.

.

Master Probabilité et Finance

Rapport sur article.

"Simulating and analyzing order book data : The queue-reactive model"

Auteurs :
M. Zakaria Iraqi

14 avril 2019

Table des matières

Introduction

Le but de l'article est de simuler la dynamique d'un carnet d'ordres limites (Limit Order Book -LOB), du fait de son rôle dans la régulation, ainsi que de détecter les tendances qui gouvernent le orix. Les carnets d'ordre sont l'endroit où les ordrent se croisent. Un ordre inséré est alors comparé à un ordre déjà existant dans la machine, si on a une intersection, une transaction est exécute, sinon, l'ordre est stocké dans le carnet d'ordres.

L'étude de la dunamique des ordres est l'étude et la modélisation des insertions et des annulations dans le carnet d'ordre conditionnellement à son état. PLus l'état du carnet 'ordre est décrit avecc un nombre élevé de paramètres, plus le modèle devient complexe (et fiable ?).

On trouve actuellement dans la littérature beaucoup de modèle automatiques du traitement des carnets d'ordres : des modèles de type "zéro intelligence", à d'autres de mean-field games ou d'autres basés sur des EDPs.

Dans cet article, nous allons étudeir quatre différents modèles, où on modélisera les modèles du carnet d'ordre en incluant de plus en plus de facteurs.

En pratique, les acteurs du marché (ou leurs algorithmes) analysent de nombreuses quantités avant d'envoyer un ordre donné à un niveau donné. L'une des variables les plus importantes de ce processus de décision est probablement la distance entre leur prix cible et leur "prix de référence du marché", typiquement le prix moyen. Ce prix de référence est lié aux flux d'ordres puisqu'il est généralement déterminé par l'État du LOB. Cette interconnexion rend la conception des modèles LOB assez complexe. Pour surmonter cette difficulté, les intervalles de temps d'intérêt sont divisés en périodes de référence constante, et on considére deux parties dans la modélisation. La LOB est étudié comme une file d'attente de Markov pendant les périodes où le prix de référence est constant. Ensuite, la dynamique des prix de référence est étudiée. Un tel cadre est particulièrement adapté aux actifs de large tick pour lesquels les périodes de prix de référence constants sont assez longs et permettent des estimations précises des paramètres.

Deux informations sont disponibles aux participants à haute fréquence : le flux historique et l'état actuel du carnet d'ordre. Le papier étudie plus en détail l'impact du carnet d'ordre dans les décisions des acteurs du marché. Ce qui permettra d'analyser la réaction des acteurs du marché en fonction de la configuration du carnet d'ordre.

Le premier chapitre consistera en une description générale des modèles choisis, et quelques résultats donnés par l'article.

Nous testerons finalement le modèle avec les résultats finaux pour déterminer la véracité de ces derniers

Chapitre 1

Dynamique du carnet d'ordre à prix de référence constant

De manière générale, la manière de capter et d'estimer les données sera expliquée dans une prochaine partie.

1.1 Mise en situation

1.1.1 Le carnet d'ordre

Le carnet d'ordre peut être modélisé, à prix de référence fixé, comme un vecteur de dimension 2K, où K est le nombre de limites disponibles de chaque côté. La partie gauche représentera le bid, la partie droite représentera l'ask, en pratique, en prendra K=4, car on remarque peu de significativité des autres limites et de leur dynamique.

— bid : $(Q_i)_i : i \in [\![-K; -1]\!]$

— ask : $(Q_i)_i : i \in [\![1; K]\!]$

Le nombre d'ordres à l'échelle Q_i sont notés q_i. Les actions possibles dans le carnet d'ordre sont :

— Ordre d'achat pour le bid et annulation de cet ordre.

— Ordre de vente pour l'ask et annulation de cet ordre.

— Achat et Vente d'ordres de marché.

On modèlisera $(q_i)_i$ de ce fait comme un processus de markov avec trois sauts possibles par position, et seulement un seul, tous de quantité 1 : Un saut vers l'avant (de la quantité q_i à la quantité $q_i + 1$), un saut vers l'arière (de la quantité q_i à la quantité $q_i - 1$), et un saut "nul" (rester à la quantité q_i à l'instant i), et ce pour chaque position i. On note

$e_i = \delta_i$, c'est à dire que le saut s'est fait suivant la position i. On obtient :

$$\begin{aligned}
\mathrm{Q}_{q,q+e_i} &= f_i(q) \\
\mathcal{Q}_{q,q-e_i} &= g_i(q) \\
\mathcal{Q}_{q,q} &= -\sum_i f_i(q) - g_i(q) \\
\mathcal{Q}_{q,p} &= 0 \ sinon
\end{aligned} \tag{1.1}$$

Sous certaines hypothèses sur les intensités des sauts, on a des résultats d'ergodicité. Ce qui assure que le système n'explose pas, et que la taille des queues est décroissante quand elle devient grande.

1.1.2 Le prix de référence

Dnas nos cas de départ, on considère un prix de référence fixé, p_{ref} doir représenter le "centre" du carnet d'ordres. Si le bid-ask spread est de 1 tick :

$$p_{ref} = \frac{p_1 + p_{-1}}{2}$$

Si le spread est plus grand qu'un tick, le mid-price n'est plus approprié, on choisit donc entre deux prix : $p_{mid} \pm \frac{tick\ size}{2}$.

1.2 Modèle 1 : Queues du carnet d'ordres indépendantes

On considère un prix de référence p_ref fixée.

Les principales propriétés du premie rmodèle sont :

- Les flots entrants, sortants à chaque niveau sont indépendants.
- Les intensité de ces processus ne dépendent que de la portion du volume où ils sont placés.
- A partir d'une certaine limite, les ordres d'arrivée des ordres limites sont indépendantes(on a atteint la stationnarité).
- Le carnet d'ordre est symétrique(et donc les intensités relatives aussi).
- les intensités sont notées $\lambda^L, \lambda^C, \lambda^M$, relativement aux ordres limites, d'annulation et de marché.
- De ce fait :

$$\begin{aligned}
\mathrm{f}_i(q) &= \lambda_i^L(q_i) \\
g_i(q) &= \lambda_i^C(q_i) + \lambda_i^M(q_i)
\end{aligned} \tag{1.2}$$

Ce modèle sera le plus simple : chaque queue a sa propre dynamique, les intensités des trois flots seront en fonction de leurs états, ce qui dépassera le cadre "zéro-intelligence".

On aura aussi intérêt à renormaliser les tailles moyennes des évenement de la queue, et de travailler en ticks et non en limites.

D'un autre côté, on peut déduire une distribution du prix moyen d'un actif en utilisant les données relatives aux stocks, fitter notre modèle par rapport à ce dernier, et estimer ensuite la loi de Poisson du processus. Ceci tout en prenant en compte que le prix moyen change quand l'une des queues est vide.

Après estimation des paramètres selon ce modèle, le résultat constaté est que les distributions asymptottiques des tailles des queues approchent les distributions réelles. Néanmoins, la probabilité d'éxecution d'un ordre limite par ce modèle n'est pas proche du cas réel. Cela nous fait penser que l'éxecution d'un ordre limite dépend aussi des autres tailles des queues avoisinantes. Ce que l'on explorera dans le modèle 2.

1.3 Modèle 2 : Cas de dépendances

1.3.1 Modèle 2-a : deux ensembles de queues indépendantes

Dans ce contexte, on considère le même cas général que précedemment, de plus :

— De chaque côté, la première queue est une fonction de son état (markovienne).
— la 2-ème queue dépend en même temps de son état et de l'état de la première, que ce soit pour le bid que pour l'ask.
— Les autres queues (à partir de la 3ème) se comportent comme dans le cas du modèle 1. En supposant que le comportement des 2 premières queues est indépendant du reste.

Ces assomptions sont justifiées du fait que les traders en général et les traders haute fréquence en particulier placent leurs ordres en fonction des évèneements qui précèdent. De plus, le fait de rendre indépendantes les deux premières queues par rapport au reste suppose aussi que l'on se place dans un cadre de large ticks.

Le cas du modèle 2-a est un modèle de quasi naissance-décès(QBD). On doit étudier la densité jointe du couple (q_1, q_2)

1.3.2 Modèle 2-a : Modélisation de la dépendance bid/ask

Ce modèle est facilement interprétable dans un classifieur de type réseau de neuronnes : On introduit une fonction de classification $S_{m,,l}(x)$ qui prend 4 valeurs distinctes :

— $S_{m,,l}(x) = Q^0$ si $x = 0$.

— $S_{m,,l}(x) = Q^-$ si $0 < x \leq m$, c.a.d si la quantité en x est inférieure à un seuil donné.(En pratique, on choisra m le quantile inférieur à 33% de la première queue).
— $S_{m,,l}(x) = \hat{Q}$ si $m < x \leq l$ c.a.d si la quantité en x est entre deux seuils donnés.(En pratique, on choisra l le quantile supérieur à 33% de la première queue).
— $S_{m,,l}(x) = Q^+$ si $l < x$.

De ce fait, les intensités de sauts des différents ordres est modélisée en fonction des tailles du carnet d'ordre, et aussi du côté opposé du carnet d'ordre. De plus, on suppose aussi que les ordres limites et de marché ne sont envoyés qu'aux premiers et seconds ordres. Par exemple les intensité en Q_1 dépendent de q_1 et de $S_{m,l}(q_{-1})$. Le changement de régime instauré dans le modèle 2-a reste valable pour ce modèle.

De plus, il est à noter que l'on peut trouver d'autres fonctions des intensité $Q_\pm$ en fonctuin du bid/ask imbalance, ou encore du spread...

En pratique, la densité jointe du modèle 2-b Se comporte très bien en dehors des zones à faible quantités par rapport au quantile. Vu que ce modèle remplace ces dernières valeurs par des couples (x,0) ou (0, y), et que le marché ne peut pas rester longtemps dans ces états, sans pour le moins changer le fait que les prix de référence sont constants. Ce qui va être implémenté dans les modèles suivants.

1.3.3 Le modèle à queue réactive

Ceci est le dernier modèle de l'article, il inclut aussi un changement du prix de référence comme suit :

Soit δ la valeur d'un tick, on suppose que le prix de reférence p_{ref} change avec une probabilité θ quand le prix moyen p_{mid} change : quand le prix moyen augmente(resp. diminue), le prix de référence augmente(resp. diminue) avec probabilité θ, si $q_1 = 0$(resp. $q_{-1} = 0$. De ce fait, la modification du prix de référence est causée par l'un de ces trois évènements :

— L'insertion d'un ordre d'achat(resp. de vente) limite dans le spread bid-ask alors que Q_1(resp. $Q_{-1})estvideaummemoment.Annulationdudenierordrelimitel'unedesdeuxqueues.$
— Un ordre de marché qui consomme toute la liquidité eu best bid ou best ask.

De plus, quand le prix de référence change, tous les q_i sont translatés dans les sens de changement de ce dernier.

Pour rajouter de l'information supplémentaire, on suppose que, avec $\theta^{reinitialisation}$, est retiré autour du nouveau prix de référence. C'est la probabilité qu'un évènement exogène soit arrivé, ce qui induit un changement réel du carnet d'ordre en long terme. De ce fait, la dynamique du marché est un (2K+1)-processus de Markov, vu que l'on rajoute p_{ref} et son impact à différentes limites.

Chapitre 2

Implémentation

2.1 Données utilisées

Nous allons utiliser les données de Google disponibles dans le site `https://lobsterdata.com/info/DataSamples.php`. Nous allons utiliser les données qui disposent de 5 limites des deux côtés pour inférer nos résultats (voir notebook).

Tous les résultats de l'implémentation sont synthétisés dans le notebook accompagnant le rapport.

2.2 Modèle 1

Nos résultats différent des résultats trouvés dans l'article, forcément du fait que notre manière de définir Q_i en terme de meilleurs limites, sans indéxer par rapport à la distance du tick, doit rendre nos résultats faux. Mais dans notre cas, le tick est très petit (de l'ordre de 10^{-5} de la valeur du stock).De plus, le spread moyen est de 30 ticks.

Même en choisissant un modèle où $q_{\pm 1}$ peut être égal à 0, on n'obtient tout de même pas le même comportement que celui de l'article.

2.3 Modèle 2

2.3.1 Modèle 2-a

Comme trouvé dans l'article, nous avons trouvé que les intensités étaient moindres quand la valeur opposée est non vide. Comparée à lorsque cette dernière ne l'est pas.

2.3.2 Modèle 2-b

Le problème que l'on retrouve dans notre cas est que les quantités sont trop centrées sur la valeur moyenne, de ce fait, on trouve rarement $q > 2$, ce qui se traduit sur les quantiles retrouvés. On ne trouve pas les mêmes résultats que ceux de l'article : le cas $q_{-1} > 1$ qui correspond au valeur supérieures au quantile supérieur à 30% sont celles dont les intensités sont dominantes.

Conclusion et perspectives

Cette étude avait pour but de modéliser un carnet d'ordre sous différents modèles et de tester la significativité de ces derniers. Dans ce cadre, nous avons étudié des carnets d'ordres, mais nos données étaient comprises de stocks avec de très petits ticks, ce qui rendait l'interprétabilité de nos données assez différente de celle donnée dans l'article, tout en gardant néanmoins plusieurs points communs. Il aurait été préférable d'utiliser un stock à large tick, mais je n'avais pas trouvé de données avec plusieurs limites dans le carnet d'ordre qui reflètent ceci. On pourrait aussi développer alors un modèle de type queue réactive dans ce sens, qi aura plus d'intérêt pratique.

Bibliographie

[1]- "Simulating and analyzing order bookdata : The queue-reactive model" - [Weibing Huang , Charles-Albert Lehalle and Mathieu Rosenbaum] - 2014.

[2] - "Time-varying Limit Order Book Networks" [Wolfgang Karl Härdle , Shi Chen , Chong Liang, Melanie Schienle] IRTG 1792 Discussion Paper 2018-016