Université de Montréal

Modélisation du carnet d'ordres limites et prévision de séries temporelles

par

Clarence Simard

Département de mathématiques et de statistique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en mathématiques

15 décembre 2014

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

Modélisation du carnet d'ordres limites et prévision de séries temporelles

présentée par

Clarence Simard

a été évaluée par un jury composé des personnes suivantes :

Benjamin Avanzi
(président-rapporteur)

Bruno Rémillard
(directeur de recherche)

Geneviève Gauthier
(membre du jury)

Matt Davison
(examinateur externe)

Thèse acceptée le
27 octobre 2014

SOMMAIRE

Le contenu de cette thèse est divisé de la façon suivante. Après un premier chapitre d'introduction, le Chapitre 2 est consacré à introduire aussi simplement que possible certaines des théories qui seront utilisées dans les deux premiers articles. Dans un premier temps, nous discuterons des points importants pour la construction de l'intégrale stochastique par rapport aux semimartingales avec paramètre spatial. Ensuite, nous décrirons les principaux résultats de la théorie de l'évaluation en monde neutre au risque et, finalement, nous donnerons une brève description d'une méthode d'optimisation connue sous le nom de dualité.

Les Chapitres 3 et 4 traitent de la modélisation de l'*illiquidité* et font l'objet de deux articles. Le premier propose un modèle en temps continu pour la structure et le comportement du carnet d'ordres limites. Le comportement du portefeuille d'un investisseur utilisant des ordres de marché est déduit et des conditions permettant d'éliminer les possibilités d'arbitrages sont données. Grâce à la formule d'Itô généralisée il est aussi possible d'écrire la valeur du portefeuille comme une équation différentielle stochastique. Un exemple complet de modèle de marché est présenté de même qu'une méthode de calibrage.

Dans le deuxième article, écrit en collaboration avec Bruno Rémillard, nous proposons un modèle similaire mais cette fois-ci en temps discret. La question de tarification des produits dérivés est étudiée et des solutions pour le prix des options européennes de vente et d'achat sont données sous forme explicite. Des conditions spécifiques à ce modèle qui permettent d'éliminer l'arbitrage sont aussi données. Grâce à la méthode duale, nous montrons qu'il est aussi possible d'écrire le prix des options européennes comme un problème d'optimisation d'une espérance sur en ensemble de mesures de probabilité.

Le Chapitre 5 contient le troisième article de la thèse et porte sur un sujet différent. Dans cet article, aussi écrit en collaboration avec Bruno Rémillard, nous proposons une méthode de prévision des séries temporelles basée sur les copules multivariées. Afin de mieux comprendre le gain en performance que donne cette méthode, nous étudions à l'aide d'expériences numériques l'effet de la force et la structure de dépendance sur les prévisions. Puisque les copules permettent d'isoler la structure de dépendance et les distributions marginales, nous étudions l'impact de différentes distributions marginales sur la performance des prévisions. Finalement, nous étudions aussi l'effet des erreurs d'estimation sur la performance des prévisions. Dans tous les cas, nous comparons la performance des prévisions en utilisant des prévisions provenant d'une série bivariée et d'une série univariée, ce qui permet d'illustrer l'avantage de cette méthode. Dans un intérêt plus pratique, nous présentons une application complète sur des données financières.

SUMMARY

This thesis is structured as follows. After a first chapter of introduction, Chapter 2 exposes as simply as possible different notions that are going to be used in the two first papers. First, we discuss the main steps required to build stochastic integrals for semimartingales with space parameters. Secondly, we describe the main results of risk neutral evaluation theory and, finally, we give a short description of an optimization method known as duality.

Chapters 3 and 4 consider the problem of modelling illiquidity, which is covered by two papers. The first one proposes a continuous time model for the structure and the dynamic of the limit order book. The dynamic of a portfolio for an investor using market orders is deduced and conditions to rule out arbitrage are given. With the help of Itô's generalized formula, it is also possible to write the value of the portfolio as a stochastic differential equation. A complete example of market model along with a calibration method is also given.

In the second paper, written in collaboration with Bruno Rémillard, we propose a similar model with discrete time trading. We study the problem of derivatives pricing and give explicit formulas for European option prices. Specific conditions to rule out arbitrage are also provided. Using the dual optimization method, we show that the price of European options can be written as the optimization of an expectation over a set of probability measures.

Chapter 5 contained the third paper and studies a different topic. In this paper, also written with Bruno Rémillard, we propose a forecasting method for time series based on multivariate copulas. To provide a better understanding of the proposed method, with the help of numerical experiments, we study the effect of the strength and the structure of the different dependencies on predictions performance. Since copulas allow to isolate the dependence structure and marginal distributions, we study the impact of different marginal distributions on predictions performance. Finally, we also study the effect of estimation errors on the predictions. In all the cases, we compare the performance of predictions by

using predictions based on a bivariate series and predictions based on a univariate series, which allows to illustrate the advantage of the proposed method. For practical matters, we provide a complete example of application on financial data.

TABLE DES MATIÈRES

REMERCIEMENTS

Je remercie mon directeur de recherche Bruno Rémillard pour son rôle de guide, de personne ressource et aussi de collaborateur. Bruno est un mathématicien d'une efficacité impressionnante et un chercheur énergique. Collaboré avec lui fut un apprentissage inestimable et un source de motivation.

Je voudrais remercier mes collègues, Frédéric Godin, Maciej Augustyniak, Mbaye Ndoye et tous les autres présents à nos séminaires de finances pour vos présentations, vos commentaires sur mon travail et les discussions enrichissantes. J'espère que ces liens se poursuivront à travers nos carrières respectives.

La vie d'un doctorant comporte des hauts et des bas, mais est en grande partie un long travail solitaire. Merci à mes collègues du département de mathématiques d'avoir transformé ces années en un parcours rempli de discussions intéressantes et de rencontres stimulantes. Mais surtout pour ces rencontres amicales, les discussions légères et inutiles, les rires et les liens d'amitié qui se sont créés. Je pense à Raymond, Dimitri, Laurent, Louis-Xavier, Guillaume, Alexandre et tous les autres qui mériteraient autant d'être nommés.

La vie en générale comporte aussi des hauts et des bas. Pour rendre les hauts plus hauts et les bas moins bas, je remercie ma copine, ma meilleure amie, Josianne, d'être là, avec moi.

Je remercie évidemment mes parents, pour leur soutien constant et indefectible. André et Hélène, merci pour tout.

Finalement, pour m'avoir permis de me consacrer à ce projet, je tiens à remercier le Fonds québecois de la recherche sur la nature et les technologies et l'Institut de finance mathématique de Montréal pour leur soutien financier. Je remercie aussi Bruno Rémillard pour son soutien financier et les opportunités professionnelles.

Chapitre 1

INTRODUCTION

La finance mathématique est un domaine de recherche qui couvre un large spectre qui va de la recherche fondamentale jusqu'aux questions purement empiriques. Conséquemment, les chercheurs du domaine sont régulièrement en contact avec une grande diversité de problèmes, ce qui favorise le mélange des genres et pousse les chercheurs vers un certain niveau de polyvalence.

Les résultats de cette thèse reflètent bien cette réalité où deux sujets différents sont présentés. Dans un premier temps, le sujet principal de la thèse porte sur la modélisation de l'*illiquidité*[1] des actions dans le marché financier. Dans ce contexte, les questions d'arbitrage ainsi que de tarification de produits dérivés sont étudiées. L'approche est très générale et, bien que le développement de ces résultats pourra éventuellement mener à des applications, il en reste que ces travaux répondent des questions fondamentalement théoriques. En second lieu, nous proposons une méthode pour la prévision de séries temporelles basée sur les copules. Les résultats présentés sont justifiés par des méthodes numériques et ont un potentiel d'application direct en finance et dans d'autres domaines.

Les plupart des résultats classiques en finance mathématique considèrent un modèle de marché qui repose sur deux hypothèses : le marché est parfaitement liquide et sans friction. Sous l'hypothèse de liquidité parfaite, un investisseur a la possibilité d'acheter ou vendre autant d'actifs qu'il le désire au prix donné par le marché, alors que l'hypothèse d'un marché sans friction fait en sorte qu'il n'y a pas de coûts de transaction. Ces deux hypothèses rendent le mécanisme de transactions plus facile à représenter, ce qui a permis le développement d'une théorie de l'évaluation connue sous le nom d'évaluation en monde neutre au risque et

1. Bien que *illiquidité* ne soit pas un mot français, nous l'employons parce qu'il évoque clairement le concept d'un marché financier où les actifs ne sont pas parfaitement liquides. C'est aussi la francisation du terme anglais *illiquidity* qui est couramment utilisé dans la litérature.

cette théorie peut être appliquée dans un modèle de marché très général. Cependant, ces deux hypothèses restent des simplifications qui ne représentent pas parfaitement la réalité.

Au milieu des années 90, afin de rendre les modèles de marché plus réalistes, les chercheurs ont commencé à s'intéresser à la modélisation de l'*illiquidité*, ce qui correspond à l'étude de modèles de marchés où les actifs ne sont pas parfaitement liquide. Sous l'hypothèse qu'un actif est *illiquide*, le prix d'un actif varie en fonction de la quantité qui est achetée ou vendue.

Pour le marché des actions, ce phénomène s'observe dans le comportement du carnet d'ordres limites. Un ordre limite est un ordre d'achat ou de vente contenant le nombre d'actions à transiger ainsi que le prix par action. Ces ordres limites sont notés dans le carnet d'ordres limites en attendant d'être exécutés ou retirés. Par opposition, un ordre de marché est un ordre d'achat ou de vente contenant seulement le nombre d'actions à transiger, le prix par action étant celui disponible sur les marchés. Une façon de modéliser l'*illiquidité* du marché est donc de bâtir un modèle permettant de représenter la structure et le comportement du carnet d'ordres limites. Un des articles fondateurs pour la modélisation du carnet d'ordres est celui de Cetin et al. (2004) où l'utilisation de semimartingales avec paramètre spacial permet de représenter une courbe de prix qui dépend de la taille des transactions. Avec cette version simplifiée du carnet d'ordres limites, les auteurs arrivent à reconstruire les théorèmes fondamentaux de l'évaluation qui sont au coeur de la théorie de l'évaluation en monde neutre au risque. Cependant, ces résultats reposent sur le fait que l'investisseur, en utilisant des transactions suffisamment lisses, peut transiger au prix marginal, ce qui revient d'une certaine manière à transiger sur un actif parfaitement liquide. La qualité et l'originalité des idées présentées a cependant inspiré plusieurs articles dont celui présenté au chapitre 3.

Le premier article présenté dans cette thèse, inclus dans le chapitre 3, traite de la modélisation du carnet d'ordres limites et est inspiré de l'article de Cetin et al. (2004). La contribution principale de mon article est de construire un modèle général permettant de mieux représenter la structure des prix dans le carnet d'ordres limites et de modéliser l'impact des ordres de marché sur cette structure de prix. L'idée étant que les ordres de marché vide le carnet d'ordres et poussent les prix d'achat (resp. vente) vers le haut (resp. bas) alors que les nouvels ordres limites ramènent les prix vers un état fondamental. Une conséquence importante du modèle est que l'investisseur ne peut pas éviter d'avoir un impact sur la structure des prix. Il est aussi démontré que l'existence d'une certaine mesure équivalente de probabilité est suffisante pour éliminer les possibilités d'arbitrages. Ce résultat

FIGURE 1.1. Les flèches représentent les différentes dépendances modélisés.

peut-être relié au premier théorème fondamental de l'évaluation. Finalement, un exemple complet de modèle de marché est donné de même qu'une méthode de calibrage.

Ce premier article est aussi à la base du second, écrit en collaboration avec Bruno Rémillard et qui est l'objet du Chapitre 4. La question de tarification des produits dérivés est l'un des sujets les plus importants en finance mathématique et il est donc intéressant d'étudier l'effet de l'*illiquidité* pour ce problème. Cependant, il est peu envisageable de résoudre le problème de tarification de produits dérivés dans un modèle aussi général que celui présenté au Chapitre 3. Nous proposons donc une version en temps discret d'un modèle similaire pour lequel il nous est possible de calculer le prix des options européennes de vente et d'achat. Les solutions étant données dans une forme explicite, il est possible de comprendre l'effet des différents paramètres du modèle sur les prix. Puisque le modèle binomial classique est un cas particulier de notre modèle, nous sommes en mesure de comparer nos résultats avec ceux obtenus sous l'hypothèse de liquidité parfaite et ainsi mieux comprendre l'effet de l'*illiquidité* lors de l'évaluation de produits dérivés. Nous présentons aussi des conditions permettant d'éliminer les possibilités d'arbitrage. Finalement, grâce à la méthode d'optimisation duale, nous montrons qu'il est possible de réécrire le prix des options européennes comme une optimisation d'une espérance sur un ensemble de mesures de probabilité.

Le dernier sujet qui est traité dans cette thèse, qui fait l'objet du Chapitre 5, contient les résultats d'un article écrit en collaboration avec Bruno Rémillard. Dans cet article, nous proposons une méthode de prévision des séries temporelles basée sur les copules multivariées. L'avantage de cette méthode est d'utiliser des copules pour modéliser conjointement la dépendance temporelle de même que l'interdépendance de séries temporelles multivariées. En modélisant la dépendance temporelle et la l'interdépendance des séries, il est possible d'extraire plus d'informations et donc d'améliorer la performance des prévisions comparativement à des prévisions basées seulement sur la dépendance temporelle. Par exemple, pour une série bivariée $\mathbf{X}_t = (X_{1,t}, X_{2,t})$, sous les conditions que la série est stationnaire et Markovienne, l'utilisation de copules permet de modéliser la dépendance du vecteur $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$ pour $t = 1, \ldots$ Ces différentes dépendances sont représentées par les flèches dans la Figure 1.1.

En théorie, plus d'informations doit générer une plus grande précision dans les prévisions. Cependant, en pratique, il reste à démontrer si cette augmentation de la précision est significative et dans quelles circonstances notre méthode devrait être appliquée. Afin de répondre à ces questions, nous avons effectué plusieurs expériences numériques permettant de comparer la performance des prévisions basées sur une série multivariée avec la performance des prévisions basées sur la version univariée de la méthode. Nos expériences ont permis d'identifier quelles combinaisons des dépendances possibles étaient plus ou moins favorables à l'utilisation de la méthode multivariée. Nous avons aussi démontré que certaines structures de dépendance, notamment la copule de Clayton, faisaient en sorte que l'avantage des prévisions basées sur la série bivariée était très faible, peu importe la force des dépendances en jeu.

La grande force des copules est la possibilité d'isoler la structure de dépendance et les distributions marginales. Pour cette raison, nous avons aussi testé si l'effet des distributions marginales des séries avaient un impact sur la performance des prévisions. Cependant, les prévisions pour des distributions ayant une plus grande variance devraient naturellement être moins précises. Nous proposons donc une nouvelle mesure de performance permettant d'éliminer l'effet des distributions marginales et permettant de comparer la précision des prévisions.

Finalement, pour des considérations plus pratiques, nous avons testé l'effet des erreurs d'estimation sur les paramètres des copules et nous présentons aussi un exemple complet d'application sur des séries financières.

Chapitre 2

NOTIONS PRÉLIMINAIRES

Ce chapitre présente quelques notions qui seront utiles pour les deux prochains chapitres. Dans un premier temps, nous présentons quelques éléments de la théorie des semimartingales avec paramètre spatial. Cette théorie généralise l'intégrale stochastique et la formule d'Itô, voir Protter (2004) ou Durrett (1996) pour la théorie classique du calcul stochastique. En second lieu, nous présentons un rappel sur la théorie de l'évaluation en monde neutre au risque. Cette théorie est une des plus importante en mathématiques financières. Les articles formant les Chapitres 3 et 4 généralisent certains résultats de cette théorie pour un marché *illiquide*, plus précisément la relation entre absence d'arbitrage et l'existence d'une certain mesure de probabilité équivalente. Finalement, nous décrivons brièvement une méthode d'optimisation, appelée méthode duale, qui sera utilisée au Chapitre 4.

1. Semimartingales avec paramètre spatial

Dans cette section nous présentons une brève introduction à la théorie des semimartingales avec paramètre spatial. L'objectif est de donner rapidement au lecteur une vision intuitive de cette théorie. Pour cette raison, les résultats sont présentés sans preuves et le lecteur est référé au livre de Kunita (1990) pour tous les détails.

Soit $(\Omega, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$ un espace de probabilité filtré tel que $\mathcal{F}_0$ contient tous les ensembles nuls de la filtration et $\mathcal{F}_t$ est continue à droite, i.e. $\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u$. Une semimartingale avec paramètre spatial est une famille de semimartingales $\{F(x,t); x \in \mathbb{R}\}$ indexée par le paramètre $x \in \mathbb{D} \subset \mathbb{R}$. Soit X un processus prévisible, notre objectif est de construire une intégrale de la forme

$$\int_0^T dF(X_t, dt). \tag{1.1}$$

Cette intégrale permettra de définir la différentielle $dF(X_t, t)$, qui est en fait une généralisation de la formule d'Itô.

Avant d'aller plus loin, voyons quelle est la difficulté de ce problème. Par exemple, soit M une martingale locale continue et $f_t(\lambda)$ un processus prévisible pour chaque paramètre $\lambda \in \mathbb{D}$ où $\mathbb{D} \subset \mathbb{R}$. Supposons que $\int_0^T f_s(\lambda)^2 \langle M \rangle_s < \infty$ pour chaque λ. Alors, l'intégrale $M(\lambda, t) := \int_0^t f_s(\lambda) dM_s$ est une martingale locale continue, bien définie sauf pour un ensemble de mesure nulle N_λ. Le problème est que, vue comme une fonction de λ, $M(\cdot, t)$ est bien définie sauf sur l'ensemble $\bigcup_\lambda N_\lambda$. Cependant, comme $\mathbb{D}$ peut être non-dénombrable, cet ensemble n'est pas nécessairement de mesure nulle. Pour résoudre ce problème, on suppose que le processus $t \mapsto M(\cdot, t)$ est continu dans l'espace des fonctions m fois continuement dérivables $(C^{m,\delta}(\mathbb{D}), ||\cdot||_{m,\delta})$ où

$$||f||_{m,\delta} = \sup_x \frac{|f(x)|}{1+|x|} + \sum_{i=1}^m \sup_x \left| \frac{d^i}{dx^i} f(x) \right| + \sup_{x,y} \frac{\left| \frac{d^m}{dx^m} f(x) - \frac{d^m}{dy^m} f(y) \right|}{|x-y|^\delta}$$

où $\delta \in (0,1)$. En conséquence, on peut montrer que le processus de variation quadratique $A(x,y,t) = \langle M(x,t), M(y,t) \rangle$ est continu par rapport à la norme

$$\begin{aligned} ||g||^\sim_{m,\epsilon} &= \sup_{x,y} \frac{|g(x,y)|}{(1+|x|)(1+|y|)} + \sum_{1 \le \alpha \le m} \sup_{x,y} \left| \partial_x^\alpha \partial_y^\alpha f(x,y) \right| \\ &+ \sum_{0 \le \alpha \le m} ||g||^\sim_{m+\epsilon}, \end{aligned}$$

où

$$||g||^\sim_{m+\epsilon} = \sup_{x,x',y,y'} \frac{|g(x,y) - g(x',y) - g(x,y') - g(x',y')|}{|x-y|^\epsilon |x'-y'|^\epsilon}$$

et $\epsilon < \delta$. C'est en utilisant cette continuité de la variation quadratique $A(x,y,t)$ qu'il est possible de définir l'intégrale 1.1 comme la limite d'intégrales de processus simples. Comme pour l'intégrale stochastique classique, la convergence est uniforme en probabilité.

On peut maintenant définir l'intégrale pour un processus simple. Soit $\{F(x,t); x \in \mathbb{D}, t \in [0,T]\}$ une famille de semimartingales continues dans $(C^{m,\delta}(\mathbb{D}), ||\cdot||_{m,\delta})$ et X^l un processus simple, i.e. il existe une partition $\Delta = \{0 = t_0 < t_1 < \cdots < t_l = T\}$ de l'intervalle $[0,T]$ telle que $X_t^l = X_{t_k}^l$ pour tout $t \in [t_k, t_{k+1})$. On définit l'intégrale pour le processus simple X^l par

$$F(X^l, t) = \int_0^t F(X_s, ds) = \sum_{k=0}^l \left\{ F(X^l_{t_k \wedge t, t_{k+1} \wedge t}) - F(X^l_{t_k \wedge t, t_k \wedge t}) \right\}.$$

Par la suite, on construit l'intégrale stochastique $F(X,t)$ pour X un processus prévisible de façon similaire à l'intégrale stochastique classique. On commence par supposer que $F(x,t)$ est une martingale continue de carré intégrable et on montre que $F(X^l, t)$ est aussi une martingale continue de carré intégrable.

On suppose ensuite que $F(x,t)$ est une martingale locale continue et on refait le même exercise en utilisant une suite croissante de temps d'arrêts $\{\tau^n\}$ tels que $F(x,\tau^n \wedge t)$ est une martingale. Finalement, en supposant que $X = \lim_{l\to\infty} X^l$ on peut montrer que $\langle F(X^m,t) - F(X^n,t)\rangle \to 0$ p.s. lorsque $m,n \to \infty$. Puisque le variation quadratique définit une norme équivalente à la norme L^2, nous avons que $F(X^l,t)$ converge uniformément en probabilité vers $F(X,t)$. La limite $\langle F(X^m,t) - F(X^n,t)\rangle \to 0$ est obtenue grâce à la continuité par rapport à la norme mentionnée plus haut. Finalement, il est possible de construire la formule d'Itô généralisée, voir l'Annexe B du Chapitre 3,

$$dF(X_t,t) = F(X_t,dt) + \frac{\partial F}{\partial x}(X_t,t)dX_t + \frac{1}{2}\frac{\partial F}{\partial x^2}(X_t,t)d\langle X\rangle_t + \left\langle \frac{\partial F}{\partial x}(X_t,dt), X_t \right\rangle.$$

2. Théorie de l'évaluation en monde neutre au risque

Une des théories la plus importante en finance mathématique est la théorie de l'évaluation en monde neutre au risque, (Harrison and Kreps, 1979) et (Harrison and Pliska, 1981). Cette théorie démontre bien le niveau de maturité qu'a atteind la finance mathématique lorsque l'on se place sous les hypothèses d'un marché parfaitement liquide et sans friction. Cette théorie s'applique à des modèles de marché en temps continu ou discret très généraux dans lesquels les actifs financiers peuvent être représentés par des semimartingales.

La théorie de l'évaluation en monde neutre au risque contient deux principaux résultats. Ces derniers sont souvent appelés théorèmes fondamentaux de l'évaluation. Le premier affirme que l'absence d'arbitrage dans un modèle de marché est équivalente à l'existence d'une mesure de probabilité équivalente pour laquelle les actifs actualisés sont des martingales. Sans ce résultat, pour chaque modèle de marché, il faut démontrer que parmi toutes les stratégies d'investissments, aucune ne donne une opportunité d'arbitrage. Ce résultat donne donc une méthode systématique pour montrer que le marché ne contient pas d'arbitrage. Par exemple, si l'on représente les actifs par des processus d'Itô, le théorème de Girsanov permet de déterminer la mesure martingale en question, voir Shreve (2004).

Le second théorème fondamental dit que le marché est complet si et seulement si il existe une unique mesure martingale équivalente. Cela signifie que tous les droits contingents peuvent être répliqués, c'est-à-dire qu'il existe une stratégie d'investissement telle que la valeur du portefeuille est la même que la valeur du droit contingent.

En conséquence, on obtient que la valeur d'un droit contingent est l'espérance sous la mesure neutre au risque des flux monétaires actualisés à l'échéance.

On note que s'il existe plus d'une mesure martingale, alors la valeur du droit contingent n'est pas unique.

Un bon exemple de la force de cette théorie est son application au modèle de Black-Scholes, (Black and Scholes, 1973). Le modèle de Black-Scholes suppose que la valeur de l'action est un mouvement Brownien géométrique

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t},$$

avec $S_0, \mu, \sigma > 0$, et que le taux d'intérêt instantané est constant. C'est-à-dire que l'évolution d'une unité dans le marché monétaire est donnée par $dB_t = rB_t dt$, avec $r > 0$. Dans l'article de Black and Scholes (1973), la solution du problème de réplication et de tarification d'une option européenne passe par la résolution d'une équation aux dérivées partielles où la condition terminale est la valeur de l'option à l'échéance. En utilisant la théorie de l'évaluation en monde neutre au risque, la solution du problème de tarification se résume à prendre l'espérance de la valeur actualisée de l'option à l'échéance en utilisant la mesure neutre au risque,

$$\frac{1}{B_T} E^Q \left[C(S_T) \right]$$

où la dérivée de Radon-Nykodim est donnée par $\frac{dQ}{dP} = e^{-\Theta W_T - \frac{1}{2}\Theta^2 T}$ et $\Theta = \left(\frac{\mu - r}{\sigma}\right)$. Sous cette mesure, la valeur actualisée de l'action suit une distribution Log-Normale. Dans le cas d'une option européenne, nous avons que $C(s) = \max(s - K; 0)$ pour un certain $K > 0$. Il est donc très facile de calculer la valeur de l'option. Dans des cas plus généraux, puisque la valeur de l'option s'exprime comme une espérance, il est possible d'obtenir des approximations numériques par la méthode de simulation Monte Carlo, ce qui donne la possibilité d'avoir des résultats numériques pour des modèles où l'espérance ne peut pas être calculée explicitement.

3. Dualité

Nous concluons ce chapitre avec une méthode d'optimisation connue sous le nom de dualité, dualité Lagrangienne ou simplement méthode duale. Cette technique permet parfois de simplifier le problème d'optimisation puisque le problème dual est toujours concave. De plus, dans certains cas, elle apporte une interprétation différente de la solution. C'est d'ailleurs le cas pour notre problème de tarification d'options européennes à la Section 5 du Chapitre 4. Dans ce qui suit, nous allons présenter brièvement cette méthode et nous terminerons avec un exemple. La présentation suit les idées du Chapitre 5 de Boyd and Vandenberghe (2009).

Supposons le problème d'optimisation avec contraintes suivant. Soit $f_0 : \mathbb{R}^n \to \mathbb{R}$,

$$\begin{array}{cl} \text{minimiser} & f_0(x), \\ \text{sous les contraintes} & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_j(x) = 0, j = 1, \ldots, p, \end{array} \tag{3.1}$$

où le domaine $\mathcal{D} = \left\{\bigcap_{i=0}^{m} \mathrm{dom} f_i(x)\right\} \bigcap \left\{\bigcap_{j=0}^{p} \mathrm{dom} h_j(x)\right\}$ est non-vide. Notons par p^* la valeur optimale du problème d'optimisation. La fonction Lagrangienne associée à ce problème est définie par

$$L(x, \lambda, \nu) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x).$$

De sorte que la fonction duale, ou fonction Lagrangienne duale, est donnée par

$$g(\lambda, \nu) := \inf_{x \in \mathcal{D}} L(x, \lambda, \nu).$$

Il est important de noter que pour $\lambda \geq 0$ et $\nu \in \mathbb{R}^p$, $g(\lambda, \nu) \leq p^*$. En effet, soit $\tilde{x} \in \mathcal{D}$, alors

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{j=1}^{p} \nu_j h_j(\tilde{x}) \leq f_0(\tilde{x}),$$

puisque $f_i(\tilde{x}) \leq 0$ pour $i = 1, \ldots, m$ et $h_j(\tilde{x}) = 0$ pour $j = 1, \ldots, p$.

Comme on le voit, la fonction duale donne une borne inférieure pour p^*. Le problème dual et donc de déterminer la plus grande borne, soit

$$d^* = \sup_{\lambda \geq 0, \nu \in \mathbb{R}^p} g(\lambda, \nu).$$

Comme il a été mentionné plus haut, la fonction duale est concave puisqu'elle est l'infimum point par point d'une famille de fonctions concaves en (λ, ν). Lorsque $d^* \leq p^*$ nous disons qu'il y a dualité faible et qu'il y a dualité forte lorsque $d^* = p^*$. Dans ce cas, le problème dual est équivalent au problème initial.

Une question importante est évidemment de déterminer si l'on a la dualité forte. Voici une condition suffisante pour la dualité forte qui s'applique à notre problème. Supposons que le problème 3.1 est convexe et est de la forme suivante

$$\begin{array}{cl} \text{minimiser} & f_0(x), \\ \text{sous les contraintes} & f_i(x) \leq 0, i = 1, \ldots, m \\ & Ax = b, \end{array}$$

avec $f_i(x)$ convexes pour tout $i = 1, \ldots, m$. S'il existe x_0 dans l'intérieur de $\mathcal{D}$ tel que $f_i(x_0) < 0$ pour tout $i = 1, \ldots, m$ et $Ax_0 = b$ alors il y a dualité forte. Cette dernière condition est la condition de Slater.

Remarque 3.1. *En fait, la condition exacte de Slater est que x_0 appartient à l'intérieur relatif de $\mathcal{D}$, mais pour notre probème la formulation précédente est suffisante.*

Afin d'illustrer cette méthode, voici un exemple d'application sur un problème d'optimisation linéaire.

Exemple 3.1. *Soit le problème d'optimisation suivant.*

$$\begin{array}{ll} \textit{minimiser} & c^T x, \\ \textit{sous les contraintes} & x \leq 0, \\ & Ax = b. \end{array}$$

La fonction Lagrangienne est donnée par

$$L(x, \lambda, \nu) = -b^T \nu + (c + A^T \nu - \lambda)^T x.$$

De sorte que la fonction duale est

$$g(\lambda, \nu) = -b^T \nu + \inf_x \left(c + A^T \nu - \lambda\right)^T x.$$

Puisque $x \in \mathbb{R}$, nous avons que

$$g(\lambda, \nu) = \begin{cases} -b^T \nu, & \textit{si } A^T \nu - \lambda + c = 0, \\ -\infty, & \textit{sinon.} \end{cases}$$

Soit $\mathcal{D}^* = \left\{\nu \in \mathbb{R}^p ; \exists \lambda \geq 0 \text{ tel que } A^T \nu - \lambda + c = 0\right\}$, le problème dual est $\sup_{\nu \in \mathcal{D}^*} -b^T \nu$. Si A est inversible, alors le problème dual est simplement $\sup_{\lambda \geq 0} -b^T \left\{(A^T)^{-1}(\lambda - c)\right\}$.

Chapitre 3

GENERAL MODEL FOR LIMIT ORDER BOOKS AND MARKET ORDERS

Résumé

Dans cet article nous construisons un modèle en temps continu pour la structure et la dynamique du carnet d'ordres limites. L'aspect novateur de cet approche est d'utiliser des processus aléatoires ayant des valeurs dans l'espace des fonctions continues pour modéliser directement le coût des transactions. Nous déduisons ensuite le comportement du portefeuille d'un gros investisseur utilisant des ordres de marchés. Le comportement de ce portefeuille tient compte de l'effet sur les prix des ordres de marché qui vident le carnet d'ordres de même que l'arrivée de nouvelles ordres limites. Nous prouvons que l'existence d'une certaine mesure équivalente de probabilité est une condition suffisante pour s'assurer que le marché ne contient pas d'arbitrage. Nous présentons finalement un exemple pratique de même qu'une méthode d'estimation pour la structure et la dynamique du carnet d'ordres limites de même que pour la vitesse d'arrivée des nouvelles ordres limites.

Abstract

In this paper we build a general model for the structure and the dynamic of the limit order book in continuous time. The novelty of our approach is to use random processes with value in the space of continuous functions to model the cost of a transaction. By taking the viewpoint of a large investor who trades using market orders, we derive the dynamic for the value of his portfolio. This dynamic takes into account the opposite forces of the market orders depleting the limit order book and the arrival of new limit orders. We prove that the existence of some equivalent probability measure is sufficient to rule out arbitrage. We also provide a practical example with a complete method of estimation for the structure and the dynamic of the order book as well as for the arrival of new limit orders.

1. Introduction

Most of the major results in the mathematical finance literature are based on the assumptions of frictionless and perfectly liquid markets. Obviously, these two assumptions simplify the trading mechanism and market models can be made more tractable. However, practitioners face transaction fees, delays before transactions are executed and that stock prices are influenced by the size of their transactions. Those are all features which go against frictionless and perfectly liquid market assumptions. Indeed, with the aim of representing more closely the reality of financial markets, researchers tried to incorporate these features in their models. Among others, it is of interest to include the impact that transactions may have on prices, which is a characteristic of an illiquid market. Inside this new paradigm, the classical problems have to be revisited and new ones appear.

One of the main topic in mathematical finance is the one of pricing derivatives and in parallel the problem of hedging. Under the hypothesis of illiquidity, some of the results we can find are : Frey and Patie (2002) and Liu and Yong (2005), where they used an EDP close to the one developed by Black and Scholes (1973) to price and hedge an European option, Cvitanić and Ma (1996) use forward-backward stochastic differential equations, Rogers and Singh (2010) solve the hedging problem using a Hamilton-Jacobi-Bellman equation and Roch (2011) uses backward stochastic differential equations. New questions also arise, such as the problem of the feedback effect. Some results on that question are : Frey and Stremme (1997), Platen and Schweizer (1998) and Schönbucher and Wilmott (2000) where illiquidity is modeled through an equilibrium solution. On a different topic, Jarrow et al. (2010) use illiquidity to study price bubbles. Since trading strategies have an impact on price, there is also the problem of how to optimally liquidate a portfolio, some results on that topic are given by Obizhaeva and Wang (2005) and Predoiu et al. (2011). Finally, on another level, Jarrow (1992) and Jarrow (1994) build a general model for price impact in a discrete time setting to study market manipulations and new arbitrage opportunities created by the impact that an investor may have on prices.

Another topic of interest is the possibility to extend to theory of risk neutral evaluation, see Harrison and Kreps (1979) and Harrison and Pliska (1981), in a model where price impact is considered. On this topic, the most complete results can be found in Cetin et al. (2004) and Bank and Baum (2004), where they give conditions for the absence of arbitrage and study the problem of pricing and hedging contingent claims. More specifically, they both use semimartingales with space parameters to build a general model for the supply curve of a stock. In the case of Cetin et al. (2004), they show a complete version of the fundamental theorems of asset pricing. However, their results are based on the possibility for the large investor to trade without impact by using trading strategies that are continuous and with finite variation.

Thereafter, many papers have been built on the model of Cetin et al. (2004) and proposed modifications so that the impact on prices cannot be avoided. In Cetin et al. (2006), they study the impact of illiquidity on a Black & Scholes type of hedging for a discrete time model. By using discrete trading times, the use of continuous trading strategies is impossible and impact on prices cannot be avoided. Later, Cetin et al. (2010) study super-replication strategies in the continuous time model of Cetin et al. (2004). They impose conditions on the trading process so that it has infinite variation, and hence, has an impact on prices. In Roch (2011) and Roch and Soner (2013), the impact is unavoidable by defining an affected supply curve which comes from modeling the lasting effect of market orders on the limit order book.

From the early work of Kyle (1985) and as explained in Roch and Soner (2013), a good model for the limit order book should include three aspects : *depth*, *resilience* and *tightness*. The *depth* is the size of the trade required to move the stock price by a certain amount, *resilience* is the speed at which the prices recover from a big transaction, and *tightness* is the cost of rapidly turning around a position. Up to now, the models found in the literature are either specific cases to insure the computability of solutions to specific problems, or do not correctly represent all three aspects previously mentioned.

In this paper, we build a general model in continuous time for the limit order book which incorporates *depth*, *resilience* and *tightness*. The approach is to use random processes with values in the space of continuous functions to model directly the cost (resp. profit) of transactions, instead of the usual approach of modeling the price of the stock. By using state variables, we are also able to represent the impact of market orders in a way that it is unavoidable.

One of our main results is to show that the existence of some equivalent probability measure rules out arbitrage. Since our model includes a perfectly liquid and frictionless market as a specific case, our no-arbitrage theorem generalizes the sufficiency part of the first fundamental theorem of asset pricing. On a more practical level, we also present an explicit example along with an estimation method.

The next section is dedicated to the description of the trading mechanism of market orders and the general structure of the limit order book structure. We define the mathematical model in Section 3 and give the value of the portfolio for discrete time trading in Section 4. In Section 5, we derive the value of the portfolio for continuous time trading and give our no-arbitrage result in Section 6. Finally, we use the generalized Itô's formula to describe the dynamic of the price processes in Section 7. In Section 8, we build a complete example and provide an estimation method based on observed values of the limit order book. The last two sections give some possible extensions and concluding remarks.

2. Description of limit order book transactions

To create our market model we suppose two categories of investors. The first category comprises small investors who send limit orders. These limit orders are compiled in the order book and contain the prices at which the small investors are willing to buy or sell and how many shares they want to trade. When two orders agree, the broker execute the transactions and delete those orders from the order book. See Table 3.1 for a snapshot of the order book for Amazon Inc.

Table 3.1. Five first orders of the limit order book for Amazon Inc., 29 June 2012, at 13 :47.

Bid		Ask	
Price	Size	Price	Size
227.53	100	227.64	100
227.47	200	227.65	100
227.45	100	227.67	300
227.44	200	227.69	400
227.43	100	227.72	100

We suppose that those small investors provide the liquidity and their actions are responsible for the uncertainty in the price dynamic. The second category is a single investor, which is sometimes referred to as the *large investor*, who trades with market orders. Those market orders contain the amount of stocks that have to be bought or sold and the transactions are executed immediately at the best price available in the limit order book. For instance, based on the order book in Table 3.1, an investor who sends a market order to buy 150 shares of Amazon will pay \$227.64 per share for the first 100 shares and \$227.65 for the 50 remaining. So, as we see, the perfect liquidity assumption is violated since the size of the transaction has an impact on the price per share. Finally, as time moves forward, new limit orders will refill the order book as well as new market orders will tend to deplete it, thus creating the dynamic of the market that we aim to model.

The term *large investor* comes from the literature about price impact models where the aim is to model the behavior of an investor whose trades are big enough to have an impact on prices. Although we will refer to the *large investor*, in the context of market orders, any size of trade has an impact, so we mainly use the term *large investor* for differentiation.

Before moving on with the mathematical definition of the model, we would like to highlight some of the features of the dynamics of the order book structure that appear in the above description. Firstly, we see that we have two sets of prices for the same stock, bid and ask prices, evolving simultaneously. Moreover, it is possible for an investor, although unlikely profitable, to buy and sell at the same time. Secondly, the large investor's transactions have an impact on the prices. For instance, if we refer to Table 3.1, if the large investor buys 100 shares, then the price to buy another share will no longer be 227.64 but rather 227.65. This phenomenon relates to the *depth* of the limit order book. In opposition to that, small investors will keep sending new orders which will more or less quickly refill the order book and bring the prices back toward their *fundamental value*. This is the *resilience*. We also see that there is a dependence on the trajectory of the transactions. For instance, if the large investor bought 1000 shares one minute ago, the impact of his transaction will be greater than if he bought those 1000 shares one day ago, since in the later case new limit orders will have refilled the order book.

3. Market Model

The first step in building our market model is to have a general representation of the ask part and bid part of the limit order book. Let $(\Omega, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$ be a filtered probability space which satisfies the usual assumptions (Protter, 2004). We define two processes with values in the space of continuous functions $C^{2,1}([0,\infty)\times[0,\infty))$.

$$\begin{aligned} F^a, F^b: \quad & [0,\infty)\times\Omega \to C^{2,1}([0,\infty)\times[0,\infty)) \\ (t,\omega) \mapsto \quad & F^a(t,\cdot,\cdot,\omega), F^b(t,\cdot,\cdot,\omega). \end{aligned} \tag{3.1}$$

To clarify the notation, for $t\in[0,\infty)$ and $\omega\in\Omega$ fixed, $(y,x)\mapsto F^a(t,y,x,\omega)$ and $(y,x)\mapsto F^b(t,y,x,\omega)$ are functions of $(y,x)\in[0,\infty)\times[0,\infty)$ that are respectively twice continuously differentiable in y and continuously differentiable in x.

Remark 3.1. *To simplify the notation we extend* $(y,x)\mapsto F^a(t,y,x,\omega)$ *and* $(y,x)\mapsto F^b(t,y,x,\omega)$ *to* $[0,\infty)\times\mathbb{R}$ *by setting* $F^a(t,y,x,\omega) = F^b(t,y,x,\omega) = 0$

with $x < 0$. With this extension, F^a and F^b are not necessarily continuously differentiable respect to x at $x = 0$. However, it is sufficient to establish our results that the right derivatives exist and are continuous at $x = 0$, which is given by (3.1).

The processes F^a and F^b are respectively the structure of the ask and the bid parts of the order book and, for all $x \geq 0$, the value $F^a(t, y, x)$ (resp. $F^b(t, y, x)$) gives the cost (resp. profit) for buying (resp. selling) x shares at time t. The variable y gives the level of impact on the order book of the large investor's past transactions. Following remark 3.1, we also define $F^a(t, y, x) = F^b(t, y, x) = 0$ if $x \leq 0$, for any $t \geq 0$ and any $y \geq 0$, which assigns a cost (profit) of zero for a transaction of negative size.

Example 3.1. *Although it does not satisfy the differentiability conditions, we give this example in order to provide a more concrete case for F^a and F^b. In Section 8, we present an example of differentiable functions that can fit on the discrete structure of the order book given in Table 3.1.*

Suppose that $F^a(0, 0, \cdot)$ and $F^b(0, 0, \cdot)$ represent the structure of the order book in Table 3.1. Then

$$F^a(0,0,x) = \begin{cases} 227.64x, & x \in [0, 100); \\ 22764 + 227.65(x - 100), & x \in [100, 200); \\ 68294 + 227.67(x - 200), & x \in [200, 500). \end{cases}$$

$$F^b(0,0,x) = \begin{cases} 227.53x, & x \in [0, 100); \\ 22753 + 227.47(x - 100), & x \in [100, 300); \\ 90994 + 227.67(x - 300), & x \in [300, 400). \end{cases}$$

For instance, the cost of buying all the 300 shares available at the price \$227.67 would be given by $F^a(0, 0, 500) - F^a(0, 0, 200)$ and the profit from selling the first 50 shares would be $F^b(0, 0, 50)$. To allow for arbitrary large transactions these functions could be extended to all $x > 0$.

Now, to have a structure which is coherent with the trading mechanism we described in Section 2 we impose the following conditions :

(P1) : For x and y fixed, $t \mapsto F^a(t, y, x)$, $t \mapsto F^b(t, y, x)$ are continuous semimartingales and for each t

$$E[F^a(t, y, x)] < \infty \text{ and } E[F^b(t, y, x)] < \infty.$$

(P2) : For all $y_1, y_2 \geq 0$ and all $t \geq 0$, $F^a(t, y_1, x) \geq F^b(t, y_2, x) > 0$ and $F^a(t, y_1, x) = F^b(t, y_2, x) = 0$ for all $x \leq 0$.

(P3) : $F^a(t, y, x)$ is increasing in y and increasing and convex in x. $F^b(t, y, x)$ is decreasing in y and increasing and concave in x.

Property **(P1)** gives technical conditions mainly required for the continuous model. Otherwise, it says that the structure of the order book incorporates all the informations from the past and that the expected cash flow of a transaction of size $x > 0$ is finite. Property **(P2)** simply states that at any time the ask price is never smaller than the bid price, so that our model satisfies the *tightness* condition. Property **(P3)** imposes a convex shape to the ask part and a concave shape to the bid part. Financially speaking, it means that the price per share increases as long as the size of the buy increases and profit per share decreases as long as the size of the sale increases. See Figure 3.1 for an example of the general shape of the cost (profit) of a transaction. The degree of convexity and concavity are related to the *depth* of the limit order book. A higher degree of convexity (resp. concavity) means less *depth*, since the price per share is more sensitive to the size of the transaction. Moreover, we know that the impact always plays against the investor, that is, the impact on ask price pushes the price upward whereas the impact on bid price pull the price downward. For that reason, we impose that $F^a(t,y,x)$ is increasing in y and $F^b(t,y,x)$ is decreasing in y. One consequence of **(P3)** is that for all $x_1, x_2 \geq 0$, $F^a(t,y,x_1+x_2) \geq F^a(t,y,x_1) + F^a(t,y,x_2)$ and $F^b(t,y,x_1+x_2) \leq F^b(t,y,x_1) + F^b(t,y,x_2)$.

FIGURE 3.1. General shape of the cost (profit) of a transaction. The higher line gives the cost of buying and the lower line gives the profit from selling.

Before defining the trading process let's recall the following mathematical notions. The variation (or total variation) of a process X is defined as $\mathcal{V}_{[t_1,t_2]}(X) = \lim_{n\to\infty} \sum_{k=1}^{n} |X_{t_k^n} - X_{t_{k-1}^n}|$, where $\{t^n\}$ is an increasing sequence of partitions of $[t_1, t_2]$ such that $\lim_{n\to\infty} \max_{k=1,\ldots,n}\{|t_k^n - t_{k-1}^n|\} = 0$. One can also define

$$X_t^a = \int_0^t (dX_s)^+ \;=\; \lim_{n\to\infty} \sum_{k=1}^{n} \left(X_{t_k^n} - X_{t_{k-1}^n}\right)^+,$$

$$X_t^b = \int_0^t (dX_s)^- \;=\; \lim_{n\to\infty} \sum_{k=1}^{n} \left(X_{t_k^n} - X_{t_{k-1}^n}\right)^-,$$

where $(x)^+ = \max\{x, 0\}$ and $(x)^- = -\min\{-x, 0\}$. The condition that $\mathcal{V}_{[0,t]}(X)$ is finite for all t implies that $\int_0^t (dX_s)^+$ and $\int_0^t (dX_s)^-$ are finite since

$$\mathcal{V}_{[0,t]}(X) \;=\; \int_0^t (dX_s)^+ + \int_0^t (dX_s)^-.$$

Definition 3.1. *A trading strategy is a predictable process with finite variation,* $\{X_t\}_{t\geq 0}$ *such that* $X_0 = 0$. *We define*

$$X_t^a = \int_0^t (dX_s)^+ \text{ and } X_t^b = \int_0^t (dX_s)^-$$

and set $X_0^a = X_0^b = 0$.

In the preceding definition, X_t is the number of shares in the portfolio at time t. The values X_t^a and X_t^b are respectively the aggregate number of shares bought and sold up to time t. We also have that $X_t = X_t^a - X_t^b$.

3.1. Level of impact

One of the main difficulty in modeling the dynamic between market orders and limit orders is the lasting effect of the impact on the order book. As we described in Section 2, after a transaction from the investor (market order), some of the limit orders are removed and it takes a certain time before new limit orders arrive to replace them. This aspect corresponds to the *resilience*. For instance, suppose an investor buys 200 shares of Amazon and the limit order book is given by Table 3.1. Immediately after this trade, the ask price is no longer 227.64$ but 227.67$. The small investors will want to take advantage of this higher price and will submit new limit orders, so that the price gradually decreases toward its *fundamental value*.

In Predoiu et al. (2011), they build a static model for the limit order book by defining a measure μ such that $\mu\{[0,x)\}$ gives the number of limit orders available between the prices $[S_t^a, S_t^a + x)$ where S_t^a is the ask price at time $t \geq 0$. Then, they keep track of the impact of the past market orders by defining the *volume effect process*

$$E_t = X_t^a - \int_0^t h(E_s)ds,$$

where $h(0) = 0$ and h is strictly increasing. The *volume effect process* takes into account the total of shares bought, X_t^a, and h models the *resilience* of the limit order book.

This idea can be generalized by two processes A, B which are state variables giving respectively the level of impact for the ask part and the bid part of the order book. Hence, the cost of buying x shares at time t is $F^a(t, A_t, x)$ and the profit from selling x shares is $F^b(t, B_t, x)$. Accordingly to the structure we imposed on F^a and F^b, the processes A and B should satisfy this simple condition.

(P4) The processes A and B, are non-negative predictable processes where, for all $t \geq 0$, $A_t = 0$ if and only if $X_t^a = 0$ and $B_t = 0$ if and only if $X_t^b = 0$.

Example 3.2. *We can define* $A_t = G(t, X_t^a)$ *and* $B_t = H(t, X_t^b)$ *where* $G(t,x)$ *et* $H(t,x)$ *are continuous non-negative deterministic functions, increasing in* x *and decreasing in* t*. We also suppose that, for all* $t \geq 0$, $G(t,0) = H(t,0) = 0$ *and* $G(t,x), H(t,x) > 0$ *if* $x > 0$*.*

Let $\delta > 0$*, then* $H(t, X_t^a + \delta) > H(t, X_t^a)$ *and* $G(t, X_t^b + \delta) > G(t, X_t^b)$*, one sees that the impact is more important if the total number of shares bought or sold is higher. The fact that* G *and* H *are decreasing in time means that if the large investor stop trading, then the impact will tend to decrease, which means that new orders are refilling the order book. For instance, let* $k > 0$ *and suppose that* $X_{t+k}^a = X_t^a$*, then* $H(t+k, X_{t+k}^a) < H(t, X_t^a)$ *(similarly for* G*). The use of these state variables allow to model the lasting effect of the impact as well as the arrival of new limit orders.*

The next example comes from Predoiu et al. (2011).

Example 3.3. *In this example,* A *and* B *are defined as stochastic differential equations.*

$$A_t = X_t^a - \int_0^t h(A_s)ds$$

and

$$B_t = X_t^b - \int_0^t h(B_s)ds$$

where g *and* h *are non-negative functions with* $g(0) = h(0) = 0$*.*

4. Discrete Time Portfolio

The focus of this paper is on defining a general price model based on the limit order book structure. For that matter, the interest rate is not of a big importance and it is set to zero. In Section 9 we discuss in more details how to include interest rates.

In this context, the portfolio value will be the difference between the liquidation value of the portfolio and the cash flow from the trades. The liquidation value at time t is the profit (resp. cost) of closing the position X_t, while the cost of the portfolio is the cash flow from rebalancing the position at each trading time. Since there is no interest rate, this is equivalent to the self-financing portfolio property.

To compute the portfolio value, we first need to determine the cash flow of the portfolio. Since we are defining the discrete time version of the model, we suppose that transactions are made at time $t_0 = 0 < t_1 < \cdots < t_n = T$ with $T > 0$. So, we restate our definition of trading strategies to satisfy the discrete time setting.

Definition 4.1. *Let $t_0 = 0 < t_1 < \cdots < t_n = T$ be a partition of $[0,T]$. A trading strategy is a process $\{X_{t_i}\}_{i=0,1,\ldots,n}$ such that $X_0 = 0$ and X_{t_i} is $\mathcal{F}_{t_{i-1}}$-measurable for $i = 1,\ldots,n$. We define $X^a_{t_i} = \sum_{j=1}^i \left(X_{t_j} - X_{t_{j-1}}\right)^+$ and $X^b_{t_i} = \sum_{j=1}^i \left(X_{t_j} - X_{t_{j-1}}\right)^-$ for $i = 1,...,n$ and set $X^a_0 = X^b_0 = 0$.*

The cash flow of the portfolio for a trading strategy X is given by

$$\sum_{i=1}^n \left\{F^a(t_i, A_{t_i}, \delta X^a_{t_{i+1}}) - F^b(t_i, B_{t_i}, \delta X^b_{t_{i+1}})\right\} \tag{4.1}$$

where $\delta X_{t_i} = X_{t_i} - X_{t_{i-1}}$.

We define the liquidation value as the profit from liquidating the portfolio in one transaction. Suppose the portfolio is liquidated at time T. Then, its liquidation value is

$$F^b(T, B_T, X_T) - F^a(T, A_T, -X_T). \tag{4.2}$$

One recalls that in (**P1**) we defined $F^a(t,y,x) = F^b(t,y,x) = 0$ for $x \leq 0$. Consequently, the value of the portfolio is given by the difference between the liquidation value, Equation (4.2), and the sum of the cash flows for each transaction, Equation (4.1), denoted by

$$\begin{aligned} V_T(X) &= F^b(T, B_T, X_T) - F^a(T, A_T, -X_T) \\ &\quad - \sum_{i=1}^n \left\{F^a(t_i, A_{t_i}, \delta X^a_{t_{i+1}}) - F^b(t_i, B_{t_i}, \delta X^b_{t_{i+1}})\right\}. \end{aligned} \tag{4.3}$$

Note that this is equivalent to the value of a self-financing portfolio if the market contains a stock and a bank account where the interest rate is zero.

The next example shows that we can recover classical models from ours.

Example 4.1. *Suppose that there is a bid/ask spread and the market is perfectly liquid. Let $\{S^a_t\}_{t\geq 0}$ and $\{S^b_t\}_{t\geq 0}$ be two non-negative processes with $S^b_t \leq S^a_t$. Set $F^a(t,y,x) = S^a_t x$ and $F^b(t,y,x) = S^b_t x$ for $x \geq 0$ and $F^a(t,y,x) = F^b(t,y,x) = 0$ if $x \leq 0$. Since the market is perfectly liquid we removed the dependence in y.*

For any trading strategy X and any time $T > 0$ the value of the portfolio is

$$\begin{aligned} V_T(X) &= S^b_T(X_T)^+ - S^a_T(X_T)^- - \sum_{i=1}^n \left\{S^a_{t_{i-1}} \delta X^a_{t_i} - S^b_{t_{i-1}} \delta X^b_{t_i}\right\} \\ &= X^a_T\left(S^b_T - S^a_{T-1}\right)\mathbf{1}_{[0,\infty)}(X_T) - X^b_T\left(S^b_T - S^b_{T-1}\right)\mathbf{1}_{[0,\infty)}(X_T) \\ &\quad + X^a_T\left(S^a_T - S^a_{T-1}\right)\mathbf{1}_{(-\infty,0]}(X_T) - X^b_T\left(S^a_T - S^b_{T-1}\right)\mathbf{1}_{(-\infty,0]}(X_T) \\ &\quad + \sum_{i=1}^{n-1} X^a_{t_i}(S^a_{t_i} - S^a_{t_{i-1}}) - \sum_{i=1}^{n-1} X^b_{t_i}(S^b_{t_i} - S^b_{t_{i-1}}) \end{aligned} \tag{4.4}$$

which is the value of the portfolio under the assumptions of perfect liquidity and with a bid/ask spread. If we suppose there is no bid/ask spread, that is $S^a_t = S^b_t =$

S_t for all t, then we have

$$V_T(X) = \sum_{i=1}^{n} X_i(S_{t_i} - S_{t_{i-1}})$$

which is the value of the portfolio in classical theory.

Remark 4.1. *Another specific aspect of price impact models is that portfolio value is an ambiguous concept. Jarrow (1992), Cetin et al. (2004) and Bank and Baum (2004) also discuss this particularity. Once again, since transactions influence the value of the stock, there are different values we can associate to the portfolio. For instance, one can use the marked-to-market value. This is the value if one considers that the portfolio is liquidated at the market price (there is no impact). In this case, we would redefine the liquidation value* (4.2) *by*

$$F^b(T, B_T, 1)X_T\mathbf{1}_{[0,\infty)}(X_T) + F^a(T, A_T, 1)X_T\mathbf{1}_{(-\infty,0]}(X_T).$$

There is also the optimal value. We understand that liquidating the portfolio in one-block transaction will create a big impact on price, which is unfavorable for the investor. A better strategy should be to divide the liquidation of the portfolio in many smaller transactions, which would create a smaller impact. However, dividing the liquidation in smaller transactions also means that the liquidation is carried over some period of time, adding uncertainty since the value of the stock might change during the liquidation process. Consequently, the optimal value of the portfolio is again ambiguous as it depends on the criteria to optimize. This problem is related to the optimal execution problem ; see, e.g. Predoiu et al. (2011) and Obizhaeva and Wang (2005). For our concern, we will use the liquidation value of the portfolio which is the value when the portfolio is liquidated in one transaction.

5. Portfolio Value In Continuous Time

Our continuous time model shares some of the features of those found in Bank and Baum (2004) and Cetin et al. (2004) as it uses semimartingales with space parameters. However, our construction differs from theirs since it is impossible for the large investor to avoid the impact.

The liquidation of the portfolio is carried in one block transaction so that the difference between the discrete time portfolio and the continuous time portfolio is that we have to define the cash flow of the portfolio, Equation (4.1), for continuous time trading strategies from Definition 3.1. Proposition A.1 in Appendix A.1 shows that if X is a continuous time trading strategy, then the cash flow of the portfolio is given by

$$\sum_{0<t\le T}\left\{F^a(t,A_{t-},\Delta X_t^a)-F^b(t,B_{t-},\Delta X_t^b)\right\} \tag{5.1}$$
$$+\int_0^T f^a(t,A_{t-})dX_t^{a,c}-\int_0^T f^b(t,B_{t-})dX_t^{b,c},$$

where $f^a(t,y)=\lim_{x\searrow 0}\frac{F^a(t,y,x)}{x}$ and $f^b(t,y)=\lim_{x\searrow 0}\frac{F^b(t,y,x)}{x}$. The superscript c stands for the continuous part of a process. Here, $f^a(t,y)$ and $f^b(t,y)$ represent the value per share for an infinitesimal transaction. This notation will be used for the remaining of the paper.

Let X be a trading strategy satisfying Definition 3.1, then the value of the portfolio is

$$\begin{aligned} V_T(X) &= F^b(T,B_T,X_T)-F^a(T,A_T,-X_T) \\ &\quad -\sum_{0<t\le T}\left\{F^a(t,A_{t-},\Delta X_t^a)-F^b(t,B_{t-},\Delta X_t^b)\right\} \\ &\quad -\left(\int_0^T f^a(t,A_t)dX_t^{a,c}-\int_0^T f^b(t,B_t)dX_t^{b,c}\right). \end{aligned}$$

Example 5.1. *This example is based on the model in Predoiu et al. (2011). Only the ask part is defined.*

Let S be a non-negative continuous martingale. Let $\mu(s)$ be a continuous increasing function with $\mu(0)=0$. Let $t\ge 0$ and suppose there is no impact from the large investor, the function $\mu(s)$ gives the number of limit orders available in the set of prices $[S_t,S_t+s)$ at time t. Let X_t^a be a continuous non-decreasing process with $X_0^a=0$, giving the number of shares bought up to time t. We define the state variable

$$A_t=X_t^a-\int_0^t h(A_s)ds.$$

where h is a non-negative function with $h(0)=0$. For all $x\ge 0$ we define

$$F^a(t,y,x) = S_t x+\int_{\mu^{-1}(y)}^{\mu^{-1}(x+y)}\xi d\left(\mu(\xi)-y\right).$$

The cost of buying $x>0$ shares at time t is $F^a(t,A_t,x)$ and the total cost of buying X_T^a shares is

$$C_T(X) = \int_0^T S_t dX_t^{c,a}+\int_0^T \mu^{-1}(A_{t-})dX_t^{a,c}+\sum_{0<t\le T}F^a(t,A_{t-},\Delta X_t^a). \tag{5.2}$$

5.1. Cost of impact

One of the main goal of this paper is to build a general model where the impact on price is unavoidable for the large investor. We define the cost of impact as the difference between the value of the portfolio with $A \equiv B \equiv 0$ and the value of the portfolio as given in Equation (5.2). In the following, we show that the cost of impact is zero if and only if the investor does not trade.

The cash flow from trading x shares at time t for the large investor if we don't consider the impact is $F^a(t,0,x) - F^b(t,0,x)$. So that the cost of impact for a trading strategy X is

$$\begin{aligned} I_T(X) &= F^b(T,0,X_T) - F^b(T,B_T,X_T) - (F^a(T,0,-X_T) - F^a(T,A_T,-X_T)) \\ &\quad - \sum_{0<t\leq T} \{F^a(t,0,\Delta X^a_t) - F^a(t,A_{t-},\Delta X^a_t)\} \\ &\quad + \sum_{0<t\leq T} \left\{F^b(t,0,\Delta X^b_t) - F^a(t,B_{t-},\Delta X^b_t)\right\} \\ &\quad - \int_0^T \{f^a(t,0) - f^a(t,A_t)\}\, dX^{a,c}_t \\ &\quad + \int_0^T \left\{f^b(t,0) - f^b(t,B_t)\right\} dX^{b,c}_t. \end{aligned}$$

Remark 5.1. *One of the consequence of **(P2)** and **(P3)** is that $f^a(t,y)$ and $f^b(t,y)$ are positive. Let $x_2 > x_1 > 0$ and $y \geq 0$, then we have that $\frac{F^a(t,y,x_1)}{x_1} > \frac{F^b(t,y,x_1)}{x_1} \geq \frac{F^b(t,y,x_2)}{x_2} > 0$.*

With the last remark and **(P4)** we see that $I_T(X) \geq 0$ and $I_T(X) = 0$ if and only if $X \equiv 0$. So, one can see that the impact on price is unavoidable for the large investor.

6. NO-ARBITRAGE THEOREM

In this section we show that the existence of some equivalent probability measure is enough to rule out arbitrage in our framework. This result extends what can be found in the seminal work of Harrison and Kreps (1979) and Harrison and Pliska (1981) to market models under illiquidity assumptions. It also shows that the sufficiency part of the result of Cetin et al. (2004) still holds in our general framework where the impact of transactions cannot be avoided. The fact that only the sufficiency part holds agrees with the results of Roch (2011) and Roch and Soner (2013) where they restrict to a linear structure for the supply curve.

For our part, our result requires that $F^a(t,0,x)$ and $F^b(t,0,x)$ are respectively supermartingale and submartingale under the equivalent measure, instead of being martingales. To show that this generalisation is required, we build an example where $F^a(t,0,x)$ and $F^b(t,0,x)$ are respectively supermartingale and submartingale, where there is no arbitrage, but there is no equivalent martingale measure.

To prove our result about arbitrage we need to restrict the class of trading strategies.

Definition 6.1. *(Admissible trading strategies) We say that a trading strategy* X *is admissible if* $E\left[F^a(t, A_t, X_t^a)\right] < \infty$ *and* $E\left[f^b(t,0)X_t^b\right] < \infty$ *for all* $t \geq 0$.

We will use the following definition for arbitrage.

Definition 6.2. *(Arbitrage) We say that a market admits arbitrage if there exists an admissible trading strategy* X *such that*

$$\mathbf{P}\left(V_T(X) \geq 0\right) = 1 \text{ and } \mathbf{P}\left(V_T(X) > 0\right) > 0.$$

for some $T > 0$.

In financial terms, an arbitrage is a strategy with no initial investment that gives a positive probability to make a profit and a probability zero of losing money.

Remark 6.1. *One notes that from our definition of a trading strategy (Definition 3.1), we exclude the possibility to invest explicitly in the bank account. The bank account, as an investment opportunity, is only there to allow the construction of a self-financing portfolio. As a result, the strategy which consists in investing a positive amount of money in the bank account at* $t = 0$ *and withdrawing it at* $t = T$ *cannot be considered as an arbitrage opportunity.*

The next theorem states that the existence of an equivalent martingale measure such that $\{F^a(t, 0, x)\}_{t\geq 0}$ and $\{F^b(t, 0, x)\}_{t\geq 0}$ are respectively supermartingales and submartingales for all $x \geq 0$, is sufficient to rule out arbitrage. This result generalizes the sufficiency part of the fundamental theorem of asset pricing for perfectly liquid assets.

To our knowledge, the only result giving an equivalence relation between the absence of arbitrage and the existence of an equivalent measure for illiquid assets is found in Cetin et al. (2004). However, the necessity part of the theorem relies on the fact that the investor can avoid the effect of illiquidity by using trading strategies which are continuous and with finite variation. So, from our Theorem 6.1, it seems that we lose the necessity part of the result when the impact cannot be avoided. The fact that only the sufficiency part holds is also seen in Roch and Soner (2013) and Roch (2011), but our theorem is not restricted to diffusion processes and linear order book structures.

Theorem 6.1. *(No Arbitrage) Suppose there exists a probability measure* $\mathbf{Q}$*, equivalent to* $\mathbf{P}$*, such that* $F^a(t, 0, x)$ *is a* $\mathbf{Q}-$*supermartingale and* $F^b(t, 0, x)$ *is a* $\mathbf{Q}-$*submartingale for each* $x \geq 0$*. Then, there is no arbitrage for admissible trading strategies.*

The proof is in Appendix C.

The interest of this result is to give a general method to rule out the existence of arbitrage strategies, a required feature of a financial model. The idea behind the theorem is the following : for an admissible trading strategy $X = (X^a, X^b)$, let $H_X = \{\omega \in \Omega : V_T(X,\omega) \geq 0\}$ and $G_X = \{\omega \in \Omega : V_T(X,\omega) = 0\}$. An admissible strategy X is an arbitrage if $\mathbf{P}(H_X) = 1$ and $\mathbf{P}(G_X) > 0$. But our theorem says that if $\mathbf{Q}(H_X) = 1$ then $\mathbf{Q}(G_X) = 1$. Since the two measures $\mathbf{Q}$ and $\mathbf{P}$ are equivalent they agree on set of probability one. In the context of the classical theory, i.e., $F^a(t,y,x) = F^b(t,y,x) = xS_t$, where S_t is the price of the stock at time t, then the existence of an equivalent measure $\mathbf{Q}$ such that S_t is a martingale rules out the existence of arbitrage strategies. This is equivalent to saying that for any transaction of size x, the process xS_t is a martingale under $\mathbf{Q}$, which is a specific case of Theorem 6.1.

Starting from Example 5.1, it is easy to build an example where we know the existence of an equivalent martingale measure. By letting S_t to be a geometric Brownian motion, that is $S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right)$ where W_t is a standard Brownian motion and μ and σ are positive constant. Then, we know that there exists an equivalent martingale measure defined by the Radon-Nikodym derivative $\frac{d\mathbf{Q}}{d\mathbf{P}} = e^{-\frac{\mu}{\sigma}W_t - \frac{1}{2}\frac{\mu^2}{\sigma^2}t}$.

To conclude this section we show that the restriction to the existence of an equivalent martingale measure is not general enough to consider all the cases.

Example 6.1. *Let F^a and F^b be defined as in* (3.1) *and satisfy properties **(P1)**, **(P2)** and **(P3)**. Moreover, we suppose that the probability space is a singleton, $\Omega = \{\omega\}$. In other words, F^a and F^b are deterministic. Finally, suppose that for all $x, y \geq 0$, $F^a(t,y,x)$ is strictly decreasing in t, $F^b(t,y,x)$ is strictly increasing in t and $F^a(T,y,x) > F^b(T,y,x)$.*

Then, we have,

$$\begin{aligned} V_T(X) &= F^b(T,B_T,X_T) - F^a(T,A_T,-X_t) \\ &\quad - \sum_{0<t\leq T}\left\{F^a(t,A_{t-},\Delta X_t^a) - F^b(t,B_{t-},\Delta X_t^b)\right\} \\ &\quad - \left(\int_0^T f^a(t,A_t)dX_t^{a,c} - \int_0^T f^b(t,B_t)dX_t^{b,c}\right) \\ &\leq f^b(T,0)X_T\mathbf{1}_{[0,\infty)}(X_T) + f^a(T,0)X_T\mathbf{1}_{(-\infty,0]} \\ &\quad - \sum_{0<t\leq T}\left\{f^a(t,0)\Delta X_t^a - f^b(t,0)\Delta X_t^b\right\} \\ &\quad - \left(\int_0^T f^a(t,0)dX_t^{a,c} - \int_0^T f^b(t,0)dX_t^{b,c}\right) \\ &= \Big(f^b(T,0) - f^a(T,0)\Big)\Big(X_T^a\mathbf{1}_{[0,\infty)} + X_T^b\mathbf{1}_{(-\infty,0]}\Big) \leq 0. \end{aligned} \tag{6.1}$$

As one sees, there is no arbitrage since the value of the portfolio is not greater than zero. Moreover, $F^a(t,0,x)$ is trivially a supermartingale, $F^b(t,0,x)$ is a submartingale, but there is no equivalent probability measure such that $F^a(t,0,x)$ and $F^b(t,0,x)$ are martingales.

7. Semimartingales with space parameters

In the classical theory, under the general assumption that the price process is a semimartingale, one can write the value of a portfolio as a stochastic integral with the price process as integrator. It is then possible to directly define the price process dynamic as a stochastic differential equation which allows for numerical simulations. Suppose that $\{S_t\}_{t\geq 0}$ is a semimartingale giving the price of the stock and that $\{X_t\}_{t\geq 0}$ is a predictable process giving the number of shares in the portfolio. Then the value of the portfolio at time T is given by $\int_0^T X_t dS_t$.

To get a similar expression in our setting we can use the theory of semimartingales with space parameters, (Kunita, 1990). This theory defines conditions such that for a process $\{F_t\}_{t\in[0,T]}$ with values in the space of continuous functions and a continuous semimartingale $\{X_t\}_{t\in[0,T]}$, the process $t\mapsto F_t(X_t)$ is a semimartingale. It also defines a generalisation of the Itô's formula in order to define the differential $dF_t(X_t)$.

To satisfy the conditions of the generalized Itô's formula, we restrict our trading strategy X to be a continuous predictable process with finite variation, while the processes A and B defining the level of impact to be continuous predictable semimartingales. The assumption of continuity for A and B could also be seen as a consequence of the continuity of the trading strategy X. Under this restriction, it is possible to show that the processes $t\mapsto f^a(t,A_t)$ and $t\mapsto f^b(t,B_t)$ are continuous semimartingales and to use these processes as integrators. Then, with the generalized Itô's formula, see Appendix B, we can use stochastic differential equations to describe explicitly the dynamic of $f^a(t,A_t)$ and $f^b(t,B_t)$.

First we use the fact that $f^a(t,A_t)$ and $f^b(t,B_t)$ are semimartingales and use the integration by parts formula to write the cash flow of the portfolio as

$$\begin{aligned}
&\int_0^T f^a(t,A_t)dX_t^a - \int_0^T f^b(t,B_t)dX_t^b \qquad (7.1)\\
=\;& f^a(T,A_T)X_T^a - f^b(T,B_T)X_T^b - \left(\int_0^T X_t^a df^a(t,A_t) - \int_0^T X_t^b df^b(t,B_t)\right).
\end{aligned}$$

Recall that X is continuous so that $\Delta X_t^a = \Delta X_t^b = 0$ for all $t\geq 0$ a.s. Then, by using the generalized Itô's formula to write the dynamic of $f^a(t,A_t)$ and $f^b(t,B_t)$

one gets that (7.1) is equal to

$$f^a(T,A_T)X^a_T - \int_0^T X^a_t f^a(dt,A_t) - \int_0^T X^a_t \frac{\partial}{\partial y} f^a(t,A_t)dA_t$$
$$-f^b(T,B_T)X^b_T + \int_0^T X^b_t f^b(dt,B_t) + \int_0^T X^b_t \frac{\partial}{\partial y} f^b(t,B_t)dB_t \qquad (7.2)$$

To our knowledge, the theory for semimartingales with space parameters and the generalized Itô's formula has not been extended yet for processes with jumps. It is beyond the scope of this paper to develop this theory here and we defer this task for a future paper. But it is reasonable to think that the theory can be adapted for the full generality of our model.

8. Complete example and calibration

Nowadays, it is possible to have data for the limit order book of different stocks with all the informations about the limit orders that are submitted. In order to bring the literature on limit order book models at a more practical level it is desirable to have realistic models which can be estimated. In this section, we present a complete example with an estimation method and suggest possible extensions. We also use this example to apply formula (7.2) from the preceding section.

Let $S^a_t = S^a_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}$ and $S^b_t = S^b_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}$ be two geometric Brownian motions where W is a standard Brownian motion where $0 < S^b_0 \leq S^a_0$ and μ, σ are positive constants. We define

$$\begin{aligned} G^a(t,x) &= S^a_t(-1+(1+x)^\alpha), \\ G^b(t,x) &= S^b_t(-1+(1+x)^\beta) \end{aligned}$$

where $\alpha > 1$ and $\beta \in (0,1)$. The function $G^a(t,x)$ (resp. $G^b(t,x)$) gives the cost (resp. profit) of buying (resp selling) x shares at time $t \geq 0$ when no transactions occurred before t. The parameters α and β define the shape of the order book. By taking $\alpha > 1$ and $\beta \in (0,1)$ we get that the ask part of the order book is convex and the bid part is concave. The processes S^a_t and S^b_t represent the uncertainty of the market and by taking $0 < S^b_0 \leq S^a_0$ we get that the bid/ask spread is non-negative.

Next, we define the structure of the order book including the impact of past transactions.

$$\begin{aligned} F^a(t,y,x) &= G^a(t,x+y) - G^a(t,y), \\ F^b(t,y,x) &= G^b(t,x+y) - G^b(t,y). \end{aligned}$$

From the definition of F^a and F^b, one sees that the effect of the impact, y, is to move higher in the structure of the order book. Since $G^a(t,x)$ is increasing and convex, we have that $G^a(t,x+y_2) - G^a(t,y_2) \geq G^a(t,x+y_1) - G^a(t,y_1)$ for all $0 \leq y_1 \leq y_2$ so that the cost of x shares increases when the impact increases. On the other hand, since $G^b(t,x)$ is increasing and concave, $G^b(t,x+y_2)-G^b(t,y_2) \leq G^b(t,x+y_1)-G^b(t,y_1)$ for al $0 \leq y_1 \leq y_2$, so that the profit from selling x shares decreases when the impact increases.

For the state variables we take

$$\begin{aligned} A_t &= X_t^a - \int_0^t g(A_s)ds, \\ B_t &= X_t^b - \int_0^t h(B_s)ds. \end{aligned}$$

where g and h are non-negative functions with $g(0) = h(0) = 0$. We see that A_t is an increasing function of X_t^a, the number of share bought, and B_t is an increasing function of X_t^b, the number shares sold. Also, if $X_{t_1+\delta}^a = X_{t_1}^a$ for some $\delta > 0$ and $t_1 > 0$, then one finds that $dA_t = -g(A_t)dt$ for all $t \in (t_1, t_1+\delta)$. Since g is non-negative, A_t is non-increasing on $(t_1, t_1+\delta)$, that is the ask part of the limit order book is recovering from the impact of the past market orders. The process B displays a similar behaviour. Those state variables were taken from Predoiu et al. (2011).

If we assume that X is continuous, we can apply equation (7.2) to get

$$\begin{aligned} & S_T X_T^a (1+A_T)^{\alpha-1} - \int_0^T X_t^a S_t (1+A_t)^{\alpha-1} \left(\mu - \frac{(\alpha-1)}{1+A_t} g(A_t) \right) dt \\ & - \int_0^T X_t^a S_t (1+A_t)^{\alpha-1} \left\{ \sigma dW_t - \frac{(\alpha-1)}{1+A_t} dX_t^a \right\} \\ - \; & \left[S_T X_T^b (1+B_T)^{\beta-1} - \int_0^T X_t^b S_t (1+B_t)^{\beta-1} \left(\mu - \frac{(\beta-1)}{1+B_t} h(B_t) \right) dt \right. \\ & \left. - \int_0^T X_t^b S_t (1+B_t)^{\beta-1} \left(\sigma dW_t - \frac{(\beta-1)}{1+B_t} dX_t^b \right) \right]. \end{aligned}$$

For practical considerations we propose a simple fitting method for this model. It is based on observed values of the order book and we suppose that we observed realizations of $F^a(t,0,x)$ and $F^b(t,0,x)$ for each observation time t. The processes A and B are set to zero for the past observations since we suppose that our investor did not trade during this period.

For the first step, we calibrate the structure of the order book independently from the dynamic, that is we use the observed value of the limit order book at $t=0$ to determine the constants S_0^a, S_0^b, α and β. Let $\{(x_i^a, p_i^a); i = 1, \ldots, m^a\}$ and $\{(x_i^b, p_i^b); i = 1, \ldots, m^b\}$ be the entries in the limit order book at time $t = 0$.

The variables x_i^a and (resp. x_i^b) are the number of shares offered (resp. asked) for a price p_i^a (resp. p_i^b). Then, the constants S_0^a, S_0^b, α and β are determined by solving

$$\inf_{0<S_0^b<S_0^a,\alpha>1,\beta\in(0,1)} \left\{ \sum_{i=1}^{m^a} \left(\frac{F^a(0,0,x_i^a)}{x_i^a} - P_i^a \right)^2 + \sum_{i=1}^{m^b} \left(\frac{F^b(0,0,x_i^b)}{x_i^b} - P_i^b \right)^2 \right\}.$$

where $P_i^a = \frac{1}{x_1^a+\cdots+x_i^a} \sum_{j=1}^i x_i^a p_i^a$ and $P_i^b = \frac{1}{x_1^b+\cdots+x_i^b} \sum_{j=1}^i x_i^b p_i^b$. For instance, if one wants to calibrate the model on the limit order book of Amazon Inc. given in Table 3.1 then $(x_1^a, P_1^a) = (100, 227.64)$, $(x_2^a, P_2^a) = (200, 227.645)$, $(x_3^a, P_3^a) = (500, 227.66)$ and so on. A similar approach works for the bid part. Applying this method leads to the following values : $S_0^a = 227.58$, $\alpha = 1.0002$, $S_0^b = 227.52$ and $\beta = 0.99972$. Since Amazon is a highly traded stock, the limit order book has a lot of *depth* and it is not surprising that the convexity and concavity parameters α and β are close to one.

The next step is to estimate the constants μ and σ. Let $\delta_t := F^a(t,0,1) - F^b(t,0,1)$, then $\delta_t = \delta_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}$. On sees that the process δ_t is the bid-ask spread given by our model. Suppose that $\{\tilde{\delta}_i; i = 1, \ldots n\}$ are past observations of the bid-ask spread with constant time interval $\Delta_t > 0$ between the observations. Then, if we define $\rho_i = \frac{\tilde{\delta}_{i+1} - \tilde{\delta}_i}{\tilde{\delta}_i}$ for $i = 1, \ldots, n-1$ then one can set

$$\mu = \frac{1}{\Delta_t(n-1)} \sum_{i=2}^{n} \rho_i \qquad \text{and} \qquad \sigma = \sqrt{\frac{1}{\Delta_t(n-2)} \sum_{i=1}^{n-1} (\rho_i - \mu)^2}.$$

Remark 8.1. *The goal of this example is to present a model which is simple enough to be tractable but yet non-trivial. The same is applicable for our choice of calibration method. In our case we chose to calibrate the structure of the limit order book and the dynamic in two independent steps. One could choose to use observed values of the limit order book at different time in order to calibrate the structure as well as the dynamic of the order book simultaneously. In any case, the precision of some aspect of the model will be favored over another. Also, the choice of a geometric Brownian motion for the bid-ask spread is not the most appropriate. A mean reverting process such as a CIR process could be a better alternative.*

Remains the question of determining h and g, which give the speed at which the order book refills. To see this, suppose that for some $t_0 > 0$, $X_t^a = X_{t_0}^a > 0$ for $t > t_0$. Then, $dA_t = -g(A_t)dt$, that is A_t is decreasing for $t > t_0$ since $A_{t_0} > 0$ and $g(x) > 0$ if $x > 0$, (the same applies for B). To calibrate g and h we use that $g(x)dt$ (resp. $h(x)dt$) is the number of new shares submitted over the interval of time dt knowing x shares were removed from the ask part (resp. bid part) of the limit order book.

To ease the description of our calibration method we only give the details for the ask part, that is, the function g. The calibration of the bid part follows the same methodology.

(1) Suppose that the observed values of the order book cover the time interval $[0,T]$ and let $t_0 = 0 < t_1 < \cdots < t_n = T$ be the observation times with $t_i - t_{i-1} = \Delta_t$ for $i = 1,\ldots,n$.

(2) At each time interval $[t_i, t_{i+1})$, we note $\{(x_{ij}, p_{ij}); j = 1, ..., m_i\}$ the new limit orders submitted, where x_{ij} is the number of shares at price p_{ij}, and we define $\tilde{P}_i = \min_{k=1,\ldots,m_j} p_{ik}$ to be the best ask price amongst the new limit orders.

(3) Let $n_j = \sum_{k=1}^{j} x_{ij}$, we define $\tilde{q}_{ij} = \frac{1}{n_j}\sum_{k=1}^{j} x_{ik}\frac{p_{ik}}{\tilde{P}_i}$, for $j = 1,\ldots,m_i$. The variable $\tilde{q}_{ij}$ is the relative price per share of buying all the shares in the first j entries in the limit order book.

(4) Let $s_j = 1 + j\Delta_s$, for $j = 0,\ldots,M$ for some positive constant Δ_s. The set of values $\{s_j\}_{j=0,\ldots,M}$ is going to be used as a discretization of the space of relative price per share.

(5) For each $i = 1,\ldots,n$, we compute the number of new shares available during the interval $[t_i, t_{i+1})$ at a relative price per share in $[s_{j-1}, s_j)$ for $j = 1,\ldots,m$ and we note this quantity N_{ij}. That is, $N_{ij} = \sum_{k=1}^{m_i} x_{ik}\mathbf{1}_{[s_{j-1},s_j)}(\tilde{q}_{ik})$.

(6) Let η_j be such that $\frac{F^a(t,0,\eta_j)}{\eta_j F^a(t,0,1)} = s_j$ for $j = 1,\ldots,M$. Note that $\frac{F^a(t,0,x)}{F^a(t,0,1)}$ is independent of t.

(7) Finally, let $\bar{N}_j = \sum_{i=0}^{n-1} \frac{N_{ij}}{(n-1)\Delta_t}$ for $j = 1,\ldots,m$, then the function g should be chosen such that the curve $\{(x, g(x)), x \geq 0\}$ approximates the set $\{(\eta_j, \bar{N}_j); j = 1,\ldots,m\}$. A possible choice is to take $g(x)$ as a polynomial and to carry a least squared approximation.

9. Other applications and further extensions.

For the construction of our model we referred to a large investor and small investors. However, the mechanism of trading we modeled applies to all investors who trade using market orders, no matter the size of the transactions. For smaller transactions, the impact and the effect of the limit order book structure would be smaller and probably negligible in many cases. On the other hand, in the case of high speed trading, it is possible that the cumulative impact of many small transactions would become non-negligible. For this situation, considering the structure of the limit order book might help to develop better trading strategies.

The model we presented can also include transaction fees. For instance, let S_t, $t \geq 0$ be a continuous non-negative semimartingale which represents the value of a stock. Suppose one wants to apply proportional cost of transactions, then we can define $F^a(t,y,x) = S_t(1+y)x$ and $F^b(t,y,x) = S_t(1-y)x$ for $x \geq 0$ and zero otherwise. We set the state variables $A_t = B_t = \alpha$ for all $t \geq 0$ with $\alpha \in (0,1)$. In this case, a transaction of size $x > 0$ incurs a cost (resp. profit) of $F^a(t,\alpha,x) = S_t(1+\alpha)x$ (resp. $F^b(t,\alpha,x) = S_t(1-\alpha)x$). Using the same notation as before, the value of the portfolio is given by

$$\begin{aligned} V_T(X) &= (X_T)^+(1-\alpha)S_T - (X_T)^-(1+\alpha)S_T \\ &\quad - \left(\int_0^T (1+\alpha)S_t dX_t^a - \int_0^T (1-\alpha)S_t dX_t^b \right). \end{aligned}$$

As mentioned at the beginning of Section 4, it is possible to include interest rate in the model. Suppose that the instantaneous rate is given by a positive predictable process $\{r_t\}_{t\in[0,T]}$. Then, it is only a matter of applying the interest rate to each transaction so that the cash flow of the portfolio becomes

$$\begin{aligned} &\sum_{0<t\leq T} e^{\int_t^T r_s ds} \left\{ F^a(t, A_{t-}, \Delta X_t^a) - F^b(t, B_{t-}, \Delta X_t^b) \right\} \\ &+ \int_0^T f^a(t, A_{t-}) e^{\int_t^T r_s ds} dX_t^{a,c} - \int_0^T f^b(t, B_{t-}) e^{\int_t^T r_s ds} dX_t^{b,c}. \end{aligned}$$

9.1. Options pricing

Another matter of importance is the question of options pricing. Here we want to highlight some of the new problems that arise in our framework. The study of these questions are set aside for a future work.

First, we suppose the large investor is long on an European call option with maturity $T > 0$, strike K and there is physical delivery. In this situation, to exercise the option, the large investor has to consider the value of liquidating the stock. Hence, he exercises the option if and only if $F^b(T, B_T, 1) > K$. On the other hand, he exercises a put option if and only if $F^a(T, A_T, 1) < K$. Surprisingly, if the large investor is long on $n > 0$ call options, then the decision to exercise is on a number of positions n^* where

$$n^* = \arg \max_{m\in[0,n]} \left\{ \frac{F^b(T, B_T, m)}{m} > K \right\}.$$

Finally, if the large investor wants to hedge a short position on a call option, then, the terminal value he has to hedge is $\{K - F^a(T, A_T, 1)\} \mathbf{1}_{[K,0)} \left(F^b(T, B_T, 1) \right)$. That is, if the option is exercise, he has to cover the difference between the strike and the cost of buying one share of the stock. One should also consider the possibility for the large investor to manipulate prices using the impact of his transactions. This problem was considered in Jarrow (1992) and Jarrow (1994).

10. Conclusion

In this paper, we used the mechanism of market orders and the limit order book structure to build a general model for price impact. The novelty of our approach was to use random functions to directly model the cost of transactions instead of the value of the asset and the use of state variables to capture the non-markovian property of the impact. It turns out that this approach generalizes classical models in perfect liquidity and some models under illiquidity assumption, the model of Cetin et al. (2004) for instance. By deriving the value of the portfolio for continuous time trading we showed that the impact on prices cannot be avoided. An important result is that the usual relationship between the existence of some equivalent martingale probability measure and the absence of arbitrage is not general enough in our setting and one as to consider supermartingales and submartingales. By using the Itô's formula for semimartingales with space parameters, we also wrote the cost of the portfolio as a stochastic integral respect to the value of the stock, alike the classical approach, which allows to use stochastic differential equations and eventually for numerical simulations. Finally, we provided a complete example with an estimation method for the limit order book as well as the arriving of new limit orders.

A. Proofs

A.1. Cash flow for continuous time trading strategies.

The following proposition is the mathematical result which allows to define the cost of the portfolio for continuous time trading strategies

Proposition A.1. *Let $T>0$ and $\{\pi^n\}_{n=1,2,...}$ be a sequence of increasing partitions of $[0,T]$, i.e., $\pi^n = \{0 = t_0^n \leq t_1^n \leq ... \leq t_n^n = T\} \subset \pi^{n+1}$, and $\max_{i=1,...,n}\{t_i^n - t_{i-1}^n\} \to 0$. Let $F(t,y,x)$, $(y,x) \in [0,\infty)\times[0,\infty)$ be a family of continuous processes and suppose that $F(t,\cdot,\cdot)$ is in $C^{2,1}([0,\infty)\times[0,\infty))$ and that $F(t,y,0)\equiv 0$. Let Y be a càdlàg process with finite variation and Z a càdlàg, non-decreasing process with $Z_0 = 0$. We suppose that $\Delta Z_s = 0$ if and only if $\Delta Y_s = 0$, where $\Delta V_s = V_s - V_{s-}$ for any process V. Then,*

$$\sum_{i=1}^{n} F(t_{i-1}^n, Y_{t_{i-1}^n}, \Delta Z_{t_i^n}) \to \int_0^T f(t, Y_{t-}, 0)dZ_t^c + \sum_{0<t\leq T} F(t, Y_{t-}, \Delta Z_t) \text{ a.s.} \quad \text{(A.1)}$$

where $f(t,y,x) = \lim_{\epsilon \searrow 0} \frac{F(t,y,x+\epsilon)-F(t,y,x)}{\epsilon}$.

Proof A.1. *For any process* V *we define the set of discontinuity points* $\mathcal{D}(V) = \{t \in [0,T]; \Delta V_t > 0\}$ *where* $V_{0-} = 0$. *Without loss of generality we can suppose that there exists* n^* *such that* $\mathcal{D}(Z) \subset \pi^n$ *for all* $n \geq n^*$. *From our hypothesis we also have that* $\mathcal{D}(Y) \subset \pi^n$.

For all $n \geq n^*$ *suppose that* $v \in \pi^n \bigcap \mathcal{D}(Z)$ *and* $u^n = \max_{i=1,\ldots,n} \{t_i^n \in \pi^n; t_i^n < v\}$, *so that* u^n *is the partition point preceding* v. *Then, we see that*

$$\lim_{n\to\infty} F(u^n, Y_{u^n}, Z_v - Z_{u^n}) = F(v-, Y_{v-}, \Delta Z_v). \tag{A.2}$$

Since F *is a continuous process we also have that* $F(v-, Y_{v-}, \Delta Z_v) = F(v, Y_{v-}, \Delta Z_v)$. *Equation* (A.2) *settle the case for discontinuity points.*

To treat continuity points we define $\phi^n = \pi^n \bigcap \mathcal{D}(Z)^c$, *so that the sequence of partitions* ϕ^n *does not contain the discontinuity points of* Z. *From the Mean Value Theorem we have*

$$\begin{aligned}\sum_{t_i^n, t_{i-1}^n \in \phi^n} F(t_{i-1}^n, Y_{t_{i-1}^n}, \delta Z_{t_i^n}) &= \sum_{t_i^n, t_{i-1}^n \in \phi^n} f(t_{i-1}^n, Y_{t_{i-1}^n}, 0)\delta Z_{t_i^n} \\ &+ \sum_{t_i^n, t_{i-1}^n \in \phi^n} \left\{ f(t_{i-1}^n, Y_{t_{i-1}^n}, \xi_i^n) - f(t_{i-1}^n, Y_{t_{i-1}^n}, 0) \right\} \delta Z_{t_i^n}\end{aligned} \tag{A.3}$$

where, $\xi_i^n \in [0, \delta Z_{t_i^n})$. *Recall that* $\delta Z_{t_i^n} = Z_{t_i^n} - Z_{t_{i-1}^n}$. *Let* I_n *be the sum in* (A.3) *and denote* $\bar{Y} = \sup_{t\in[0,T]}\{Y_t\}$ *and* $\rho^n = \max_{i=1,\ldots,n} \delta Z_{t_i^n}$. *Here we note that* $\bar{Y}$ *is almost surely finite, since* Y *is a finite variation process. Then we have*

$$|I_n| \leq \sup_{t\in[0,T], Y\in[0,\bar{Y}), x\in[0,\rho^n)} |f(t,Y,x) - f(t,Y,0)| \, Z_T \to 0$$

from the continuity of f *and since* $\rho^n \to 0$. *Now, since* Z_t *defines a Lebesgue-Stieltjes measure on* $[0,T]$, *we can define the integral using path-by-path convergence,*

$$\sum_{t_i^n, t_{i-1}^n \in \phi^n} f(t_{i-1}^n, Y_{t_{i-1}^n}, 0)\delta Z_{t_i^n} \to \int_0^T f(t, Y_{t-}, 0) dZ_t^c \ \textit{a.s.}$$

Finally, we can get the result by adding the discontinuity points in the partition,

$$\lim_{n\to\infty} \sum_{t_i^n, t_{i-1}^n \in \pi^n} F(t_{i-1}^n, Y_{t_{i-1}^n}, Z_{t_i^n} - Z_{t_{i-1}^n}) = \int_0^T f(t, Y_{t-}, 0) dZ_t^c + \sum_{0<t\leq T} F(t, Y_{t-1}, \Delta Z_t).$$

A.2. Proof of Theorem 6.1

The idea of the proof is to show that, for an admissible strategy X, $E^{\mathbf{Q}}[V_T(X)] \leq 0$. Consequently, there is no arbitrage since $\mathbf{Q}\{V_T(X) \geq 0\} = 1$ implies that $V_T(X) \equiv 0$.

The proof is divided in two major steps. At first, we show the result for the case where $F^a(t,0,x)$ and $F^b(t,0,x)$ are $\mathbf{Q}-$martingales for all $x \geq 0$. The general case with $F^a(t,0,x)$ and $F^b(t,0,x)$ being respectively supermartingales and submartingales under $\mathbf{Q}$ will follow from the Doob-Meyer decomposition.

Suppose that $F^a(t,0,x)$ and $F^b(t,0,x)$ are $\mathbf{Q}-$martingales for all $x \geq 0$, then we have that $f^a(t,0)$ and $f^b(t,0)$ are also $\mathbf{Q}$-martingales. To see this, one note that for $x \in (0,1)$, $\frac{F^a(t,0,x)}{x} \leq F^a(t,0,1)$ so that by the dominated convergence theorem $f^a(t,0)$ is a $\mathbf{Q}-$martingale. Again, we can apply the dominated convergence theorem to show that $f^b(t,0)$ is a $\mathbf{Q}-$martingale since $\frac{F^b(t,0,x)}{x} \leq f^a(t,0)$.

By using the integration by parts for stochastic integrals (Protter, 2004) one has

$$\int_0^T f^a(t,0)dX_t^a = f^a(T,0)X_T^a - \int_0^T X_t^a df^a(t,0) \tag{A.4}$$

and similarly for $\int_0^T f^b(t,0)dX_t^b$. Since $f^a(t,0)$ and $f^b(t,0)$ are $\mathbf{Q}-$martingales then the processes $\int_0^t X_s^a df^a(s,0)$ and $\int_0^t X_s^b df^b(s,0)$ are local martingales. Let $\{\alpha^n\}_{n=0,1,\dots}$ and $\{\beta^n\}_{n=0,1,\dots}$ be two sequences of stopping times which respectively reduce the local-martingales $\int_0^t X_s^a df^a(s,0)$ and $\int_0^t X_s^b df^b(s,0)$. Then, by the bounded convergence theorem we have

$$E^{\mathbf{Q}}\left[\int_0^T X_t^a df^a(t,0)\right] = \lim_{n\to\infty} E^{\mathbf{Q}}\left[\int_0^{T\wedge\alpha^n} X_t^a df^a(t,0)\right] = 0, \tag{A.5}$$

$$E^{\mathbf{Q}}\left[\int_0^T X_t^b df^b(t,0)\right] = \lim_{n\to\infty} E^{\mathbf{Q}}\left[\int_0^{T\wedge\beta^n} X_t^b df^b(t,0)\right] = 0. \tag{A.6}$$

To justify the use of the bounded convergence theorem above, we see from equation (A.4) that $\int_0^T X_t^a df^a(t,0) \leq f^a(T,0)X_T^a$. Moreover, $E^{\mathbf{Q}}[f^a(T,0)X_T^a] < \infty$ by the admissibility condition of X^a. Similarly, we have that $\int_0^T X_t^b f^b(t,0) \leq f^b(T,0)X_T^b$ and that $E^{\mathbf{Q}}[f^b(T,0)X_T^b] < \infty$.

Now we have the following lower bound for the cash flow of the portfolio. Let

$$\begin{aligned} C_T(X) &= \sum_{0<t\leq T}\left\{F^a(t,A_{t-},\Delta X_t^a) - F^b(t,B_{t-},\Delta X_t^b)\right\} \\ &\quad + \int_0^T f^a(t,A_{t-})dX_t^{a,c} - \int_0^T f^b(t,B_{t-})dX_t^{b,c}, \end{aligned} \tag{A.7}$$

be the cash flow of the portfolio. One finds that

$$\begin{aligned} C_T(X) &\geq \sum_{0<t\leq T}\left\{f^a(t,0)\Delta X_t^a - f^b(t,0)\Delta X_t^b\right\} \\ &\quad + \int_0^T f^a(t,0)dX_t^{a,c} - \int_0^T f^b(t,0)dX_t^{b,c} \end{aligned}$$

$$= \int_0^T f^a(t,0)dX_t^a - \int_0^T f^b(t,0)dX_t^b$$

where the superscript c stands for the continuous part of a process. To obtain the inequality one only has to recall property **P3**. First, we use that $F^a(t,y,x)$ is increasing in y and $F^b(t,y,x)$ is decreasing in y. Then we use that $F^a(t,0,x) \geq f^a(t,0)x$ and $F^b(t,0,x) \leq f^b(t,0)x$ by the convexity and concavity properties. Taking the expectation one finds

$$\begin{aligned} E^{\mathbf{Q}}\left[C_T(X)\right] &\geq E^{\mathbf{Q}}\left[\int_0^T f^a(t,0)dX_t^a - \int_0^T f^b(t,0)dX_t^b\right] \\ &= E^{\mathbf{Q}}\left[f^a(T,0)X_T^a - f^b(T,0)X_T^b\right]. \end{aligned} \tag{A.8}$$

We get the last equality by applying the integration by parts formula and the results in equations (A.5) and (A.6).

Finally, we can show that $E^{\mathbf{Q}}\left[V_T(X)\right] \leq 0$.

$$\begin{aligned} E^{\mathbf{Q}}[V_T(X)] &= E^{\mathbf{Q}}\left[F^b(T,B_T,X_T) - F^a(T,A_T,-X_T)\right] - E^{\mathbf{Q}}\left[C_T(X)\right] \\ &\leq E^{\mathbf{Q}}\left[f^b(T,0)X_T\mathbf{1}_{[0,\infty)}(X_T) + f^a(T,0)X_T\mathbf{1}_{(-\infty,0]}(X_T)\right] \\ &\quad - E^{\mathbf{Q}}\left[C_T(X)\right] \\ &\leq E^{\mathbf{Q}}\left[f^b(T,0)X_T\mathbf{1}_{[0,\infty)}(X_T) + f^a(T,0)X_T\mathbf{1}_{(-\infty,0]}(X_T)\right] \\ &\quad - E^{\mathbf{Q}}\left[f^a(T,0)X_T^a - f^b(T,0)X_T^b\right] \\ &= E^{\mathbf{Q}}\left[\left(X_T^a\mathbf{1}_{[0,\infty)}(X_T) + X_T^b\mathbf{1}_{(-\infty,0]}(X_T)\right)\left(f^b(T,0) - f^a(T,0)\right)\right] \leq 0. \end{aligned}$$

Which shows the result for the case where $F^a(t,0,x)$ and $F^b(t,0,x)$ are $\mathbf{Q}-$martingales.

For the general result, we suppose that $F^a(t,0,x)$ is a $\mathbf{Q}-$supermartingale and $F^b(t,0,x)$ is a $\mathbf{Q}-$submartingale. Consequently, we have that $f^a(t,0)$ and $f^b(t,0)$ are respectively supermartingale and submartingale under $\mathbf{Q}$. Using the Doob-Meyer decomposition theorem (Protter, 2004) we can write

$$\begin{aligned} f^a(t,0) &= M_t^a - N_t^a \\ f^b(t,0) &= M_t^b + N_t^b \end{aligned}$$

where M^a and M^b are martingales and N^a and N^b are increasing predictable processes. With this decomposition we find

$$\begin{aligned} E^{\mathbf{Q}}\left[C_T(X)\right] &\geq E^{\mathbf{Q}}\left[\int_0^T f^a(t,0)dX_t^a - \int_0^T f^b(t,0)dX_t^b\right] \\ &= E^{\mathbf{Q}}\left[f^a(T,0)X_T^a - \int_0^T X_t^a dM_t^a + \int_0^T X_t^a dN_t^a\right] \end{aligned}$$

$$
\begin{aligned}
&-E^{\mathbf{Q}}\left[f^b(T,0)X_T^b - \int_0^T X_t^b dM_t^b - \int_0^T X_t^b dN_t^b\right] \\
\geq\ & E^{\mathbf{Q}}\left[f^a(T,0)X_T^a - f^b(T,0)X_T^b + \int_0^T X_t^a dN_t^a + \int_0^T X_t^b dN_t^b\right] \\
\geq\ & E^{\mathbf{Q}}\left[f^a(T,0)X_T^a - f^b(T,0)X_T^b\right]
\end{aligned}
$$

since $\int_0^T X_t^a dN_t^a$ and $\int_0^T X_t^b dN_t^b$ are non-negative. One note that this lower bound for $C_T(X)$ is the same than (A.8) so that the rest of the proof remains valid for the general case.

B. Generalized Itô's formula

Let $C^m([0,\infty)^n)$ be the space of m-times continuously differentiable functions defined on $[0,\infty)^n$. For every compact $K \in [0,\infty)^n$, we set

$$||f||_{m:K} = \sup_{x\in K} \frac{|f(x)|}{1+|x|} + \sum_{1\leq|\alpha|\leq m} \sup_{x\in K} |D^\alpha f(x)|$$

where $\alpha = (\alpha_1, ..., \alpha_n)$ is a vector of n non-negative integers and D^α the differential operators $\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1},...,\partial x_n^{\alpha_n}}$. A continuous function $f(t,x)$, $t \in [0,T]$ and $x \in [0,\infty)^n$ is said to belong to $\mathbb{C}^m([0,\infty)^n)$ if for every t, $f(t,\cdot) \in C^m([0,\infty)^n)$ and $||f(t,\cdot)||_{m:K}$ is integrable respect to t on $[0,T]$ for any compact K.

Let $F(t,x)$ be a family of semimartingales with $x \in [0,\infty)$ and define $G(t,x,y) := \langle F(t,x), F(t,y)\rangle$ the quadratic variation of $F(t,x)$ and $F(t,y)$. There exists a process $g(t,x,y)$ and continuous increasing process such that $G(t,x,y) = \int_0^t g(s,x,y)dG_s$. The pair $(g(t,x,y), G_t)$ is called the *local characteristic* of $F(t,x)$.

Theorem B.1. *(Generalized Itô's formula) Let $F(t,x)$, $x \in [0,\infty)$ be a continuous $\mathbb{C}^2([0,\infty))$-process and a continuous $\mathbb{C}^1([0,\infty))$-semimartingale with $g(t,y,x)$ be a continuous $C^1([0,\infty)^2)$-process such that $||a(t,x,y)||_{1:K}$ is in $L^1(G_t)$ for every compact $K \in [0,\infty)^2$. Suppose that Z_t is a continuous semimartingale with value in $[0,\infty)$. Then $F(t,Z_t)$ is a continuous semimartingale and*

$$
\begin{aligned}
F(t,Z_t) - F(0,Z_0) = \int_0^t F(ds,Z_s) + \int_0^t \frac{\partial}{\partial x}F(s,Z_s)dZ_s + \frac{1}{2}\int_0^t \frac{\partial^2 F}{\partial x^2}(s,Z_s)d\langle Z_s\rangle \\
+ \left\langle \int_0^t \frac{\partial F}{\partial x}(ds,Z_s), Z_s \right\rangle
\end{aligned}
$$

The proof of this theorem can be found in Kunita (1990).

Chapitre 4

PRICING EUROPEAN OPTIONS IN A DISCRETE TIME MODEL FOR THE LIMIT ORDER BOOK

Résumé

Dans cet article, nous construisons un modèle pour la structure du carnet d'ordres limites de sorte que le prix par action dépend de la taille des transactions. Nous déduisons la valeur du portefeuille lorsque l'investisseur utilise des ordres de marché et un compte bancaire avec des taux d'intérêts différents pour prêter et emprunter. Dans ce contexte de marché, nous déduisons des conditions permettant d'éliminer les possibilités d'arbitrages et nous résolvons les problèmes de tarification et de couverture pour les options européennes de vente et d'achat lorsque l'échéance est d'une période et la livraison est physique. En utilisant la méthode d'optimisation duale nous montrons que le prix des options européennes peut être écrit comme un problème d'optimisation sur un certain ensemble de mesures de probabilités.

Abstract

In this paper we build a discrete time model for the structure of the limit order book, so that the price per share depends on the size of the transaction. We deduce the value of a portfolio when the investor trades using market orders and a bank account with different interest rates for lending and borrowing. In this setting, we deduce conditions to rule out arbitrage and solve the problem of pricing and hedging an European call and put option with maturity one and physical delivery. By using primal-dual optimization we show that the price of European options can be written as an optimization problem over some set of probability measures.

1. Introduction

One of the most important topic in mathematical finance is the problem of pricing and hedging contingent claims. Between the seminal work of Black and Scholes (1973) and the early 1990's, most of the research on this topic were done under the assumption that the market is perfectly liquid. Since then, researchers started to study this problem for illiquid markets and many approaches have been tried. For instance, Liu and Yong (2005) proposed an extension of the Black and Scholes model where the asset price depends on the size of the transaction. The solution to the pricing problem is a non-linear partial differential equation. Cetin et al. (2004) built a supply curve for stocks and show a version of the second fundamental theorem of asset pricing and that a smoothed version of delta hedging can be used to hedge contingent claims, while Bank and Baum (2004) discuss the case of superreplication also in a similar model. Another solution for the hedging problem comes from forward backward stochastic differential equations as presented in Cvitanić and Ma (1996), or backward stochastic differential equations as in Roch (2011). In Cetin et al. (2010) they study the superreplication problem of a contingent claim depending on the fundamental value of the stock. The two last papers use an extension of the model in Cetin et al. (2004).

All these approaches consider continuous time models and the solutions to the hedging and pricing problem that arise are rarely explicit and difficult to analyze. They also define an *a priori* value function for the contingent claims which might not reflect the actual value that the seller or the buyer of the contingent claim would define.

The purpose of this paper is to present a simplified framework, which still keeps the characteristics of an illiquid market, such that the solution of the problem of hedging and pricing contingent claims can be found using a more fundamental approach. This way, the solutions we obtain present more explicitly the relationship between the different inputs, and thus, it provides a better understanding of the effect of illiquidity on the price and the hedging strategy. For one period European call and put, we derive an explicit expression for the price by taking alternatively the point of view of the seller and the buyer which leads to different hedging portfolios. The expressions for the prices that we find show a completely different behaviour than in a perfectly liquid market. Since our model includes the standard binomial model as a specific case, it makes for interesting comparisons. Although some aspects of the model are simplistic, our proposed model considers bid-ask prices as well as different interest rates for lending and borrowing, thus creating a somewhat realistic market.

In Section 2.3, we give a new result in the form of a no-arbitrage theorem. For perfectly liquid markets, it is well known that the existence of an equivalent martingale measure is equivalent to the absence of arbitrage, see Harrison and Kreps (1979). In the literature about illiquid markets, most results state that the existence of an equivalent martingale (resp. submartingale or supermartingale) measure for the marginal price is a sufficient condition to rule out arbitrage, e.g. Cetin et al. (2004), Bank and Baum (2004), Roch (2011), Roch and Soner (2013), Simard (2014). In Theorems 2.1 and 2.2 we provide new conditions which make the market free of arbitrage.

The theoretical setting is presented in the next section ; in Section 2.1 we define the market model, and then compute the value of a general portfolio in Section 2.2, while some conditions of no-arbitrage are stated in Section 2.3. In Section 3 we consider the value call and put options, together with hedging portfolios for the general model. These results are specialized for the one-period model in Section 4, where the condition of no-arbitrage are stated in Section 4.1, while the price of European call and put options are given in Sections 4.2 and 4.3. Also, in Section 4.4, we discuss the results and we present an explicit example along with numerical results in Section 4.5. Finally, in Section 5, we explore another approach to solve the problem of pricing European options through primal-dual optimization.

2. Theoretical setting

In what follows we first introduce the model, then we describe some properties.

2.1. The Model

We assume that the market is composed of two types of investors. First, we suppose that there exists a large number of investors trading using limit orders. These investors are liquidity providers and are responsible for the uncertainty. The second type of investors are the buyers and the sellers of European call or put options and they aim to hedge their position with market orders. As a result, the size of their transactions has an influence on the value per share of the traded asset. More specifically, when buying the stock, the cost per share increases with the size of the transaction while, for selling the stock, the value per share decreases with the size of the transaction. Finally, transactions are held at discrete time and we suppose there is enough liquidity to insure that the impact of market orders have no lasting effect.

Let $(\Omega, \mathbb{F} = \{\mathcal{F}_t\}_{t=0,1,\ldots,T}, P)$ be a filtered probability space. The evolution of the limit order book is determined by two processes F^a and F^b defined on $\{0, 1, \ldots, T\} \times \Omega$, with values in the space $C([0, \infty))$ of continuous functions. The cost of buying $x \geq 0$ shares at time t is given by $F^a_t(x)$, while the profit of selling $x \geq 0$ shares at time t is given by $F^b_t(x)$. To simplify some formulas, we extend $F^a_t(x)$ and $F^b_t(x)$ to $\mathbb{R}$ by setting $F^a_t(x) = F^b_t(x) = 0$ for all $x \leq 0$ and all $t \in \{0, 1, \ldots, T\}$. To satisfy the general structure of the limit order book we impose the following properties :

(**P1**) For all $x \in \mathbb{R}$, $t \mapsto F^a_t(x)$ and $t \mapsto F^b_t(x)$ are adapted processes.

(**P2**) For all $t \in \{0, 1, \ldots, T\}$, $x \mapsto F^a_t(x)$ is non-negative, non-decreasing and convex while $x \mapsto F^b_t(x)$ is non-negative, non-decreasing and concave. Moreover, the right derivatives f^a_t and f^b_t of F^a_t and F^b_t are such that for any $x \geq 0$, $f^a_t(x) \geq f^b_t(x)$.

Remark 2.1. *The property (**P1**) states that the structure of the limit order book at time t incorporates all the informations up to time t. Property (**P2**) is necessary to insure proper ordering of the limit orders in the order book. In fact, for any $x, y, z \geq 0$, one must have*

$$F^a_t(x + y + z) - F^a_t(x + y) \geq F^a_t(x + z) - F^a_t(x), \tag{2.1}$$

since the cost of buying z additional shares if one has already bought $x + y$ is at least as much as buying z shares while one has bought x. This implies that $f^a_t(x) = D^+F^a_t(x)$ is non decreasing so F^a_t is convex. Similarly, one must have

$$F^b_t(x + y + z) - F^b_t(x + y) \leq F^b_t(x + z) - F^b_t(x), \tag{2.2}$$

for any $x, y, z \geq 0$, implying that $f^b_t(x) = D^+F^b_t(x)$ is non increasing, so F^b_t is concave. Finally, one must also have

$$F^a_t(x + y) - F^a_t(x) \geq F^b_t(x + y) - F^b_t(x), \quad \textit{for all } x, y \geq 0, \tag{2.3}$$

which is equivalent to $f^a_t(x) \geq f^b_t(x)$ for any $x \geq 0$. In fact, one just need to check that $f^a_t(0) \geq f^b_t(0)$ since f^a_t is non decreasing and f^b_t is non increasing. In particular, one gets that $F^a_t(x) \geq F^b_t(x)$ for any $x \geq 0$.

An interesting consequence of the model is given in the following proposition, following from (2.1)–(2.2) by putting $x = 0$.

Proposition 2.1. *Let $g, h : [0, \infty) \to \mathbb{R}$ be such that g is a convex function, h is a concave function and $g(0) = h(0) = 0$. Then g is superadditive and h is* *subadditive, i.e., for all* $x, y \geq 0$

$$g(x+y) \geq g(x) + g(y),$$

and

$$h(x+y) \leq h(x) + h(y).$$

In particular, F_t^a is superadditive and F_t^b is subadditive. These properties will be used often.

To complete the description of the model, we also assume that the investors can lend or borrow money from a bank account with possibly different interest rates for lending and borrowing. To this end, we define two processes $\{r_t^a\}_{t\in\{1,\ldots,T\}}$ and $\{r_t^b\}_{t\in\{1,\ldots,T\}}$ where r^a is the interest rate for lending and r^b for borrowing. As usual, we suppose that $0 \leq r_t^a \leq r_t^b$, and both r^a, r^b are predictable, i.e., $r_t^a, r_t^b \in \mathcal{F}_{t-1}$ for $t \in \{1, \ldots, T\}$.

2.2. Value of the portfolio

Let X be a predictable process representing the number of shares X_t in the investor's portfolio during period $(t-1, t]$, $t \in \{1, \ldots, T\}$. Further set $X_0 = 0$. Since we have two set of prices, we need to keep track of the total number of shares bought and sold. But first, we establish the following notation : $(x)^+ = \max\{x, 0\}$ and $(x)^- = \max\{-x, 0\}$. Then, we define $X_t^a = \sum_{i=1}^t (\Delta X_i)^+$ and $X_t^b = \sum_{i=1}^t (\Delta X_i)^-$, where $\Delta X_t = X_t - X_{t-1}$. The variable X_t^a (resp. X_t^b) gives the total number of shares bought (resp. sold) up to time t.

The cash flow of a transaction at time t is defined as

$$F_t^a(\Delta X_{t+1}^a) - F_t^b(\Delta X_{t+1}^b). \tag{2.4}$$

Then, we define the process Y as the amount of money in the bank account. For each $t \in \{1, \ldots, T\}$, define $\ell_{t-1}(y) = y^+(1 + r_t^a) - y^-(1 + r_t^b)$, and set

$$Y_t = \ell_{t-1}(Y_{t-1}) - \left\{ F_t^a(\Delta X_{t+1}^a) - F_t^b(\Delta X_{t+1}^b) \right\}, \tag{2.5}$$

where $Y_0 = \pi_0 - \left\{ F_0^a(\Delta X_1^a) - F_0^b(\Delta X_1^b) \right\}$ and π_0 is the initial value of the bank account. The function ℓ_t provides the interest rate for the period $(t, t+1]$ accordingly to the positive or negative value of the amount in the bank account, and Y_t gives the value of the position in the bank account immediately after the transaction at time t. From the definition of Y_t, all the changes of positions in the stock are financed by the bank account, so that we have a self-financing portfolio. The next step is to determine the liquidation value of the portfolio.

As opposed to a perfectly liquid market, in our setting, the liquidation value of the portfolio is not unique. It is well known in the literature, e.g. Bank and

Baum (2004), Obizhaeva and Wang (2005), Predoiu et al. (2011), that an investor should not trade a large amount of shares in one transaction since it will require to go deeper in the limit order book. Therefore, an investor could choose to liquidate its portfolio in many small trades instead of one large transaction. However, this strategy includes uncertainty in the liquidation value and highly complicates the computation of the portfolio value. To overcome this difficulty, some authors assume, as in Cetin et al. (2004), that the last transaction is not affected by the limit order book. In our case we consider that the impact of the last transaction should be considered, and we define the liquidation value as

$$F_t^b(X_t) - F_t^a(-X_t). \tag{2.6}$$

This value is the cash flow from closing the position in the stock in one transaction. Finally, the value of the portfolio is given by

$$V_t(X, \pi_0) = F_t^b(X_t) - F_t^a(-X_t) + \ell_{t-1}(Y_{t-1}) \tag{2.7}$$

for all $t \geq 1$ and $V_0(X, \pi_0) = \pi_0$.

Remark 2.2. *One should note that* (2.7) *is a self-financing portfolio condition. At time t, before rebalancing the portfolio, the amount in the bank account is $\ell_{t-1}(Y_{t-1})$. The cost of rebalancing the portfolio is $-\left\{F_t^a(\Delta X_{t+1}^a) - F_t^b(\Delta X_{t+1}^b)\right\}$ which is equal to the variation in the bank account*

$$Y_t - \ell_{t-1}(Y_{t-1}) = -\{F_t^a(\Delta X_{t+1}^a) - F_t^b(\Delta X_{t+1}^b)\}.$$

2.3. Non arbitrage conditions

The question of finding conditions to rule out arbitrage in a market model is crucial for the validity of a model. In the paradigm of a perfectly liquid and frictionless market, this question has been solved in the work of Harrison and Kreps (1979) and Harrison and Pliska (1981) and in a more general version of arbitrage in Delbaen and Schachermayer (1994). It has been showed that the absence of arbitrage is equivalent to the existence of an equivalent martingale probability measure for the former and that the condition of no-free-lunch-with-vanishing-risk is equivalent with the existence of an equivalent local-martingale probability measure in Delbaen and Schachermayer (1994).

In the context of illiquidity models, Cetin et al. (2004) show that the equivalence between the existence of an equivalent local-martingale probability measure for the marginal price and the condition of no-free-lunch-with-vanishing-risk still holds in their supply curve model. However, this result is based on the fact that the investor can trades at the marginal price by using sufficiently smooth trading strategies. Recently, Roch (2011) and Cetin et al. (2010) improved the model of Cetin et al. (2004) so that the investor cannot avoid the impact of the transactions and, in this case, they can only show the sufficiency part of the result. Simard (2014) presents a more general model where the investor cannot avoid the impact. It is shown that the existence of an equivalent probability measure such that the marginal price for the ask part and the bid part are respectively supermartingale and submartingale is sufficient to rule out arbitrage. It is also shown that this generalization to supermartingales and submartingales is not trivial. This result suggests that the martingale paradigm from the perfect liquidity models is not general enough. Our results, found in Theorems (2.1)-(2.2), bring a new kind of condition to rule out arbitrage.

Before going further, we establish some definitions and some notation that will be used throughout the paper. First, we restrict the set of trading strategies to assure the integrability of the portfolio value at maturity.

Definition 2.1. *A trading strategy* (X, π_0) *is admissible if* X *is almost surely bounded. The set of all admissible trading strategies is denoted by* $\mathcal{A}$.

Note that from (2.7), an admissible trading strategy is self-financing by definition.

Recall from (**P2**) that

$$
\begin{aligned}
f_t^a(x) &= D^+F_t^a(x) = \lim_{\epsilon \downarrow 0} \frac{F_t^a(x+\epsilon) - F_t^a(x)}{\epsilon}, \\
f_t^b(x) &= D^+F_t^b(x) = \lim_{\epsilon \downarrow 0} \frac{F_t^b(x+\epsilon) - F_t^b(x)}{\epsilon},
\end{aligned}
$$

and for simplicity, set $f_t^a = f_t^a(0)$ and $f_t^b = f_t^b(0)$. The function $f_t^a(x)$ (resp. $f_t^b(x)$) gives the price of the stock for buying (resp. selling) an infinitesimal number of shares in excess of x. For instance, the cost of buying $x + dx$ shares, when dx is "small" , is approximately $F_t^a(x + dx) = F_t^a(x) + f_t^a(x)dx$.

We are now in a position to define the notion of arbitrage.

Definition 2.2. *An arbitrage opportunity is an admissible trading strategy* $(X, 0)$ *such that*

$$P\{V_T(X, 0) \geq 0\} = 1 \text{ and } P\{V_T(X, 0) > 0\} > 0.$$

Consequently, there is no arbitrage if and only if for any admissible trading strategy $(X, 0)$, *one of the following (mutually exclusive) conditions is met :*

(A1) $P\{V_T(X, 0) = 0\} = 1$,

(A2) $P\{V_T(X, 0) < 0\} > 0$.

In the following, one gives conditions such that for a given probability measure, the expectation of the portfolio value at maturity is not greater than its initial value.

Let $\beta_0 = 1$ and $\beta_t = \frac{\beta_{t-1}}{1+r_t}$, where $r_t \in [r_t^a, r_t^b]$ for all $t \in \{1, \dots, T\}$. The process β can be seen as a discount factor associated to the intermediate interest rate process r. For a given probability measure Q on the filtered space $(\Omega, \mathbb{F})$, set

$$\begin{aligned} H_T^a &= \beta_T f_T^a, \\ H_{t-1}^a &= \min\{\beta_{t-1} f_{t-1}^a, E_Q(H_t^a|\mathcal{F}_{t-1})\}, \quad t \in \{1, \dots, T\}, \\ H_T^b &= \beta_T f_T^b, \\ H_{t-1}^b &= \max\{\beta_{t-1} f_{t-1}^b, E_Q(H_t^b|\mathcal{F}_{t-1})\}, \quad t \in \{1, \dots, T\}. \end{aligned} \tag{2.8}$$

Note that for each $t \in \{0, \dots, T\}$, $H_t^a \le \beta_t f_t^a$ and $H_t^b \ge \beta_t f_t^b$. Moreover H^a is a submartingale, while H^b is a supermartingale. The properties of the processes H^a and H^b will be used in Theorem 2.2 as a sufficient condition to rule out arbitrage. To compare with the classical theory of risk neutral evaluation, the condition for the absence of arbitrage is usually the existence of an equivalent probability measure such that the discounted value of stocks are martingales. In our case, we cannot define a unique discount factor, since there might be two different rates. As a result, one has to find an intermediate interest rate $r_t \in [r_t^a, r_t^b]$, leading to an intermediate discount factor β_t. Accordingly, the processes H^a and H^b are extension of the discounted value of the assets.

The proof of the following theorem is given is Appendix B.

Theorem 2.1. *Let Q be a given probability measure on the filtered space $(\Omega, \mathbb{F})$ and assume that $E_Q\{F_t^a(x)\} < \infty$ for all x and for all $t \in \{0, \dots, T\}$. Then $E_Q\{\beta_T V_T(X, \pi_0)\} \le \pi_0$ for all admissible strategies (X, π_0) if the following conditions are satisfied : for all $t \in \{1, \dots, T\}$,*

$$E_Q\left(H_t^a|\mathcal{F}_{t-1}\right) \ \ge \ \beta_{t-1} f_{t-1}^b, \tag{2.9}$$

$$E_Q\left(H_t^b|\mathcal{F}_{t-1}\right) \ \le \ \beta_{t-1} f_{t-1}^a. \tag{2.10}$$

Moreover, if $r_t = r_t^a = r_t^b$ for all $t \in \{1, \dots, T\}$, then the conditions are also necessary, i.e., if (2.9)–(2.10) are met, then $E_Q\{\beta_T V_T(X, \pi_0)\} \le \pi_0$ for all admissible strategies (X, π_0).

Next, the following no-arbitrage result is a direct consequence of Theorem 2.1. This result gives a systematic way to rule out arbitrage in the market model. The absence of arbitrage is an essential condition in the theory of pricing, since it implies that assets producing the same returns must have the same value.

Theorem 2.2. *Let the predictable process r be such that $r_t^a \le r_t \le r_t^b$, $i \in \{1, \dots, T\}$. Suppose there exists an equivalent probability measure Q on the filtered* *space* $(\Omega, \mathbb{F})$ *such that* $E_Q\{F_t^a(x)\} < \infty$ *for all* x *and for all* $t \in \{0, \ldots, T\}$*, and such that the processes* H^a *and* H^b *defined in* (2.8) *satisfy* (2.9)–(2.10)*. Then, there is no arbitrage.*

Remark 2.3. *It is interesting to note that in the classical case of perfectly liquid markets, i.e.,* $F_t^a(x) = F_t^b(x) = f_t x$*, for some adapted process* f_t*, conditions* (2.9)–(2.10) *are equivalent to* $H^a = H^b = \beta f$ *and* βf *is a* $Q-$*martingale for some probability measure* Q *equivalent to* P*, recovering the first fundamental theorem of asset pricing.*

2.4. Possible extension

Definition 2.2 states that an arbitrage opportunity must start with zero initial wealth, that is $\pi_0 = 0$. However, if $r^a \equiv r^b$, then $V_T(X, \pi_0) = V_T(X, 0) + V_T(0, \pi_0)$ and we have that

$$P\{V_T(X, 0) \geq 0\} = P\{V_T(X, \pi_0) \geq V_T(0, \pi_0)\}$$

and

$$P\{V_T(X, 0) > 0\} = P\{V_T(X, \pi_0) > V_T(0, \pi_0)\}.$$

We see that, in this case, the definition of arbitrage given in 2.2 could be generalized to $\pi_0 \in \mathbb{R}$ and Theorem 2.2 would still be valid, which is not true if $r^a \neq r^b$. It is also the case that in a perfectly liquid market with a unique interest rate, the definition of arbitrage is equivalent whether the initial wealth is set to zero or not. This suggests that, in our model, the definition of arbitrage opportunity could be generalized and that other conditions would have to be defined to rule out arbitrage in the general case. This problem will be studied in a future paper.

3. Value of call and put options for buyers and sellers

We now propose a way to evaluate European call and put options. Assume that the maturity of the option is $T \geq 1$, the strike is K, the option is written for 1 share of the underlying stock and the settlement is in-kind, not a cash settlement. The latter occurs for example for options on indices.

In a perfectly liquid market, the price of an option, is the same whether the settlement is in-kind or cash, since the payoff is the same. The latter is not true in our setting, so we begin by assuming that the settlement is in-kind. Actually, it is the case for most of the stock options in the market. The case of in-kind settlement also allows to deduce the payoff function of the options. The case of cash settlement is discussed in Section 3.4.

To determine the value of the call or the put, one needs to take the viewpoint of the buyer and the seller, thought the value of their hedging portfolios. We consider that a price of the option is acceptable for the buyer or the seller if they can build an hedging portfolio which is non-negative with probability one.

3.1. Hedging portfolios

In what follows we define the hedging portfolios for buyers and sellers of call and put options. First, we consider the value of the portfolio for the buyer of a call option.

3.1.1. *Buyer of a call option*

In order to realize a profit if she exercises, the buyer of the call has to sell the share of the stock she receives from exercising the option. So, the decision to exercise is based on the difference between the strike K and the value of selling one share at market value at time T, i.e., $F_T^b(1)$. Consequently, we assume that the call is exercised if and only if $F_T^b(1) > K$.

The value of the buyer's portfolio is

$$\begin{aligned} V_T(X, -\pi_0) &= F_T^b(X_T + \mathbf{1}_{\{F_T^b(1)>K\}}) - F_T^a(-X_T - \mathbf{1}_{\{F_T^b(1)>K\}}) \\ &\quad + \ell_{T-1}(Y_{T-1}) - K\mathbf{1}_{\{F_T^b(1)>K\}}, \end{aligned} \tag{3.1}$$

where $Y_0 = -\pi_0 - \left\{F_0^a(\Delta X_1^a) - F_0^b(\Delta X_1^b)\right\}$ and π_0 is the price of the call. Here, $\mathbf{1}_A$ is defined as indicator function of the set A. In (3.1), one sees that the share the buyer receives from exercising the option is added to the trading strategy in the liquidation value of the portfolio.

Now we determine the value of the portfolio for the seller of a call.

3.1.2. *Seller of a call contract*

The seller must consider that she has to deliver one share of the underlying stock when the buyer exercises, so that the value of the seller's portfolio is given by

$$\begin{aligned} V_T(X, \pi_0) &= F_T^b(X_T - \mathbf{1}_{\{F_T^b(1)>K\}}) - F_T^a(-X_T + \mathbf{1}_{\{F_T^b(1)>K\}}) \\ &\quad + \ell_{T-1}(Y_{T-1}) + K\mathbf{1}_{\{F_T^b(1)>K\}}, \end{aligned} \tag{3.2}$$

where $Y_0 = \pi_0 - \left\{F_0^a(\Delta X_1) - F_0^b(-\Delta X_1)\right\}$ and π_0 is the price of the call. This time the sign before $\mathbf{1}_{\{F_T^b(1)>K\}}$ is different than in the buyer's case since the seller has to deliver the stock. The portfolios defined for the buyer and seller of the put follow the same logic.

3.1.3. *Buyer of a put contract*

In the case of a put option, for the buyer to realize a profit, she has to consider the difference between the value of the strike K and the value $F_T^a(1)$ of buying one share at period T. Consequently, one assumes that the put is exercised if and only if $F_T^a(1) < K$.

Then the value of the buyer's portfolio is given by

$$\begin{aligned} V_T(X, -\pi_0) &= F_T^b(X_T - \mathbf{1}_{\{F_T^a(1)<K\}}) - F_T^a(-X_T + \mathbf{1}_{\{F_T^a(1)<K\}}) \\ &\quad + \ell_{T-1}(Y_{T-1}) + K\mathbf{1}_{\{F_T^a(1)<K\}}, \end{aligned} \tag{3.3}$$

where $Y_0 = -\pi_0 - \left\{F_0^a(\Delta X_1) - F_0^b(-\Delta X_1)\right\}$ and π_0 is the price of the put.

3.1.4. *Seller of a put contract*

Finally, using the same reasoning as before, the value of the seller's portfolio is given by

$$\begin{aligned} V_T(X, \pi_0) &= F_T^b(X_T + \mathbf{1}_{\{F_T^a(1)<K\}}) - F_T^a(-X_T - \mathbf{1})_{\{F_T^a(1)<K\}}) \\ &\quad + \ell_{T-1}(Y_{T-1}) - K\mathbf{1}_{\{F_T^a(1)<K\}}, \end{aligned} \tag{3.4}$$

where $Y_0 = \pi_0 - \left\{F_0^a(\Delta X_1) - F_0^b(-\Delta X_1)\right\}$ and π_0 is the price of the put.

3.2. Hedging value of the contracts

The hedging portfolio values are now used to define bid and ask prices for the call and put options.

3.2.1. *Bid price of the option*

For the buyer of the option, the bid price is the highest π_0 for which there exists an admissible trading strategy (X, π_0) such that $V_T(X, -\pi_0) \geq 0$, where $V_T(X, -\pi_0)$ is given by (3.1) or (3.3), depending if it is a call or a put option.

3.2.2. *Ask price of the option*

For the seller, the ask price is the smallest π_0 for which there exists an admissible trading strategy (X, π_0) such that $V_T(X, \pi_0) \geq 0$, where $V_T(X, \pi_0)$ is given by (3.2) or (3.4), depending if it is a call or a put option.

Remark 3.1. *In this section we defined how to determine the price of call and put options written for one share of the underlying stock. In a perfectly liquid market, the price of options (per unit of the underlying stock) does not depend on the number of shares of the underlying stock specified in the option contract. Consequently, there is no loss of generality considering an option for one share of* *the underlying stock. In our case, the price of an option (per unit of the underlying) depends, in general, on the number of shares of the underlying. The choice of pricing options for one share of the underlying is a matter of rescaling the problem.*

Again, in a perfectly liquid market, the problem of pricing options does not depend on the number of options in the hedging portfolio, which is not true in our setting. Suppose we take the case of a buyer of $n > 1$ call options for one share of the underlying stock with strike K and physical delivery. If $F_T^b(1) > K$, then it is not true in general that $\frac{F_T^b(n)}{n} > K$ since the function $F_T^b(x)/x$ is non-increasing. Thus, the buyer will exercise $m \leq n$ options where $m = \arg\sup_{m \leq n} \frac{F_T^b(m)}{m} > K$. Consequently, the price the buyer is willing to pay will depend on the number of options she bought.

3.3. Non linearity

A particularity of the model is that, in general, the portfolio value is not a linear function of the hedging strategy, meaning that in general one does not have

$$V_T(X + \tilde{X}, \pi_0 + \tilde{\pi}_0) = V_T(X, \pi_0) + V_T(\tilde{X}, \tilde{\pi}_0)$$

for $(X, \pi_0), (\tilde{X}, \tilde{\pi}_0) \in \mathcal{A}$. This absence of linearity can creates a lot of problems. For example, there is no more put-call parity formula, since it is based on a payoff giving at maturity one share minus K dollars. Similarly, pricing forward contracts or futures is not as easy as in the classical case. Also, by setting $K = 0$ in a call option, one does not necessarily recover the value $F_0^a(1)$. This last fact will be discussed later.

3.4. Cash settlement

We mentioned at the beginning of this section that the problem of pricing options with cash or physical settlement is no longer the same. Suppose we take a call option with payoff function $\max\{F_T^b(1) - K; 0\}$ and with cash settlement. The value of the seller's hedging portfolio is defined by

$$V(X, \pi_0) = L_T(X) - \ell_{T-1}(Y_{T-1}) - \max\{F_T^b(1) - K; 0\},$$

where π_0 is the price of the option. This value is different than the one defined in (3.2) since the seller no longer has to deliver the stock. Indeed, the seller's price of the option should be different than in the case of physical delivery.

Moreover, the case of in-kind settlement leads naturally to the payoff function of the option, since the buyer has to sell (resp. buy) one share of the underlying when exercising the call (resp. put) option. It is no longer the case with cash settlement options. Since there is two sets of prices and that the buyer of the option receives money upon exercise, there is no reason to restrict the payoff functions to $\max\{F_T^b(1) - K; 0\}$ for the call and $\max\{K - F_T^a(1); 0\}$ for the put. For instance, one could defined the payoff of the call as $\max\{F_T^a(1) - K; 0\}$ or $\max\left[\frac{1}{2}\left\{F_T^a(1) + F_T^b(1)\right\} - K; 0\right]$.

4. One-period model

One now consider the question of pricing and hedging a European call or put option in a one-period model. Most of the past researches on the topic consider continuous time models and the solution to the hedging problem often turns out to be equivalent to solving a complicated differential equation ; see, e.g., Liu and Yong (2005) and Cvitanić and Ma (1996). The approach here is to simplify the model by restricting the space Ω to be countable, i.e., $\Omega = \{\omega_i;\ i \in I\}$, where I is finite or $I = \mathbb{N}$, and $T = 1$. Since I can be finite, one may assume that $P(\omega_i) > 0$ for all $i \in I$. In this framework, the solution to the pricing problem can be made more explicit so that it is easier to analyze the behaviour of the price and the hedging strategy. By assuming that the investment horizon is $T = 1$, one can also refine the no-arbitrage result by providing sufficient and necessary conditions to rule out arbitrage, including different interest rates for the lender and the borrower.

4.1. No-arbitrage conditions

Since X_1 is $\mathcal{F}_0$-measurable it is constant, so one can set $X_1 = x$ as a non-random variable. Setting $\pi_0 = 0$, the value of the one-period self-financing portfolio reduces to these two cases. Since it only depends on x, one writes $V_1(x)$ for the value of the portfolio.

If $x > 0$ then $Y_1 = -F_0^a(x)$ and

$$V_1(x) = F_1^b(x) - (1 + r_1^b)F_0^a(x).$$

If $x = -y < 0$ then $Y_1 = F_0^b(y)$ and

$$V_1(-y) = -F_1^a(y) + (1 + r_1^a)F_0^b(y).$$

Then $V_1(x) \geq 0$ if $F_1^b(x) \geq (1+r_1^b)F_0^a(x)$ for some $x \geq 0$ or $F_1^a(x) \leq (1+r_1^a)F_0^b(x)$ for some $x \geq 0$.

>From Definition 2.2, we see that there is no arbitrage opportunity if and only if

$$P\{V_1(x) = 0\} = 1 \text{ or } P\{V_1(x) < 0\} > 0.$$

This last condition for the absence of arbitrage can be stated in a more practical way, which is the point of the next theorem, proven in Appendix C.

Theorem 4.1. *There is no arbitrage for a one-period maturity portfolio if, from the following conditions, either* $(NA1)$ *or* $(NA2)$ *is satisfied and either* $(NA3)$ *or* $(NA4)$ *is also satisfied.*

$$\begin{aligned}
(NA1): \sup_{\omega\in\Omega} f_1^b(0,\omega) &\leq (1+r_1^b) f_0^a(0),\\
(NA2): \inf_{\omega\in\Omega} f_1^b(0,\omega) &< (1+r_1^b) f_0^a(0) < \sup_{\omega\in\Omega} f_1^b(0,\omega),\\
(NA3): \inf_{\omega\in\Omega} f_1^a(0,\omega) &\geq (1+r_1^a) f_0^b(0),\\
(NA4): \sup_{\omega\in\Omega} f_1^a(0,\omega) &> (1+r_1^a) f_0^b(0) > \inf_{\omega\in\Omega} f_1^a(0,\omega).
\end{aligned}$$

Remark 4.1. *Note that in the classical binomial tree model where* $f_1 = f_0 U$ *or* $f_1 = f_0 D$, *one recovers the usual conditions of non arbitrage, i.e.,* $U > 1+r > D$ *or* $U = D = 1+r$.

In the next two sections, one gives the price of the European call and put options from the seller's and buyer's viewpoint. The results are then discussed in Section 4.4.

4.2. Pricing European calls

In this section one determines the price for the European call with maturity $T = 1$, strike price K, for physical delivery of one share of the underlying asset. The expressions we derive are actually for the bid and the ask price of the options, that is the highest and lowest price which allow the buyer or the seller of the option to build a non-negative hedging portfolio. We sometimes refer to these values as *the price* since the meaning will be clear from the context.

As discussed in Section 3, one assumes that the option is exercised if and only if $F_1^b(1) > K$.

To treat this problem one imposes the following conditions :

(**C1**) : There is no possibility of arbitrage.

(**C2**) : $\sup_{\omega\in\Omega} F_1^a(x,\omega) = \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(x,\omega)$.

(**C3**) : For any $\omega,\tilde{\omega}\in\Omega$, $F_t^b(x_0,\omega) \leq F_t^b(x_0,\tilde{\omega})$ for some $x_0 > 0$ if and only if $F_t^b(x,\omega) \leq F_t^b(x,\tilde{\omega})$ for all $x \geq 0$. Also, for any $\omega,\tilde{\omega}\in\Omega$, $F_t^a(x_0,\omega) \leq F_t^a(x_0,\tilde{\omega})$ for some $x_0 > 0$ if and only if $F_t^a(x,\omega) \leq F_t^a(x,\tilde{\omega})$ for all $x \geq 0$.

Conditions (**C2**) and (**C3**) simplify some computations, but can also be justified financially. We can interpret the set $\{F_1^b(1) \leq K\}$ as bear market states and $\{F_1^b(1) > K\}$ as bull market states which explains (**C2**) while we suppose that an upward (resp. downward) movement on prices applies to the whole limit order book structure, which explains (**C3**).

One needs to consider three cases : $P\{F_1^b(1) > K\} = 1$, $P\{F_1^b(1) > K\} = 0$, and $0 < P\{F_1^b(1) > K\} < 1$. The case $P\{F_1^b(1) > K\} = 1$ corresponds to the situation where the buyer always exercises, while $P\{F_1^b(1) > K\} = 0$ is the case where the buyer never exercises. Some consequences of these two limiting, and rather extreme cases will be discussed in turn.

4.2.1. *Seller's call price*

One first looks at the price of the option from the seller's point of view. Using (3.2) with $T = 1$, the value of the hedging portfolio of the seller of the call option is

$$V_1(x,\pi_0) = \begin{cases} F_1^b(x) - F_1^a(-x) + \ell_0(y), & \text{if } F_1^b(1) \le K, \\ F_1^b(x-1) - F_1^a(-x+1) + \ell_0(y) + K, & \text{if } F_1^b(1) > K, \end{cases}$$

where $y = \pi_0 - \left\{F_0^a(x) - F_0^b(x)\right\}$, and $\ell_0(y) = y^+(1+r_1^a) - y^-(1+r_1^b)$. The ask price, i.e., the lowest price the seller is willing to accept, is the smallest π_0 such that there exists x for which $V_1(x,\pi_0,\omega) \ge 0$ for every $\omega \in \Omega$.

Set

$$\begin{aligned} \ell^{(ca)}(x) = \min&\left\{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega), K + \inf_{\omega\in\{F_1^b(1)>K\}} F_1^b(x-1,\omega)\right\}\mathbf{1}_{[1,\infty)}(x) \\ +\min&\left\{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega), K - \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega)\right\}\mathbf{1}_{[0,1]}(x) \\ -\max&\left\{\sup_{\omega\in\{F_1^b(1)\le K\}} F_1^a(-x,\omega), \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega) - K\right\}\mathbf{1}_{(-\infty,0]}(x). \end{aligned}$$

Then $V_1(x,\pi_0,\omega) \ge 0$ for all $\omega \in \Omega$ if and only if $\pi_0 \ge F_0^a(x) - F_0^b(x) + \ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\}$.

In what follows, we will use the convention that the infimum and the supremum of non negative functions over empty sets are respectively $+\infty$ and $-\infty$. This way, $\ell^{(ca)}$ is well-defined even if $P\{F_1^b(1) > K\} = 0$ or 1. For if $P\{F_1^b(1) > K\} = 0$, then

$$\ell^{(ca)}(x) = \inf_\omega F_1^b(x,\omega)\mathbf{1}_{[0,\infty)}(x) - \sup_\omega F_1^a(-x,\omega)\mathbf{1}_{(-\infty,0]}(x), \tag{4.1}$$

while

$$\ell^{(ca)}(x) = K + \inf_\omega F_1^b(x-1,\omega)\mathbf{1}_{[1,\infty)}(x) - \sup_\omega F_1^a(1-x,\omega)\mathbf{1}_{(-\infty,1]}(x), \tag{4.2}$$

if $P\{F_1^b(1) > K\} = 1$.

Theorem 4.2. *Suppose that (**C1**)–(**C3**) hold. The ask price π_0^a of the European call is given by*

(1) $\pi_0^a = 0$, *if* $P\{F_1^b(1) > K\} = 0$ *;*

(2)

$$\begin{aligned}\pi_0^a &= \min\left[\inf_{x\in[0,x_0]}\left\{F_0^a(x) + \frac{\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega)-K}{1+r_1^a}\right\},\right.\\ &\left.\inf_{x\in[x_0,x_1]}\left\{F_0^a(x) + \frac{\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega)-K}{1+r_1^b}\right\}\right],\end{aligned}$$

if $0 < P\{F_1^b(1) > K\} < 1$, *where* x_0, x_1 *are such that* $\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x_0,\omega) = K$, $x_1 = \inf\left\{x \in (x_0,1); \ell^{(ca)}(x) = \inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega)\right\}$, *and* $0 < x_0 < x_1 < 1$ *;*

(3)

$$\begin{aligned}\pi_0^a &= \min\left[\inf_{x\in[0,x_0]}\left\{F_0^a(x) + \frac{\sup_{\omega} F_1^a(1-x,\omega)-K}{1+r_1^a}\right\},\right.\\ &\left.\inf_{x\in[x_0,1]}\left\{F_0^a(x) + \frac{\sup_{\omega} F_1^a(1-x,\omega)-K}{1+r_1^b}\right\}\right],\end{aligned}$$

if $P\{F_1^b(1) > K\} = 1$.

Remark 4.2. *Note that the optimal number of shares to construct the hedging portfolio is always between* 0 *and* 1.

The proof is given in Appendix D.

4.2.2. *Buyer's call price*

We now consider the price of the call from the buyer's point of view. Using (3.1) with $T = 1$, the value of the hedging portfolio is given by

$$V_1(x,-\pi_0) = \begin{cases} F_1^b(x) - F_1^a(-x) + \ell_0(y), & \text{if } F_1^b(1) \le K,\\ F_1^b(x+1) - F_1^a(-x-1) + \ell_0(y) - K, & \text{if } F_1^b(1) > K,\end{cases}$$

where $y = -\pi_0 - \left\{F_0^a(x) - F_0^b(x)\right\}$ and $\ell_0(y) = y^+(1+r_1^a) - y^-(1+r_1^b)$. The bid price, i.e., the highest price the buyer is willing to accept, is the highest π_0 such that there exists x for which $V_1(x,-\pi_0) \ge 0$ a.s.

Before stating the next theorem, set

$$\ell^{(cb)}(x) = \min\left\{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega), \inf_{\omega\in\{F_1^b(1)>K\}} F_1^b(x+1,\omega) - K\right\}\mathbf{1}_{[0,\infty)}(x)$$

$$
\begin{aligned}
&+ \quad \min\left\{ - \sup_{\omega\in\{F_1^b(1)\le K\}} F_1^a(-x,\omega), \inf_{\omega\in\{F_1^b(1)>K\}} F_1^b(x+1,\omega) - K \right\} \mathbf{1}_{[-1,0]}(x) \\
&- \quad \max\left\{ \sup_{\omega\in\{F_1^b(1)\le K\}} F_1^a(-x,\omega), \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(-x-1,\omega) + K \right\} \mathbf{1}_{(-\infty,-1]}(x).
\end{aligned}
$$

We retain the same conventions for the infimum and supremum over empty sets. If $P\{F_1^b(1) > K\} = 0$, then

$$
\ell^{(cb)}(x) \quad = \quad \inf_\omega F_1^b(x,\omega)\mathbf{1}_{[0,\infty)}(x) - \sup_\omega F_1^a(-x,\omega)\mathbf{1}_{(-\infty,0]}(x),
$$

while

$$
\begin{aligned}
\ell^{(cb)}(x) \quad &= \quad \inf_\omega F_1^b(x+1,\omega)\mathbf{1}_{[-1,\infty)}(x) \\
&\quad - \sup_\omega F_1^a(-x-1,\omega)\mathbf{1}_{(-\infty,-1]}(x) - K, \qquad (4.3)
\end{aligned}
$$

if $P\{F_1^b(1) > K\} = 1$.

Theorem 4.3. *Suppose that (**C1**)–(**C3**) hold. The bid price π_0^b of the European call option is given by*

(1) $\pi_0^b = 0$, *if* $P\{F_1^b(1) > K\} = 0$,

(2)

$$
\pi_0^b = \sup_{x\le 0}\left\{ F_0^b(-x) + \frac{\ell^{(cb)}(x)}{1+r_1^a} \right\},
$$

if $0 < P\{F_1^b(1) > K\} < 1$, *and if* $\sup_{\omega\in\{F_1^b(1)\le K\}} F_1^a(1,\omega) < K$, *then*

$$
\pi_0^b = \sup_{x\in[-1,0]}\left\{ F_0^b(-x) + \frac{\ell^{(cb)}(x)}{1+r_1^a} \right\}, \qquad (4.4)
$$

(3)

$$
\begin{aligned}
\pi_0^b \quad = \quad \max\Bigg[&\sup_{x\in[-1,x_0]}\left\{ F_0^b(-x) + \frac{\inf_\omega F_1^b(x+1,\omega) - K}{1+r_1^a} \right\}, \\
&\sup_{x\in[x_0,0]}\left\{ F_0^b(-x) + \frac{\inf_\omega F_1^b(x+1,\omega) - K}{1+r_1^b} \right\} \Bigg],
\end{aligned}
$$

if $P\{F_1^b(1) > K\} = 1$, *where* $x_0 \in [-1,0]$, *is defined by* $\inf_\omega F_1^b(x_0+1,\omega) = K$.

The proof is given in Appendix E.

4.3. Pricing European puts

In this section we determine the price for the European put with maturity $T = 1$, strike price K, for physical delivery of one share of the underlying asset.

As discussed in Section 3, we assume that the option is exercised if and only if $F_1^a(1) < K$.

To treat this problem, the following conditions are imposed :

(**C1**) There is no possibility of arbitrage.

(**C2′**) $\inf_{\omega\in\Omega} F_1^b(x,\omega) = \inf_{\omega\in\{F_1^a(1)<K\}} F_1^b(x,\omega)$.

(**C3′**) For $\omega, \tilde{\omega} \in \Omega$, $F_t^b(x_0,\omega) \le F_t^b(x_0,\tilde{\omega})$ for some $x_0 > 0$ if and only if $F_t^b(x,\omega) \le F_t^b(x,\tilde{\omega})$ for all x. Also, for any $\omega, \tilde{\omega} \in \Omega$, $F_t^a(x_0,\omega) \le F_t^a(x_0,\tilde{\omega})$ for some $x_0 > 0$ if and only if $F_t^a(x,\omega) \le F_t^a(x,\tilde{\omega})$ for all $x \ge 0$.

Once again, conditions (**C2′**) and (**C3′**) simplify the computations, and are justified financially as in Section 4.2. We interpret $\{F_1^a(1) < K\}$ as bear market states and $\{F_1^a(1) \ge K\}$ as bull market states, which explains $(C2')$ while we suppose that upward and downward movements affect the whole limit order book structure, which explains (**C3′**).

Also, one has to consider the three cases : $P\{F_1^a(1) < K\} = 1$ and $P\{F_1^a(1) < K\} = 0$ corresponding respectively to the situations where the buyer exercises with probability one and probability zero and finally the case $0 < P\{F_1^a(1) < K\} < 1$.

4.3.1. *Seller's put price*

One first looks at the price of the option from the seller point of view. Using (3.4) with $T = 1$, the value of the hedging portfolio of the seller of the put option is

$$V_1(x,\pi_0) = \begin{cases} F_1^b(x) - F_1^a(-x) + \ell_0(y), & \text{if } F_1^a(1) \ge K, \\ F_1^b(x+1) - F_1^a(-x-1) + \ell_0(y) - K, & \text{if } F_1^a(1) < K, \end{cases}$$

where $y = \pi_0 - \{F_0^a(x) - F_0^b(x)\}$ and $\ell_0(y) = y^+(1+r_1^a) - y^-(1+r_1^b)$. The ask price, i.e., the lowest price the seller is willing to accept, is the smallest π_0 such that there exists x for which $V_1(x,\pi_0) \ge 0$ a.s.

Set

$$\begin{aligned}\ell^{(pa)}(x) &= \min\left\{\inf_{\omega\in\{F_1^a(1)<K\}} F_1^b(x+1,\omega) - K, \inf_{\omega\in\{F_1^a(1)\ge K\}} F_1^b(x,\omega)\right\}\mathbf{1}_{[0,\infty)}(x) \\ &+ \min\left\{\inf_{\omega\in\{F_1^a(1)<K\}} F_1^b(x+1,\omega) - K, -\sup_{\omega\in\{F_1^a(1)\ge K\}} F_1^a(-x,\omega)\right\}\mathbf{1}_{[-1,0]}(x) \\ &- \max\left\{\sup_{\omega\in\{F_1^a(1)<K\}} F_1^a(-x-1,\omega) + K, \sup_{\omega\in\{F_1^a(1)\ge K\}} F_1^a(-x,\omega)\right\}\mathbf{1}_{(-\infty,-1]}(x).\end{aligned}$$

In particular, if $P\{F_1^a(1) < K\} = 0$, then

$$\ell^{(pa)}(x) = \inf_\omega F_1^b(x,\omega)\mathbf{1}_{[0,\infty)}(x) - \sup_\omega F_1^a(-x,\omega)\mathbf{1}_{(-\infty,0]}(x),$$

while

$$\ell^{(pa)}(x) = \inf_\omega F_1^b(x+1,\omega)\mathbf{1}_{[-1,\infty)}(x) - \sup_\omega F_1^a(-x-1,\omega)\mathbf{1}_{(-\infty,-1]}(x) - K,$$

if $P\{F_1^a(1) < K\} = 1$.

Theorem 4.4. *Suppose that (**C1**), (**C2′**)–(**C3′**) hold. The ask price π_0^a of the European put is given by*

(1) $\pi_0^a = 0$, *if* $P\{F_1^a(1) < K\} = 0$,

(2)

$$\begin{aligned}\pi_0^a &= \min\Bigg[\inf_{x\in[x_0,0]}\left\{\frac{K - \inf_{\omega\in\{F_1^a(1)<K\}} F_1^b(1+x,\omega)}{1+r_1^a} - F_0^b(-x)\right\}, \\ &\qquad \inf_{x\in[0,x_1]}\left[\ell_0^{-1}\left\{K - \inf_{\omega\in\{F_1^a(1)<K\}} F_1^b(1+x,\omega)\right\} + F_0^a(x)\right]\Bigg],\end{aligned}$$

if $0 < P\{F_1^a(1) < K\} < 1$, *where*

$$x_0 = \sup\left\{x \in ([-1,0]; \ell^{(pa)}(x) = -\sup_{\omega\in\{F_1^a(1)\geq K\}} F_1^a(-x,\omega)\right\},$$

and x_1 is such that $\inf_\omega F_1^b(x_1+1,\omega) = K$. *If in addition* $\frac{\inf_{\omega\in\{F_1^a(1)<K\}} f_1^b(1,\omega)}{1+r_1^a} \leq f_0^a(0)$, *then*

$$\pi_0^a = \inf_{x\in[x_0,0]}\left\{\frac{K - \inf_{\omega\in\{F_1^a(1)<K\}} F_1^b(1+x,\omega)}{1+r_1^a} - F_0^b(-x)\right\}.$$

(3) If $P\{F_1^a(1) < K\} = 1$,

$$\begin{aligned}\pi_0^a &= \min\Bigg[\inf_{x\in[-1,0]}\left\{\frac{K - \inf_\omega F_1^b(1+x,\omega)}{1+r_1^a} - F_0^b(-x)\right\} \\ &\qquad \inf_{x\in[0,x_1]}\left[\ell_0^{-1}\left\{K - \inf_\omega F_1^b(1+x,\omega)\right\} + F_0^a(x)\right]\Bigg].\end{aligned}$$

If in addition $\frac{\inf_\omega f_1^b(x+1,\omega)}{1+r_1^a} \leq f_0^a(0)$, *then*

$$\pi_0^a = \inf_{x\in[-1,0]}\left\{\frac{K - \inf_\omega F_1^b(1+x,\omega)}{1+r_1^a} - F_0^b(-x)\right\}.$$

The proof is given in Appendix F.

4.3.2. *Buyer's put price*

One now looks at the price of the put from the buyer's point of view. Using (3.3) with $T=1$ the value of the hedging portfolio is given by

$$V_1(x,-\pi_0) = \begin{cases} F_1^b(x) - F_1^a(-x) + \ell_0(y), & \text{if } F_1^a(1) \geq K, \\ F_1^b(x-1) - F_1^a(-x+1) + \ell_0(y) + K, & \text{if } F_1^a(1) < K, \end{cases}$$

where $y = -\pi_0 - \{F_0^a(x) - F_0^b(x)\}$ and $\ell_0(y) = y^a(1+r_1^a) - y^b(1+r_1^b)$. The bid price, i.e., the highest price the buyer is willing to accept is the biggest $\pi_0 \geq 0$ such that there exists x for which $V_1(x,-\pi_0) \geq 0$ a.s.

Before stating the next theorem, set

$$\ell^{(pb)}(x) = \min\left\{ \inf_{\omega\in\{F_1^a(1)<K\}} F_1^b(x-1,\omega) + K, \inf_{\omega\in\{F_1^a(1)\geq K\}} F_1^b(x,\omega) \right\} \mathbf{1}_{[1,\infty)}(x)$$
$$+ \min\left\{ - \sup_{\omega\in\{F_1^a(1)<K\}} F_1^a(-x+1,\omega) + K, \inf_{\omega\in\{F_1^a(1)\geq K\}} F_1^b(x,\omega) \right\} \mathbf{1}_{[0,1]}(x)$$
$$- \max\left\{ \sup_{\omega\in\{F_1^a(1)<K\}} F_1^a(-x+1,\omega) - K, \sup_{\omega\in\{F_1^a(1)\geq K\}} F_1^a(-x,\omega) \right\} \mathbf{1}_{(-\infty,0]}(x).$$

If $P\{F_1^a(1) < K\} = 0$,

$$\ell^{(pb)}(x) = \inf_\omega F_1^b(x,\omega)\mathbf{1}_{[0,\infty)}(x) - \sup_\omega F_1^a(-x,\omega)\mathbf{1}_{(-\infty,0]}(x),$$

while

$$\ell^{(pb)}(x) \quad = \quad \inf_\omega F_1^b(x-1,\omega)\mathbf{1}_{[1,\infty)}(x) - \sup_\omega F_1^a(-x+1,\omega)\mathbf{1}_{(-\infty,1]}(x) + K,$$

if $P\{F_1^a(1) < K\} = 1$.

Theorem 4.5. *Suppose that (**C1**), (**C2′**)–(**C3′**) hold. The bid price π_0^b of the European put is given by*

(1) $\pi_0^b = 0$, *if* $P\{F_1^a(1) < K\} = 0$.

(2)

$$\pi_0^b = \sup_{x\geq 0} \left\{ \frac{\ell^{(pb)}(x)}{1+r_1^b} - F_0^a(x) \right\},$$

if $0 < P\{F_1^a(1) < K\} < 1$. *If in addition* $\inf_{\omega\in\{F_1^a(1)\geq K\}} F_1^b(1,\omega) > K$, *then*

$$\pi_0^b = \sup_{[0,1]} \left\{ \frac{\ell^{(pb)}(x)}{1+r_1^b} - F_0^a(x) \right\}.$$

(3)

$$\pi_0^b = \sup_{x\in[0,1]} \left\{ \frac{K - \sup_\omega F_1^a(1-x,\omega)}{1+r_1^b} - F_0^a(x) \right\},$$

if $P\{F_1^a(1) < K\} = 1$.

The proof is given in Appendix G.

4.4. Discussion and specific cases

As we saw previously, even for a one maturity option, the problem of pricing an European option is much more complex when we consider the structure of the limit order book. Moreover, we see significant differences between those results and corresponding results from perfect liquidity models. Here we highlight some particularities of our results

As it is expected, the value of the option generally depends on the value of the strike. However, for in-the-money options, the dependence is not necessarily linear as it is the case for the classical one period binomial model. In the seller's case, the value of x_0, with its respective definition in Theorems 4.2 and 4.4, depends on the value of the strike. On the other hand, one surprising result is that it is possible that the buyer's price, as well as the hedging strategy, do not depend on the strike for values in some open interval. For instance, let's look at the buyer's call price in Theorem 4.3, for the case $0 < P\{F_1^b(1) > K\} < 1$. If $\sup_{\omega \in \{F_1^b(1) \leq K\}} F_1^a(1, \omega) > K$, then there exists $\epsilon > 0$ such that $\ell^{(cb)}(x) = -\sup_{\omega \in \{F_1^b(1) \leq K\}} F_1^a(-x, \omega)$ for all $x \in [-1, -1 + \epsilon)$. Moreover, one can define $F_0^b(x)$ such that

$$\pi_0^b = \sup_{x \leq 0} \left\{ F_0^b(-x) + \frac{\ell^{(cb)}(x)}{1 + r_1^a} \right\} = F_0^b(-x^*) - \frac{\sup_{\omega \in \{F_1^b(1) \leq K\}} F_1^a(-x^*, \omega)}{1 + r_1^a}$$

for some $x^* \in [-1, -1+\epsilon)$. This way, we have that the price is going to be constant for all K in some open interval. A similar result applies for the buyer's price of a put.

Suppose now that $P\{F_1^b(1) > K\} = 1$ and take the viewpoint of the seller of a call. Using Theorem 4.2 and setting $x = 1$, one finds that

$$\pi_0^a \leq F_0^a(1) - \frac{K}{1 + r_1^b}.$$

Consequently, if $F_0^a(1) < \frac{K}{1+r_1^b}$, then $\pi_0^a < 0$. One can build a simple example to illustrate this result.

Let $\Omega = \{u, d\}$ where u and d stand for an upward movement and a downward movement in the price and set $r_1^a = r_1^b = 0$. From Theorem 4.1 one needs that $f_0^b < f_1^a(0, u)$ and $f_0^a > f_1^b(0, d)$. Then, assuming the price satisfy the following inequalities

$$F_1^b(1, d) < F_0^b(1) < F_1^a(1, d) < F_1^b(1, u) < K < F_0^a(1) < F_1^a(1, u)$$

one finds that $\pi_0^a < 0$ from Theorem 4.2.

From a financial point, the seller knows that the buyer will exercise. To hedge her position the seller pays $F_0^a(1)$, but she will receive $\frac{K}{1+r_1^b}$ at $t = 1$ when the buyer will exercise the option, which covers the hedging cost. One can get a similar result for the seller of a put by taking $P\{F_1^a(1) < K\} = 1$ and $\frac{K}{1+r_1^a} < F_0^b(1)$. Although it does not satisfy our definition of arbitrage in 2.2, this situation corresponds to the intuitive idea of arbitrage since it is an opportunity of riskless profit. However, the investor cannot exploit this arbitrage opportunity to create an arbitrary large profit, as it is the case for arbitrage in a perfectly liquid market. In order to make a larger profit, the seller could sell $n > 1$ call options and covers her position in buying n shares of the underlying stock at a cost $F_T^a(n)$. However, although she realizes a profit with one call option, recall that $F_1^a(1) < \frac{K}{1+r_1^b}$, it is not true in general that $F_1^a(n) < n\frac{K}{1+r^b}$ since $\frac{F_1^a(n)}{n}$ is non-decreasing in n. As we see, the addition of the option as an investment opportunity creates a new possibility of arbitrage. The fact that including options in the portfolio might creates new arbitrage opportunities in a model of illiquidity were studied in Jarrow (1994).

4.5. Example and numerical results

In this section, we present an example where we price European options of maturity $T = 1$ within a two parameter model for the structure of the order book and with interest rates r_1^a for the lender and r_1^b the borrower. We suppose that $\Omega = \{\omega_1, \omega_2\}$ where ω_1 and ω_2 correspond to downward and upward movement in the prices. It is possible to see this example as a generalisation of the classical one period binomial model under the perfect liquidity assumption. The structure of the limit order book at time $t = 0$ is defined as follows :

$$\begin{aligned} F_0^a(x) &= S_0^a \left\{(1+x)^{\alpha_0} - 1\right\}, \\ F_0^b(x) &= S_0^b \left\{(1+x)^{\beta_0} - 1\right\}, \end{aligned}$$

where $0 < S_0^b \leq S_0^a$, $\alpha_0 \geq 1$ and $\beta_0 \in (0, 1)$, while

$$\begin{aligned} F_1^a(x,\omega) &= S_1^a(\omega) \left\{(1+x)^{\alpha_1(\omega)} - 1\right\}, \\ F_1^b(x,\omega) &= S_1^b(\omega) \left\{(1+x)^{\beta_1(\omega)} - 1\right\}, \end{aligned}$$

with $0 < S_1^b(\omega_1) \leq S_1^b(\omega_2)$, $0 < S_1^a(\omega_1) \leq S_1^a(\omega_2)$, $\alpha_1(\omega_1) \leq \alpha_1(\omega_2)$, $\beta_1(\omega_1) \leq \beta_1(\omega_2)$, $\alpha_1(\omega) \geq 1$ and $\beta_1(\omega) \in (0, 1]$.

We also suppose that (**C2**)-(**C1**), (**C2′**)-(**C1′**) as well as the no-arbitrage conditions in Theorem 4.1 are satisfied. We note that setting $S_t^a = S_t^b$, $\alpha_t = \beta_t = 1$ and $r_1^a = r_1^b$ one recovers the binomial model.

The first set of parameters satisfying these conditions is given in Table 4.1.

In Figure 4.1, we show the price of the call and the hedging strategy for the seller and the buyer for different values of the strike. The buyer exercises with probability one if $K \leq 0.953$ and probability zero if $K \geq 1.0059$. As we mentioned in Section 4.4, Figure 4.1 provides an example where the buyer's price depends non-linearly on the strike. The parameters of the model are given in Table 4.1.

TABLE 4.1. Parameters used in the first example.

	S_0^a	α_0	α_1	r_1^a	S_1^a	S_0^b	β_0	β_1	r_1^b	S_1^b
ω_1	1	1.01	1.01	0.03	0.98	1	0.99	0.98	0.04	0.98
ω_2	1	1.01	1.02	0.03	1.02	1	0.99	0.99	0.04	1.02

FIGURE 4.1. The value of the call for different strike values is displayed on the left panel, while one the right panel, the number of shares X_1 for the hedging portfolios are displayed as a function of the strike. For both graphs, the dashed lines denotes the seller while the plain line is for the buyer.

In Figure 4.2 we show the effect of the convexity and concavity of the limit order book structure. We start from the binomial model for a perfectly liquid market (parameters are given in Table 4.2) and we change the value of $\alpha_1(\omega)$ from 1 to 1.3 while $\beta_1(\omega)$ varies from 1 to 0.97, at which point the buyer exercises with probability zero.

TABLE 4.2. Parameters used in the first example.

	S_0^a	α_0	α_1	r_1^a	S_1^a	S_0^b	β_0	β_1	r_1^b	S_1^b
ω_1	1	1	1	0.01	0.98	1	1	1	0.01	0.98
ω_2	1	1	1	0.01	1.02	1	1	1	0.01	1.02

FIGURE 4.2. The top figure shows the price of the call for $\alpha_1(\omega)$ going from 1 to 1.3 while $\beta_1(\omega)$ goes from 1 to 0.97. The scale of the x-axis is $\beta_1(\omega)$. The dashed line give the seller's price while the plain line gives the buyer's price. The two following figures are the position in the stock for hedging the option for the seller and the buyer.

5. Pricing an option with primal-dual optimization

In this last section we show that we can use primal-dual optimization to express the bid and ask prices of European call and put options as optimization problems related to probability measures. We refer the reader to Boyd and Vandenberghe (2009) for details about primal-dual optimization.

To simplify the presentation, we only consider the case of the seller of a call. However, all the results of Theorems 4.2 to 4.5 can be deduced using primal-dual optimization.

Consider the problem of pricing a European call with the same characteristics as in Section 4.2 and assume that the probability space is given by $\Omega = \{\omega_i; i \in I\}$ where I is finite, and $P(\omega_i) > 0$ for all $i \in I$.

Recall that the value of the hedging portfolio for the seller is given by,

$$V_1(x, \pi_0) = \begin{cases} F_1^b(x) - F_1^a(-x) + \ell_0(y), & \text{if } F_1^b(1) \leq K, \\ F_1^b(x-1) - F_1^a(-x+1) + \ell_0(y) + K, & \text{if } F_1^b(1) > K, \end{cases}$$

where $y = \pi_0 - \{F_0^a(x) - F_0^b(x)\}$, and $\ell_0(y) = y^+(1 + r_1^a) - y^-(1 + r_1^b)$. We also define the set $B_K = \{i \in I; F_1^b(1, \omega_i) > K\}$.

The pricing problem, which is the primal form of the optimization problem, is written as

$$\pi_0^a = \inf_{x,y \in \mathbb{R}} \left\{ F_0^a(x) - F_0^b(-x) + y \right\},$$

subject to the constraints $\ell^{(i)}(x) + \ell_0(y) \geq 0$ for all $i \in I$, where

$$\ell^{(i)}(x) = \begin{cases} F_1^b(x, \omega_i), & i \in B_K^c, x \geq 0, \\ -F_1^a(-x, \omega_i), & i \in B_K^c, x < 0, \\ F_1^b(x-1, \omega_i) + K, & i \in B_K, x \geq 1, \\ -F_1^a(-x+1, \omega_i) + K, & i \in B_K, x < 1. \end{cases}$$

To construct the dual problem, we define the Lagrangian $H(x, y, \xi)$ for $\xi \geq 0$ by

$$H(x, y, \xi) = F_0^a(x) - F_0^b(-x) + y - \ell_0(y) \sum_{i \in I} \xi_i - \sum_{i \in I} \ell^{(i)}(x) \xi_i.$$

Further, let the Lagrange dual function be defined by $G(\xi) = \inf_{x,y} H(x, y, \xi)$. Then the dual problem is to find $\sup_{\xi \geq 0} G(\xi)$. In general on has that $\sup_{\xi \geq 0} G(\xi) \leq \pi_0^a$.

However, in the primal problem, the function to optimize is convex, the constraints are concave and by taking $x = 0$ and $y > 0$ one has that $\ell^{(i)}(0) + \ell_0(y) > 0$ for all $i \in I$. These conditions are sufficient to have strong duality, which means that

$$\pi_0^a = \sup_{\xi \geq 0} G(\xi). \tag{5.1}$$

These conditions for strong duality are known as Slater's theorem ; see, e.g., Boyd and Vandenberghe (2009)[Section 5.2.3].

Next, since

$$y - \ell_0(y) \sum_{i \in I} \xi_i = y \left[1 - \left\{ (1 + r_1^a) \mathbf{1}(y)_{[0,\infty)} - (1 + r_1^b) \mathbf{1}_{(-\infty,0)} \right\} \sum_{i \in I} \xi_i \right],$$

it follows that $G(\xi) = -\infty$ whenever $\sum_{i \in I} \xi_i < \frac{1}{1+r_1^b}$ or $\sum_{i \in I} \xi_i > \frac{1}{1+r_1^a}$. Let $\mathcal{D} = \left\{ \xi \geq 0; \sum_{i \in I} \xi_i \in \left[\frac{1}{1+r_1^b}, \frac{1}{1+r_1^a} \right] \right\}$. Note that $\mathcal{D}$ is compact.

Using (5.1), one may assume that $\xi \in \mathcal{D}$, from which it follows that $G(\xi) = \inf_x H(x, 0, \xi)$.

At this point, we would like to simplify the expression of the Lagrange dual problem by restricting the domain of ξ and the domain of x for the function $H(x, 0, \xi)$. To this end, let $\mathcal{H} = \{\xi \in \mathcal{D}; \inf_x H(x, 0, \xi) > -\infty\}$. As a result,

$$\pi_0^a = \sup_{\xi \in \mathcal{H}} \inf_x H(x, 0, \xi).$$

We first try to describe $\mathcal{H}$ more precisely. Let $f_t^b(\infty) = \lim_{x \to \infty} \frac{F_t^b(x)}{x}$ and $f_t^a(\infty) = \lim_{x \to \infty} \frac{F_1^a(x)}{x}$. Note that both limits exist, $f_t^b(\infty) < \infty$ while $f_t^a(\infty)$ might be infinite. Thus

$$\lim_{x \to \infty} \frac{H(x, 0, \xi)}{x} = f_0^a(\infty) - \sum_{i \in I} \xi_i f_1^b(\infty, \omega_i). \tag{5.2}$$

Similarly

$$\lim_{x \to -\infty} \frac{H(x, 0, \xi)}{-x} = \sum_{i \in I} \xi_i f_1^a(\infty, \omega_i) - f_0^b(\infty). \tag{5.3}$$

Using (5.2) and (5.3), one finds that $G(\xi) > -\infty$ implies that ξ satisfies

$$f_0^a(\infty) \geq \sum_{i \in I} \xi_i f_1^b(\infty, \omega_i) \text{ and } f_0^b(\infty) \leq \sum_{i \in I} \xi_i f_1^a(\infty, \omega_i).$$

In addition, if

$$f_0^a(\infty) > \sum_{i \in I} \xi_i f_1^b(\infty, \omega_i) \text{ and } f_0^b(\infty) < \sum_{i \in I} \xi_i f_1^a(\infty, \omega_i),$$

then $G(\xi) > -\infty$.

As a result, one can replace $\mathcal{H}$ by

$$\tilde{\mathcal{H}} = \left\{ \xi \in \mathcal{D}; f_0^a(\infty) \geq \sum_{i \in I} \xi_i f_1^b(\infty, \omega_i) \text{ and } f_0^b(\infty) \leq \sum_{i \in I} \xi_i f_1^a(\infty, \omega_i) \right\},$$

so that

$$\pi_0^a = \sup_{\xi \in \tilde{\mathcal{H}}} G(\xi).$$

We now show that we can sometimes restrict the problem to definition of $G(\xi)$ to the infimum over the set $x \in [0,1]$.

First, note that for any $\xi \in \mathcal{D}$, as $x \downarrow 1$,

$$\begin{aligned} \frac{H(x,0,\xi) - H(1,0,\xi)}{x-1} &\downarrow f_0^a(1) - \sum_{i \in B_K^c} \xi_i f_1^b(1,\omega_i) - \sum_{i \in B_K} \xi_i f_1^b(0,\omega_i) \\ &\geq f_0^a(0) - \frac{1}{1+r_1^a} \max_{i \in I} f_1^b(0,\omega_i). \end{aligned}$$

So if

$$f_0^a(0) \geq \frac{1}{1+r_1^a} \max_{i \in I} f_1^b(0,\omega_i), \tag{5.4}$$

then $\inf_{x \geq 1} H(x,0,\xi) = H(1,0,\xi)$.

Similarly, as $x \uparrow 0$,

$$\begin{aligned} \frac{H(x,0,\xi) - H(0,0,\xi)}{-x} &\downarrow -f_0^b(0) + \sum_{i \in B_K^c} \xi_i f_1^a(0,\omega_i) + \sum_{i \in B_K} \xi_i f_1^a(1,\omega_i) \\ &\geq \frac{1}{1+r_1^b} \min_{i \in I} f_1^a(0,\omega_i) - f_0^b(0). \end{aligned}$$

As a result, if

$$\frac{1}{1+r_1^b} \min_{i \in I} f_1^a(0,\omega_i) \geq f_0^b(0), \tag{5.5}$$

then $\inf_{x \leq 0} H(x,0,\xi) = H(0,0,\xi)$.

It follows that under (5.4)–(5.5),

$$G(\xi) = \inf_{x \in [0,1]} \left\{ F_0^a(x) - \sum_{i \in B_K^c}^{m} \xi_i F_1^b(x,\omega_i) + \sum_{i \in B_K} \xi_i F_1^a(1-x,\omega_i) \right\} - K \sum_{i \in B_K} \xi_i,$$

for any $\xi \in \mathcal{D}$.

5.1. The dual problem and its relationship with expectations

Let $\mathcal{Q}$ be the set of probability measures. Then to any $\xi \in \mathcal{D}$, one can associate a unique couple $(Q,r) \in \mathcal{Q} \times [r_1^a, r_1^b]$ so that $\sum_{i \in I} \xi = \frac{1}{1+r}$ and $Q(\omega_i) = (1+r)\xi_i$, $i \in I$.

Next, set

$$\mathcal{L}(x,\omega) = \begin{cases} F_1^b(x,\omega), & \omega \in B_K^c, x \geq 0, \\ -F_1^a(-x,\omega), & \omega \in B_K^c, x < 0, \\ F_1^b(x-1,\omega) + K, & \omega \in B_K, x \geq 1, \\ -F_1^a(-x+1,\omega) + K, & \omega \in B_K, x < 1. \end{cases}$$

It then follows that $H(x,0,\xi)$ can be written as

$$H(x,0,\xi) = F_0^a(x) - F_0^b(-x) - \frac{1}{1+r} E^Q \left\{ \mathcal{L}(x) \right\}.$$

One can redefine $\tilde{\mathcal{H}}$ in terms of (Q,r) viz.

$$\begin{aligned} \tilde{\mathcal{H}} &= \Bigg\{ (Q,r) \in \mathcal{Q} \times [r_1^a, r_1^b]; E^Q \left\{ \frac{f_1^b(\infty)}{1+r} \right\} \leq f_0^a(\infty) \\ &\qquad \text{and} \quad E^Q \left\{ \frac{f_1^a(\infty)}{1+r} \right\} \geq f_0^b(\infty) \Bigg\}. \end{aligned} \tag{5.6}$$

Finally, one gets

$$\pi_0^a = \sup_{(Q,r)\in\tilde{\mathcal{H}}} \inf_x \left[F_0^a(x) - F_0^b(-x) - \frac{1}{1+r} E^Q \left\{ \mathcal{L}(x) \right\} \right]. \tag{5.7}$$

Example 5.1. *Assume a perfectly liquid market with no bi-ask spread. Then one has* $F_0^a(x) = F_0^b(x) = S_0 x$, $F_1^a(x) = F_1^b(x) = S_1 x$.

In this case, $\tilde{\mathcal{H}}$ *is the set of probability measures such that the discounted price is a martingale. Also,* $\mathcal{L}(x) = S_1 x - (S_1 - K)^+$, *so*

$$\begin{aligned} \pi_0^a &= \sup_{Q\in\tilde{\mathcal{H}}} \inf_x \left[\frac{1}{1+r} E^Q \left\{ (S_1 - K)^+ \right\} + x \left\{ S_0 - \frac{1}{1+r} E^Q(S_1) \right\} \right] \\ &= \sup_{Q\in\tilde{\mathcal{H}}} \frac{1}{1+r} E^Q \left\{ (S_1 - K)^+ \right\}. \end{aligned}$$

6. Conclusion

In this paper, we studied a market model where illiquidity is modeled through the structure of the limit order book. We provided sufficient conditions to rule out arbitrage for our general model and sufficient and necessary conditions for the one period version. The conditions for the absence of arbitrage of Theorem (2.2) are a weaker version of those given in the continuous time model of Simard (2014). But again they show that the theory based on equivalent martingale measures of perfectly liquid market is not general enough for illiquidity model. Sufficient and necessary conditions for the absence of arbitrage in a general illiquid market model are yet to be found.

Also, we discussed the pricing of one-period European call and put options. The literature about options pricing under illiquidity assumptions is mainly focused on continuous time models and the solutions that are given are rarely explicit. Alike the binomial model in perfectly liquid market, our framework allows for a more intuitive approach and simpler solutions. By showing more explicitly how the different inputs of the model affect the price and hedging strategies, our solutions provide a better understanding of the impact of illiquidity in options pricing so that we could highlight some particularities of the results. The last section also demonstrate that the pricing problem can be written as a problem of maximizing an expectation over a particular set of probability measures.

A similar approach for options with longer maturity should be considered. However, since we used different interest rates for lending and borrowing, at each time step, the problem of option pricing required to split into cases where the bank account is positive or negative. The complexity it creates might forbid our method to be used in a multiperiod setting, so that more powerful tools might be required.

A. Additional notations

To write the proofs in a more compact form, we introduce additional notations.

The liquidation value of the portfolio is denoted by

$$L_t(X) = F_t^b(X_t) - F_t^a(-X_t)$$

for $t \geq 1$, while the cash flow of a transaction x at time t is given by

$$c_t(X) = F_t^a(\Delta X_{t+1}^a) - F_t^b(\Delta X_{t+1}^b).$$

Recall that $\ell_{t-1}(y) = y^+(1+r_t^a) - y^-(1+r_t^b)$. The amount of money in the bank account is written as

$$Y_t = \ell_{t-1}(Y_{t-1}) - c_t(X)$$

with $Y_0 = \pi_0 - c_0(X)$. Finally, the value of the portfolio at time t is given by

$$V_t(X, \pi_0) = L_t(X) + \ell_{t-1}(Y_{t-1})$$

where $V_0(X, \pi_0) = \pi_0$.

B. Proof of Theorem 2.1

For sake of simplicity, the subscript Q is eliminated from the expectations.

B.1. Sufficiency

First, one shows that conditions (2.9)–(2.10) are sufficient. Note that

$$L_T(X) = F_T^b(X_T) - F_T^a(-X_T) \leq X_T^+ f_T^b - X_T^- f_T^a, \tag{B.1}$$

since $F_T^a(x) \geq x^+ f_T^a$ and $F_T^b(x) \leq x^+ f_T^b$, using the convexity and concavity of F^a and F^b respectively. Similarly, if $\tilde{c}_i(X) = f_i^a\,(\Delta X_{i+1})^+ - f_i^b\,(\Delta X_{i+1})^-$, we have

$$c_i(X) = F_i^a\,(\Delta X_{i+1}) - F_i^b\,(-\Delta X_{i+1}) \geq \tilde{c}_i(X). \tag{B.2}$$

Also, it is easy to see that $\ell_{i-1}(y) \leq y(1+r_i)$, for all $i \geq 1$. Hence, using (2.5), one finds that

$$\begin{aligned} \ell_{t-1}(Y_{t-1}) &\leq \frac{\pi_0}{\beta_t} - \frac{1}{\beta_t}\sum_{i=0}^{t-1}\beta_i c_i(X) \qquad \text{(B.3)} \\ &\leq \frac{\pi_0}{\beta_t} - \frac{1}{\beta_t}\sum_{i=0}^{t-1}\beta_i \tilde{c}_i(X). \qquad \text{(B.4)} \end{aligned}$$

Note that (B.3) is indeed an equality when $r^a = r^b$.

Next, using (B.2) and (B.4), one obtains

$$\begin{aligned} \beta_T V_T(X,\pi_0) &= \beta_T L_T(X) + \beta_T \ell_{T-1}(Y_{T-1}) \\ &\leq \pi_0 + \beta_T f_T^b X_T^+ - \beta_T f_T^a X_T^- \\ &\quad - \sum_{i=0}^{T-1} \beta_i \left\{ f_i^a (\Delta X_{i+1})^+ - f_i^b (\Delta X_{i+1})^- \right\}. \end{aligned}$$

For the next step, we use induction to show that

$$\begin{aligned} \pi_0 \quad &+ \quad E\left(\beta_T f_T^b X_T^+ - \beta_T f_T^a X_T^- | \mathcal{F}_i\right) - \sum_{j=0}^{T-1} E\left(\beta_j \tilde{c}_j(X) | \mathcal{F}_i\right) \\ &\leq \pi_0 + X_i^+ H_i^b - X_i^- H_i^a - \sum_{j=0}^{i-1} \beta_j \tilde{c}_j(X). \end{aligned}$$

As a result, one may assume that $\pi_0 = 0$ and $\beta \equiv 1$. For sake of simplicity set $V_T = V_T(X, 0)$.

Under assumptions (2.9)–(2.10), we can prove by induction that

$$E(V_T|\mathcal{F}_i) \leq X_i^+ H_i^b - X_i^- H_i^a - \sum_{j=0}^{i-1} \tilde{c}_j(X), \quad i \in \{1, \dots, T\}. \qquad \text{(B.5)}$$

Note that the expectation in (B.5) is finite by assumption and by the definition of an admissible strategy. The result in (B.5) is trivial for $i = T$. Suppose now that (B.5) holds true for i. Then one will prove that (B.5) holds true for $i-1$. To do so, one must consider the following six cases :

(I1) $X_{i-1} \geq 0$, $\Delta X_i \geq 0$;

(I2) $X_{i-1} \geq 0$, $\Delta X_i \leq 0$, $X_i \geq 0$;

(I3) $X_{i-1} \geq 0$, $\Delta X_i \leq 0$, $X_i \leq 0$;

(I4) $X_{i-1} \leq 0$, $\Delta X_i \geq 0$, $X_i \geq 0$;

(I5) $X_{i-1} \leq 0$, $\Delta X_i \geq 0$, $X_i \leq 0$;

(I6) $X_{i-1} \leq 0$, $\Delta X_i \leq 0$.

So assume that $E\left(V_T|\mathcal{F}_i\right) \leq X_i^+ H_i^b - X_i^- H_i^a - \sum_{j=0}^{i-1} \tilde{c}_j(X)$.

(1) If (**I1**) holds, then

$$\begin{aligned}E\left(V_T|\mathcal{F}_i\right) &\leq X_i^+H_i^b - X_i^-H_i^a - \sum_{j=0}^{i-1}\tilde{c}_j(X) \\ &= X_i^+H_i^b - f_{i-1}^a\left(\Delta X_i\right)^+ - \sum_{j=0}^{i-2}\tilde{c}_j(X) \\ &= X_{i-1}^+H_i^b + \left(\Delta X_i\right)^+\left(H_i^b - f_{i-1}^a\right) - \sum_{j=0}^{i-2}\tilde{c}_j(X).\end{aligned}$$

Consequently, using (2.10) and the very definition of H^b, one gets

$$\begin{aligned}E\left(V_T|\mathcal{F}_{i-1}\right) &\leq X_{i-1}^+E\left(H_i^b|\mathcal{F}_{i-1}\right) + \left(\Delta X_i\right)^+E\left(H_i^b - f_{i-1}^a|\mathcal{F}_{i-1}\right) \\ &\quad - \sum_{j=0}^{i-2}\tilde{c}_j(X) \\ &\leq X_{i-1}^+H_{i-1}^b - \sum_{j=0}^{i-2}\tilde{c}_j(X) \\ &= X_{i-1}^+H_{i-1}^b - X_{i-1}^-H_{i-1}^a - \sum_{j=0}^{i-2}\tilde{c}_j(X).\end{aligned}$$

(2) If (**I2**) holds, then

$$E\left(V_T|\mathcal{F}_i\right) \leq X_i^+H_i^b + f_{i-1}^b\left(\Delta X_i\right)^- - \sum_{j=0}^{i-2}\tilde{c}_j(X).$$

As a result, from the very definition of H^b, one obtains

$$\begin{aligned}E\left(V_T|\mathcal{F}_{i-1}\right) &\leq X_i^+E\left(H_i^b|\mathcal{F}_{i-1}\right) + \left(\Delta X_i\right)^-f_{i-1}^b - \sum_{j=0}^{i-2}\tilde{c}_j(X) \\ &\leq X_i^+H_{i-1}^b + \left(\Delta X_i\right)^-f_{i-1}^b - \sum_{j=0}^{i-2}\tilde{c}_j(X) \\ &= X_{i-1}^+H_{i-1}^b + \left(\Delta X_i\right)^-\left(f_{i-1}^b - H_{i-1}^b\right) - \sum_{j=0}^{i-2}\tilde{c}_j(X) \\ &\leq X_{i-1}^+H_{i-1}^b - X_{i-1}^-H_{i-1}^a - \sum_{j=0}^{i-2}\tilde{c}_j(X).\end{aligned}$$

(3) If (**I3**) holds, then

$$E\left(V_T|\mathcal{F}_i\right) \leq -X_i^-H_i^a + \left(\Delta X_i\right)^-f_{i-1}^b - \sum_{j=0}^{i-2}\tilde{c}_j(X)$$

Consequently, using (2.9) and the very definition of H^b,

$$\begin{aligned}
E\left(V_T|\mathcal{F}_{i-1}\right) &\leq -X_i^- E\left(H_i^a|\mathcal{F}_{i-1}\right) + (\Delta X_i)^- f_{i-1}^b - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\
&\leq -X_i^- f_{i-1}^b + (\Delta X_i)^- f_{i-1}^b - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\
&= X_{i-1}^+ f_{i-1}^b - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\
&\leq X_{i-1}^+ H_{i-1}^b - X_{i-1}^- H_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X).
\end{aligned}$$

(4) Under (**I4**), one has

$$E\left(V_T|\mathcal{F}_i\right) \leq X_i^+ H_i^b - (\Delta X_i)^+ f_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X).$$

Hence, using (2.10) and the very definition of H^a,

$$\begin{aligned}
E\left(V_T|\mathcal{F}_{i-1}\right) &\leq X_i^+ E\left(H_i^b|\mathcal{F}_{i-1}\right) - (\Delta X_i)^+ f_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\
&\leq X_i^+ f_{i-1}^a - (\Delta X_i)^+ f_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\
&= -X_{i-1}^- f_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\
&\leq X_{i-1}^+ H_{i-1}^b - X_{i-1}^- H_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X).
\end{aligned}$$

(5) If (**I5**) is true, then

$$E\left(V_T|\mathcal{F}_i\right) \leq -X_i^- H_i^a - (\Delta X_i)^+ f_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X).$$

Consequently, from the very definition of H^a,

$$\begin{aligned}
E\left(V_T|\mathcal{F}_{i-1}\right) &\leq -X_i^- E\left(H_i^a|\mathcal{F}_{i-1}\right) - (\Delta X_i)^+ f_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\
&\leq -X_i^- H_{i-1}^a - (\Delta X_i)^+ f_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\
&= -X_{i-1}^- H_{i-1}^a - (\Delta X_i)^+ \left(f_{i-1}^a - H_{i-1}^a\right) - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\
&\leq -X_{i-1}^- H_{i-1}^a - \sum_{j=0}^{i-2} \tilde{c}_j(X)
\end{aligned}$$

$$= \quad X_{i-1}^{+} H_{i-1}^{b} - X_{i-1}^{-} H_{i-1}^{a} - \sum_{j=0}^{i-2} \tilde{c}_j(X).$$

(6) Finally, if (**I6**) holds true, then

$$E\left(V_T|\mathcal{F}_i\right) \leq -X_i^{-} H_i^{a} + f_{i-1}^{b} \beta_{i-1} \left(\Delta X_i\right)^{-} - \sum_{j=0}^{i-2} \tilde{c}_j(X).$$

Consequently, using (2.9) and the very definition of H^a,

$$\begin{aligned} E\left(V_T|\mathcal{F}_{i-1}\right) &\leq -X_{i-1}^{-} E\left(H_i^{a}|\mathcal{F}_{i-1}\right) - \left(\Delta X_i\right)^{-} E\left(H_i^{a} - f_{i-1}^{b}|\mathcal{F}_{i-1}\right) \\ &\quad - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\ &\leq -X_{i-1}^{-} H_{i-1}^{a} - \sum_{j=0}^{i-2} \tilde{c}_j(X) \\ &= X_{i-1}^{+} H_{i-1}^{b} - X_{i-1}^{-} H_{i-1}^{a} - \sum_{j=0}^{i-2} \tilde{c}_j(X). \end{aligned}$$

It follows from the previous 6 cases that

$$E\left(V_T|\mathcal{F}_{i-1}\right) \leq X_{i-1}^{+} H_{i-1}^{b} - X_{i-1}^{-} H_{i-1}^{a} - \sum_{j=0}^{i-2} \tilde{c}_j(X),$$

completing the proof of (B.5). In particular, taking $i = 1$ in (B.5), one obtains

$$\begin{aligned} E\left(V_T|\mathcal{F}_1\right) &\leq X_1^{+} H_1^{b} - X_1^{-} H_1^{a} - (\Delta X_1)^{+} f_0^{a} + (\Delta X_1)^{-} f_0^{b} \\ &= X_1^{+}(H_1^{b} - f_0^{a}) - X_1^{-}(H_1^{a} - f_0^{b}). \end{aligned}$$

As a result, $E(V_T) \leq 0$, using the last inequality together with (2.9)–(2.10). This proves that these conditions are sufficient.

B.2. Necessity

One will now prove that if $r_i = r_i^a = r_i^b$ for all $i \in \{1, \ldots, T\}$, then conditions (2.9)–(2.10) are necessary to have $E\{\beta_T V_T(X, \pi_0)\} \leq \pi_0$ for all admissible strategy X.

Before going further, we need the following result. From the dominated convergence theorem, one has that, for all $t \geq 0$,

$$\lim_{n\to\infty} E\left\{nF_t^{a}(1/n)\right\} = E(f_t^{a}) \tag{B.6}$$

and

$$\lim_{n\to\infty} E\left\{nF_t^{b}(1/n)\right\} = E(f_t^{b}), \tag{B.7}$$

since $0 \leq nF_t^b(1/n) \leq nF_t^a(1/n) \leq F_T^a(1)$ for all $n \geq 1$ and $E_Q\left\{F_T^a(1)\right\} < \infty$ by assumption and (P2).

Recall from (B.3) and the assumption $r^a \equiv r \equiv r^b$ that

$$\ell_{i-1}(Y_{i-1}) = \frac{\pi_0}{\beta_i} - \frac{1}{\beta_i}\sum_{j=0}^{i-1} \beta_j c_j(X), \quad i = 1, \ldots, T,$$

so

$$\beta_T V_T(X) = \pi_0 + \beta_T\{F_T^b(X_T) - F_T^a(-X_T)\} - \sum_{j=0}^{T-1} \beta_j c_j(X). \tag{B.8}$$

Based on (B.8), there is no loss of generality in assuming that $\pi_0 = 0$.

First, take $X_i = 0$, $i \in \{1, \ldots, T-1\}$, and $X_T = \mathbf{1}_A x$ with $A \in \mathcal{F}_{T-1}$.

On one hand, if $x \geq 0$,

$$\begin{aligned} \beta_T V_T(X) &= \beta_T\left\{F_T^b(\mathbf{1}_A x) - F_{T-1}^a(\mathbf{1}_A x)(1 + r_T)\right\} \\ &= \mathbf{1}_A\left\{\beta_T F_T^b(x) - \beta_{T-1} F_{T-1}^a(x)\right\}. \end{aligned}$$

By taking the limit $\lim_{X \downarrow 0} E\left\{X^{-1}\beta_T V_T(X)\right\}$ and using (B.6) and (B.7), one finds that $E\left\{\beta_T V_T(X)\right\} \leq 0$ implies that $E\left\{\mathbf{1}_A\left(\beta_T f_T^b - \beta_{T-1} f_{T-1}^a\right)\right\} \leq 0$ for any $A \in \mathcal{F}_{T-1}$. Hence, $E\left(\beta_T f_T^b | \mathcal{F}_{T-1}\right) \leq \beta_{T-1} f_{T-1}^a$, proving (2.10) for $t = T$.

On the other hand, if $x = -y \leq 0$, then

$$\beta_T V_T(X) = \mathbf{1}_A\left\{-\beta_T F_T^a(y) + \beta_{T-1} F_{T-1}^b(y)\right\},$$

so, using (B.6) and (B.7) again, $E\left\{\beta_T V_T(X)\right\} \leq 0$ implies that

$$E\left\{\mathbf{1}_A\left(\beta_T f_T^a - \beta_{T-1} f_{T-1}^b\right)\right\} \geq 0$$

for any $A \in \mathcal{F}_{T-1}$. Hence,

$$E\left(\beta_T f_T^a | \mathcal{F}_{T-1}\right) \geq \beta_{T-1} f_{T-1}^b,$$

proving (2.9) for $t = T$.

Next, take $A \in \mathcal{F}_{T-2}$, and let ζ be a $\mathcal{F}_{T-1}$-measurable random variable with $\zeta \in [0, 1]$. Set $X_i = 0$, $i \in \{1, \ldots, T-2\}$, $X_{T-1} = x\mathbf{1}_A$, and $X_T = x(1-\zeta)\mathbf{1}_A$. Hence $\Delta X_T = -x\zeta\mathbf{1}_A$ and $\Delta X_{T-1} = x\mathbf{1}_A$.

Fox $x \geq 0$, one gets,

$$\beta_T V_T(X) = \beta_T \mathbf{1}_A \left\{F_T^b(x - x\zeta) - F_{T-2}^a(x)(1 + r_{T-1})(1 + r_T)\right.$$

$$\begin{aligned} & +F^b_{T-1}(x\zeta)(1+r_T)\Big\} \\ = \;& \mathbf{1}_A \left\{ \beta_T F^b_T(x - x\zeta) - \beta_{T-2} F^a_{T-2}(x) + \beta_{T-1} F^b_{T-1}(x\zeta) \right\}. \end{aligned}$$

Hence, $E\{\beta_T V_T(X)\} \le 0$ implies that

$$E\left[\mathbf{1}_A \left\{(1-\zeta)\beta_T f^b_T + \zeta\beta_{T-1} f^b_{T-1} - \beta_{T-2} f^a_{T-2}\right\}\right] \le 0.$$

Since the latter holds for any $\zeta \in [0,1]$ with ζ being $\mathcal{F}_{T-1}-$measurable, on may conclude that

$$(1-\zeta)E\left(\beta_T f^b_T | \mathcal{F}_{T-1}\right) + \zeta\beta_{T-1} f^b_{T-1} \le \max\left\{E\left(\beta_T f^b_T | \mathcal{F}_{T-1}\right), \beta_{T-1} f^b_{T-1}\right\},$$

and the upper bound is attained by setting $\zeta = \mathbf{1}_{\left\{\beta_{T-1} f^b_{T-1} \ge E\left(\beta_T f^b_T | \mathcal{F}_{T-1}\right)\right\}}$. As a result, $E\left(H^b_{T-1} | \mathcal{F}_{T-2}\right) \le \beta_{T-2} f^a_{T-2}$, proving (2.10) for $t = T-1$.

For $x = -y \le 0$, $X_T = -y(1-\zeta)\mathbf{1}_A$ and $X_{T-1} = -y\mathbf{1}_A$ one gets

$$\beta_T V_T(X) = \mathbf{1}_A \left\{-\beta_T F^a_T(y - y\zeta) - \beta_{T-1} F^a_{T-1}(y\zeta) + \beta_{T-2} F^b_{T-2}(y)\right\}.$$

Hence, $E\left(\{\beta_T V_T(X)\} \le 0\right.$ implies that

$$E\left[\mathbf{1}_A \left\{(1-\zeta)E\left(\beta_T f^a_T | \mathcal{F}_{T-1}\right) + \zeta\beta_{T-1} f^a_{T-1} - \beta_{T-2} f^b_{T-2}\right\}\right] \ge 0.$$

Since the latter holds for any $\zeta \in [0,1]$ and ζ is $\mathcal{F}_{T-1}$-measurable, one may conclude that

$$E\left[\mathbf{1}_A \left[\min\left\{E\left(\beta_T f^a_T | \mathcal{F}_{T-1}\right), \beta_{T-1} f^a_{T-1}\right\} - \beta_{T-2} f^b_{T-2}\right]\right] \ge 0,$$

and it follows that $E\left(H^a_{T-1} | \mathcal{F}_{T-2}\right) \ge \beta_{T-2} f^b_{T-2}$, proving (2.9) for $t = T-1$.

Finally, for a given $i \in \{1, \ldots, T-2\}$, take $A \in \mathcal{F}_{i-1}$ and set $X_i = x\mathbf{1}_A$, with $X_j = 0$ for $j < i$. Further, set $\Delta X_j = -x\zeta_j \mathbf{1}_A$, for $j = i+1, \ldots, T$, where $\zeta_j \ge 0$, ζ_j is $\mathcal{F}_{j-1}-$measurable and $\sum_{j=i+1}^T \zeta_j \le 1$. Then, $X_T = x\mathbf{1}_A\left(1 - \sum_{j=i+1}^T \zeta_j\right)$. If $x \ge 0$, then

$$\begin{aligned} \beta_T V_T(X) \;=\; & \mathbf{1}_A \Bigg\{ \beta_T F^b_T \left(x - x\sum_{j=i+1}^T \zeta_j\right) + \sum_{j=i}^{T-1} \beta_j F^b_j(x\zeta_{j+1}) \\ & -\beta_{i-1} F^a_{i-1}(x) \Bigg\}. \end{aligned}$$

The condition $E\{\beta_T V_T(X)\} \le 0$ implies that

$$E\left[\mathbf{1}_A \left\{\beta_T\left(1 - \sum_{j=i+1}^T \zeta_j\right) f^b_T + \sum_{j=i}^{T-1} \beta_j \zeta_{j+1} f^b_j - \beta_{i-1} f^a_{i-1}\right\}\right] \le 0. \qquad \text{(B.9)}$$

Next, we set $\zeta_{i+1} = \xi_{i+1}$ and $\zeta_k = \xi_k \prod_{j=i+1}^{k-1}(1-\xi_j)$, for $k = i+2, \ldots, T$, where $\xi_j \in [0,1]$ and ξ_j is $\mathcal{F}_{j-1}-$measurable. This way we find that $1 - \sum_{j=i+1}^{T} \zeta_j = \prod_{j=i+1}^{T}(1-\xi_j)$ and that (B.9) is equivalent to

$$E\left[\mathbf{1}_A \left\{ \beta_T f_T^b \prod_{j=i+1}^{T}(1-\xi_j) + \sum_{j=i}^{T-1} \beta_j f_j^b \xi_{j+1} \prod_{k=i+1}^{j}(1-\xi_k) - \beta_{i-1} f_{i-1}^a \right\}\right] \le 0. \tag{B.10}$$

Finally, one can show that $E\left(H_i^b|\mathcal{F}_{i-1}\right) \le \beta_{i-1} f_{i-1}^a$ by choosing the ξ's correctly, e.g., by setting $\xi_j = \mathbf{1}_{\left\{\beta_{j-1} f_{j-1}^b > E\left(H_j^b|\mathcal{F}_{j-1}\right)\right\}}$ for $i+1 \le j \le T$. Then one gets (2.10) for $i \in \{1, \ldots, T-2\}$, by taking conditional expectations recursively.

For instance, take the case $i = T-2$. Using the iterated conditional property together (B.10), one gets

$$\begin{aligned} &E\left[\mathbf{1}_A(1-\xi_{T-1})E\left[(1-\xi_T)E\left[\beta_T f_T^b|\mathcal{F}_{T-1}\right]|\mathcal{F}_{T-2}\right]|\mathcal{F}_{T-3}\right] \\ &+E\left[\mathbf{1}_A(1-\xi_{T-1})E\left[\beta_{T-1} f_{T-1}^b|\mathcal{F}_{T-2}\right]|\mathcal{F}_{T-3}\right] \\ &\quad +E\left[\mathbf{1}_A \beta_{T-2} f_{T-2}^b \xi_{T-1}|\mathcal{F}_{T-3}\right] \le \beta_{T-3} f_{T-3}^a. \end{aligned}$$

By setting $\xi_t = \mathbf{1}_{\left\{\beta_{t-1} f_{t-1}^b > E(H_t^b|\mathcal{F}_{t-1})\right\}}$, for $t = T-1, T$ one finds that

$$(1-\xi_T)E\left(\beta_T f_T^b|\mathcal{F}_{T-1}\right) \ge H_{T-1}^b.$$

Similarly,

$$(1-\xi_{T-1})E\left(\beta_{T-1} f_{T-1}^b|\mathcal{F}_{T-2}\right) \ge H_{T-2}^b.$$

Replacing the two last inequalities in (B.11) one gets (2.10) for $t = T-2$.

As expected, by setting $x = -y \le 0$ one can use the same method to show that $E\left(H_i^a|\mathcal{F}_{i-1}\right) \ge f_{i-1}^b \beta_{i-1}$, which completes the proof.

C. Proof of Theorem 4.1

Suppose there is no arbitrage. Then (A1) or (A2) must hold for any $x \ge 0$. Suppose first that $P\{V_1(x_0) = 0\} = 1$ for some $x_0 > 0$. Then, for any $x \in [0, x_0]$, $P\{V_1(x) \ge 0\} = 1$, since $\frac{V_1(x)}{x}$ is non-increasing from (P2) for $x > 0$. In particular, by taking the limit $\lim_{x \downarrow 0} \frac{V_1(x)}{x}$ one finds that $P\{f_1^b - (1+r_1^b) f_0^a \ge 0\} = 1$. If $P\{f_1^b - (1+r_1^b) f_0^a = 0\} = 1$, then for any $x \ge 0$, $P\{V_1(x) = 0\} = 1$. However, if $P\{f_1^b - (1+r_1^b) f_0^a > 0\} > 0$, then there exists at least one $x \in [0, x_0]$ so that $P\{V_1(x) \ge 0\} = 1$ and $P\{V_1(x) > 0\} > 0$, so there is an arbitrage opportunity, which is a contradiction. Hence one must have $P\{f_1^b - (1+r_1^b) f_0^a = 0\} = 1$. Similarly, if $P\{V_1(-y_0) = 0\} = 1$ for some $y_0 > 0$, then one must have $P\{f_1^a - (1+r_1^a) f_0^b = 0\} = 1$. To summarize, if for $x_0 \ge 0$ (resp. $x_0 < 0$) the no arbitrage condition is established by $P\{V_1(X_0)=0\}=1$, then $P\left\{f_1^b-(1-r_1^b)f_0^a=0\right\}=1$ (resp. $P\left\{f_1^a-(1+r_1^a)f_0^b=0\right\}=1$). Next, we study the consequences of the condition $P\{V_1(x)<0\}>0$.

Suppose now that there is no arbitrage, and that for all $x\neq 0$, $P\{V_1(x)<0\}>0$. If there exists $x_0>0$ so that $P\{V_1(x_0)>0\}>0$, then for any $x\in(0,x_0]$,

$$P\left\{f_1^b-(1+r_1^b)f_0^a>0\right\}\geq P\left\{\frac{V_1(x)}{x}>0\right\}\geq P\left\{\frac{V_1(x_0)}{x_0}>0\right\}>0.$$

But since we showed that one cannot have $P\left\{f_1^b-(1+r_1^b)f_0^a\geq 0\right\}=1$, it follows that one must also have $P\left\{f_1^b-(1+r_1^b)f_0^a<0\right\}>0$. On the other hand, if $P\{V_1(x)\leq 0\}=1$, for all $x\geq 0$, then $nV_1(1/n)\to f_1^b-(1+r_1^b)f_0^a$ a.s., and it follows from the properties of convergence in law that

$$1=\limsup_{n\to\infty}P\{nV_1(1/n)\leq 0\}\leq P\{f_1^b-(1+r_1^b)f_0^a\leq 0\}.$$

Thus $P\{f_1^b-(1+r_1^b)f_0^a\leq 0\}=1$. Similarly, if there exists $y_0>0$ so that $P\{V_1(-y_0)>0\}>0$, then one must have $P\left\{f_1^a-(1+r_1^a)f_0^b<0\right\}>0$ and $P\left\{f_1^a-(1+r_1^a)f_0^b>0\right\}>0$. On the other hand, if $P\{V_1(-y)\leq 0\}=1$, for all $y\geq 0$, then $nV_1(-1/n)\to -f_1^a+(1+r_1^a)f_0^b$ a.s., and it follows from the properties of convergence in law that

$$1=\limsup_{n\to\infty}P\{nV_1(-1/n)\leq 0\}\leq P\{-f_1^a+(1+r_1^a)f_0^b\leq 0\}.$$

Thus $P\{f_1^a-(1+r_1^a)f_0^b\geq 0\}=1$.

Summarizing, non arbitrage implies that either (NA1) : $P\{f_1^b-(1+r_1^b)f_0^a\leq 0\}=1$ or (NA2) : $P\{f_1^b-(1+r_1^b)f_0^a<0\}>0$ and $P\{f_1^b-(1+r_1^b)f_0^a>0\}>0$, and either (NA3) : $P\{f_1^a-(1+r_1^a)f_0^b\geq 0\}=1$ or (NA4) : $P\{f_1^a-(1+r_1^a)f_0^b<0\}>0$ and $P\{f_1^a-(1+r_1^a)f_0^b>0\}>0$.

On the other hand, if (NA1) holds then $P\{V_1(x)\leq 0\}=1$ for any $x\geq 0$, while if (NA2) holds then $P\{V_1(x)\leq 0\}>0$ for any $x\geq 0$, and there is $x_0>0$ so that $P\{V_1(x_0)>0\}>0$. Similarly, if (NA3) holds then $P\{V_1(x)\leq 0\}=1$ for any $x\leq 0$, while if (NA4) holds then $P\{V_1(x)\leq 0\}>0$ for any $x\leq 0$, and there is $x_0<0$ so that $P\{V_1(x_0)>0\}>0$. As a result conditions (NA1) or (NA2) and (NA3) or (NA4) are necessary and sufficient for non arbitrage in the market.

D. Proof of Theorem 4.2

Recall that

$$\ell^{(ca)}(x) = \min\left\{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega), K + \inf_{\omega\in\{F_1^b(1)>K\}} F_1^b(x-1,\omega)\right\}\mathbf{1}_{[1,\infty)}(x)$$
$$+\min\left\{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega), K - \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega)\right\}\mathbf{1}_{[0,1]}(x)$$
$$-\max\left\{\sup_{\omega\in\{F_1^b(1)\le K\}} F_1^a(-x,\omega), \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega) - K\right\}\mathbf{1}_{(-\infty,0]}(x).$$

One notes that the function $\ell^{(ca)}(x)$ regroups the worst case scenarios for the portfolio value of the seller, so that one can write the condition $V_1(x,\pi_0)\ge 0$ a.s. in the simpler form $\pi_0 \ge c_0(x) + \ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\}$. As a result, the problem of finding the lowest price of the option for which there exists x such that $V_1(x,\pi_0)\ge 0$ a.s. can be written as

$$\pi_0^s = \inf_x\left[\ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\} + c_0(x)\right].$$

First, consider the case $0 < P\{F_1^b(1) > K\} < 1$, and note that $\ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\} + c_0(x)$ is non-decreasing for $x\ge 1$. In fact, for $x\ge 1$, one has

$$G(x) = \ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\} + c_0(x) \tag{D.1}$$

$$= F_0^a(x) - \frac{1}{1+r_1^b}\min\left\{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega), K + \inf_{\omega\in\{F_1^b(1)>K\}} F_1^b(x-1,\omega)\right\}$$
$$= F_0^a(x) - \frac{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega)}{1+r_1^b},$$

since

$$\begin{aligned}\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega) &\le \inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x-1,\omega) + \inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(1,\omega)\\ &\le \inf_{\omega\in\{F_1^b(1)>K\}} F^b(x-1,\omega) + K.\end{aligned}$$

The first inequality follows from (**P2**), while the second one follows from (**C3**). Then, we see from (D.1) that G is non-decreasing since, for all $x,y\ge 0$,

$$\begin{aligned}F_0^a(x+y) &- \frac{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x+y,\omega)}{1+r_1^b} \ge F_0^a(x) + F_0^a(y)\\ &- \frac{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega)}{1+r_1^b} - \frac{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(y,\omega)}{1+r_1^b},\end{aligned}$$

and $F_0^a(y) - \frac{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(y,\omega)}{1+r_1^b} \ge 0$, by Theorem 4.1.

Next, note that $\ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\} + c_0(x)$ is non-increasing for $x \in (-\infty, 0]$. In fact, for $x \le 0$ one gets

$$\begin{aligned}
\ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\} + c_0(x) &= \frac{1}{1+r_1^a}\max\left\{\sup_{\omega\in\{F_1^b(1)\le K\}} F_1^a(-x,\omega),\right.\\
&\qquad \left.\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega) - K\right\} - F_0^b(-x)\\
&= \frac{\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega) - K}{1+r_1^a} - F_0^b(-x), \quad \text{(D.2)}
\end{aligned}$$

since

$$\begin{aligned}
\sup_{\omega\in\{F_1^b(1)\le K\}} F_1^a(-x,\omega) &\le \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(-x,\omega) + \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1,\omega) - K\\
&\le \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega) - K.
\end{aligned}$$

The first inequality follows from (**C2**) and (**C3**) while the last inequality is a consequence of (**P2**). Then, (D.2) is non-increasing since for all $x, y \le 0$,

$$\begin{aligned}
&\left\{\frac{\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x-y,\omega)}{1+r_1^a} - F_0^b(-x-y)\right\}\\
&\qquad -\left\{\frac{\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega)}{1+r_1^a} - F_0^b(-x)\right\}\\
&\ge \frac{\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(-y,\omega)}{1+r_1^a} - F_0^b(-y) \ge 0,
\end{aligned}$$

by Theorem 4.1.

So far, one has seen that

$$\pi_0^s = \inf_{x\in[0,1]} \left\{\ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\} + F_0^a(x)\right\}.$$

For $x \in [0,1]$, recall that

$$\ell^{(ca)}(x) = \min\left\{\inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(x,\omega), K - \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1-x,\omega)\right\}.$$

Now, $\ell^{(ca)}(0) = K - \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(1,\omega) < 0$ and $\ell^{(ca)}(1) = \inf_{\omega\in\{F_1^b(1)\le K\}} F_1^b(1,\omega) > 0$. Set $x_0 = \sup\{x \in [0,1]; \ell^{(ca)}(x) \le 0\}$. From the continuity of $\ell^{(ca)}$ and the monotonicity of F_1^a, $\ell^{(ca)}(x_0) = 0$. Moreover $\ell^{(ca)}(x) < 0$ for all $0 \le x < x_0$, and $\ell^{(ca)}(x) > 0$ for all $x_0 < x \le 1$. Consequently, $\ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\} = \frac{\ell^{(ca)}(x)}{1+r_1^a}$ for all $x \in [0, x_0]$, and $\ell_0^{-1}\left\{-\ell^{(ca)}(x)\right\} = -\frac{\ell^{(ca)}(x)}{1+r_1^b}$ for all

$x \in [x_0, 1]$. Further set

$$x_1 = \inf \left\{ x \in (x_0, 1); \ell^{(ca)}(x) = \inf_{\omega \in \{F_1^b(1) \le K\}} F_1^b(x, \omega) \right\}.$$

Using the fact that $\ell^{(ca)}(x)$ is concave on $[0, 1]$, one finds that

$$\ell^{(ca)}(x) = \frac{\inf_{\omega \in \{F_1^b(1) \le K\}} F_1^b(x, \omega)}{1 + r_1^b}$$

for $x \in [x_1, 1]$, and

$$\ell^{(ca)}(x) = \frac{K - \sup_{\omega \in \{F_1^b(1) > K\}} F_1^a(1 - x, \omega)}{1 + r_a^b}$$

for $x \in [x_0, x_1, 1]$. Hence $F_0^a(x) - \frac{\ell^{(ca)}(x)}{1+r_1^b}$ is non-decreasing for $x \ge x_1$.

As a consequence,

$$\begin{aligned} \pi_0^a &= \min \left[\inf_{x \in [0, x_0]} \left\{ F_0^a(x) + \frac{\sup_{\omega \in \{F_1^b(1) > K\}} F_1^a(1 - x, \omega) - K}{1 + r_1^a} \right\}, \right. \\ & \left. \inf_{x \in [x_0, x_1]} \left\{ F_0^a(x) + \frac{\sup_{\omega \in \{F_1^b(1) > K\}} F_1^a(1 - x, \omega) - K}{1 + r_1^b} \right\} \right], \end{aligned}$$

proving the result when $0 < P\{F_1^b(1) > K\} < 1$.

Suppose that $P\{F_1^b(1) > K\} = 0$. Then, using (4.1),

$$\pi_0^a = \min \left[\inf_{x \ge 0} \left\{ F_0^a(x) - \frac{\inf_\omega F_1^b(x, \omega)}{1 + r_1^b} \right\}, \inf_{x \ge 0} \left\{ -F_0^b(x) + \frac{\sup_\omega F_1^a(x, \omega)}{1 + r_1^a} \right\} \right].$$

Since $\left\{ F_0^a(x) - \frac{\inf_\omega F_1^b(x,\omega)}{1+r_1^b} \right\} / x$ and $\left\{ -F_0^b(x) + \frac{\sup_\omega F_1^a(x,\omega)}{1+r_1^a} \right\} / x$ converges respectively to $\left\{ f_0^a(0) - \frac{\inf_\omega f_1^b(0,\omega)}{1+r_1^b} \right\}$ and $\left\{ -f_0^b(0) + \frac{\sup_\omega f_1^a(0,\omega)}{1+r_1^a} \right\}$, as $x \downarrow 0$, and both limits are non-negative by the no-arbitrage assumption, it follows that $\pi_0^a = 0$.

Finally, suppose that $P\{F_1^b(1) > K\} = 1$. Then, according to (4.2),

$$\ell^{(ca)}(x) = K + \inf_\omega F_1^b(x - 1, \omega) \mathbf{1}_{[1,\infty)}(x) - \sup_\omega F_1^a(1 - x, \omega) \mathbf{1}_{(-\infty,1]}(x),$$

so

$$\pi_0^a = \min \left[\inf_{x \ge 1} G_1(x), \inf_{x \in [0, x_0]} G_{0x_0}(x), \inf_{x \in [x_0, 1]} G_{x_1 1}(x), \inf_{x \le 0} G_{-1}(x) \right]$$

where $G_1(x) = F_0^a(x) - \frac{K+\inf_\omega F_1^b(x-1,\omega)}{1+r_1^b}$, $G_{0x_0}(x) = F_0^a(x) + \frac{\sup_\omega F_1^a(1-x,\omega)-K}{1+r_1^a}$, $G_{x_0 1}(x) = F_0^a(x) + \frac{\sup_\omega F_1^a(1-x,\omega)-K}{1+r_1^b}$, $G_{-1}(x) = -F_0^b(-x) + \frac{\sup_\omega F_1^a(1-x,\omega)-K}{1+r_1^a}$, and x_0 is defined by $\sup_\omega F_1^a(1 - x_0) = K$.

First, for any $x \geq 1$, $G_1(x) - G_1(1) \geq F_0^a(x-1) - \frac{\inf_\omega F_1^b(x-1,\omega)}{1+r_1^b} \geq 0$ by convexity and the no-arbitrage conditions. Similarly, $G_{-1}(x) \geq G_{-1}(0)$ for any $x \leq 0$.

It then follows that

$$\begin{aligned}\pi_0^a &= \min\left[\inf_{x\in[0,x_0]}\left\{F_0^a(x) + \frac{\sup_\omega F_1^a(1-x,\omega) - K}{1+r_1^a}\right\},\right.\\ &\qquad \left.\inf_{x\in[x_0,1]}\left\{F_0^a(x) + \frac{\sup_\omega F_1^a(1-x,\omega) - K}{1+r_1^b}\right\}\right].\end{aligned}$$

E. Proof of Theorem 4.3

Recall that

$$\begin{aligned}\ell^{(cb)}(x) &= \min\left\{\inf_{\omega\in\{F_1^b(1)\leq K\}} F_1^b(x,\omega), \inf_{\omega\in\{F_1^b(1)>K\}} F_1^b(x+1,\omega) - K\right\}\mathbf{1}_{[0,\infty)}(x)\\ &+ \min\left\{-\sup_{\omega\in\{F_1^b(1)\leq K\}} F_1^a(-x,\omega), \inf_{\omega\in\{F_1^b(1)>K\}} F_1^b(x+1,\omega) - K\right\}\mathbf{1}_{[-1,0]}(x)\\ &- \max\left\{\sup_{\omega\in\{F_1^b(1)\leq K\}} F_1^a(-x,\omega), \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(-x-1,\omega) + K\right\}\mathbf{1}_{(-\infty,-1]}(x).\end{aligned}$$

The function $\ell^{(cb)}(x)$ represents the worst case scenarios for the portfolio value of the buyer, so one can write the condition $V_1(x,-\pi_0) \geq 0$ a.s. in the simpler form $\pi_0 \leq -\ell_0\left\{-\ell^{(cb)}(x)\right\} - c_0(x)$. As a result, the problem of finding the highest acceptable price of the option for which there exists x such that $V_1(x,-\pi_0) \geq 0$ a.s., can be written as

$$\begin{aligned}\pi_0^b &= \sup_x\left[-\ell_0^{-1}\left\{-\ell^{(cb)}(x)\right\} - c_0(x)\right]\\ &= \sup_x\left\{\frac{(\ell^{(cb)}(x))^+}{1+r^b} - \frac{(\ell^{(cb)}(x))^-}{1+r^a} - c_0(x)\right\}.\end{aligned}$$

First, one considers the case $0 < P\{F_1^b(1) > K\} < 1$ and one shows that the supremum is attained for $x \leq 0$.

For $x \geq 0$, set $G_1(x) = \frac{\ell^{(cb)}(x)}{1+r_1^b} - F_0^a(x)$. Then, $G_1(0) = 0$ and one only needs to show that $G_1(x) \leq 0$ for all $x > 0$. By taking the limit one finds that

$$\lim_{x\downarrow 0}\frac{G_1(x)}{x} = \frac{\inf_{\omega\in\{F_1^b(1)\leq K\}} f_1^b(0,\omega)}{1+r_1^b} - f_0^a(0) \leq 0 \tag{E.1}$$

where the inequality is given by Theorem 4.1. Since $G_1(x)$ is concave, one can conclude that $G_1(x) \leq 0$ for all $x \geq 0$.

Now, we study the behaviour of $\frac{(\ell^{(cb)}(x))^+}{1+r^b} - \frac{(\ell^{(cb)}(x))^-}{1+r^a} - c_0(x)$ for $x \leq -1$. In this case,

$$\ell^{(cb)}(x) = -\max\left\{ \sup_{\omega\in\{F_1^b(1)\leq K\}} F_1^a(-x,\omega), \sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(-x-1,\omega)+K \right\} \leq 0.$$

Set $H_1(x) = F_0^b(-x) + \frac{\ell^{(cb)}(x)}{1+r_1^a}$ for $x \leq -1$. If $\sup_{\omega\in\{F_1^b(1)\leq K\}} F_1^a(1,\omega) < K$, then $\ell^{(cb)}(-1) = -K$ and one has

$$\begin{aligned}\lim_{x\uparrow -1} \frac{H_1(x)-H_1(-1)}{x+1} &= -f_0^b(1) + \frac{\sup_{\omega\in\{F_1^b(1)>K\}} f_1^a(0,\omega)}{1+r_1^a} \\ &\geq -f_0^b + \frac{\sup_{\omega\in\{F_1^b(1)>K\}} f_1^a(0,\omega)}{1+r_1^a} \geq 0.\end{aligned}$$

The first inequality follows from (P2) while the second is a due to **(C1)** and Theorem 4.1. Since $H_1(x)$ is concave, one concludes that $H_1(x) \leq H_1(-1)$ for all $x \leq -1$, proving that

$$\pi_0^b = \sup_{x\in[-1,0]} \left\{ F_0^b(-x) + \frac{\ell^{(cb)}(x)}{1+r_1^a} \right\}.$$

Finally, suppose that $\sup_{\omega\in\{F_1^b(1)\leq K\}} F_1^a(1,\omega) \geq K$. Then

$$\begin{aligned}\sup_{x\leq -1} \left\{ F_0^b(-x) + \frac{\ell^{(cb)}(x)}{1+r_1^a} \right\} &\leq \sup_{y\geq 0} \left\{ F_0^b(1+y) - \frac{\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(y,\omega)+K}{1+r_1^a} \right\} \\ &\leq F_0^b(1) - \frac{K}{1+r_1^a} \\ &\quad + \sup_{y\geq 0} \left\{ F_0^b(y) - \frac{\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(y,\omega)}{1+r_1^a} \right\}.\end{aligned}$$

Since $H_3(y) = F_0^b(y) - \frac{\sup_{\omega\in\{F_1^b(1)>K\}} F_1^a(y,\omega)}{1+r_1^a}$ is concave, it follows that

$$\lim_{y\downarrow 0} H_3(y)/y = f_0^b - \frac{\sup_\omega f_1^a}{1+r_1^a} \leq 0,$$

using **(C1)** and Theorem 4.1. As a result, $H_3(y) \leq 0 = H_3(0)$ for any $y \geq 0$, so

$$\sup_{x\leq -1} \left\{ F_0^b(-x) + \frac{\ell^{(cb)}(x)}{1+r_1^a} \right\} \leq F_0^b(1) - \frac{K}{1+r_1^a},$$

This completes the proof of the case $0 < P\{F_1^b(1) > K\} < 1$.

If $P\{F_1^b(1) > K\} = 1$, recall that

$$\begin{aligned}\ell^{(cb)}(x) &= \inf_\omega F_1^b(x+1,\omega)\mathbf{1}_{[-1,\infty)}(x) \\ &\quad - \sup_\omega F_1^a(-x-1,\omega)\mathbf{1}_{(-\infty,-1]}(x) - K.\end{aligned}$$

For $x \geq 0$, one has $-\ell_0^{-1}\left\{-\ell^{(cb)}(x)\right\} - c_0(x) = G_2(x)$, where

$$G_2(x) = \frac{\inf_\omega F_1^b(x+1,\omega) - K}{1+r_1^b} - F_0^a(x).$$

Then

$$\lim_{x\downarrow 0} \frac{G_2(x) - G_2(0)}{x} \leq \frac{\inf_\omega f_1^b(0,\omega)}{1+r_1^b} - f_0^a \leq 0,$$

where the first inequality comes from (P2) and the second from Theorem 4.1. Since $G_2(x)$ is concave, one concludes that $\sup_{x\geq 0} G_2(x) = G_2(0)$.

Next, for $x \leq -1$, one has $-\ell_0^{-1}\left\{-\ell^{(cb)}(x)\right\} - c_0(x) = H_2(x)$, where $H_2(x) = F_0^b(-x) - \frac{\sup_\omega F_1^a(-x-1,\omega)+K}{1+r_1^a}$. Then

$$\begin{aligned} \lim_{x\uparrow -1} \frac{H_2(x) - H_2(-1)}{x} &= \frac{\sup_\omega f_1^a(0,\omega)}{1+r_1^a} - f_0^b(1) \\ &\geq \frac{\sup_\omega f_1^a(0,\omega)}{1+r_1^a} - f_0^b(0) \geq 0. \end{aligned}$$

Again, the first inequality comes from (P2) while the second comes from Theorem 4.1. One concludes that $\sup_{x\leq -1} H_2(x) = H_2(-1)$ since $H_2(x)$ is concave.

Finally, let $x_0 \in [-1,0]$ be defined by $\inf_\omega F_1^b(x_0+1,\omega) = K$, so that $\ell^{(cb)}(x) \leq 0$ if $x \in [-1,x_0]$ and $\ell^{(cb)}(x) > 0$ if $x \in (x_0,0]$. Then, one has

$$\begin{aligned} \pi_0^b &= \max\left[\sup_{x\in[-1,x_0]}\left\{F_0^b(-x) + \frac{\inf_\omega F_1^b(x+1,\omega) - K}{1+r_1^a}\right\},\right. \\ &\qquad \left.\sup_{x\in[x_0,0]}\left\{F_0^b(-x) + \frac{\inf_\omega F_1^b(x+1,\omega) - K}{1+r_1^b}\right\}\right], \end{aligned}$$

which completes the proof when $P\{F_1^b(1) > K\} = 1$.

If $P\{F_1^b(1) > K\} = 0$, then one proves that $\pi_0^b = 0$ by showing that the supremum is attained at $x = 0$. Recall that

$$\ell^{(cb)}(x) = \inf_\omega F_1^b(x,\omega)\mathbf{1}_{[0,\infty)}(x) - \sup_\omega F_1^a(-x,\omega)\mathbf{1}_{(-\infty,0]}(x).$$

$$\ell^{(cb)}(x) = \begin{cases} \inf_\omega F_1^b(x,\omega), & x \geq 0, \\ -\sup_\omega F_1^a(-x,\omega), & x \leq 0. \end{cases}$$

Then $\ell^{(cb)}(0) = 0$. Next, one can apply (E.1) to show that $-\ell_0^{-1}\left\{-\ell^{(cb)}(x)\right\} - c_0(x) \leq 0$ for $x \geq 0$. For $x \leq 0$, $-\ell_0^{-1}\left\{-\ell^{(cb)}(x)\right\} - c_0(x) = F_0^b(-x) - \frac{\sup_\omega F_1^a(-x,\omega)}{1+r_1^a}$

and

$$\lim_{x\uparrow 0}\frac{1}{x}\left\{F_0^b(-x)-\frac{\sup_\omega F_1^a(-x,\omega)}{1+r_1^a}\right\}=\frac{\sup_\omega f_1^a(0,\omega)}{1+r_1^a}-f_0^b(0)\geq 0.$$

The latter shows that $F_0^b(-x)-\frac{\sup_\omega F_1^a(-x,\omega)}{1+r_1^a}\leq 0$. Since $-\ell_0^{-1}\left\{-\ell^{(cb)}(x)\right\}-c_0(x)$ is non-decreasing for $x\leq 0$, worths zero if $x=0$ and is non-positive for $x\geq 0$, we have that $\pi_0^b=0$.

F. Proof of Theorem 4.4

Recall that

$$\begin{aligned}\ell^{(pa)}(x)&=\min\left\{\inf_{\omega\in\{F_1^a(1)<K\}}F_1^b(x+1,\omega)-K,\inf_{\omega\in\{F_1^a(1)\geq K\}}F_1^b(x,\omega)\right\}\mathbf{1}_{[0,\infty)}(x)\\&+\min\left\{\inf_{\omega\in\{F_1^a(1)<K\}}F_1^b(x+1,\omega)-K,-\sup_{\omega\in\{F_1^a(1)\geq K\}}F_1^a(-x,\omega)\right\}\mathbf{1}_{[-1,0]}(x)\\&-\max\left\{\sup_{\omega\in\{F_1^a(1)<K\}}F_1^a(-x-1,\omega)+K,\sup_{\omega\in\{F_1^a(1)\geq K\}}F_1^a(-x,\omega)\right\}\mathbf{1}_{(-\infty,-1]}(x).\end{aligned}$$

Since the function $\ell^{(pa)}(x)$ represents the worst case scenarios for the portfolio value of the seller, one can write the condition $V_1(x,\pi_0)\geq 0$ a.s. in the simpler form $\pi_0\geq\ell_0^{-1}\left\{-\ell^{(pa)}(x)\right\}+c_0(x)$. As a result, the problem of finding the lowest price of the option for which there exists x such that $V_1(x,\pi_0)\geq 0$ a.s. can be written as

$$\pi_0^a=\inf_x\left[\ell_0^{-1}\left\{-\ell^{(pa)}(x)\right\}+c_0(x)\right].$$

First, consider the case $0<P\{F_1^a(1)<k\}<1$. Note that for all $x\leq -1$, $\ell^{(pa)}(x)=-\sup_{\omega\in\{F_1^a(1)\geq K\}}F_1^a(-x,\omega)$ since

$$\begin{aligned}\sup_{\omega\in\{F_1^a(1)<k\}}F_1^a(-x-1,\omega)+K&\leq\sup_{\omega\in\{F_1^a(1)\geq K\}}F_1^a(-x-1,\omega)\\&+\sup_{\omega\in\{F_1^a(1)\geq K\}}F_1^a(1,\omega\leq\sup_{\omega\{F_1^a(1)\geq K\}}F_1^a(-x,\omega).\end{aligned}$$

The first inequality comes from (**C2′**) and (**C3′**) while the second inequality comes from (**P2**).

For $x\leq -1$, set $G_1(x)=\frac{\sup_{\omega\in\{F_1^a(1)\geq k\}}F_1^a(-x,\omega)}{1+r_1^a}-F_0^b(-x)$. One finds that

$$\begin{aligned}\lim_{x\uparrow -1}\frac{G_1(x)-G_1(-1)}{x+1}&=f_0^b(1)-\frac{\sup_{\omega\in\{F_1^a(1)\geq k\}}f_1^a(0,\omega)}{1+r_1^a}\\&\leq f_0^b(0)-\frac{\sup_{\omega\in\{F_1^a(1)\geq k\}}f_1^a(0,\omega)}{1+r_1^a}\leq 0,\end{aligned}$$

where the first inequality comes from (P2) and the second from Theorem 4.1. Since $G_1(x)$ is convex, one concludes that $G_1(x) \geq G_1(-1)$ for $x \leq -1$.

For $x \in [-1, 0]$, one has that

$$\ell^{(pa)}(x) = \min\left\{ \inf_{\omega \in \{F_1^a(1) < k\}} F_1^b(x+1, \omega) - K, - \sup_{\omega \in \{F_1^a(1) \geq k\}} F_1^a(-x, \omega) \right\} \leq 0.$$

Let $x_0 = \sup\left\{x \in [-1,0]; \ell^{(pa)}(x) = -\sup_{\omega \in \{F_1^a(1) \geq k\}} F_1^a(-x, \omega)\right\}$. One knows that x_0 exists since $\sup_{\omega \in \{F_1^a(1) \geq k\}} F_1^a(1, \omega) \geq K$. Set $H_1(x) = -F_0^b(-x) - \frac{\ell^{(pa)}(x)}{1+r_1^a}$, then

$$\lim_{x \uparrow x_0} \frac{H_1(x) - H_1(x_0)}{x - x_0} \leq f_0^b - \frac{\sup_{\omega \in \{F_1^a(1) \geq k\}} f_1^a(0, \omega)}{1 + r_1^a} \leq 0.$$

where the first inequality comes from (P2) and the second comes from Theorem 4.1. Since $H_1(x)$ is convex, one concludes that $H_1(x) \geq H_1(x_0)$, for $x \in [-1, x_0]$. Hence

$$\pi_0^a = \inf_{x \geq x_0} \left[\ell_0^{-1}\left\{-\ell^{(pa)}(x)\right\} + c_0(x)\right],$$

since

$$\inf_{x \in [-1, x_0]} \left[\ell_0^{-1}\left\{-\ell^{(pa)}(x)\right\} + c_0(x)\right] = - \sup_{x \in [-1, x_0]} H_1(x) = -H_1(x_0).$$

For $x \geq 0$, one finds that

$$\begin{aligned} \ell^{(pa)}(x) &= \min\left\{ \inf_{\omega \in \{F_1^a(1) < k\}} F_1^b(x+1, \omega) - K, \inf_{\omega \in \{F_1^a(1) \geq k\}} F_1^b(x, \omega) \right\} \\ &= \inf_{\omega \in \{F_1^a(1) < k\}} F_1^b(x+1, \omega) - K, \end{aligned}$$

since

$$\begin{aligned} \inf_{\omega \in \{F_1^a(1) < k\}} F_1^b(x+1, \omega) - K &\leq \inf_{\omega \in \{F_1^a(1) < k\}} F_1^b(x, \omega) + \inf_{\omega \in \{F_1^a(1) < k\}} F_1^b(1, \omega) - K \\ &\leq \inf_{\omega \in \{F_1^a(1) \geq k\}} F_1^b(x, \omega). \end{aligned}$$

Set $J(x) = \ell_0^{-1}\left\{K - \inf_{\omega \in \{F_1^a(1) < k\}} F_1^b(x+1, \omega)\right\} + F_0^a(x)$. Further let x_1 be such that $\inf_{\omega \in \{F_1^a(1) < k\}} F_1^b(x_1 + 1, \omega) = K$. Then

$$\lim_{x \downarrow x_1} \frac{J(x) - J(x_1)}{x - x_1} \geq f_0^a - \frac{\inf_{\omega \in \{F_1^a(1) < k\}} f_1^b(0, \omega)}{1 + r_1^b} \geq 0, \tag{F.1}$$

by convexity, (**C1**) and Theorem 4.1. Hence, $\inf_{x \geq x_1} J(x) = J(x_1)$. Now,

$$\lim_{x \downarrow 0} \frac{J(x) - J(0)}{x} \geq f_0^a - \frac{\inf_{\omega \in \{F_1^a(1) < k\}} f_1^b(0, \omega)}{1 + r_1^a}. \tag{F.2}$$

Hence, if $\frac{\inf_{\omega\in\{F_1^a(1)<k\}} f_1^b(1,\omega)}{1+r_1^a} \le f_0^a(0)$, then $J(x) \ge J(0)$ for $x \ge 0$, since $J(x)$ is convex.

>From these last observations, one may conclude that

$$\begin{aligned}\pi_0^a &= \min\left[\inf_{x\in[x_0,0]}\left\{\frac{K-\inf_{\omega\in\{F_1^a(1)<k\}} F_1^b(1+x,\omega)}{1+r_1^a} - F_0^b(-x)\right\},\right.\\ &\qquad \left.\inf_{x\in[0,x_1]}\left[\ell_0^{-1}\left\{K-\inf_{\omega\in\{F_1^a(1)<k\}} F_1^b(1+x,\omega)\right\}+F_0^a(x)\right]\right],\end{aligned}$$

and if $\frac{\inf_{\omega\in\{F_1^a(1)<k\}} f_1^b(x+1,\omega)}{1+r_1^a} \le f_0^a(0)$, then

$$\pi_0^a = \inf_{x\in[x_0,0]}\left\{\frac{K-\inf_{\omega\in\{F_1^a(1)<k\}} F_1^b(1+x,\omega)}{1+r_1^a} - F_0^b(-x)\right\}.$$

If $P\{F_1^a(1)<k\}=1$, recall that

$$\ell^{(pa)}(x) = \inf_\omega F_1^b(x+1,\omega)\mathbf{1}_{[-1,\infty)}(x) - \sup_\omega F_1^a(-x-1,\omega)\mathbf{1}_{(-\infty,-1]}(x) - K.$$

For $x \le -1$, set

$$G_2(x) = \ell_0^{-1}\left\{-\ell^{(pa)}(x)\right\} + c_0(x) = \frac{\sup_\omega F_1^a(-x-1,\omega)+K}{1+r_1^a} - F_0^b(-x).$$

Then,

$$\begin{aligned}\lim_{x\uparrow -1}\frac{G_2(x)-G_2(-1)}{x+1} &= f_0^b(1) - \frac{\sup_\omega f_1^a(0,\omega)}{1+r_1^a}\\ &\le f_0^b(0) - \frac{\sup_\omega f_1^a(0,\omega)}{1+r_1^a} \le 0,\end{aligned}$$

where the first inequality comes from (P2) and the second comes from **(C1')** and Theorem 4.1. Since $G_2(x)$ is convex, one concludes that $G_2(x) \ge G_2(-1)$ for $x \le -1$.

For $x \ge 0$, one has that $J(x) = \ell_0^{-1}\left\{K - \inf_\omega F_1^b(x+1,\omega)\right\} + F_0^a(x)$ since $\inf_\omega F_1^b(x,\omega) = \inf_{\omega\in\{F_1^a(1)<k\}} F_1^b(x,\omega)$. Consequently, using (F.2), one finds that

$$\begin{aligned}\pi_0^a &= \min\left[\inf_{x\in[-1,0]}\left\{\frac{K-\inf_\omega F_1^b(1+x,\omega)}{1+r_1^a} - F_0^b(-x)\right\},\right.\\ &\qquad \left.\inf_{x\in[0,x_1]}\left[\ell_0^{-1}\left\{K-\inf_\omega F_1^b(1+x,\omega)\right\}+F_0^a(x)\right]\right],\end{aligned}$$

and if $\frac{\inf_\omega f_1^b(x+1,\omega)}{1+r_1^a} \le f_0^a(0)$, then

$$\pi_0^a = \inf_{x\in[-1,0]}\left\{\frac{K-\inf_\omega F_1^b(1+x,\omega)}{1+r_1^a} - F_0^b(-x)\right\}.$$

Finally, if $P\{F_1^a(1) < k\} = 0$, recall that

$$\ell^{(pa)}(x) = \begin{cases} \inf_\omega F_1^b(x,\omega), & x \geq 0, \\ -\sup_\omega F_1^a(-x,\omega), & x \leq 0. \end{cases}$$

For $x \leq 0$, set $G_3(x) = \ell_0^{-1}\left\{-\ell^{(pa)}(x)\right\} + c_0(x) = \frac{\sup_\omega F_1^a(-x,\omega)}{1+r_1^a} - F_0^b(-x)$. Then, using **(C1)** and Theorem 4.1, one gets

$$\lim_{x\uparrow 0} \frac{G_3(x)}{x} = f_0^b(0) - \frac{\sup_\omega f_1^a(0,\omega)}{1+r_1^a} \leq 0,$$

which shows that $G_3(x) \geq G_3(0) = 0$ for $x \leq 0$, since $G_3(x)$ is convex.

For $x \geq 0$, set $H_2(x) = \ell_0^{-1}\left\{-\ell^{(pa)}(x)\right\} + c_0(x) = F_0^a(x) - \frac{\inf_\omega F_1^b(x,\omega)}{1+r_1^b}$. Then,

$$\lim_{x\downarrow 0} \frac{H_2(x)}{x} = f_0^a(0) - \frac{\inf_\omega f_1^b(0,\omega)}{1+r_1^b} \geq 0,$$

where the inequality comes from Theorem 4.1. Again, since $H_2(x)$ is convex, one has that $H_2(x) \geq H_2(0) = 0$ for $x \geq 0$. Hence one may conclude that $\pi_0^a = 0$.

G. Proof of Theorem 4.5

Recall that

$$\begin{aligned}\ell^{(pb)}(x) &= \min\left\{\inf_{\omega\in\{F_1^a(1)<K\}} F_1^b(x-1,\omega)+K, \inf_{\omega\in\{F_1^a(1)\geq K\}} F_1^b(x,\omega)\right\}\mathbf{1}_{[1,\infty)}(x) \\ &+ \min\left\{-\sup_{\omega\in\{F_1^a(1)<K\}} F_1^a(-x+1,\omega)+K, \inf_{\omega\in\{F_1^a(1)\geq K\}} F_1^b(x,\omega)\right\}\mathbf{1}_{[0,1]}(x) \\ &- \max\left\{\sup_{\omega\in\{F_1^a(1)<K\}} F_1^a(-x+1,\omega)-K, \sup_{\omega\in\{F_1^a(1)\geq K\}} F_1^a(-x,\omega)\right\}\mathbf{1}_{(-\infty,0]}(x).\end{aligned}$$

The function $\ell^{(pb)}(x)$ represents the worst case scenarios for the portfolio value of the buyer, so one can write the condition $V_1(x,-\pi_0) \geq 0$ a.s. in the simpler form $\pi_0 \leq -\ell_0^{-1}\left\{-\ell^{(pb)}(x)\right\} - c_0(x)$. As a result, the problem of finding the highest acceptable price of the option for which there exists x such that $V_1(x,-\pi_0) \geq 0$ a.s., can be written as

$$\begin{aligned}\pi_0^b &= \sup_x \left\{-l_0^{-1}\left\{-\ell^{(pb)}(x)\right\} - c_0(x)\right\} \\ &= \sup_x \left\{\frac{(\ell^{(pb)}(x))^+}{1+r_1^b} - \frac{(\ell^{(pb)}(x))^-}{1+r_1^a} - c_0(x)\right\}.\end{aligned}$$

First, consider the case $0 < P\{F_1^a(1) < k\} < 1$. One can show that $\frac{(\ell^{(pb)}(x))^+}{1+r^b} - \frac{(\ell^{(pb)}(x))^-}{1+r^a} - c_0(x) \leq 0$ for $x \leq 0$. Recall that for $x \leq 0$,

$$\ell^{(pb)}(x) = -\max\left\{\sup_{\omega\in\{F_1^a(1)<k\}} F_1^a(-x+1,\omega) - K, \sup_{\omega\in\{F_1^a(1)\geq k\}} F_1^a(-x,\omega)\right\} \leq 0,$$

so that

$$\frac{(\ell^{(pb)}(x))^+}{1+r_1^b} - \frac{(\ell^{(pb)}(x))^-}{1+r_1^a} - c_0(x) = G_1(x),$$

where $G_1(x) = F_0^b(-x) + \frac{\ell^{(pb)}(x)}{1+r_1^a}$. Then

$$\lim_{x\uparrow 0} \frac{G_1(x)}{x} = \frac{\sup_{\omega\in\{F_1^a(1)\geq k\}} f_1^a(0,\omega)}{1+r_1^a} - f_0^b(0) \geq 0,$$

where the inequality follows from (**C1**) and Theorem 4.1. Since $G_1(x)$ is concave, one concludes that $G_1(x) \leq G_1(0) = 0$ for $x \leq 0$.

One now studies the behaviour of $\frac{(\ell^{(pb)}(x))^+}{1+r_1^b} - \frac{(\ell^{(pb)}(x))^-}{1+r_1^a} - c_0(x)$ for $x \geq 1$. Recall that for $x \geq 1$,

$$\ell^{(pb)}(x) = \min\left\{\inf_{\omega\in\{F_1^a(1)<k\}} F_1^b(x-1,\omega) + K, \inf_{\omega\in\{F_1^a(1)\geq k\}} F_1^b(x,\omega)\right\} \geq 0,$$

so that

$$\frac{(\ell^{(pb)}(x))^+}{1+r_1^b} - \frac{(\ell^{(pb)}(x))^-}{1+r_1^a} - c_0(x) = \frac{\ell^{(pb)}(x)}{1+r_1^b} - F_0^a(x).$$

One has to split the problem in two cases, relatively to the value of $\inf_{\omega\in\{F_1^a(1)\geq k\}} F_1^b(1,\omega)$ and K.

Set $H_1(x) = \frac{\ell^{(pb)}(x)}{1+r_1^b} - F_0^a(x)$ and assume that $\inf_{\omega\in\{F_1^a(1)\geq k\}} F_1^b(1,\omega) > K$. Then,

$$\begin{aligned}\lim_{x\downarrow 1} \frac{H_1(x)-H_1(1)}{x-1} &= \frac{\inf_{\omega\in\{F_1^a(1)<k\}} f_1^b(0,\omega)}{1+r_1^b} - f_0^a(1) \\ &\leq \frac{\inf_{\omega\in\{F_1^a(1)<k\}} f_1^b(0,\omega)}{1+r_1^b} - f_0^a(0) \leq 0,\end{aligned}$$

where the first inequality comes from (P2) and the second from (**C1**) and Theorem 4.1. Using the fact that $H_1(x)$ is concave, one concludes that $H_1(x) \leq H_1(1)$ for $x \geq 1$, proving that

$$\pi_0^b = \sup_{[0,1]} \left\{\frac{\ell^{(pb)}(x)}{1+r_1^b} - F_0^a(x)\right\}.$$

Finally, suppose that $\inf_{\omega\in\{F_1^a(1)\geq k\}} F_1^b(1,\omega) < K$. Then, for $x \geq 1$, one has that

$$\frac{\ell^{(pb)}(x)}{1+r_1^b} - F_0^a(x) \leq \frac{\inf_{\omega\in\{F_1^a(1)<k\}} F_1^b(x-1,\omega) + K}{1+r_1^b} - F_0^a(x)$$

$$\leq \frac{\inf_{\omega\in\{F_1^a(1)<k\}} F_1^b(x-1,\omega)+K}{1+r_1^b} - F_0^a(x-1) - F_0^a(1)$$
$$\leq \frac{K}{1+r_1^b} - F_0^a(1),$$

using (**C1**), Theorem 4.1, and convexity. This shows that the supremum is finite, concluding the proof for the case $0 < P\{F_1^a(1) < k\} < 1$.

If $P\{F_1^a(1) < k\} = 1$, recall that

$$\ell^{(pb)}(x) = \inf_{\omega} F_1^b(x-1,\omega)\mathbf{1}_{[1,\infty)}(x) - \sup_{\omega} F_1^a(-x+1,\omega)\mathbf{1}_{(-\infty,1]}(x) + K.$$

For $x \geq 1$, set

$$H_2(x) = \frac{\ell^{(pb)}(x)}{1+r_1^b} - c_0(x) = \frac{\inf_\omega F_1^b(x-1,\omega)+K}{1+r_1^b} - F_0^a(x).$$

Then

$$\lim_{x\downarrow 1} \frac{H_2(x)-H_2(1)}{x-1} = \frac{\inf_\omega f_1^b(0,\omega)}{1+r_1^b} - f_0^a(1) \leq \frac{\inf_\omega f_1^b(0,\omega)}{1+r_1^b} - f_0^a(0) \leq 0,$$

where the first inequality is given by (P2) and the second by (**C1**), and Theorem 4.1. Since $H_2(x)$ is concave, one concludes that $H_2(x) \leq H_2(1)$ for $x \geq 1$.

For $x \leq 0$, set

$$G_2(x) = -\ell_0^{-1}\left\{-\ell^{(pb)}(x)\right\} - c_0(x) = \ell_0^{-1}\left\{K - \sup_\omega F_1^a(1-x,\omega)\right\} - F_0^b(-x).$$

Then, one has

$$\lim_{x\uparrow 0} \frac{G_2(x)-G_2(0)}{x} \geq \frac{\sup_\omega f_1^a(0,\omega)}{1+r_1^a} - f_0^b(0) \geq 0,$$

where the first inequality comes from (P2) and the second comes from (**C1**) and Theorem 4.1. Finally, one concludes that $G_2(x) \leq G_2(0)$ for $x \leq 0$ since $G_2(x)$ is concave. The latter completes the proof for the case $P\{F_1^a(1) < k\} = 1$.

If $P\{F_1^a(1) < k\} = 0$, recall that

$$\ell^{(pb)}(x) = \inf_\omega F_1^b(x,\omega)\mathbf{1}_{[0,\infty)}(x) - \sup_\omega F_1^a(-x,\omega)\mathbf{1}_{(-\infty,0]}(x).$$

For $x \geq 0$, set

$$H_3(x) = \frac{\ell^{(pb)}(x)}{1+r_1^b} - c_0(x) = \frac{\inf_\omega F_1^b(x,\omega)}{1+r_1^b} - F_0^a(x).$$

Then

$$\lim_{x\downarrow 0} \frac{H_3(x)}{x} = \frac{\inf_\omega f_1^b(0,\omega)}{1+r_1^b} - f_0^a(x) \leq 0,$$

where the inequality comes (**C1**) and Theorem 4.1. Since $H_3(x)$ is concave, one has that $H_3(x) \leq H_3(0) = 0$ for $x \geq 0$.

Finally, for $x \leq 0$, set

$$G_3(x) = \frac{\ell^{(pb)}(x)}{1 + r_1^a} - c_0(x) = F_0^b(-x) - \frac{\sup_\omega F_1^a(-x, \omega)}{1 + r_1^a}.$$

Then

$$\lim_{x \uparrow 0} \frac{G_3(x)}{x} = \frac{\sup_\omega f_1^a(0, \omega)}{1 + r_1^a} - f_0^b(0) \geq 0,$$

where the inequality comes from (**C1**) and Theorem 4.1. Since G_3 is concave, one concludes that $G_3(x) \leq G_3(0) = 0$ for $x \leq 0$, showing that $\pi_0^b = 0$ when $P\{F_1^a(1) < k\} = 0$.

Chapitre 5

FORECASTING TIME SERIES WITH MULTIVARIATE COPULAS

Résumé

Dans cet article, nous présentons une méthode de prévision pour les séries temporelles qui est basée sur les modèles de copules pour les séries à plusieurs dimensions. Nous étudions comment la performance des prévisions évolue lorsque l'on fait varier la force des différentes dépendances possibles et nous les comparons avec une version unidimensionnelle de la méthode qui a été introduite par Sokolinskiy and Van Dijk (2011). De plus, nous étudions aussi l'influence des distributions marginales grâce à une nouvelle mesure de performance. Finalement, nous étudions comment la structure de la dépendance affecte les prévisions et l'effet des erreurs d'estimations. Nous présentons aussi un exemple d'application sur des données financières.

Abstract

In this paper, we present a forecasting method for time series using copula-based models for multivariate time series. We study how the performance of the predictions evolve when changing the strength of the different possible dependencies and compare it with the univariate version of the forecasting method introduced by Sokolinskiy and Van Dijk (2011). Moreover, we also study the influence of the marginal distributions with the help of a new performance measure. Lastly we look at the impact of the dependence structure on the predictions performance and the effect of estimation errors. We also give an example of practical implementation with financial data.

Key words : *copulas, time series, forecasting, realized volatility.*

1. Introduction

For many years, copulas have been used for modeling dependence between random variables. See e.g. Genest et al. (2009) for a survey on copulas in finance. The possibility to model the dependence structure independently from marginal distributions allows for a better understanding of the dependence structure and a wide range of joint distributions. More recently, copulas have been used to model the temporal dependence in time series, first in the univariate case, as in Chen and Fan (2006) and Beare (2010), and then in a multivariate setting, Rémillard et al. (2012). Once again, the flexibility of copulas allows to model more complex dependence structures and thus to better capture the evolution of the time series. In the recent work of Sokolinskiy and Van Dijk (2011), copulas were used to forecast the realized volatility associated with a univariate financial time series and it was shown there that copula-based forecasts perform better than forecasts based on heterogeneous autoregressive (HAR) model, Corsi (2009). The later method had been proven successful in Andersen et al. (2007), Corsi (2009) and Bush et al. (2011).

Looking at the literature, we see that all the tools are there to build forecasts based on copula models for multivariate time series. For instance, Patton (2013) suggests the idea of multivariate forecasting based on copulas. However, although he makes a good presentation of the different aspects of modeling multivariate time series using copulas, he never describes an actual forecasting method. Beyond that, there is no understanding about how would perform predictions based on multivariate time series respect to univariate copula-based predictions and what could impact the performance of the predictions.

Consequently, our first goal is to extend the methodology of Sokolinskiy and Van Dijk (2011) by proposing a forecasting method using copula-based models for multivariate time series, as in Rémillard et al. (2012). As one can guess, we show that forecasting multivariate time series using copula-based models gives better results than forecasting a single time series, as long as there is dependence between the series. For example, let $\{(X_{1,t}, X_{2,t}); t = 0, 1, ...\}$ be two dependent time series with both series showing temporal dependence. Suppose one wants to forecast $X_{1,T+1}$ based on the information available at period T. We show that forecasting the joint values of $(X_{1,T+1}, X_{2,T+1})$ using the observed values $(X_{1,T}, X_{2,T})$ gives significantly better predictions of $X_{1,T+1}$ in general than predictions on $X_{1,T+1}$ based only on the single value of $X_{1,T}$, which of course has to be expected. Since $\{X_{1,t}\}$ and $\{X_{2,t}\}$ are dependent and temporally dependent, the knowledge of $(X_{1,T}, X_{2,T})$ gives more informations than the knowledge of $X_{1,T}$ alone.

In practice, many factors might affect the quality of the predictions. However, it would be impossible to test all practical aspects of the implementation of our method. So, we designed some numerical experiments to test the impact of, what we think, are the most important factors that could affect the performance of the predictions.

Our first numerical experiment studies what is the impact on the predictions of the strength of the different dependencies of the vector $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$. Similarly, we study the structure of the dependencies as well as the impact of marginal distributions affect the performance of the predictions. Another important aspect of the implementation of our method is the problem of estimation errors, considering that a multivariate model might require more parameters to estimate. We also study this question.

Although our numerical experiments focus on the bivariate case, our presentation can be readily extended to an arbitrary number of dimensions. Actually, the results of Rémillard et al. (2012), which provide the estimation methods, are given for an arbitrary number of time series and most of the theoretical background is going to be presented in the general case. Moreover, the results of our numerical experiments should naturally extend to the multivariate case.

The rest of the paper is structured as follows. In Section 2 we give some basic results about copulas and apply the results to model time series. In Sections 2.3 we define our forecasting methods. Section 3 contains the result of our numerical experiments as well as the analysis of the results. We also give a complete example of practical implementation with financial data in Section 4. The last section contains some concluding remarks.

2. Modeling time series with copulas

2.1. Copulas

We begin by giving some definitions and basic results about copulas. More details about copulas can be found in Nelsen (1999) and Rémillard (2013).

Definition 2.1. *(Copula)*

A d-dimensional copula is a distribution function with domain $[0,1]^d$ and uniform margins.

Equivalently, the function $C : [0,1]^d \to [0,1]$ is a d-dimensional copula if and only if there exists random variables $U_1, \ldots, U_d$ such that $P(U_i \le u) = u_i$ for $i = 1, \ldots, d$ and $C(\mathbf{u}) = P(U_1 \le u_1, \ldots, U_d \le u_d)$ for all $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$. The existence of a copula function for any joint distribution is given by Sklar's theorem.

Theorem 2.1. *(Sklar's theorem)*

Let $X_1, \ldots, X_d$ be d random variables with joint distribution function H and margins $F_1, \ldots, F_d$. Then, there exists a d-dimensional copula C such that for all $(x_1, \ldots, x_d) \in \bar{\mathbb{R}}^d$,

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), \tag{2.1}$$

where $\bar{\mathbb{R}} = \mathbb{R} \bigcup \{-\infty, \infty\}$.

We note that the copula function in (2.1) is uniquely defined on the set $Range(F_1) \times \cdots \times Range(F_d)$. Hence, if $Range(F_i) = [0, 1]$ for $i = 1, \ldots, d$ the copula is unique.

We define the left continuous inverse of a distribution function F as

$$F^{-1}(u) = \inf \{x; F(x) \geq u\}, \text{ for all } u \in (0, 1).$$

Using this inverse and Sklar's theorem, we have a way to define the copula function in terms of the quasi-inverses and the joint distribution.

Assuming that the density f_i of F_i exists for each $i = 1, \ldots, d$, then the density c of C exists if and only if the density h of H also exists. In this case, differentiating equation (2.1), we get

$$h(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d))\Pi_{i=1}^d f_i(x_i).$$

Furthermore, for all $(u_1, \ldots, u_d) = (0, 1)^d$,

$$c(u_1, \ldots, u_d) = \frac{h(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{\Pi_{i=1}^d f_i(F_i^{-1}(u_i))}. \tag{2.2}$$

Following the example of conditional distributions it is also possible to define conditional copulas. Let $(\mathbf{X}, \mathbf{Y})$ be a $(d_1 + d_2)$-dimensional random vector with joint distribution H, where $\mathbf{X}$ has marginal distributions $F_1, \ldots, F_{d_1}$ and $\mathbf{Y}$ has marginal distributions $G_1, \ldots, G_{d_2}$.

Setting $\mathbf{F}(X) = (F_1(X_1), \ldots, F_{d_1}(X_{d_1}))$, $\mathbf{G}(Y) = (G_1(Y_1), \ldots, G_{d_2}(Y_{d_2}))$ and defining the random vector $(\mathbf{U}, \mathbf{V}) = (\mathbf{F}(X), \mathbf{G}(Y))$ we can define the copula $C_{\mathbf{UV}}$ of the vector (X, Y) as the joint distribution function of $(\mathbf{U}, \mathbf{V})$.

Assuming that the density functions exist and applying equation (2.2), one obtains that the conditional copula $C_{\mathbf{U}|\mathbf{V}}$, i.e., the conditional distribution of U given V, is given by

$$C_{\mathbf{U}|\mathbf{V}}(\mathbf{u}; \mathbf{v}) = \frac{\partial_{v_1} \cdots \partial v_{d_2} C_{\mathbf{UV}}(u, v)}{c_{\mathbf{V}}(\mathbf{v})},$$

with density

$$c_{\mathbf{U}|\mathbf{V}}(\mathbf{u}; \mathbf{v}) = \frac{c_{\mathbf{UV}}(u, v)}{c_{\mathbf{V}}(\mathbf{v})},$$

where $c_{\mathbf{V}}$ is the density of the copula $C_{\mathbf{V}}(\mathbf{v}) = C_{\mathbf{UV}}(1, \ldots, 1, v)$ associated to $\mathbf{Y}$ or $\mathbf{V}$.

Having defined conditional copulas, one can now look at how to obtain copula-based models for multivariate time series.

2.2. Modeling time series

In order to get a prediction method, we first need to present how to use copulas for modeling time series. The ideas presented here were developed in Soustra (2006) and Rémillard et al. (2012), extending the results of Chen and Fan (2006) to the multivariate case.

Let $\mathbf{X} = \{\mathbf{X}_t; t = 0, 1, \ldots\}$ be a d-dimensional time series and assume that $\mathbf{X}$ is Markovian and stationary. We note F_i the marginal distribution of $X_{i,t}$ for $i = 1, ..., d$ and H the joint distribution of $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ and assume that all distributions are continuous. From the stationarity assumption, it follows that all distribution functions $F_1, \ldots, F_d$ and H are time-independent. Using Sklar's theorem, there is a unique copula C associated to $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ and unique copula Q associated to $\mathbf{X}_{t-1}$, viz.

$$Q(\mathbf{u}) = C(\mathbf{1}_d, u) = C(\mathbf{u}, \mathbf{1}_d) \text{ for all } \mathbf{u} \in [0, 1]^d,$$

where $\mathbf{1}_d$ is the d-dimensional unit vector. The second equality above comes from the hypothesis of stationarity, i.e. the distribution of U_t is the same as the distribution of U_{t-1}.

Set $\mathbf{U}_t = \mathbf{F}(\mathbf{X}_t)$, for $t \geq 0$. The next step is to deduce the conditional copula of X_t given X_{t-1}, which is

$$\mathcal{C}(\mathbf{u}; \mathbf{v}) = C_{\mathbf{U}_t|\mathbf{U}_{t-1}}(\mathbf{u}; \mathbf{v}) = \frac{\partial_{v_1} \cdots \partial v_d C(u, v)}{q(\mathbf{v})},$$

with density

$$c_{\mathbf{U}_t|\mathbf{U}_{t-1}}(\mathbf{u}; \mathbf{v}) = \frac{c(u, v)}{q(\mathbf{v})},$$

where q is the density of Q.

Combining the knowledge of the marginal distributions and the conditional copula above we can get the conditional distribution of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$. This is what we use to define our predictions.

2.3. Forecasting method

To expose our forecasting method, we first make the assumption that the joint distribution as well as the marginal distributions of the time series are known. However, for practical implementation, these distributions are unknown and estimations are to be done. This will be treated next.

Let $\mathbf{X} = \{\mathbf{X}_t; t = 0, 1, .., T\}$ be a d-dimensional time series. Our goal is to forecast X_{T+1} based on the information available at time T. Suppose that for all $t \geq 0$, F_i is the marginal distribution of $X_{i,t}$ for $i = 1, ..., d$ and the $2d$-dimensional vector $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ as joint distribution H and copula C. Using the preceding section we can define the conditional copula $\mathcal{C}$ of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$, namely $C_{\mathbf{U}_t|\mathbf{U}_{t-1}}(\mathbf{u}; \mathbf{v})$.

Now suppose we observe the value $\mathbf{X}_T = \mathbf{y}$ for the time series at time T. The prediction of $\mathbf{X}_{T+1}$ goes as follows :

(1) Set $\mathbf{v} = \mathbf{F}(\mathbf{y})$.

(2) Simulate n realizations of the conditional copula, $\mathbf{U}^{(i)} \sim \mathcal{C}(\cdot; \mathbf{v})$, $i = 1, \ldots, n$.

(3) For $i = 1, \ldots, n$, set $\mathbf{X}_{T+1}^{(i)} = \mathbf{F}^{-1}(\mathbf{U}^{(i)})$.

(4) Set

$$\hat{\mathbf{X}}_{T+1} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_{T+1}^{(i)}. \tag{2.3}$$

We use $\hat{X}_{1,T+1}$ as a predictor for $X_{1,T+1}$.

4′ One can also define a prediction interval of level $1 - \alpha \in (0, 1)$ for $X_{1,T+1}$ by taking the estimated quantiles of order $\alpha/2$ and $1 - \alpha/2$ amongst $\{X_{1,T+1}^{(i)}; i = 1, \ldots, n\}$. We denote by $\widehat{LB}_{T+1}^{\alpha}$ and $\widehat{UB}_{T+1}^{\alpha}$ the lower and upper values for the prediction interval.

As mentioned previously, we are going to compare our predictions performance with the univariate version presented in Sokolinskiy and Van Dijk (2011). Let D be the copula associated with $(X_{1,t-1}, X_{1,t})$ for $t = 0, 1, ..$, i.e., D is the copula of $(U_{1,t-1}, U_{1,t})$.

Suppose we observe the value $X_{1,T} = y_1$; The predictor presented in Sokolinskiy and Van Dijk (2011) is defined as

$$\bar{X}_{1,T+1} = n^{-1} \sum_{i=1}^{n} F_1^{-1}(Z^{(i)}) \tag{2.4}$$

where $Z^{(i)}$ are realizations of $\mathcal{D}(\cdot; v_1)$ where $v_1 = F_1(y_1)$ and $\mathcal{D}$ is the conditional copula associated with $X_{1,t}$ given $X_{1,t-1}$. As before, we can define the prediction interval using the estimated $\alpha/2$ and $1 - \alpha/2$ quantiles from the values $\{F_1^{-1}(Z^{(i)}), i = 1, \ldots, n\}$. We note the upper and lower bound of predictions interval at time $T + 1$ by $\overline{UB}_{T+1}^{\alpha}$ and $\overline{LB}_{T+1}^{\alpha}$.

2.3.1. *Implementation in practice*

For practical implementation, one has to replace the known distributions $\mathbf{F}$ and the copula C by estimated versions. The estimation method for copula-based model for time series is presented in Rémillard et al. (2012) which use non-parametrical estimation for marginal distribution and parametrical estimation for conditional copula where the copula parameters are estimated through pseudo maximum likelihood. Goodness-of-fit tests are also provided to help choose the right copula family, but in the context of forecasting, one can also choose copulas by their prediction power.

2.4. Including more information

The methodology proposed here can also be applied to predict $X_{1,T+1}$ given $\mathbf{X}_T$ and $X_{2,T+1}, \ldots, X_{d,T+1}$, since the joint copula of $(\mathbf{X}_t, \mathbf{X}_{t+1})$ is given. For instance, suppose one estimated a copula model for the series $\{(X_{1,t}, X_{2,t}, X_{1,t+1}, X_{2,t+1}); t = 0, 1, \ldots\}$ where $X_{1,t}$ is the realized volatility and $X_{2,t}$ is the volume of transaction. Using the conditional copula of $X_{1,t+1}$ given $(X_{1,t}, X_{2,t}, X_{2,t+1})$ one can create different predictions of the realized volatility at time $t+1$ based on the observed values of the volume and the realized volatility at time t and different possible values of the volume at time $t+1$. This way, one gets a curve of predictions for the realized volatility.

3. Analysis of the performance

Here we compare the performance of our forecasting method versus the method proposed by Sokolinskiy and Van Dijk (2011). The analysis is restricted to bivariate time series but the results can easily be extrapolated to higher dimensions. For these experiments we suppose that the copula and the marginal distributions are known so that the predictions are not affected by estimation error. We will see in Section 3.5 the effect of parameter misspecification.

Theoretically, our proposed methodology should give better results because we are using the additional information contained a second series, provided the dependence is strong enough and we let aside the possible problems of model and parameters misspecification. Consequently, the gain in performance should be affected by the strength of the dependencies between and within time series, i.e., the overall dependence associated with the vector $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$.

To understand how these factors come into play, we consider two copulas : The Student copula and the Clayton copula. The choice of the Student copula is motivated by the fact that it seems to fit data well in practice and also that we have a lot of flexibility in specifying the correlation matrix for the Student distribution, which in turn defines the strength of the dependencies in the related copula. Actually, there is a bijection between the correlation matrix and the Kendall's tau matrix. If $R = [R_{i,j}]$ for $i, j = 1, ..., d$ are the elements of the correlation matrix, the Kendall's tau matrix for the Student copula is then given by $\tau_{i,j} = \frac{2}{\pi}\arcsin(R_{i,j})$, (Fang et al., 2002). Note that in general the correlation matrix R is not the correlation matrix of the observations. However, this is true when the margins have a Student distribution with the same degrees of freedom, i.e., if the joint law is Student. This is why in practice Kendall's tau is used to characterize the pairwise dependence for the Student copula.

In order to test a different dependence structure we also use data simulated from the Clayton copula. See Appendix A for details about simulating multivariate copula-based time series using the Student and Clayton copulas.

We also want to examine the impact of the marginal distributions on the performance. To make things simpler, we chose the same margins for both series. We can expect that the predictions of a random variable with large variance should be less precise than when the variance is small. To try to eliminate this effect, we propose a new measure of performance.

3.1. Performance measures

For most of the numerical experiments we use prediction intervals with $\alpha = 0.05$. To measure the performance of the predictions we compute the mean length of the prediction intervals and also give the proportion of observed values outside the prediction intervals. Let UB_t^α and LB_t^α be the upper bound and lower bound of the prediction interval with confidence level $1-\alpha$, for $t = 1, \dots N$. We define the mean length as

$$ML^\alpha = \frac{1}{N}\sum_{t=1}^{N}\left(UB_t^\alpha - LB_t^\alpha\right).$$

We will use $\widehat{ML}^\alpha$ for the mean length of prediction intervals based on the bivariate series and $\overline{ML}^\alpha$ if predictions are based on the univariate series. Indeed, smaller values for the mean length of prediction intervals means better precision. However, the proportion of observed values outside the prediction intervals must be close to $\alpha = 0.05$. For our numerical experiments, we put $N = 10\,000$ so that using the normal approximation for the binomial distribution one can find that a 95% confidence interval for the proportion of observed values outside the prediction intervals is approximately 0.05 ± 0.0042.

Remark 3.1. *Instead of prediction intervals, one can choose to make pointwise predictions using* (2.3) *and* (2.4). *In this case, a natural performance measure* *for the predictions is the mean absolute prediction errors. We also used this performance measure for our numerical experiments of Sections 3.2.1 - 3.2.3 and we found out that the results were the same as for the mean length of prediction intervals. Consequently, we did not include these results in the paper since they did not bring additional information.*

In Section 3.3 we test the effect of different marginal distributions on the quality of predictions. However, we expect that the mean length of prediction intervals should be larger for distribution with bigger variance. To eliminate the margins effect, we use pointwise predictions defined at (2.3) and (2.4) and we propose the following performance measure called the mean absolute rank error. Let $\tilde{X}_t$ be pointwise predictions of X_t for $t = 1, \ldots, N$, the mean absolute rank error is defined by

$$MARE = N^{-1} \sum_{t=1}^{N} |F_1(X_{1,t}) - F_1(\widetilde{X}_{1,t})|,$$

where F_1 is the marginal distribution of $X_{1,t}$ for all $t = 1, \ldots, N$. As before, we will use $\widehat{MARE}$ if predictions are based on the bivariate series and $\overline{MARE}$ for the univariate case.

All those performance measures are summarized in Table 5.1. The table also includes the mean absolute error that will be used in Section 4.

Table 5.1. Summary of the performance measures. $X_{1,t}$ is the observed value, $\tilde{X}_{1,t}$ is the predicted value, F_1 is the distribution function of $X_{1,t}$ for all $t \geq 0$ and UB_t^α and LB_t^α are respectively the upper and lower bound of the prediction intervall with confidence level α.

Performance measure	Symbol	Definition
Mean absolute error	MAE	$N^{-1} \sum_{t=1}^{N} \left\| X_{1,t} - \tilde{X}_{1,t} \right\|$
Mean absolute rank error	$MARE$	$N^{-1} \sum_{t=1}^{N} \left\| F_1(X_{1,t}) - F_1(\tilde{X}_{1,t}) \right\|$
Mean length of confidence interval with confidence level α	ML^α	$N^{-1} \sum_{t=1}^{N} \left(UB_t^\alpha - LB_t^\alpha \right).$

3.2. Impact of the dependencies strength

As we said previously, the structure and the strength of the dependencies of the vector $\mathbf{X}_t = (X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$ should have an impact on the performance of our predictor. To understand and quantify this impact, we first simulate $\mathbf{X}_t$ with a Student copula and study the impact for a set of possible correlations. We choose degrees of freedom $\nu = 8$ and fix the initial values $(X_{1,0}, X_{2,0}) = (0, 0)$. To isolate the effect of the correlation, we take the margins as Student with $\nu = 8$

degrees of freedom as well, so the distribution of $\mathbf{X}_t$ is a multivariate Student distribution with a correlation ρ between $X_{1,t}$ and $X_{2,t}$. For these experiments, we take $N = 10\,000$ and we generate a sample of $n = 1000$ to compute prediction intervals with error $\alpha = 0.05$. For three different simulations we will take different correlation values between $(X_{1,t}, X_{2,t})$, $(X_{1,t}, X_{1,t-1})$ and finally $(X_{1,t}, X_{2,t-1})$.

Remark 3.2. *Recall that our method applies to stationary series. In all the following numerical experiments, when we have to generate stationary series, we always discard the first 100 values of the series, so we can consider that the series is stationary. To be precise, for a series of length N, we actually generate $N+100$ values and discard the first* 100 *values.*

In all the following numerical experiments, we compare $\widehat{ML}^{\alpha}$ and $\overline{ML}^{\alpha}$ and we see that $\widehat{ML}^{\alpha}$ is always less than or equal to $\overline{ML}^{\alpha}$. This observation shows that predictions based on bivariate series outperform predictions based on univariate series since prediction intervals are narrower. However, this comparison is valid only if the proportion of observed values outside the prediction intervals is close to 0.05. As one will see these proportions are most of the time in the 95% confidence interval $[0.0458, 0.0542]$.

3.2.1. *First numerical experiment*

The first simulation study the impact of the dependence between both series. We simulate the series using correlation matrices

$$R_\rho = \begin{bmatrix} 1 & \rho & 0.25 & 0.25 \\ \rho & 1 & 0.25 & 0.25 \\ 0.25 & 0.25 & 1 & \rho \\ 0.25 & 0.25 & \rho & 1 \end{bmatrix}$$

with

$$\begin{aligned} \rho \in \; & \{-0.4, -0.3, -0.2, -0.1, -0.05, -0.01, 0.01, 0.05, 0.1, 0.2 \\ & 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}\,. \end{aligned}$$

Remark 3.3. *As one can see, we introduce a correlation of* 0.25 *between the pairs $(X_{1,t}, X_{1,t-1})$, $(X_{1,t}, X_{2,t-1})$ and $(X_{2,t}, X_{2,t-1})$, so that the effect of ρ is not completely isolated in the performance of the predictions. The reason of this choice is that, in practice, one will implement this forecasting method only if there is dependence between series. In this case, we consider that it is more likely that one will find a minimum of dependence between all pairs in $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$. This additional dependence which is introduced should be kept in mind when analysing the results in Sections 3.2.1-3.2.3.*

As seen in Figure 5.1, the bivariate method generates smaller prediction intervals. Since they are prediction intervals with confidence level 95%, we expect that the proportion of observed values out of sample should be around $\alpha = 0.5$. From Figure 5.2, we see that the proportion of observed values out of prediction intervals is similar for both methods. Consequently, predictions from the bivariate method are more precise. Also, from Figure 5.1, one sees that the mean length of prediction intervals $\widehat{ML}^{\alpha}$ of our predictions increases when τ (Kendall's tau) increases. To explain this result, consider the extreme case where the Kendall's tau between $X_{1,t}$ and $X_{2,t}$ is one. Then, the two series are identical and hence our predictor has no additional information coming from the second series. This also explain why the difference between mean length of prediction intervals gets closer to zero when τ is high. Again, as τ increases, the predictor $\widehat{ML}^{\alpha}$ tends to $\overline{ML}^{\alpha}$, since the information from the second series becomes irrelevant.

FIGURE 5.1. Impact of the dependence between $X_{1,t}$ and $X_{2,t}$. The plain line gives the value of $\overline{ML}^{\alpha}$ and the dashed gives the value of $\widehat{ML}^{\alpha}$. The x-axis gives the Kendall's tau for $X_{1,t}$ and $X_{2,t}$.

FIGURE 5.2. Proportion of observed values outside of prediction intervals. The circles give the results for prediction intervals based on univariate series while plus signs are for prediction intervals based on bivariate series. The x-axis gives the Kendall's tau for $X_{1,t}$ and $X_{2,t}$. The horizontal lines give the 95% confidence interval.

3.2.2. *Second numerical experiment*

For the second simulation, we study the impact of the serial dependence, i.e., the dependence between $X_{1,t}$ and $X_{1,t-1}$ through $\tau = \frac{2}{\pi}\arcsin(\rho)$, Kendall's tau between $X_{1,t}$ and $X_{1,t-1}$. We simulate the series using correlation matrices

$$R_\rho = \begin{bmatrix} 1 & 0.25 & \rho & 0.25 \\ 0.25 & 1 & 0.25 & 0.25 \\ \rho & 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1 \end{bmatrix}$$

with

$$\begin{aligned} \rho \;\in\; & \{-0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, -0.05, -0.01, \\ & 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}\,. \end{aligned}$$

Remark 3.4. *We remark that the range of ρ is not the same than before. The reason is that we have to keep the correlation matrices positive semi-definite.*

A first observation from Figure 5.3 is that prediction intervals are smaller for the bivariate method. Since the proportion of observed values outside the prediction intervals is similar for both methods, according to Figure 5.4, we conclude that predictions from the bivariate method are more precise. As expected theoretically, Figure 5.3 shows that the predictions are better when the dependence is strong. However, it is interesting to note that, for this dependence structure, $\widehat{ML}^{\alpha}$ benefits more from the negative dependence than $\overline{ML}^{\alpha}$. This is probably explained by the fact that there is a positive dependence between the other pairs of random variables, i.e. $(X_{1,t}, X_{2,t})$ and $(X_{1,t}, X_{2,t-1})$. Finally, when the correlation is close to one, the information given by the first lag dictates almost completely the succeeding value and the information given by the second series then becomes marginal. This is why the difference in prediction error is close to zero when the correlation is close to one.

FIGURE 5.3. Impact of the dependence between $X_{1,t}$ and $X_{1,t-1}$. The plain line gives the value of $\overline{ML}^{\alpha}$ and the dashed gives the value of $\widehat{ML}^{\alpha}$. The x-axis gives the Kendall's tau for $X_{1,t}$ and $X_{1,t-1}$.

FIGURE 5.4. Proportion of observed values outside of prediction intervals. The circles give the results for prediction intervals based on univariate series while plus signs are for prediction intervals based on bivariate series. The x-axis gives the Kendall's tau for $X_{1,t}$ and $X_{1,t-1}$. The horizontal lines give the 95% confidence interval.

3.2.3. *Third numerical experiment*

The last simulation is about the impact of the strength of the dependence between $X_{1,t}$ and $X_{2,t-1}$. We simulate the series using correlation matrices

$$R_\rho = \begin{bmatrix} 1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 1 & \rho & 0.25 \\ 0.25 & \rho & 1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1 \end{bmatrix}$$

with

$$\rho \in \{-0.4, -0.3, -0.2, -0.1, -0.05, -0.01, 0.01, 0.05, 0.1, 0.2$$

$$0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}.$$

As one should expect, this is the most important dependence in the comparative performance of our predictor. Since the information given by $X_{2,t-1}$ cannot be used by predictions based only on univariate series, our predictor gives much better performance when this dependence is strong, as seen in Figure 5.5. In the case where the dependence is strong, the information of $X_{2,t-1}$ almost completely dictates the value of $X_{1,t}$ and this is why we observe a great difference between $\widehat{ML}^{\alpha}$ and $\overline{ML}^{\alpha}$. Once again, the proportion of observed values outside prediction intervals, as seen from Figure 5.6, is similar for both methods.

FIGURE 5.5. Impact of the dependence between $X_{1,t}$ and $X_{2,t-1}$. The plain line gives the value of $\overline{ML}^{\alpha}$ and the dashed gives the value of $\widehat{ML}^{\alpha}$. The x-axis gives the Kendall's tau for $X_{1,t}$ and $X_{2,t-1}$.

FIGURE 5.6. Proportion of observed values outside of prediction intervals. The circles give the results for prediction intervals based on univariate series while plus signs are for prediction intervals based on bivariate series. The x-axis gives the Kendall's tau for $X_{1,t}$ and $X_{2,t-1}$. The horizontal lines give the 95% confidence interval.

3.3. Impact of the marginal distributions

Another question we want to tackle is the impact of the margins. In order to isolate more closely the impact of the different correlations in the first set of simulations, we used only the Student copula with Student margins. In the next experiment, we still use Student copula, but we are going to use different marginal distributions. The parameters of the Student copula are the same as before and we fix the correlation matrix at $R_{i,j} = 0.25$ for all $i \neq j$.

Looking at the results of Table 5.2, our first observation is that for the same marginal distributions, the the mean length of prediction intervals are affected by the change of parameters while the $MARE$ is not. This observation seems to confirm that our performance measure behaves as intended. So, looking at the $MARE$ values, we conclude that the marginal distributions do affect the performance of the predictions.

Table 5.2. Evolution of $MARE$ and ML as a function of the marginal distributions. Numbers in parenthesis are the proportion of observed values out of prediction intervals.

Margins	$\overline{MARE}$	$\widehat{MARE}$	$\overline{ML}^{\alpha}$	$\widehat{ML}^{\alpha}$
T_5	0.24	0.21	4.96 (0.0506)	4.35 (0.0514)
T_8	0.24	0.21	4.44 (0.0526)	3.91 (0.0520)
T_{10}	0.24	0.21	4.29 (0.0515)	3.78 (0.0514)
T_{15}	0.24	0.21	4.11 (0.0527)	3.63 (0.0512)
T_{20}	0.24	0.21	4.02 (0.0519)	3.56 (0.0513)
T_{30}	0.24	0.21	3.94 (0.0534)	3.49 (0.0525)
T_{50}	0.24	0.21	3.87 (0.0534)	3.43 (0.0518)
$LN(0,1)$	0.28	0.23	6.69 (0.0541)	5.78 (0.0521)
χ^2_8	0.25	0.21	14.79 (0.0530)	13.11 (0.0515)
Exp_3	0.26	0.22	10.55 (0.0527)	9.26 (0.0526)
$N(0,1)$	0.24	0.21	3.78 (0.0524)	3.36 (0.0512)
$N(0,2)$	0.24	0.21	5.34 (0.0518)	4.75 (0.0511)
$N(0,4)$	0.24	0.21	7.56 (0.0516)	6.71 (0.0502)
$N(0,8)$	0.24	0.21	10.68 (0.0526)	9.49 (0.0513)
$N(2,1)$	0.24	0.21	3.78 (0.0520)	3.36 (0.0507)
$N(4,1)$	0.24	0.21	3.78 (0.0530)	3.36 (0.0513)
$N(8,1)$	0.24	0.21	3.78 (0.0510)	3.36 (0.0524)

3.4. Impact of the dependence structure

Our last numerical experiment shows that the dependence structure has an impact on the gain in performance of predictions based on bivariate series compare to predictions based on univariate series. At first sight, we could think that using the information provided by the series X_2 might always give better predictions but we find that the dependence structure of the Clayton copula almost negate this advantage. From the definition of the Clayton copula we see that the dependence structure is symmetric, that is, all the dependencies of the vector $\mathbf{X}_t = (X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$ are the same. Moreover, the strength of the dependencies increases when θ increases. When θ is close to zero, the elements of the vector $\mathbf{X}_t$ are close to be independent and so, there is not much information to use to predict the next value. In contrast, when θ is high, both series are almost the same and, this time, the series X_2 cannot provides useful information to our predictor.

The results illustrated in Figure 5.7 show the evolution of prediction performances in terms of the parameter θ. First, Figure 5.8 shows that the proportion of values outside prediction intervals corresponds to the confidence level of 95%. Then, from Figure 5.7, we see that both predictors perform badly when θ is small and perform better as long as θ becomes bigger. We also see that that the difference between both prediction errors is slowly decreasing for high values of θ. This is due to the fact that the Kendall's tau between $X_{1,t}$ and it's first lag $X_{1,t-1}$ is close to one, and so, the additional information provided by X_2 becomes marginal. As we see, with the dependence structure given by the Clayton copula the advantage of using predictions based on the bivariate series is minor.

FIGURE 5.7. Impact of the parameter θ in the Clayton copula. The plain line gives the value of $\overline{ML}^{\alpha}$ and the dashed gives the value of $\widehat{ML}^{\alpha}$. The x-axis gives the values of θ.

FIGURE 5.8. Proportion of observed values outside of prediction intervals. The circles give the results for prediction intervals based on univariate series while plus signs are for prediction intervals based on bivariate series. The x-axis gives the values of the parameter θ. The horizontal lines give the 95% confidence interval.

3.5. Impact of estimation errors

In all the previous numerical experiments, we supposed that copulas and marginal distributions were known, i.e. there was no need to estimate parameters. This way, we were able to isolate the effect of using the information provided by the second series. We saw that using multivariate predictions outperform univariate predictions, but in some cases the improvement is rather small. In these cases, one can ask if the errors caused by parameters estimation might negate the advantage of the multivariate forecasting method. Mostly when the multivariate method requires more parameters to estimate. In this section, perform two numerical experiments to test the impact of estimation errors on the performance of the predictions.

For our first experiment we generate a bivariate series $\mathbf{X}_t$ of length $N+1$ from a Student copula with $\nu = 8$ degrees of freedom and correlation matrix

$$R = \begin{bmatrix} 1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 1 & 0.7 & 0.25 \\ 0.25 & 0.7 & 1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1 \end{bmatrix}.$$

We estimate the parameters of the copula using the N first values of the series and predict the value $N+1$. We repeat this experiment 10 000 times and compute the performance of the 10 000 predictions. To generate the series, we also take Student marginal distributions with the same degrees of freedom as the copula.

In order to have different precisions for the estimated parameters we perform this experiment with sample sizes $N \in \{100, 250, 500, 750\}$. The results are displayed in Table 5.3.

TABLE 5.3. Impact of estimation errors for the Student copula. From left to right we have : the length of the series, the mean absolute error for the bivariate and univariate method and the mean length of prediction intervals for the bivariate and univariate method. The numbers in parenthesis are the proportion of observed values out of the prediction intervals.

Student copula				
Length	$\overline{MAE}$	$\widehat{MAE}$	$\overline{ML}^{\alpha}$	$\widehat{ML}^{\alpha}$
100	0.8917	0.6936	4.5897 **(0.0589)**	3.3043 **(0.0643)**
250	0.8572	0.6956	4.4658 (0.0532)	3.242 **(0.0592)**
500	0.8597	0.6820	4.4505 (0.0539)	3.2366 **(0.0554)**
750	0.8726	0.6817	4.413 **(0.0573)**	3.3241 (0.0534)

The first observation from Table 5.3 is that pointwise predictions are more precise for the bivariate method. Our second observation is that, even though the bivariate method still creates smaller prediction intervals, the proportion of observed values outside the prediction intervals are inside the confidence interval only for the series of length 750. This means that the predictions of the quantiles using the bivariate method is more sensitive to estimation errors than the univariate method.

Our second experiment retains the same idea as before, except we generate a bivariate series following a Clayton copula with parameter $\theta = 5$ and uniform marginal distributions. Looking at the results in Table 5.4, we see that the bivariate method outperforms the univariate method since the mean absolute error for pointwise predictions as well as the mean length of prediction intervals are smaller and the number of observed values outside predictions intervals are all within the confidence interval.

From these results, we conclude that predictions based on the bivariate method still gives better predictions when considering estimation errors. However, in the case of the Student copula, it seems that prediction intervals using the bivariate method are more sensitive to estimation errors, which is probably caused by the fact that it uses more parameters. This explanation is also supported by the results on the Clayton copula. In this case, all the predictions rely on one estimated parameter and the bivariate method always gives the best predictions.

Table 5.4. Impact of estimation errors for the Clayton copula. From left to right we have : the length of the series, the mean absolute error for the bivariate and univariate method and the mean length of prediction intervals for the bivariate and univariate method. The numbers in parenthesis are the proportion of observed values outside the prediction intervals.

Clayton copula				
Length	$\overline{MAE}$	$\widehat{MAE}$	$\overline{ML}^{\alpha}$	$\widehat{ML}^{\alpha}$
100	0.0983	0.0845	0.4730 (0.049)	0.4129 (0.0474)
250	0.0956	0.0831	0.4712 (0.0504)	0.411 (0.048)
500	0.0958	0.0833	0.4664 (0.049)	0.4076 (0.05)
750	0.0962	0.0833	0.4683 (0.0509)	0.4086 (0.0529)

It would be interesting to know if this conclusion can be generalized, but it would require an exhaustive study to fully understand the effect of estimation errors.

4. Application

In this section we present an application of our method for forecasting realized volatility. Realized volatility might be defined as an empirical measure of returns volatility. In a general setting, if we suppose that the value of an asset is a semimartingale X, then the realized volatility of X over the period $[0,T]$ is its quadratic variation at time T, $[X]_T$. Thus, an estimator of the realized volatility can be defined as the sum of squared returns

$$\hat{RV}(X)_{[0,T]} = \sum_{i=1}^{N} \left(X_{t_i} - X_{t_{i-1}}\right)^2, \tag{4.1}$$

where X_{t_i}, $i = 0, ..., n$, are observed values and $0 = t_0 \leq t_1 \leq \cdots \leq t_n = T$. The first mention of realized volatility is probably Zhou (1996) but we refer the reader to Andersen et al. (2001) for a detailed justification of the realized volatility estimation.

Since each price observation is noisy (bid-ask spread, etc), a more realistic model for observed price should be $Y_{t_i} = X_{t_i} + \epsilon_{t_i}$, where ϵ_{t_i} is a random variable. In this context, it is easy to show that (4.1) is an inconsistent estimator. A common practice to estimate realized volatility is to use (4.1) and to take observations every 5 to 30 minutes. In using less observations the bias due to noise is somewhat diminished and the estimation precision becomes acceptable. However, to perform our realized volatility estimation we used the estimator of Zhang et al. (2005) which is an asymptotically unbiased estimator that allows using high-frequency data. Another good estimator is given by Martens and van Dijk (2007) which makes use of high and low observed values. The reason we prefer the former estimator is that we used trade prices and it seems that this estimator is less affected by bid/ask spread.

FIGURE 5.9. Estimated realized volatility (top panel) and volume of transactions (bottom panel).

FIGURE 5.10. Scatter plot for the first difference of the log for the realized volatility and the volume of transaction.

The data we are using are from the Trade and Quote database. We used Apple (APPL) trade prices from 2006/08/08 to 2008/02/01, which consists of 374 trading days. In order to avoid periods of lower frequency trading we used data from 9 :00 :00 to 15 :59 :59. In combination with the estimation of realized volatility we also computed the aggregated volume of transactions, see Figure 5.9.

For our time series to satisfy the required hypothesis of stationarity, we had to take the first difference of the logarithm of both series. We define $X_{1,t} = \log(\hat{\mathrm{rv}}_t) - \log(\hat{\mathrm{rv}}_{t-1})$ and $X_{2,t} = \log(\hat{\mathrm{vol}}_t) - \log(\hat{\mathrm{vol}}_{t-1})$ where $\hat{\mathrm{rv}}_t$ is the estimated volatility, and $\hat{\mathrm{vol}}_t$ is the aggregated volume of transaction and the time scale is in days.

To verify the stationarity assumption of both series we used a non-parametric change point test using the Kolmogorov-Smirnov statistic. With p-values of 21.7% and 34.1% for the series X_1 and X_2, one cannot reject the null hypothesis of stationarity.

Next, we performed parameters estimation and goodness-of-fit tests for Clayton, Frank, Gaussian and Student copulas, as proposed in Rémillard et al. (2012). From the p-values given in Table 5.5, we selected the Student copula as the best model for the copula of $(X_{1,t-1}, X_{2,t-1}, X_{2,t}, X_{2,t})$. The estimated parameters for the Student copula distribution are the degrees of freedom, $\hat{\nu} = 12.6451$, and the correlation matrix

$$\hat{R} = \begin{bmatrix} 1 & 0.6936 & -0.3628 & -0.1234 \\ 0.6936 & 1 & -0.2960 & -0.3035 \\ -0.3628 & -0.2960 & 1 & 0.6936 \\ -0.1234 & -0.3035 & 0.6936 & 1 \end{bmatrix}.$$

The associated Kendall's tau matrix is then given by

$$\tau = \begin{bmatrix} 1 & 0.4880 & -0.2364 & -0.0788 \\ 0.4880 & 1 & -0.1913 & -0.1963 \\ -0.2364 & -0.1913 & 1 & 0.4880 \\ -0.0788 & -0.1963 & 0.4880 & 1 \end{bmatrix}.$$

TABLE 5.5. P-values for tests of goodness-of fit.

Copula	p-value
Clayton	0
Frank	0
Gaussian	0.037
Student	0.0931

Finally we used our methodology to make one-period ahead predictions for out-of-sample values of the series X_1. The results of the prediction are displayed in Figure 5.11. It is not clear from the graphics but if we take the mean length of the confidence interval over the 100 forecasts, we get 2.1260 for the predictions using $\hat{X}$ while we obtain an average length of 2.1682 using $\bar{X}$, showing our proposed methodology is slightly better in this case.

FIGURE 5.11. 95% confidence interval for X_1, using (X_1, X_2) (top panel) and and using X_1 only (bottom panel).

5. Conclusion

In this paper, we presented a forecasting method for time series based on multivariate copulas and compare the performance of the predictions with the univariate version. Using the Student copula, we studied the impact of different combinations of dependencies for the vector $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$ and we saw that some combinations are more favorable for the multivariate method than others. In a similar fashion, we also saw that with the symmetrical dependence structure of the Clayton copula, the multivariate forecasting method shows only a minor advantage. We also tested the effect of estimation errors. We observed that the multivariate method keeps its advantage, but we saw that for series following a Student copula, the sample size should be taken sufficiently large in order to provides good estimated parameters if one wants to use prediction intervals. We conjectured that the results we observed might be explained by the number of estimated parameters used for the predictions but this matter would need a thorough study before to be settled. Finally, we presented a complete application with parameters estimation and goodness-of-fit test on the bivariate series of realized volatility and volume of transactions.

A. Simulation

A.1. Simulation of multivariate time series with Student copula

The Student copula is based on a multivariate Student distribution. Suppose $(\mathbf{X}, \mathbf{Y})$ is a $d = (d_1 + d_2)$-dimensional random vector which follows a Student distribution with mean 0, correlation matrix R and ν degrees of freedom. We write the matrix R as a block matrix

$$R = \begin{bmatrix} R_{\mathbf{X}} & R_{\mathbf{XY}} \\ R_{\mathbf{YX}} & R_{\mathbf{Y}} \end{bmatrix}$$

where $R_{\mathbf{X}}$, $R_{\mathbf{Y}}$, $R_{\mathbf{YX}}$ and $R_{\mathbf{XY}}$ are respectively the correlation matrices of the variables in subscript. It is easy to check that all possible joint distributions of a multivariate Student vector are also of Student distributions with respective correlation matrix and the same degrees of freedom. Let $T_{\nu,R}$ be the distribution function of a multivariate Student vector. The Student copula, noted $C_{\nu,R}$ is defined, for all $(\mathbf{u}, \mathbf{v}) \in (0,1)^{d+1+d_2}$, by

$$C_{\nu,R}(\mathbf{u}, \mathbf{v}) = T_{\nu,R}\left\{T_\nu^{-1}(u_1), ..., T_\nu^{-1}(u_{d_1}), T_\nu^{-1}(v_1), ..., T_\nu^{-1}(v_{d_2})\right\}.$$

Using Schur's complement on the correlation matrix R, it is possible to show that the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ is also a Student distribution with $\tilde{\nu} = \nu + d_1$ degrees of freedom, mean $\boldsymbol{\mu} = B\mathbf{X}$, and scale matrix $\tilde{R} = \frac{\lambda}{\tilde{\nu}}\Omega$, where $\lambda = \nu + \mathbf{x}^\top \Sigma_{\mathbf{X}}^{-1}\mathbf{x}$, $\Omega = R_{\mathbf{Y}} - R_{\mathbf{YX}} R_{\mathbf{X}}^{-1} R_{\mathbf{XY}}$, and $B = R_{\mathbf{YX}} R_{\mathbf{X}}^{-1}$. The details of the derivations are given in Appendix B.

To generate a d-dimensional time series $\{\mathbf{X}_t\}_{t=0,1,...}$ such that $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ has a Student conditional copula $C_{\nu,R}$ with marginal distributions $F_1, ...F_d$, and

$$R = \begin{bmatrix} R_1 & R_{12} \\ R_{21} & R_1 \end{bmatrix},$$

we use the following algorithm :

(1) Generate $\mathbf{Y}_0$ from a d-dimensional Student distribution with ν degrees of freedom and correlation matrix R_1.

(2) For all $t = 1, 2, ...$, generate $\mathbf{Y}_t$ from a d-dimensional Student distribution with $\tilde{\nu}$ degrees of freedom, scale matrix $\tilde{R}$ and mean $B\mathbf{Y}_{t-1}$, where $B = R_{21} R_1^{-1}$.

(3) Compute $U_{it} = T_\nu(Y_{it})$, for all $i = 1, \dots, d$.

(4) Set $(X_{1t}, ..., X_{dt}) = (F_1^{-1}(U_{1t}), ..., F_d^{-1}(U_{dt}))$.

Recall that to generate a d-dimensional random vector Y from the Student distribution $T_{\nu,\mu,R}$, one can generate V from the χ^2_ν distribution and set $Y = Z\sqrt{\nu/V} + \mu$ where Z is a d-dimensional normal vector independent of V with mean 0 and correlation matrix R.

A.2. Simulation of multivariate time series with Clayton copula

The Clayton copula is a member of the Archimedean family. A copula C_ϕ is said to be Archimedean with generator ϕ if

$$C_\phi(u) = \phi^{-1}\left\{\phi(u_1) + \cdots + \phi(u_d)\right\}$$

for any bijection $\phi : [0, 1) \to [0, \infty)$ such that $(-1)^i \frac{d^i}{d^i s}\phi^{-1}(s) \geq 0$ for all $s \geq 0$ and all $i = 0, \dots, d-1$. Archimedean copulas are uniquely defined by the generator, up to a positive scaling factor. The Clayton copula is part of the Archimedean family and is defined by the generator $\phi_\theta(t) = (t^{-\theta} - 1)/\theta$ with $\theta > 0$. Note that more generally it is possible to define a generator for the Clayton copula with parameter $\theta \geq -\frac{1}{d-1}$ but we restrict ourself to the case with positive parameter. Suppose that $(\mathbf{U}, \mathbf{V})$ is a $(d_1 + d_2)-$dimensional random vector which follows a Clayton copula $C_{d,\theta}$, where $d = d_1 + d_2$. Then it is possible to show that the conditional copula of $\mathbf{V}$ given $\mathbf{U}$ is a Clayton copula with parameter $\tilde{\theta} = \frac{\theta}{1+d_1\theta}$.

To generate a $2d$-dimensional time series $\{\mathbf{X}_t\}_{t=0,1,...}$ such that $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ follows a Clayton copula $C_{2d,\theta}$ with marginal distributions $F_1, ..., F_d$ we use the following algorithm :

(1) Generate $\mathbf{U}_0$ from the distribution $C_{d,\theta}$.

(2) For all $t = 1, 2, ...$ and $i = 1, ..., d$ compute

$$U_{it} = \left[\left(\sum_{i=1}^{d} U_{it-1}^{-\theta} - d + 1\right)\left(V_{it}^{-\tilde{\theta}} - 1\right) + 1\right]^{-1/\theta}$$

where $\mathbf{V}_t \sim C_{d,\tilde{\theta}}$ with $\tilde{\theta} = \frac{\theta}{1+d\theta}$.

(3) Set $X_{it} = F_i^{-1}(U_{it})$ for all $i = 1, ..., d$ and all $t = 1, 2, ...$

Recall that to generate a d-dimensional random vector Y from a Clayton copula $C_{d,\theta}$, we can simulate independently S from a Gamma$(1/\theta, 1)$ and $E_1, ..., E_d$ from a Exp(1), and then we set $Y_i = (1 + E_i/S)^{-\theta}$ for $i = 1, ..., d$.

B. Conditional Student distribution

Let $\mathbf{Z}^\top = (\mathbf{X}^\top, \mathbf{Y}^\top)$ be a $d = (d_1 + d_2)$-dimensional random vector which follows a multivariate Student distribution $T_d(x; \nu, b\mu, \Sigma)$, where ν is the degrees of freedom, $\boldsymbol{\mu}^\top = (\boldsymbol{\mu}_\mathbf{X}^\top, \boldsymbol{\mu}_\mathbf{Y}^\top)$ is a $(d_1 + d_2)$-dimensional real vector which is the location vector and

$$\Sigma = \begin{bmatrix} \Sigma_\mathbf{X} & \Sigma_\mathbf{XY} \\ \Sigma_\mathbf{YX} & \Sigma_\mathbf{Y} \end{bmatrix}$$

is the scale block matrix. The density function of the above multivariate Student distribution is defined as

$$t_d(\mathbf{x}, \mathbf{y}; \nu, \mu, \Sigma) = \frac{\Gamma(\frac{\nu}{2} + \frac{d}{2})}{|\Sigma|^{1/2}\Gamma(\frac{\nu}{2})(\pi\nu)^{-d/2}} \left(1 + \frac{(\mathbf{X} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})}{\nu}\right)^{-(\frac{\nu}{2} + \frac{d}{2})},$$

where $\Gamma(x)$ is the gamma function. Moreover, it is well know that all joint distributions of a multivariate Student distribution are also Student. For our concern, X follows a d_1-dimensional multivariate distribution with parameters ν, μ_1 and Σ_X.

Let $\mathbb{I}_d$ and 0_d be respectively the $d-$dimensional identity matrix and null matrix. Using Schur's method we can write $\Sigma = A \times M \times B$ where

$$A = \begin{bmatrix} \mathbb{I}_{d_1} & 0_{d_1 \times d_2} \\ \Sigma_{YX}\Sigma_X^{-1} & \mathbb{I}_{d_2} \end{bmatrix}$$

$$M = \begin{bmatrix} \Sigma_X & 0_{d_1 \times d_2} \\ 0_{d_2 \times d_1} & \Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY} \end{bmatrix}$$

$$B = \begin{bmatrix} \mathbb{I}_{d_1} & \Sigma_X^{-1}\Sigma_{XY} \\ 0_{d_2 \times d_1} & \mathbb{I}_{d_2} \end{bmatrix}.$$

Then we see that we can write the inverse of Σ the following way,

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_X^{-1} + \tilde{B}\tilde{M}^{-1}\tilde{A} & -\tilde{B}\tilde{M}^{-1} \\ -\tilde{M}^{-1}\tilde{A} & \tilde{M}^{-1} \end{bmatrix} \tag{B.1}$$

where $\tilde{A} = \Sigma_\mathbf{YX}\Sigma_\mathbf{X}^{-1}$, $\tilde{M} = \Sigma_\mathbf{Y} - \Sigma_\mathbf{YX}\Sigma_X^{-1}\Sigma_\mathbf{XY}$ and $\tilde{B} = \Sigma_\mathbf{X}^{-1}\Sigma_\mathbf{XY}$. Using (B.1), we have the decomposition

$$\begin{aligned} (\mathbf{Z} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{Z} - \boldsymbol{\mu}) \quad &= \quad (\mathbf{Y} - \boldsymbol{\mu}_\mathbf{Y} - \tilde{A}\mathbf{X})^\top \tilde{M}^{-1} (Y - \boldsymbol{\mu}_\mathbf{Y} - \tilde{A}\mathbf{X}) \\ &\quad + (\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})^\top \Sigma_\mathbf{X}^{-1} (\mathbf{X} - \boldsymbol{\mu}_\mathbf{X}). \end{aligned} \tag{B.2}$$

It then follows from (B.2) and some algebraic manipulation that the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ is a d_2-dimensional Student distribution with degrees of freedom $\tilde{\nu} = \nu + d_1$, location parameter $\tilde{\mu} = \boldsymbol{\mu}_2 + \tilde{A}\mathbf{x}$ and scale matrix $\frac{\lambda}{\tilde{\nu}}\tilde{M}$, where $\lambda = \nu + (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^\top \Sigma_{\mathbf{X}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})$.

Chapitre 5

CONCLUSION

Notre travail sur les séries temporelles donne une méthode de prévision avec un grand potentiel d'application, en finance comme dans d'autres domaines. Ces résultats ouvrent donc la porte à plusieurs projets empiriques. On peut facilement imaginer différentes séries pour lesquelles on pourrait tenter de faire des prévisions afin de construire de stratégies de gestion de portefeuille. Par exemple, la volatilité du S& P500 et la valeur des CDS (Credit Default Swap) ou les prix à l'ouverture d'un titre lors de l'ouverture de deux différentes bourses sur deux fuseaux horaires différents. D'un autre côté, nos résultats ont montré que certaines structures de dépendance peuvent grandement minimiser l'avantage que peut avoir l'utilisation des séries multivariées. Il serait donc intéressant de refaire des expériences numériques sur une liste plus exhaustive de copules afin de guider l'application pratique de notre méthode. En parallèle, il serait envisageable d'organiser nos codes et d'en faire une librairie afin de faciliter l'application.

Le modèle de marché du chapitre 3 représente bien la mécanique de transaction des ordres de marché et l'impact sur la structure de carnet d'ordres limites. Cependant, le modèle, tel qu'il est présenté dans toute sa généralité, est complexe et difficile à manipuler. Par contre, il serait possible de construire des cas particuliers pour lesquels certaines questions financières pourraient être étudiées. Par exemple, puisque le modèle généralise celui de Predoiu et al. (2011), il serait peut-être possible d'utiliser leur solution pour le problème d'execution optimale et de l'appliquer à d'autres cas que notre modèle pourrait générer. D'un point de vue plus pratique, l'article contient un exemple avec une méthode de calibration, il serait intéressant de calibrer le modèle sur des données réelles et voir comment il est possible de bien représenter ces données. Finalement, d'un intérèt strictement théorique, la théorie des semimartingales avec paramètre spatial est développée seulement pour les semimartingales continues. À l'instar de la théorie classique du calcul stochastique, il serait intéressant de généraliser cette théorie aux processus avec sauts.

Le Chapitre 4 solutionne le problème de tarification et de couverture pour des options européennes d'achat et de vente dans un modèle de marché à une période pour le carnet d'ordres limites. La première idée qui nous vient en tête est évidemment d'étendre le problème au cas multipériode. Cependant, pour le cas multipériode, notre approche nécessite de de considérer séparément les cas où la valeur du compte bancaire est positive ou négative, et ce, pour chaque période. Le nombre de cas à traiter croît donc de manière exponentielle. De plus, le problème d'optimisation n'est plus convexe (resp. concave) en général pour le cas multipériode, ce qui complexifie encore le problème. L'extension de notre méthode est donc peu envisageable et une autre approche devra être développée. Il est aussi possible d'obtenir les mêmes solutions en utilisant la méthode d'optimisation primal-dual, mais la difficulté reste la même. D'un autre côté, la simplicité du modèle nous a permis d'obtenir des conditions moins sévères pour l'absence d'arbitrage. Il serait intéressant d'étudier si ces conditions peuvent se transposer au cas continu.

Bibliographie

Andersen, T., Bollerselv, T., and Diebold, F. (2007). Roughing it up : Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, Vol 89(4) :701–720.

Andersen, T., Bollerselv, T., Diebold, F., and Labys, P. (2001). The distribution of realzed exchange rate volatility. *Journal of the American Statistical Association*, 96(453) :42–55.

Bank, P. and Baum, D. (2004). Hedging and portfolio optimization in illiquid financial markets with a large trader. *Mathematical Finance*, 14 :1–18.

Beare, B. (2010). Copulas and temporal denpendence. *Econometrica*, 78(1) :395–410.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81 :637–654.

Boyd, S. and Vandenberghe, L. (2009). *Convex Optimization*. Cambridge University Press, 7th edition.

Bush, T., Christensen, B., and Nielsen, M. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 60(1) :48–57.

Cetin, U., Jarrow, R., and Protter, P. (2004). Liquidity risk and arbitrage pricing theory. *Finance and Stochastics*, 8 :311–341.

Cetin, U., Jarrow, R., Protter, P., and Warachka, M. (2006). Pricing options in an extended black scholes economy with illiquidity : theory and empirical evidence. *The Review of Financial Studies*, 19 :493–529.

Cetin, U., Soner, H., and Touzi, N. (2010). Option hedging for small investors under liquidity costs. *Finance and Stochastics*, 14 :317–341.

Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric model time series models. *Journal of Econometrics*, 130 :307–335.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2) :174–196.

Cvitanić, J. and Ma, J. (1996). Hedging options for a large investor and forward-backward sde's. *The Annals of Applied Probability*, 6 :370–398.

Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(3) :463–520.

Durrett, R. (1996). *Probability : theory and examples*. Duxbury Press, Belmont, CA, second edition.

Fang, H.-B., Fang, K.-T., and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.*, 82(1) :1–16.

Frey, R. and Patie, P. (2002). Risk management for derivatives in illiquid markets : a simulation study. *Advances in Finance and Stochastics*, pages 137–159.

Frey, R. and Stremme, A. (1997). Market volatility and feedback effects from dynamic hedging. *Mathematical Finance*, 7 :351–376.

Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009). The advent of copula in finance. *The European Journal of Finance*, 15 :609–618.

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

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

Jarrow, R. (1992). Market manipulation, bubbles, corners and short squeezes. *Journal of financial and quantitative analysis*, 27(3) :311–336.

Jarrow, R. (1994). Derivative security markets, market manipulation and option pricing. *Journal of financial and quantitative analysis*, 29 :241–261.

Jarrow, R., Protter, P., and Roch, A. (2010). A liquidity based model for asset price bubbles. *Johnson school research paper series*. Electronic version available at http ://ssrn.com/abstract=1561574.

Kunita, H. (1990). *Stochastic flows and stochastic differential equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Kyle, A. (1985). Continuous acutions and insider trading. *Econometrica*, 53 :1315–1335.

Liu, H. and Yong, J. (2005). Option pricing with an illiquid underlying asset market. *Journal of Economic, Dynamics and Control*, 29 :2125–2156.

Martens, M. and van Dijk, D. (2007). Measuring volatility with the realized range. *Journal of Econometrics*, 13 :181–207.

Nelsen, R. B. (1999). *An introduction to copulas*, volume 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Obizhaeva, A. and Wang, J. (2005). Optimal trading strategy and suply/demand dynamics. *National bureau of economics research*. Working paper.

Patton, A. (2013). Copulas methods for forecasting multivariate time series. *Handbook of Economic Forecasting, Volume 2*.

Platen, E. and Schweizer, M. (1998). On feedback effects from hedging derivatives. *Mathematical finance*, 8(1) :67–84.

Predoiu, S., Shaikhet, G., and Shreve, S. (2011). Optimal execution in a general one-sided limit-order book. *SIAM Journal of Financial Mathematics*, 2 :183–212.

Protter, P. E. (2004). *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition. Stochastic Modelling and Applied Probability.

Rémillard, B., Papageorgiou, N., and Soustra, F. (2012). Copula-based semiparametric models for multivariate time series. *Journal of Multivariate Analysis*, 3(1).

Roch, A. (2011). Liquidity risk, price impacts and the replication problem. *Finance and Stochastics*, 15(3) :399–419.

Roch, A. and Soner, H. (2013). Resilient price impact of trading and the cost of illiquidity. *International Journal of Theoretical and Applied Finance*, 16(6).

Rogers, L. and Singh, S. (2010). The cost of illiquidity and its effects on hedging. *Mathematical Finance*, 20(4) :597–515.

Rémillard, B. (2013). *Statistical Methods For Financial Engineering*. Chapman & Hall.

Schönbucher, P. and Wilmott, P. (2000). The feedback effect of hedging in illiquid markets. *SIAM Journal on Applied Mathematics*, 61(1) :237–272.

Shreve, S. E. (2004). *Stochastic calculus for finance. II*. Springer Finance. Springer-Verlag, New York. Continuous-time models.

Simard, C. (2014). General model for limit order books and market orders. *SSRN*. available at SSRN.

Sokolinskiy, O. and Van Dijk, D. (2011). Forecasting volatility with copula-based time series models. *Tinbergen Institute Discussion Paper*.

Soustra, F. (2006). Pricing of synthetic CDO tranches, analysis of base correlations and an introduction to dynamic copulas. Master's thesis, HEC Montréal.

Zhang, L., Mykland, P., and Aït-Sahalia, Y. (2005). A tale of two time scales : Determining integrated volatility with noisy high-frequency data. *Journal of The American Statistical Association*, 100(472) :1394–1414.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1) :45–52.