac{X}{S_0}\right) - (r-q)T$. The advantage of using implied volatilities rather than prices (as previously done in Chataigner, Crépey, and Dixon (2020)), both being in bijection via the Black-Scholes put pricing formula as well known, is their lower variability, hence better performance as we will see.

The corresponding local volatility surface σ is given by the following local volatility implied variance formula, i.e. the Dupire formula stated in terms of the implied total variance $\Theta(T,\kappa) = \Sigma^2(T,\kappa)T$ (assuming Θ of class $\mathcal{C}^{1,2}$ on $\{T > 0\}$):[1]

$$\sigma^2(T,X) = \frac{\partial_T \Theta}{1 - \frac{\kappa}{\Theta}\partial_\kappa \Theta + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{\Theta} + \frac{\kappa^2}{\Theta^2}\right)(\partial_\kappa \Theta)^2 + \frac{1}{2}\partial_{\kappa^2}\Theta}(T,\kappa) =: \frac{\text{cal}_T(\Theta)}{\text{butt}_k(\Theta)}(T,\kappa) \qquad (7.3.1)$$

We use a feedforward NN with weights $\mathbf{W}$, biases $\mathbf{b}$ and smooth activation functions for parameterizing the Σ (hence the total variance), which we denote by

$$\Sigma = \Sigma_{\mathbf{W},\mathbf{b}}^\Theta = \Theta_{\mathbf{W},\mathbf{b}}.$$

The terms $\text{cal}_T(\Theta_{\mathbf{W},\mathbf{b}})$ and $\text{butt}_k(\Theta_{\mathbf{W},\mathbf{b}})$ are available analytically, by automatic differentiation, which we exploit below to penalize calendar spread arbitrages, i.e. negativity of $\text{cal}_T(\Theta)$, and butterfly arbitrage, i.e. negativity of $\text{butt}_k(\Theta)$.

The training of NNs is a non-convex optimization problem and hence does not guarantee convergence to a global optimum. We must therefore guide the NN optimizer towards a local optima that has desirable properties in terms of interpolation error and arbitrage constraints. This motivates the introduction of an arbitrage penalty function into the loss function to select the most appropriate local minima. An additional challenge is that maturity-log moneyness pairs with quoted option prices are unevenly distributed and the NN may favor fitting to a cluster of quotes to the detriment of fitting isolated points. Consequently large pointwise errors may arise where the NN has favored a local minima with low interpolation accuracy and no arbitrage violation. To remedy this non-uniform data fitting problem, we propose a novel solution which involves re-weighting the observations by the Euclidean distance between neighboring points.

---

[1]This follows from the Dupire formula by simple transforms detailed in Gatheral (2011, p.13).

More precisely, given n observations $\chi_i = (T_i, \kappa_i)$ of maturity-log moneyness pairs and of the corresponding market implied volatilities $\Sigma_*(\chi_i)$, we construct the $n \times n$ distance matrix where each coefficient $d(\chi_i, \chi_j)$ is the euclidean distance between points :

$$d(\chi_i, \chi_j) = \sqrt{(T_j - T_i)^2 + (\kappa_j - \kappa_i)^2}.$$

We then define the loss weighting w_i for each point χ_i as the distance with the closest point:

$$w_i = \min_{j, j \neq i} d(\chi_i, \chi_j).$$

This weighting aims at reducing error for any isolated points. In order to adjust the weight of penalization, we multiply our penalties by the weighting mean $\mu_w := \frac{1}{h} \sum_i w_i$. Learning the weights $\mathbf{W}$ and biases $\mathbf{b}$ to the data subject to no arbitrage soft constraints (i.e. with penalization of arbitrages), takes the form of the following (nonconvex) loss minimization problem:

$$\underset{\mathbf{W}, \mathbf{b}}{\arg\min} \sqrt{\frac{1}{n} \sum_i \left(w_i \frac{\Sigma_{\mathbf{W}, \mathbf{b}}(\chi_i) - \Sigma_*(\chi_i)}{\Sigma_*(\chi_i)} \right)^2} + \frac{\mu_w}{h} \sum_{\xi \in \mathcal{D}_h} \lambda^\intercal \mathcal{R}(\Theta_{\mathbf{W}, \mathbf{b}})(\xi), \qquad (7.3.2)$$

where $\lambda = [\lambda_1, \lambda_2, \lambda_3]^\top \in \mathbb{R}_+^3$ and

$$\mathcal{R}(\Theta) = [\mathrm{cal}_T^-(\Theta), \mathrm{butt}_k^-(\Theta), \left(\frac{\mathrm{cal}_T}{\mathrm{butt}_k}(\Theta) - \bar{a} \right)^+ + \left(\frac{\mathrm{cal}_T}{\mathrm{butt}_k}(\Theta) - \underline{a} \right)^-]^\top$$

is a regularization penalty vector evaluated over a penalty grid $\mathcal{D}_h$ with $h = 50 \times 100$ nodes, which extends well beyond the unit square domain of the IV interpolation. In the unscaled moneyness and maturity coordinates, the domain of the penalty grid is $[0.5, 2] \times [0.005, 10Y]$. This is intended so that the penalty term penalizes arbitrages outside of the domain used for IV Interpolation. Even on such an extended penalty grid, we found no arbitrage violation in our experiments (after training $\mathrm{cal}_T(\Theta_{\mathbf{W}, \mathbf{b}})$ and $\mathrm{butt}_k(\Theta_{\mathbf{W}, \mathbf{b}})$ are even > 0 at all nodes of $\mathcal{D}_h$).

Note that the the error criterion is calculated as a root mean square error on relative difference, chosen here, so that it does not discriminate high or low implied volatilities.

The first two elements in the penalty vector favor the no-arbitrage conditions and the third element favors desired lower and upper bounds $0 < \underline{a} < \bar{a}$ (constants or functions of T) on the estimated local variance $\sigma^2(T, X)$. Suitable values of the "Lagrange multipliers" λ, ensuring the right balance between fit to the market implied volatilities and the constraints, is obtained by grid search. Of course a soft constraint (penalization) approach does not fully prevent arbitrages. However, for large λ, arbitrages are extremely unlikely to occur (except perhaps very far from $\mathcal{D}$).

7.4 Numerical results

7.4.1 Experimental design

Our training set is prepared using SPX European puts with different available strikes and maturities ranging from 0.005 to 2.5 years, listed on 18th May 2019, with $S_0 = \$2859.53$. Each contract is listed with a bid/ask price and an implied volatility corresponding to the mid-price. The associated interest rate is constructed from US treasury yield curve and dividend yield curve rates are then obtained from call/put parity applied to the option market prices and forward prices. We preprocess the data by removing the shortest maturity options, with $T < 0.055$, and the numerically inconsistent observations for which the gap between the listed implied volatility and the implied volatility calibrated from mid-price with our interest/dividend curves exceeds 5% of the listed implied volatility. But we do not remove arbitrable observations. The preprocessed

training set is composed of 1720 market put prices. The testing set consists of a disjoint set of 1725 put prices.

All results for the GP method are based on using Matern $\nu = 5/2$ kernels over a $[0,1]^2$ domain with fitted kernel standard-deviation hyper-parameter $\hat{\sigma} = 185.7611$, length-scale hyper-parameters $\hat{\theta}_k = 0.3282$ and $\hat{\theta}_T = 0.2211$, and homoscedastic noise standard deviation, $\hat{\varsigma} = 0.6876$.[2] The grid of basis functions for constructing the finite-dimensional process p^h has 100 nodes in the modified strike direction and 25 nodes in the maturity direction.

Regarding the NN approach, we use a three layer architecture similar to the one based on prices (instead of implied volatilities in Section 7.3) in Chataigner, Crépey, and Dixon (2020), to which we refer the reader for implementation details.

7.4.2   Arbitrage-free SVI

We benchmark the machine learning results with the industry standard provided by the arbitrage free stochastic volatility inspired (SVI) model of Gatheral and Jacquier (2014). Under the "natural parameterization" $\mathrm{SVI} = (\Delta, \mu, \rho, \omega, \zeta)$, the implied total variance is given, for any fixed T, by

$$\Theta_{\mathrm{SVI}}(\kappa) = \Delta + \frac{\omega}{2}\left(1 + \rho(\kappa - \mu)\zeta + \sqrt{(\zeta(\kappa - \mu) + \rho)^2 + (1 - \rho^2)}\right). \tag{7.4.1}$$

SSVI is the parameterization of a full surface given as $\mathrm{SVI}_T = (0, 0, \rho, \Theta_T, \phi(\Theta_T))$ for each T, where Θ_T is the at-the-money total implied variance and we use for ϕ a power law function $\phi(\vartheta) = \frac{\eta}{\vartheta^\gamma (1+\vartheta)^{1-\gamma}}$. Gatheral and Jacquier (2014, Remark 4.4) provides sufficient conditions on SSVI parameters ($\eta(1 + |\rho|) \leq 2$ with $\gamma = 0.5$) that rule out butterfly arbitrage, whereas SSVI is free of calendar arbitrage when Θ_T is nondecreasing.

We calibrate the model as in Gatheral and Jacquier (2014):[3] First, a guess on SVI is obtained by fitting the SSVI model; Second, for each maturity in the training grid, the five SVI parameters are calibrated (starting in each case from the SSVI calibrated values). The IV is obtained for new maturities by a weighted average of the parameters associated with the two closest maturities in the training grid, T and U, say, with weights determined by Θ_T and Θ_U. The corresponding local volatility is extracted by finite difference approximation of (7.3.1).

Note that as, in practice, no arbitrage constraints are implemented for SSVI by penalization (see Gatheral and Jacquier, 2014, Section 5.2), the SSVI approach is in fact only practically arbitrage-free, much like our NN approach, whereas it is only the GP approach that is proven arbitrage-free.

7.4.3   Calibration results

Training times for SSVI, GP, and NNs are reported in the last line of Table 7.1 which, for completeness, also includes numerical results obtained by NN interpolation of the prices as per Chataigner, Crépey, and Dixon (2020). Because price based NN results are outperformed by IV based NN results we only focus on the IV based NN in the figures that follow, referring to Chataigner, Crépey, and Dixon (2020) for every detail on the price based NN approach. Again, in contrast to the SSVI and NNs which fit to mid-quotes, GPs fit to the bid-ask prices.

The GP implementation is in Matlab whereas the SSVI and NN approaches are implemented in Python. On our large dataset, the constrained GP has the longest training time. Training is longer for constrained SSVI than for unconstrained SSVI because of the ensuing amendments to the optimization routine. There are no static arbitrage violations observed for any of the constrained methods in neither the training or the testing grid. Unconstrained methods yield 18 violations with NN and 177 with SSVI on the testing set, out of a total of 1725 testing points, i.e. violations in 1.04% and 10.26% of the test nodes. The unconstrained GP

---

[2]When re-scaled back to the original input domain, the fitted length scale parameters of the 2D Matern $\nu = 5/2$ are $\hat{\theta}_k = 973.1901$ and $\hat{\theta}_T = 0.5594$.

[3]Building on https://www.mathworks.com/matlabcentral/profile/authors/4439546

IV RMSE (Price RMSE)	SSVI	GP	IV based NN	Price based NN	SSVI Unconstr.	GP Unconstr.	IV based NN Unconstr.	Price based NN Unconstr.
Calibr. fit on the training set	1.37% (2.574)	0.58% (0.338)	1.23% (2.897)	13.70% (9.851)	1.04% (2.691)	0.60% (0.321)	0.84% (2.163)	5.65 % (2.456)
Calibr. fit on the testing set	1.52% (2.892)	0.57% (0.355)	1.29% (2.966)	14.27% (10.347)	1.09% (2.791)	0.57% (0.477)	0.86% (2.045)	6.14% (2.888)
MC backtest	8.69% (22.826)	19.76% (74.017)	2.95% (4.989)	6.37% (11.764)	N/A	N/A	N/A	N/A
FD backtest	6.88% (33.545)	7.86% (35.270)	3.43% (11.976)	5.56% (26.785)	N/A	N/A	N/A	N/A
Comput. time (seconds)	33	856	191	185	1	16	76	229

TABLE 7.1: The IV and price RMSEs of the SSVI, GP and NN approaches. Last line: computation times.

approach yields constraint violations on 12.5% of the basis function nodes. The NN penalizations $(\mathrm{cal}_T)^-$ and $(\mathrm{butt}_k)^-$ vanish identically on $\mathcal{D}_h$ in the constrained case, whereas in the unconstrained case their averages across grid nodes in $\mathcal{D}_h$ are $(\mathrm{cal}_T)^- = 3.91 \times 10^{-6}$ and $(\mathrm{butt}_k)^- = 1.60 \times 10^{-2}$ with the IV based NN.

Fig. 7.2(a-b) respectively compare the fitted IV surfaces and their errors with respect to the market mid-implied volatilities, among the constrained methods. The surface is sliced at various maturities and the IVs corresponding to the bid-ask price quotes are also shown – the blue and red points respectively denote training and test observations.

We generally observe good correspondence between the models and that each curve typically falls within the bid-ask spread, except for the shortest maturity contracts where there is some departure from the bid-ask spreads for observations with the lowest log-moneyness values. We see on Fig. 7.2(b) that the GP IV errors are small and mostly less than 5 volatility points whereas NN and SSVI exhibit IV error that may exceed 15 volatility points. The green line and the red shaded envelopes respectively denote the GP MAP estimates and the posterior uncertainty bands under 100 samples per observation. The support of the posterior GP process assessed on the basis of 100 simulated paths of the GP captures the majority of bid-ask quotes. The GP MAP estimate occasionally corresponds to the boundary of the support of the posterior simulation. This indicates that the posterior truncated Gaussian distribution is heavily skewed for some points, and that the MAP estimate consequently saturates the AOA constraints. This indicates a tension between the AOA constraint and the bid-ask constraint which cannot be fully reconciled, most likely because some of the (short maturity) data are arbitrable (they are at least illiquid and hence noisy). See notebook for location of arbitrages in the unconstrained approach.

All constrained methods find a no-arbitrage IV surface. Fig. 7.2(a-b) suggest that the data may exhibit arbitrage at the lowest maturities where the methods depart from the bid-ask spreads. These observations are further supported in Fig. 7.3(a-b) which shows the corresponding methods without the no-arbitrage constraints. In Fig. 7.3(a-b) we observe that the estimated IVs now fall within close proximity of the bid-ask spreads–all methods exhibit an error typically less than 5 volatility points. Note that the y-axis has been scaled for each plot in Fig. 7.3(b) to accommodate the wide uncertainty band of the posterior for the unconstrained GP. Whereas the uncertainty band of the constrained GP spanned at most 10 volatility points, the uncertainty band of the unconstrained GP is an order of magnitude larger, sometimes spanning more than 100 volatility points.

Fig. 7.4 shows the local volatility surfaces that stem from the three constrained approaches. Fig. 7.4(a) shows the spiky local volatility surface generated by SSVI, capped at the 200% level for scaling convenience. Fig. 7.4(b) shows the capped local volatility surface constructed from the GP MAP price estimate. Fig. 7.4(c) shows the (complete) NN local volatility surface.
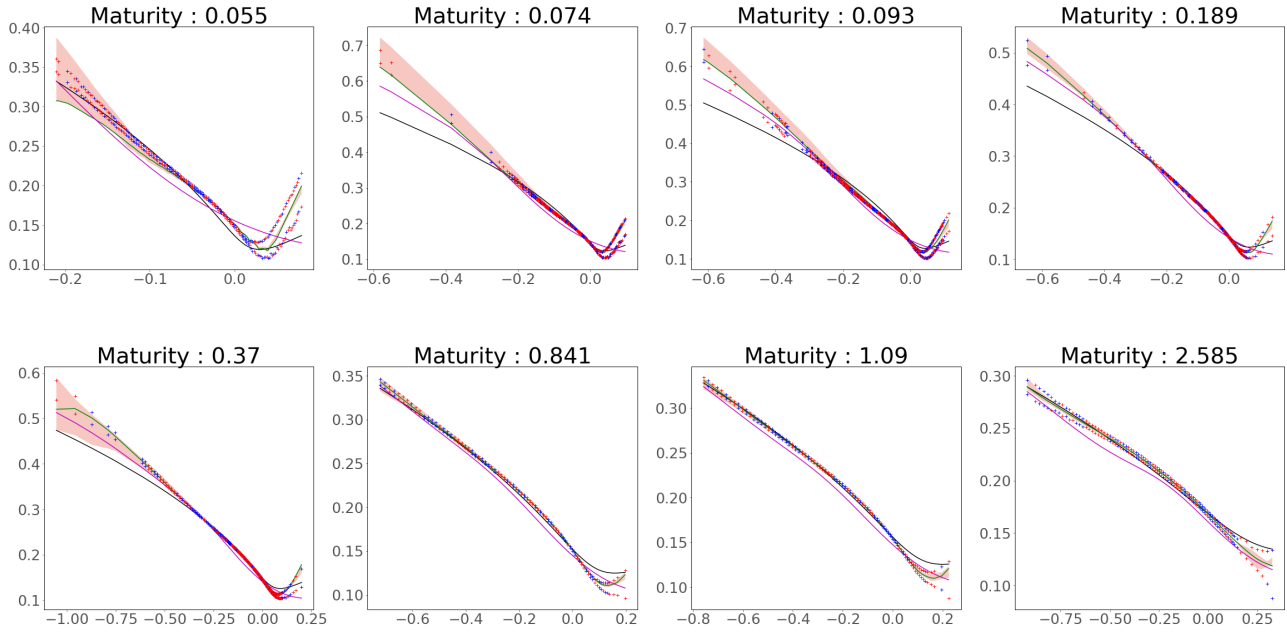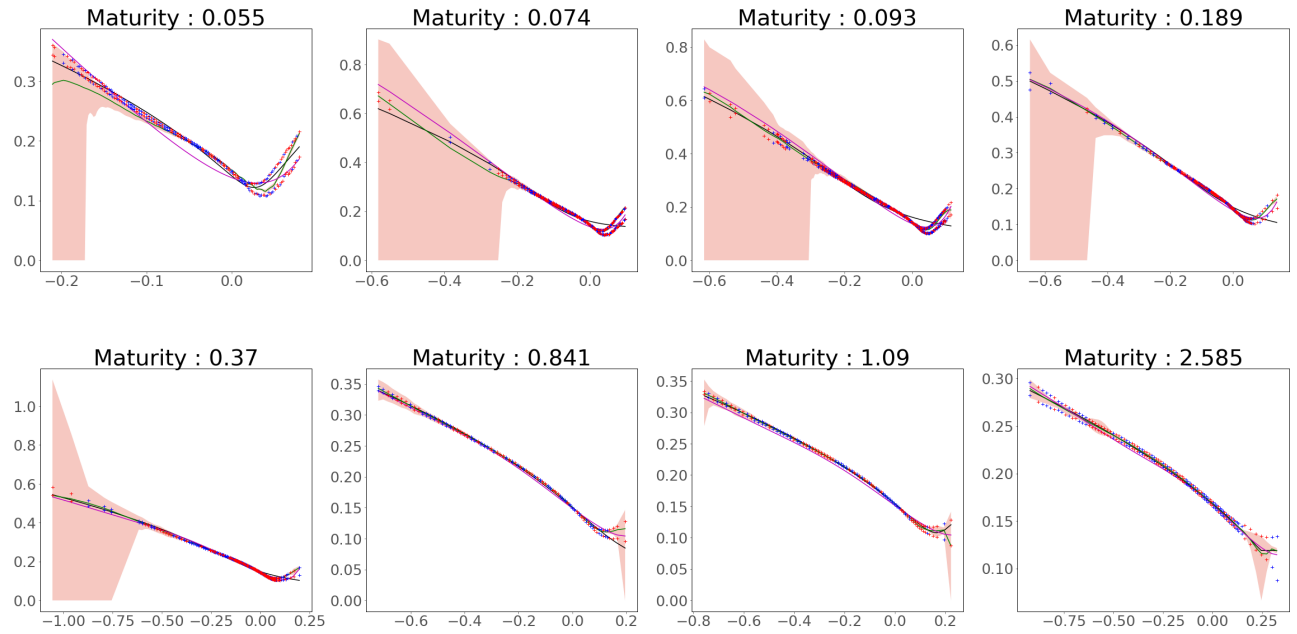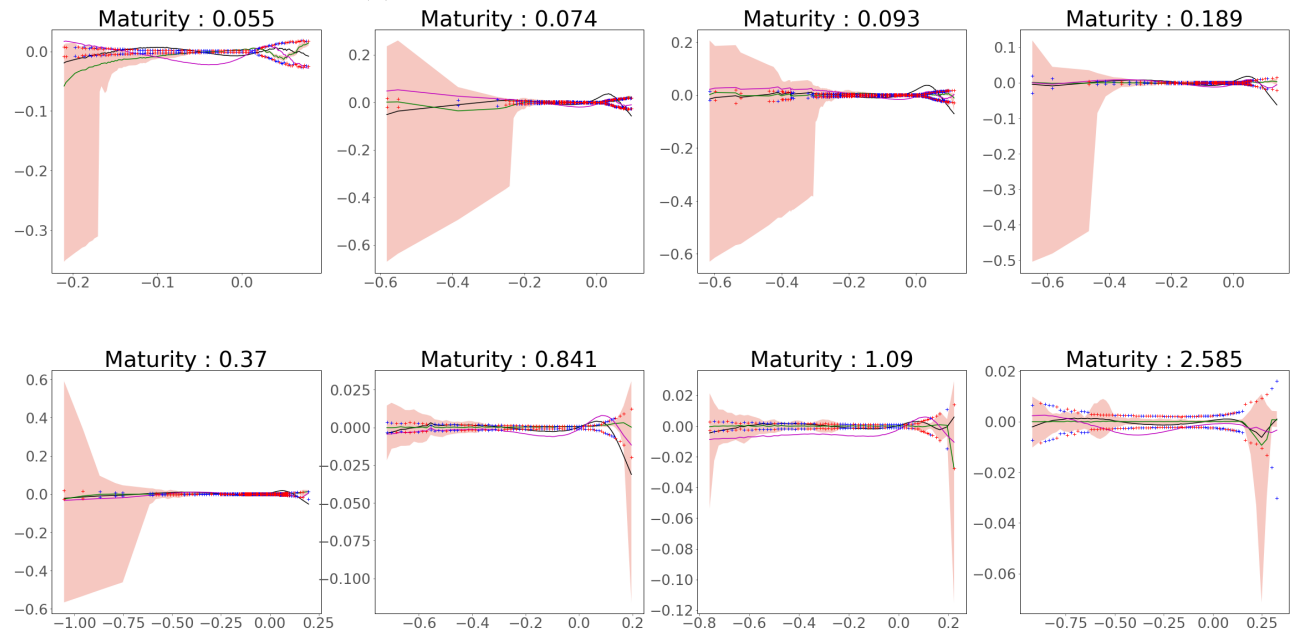
(A) Implied volatilities.
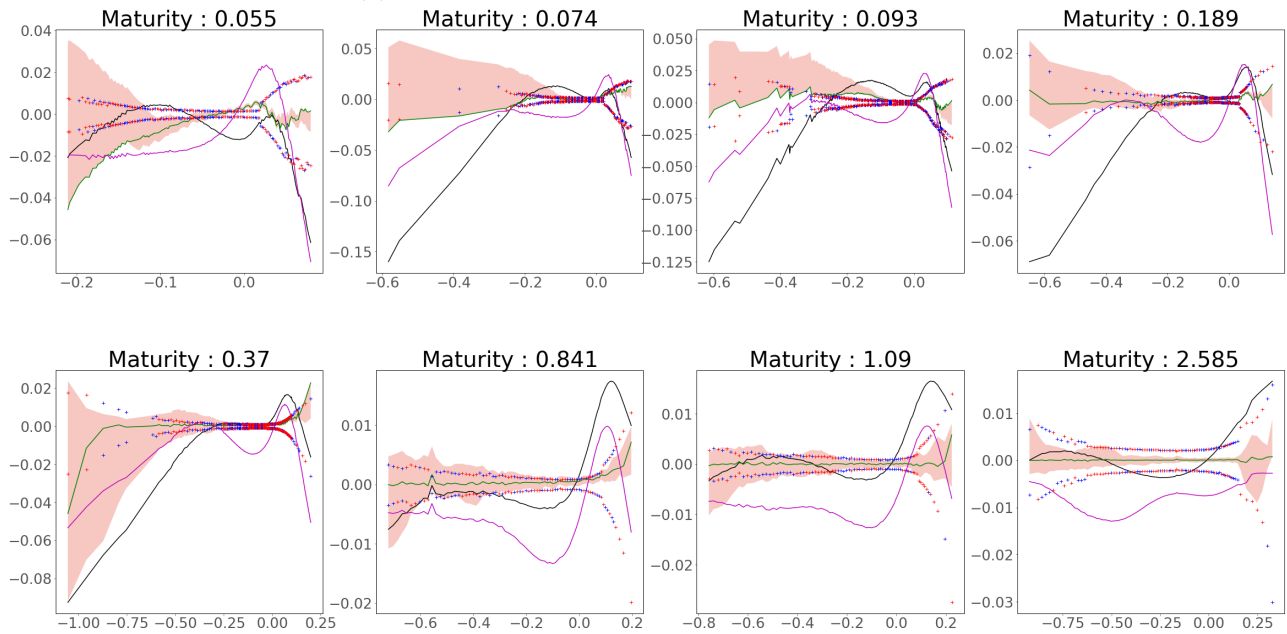
(B) Fitted IV errors with respect to mid-price IVs.

FIGURE 7.2: *Slices of constrained GP (green), NN (purple), and SSVI (black) models of SPX puts with training bid-asks IVs ($\pm$) and testing bid-asks IVs ($\frac{+}{}$) (the bid-ask IVs are reconstructed numerically from the corresponding bid-ask market prices). The shaded envelopes show 100 paths of the constrained GP's posterior.*

(A) Implied volatilities.

(B) Fitted IV errors with respect to mid-price IVs.

FIGURE 7.3: *Slices of unconstrained GP (green), NN (purple), and SSVI (black) models of SPX puts with training bid-ask IVs ($\pm$) and testing bid-ask IVs ($\dagger$) (the bid-ask IVs reconstructed numerically from the corresponding bid ask market prices). The shaded envelopes show 100 paths of the unconstrained GP's posterior.*
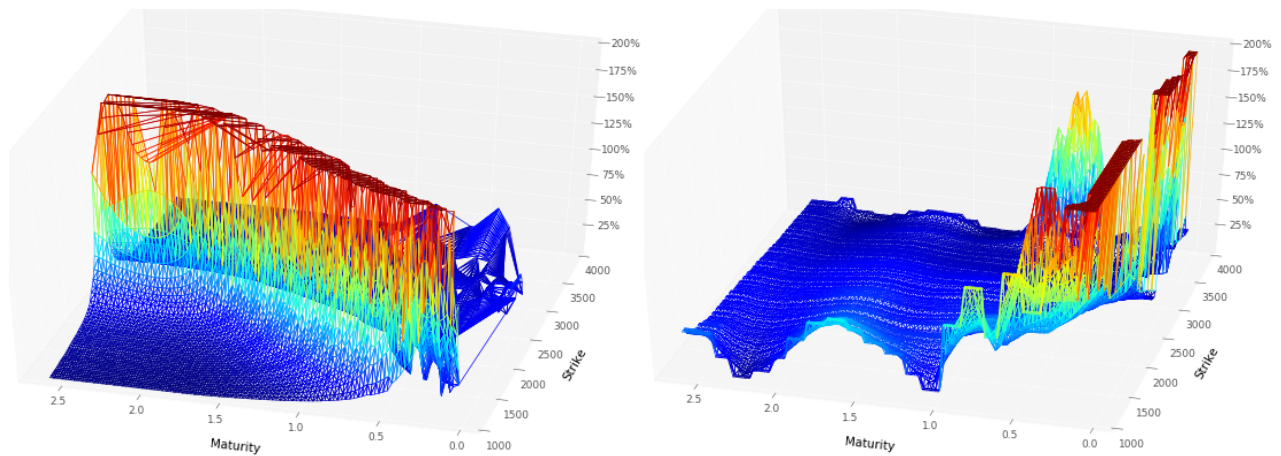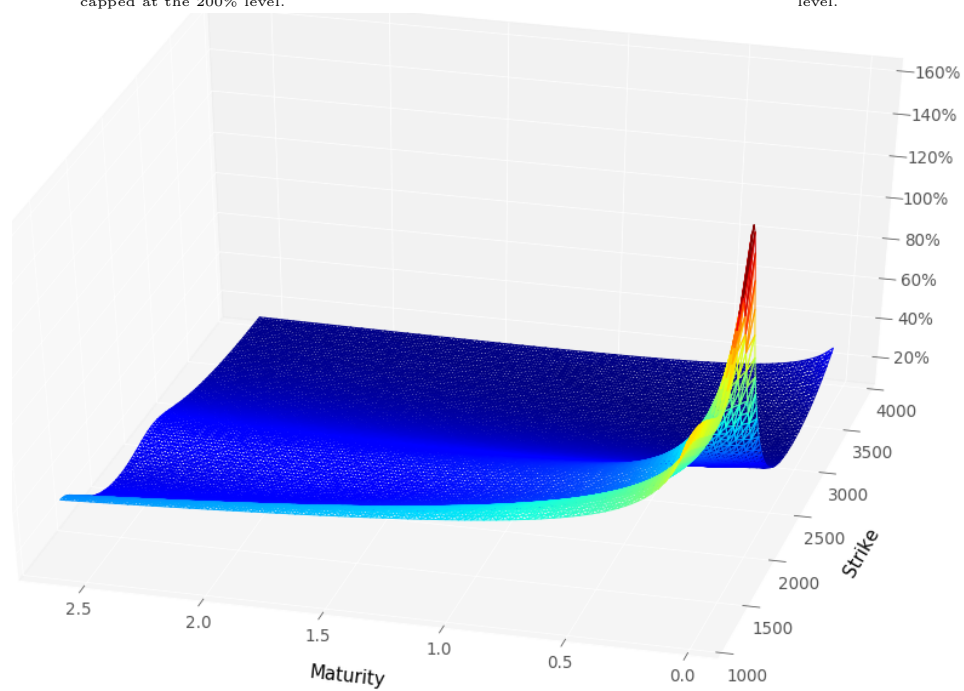
(A) The local volatility surface generated by SSVI with finite differences, capped at the 200% level.

(B) The MAP estimate of the GP local volatility surface, capped at the 200% level.

(C) The implied volatility based NN local volatility surface (with the local volatility penalization).

FIGURE 7.4: *The GP, SSVI, and NN local volatility estimate.*

7.4.4   In-sample and out-of-sample calibration errors

The error between the prices of the calibrated models and the market data are evaluated on both the training and the out-of-sample data set. The first two rows of Table 7.1 compare the in-sample and out-of-sample RMSEs of the prices and implied volatilities across the different approaches. The differences between the training and testing RMSEs are small, suggesting that all approaches are not over-fitting the training set. We observe that the GP exhibits the lowest price RMSEs.

7.4.5   Backtesting results

The first repricing backtest estimates the prices of the European options corresponding to the testing set, by Monte Carlo sampling in each calibrated local volatility model (same methodology as described in Chataigner, Crépey, and Dixon (2020, Section 7.2)). The second approach uses finite differences (FD) to price the options with the calibrated local volatility surfaces. The pricing PDEs with local volatility are discretized using a Crank-Nicolson scheme implemented on a 100×100 backtesting grid. The last two rows in Table 7.1 compare the resulting price backtest RMSEs across the different approaches. The NN fitted to implied volatilities exhibit significantly lower errors in the backtests, followed by NN based on prices, SSVI and GP. To quantify discretization error in these backtesting results (as opposed to the part of the error stemming from a wrong local volatility), we ran the same backtests in a Black-Scholes model with 20% volatility and the associated prices. The corresponding Monte Carlo and PDE backtesting IV (price) RMSEs are 2.90% (1.56) and 0.846% (4.10).

7.5   Conclusion

In this chapter, we develop a finite dimensional kriging approach for no-arbitrage interpolation of European vanilla option prices which jointly yields the full surface of local volatilities with uncertainty bands, even in the presence of arbitrage in the data. We demonstrate the performance relative to SSVI and deep learning of implied volatilities. The latter uses shape constraint regularization through the local volatility implied variance formula to penalize arbitrages. The GPs is found to produce a better fit to out-of-sample prices than SSVI, with less evidence of overfitting, although the performance is less competitive than neural networks trained to implied volatilities. Backtesting results show that kriging is competitive although it suggests that fitting to implied volatilities would be advantageous. However, in contrast to the SSVI and neural networks, the GPs provides uncertainty quantification, which promises much potential for model risk aspects of local volatility construction (cf. Cont, 2006, Example 4.3) and no-arbitrage price interpolation. The best of all worlds could be GPs trained to implied volatilities.

Conslusions and perpectives

Conclusions

In this thesis, we have addressed some aspects which contribute in quantitative financial risk management through two different parts.

In the first part, we have guided our thinking about default times modeling. As such, after having recalled some well known results about stochastic calculus and models of default times in Chapter 1, we have investigated the Gaussian example in conditional densities by showing that this example is unique in a large class of diffusion processes.

Very often, in credit risk modeling, the construction of the default time τ using the enlargement of filtration is based on the hypothesis that τ avoids the stopping times of the reference filtration. However, this framework does not allow to cover a wide range of situations, particularly in modeling defaultable claims under some economic shocks under which the default time could coincide with the shock times.

This allows us to expand our thinking, in Chapter 2, toward the model of Jiao and Li (2018) which is designed for modeling default time which could coincide with some predictable stopping times in the reference filtration. By this way, we proposed an extension of that model when the stopping times in the reference filtration are no more predictable. We have concluded that despite its attractiveness in modeling defaultable claims under some predictable shocks, the model of Jiao and Li (2018) is difficult to implement if the shocks are not predictable and does not always allow to capture the jumps of the zero-coupon bond prices.

We then proposed the generalized Cox model in Chapter 3 which extends the one of Lando (1998) in which the default time τ is the first time when an increasing process K adapted to a given filtration $\mathbb{F}$, absolutely continuous with respect to Lebesgue's measure hits a level, which is a random variable independent of $\mathbb{F}$. It follows that this random time avoids all $\mathbb{F}$-stopping times. We have relaxed the assumption that K is absolutely continuous by working in a more general case, where that process K being adapted, increasing and continuous on right with limits on left or continuous on left with limits on right. This led us to a random time which does not avoid the stopping times of the reference filtration. We were interested in the computations of characteristics of the default time such as the conditional survival martingale, the Azéma supermartingale, the compensator of the default time as well as its predictable reduction process and the multiplicative decomposition of the Azéma supermartingale.

We have also given many examples which illustrated our construction. The Generalized Cox model extends the one of Jiao and Li (2018) when K is $\mathbb{F}$-predictable.

In Chapter 4, we have investigated some applications of the generalized Cox model in credit risk. We were also interested in the effects of the jumps of the process K on the prices of the zero-coupon bonds through different configurations of K. The Generalized Cox model where K is a shot-noise allows to capture the jumps of the zero coupon bond when the stopping times of the reference filtration are not predictable.

In the second part of this thesis, we have presented the kriging techniques for solving some problems in quantitative finance such as option pricing and construction of volatility surfaces. Chapter 5 described some useful tools which have been used in Chapters 6 and 7. Particularly, we have recalled the constrained kriging in 1d using the finite dimensional approximation of GPs developed by Maatouk and Bay, 2014. A particular attention has been paid in the simulation of the Gaussian coefficients when some monotonicity constraints

are saturated and we have proposed a numerical solution to this problem which consists in reducing the number of basis function in the regions where constraints are saturated.

Chapter 6 was devoted to the construction of prices and implied volatility surfaces under no-arbitrage constraints using kriging. A comparison of the classical and constrained kriging for this task has been done. It turned out that the surfaces constructed using constrained kriging respect the no arbitrage conditions while the ones fitted by the classical kriging can lead to some arbitrages. We have also proposed a method for identifying the locations with the most likely arbitrages (which are the ones with largest noises) in the data by computing the joint MAP of the truncated Gaussian coefficient and the Gaussian noise vector. We demonstrated our construction using the Euro Stoxx 50 data of January 10, 2019. Our results show that kriging is a suitable tool for constructing option prices and quantifying uncertainty in the presence of noisy data, and for computing the associated local and implied volatility. Hence a promised tool for quantifying model uncertainty.

In Chapter 7, we have demonstrated the performance of kriging relative to various popular alternative interpolation techniques, we have benchmarked it against SSVI and deep learning techniques for local volatility surface construction. The fitting of kriging in the replication of bid ask data is an important result. Indeed, it allows to get a price surface which lies between the bid and ask prices.

Backtesting results show that kriging is competitive although it suggests that fitting to implied volatilities would be advantageous. However, in contrast to the SSVI and neural networks, the Kriging provides uncertainty quantification, which promises much potential for model risk aspects of local volatility construction and no-arbitrage price interpolation.

Perspectives

Here we present some new avenues for future researches.

Using Generalized Cox model for modeling catastrophe (CAT) bonds

The issuing of catastrophe (CAT) bonds is essential for insurance companies. These products allow them to hedge against the risks incurred following a natural disasters such as earthquakes, pandemics, etc. Recently, we are witnessing the occurrence of the Coronavirus disease COVID-19 pandemic that causes significant losses. A number of good examples have already been developed for catastrophe bonds. These include the model of Burnecki and Kukla (2003) who used a compound doubly stochastic Poisson loss process for aggregate losses where the catastrophe event times are jump times of standard Poisson Process (i.e., with a constant intensity). Their approach has been extended by Ma, Ma, and Xiao (2017) where they use a stochastic intensity rate of the Poisson process and provide an explicit intensity of the default time as well as a semi-analytical solution for evaluating zero-coupon CAT bonds. In the double stochastic Poisson loss model, default occurs when the aggregate claims process exceeds a specified level (called threshold value). Schmidt (2014) used shot-noise process for modeling the aggregated losses process.

Jarrow (2010) has been based on the reduced form model for pricing credit derivatives. He modeled the time of the catastrophe event underlying the Cat bond as a standard Cox time (i.e., the case where the increasing adapted process K is continuous). However, his model can be applied only in the case of one occurrence of a catastrophe event in the time interval of the Bond contract. This can be understood through the illustrative example on his introduction where an insurance issues CAT-Bonds with the hurricane in Florida as the underlying catastrophe event for covering eventual losses that could be included in the amount insured for home insurance written by that company at the first occurrence of the hurricane in the time interval of the contract. However after the first event, the CAT-Bonds come to an end.

Nevertheless, when one considers the case with possible successive catastrophe events in life time of a CAT-Bond contract, the model of Jarrow (2010) is not enable to cover this kind of framework, specially in the case of long maturity CAT-Bonds contracts. Indeed the default time in Jarrow (2010) represents only the

first time of the ones of these events, while others severe losses can be noted after the nth occurrence of an event.

To bridge this gap, we can adopt the generalized Cox framework where the sequence of catastrophe events can be modeled as an increasing sequence of stopping times in the reference filtration which are not avoided by the default time. This setting may offer some nice mathematical perspectives to the CAT-Bonds. However, how it could be implemented from the practical point of view?

Kriging and theory of optimal martingale transport

We have seen that kriging is a suitable tool for constructing option prices and quantifying uncertainty in the presence of noisy data by constructing some confidence bounds.

In addition, this latest example of kriging could offer new perspectives for studying risk model in quantitative finance. Recent works on the subject concern in particular the derivation of non-arbitrage bounds (and the identification of strategies of over- or under- replication to reach those boundaries) for option prices (path-dependent) knowing the marginal laws of the underlying at some maturities. This problem can for example be studied under the angle of theory of optimal martingale transport (see, e.g., Henry-Labordere and Touzi, 2013; Hobson and Klimmek, 2015; Beiglböck, Juillet, et al., 2016; Beiglböck, Henry-Labordère, and Penkner, 2013). However, in this approach, the non-arbitrage bounds can be obtained efficiently only from the knowledge of all marginal distributions.

In practice, the marginal law is only partially observed at some time horizons through the observation of traded European option prices.

As we illustrated in Chap 6, 7, constrained kriging allows to quantify uncertainty on marginals laws of a martingale measures. A more realistic uncertainty quantification of path-dependent option prices could be achieved by combining our approach and the optimal martingale transport approach. This is a perspective for future research.

Chapter 8

Finite Element Approximation

The elements of the matrix F are given by

$$F_{\imath_1,\imath_2} = \int_{T_{i-1}}^{T_{i+1}} \int_{\kappa_{j-1}}^{\kappa_{j+1}} \partial_T \phi_{\imath_2}(T,\kappa)\phi_{\imath_1}(T,\kappa)d\kappa dT, \ \forall \imath_1, \imath_2 \in \{0,\dots,N-1\}, \tag{8.0.1}$$

which simplifies to the following 6-diagonal square matrix, where the central tri-diagonal band is zero:

$$F_{\imath_1,\imath_2} = \begin{cases} 0, & i_1 = i_2, |j_1 - j_2| \leq 1, \\ \pm h_k/3, & i_1 = i_2 \pm 1, j_1 = j_2, \\ \pm h_k/12, & i_1 = i_2 \pm 1, |j_1 - j_2| = 1. \end{cases} \tag{8.0.2}$$

The elements of the stiffness matrix H are given by

$$H_{\imath_1,\imath_2} = -\sum_{\imath_2} \xi_{\imath_2} \int_{T_{i-1}}^{T_{i+1}} \int_{\kappa_{j-1}}^{\kappa_{j+1}} \partial_\kappa \phi_{\imath_2}(T,\kappa)\partial_\kappa \left(\mathrm{dup}^h(T,\kappa)\phi_{\imath_1}(T,\kappa) \right) d\kappa dT \tag{8.0.3}$$

which evaluates to the 9-diagonal square matrix:

$$H_{\imath_1,\imath_2} = \begin{cases} -\frac{1}{3}\frac{h_T}{h_\kappa}(\mathrm{dup}_{i_1,j_1} + \mathrm{dup}_{i_1-1,j_1} + \mathrm{dup}_{i_1,j_1-1} + \mathrm{dup}_{i_1-1,j_1-1}), & \imath_2 = \imath_1, \\ \frac{1}{3}\frac{h_T}{h_\kappa}(\mathrm{dup}_{i_1,j_1} + \mathrm{dup}_{i_1-1,j_1}), & i_2 = i_1, j_2 = j_1 + 1, \\ \frac{1}{3}\frac{h_T}{h_\kappa}(\mathrm{dup}_{i_1,j_1-1} + \mathrm{dup}_{i_1-1,j_1-1}), & i_2 = i_1, j_2 = j_1 - 1, \\ -\frac{1}{6}\frac{h_T}{h_\kappa}(\mathrm{dup}_{i_1,j_1} + \mathrm{dup}_{i_1,j_1-1}), & i_2 = i_1 - 1, j_2 = j_1, \\ -\frac{1}{6}\frac{h_T}{h_\kappa}(\mathrm{dup}_{i_1-1,j_1} + \mathrm{dup}_{i_1-1,j_1-1}), & i_2 = i_1 + 1, j_2 = j_1, \\ \frac{1}{6}\frac{h_T}{h_\kappa}\mathrm{dup}_{i_1,j_1}, & i_2 = i_1 - 1, j_1 = j_2 + 1 \\ \frac{1}{6}\frac{h_T}{h_\kappa}\mathrm{dup}_{i_1,j_1-1}, & i_2 = i_1 - 1, j_1 = j_2 - 1, \\ \frac{1}{6}\frac{h_T}{h_\kappa}\mathrm{dup}_{i_1-1,j_1}, & i_2 = i_1 + 1, j_1 = j_2 + 1 \\ \frac{1}{6}\frac{h_T}{h_\kappa}\mathrm{dup}_{i_1-1,j_1-1}, & i_2 = i_1 + 1, j_1 = j_2 - 1, \end{cases}$$

If we multiply each side of the equation $H(\mathbf{dup})\boldsymbol{\xi} = F\boldsymbol{\xi}$ by $\frac{3}{2h_T h_k}$ then the second line of the right hand side has terms similar to a first order central finite difference stencil $\frac{\xi_{i+1,j}-\xi_{i-1,j}}{2h_T}$. Additionally the stiffness matrix has terms which correspond to a second order finite difference stencil with local averaging of dup. To see this, suppose that dup were constant over each quadrant then the first three lines of the specification of H above correspond to the finite difference stencil $\frac{\xi_{i,j+1}-2\xi_{i,j}+\xi_{i,j-1}}{h_\kappa^2}$.

The four-diagonal rectangular matrix A is given by

$$A_{\imath_1,\imath_2} = \begin{cases} -\frac{h_T}{h_\kappa}(\frac{1}{3}\xi_{i_1,j_1} - \frac{1}{3}\xi_{i_1+1,j_1} + \frac{1}{6}\xi_{i_1,j_1+1} - \frac{1}{6}\xi_{i_1+1,j_1+1}), & \imath_1 = \imath_2, \\ -\frac{h_T}{h_\kappa}(\frac{1}{3}\xi_{i_1,j_1} - \frac{1}{3}\xi_{i_1,j_1-1} + \frac{1}{6}\xi_{i_1+1,j_1} - \frac{1}{6}\xi_{i_1+1,j_1-1}), & \imath_1 = \imath_2 + N_T, \\ -\frac{h_T}{h_\kappa}(\frac{1}{3}\xi_{i_1,j_1} - \frac{1}{3}\xi_{i_1,j_1+1} + \frac{1}{6}\xi_{i_1-1,j_1} - \frac{1}{6}\xi_{i_1-1,j_1+1}), & \imath_1 = \imath_2 + 1, \\ -\frac{h_T}{h_\kappa}(\frac{1}{3}\xi_{i_1,j_1} - \frac{1}{3}\xi_{i_1,j_1-1} + \frac{1}{6}\xi_{i_1-1,j_1} - \frac{1}{6}\xi_{i_1-1,j_1-1}), & \imath_1 = \imath_2 + N_T + 1. \end{cases} \tag{8.0.4}$$

Bibliography

Ackerer, Damien, Natasa Tagasovska, and Thibault Vatter (2019). "Deep smoothing of the implied volatility surface". In: *Available at SSRN 3402942*.

Agrell, Christian (2019). "Gaussian processes with linear operator inequality constraints". In: *arXiv preprint arXiv:1901.03134*.

Aksamit, Anna, Tahir Choulli, Jun Deng, and Monique Jeanblanc (2014). "Arbitrages in a progressive enlargement setting". In: *Arbitrage, Credit and Informational Risks*. World Scientific, pp. 53–86.

Aksamit, Anna, Tahir Choulli, and Monique Jeanblanc (2016). "Thin times and random times' decomposition". In: *arXiv preprint arXiv:1605.03905*.

Aksamit, Anna and Monique Jeanblanc (2017). *Enlargement of filtration with finance in view*. Springer.

Aksamit, Anna and Libo Li (2016). "Projections, pseudo-stopping times and the immersion property". In: *Séminaire de Probabilités XLVIII*. Springer, pp. 459–467.

Altmann, Timo, Thorsten Schmidt, and Winfried Stute (2008). "A shot noise model for financial assets". In: *International Journal of Theoretical and Applied Finance* 11.01, pp. 87–106.

Asgharian, Hossein, Wolfgang Hess, and Lu Liu (2013). "A spatial analysis of international stock market linkages". In: *Journal of Banking & Finance* 37.12, pp. 4738–4754.

Aubin-Frankowski, Pierre-Cyril and Zoltan Szabo (2020). "Hard Shape-Constrained Kernel Machines". In: arXiv:2005.12636.

Azéma, Jacques (1972). "Quelques applications de la théorie générale des processus. I". In: *Inventiones mathematicae* 18.3, pp. 293–336.

Bachoc, François (2013). "Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification". In: *Computational Statistics & Data Analysis* 66, pp. 55–69.

Bachoc, François, Agnès Lagnoux, and Andrés F López-Lopera (2018). "Maximum likelihood estimation for Gaussian processes under inequality constraints". In: *arXiv preprint arXiv:1804.03378*.

Bachoc, François, Agnes Lagnoux, Andrés F López-Lopera, et al. (2019). "Maximum likelihood estimation for Gaussian processes under inequality constraints". In: *Electronic Journal of Statistics* 13.2, pp. 2921–2969.

Barlow, Martin T (1978). "Study of a filtration expanded to include an honest time". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 44.4, pp. 307–323.

Beiglböck, Mathias, Pierre Henry-Labordère, and Friedrich Penkner (2013). "Model-independent bounds for option prices—a mass transport approach". In: *Finance and Stochastics* 17.3, pp. 477–501.

Beiglböck, Mathias, Nicolas Juillet, et al. (2016). "On a problem of optimal transport under marginal martingale constraints". In: *The Annals of Probability* 44.1, pp. 42–106.

Bélanger, Alain, Steven Shreve, and Dennis Wong (2001). "A unified model for credit derivatives". In: *Mathematical Finance*.

Bélanger, Alain, Steven E. Shreve, and Dennis Wong (2004). "A general framework for pricing credit risk". In: *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 14.3, pp. 317–350.

Bielecki, Tomasz R., Monique Jeanblanc, and Marek Rutkowski (2009). "Credit risk modeling". In: *Center of the study of finance and Insurance, Osaka University, Japan*.

Bielecki, Tomasz R. and Marek Rutkowski (2002). *Credit risk: modeling, valuation and hedging*. Springer.

Black, Fischer and Myron Scholes (1973). "The pricing of options and corporate liabilities". In: *Journal of political economy* 81.3, pp. 637–654.

Botev, ZI (2017). "The normal law under linear restrictions: simulation and estimation via minimax tilting Series B Statistical methodology". In:

Brémaud, Pierre and Marc Yor (1978). "Changes of filtrations and of probability measures". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 45.4, pp. 269–295.

Brent, Richard P. (1971). "An algorithm with guaranteed convergence for finding a zero of a function". In: *The Computer Journal* 14.4, pp. 422–425.

Burnecki, Krzysztof and Grzegorz Kukla (2003). "Pricing of zero-coupon and coupon CAT bonds". In: *Applicationes Mathematicae* 30, pp. 315–324.

Campbell, Norman (1909a). "Discontinuities in light emission". In: *Proc. Cambridge Phil. Soc.* Vol. 15. 310-328, p. 3.

— (1909b). "The study of discontinuous phenomena". In: *Proceedings of the Cambridge Philosophical Society*. Vol. 15, pp. 117–136.

Chataigner, Marc, Stéphane Crépey, and Matthew Dixon (2020). "Deep Local Volatility". In: *Risks* 8.3, p. 82. URL: http://dx.doi.org/10.3390/risks80300.

Chaudhury, Mo (2014). "Option bid-ask spread and liquidity". In:

Coculescu, Delia (2017). "From the decompositions of a stopping time to risk premium decompositions". In: *ESAIM: Proceedings and Surveys* 60, pp. 1–20.

Cont, Rama (2006). "Model uncertainty and its impact on the pricing of derivative instruments". In: *Mathematical finance* 16.3, pp. 519–547.

Cont, Rama and Peter Tankov (2004). "Financial Modelling with jump processes". In: *Chapman & Hall/CRC*.

Cousin, Areski, Hassan Maatouk, and Didier Rullière (2016). "Kriging of financial term-structures". In: *European J. Oper. Res.* 255.2, pp. 631–648. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2016.05.057. URL: https://doi.org/10.1016/j.ejor.2016.05.057.

Crepey, Stephane (2003). "Calibration of the local volatility in a generalized black–scholes model using tikhonov regularization". In: *SIAM Journal on Mathematical Analysis* 34.5, pp. 1183–1206.

Cressie, Noel (1990). "The origins of kriging". In: *Math. Geol.* 22.3, pp. 239–252. ISSN: 0882-8121. DOI: 10.1007/BF00889887. URL: https://doi.org/10.1007/BF00889887.

Da Veiga, Sébastien and Amandine Marrel (2012). "Gaussian process modeling with inequality constraints". In: *Annales de la Faculté des sciences de Toulouse: Mathématiques*. Vol. 21. 3, pp. 529–555.

Dassios, Angelos and Ji-Wook Jang (2003). "Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity". In: *Finance and Stochastics* 7.1, pp. 73–95.

De Spiegeleer, Jan, Dilip B Madan, Sofie Reyners, and Wim Schoutens (2018). "Machine learning for quantitative finance: fast derivative pricing, hedging and fitting". In: *Quantitative Finance* 18.10, pp. 1635–1643.

Dekker, TJ (1969). *Finding a zero by means of successive linear interpolation, Constructive Aspects of the Fundamental Theorem of Algebra (B. Dejon and P. Henrici, eds.)*

Dixon, Matthew F and Stéphane Crépey (2018). "Multivariate Gaussian Process Regression for Derivative Portfolio Modeling: Application to CVA". In:

Duffie, Darrell, Damir Filipović, and Walter Schachermayer (2003). "Affine processes and applications in finance". In: *The Annals of Applied Probability* 13.3, pp. 984–1053.

Duffie, Darrell, Mark Schroder, and Costis Skiadas (1996). "Recursive valuation of defaultable securities and the timing of resolution of uncertainty". In: *The Annals of Applied Probability* 6.4, pp. 1075–1090.

Dugas, Charles, Yoshua Bengio, François Bélisle, Claude Nadeau, and René Garcia (2009). "Incorporating functional knowledge in neural networks". In: *Journal of Machine Learning Research* 10.Jun, pp. 1239–1262.

Dupire, Bruno (1994). "Pricing with a smile". In: *Risk* 7.1, pp. 18–20.

El Karoui, Nicole, Monique Jeanblanc, and Ying Jiao (2010). "What happens after a default: the conditional density approach". In: *Stochastic processes and their applications* 120.7, pp. 1011–1032.

El Karoui, Nicole, Monique Jeanblanc, Ying Jiao, and Behnaz Zargari (2014). "Conditional default probability and density". In: *Inspired by Finance*. Springer, pp. 201–219.

Elliott, Robert J., Monique Jeanblanc, and Marc Yor (2000). "On models of default risk". In: *Mathematical Finance* 10.2, pp. 179–195.

Fengler, Matthias R (2009). "Arbitrage-free smoothing of the implied volatility surface". In: *Quantitative Finance* 9.4, pp. 417–428.

Fontana, Claudio and Thorsten Schmidt (2018). "General dynamic term structures under default risk". In: *Stochastic Processes and their Applications* 128.10, pp. 3353–3386.

Gapeev, Pavel V., Monique Jeanblanc, Libo Li, and Marek Rutkowski (2010). "Constructing random times with given survival processes and applications to valuation of credit derivatives". In: *Contemporary quantitative finance*. Springer, pp. 255–280.

Gatheral, Jim (2004). "A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives". In: *Presentation at Global Derivatives & Risk Management, Madrid*, p. 0.

— (2011). *The volatility surface: a practitioner's guide*. Vol. 357. John Wiley & Sons.

Gatheral, Jim and Antoine Jacquier (2014). "Arbitrage-free SVI volatility surfaces". In: *Quantitative Finance* 14.1, pp. 59–71.

Gehmlich, Frank and Thorsten Schmidt (2018). "Dynamic defaultable term structure modeling beyond the intensity paradigm". In: *Mathematical Finance* 28.1, pp. 211–239.

Giesecke, K. and S. Zhu (2013). "Affine LIBOR models with multiple curves: theory, examples and calibration". In: *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 23.4, pp. 742–762.

Giesecke, Kay and Lisa R. Goldberg (2008). "The market price of credit risk: the impact of asymmetric information". In: *Available at SSRN 450120*.

Golchi, Shirin, Derek R Bingham, Hugh Chipman, and David A Campbell (2015). "Monotone emulation of computer experiments". In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1, pp. 370–392.

Gonzalvez, Joan, Edmond Lezmi, Thierry Roncalli, and Jiali Xu (2019). "Financial applications of Gaussian processes and Bayesian optimization". In: *arXiv preprint arXiv:1903.04841*.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. MIT Press.

Grbac, Zorana, Antonis Papapantoleon, John Schoenmakers, and David Skovmand (2015). "Affine LIBOR models with multiple curves: theory, examples and calibration". In: *SIAM Journal on Financial Mathematics* 6.1, pp. 984–1025.

Gueye, Djibril, Monique Jeanblanc, and Libo Li (2019). "Models of default times and cox model revisited". In: *Freiburg FRIAS: Finance and Insurance*.

He, Sheng-Wu, Jia-Gang Wang, and Jia-An Yan (2018). *Semimartingale theory and stochastic calculus*. CRC Press.

Henry-Labordere, Pierre and Nizar Touzi (2013). "An explicit martingale version of Brenier's theorem". In:

Hentschel, Ludger (2003). "Errors in implied volatility estimation". In: *Journal of Financial and Quantitative analysis* 38.4, pp. 779–810.

Herbertsson, Alexander, Jiwook Jang, and Thorsten Schmidt (2011). "Pricing basket default swaps in a tractable shot noise model". In: *Statistics & probability letters* 81.8, pp. 1196–1207.

Hillairet, Caroline and Ying Jiao (2012). "Credit risk with asymmetric information on the default threshold". In: *Stochastics An International Journal of Probability and Stochastic Processes* 84.2-3, pp. 183–198.

Hobson, David and Martin Klimmek (2015). "Robust price bounds for the forward starting straddle". In: *Finance and Stochastics* 19.1, pp. 189–214.

Homescu, Cristian (2011). "Implied volatility surface: Construction methodologies and characteristics". In: *Available at SSRN 1882567*.

Ioffe, Ioulia D and Eliezer Z Prisman (2013). "Arbitrage violations and implied valuations: the option market". In: *The European Journal of Finance* 19.4, pp. 298–317.

Itkin, Andrey (2019). "Deep learning calibration of option pricing models: some pitfalls and solutions". In: *arXiv preprint arXiv:1906.03507*.

Itô, Kiyosi (1978). "Extension of stochastic integrals". In: *Proc. Internat. Symp. Stoch. Diff. Equations (Res. Inst. Mathemat. Sci., Kyoto Univ., Kyoto, 1976)*. Wiley.

Jacod, Jean (1985). "Grossissement initial, hypothèse H' et théorème de Girsanov". In: *Grossissements de filtrations: exemples et applications*. Springer, pp. 15–35.

Jarrow, Robert A (2010). "A simple robust model for CAT bond valuation". In: *Finance Research Letters* 7.2, pp. 72–79.

Jeanblanc, Monique and Libo Li (2020). "Characteristics and constructions of default times Characteristics and constructions of default times". In: *Hal 02183097, to appear in SIAM Journal on Financial Mathematics*.

Jeanblanc, Monique and Marek Rutkowski (2000). "Modelling of default risk: an overview". In: *Mathematical finance: theory and practice*, pp. 171–269.

Jeanblanc, Monique and Shiqi Song (2011a). "An explicit model of default time with given survival probability". In: *Stochastic Processes and their Applications* 121.8, pp. 1678–1704.

— (2011b). "Random times with given survival probability and their F-martingale decomposition formula". In: *Stochastic Processes and their Applications* 121.6, pp. 1389–1410.

Jeanblanc, Monique, Marc Yor, and Marc Chesney (2009). *Mathematical methods for financial markets*. Springer Science & Business Media.

Jeulin, Thierry and Marc Yor (1978). "Grossissement d'une filtration et semi-martingales: formules explicites". In: *Séminaire de Probabilités XII*. Springer, pp. 78–97.

Jiao, Ying and Shanqiu Li (2015). "Generalized density approach in progressive enlargement of filtrations". In: *Electronic Journal of Probability* 20.

— (2018). "Modeling sovereign risks: From a hybrid model to the generalized density approach". In: *Mathematical Finance* 28.1, pp. 240–267.

Jones, Donald R, Matthias Schonlau, and William J Welch (1998). "Efficient global optimization of expensive black-box functions". In: *Journal of Global optimization* 13.4, pp. 455–492.

Karatzas, Ioannis (1997). *Lectures on the Mathematics of Finance*. Vol. 8. American Mathematical Soc.

Kchia, Y. and P. Protter (2014). "Progressive Filtration Expansions via a Process, with Applications to Insider Trading". In: *Int. J. Theor. Appl. Finan.* 18, p. 1550027.

Keller-Ressel, Martin, Antonis Papapantoleon, and Josef Teichmann (2013). "The affine LIBOR models". In: *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 23.4, pp. 627–658.

Krige, Danie G (1951). "A statistical approach to some mine valuation and allied problems on the Witwatersrand: By DG Krige". PhD thesis. University of the Witwatersrand.

Krige, Daniel G. and Eduardo J. Magri (1982). "Geostatistical case studies of the advantages of lognormal-de Wijsian Kriging with mean for a base metal mine and a gold mine". In: *J. Internat. Assoc. Math. Geol.* 14.6, pp. 547–555. ISSN: 0020-5958. DOI: 10.1007/BF01033878. URL: https://doi.org/10.1007/BF01033878.

Kunita, H. (1981). "Some extensions of Itô's formula". In: *Séminaire de Probabilités XV*. Vol. 850. Lecture Notes in Mathematics. Springer-Verlag, pp. 118–141.

Lando, David (1995). "Three essays on contingent claims pricing". PhD thesis.

— (1998). "On Cox processes and credit risky securities". In: *Review of Derivatives research* 2.2-3, pp. 99–120.

Last, G. and A. Brandt (1995a). "Marked Point Processes on the Real Line. The Dynamic Approach". In: *Springer, Berlin.*

Last, Günter and Andreas Brandt (1995b). *Marked Point Processes on the real line: the dynamical approach.* Springer Science & Business Media.

Laurini, Márcio Poletti (2011). "Imposing no-arbitrage conditions in implied volatilities using constrained smoothing splines". In: *Applied Stochastic Models in Business and Industry* 27.6, pp. 649–659.

Li, Libo and Marek Rutkowski (2014). "Progressive enlargements of filtrations with pseudo-honest times". In: *The Annals of Applied Probability* 24.4, pp. 1509–1553.

Linetsky, Vadim (2006). "Pricing equity derivatives subject to bankruptcy". In: *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 16.2, pp. 255–282.

Liu, Shuaiqiang, Cornelis W Oosterlee, and Sander M Bohte (2019). "Pricing options and computing implied volatilities using neural networks". In: *Risks* 7.1, p. 16.

Löhner, Rainald and Joseph D Baum (1992). "Adaptive h-refinement on 3D unstructured grids for transient problems". In: *International Journal for Numerical Methods in Fluids* 14.12, pp. 1407–1419.

López-Lopera, Andrés F, François Bachoc, Nicolas Durrande, and Olivier Roustant (2018). "Finite-dimensional Gaussian approximation with linear inequality constraints". In: *SIAM/ASA Journal on Uncertainty Quantification* 6.3, pp. 1224–1255.

Ludkovski, Michael (2018). "Kriging metamodels and experimental design for Bermudan option pricing". In:

Ludkovski, Mike, Jimmy Risk, and Howard Zail (2018). "Gaussian process models for mortality rates and improvement factors". In: *ASTIN Bulletin: The Journal of the IAA* 48.3, pp. 1307–1347.

Ma, Zonggang, Chaoqun Ma, and Shisong Xiao (2017). "Pricing zero-coupon catastrophe bonds using EVT with doubly stochastic Poisson arrivals". In: *Discrete Dynamics in Nature and Society* 2017.

Maatouk, Hassan (2017). "Finite-dimensional approximation of Gaussian processes with inequality constraints and errors measurements". In:

Maatouk, Hassan and Xavier Bay (2014). *Gaussian Process Emulators for Computer Experiments with Inequality Constraints. hal-01096751.*

— (2016). "A new rejection sampling method for truncated multivariate Gaussian random variables restricted to convex sets". In: *Monte carlo and quasi-monte carlo methods.* Springer, pp. 521–530.

— (2017). "Gaussian process emulators for computer experiments with inequality constraints". In: *Mathematical Geosciences* 49.5, pp. 557–582.

Maatouk, Hassan, Olivier Roustant, and Yann Richet (2015). "Cross-validation estimations of hyper-parameters of Gaussian processes with inequality constraints". In: *Procedia Environmental Sciences* 27, pp. 38–44.

Maghsoodi, Yoosef (1996). "Solution of the extended CIR term structure and bond option valuation". In: *Mathematical finance* 6.1, pp. 89–109.

Matheron, Georges (1963). "Principles of geostatistics". In: *Economic geology* 58.8, pp. 1246–1266.

Merton, Robert C. (1974). "On the pricing of corporate debt: The risk structure of interest rates". In: *The Journal of finance* 29.2, pp. 449–470.

Metivier, Michel (1982). "Pathwise differentiability with respect to a parameter of solutions of stochastic differential equations". In: *Séminaire de Probabilités XVI 1980/81.* Springer, pp. 490–502.

Mikosch, Thomas (2009). *Non-life insurance mathematics: an introduction with the Poisson process.* Springer Science & Business Media.

Mitchell, William F (1991). "Adaptive refinement for arbitrary finite-element spaces with hierarchical bases". In: *Journal of computational and applied mathematics* 36.1, pp. 65–78.

Musiela, Marek and Marek Rutkowski (1998). *Martingale methods in financial modelling: theory and applications.* Springer.

Nikeghbali, Ashkan (2006). "An essay on the general theory of stochastic processes". In: *Probability Surveys* 3, pp. 345–412.

Øksendal, Bernt and Tusheng Zhang (2007). "The Itô-Ventzell formula and forward stochastic differential equations driven by Poisson random measures". In: *Osaka Journal of Mathematics* 44.1, pp. 207–230.

Øksendal, Bernt Karsten (2003). *Stochastic Differential Equations: An introduction with applications.* Springer-Verlag.

Orosi, Greg (2015). "Arbitrage-free call option surface construction using regression splines". In: *Applied Stochastic Models in Business and Industry* 31.4, pp. 515–527.

Pakman, Ari and Liam Paninski (2014). "Exact hamiltonian monte carlo for truncated multivariate gaussians". In: *Journal of Computational and Graphical Statistics* 23.2, pp. 518–542.

Protter, Philip E. (2005). "Stochastic differential equations". In: *Stochastic integration and differential equations.* Springer, pp. 249–361.

Redeker, Imke and Ralf Wunderlich (2019). "Credit risk with asymmetric information and a switching default threshold". In: *arXiv preprint arXiv:1910.14413.*

Riihimäki, Jaakko and Aki Vehtari (2010). "Gaussian processes with monotonicity information". In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 645–652.

Robert, Christian (2014). *Machine learning, a probabilistic perspective.*

Roper, Michael (2010). "Arbitrage free implied volatility surfaces". In: *preprint.*

Roustant, Olivier, David Ginsbourger, and Yves Deville (2012). "DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization". In:

Scherer, Matthias, Ludwig Schmid, and Thorsten Schmidt (2012). "Shot-noise driven multivariate default models". In: *European Actuarial Journal* 2.2, pp. 161–186.

Schmidt, Thorsten (2014). "Catastrophe insurance modeled by shot-noise processes". In: *Risks* 2.1, pp. 3–24.

Schönbucher, Philipp J (2003). *Credit derivatives pricing models: models, pricing and implementation.* John Wiley & Sons.

Schottky, Walter (1918). "Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern". In: *Annalen der physik* 362.23, pp. 541–567.

Shreve, Steven E (2004). *Stochastic calculus for finance II: Continuous-time models.* Vol. 11. Springer Science & Business Media.

Sousa, J Beleza, Manuel L Esquível, and Raquel M Gaspar (2012). "Machine learning Vasicek model calibration with Gaussian processes". In: *Communications in Statistics-Simulation and Computation* 41.6, pp. 776–786.

Taylor, J and Y Benjamini (2016). "RestrictedMVN: multivariate normal restricted by affine constraints". In: *R package version* 1.

Tegnér, Martin and Stephen Roberts (2019). "A probabilistic approach to nonparametric local volatility". In: *arXiv preprint arXiv:1901.06021.*

Thiébaux, H Jean and Francis W Zwiers (1984). "The interpretation and estimation of effective sample size". In: *Journal of Climate and Applied Meteorology* 23.5, pp. 800–811.

Ventzel, A.D. (1965). "On equations of the theory of conditional Markov processes". In: *Theory of Probability and its Applications* 10, pp. 357–361.

Von Neumann, John (1951). "13. various techniques used in connection with random digits". In: *Appl. Math Ser* 12.36-38, p. 5.

Wang, Xiaojing and James O Berger (2016). "Estimating shape constrained functions using Gaussian processes". In: *SIAM/ASA Journal on Uncertainty Quantification* 4.1, pp. 1–25.

Wang, Y, H Yin, and L Qi (2004). "No-arbitrage interpolation of the option price function and its reformulation". In: *Journal of Optimization Theory and Applications* 120.3, pp. 627–649.

Williams, Christopher K and Carl Edward Rasmussen (2006). "Gaussian processes for machine learning". In: *the MIT Press* 2.3, p. 4.

Yang, Yongxin, Yu Zheng, and Timothy M Hospedales (2017). "Gated neural networks for option pricing: Rationality by design". In: *Thirty-First AAAI Conference on Artificial Intelligence.*

Yor, Marc (1978). "Grossissement d'une filtration et semi-martingales: théorèmes généraux". In: *Séminaire de Probabilités XII*. Springer, pp. 61–69.

Zargari, Behnaz (2011). "Le risque de crédit et les produits dérivés de crédit: modélisation mathématique et numérique". PhD thesis. Evry-Val d'Essonne.

Zhang, Hao and Yong Wang (2010). "Kriging and cross-validation for massive spatial data". In: *Environmetrics: The official journal of the International Environmetrics Society* 21.3-4, pp. 290–304.

Cette thèse traite différentes questions liées à la gestion quantitative des risques financiers. Nous nous intéressons, dans une première partie, aux modèles de temps de défaut en risque de crédit dans le cadre de la théorie de grossissement de filtrations. Nous proposons des modèles où le temps de défaut peut coïncider avec des instants de chocs économiques. Nous mettons l'accent, dans un premier temps, sur le modèle de Jiao et Li (2018) en risque souverain où le temps de dèfaut coïncide avec des temps de chocs prévisibles. Nous étendons ce modèle dans le cas où les chocs ne sont pas prévisibles en étudiant les caractéristiques du temps de défaut. Dans un second temps, nous présentons le modèle de Cox généralisé qui est une extention de celui de Lando (voir Lando, 1998). Nous proposons une large gamme d'exemples pour ullistrer notre construction. La seconde partie porte sur la construction de surfaces de volatilités des actifs financiers sous la condition d'absence d'opportunité d'arbitrage (AOA) en utilisant les méthodologies de krigeage (où la regression par processus Gaussien). Ces surfaces permettent par exemple d'estimer à partir du prix d'options liquides, la valeur des produits financiers dont les caractéristiques sont non-standard et dont le prix n'est pas observé sur le marché. La construction de telles surfaces est une étape importante dans certains processus de gestion des risques. Elle permet également de tarifer des actifs non-liquides. Notre approche consiste á mettre en œuvre l'apprentissage du krigeage sur les prix d'options européennes en respectant les conditions de non-arbitrage. Ces conditions sont caractérisées par des contraintes de forme sur les prix, à savoir la monotonicité dans la direction des maturités et la convexité dans la direction des strikes. Etant donné que ces contraintes correspondent à un nombre fini d'inégalités linéaires, nous adoptons une technique de krigeage sous contraintes d'inégalités linéaires. Nous utilisons, pour cela, la méthode d'eveloppée par Maatouk et Bay (2016) qui est basée sur l'approximation fini-dimensionnelle du processus Gaussien. L'algorithme de Monte Carlo Hamiltonien de Pakman et Paninski (2014) sera utilisé pour simuler les coefficients Gaussiens. Nous proposons une méthode de calcul du Maximum a Posteriori (MAP) du processus Gaussien. Nous comparons notre méthode avec celles des réseaux de neuronne contraints et des SSVI.