

ENSAE Paris



ALGORITHMIC TRADING

HAWKES MODEL FOR PRICE AND TRADES HIGH-FREQUENCY DYNAMICS^[1]

EMMANUEL BACRY ET JEAN-FRANÇOIS MUZY

Professeur : Olivier Guéant

Marvin IVARD et Clément SIMON

30 avril 2024

1 Résumé de l'article

1.1 Introduction

L'article [1] intitulé 'Hawkes model for price and trades high-frequency dynamics' propose un modèle multivarié basé sur le processus de Hawkes pour analyser la dynamique des prix et des transactions sur les marchés financiers à haute fréquence. Ce modèle est principalement caractérisé par quatre noyaux qui décrivent différents aspects de l'auto-excitation des arrivées de transactions, la mean-reversion des changements de prix, l'impact des arrivées de transactions sur les changements de prix, et la rétroaction des changements de prix sur l'activité de trading.

Caractéristiques du modèle

1. **Auto-excitation des arrivées de transactions** : ce noyau capture la tendance des transactions à s'auto-stimuler, ce qui signifie que chaque transaction peut augmenter la probabilité d'une transaction future.
2. **Mean-reversion des variations de prix** : ce noyau modélise la tendance des prix à revenir à la moyenne, reflétant la volatilité et les corrections de prix après des mouvements de prix importants.
3. **Impact des arrivées de transactions sur les variations de prix** : il s'agit de comprendre comment les transactions influencent les variations directes de prix, en tenant compte de l'effet immédiat des achats et des ventes sur le marché.
4. **Rétroaction des variations de prix sur l'activité de trading** : ce noyau étudie la manière dont les variations de prix peuvent affecter le comportement des agents, en stimulant potentiellement l'activité de trading en réponse aux mouvements de prix.

Applications et résultats

Le modèle permet d'estimer l'ensemble du profil d'impact du marché à partir de données anonymes. Les auteurs montrent que ces noyaux peuvent être estimés à partir des intensités moyennes conditionnelles empiriques. Ils fournissent des exemples numériques, des applications à des données réelles et des comparaisons avec des approches antérieures.

Importance et implications

Ce modèle, phénoménologique, est particulièrement utile pour analyser la microstructure des marchés financiers, et ainsi capturer des faits stylisés tels que la mean-reversion à haute fréquence et la forme caractéristique des fonctions d'impact du marché à différentes échelles de temps. Il met en avant des faits stylisés autour du market impact, principalement la concavité, la racine carrée et la relaxation de l'impact du marché dans le cadre d'un meta-order.

1.2 Résumé parties par parties

1.2.1 Modèle basé sur Hawkes pour la microstructure du marché

Cette section construit un modèle mathématique détaillé utilisant les processus de Hawkes, reliant les données des transactions à haute fréquence à la dynamique théorique du marché. Elle aborde les complexités de la microstructure du marché grâce à un cadre sophistiqué qui intègre à la fois les mécanismes d'auto-excitation et d'excitation croisée au sein du marché.

Définition du modèle

Le modèle utilise des processus ponctuels auto-excités multivariés, représentant les transactions et les variations de prix comme des composantes d'un processus ponctuel à quatre dimensions P_t , défini comme :

$$P_t = \begin{pmatrix} T_t^- & T_t^+ & N_t^- & N_t^+ \end{pmatrix}$$

où T_t et N_t représentent respectivement les transactions et les mouvements de prix, respectivement à l'achat et à la vente.

Le vecteur d'intensité conditionnelle λ_t est déterminé par des événements passés :

$$\lambda_t = \begin{pmatrix} \lambda_t^{T-} & \lambda_t^{T+} & \lambda_t^{N-} & \lambda_t^{N+} \end{pmatrix}$$

La dynamique des intensités conditionnelles λ_t sont modélisés à l'aide du processus de Hawkes :

$$\lambda_t = M + \Phi * dP_t$$

où M représente la baseline intensity matrix et Φ est la matrice du noyau de mémoire, décomposée comme suit :

$$\Phi_t = \begin{pmatrix} \Phi_t^T & \Phi_t^F \\ \Phi_t^I & \Phi_t^N \end{pmatrix}$$

Chaque sous-matrice $(\Phi^T, \Phi^F, \Phi^I, \Phi^N)$ capture différents aspects de la dynamique du marché, tels que l'auto-excitation des transactions (Φ^T) , l'impact des transactions sur les changements de prix (Φ^I) , l'influence des changements de prix passés sur les prix futurs (Φ^N) , et la rétroaction des prix sur les transactions futures (Φ^F) . Notons que cette écriture matricielle fait ainsi état de dynamiques clusterisées en sous-matrices d'interaction.

Stabilité et estimation

La stabilité des processus ponctuels est une composante essentielle de l'analyse et de l'applicabilité du modèle. Mathématiquement, il s'agit d'étudier la stricte stationnarité du processus. Dans le cadre d'un processus de Hawkes multivarié, elle exige classiquement (la démonstration reposant sur une condition suffisante en moment d'ordre 1 de l'intensité) que les valeurs propres de la matrice du noyau intégré Φ soient inférieures à 1. Dès lors, l'estimation des paramètres du modèle peut être réalisée (entre autres) à l'aide de techniques telles que l'estimation du maximum de vraisemblance, adaptée aux données des

transactions à haute fréquence.

Fonction de covariance des prix et des échanges

La section commence par examiner les fonctions de covariance des prix sous l'angle du processus multivarié de Hawkes :

$$v_{\tau}^{(h)} = h^{-1} \text{Cov}(P_{t+h+\tau} - P_{t+\tau}, P_{t+h} - P_t),$$

où P_t représente le processus ponctuel quadridimensionnel des transactions et des changements de prix. L'article fournit une expression basée sur la transformée de Laplace pour relier la matrice de covariance $v_{\tau}^{(h)}$ aux noyaux du modèle :

$$\hat{v}_z^{(h)} = \hat{g}_z^{(h)} (I - \hat{\Phi}_z)^{-1} \Sigma (I - \hat{\Phi}_z^{\dagger})^{-1},$$

où Σ désigne la matrice diagonale composée des intensités moyennes et Φ comprend les noyaux de Hawkes.

Simulations numériques

Pour reproduire les comportements observés sur le marché, il faut alors simuler les paramètres du modèle comme cela :

- La configuration de la simulation utilise un algorithme d'amincissement (thinning) proposé par Ogata pour générer le processus ponctuel et ses trajectoires de données pour les prix et les transactions.
- Les données simulées permettent ainsi de valider la capacité du modèle à reproduire des caractéristiques clés et attendues, telles que la mean-reversion des prix et l'auto-excitation dans les processus d'arrivée des transactions. Cela est rassurant.

Analyse des propriétés de diffusion

Une analyse approfondie des propriétés diffusives du modèle permet de mieux comprendre le comportement du marché à long terme :

$$\frac{1}{\sqrt{h}} (P_{ht} - E[P_{ht}]) \xrightarrow{\text{loi}} (I - \hat{\Phi}_0)^{-1} \Sigma^{1/2} W_t,$$

où W_t est un mouvement brownien multidimensionnel standard. Ce résultat souligne la capacité du modèle à converger vers un processus gaussien à grande échelle, offrant ainsi une base théorique à la dynamique des prix et des échanges observée sur les marchés financiers. C'est un résultat notable qui relie fondamentalement la dynamique mesoscopique (en termes de microstructure) du marché à la vision plus macroscopique brownienne, d'un marché ergodique qui digère en quelque sorte l'information et tend à l'efficience.

Analyse empirique

L'analyse empirique confirme l'efficacité du modèle dans la reproduction et la prédiction des comportements des données du monde réel :

- Les structures de covariance estimées à partir des données à haute fréquence s'alignent bien sur les prédictions du modèle qui exhibe une certaine mémoire des transactions passées.
- L'autocorrélation observée dans les signes de transaction et l'impact des transactions sur la dynamique des prix sont cohérents avec le cadre théorique fourni par le modèle de Hawkes, utiles pour comprendre et expliquer la mémoire des phénomènes.

1.2.2 Modèle avec agents étiquetés

Cette partie de l'article présente une extension du modèle multivarié de Hawkes pour incorporer des agents étiquetés, fournissant un cadre pour analyser l'impact spécifique des participants, dès lors qu'il est possible de lever l'anonymat sur l'identité de certains participants, à commencer par ses propres trades et modèles.

Incorporation d'agents étiquetés

La modification du modèle comprend l'ajout d'une composante déterministe pour représenter les actions des agents étiquetés. Cette composante, A_t , est ajoutée au cadre existant, reflétant leurs activités :

$$A_t = \begin{pmatrix} A_t^- \\ A_t^+ \end{pmatrix},$$

où A_t^- et A_t^+ correspondent respectivement aux ordres de vente et d'achat de l'agent étiqueté.

L'impact de ces agents labellisés sur le marché est modélisé par un terme supplémentaire dans l'équation d'intensité :

$$\lambda_t = M + \Phi * dP_t + \Theta * dA_t,$$

où Θ représente la matrice d'influence des transactions labellisées sur la dynamique existante du marché, détaillant la manière dont ces actions spécifiques modifient le comportement des transactions et des prix.

1.2.3 Impact sur le marché dans le modèle avec agents étiquetés

Cette partie enrichit la compréhension de l'impact du marché par une approche analytique détaillée, en considérant les contributions spécifiques des agents labellisés dans le cadre d'un processus de Hawkes multivarié. Cette partie ne se contente pas d'améliorer le modèle théorique, elle fournit également des informations pratiques qui peuvent être appliquées, au bénéfice des chercheurs et des praticiens intéressés par la dynamique des marchés et l'évaluation de l'impact.

Calcul analytique de l'impact sur le marché

Le changement attendu des conditions du marché dû à ces transactions est calculé analytiquement :

$$E[\lambda_t] = (\delta I + \Psi) * (M + \Theta * dA_t),$$

indiquant comment les transactions labélisées modifient la dynamique du marché au fil du temps.

MI_t mesure l'impact des actions d'un agent étiqueté sur les prix du marché :

$$MI_t = E[X_t] = E[N_t^+ - N_t^-].$$

Ce profil est essentiel pour comprendre l'évolution temporelle des changements de prix dus à des actions spécifiques et pour quantifier les impacts permanents et temporaires de ces transactions.

Validation théorique et empirique

La validation empirique du modèle d'impact sur le marché est discutée, en se concentrant sur la comparaison des prédictions théoriques avec les données observées. La capacité du modèle à prédire et à expliquer les comportements observés sur le marché constitue un test solide de sa validité. Il est mis en évidence que le modèle capture avec succès les nuances de la dynamique du marché induite par les actions de négociation individuelles.

1.2.4 Estimation non paramétrique des fonctions du noyau

Cette partie fait progresser l'application pratique du modèle de Hawkes en détaillant une méthode d'estimation non paramétrique de ses noyaux. Cette méthode permet aux chercheurs et aux praticiens d'adapter le modèle aux données réelles du marché, améliorant ainsi sa capacité à produire un modèle explicatif "off-the-shelf". L'approche approfondie de l'estimation garantit que le modèle est non seulement solide sur le plan théorique, mais aussi viable sur le plan empirique.

Technique d'estimation

L'approche repose sur la résolution d'une équation convolutionnelle, fondamentalement issue de statistiques d'ordre 1 et 2 du processus d'intensité conditionnelle :

$$g_t = \phi * (\delta I + g_t),$$

où g_t représente la matrice des espérances conditionnelles, et δI est la matrice d'identité mise à l'échelle par la fonction delta de Dirac. Cette équation est une équation de Fredholm du second type, qu'il est possible de résoudre numériquement pour obtenir les estimations non paramétriques des noyaux de Hawkes. Ainsi :

- Les noyaux sont estimés en tenant compte des intensités conditionnelles et de l'historique des événements, en utilisant les moyennes empiriques de ces événements pour résoudre les équations du modèle.

- Un élément clé de l'estimation consiste à traiter la nature auto-excitante des processus, ce qui nécessite un traitement minutieux des dépendances temporelles dans les données (nous discuterons bien entendu cela).

Validation et application

La validation des noyaux estimés est réalisée par des études de simulation et une comparaison avec des données empiriques. Les résultats confirment que la méthode d'estimation non paramétrique permet de saisir avec précision les caractéristiques des noyaux :

- Des données simulées à partir de noyaux bien connus (exponentiels, loi de puissance, etc.) sont utilisées pour tester la robustesse de la méthode d'estimation, montrant un bon accord avec les prédictions théoriques.
- L'application à des données de trading réelles démontre l'utilité pratique de la méthode, car elle capture efficacement la dynamique observée dans les comportements du marché.

1.2.5 Application à des données réelles

Cette partie démontre efficacement l'application du modèle multivarié du processus de Hawkes dans un contexte réel, en validant ses hypothèses et en soulignant sa capacité à fournir des analyses perspicaces de la dynamique du marché. Cette application souligne le potentiel du modèle en tant qu'outil précieux d'analyse financière, contribuant de manière significative à la recherche en finance quantitative appliquée aux processus ponctuels et aux applications pratiques par les Quants.

Estimation de la matrice du noyau à partir de données réelles

Dans l'application pratique de leur modèle, Bacry et Muzy ont procédé à l'estimation des paramètres, en particulier les noyaux de Hawkes Φ et les vecteurs d'intensité exogène M , en calibrant sur les données des contrats à terme pour l'EuroStoxx et l'Euro-bund. Cette analyse s'est étalée sur 800 jours avec environ 40 000 transactions par jour. Pour ce faire, ils ont utilisé une méthode d'estimation non paramétrique. Ils arrivent alors aux résultats suivants, en faisant parler leur modèle :

- Les paramètres estimés pour l'EuroStoxx et l'Euro-bund se sont avérés stables et comparables.
- Les variations de prix ont montré un comportement de mean-reversion à haute fréquence, ce qui correspond aux observations empiriques.
- Une corrélation significative a été observée dans la directionnalité des arrivées d'ordres de marché.
- Les autocorrélations des incréments de prix ont montré des dépendances à long terme décrites par une loi de puissance.

Implications pratiques et recherches futures

L'application réussie du modèle de processus de Hawkes à des données réelles confirme non seulement sa validité théorique, mais ouvre également la voie à de futures recherches et à des mises en œuvre pratiques :

- Les résultats encouragent à poursuivre l'exploration en vue d'affiner le modèle, par exemple en incorporant des données plus granulaires ou des variables supplémentaires susceptibles d'affecter le comportement du marché.
- Les implications pratiques comprennent l'utilisation potentielle du modèle dans les stratégies de trading algorithmique, la gestion des risques et la surveillance réglementaire afin de prévoir et d'atténuer les scénarios d'abus de marché.

2 Regard critique sur l'article

2.1 Forces du papier

- **Ce papier a ouvert la voie à l'utilisation des processus de Hawkes pour analyser le market impact.** Ce papier ne se contente pas de proposer une méthode supplémentaire pour examiner les dynamiques du marché, mais il élargit la palette d'outils disponibles pour interpréter, comprendre, calibrer, et potentiellement prédire les conséquences des actions individuelles sur le comportement global du marché. Il donne une clé de lecture particulièrement intéressante en termes d'endogénéité versus exogénéité des phénomènes.
- **Le modèle intègre de multiples aspects du comportement du marché,** y compris l'auto-excitation des transactions, la mean-reversion des prix et l'impact des transactions sur les variations de prix, offrant ainsi une vision holistique de la microstructure du marché dans la formation du prix.
- **La capacité du modèle à s'adapter aux données réelles des transactions** et son utilisation pour expliquer les phénomènes observés dans les transactions à haute fréquence, comme l'impact des transactions sur le marché, le rendent très pertinent pour les praticiens et les chercheurs.

2.2 Faiblesses du papier

- **Il est toujours utile de se demander comment construire un signal de trading à partir de ces modèles.** Le recours à des processus de Hawkes n'a d'autre souhait que de vouloir faire sensiblement mieux qu'un modèle type Poisson non homogène en introduisant des auto-excitations. En intraday, et comme suggéré par Bacry et Muzy page 26, on peut réaliser un bon fit de l'intensité de base (passage de μ à μ_t) et ainsi capturer la saisonnalité exogène dans la baseline intensity du processus.
- **Or, cette astuce n'est pas réellement satisfaisante** car μ est pourtant le reflet d'activité purement exogène. Qu'est-ce qui dès lors relève de phénomènes purement endogènes ? Le papier ne propose pas de test permettant de trancher. Un test statistique permettrait de considérer un re-calibrage du modèle dès l'instant que le modèle shift drastiquement en baseline intensity. A ce titre, on peut citer le papier de M. Lallouache et D. Challet [2] :

B. Whole-day fits *The relative stability of the branching ratio and the high p-values of e.g. KS tests encourages us to fit longer time windows. As we will see, this is possible for a full day at a time. In this case, μ cannot be considered constant anymore (see Fig. 4). As suggested by Bacry and Muzy [4], a time-of-the-day dependent background intensity is a good way to account for the intraday variation of activity. This method has the advantage of not mixing data from other days like classic detrending methods do. We thus approximate, for each day, μ_t by a piecewise linear function with knots at 0 am (when the series begin), 5 am, 9 am, 12 pm, 4 pm and at the end of the series. The 6 knots values are additional fitting parameters.*

- Certes, les modèles type Hawkes sont un bon modèle phénoménologique pour se représenter le flux d'ordres en termes de microstructure de marché. En revanche, pour des **marchés OTC** comme les contrats futures en FX, on imagine assez mal comment attribuer des événements types endogènes et ses 'répliques' alors qu'on est sur un dealer market où seuls quelques acteurs traitent et impactent le marché entre eux. L'astuce des labeled agents paraît tout aussi artificielle et corrective.
- Pour intégrer les résultats du papier en termes de market impact à des **modèles d'exécution optimale**, on peut citer par exemple le papier d'Alfonsi et Blanc [3].
 - Le modèle original de Bacry et Muzy est étendu pour inclure des noyaux de décroissance généraux pour les processus de prix et d'intensité, au-delà de la décroissance exponentielle généralement utilisée dans les modèles de Hawkes. Cela permet au modèle de capturer une plus grande variété de comportements empiriques de décroissance observés dans les données du marché.
 - Un nouvel élément appelé *délai d'ajustement du marché* est introduit, qui saisit le temps de réaction initial du marché après une transaction importante. Cet élément reflète l'observation réelle selon laquelle le marché prend quelques secondes pour s'ajuster après un événement important.
 - Un protocole de calibration est développé pour estimer les paramètres de ce modèle de Hawkes étendu à partir de données réelles sur les stocks. Ce protocole comprend une analyse minutieuse de la fonction d'auto-covariance du nombre de sauts et l'application d'une méthode généralisée des moments pour l'estimation des paramètres.
- En pratique, la résolution du système liée à l'équation convolutionnelle autour des moments d'ordre 1 et 2 (covariance) de l'intensité conditionnelle est très idéaliste. **Il semble lourd de calibrer un tel modèle sur un échange pour un stock donné**, tant les dynamiques sont fines et bruitées. On aurait ainsi pu, comme l'ont fait Lee et al. en 2017 [4], incorporer des jumps de taille différents (processus de Hawkes marqués) selon le volume traité, en superposant davantage de noyaux, ou alors envisager de conditionner les processus selon les types d'ordre (cancellations, etc.) pour capturer l'agressivité du marché et différencier les phénomènes, tel que cela a été fait par Bacry, Jaisson et Muzy en 2014 [5]. Néanmoins, cela complexifie énormément le modèle et augmente la dimension, ce qui perd de vue l'objectif initial de ce modèle phénoménologique : proposer un cadre d'analyse explicatif et particulièrement intuitif pour comprendre la dynamique d'impact sur les prix et l'activité de trading. **Notons que les modèles de Hawkes sont très sensibles au fléau de la dimension dès qu'il s'agit d'augmenter le nombre de noyaux.**
- La faiblesse du modèle réside également dans le fait de **ne considérer qu'un seul actif** sur un seul échange. On peut tout à fait avoir plusieurs jumps synchrones entre les différents marchés et il serait intéressant de proposer une dynamique de la corrélation des phénomènes. C'est un peu la promesse du papier de T. Fabre et I. Muni Toke[6] qui mappe l'équation intégrale convolutive de Fredholm à un problème de Deep Learning avec Hamiltonien en vue d'affronter le problème de la dimension et ainsi proposer une vue cross-asset sur le marché des cryptos. Le modèle explicatif est notamment utile pour capturer les cross-excitations des échanges entre

eux via des effets lead-lag, dans la philosophie du market impact de Bacry et Muzy.

Bibliographie

- [1] E. BACRY et J. F MUZY. *Hawkes model for price and trades high-frequency dynamics*. 2013. arXiv : 1301.1135 [q-fin.TR].
- [2] Mehdi LALLOUACHE et Damien CHALLET. *The limits of statistical significance of Hawkes processes fitted to financial data*. 2015. arXiv : 1406.3967 [q-fin.ST].
- [3] Aurélien ALFONSI et Pierre BLANC. *Extension and calibration of a Hawkes-based optimal execution model*. 2015. arXiv : 1506.08740 [q-fin.TR].
- [4] Kyungsub LEE et Byoung Ki SEO. « Marked Hawkes process modeling of price dynamics and volatility estimation ». In : *Journal of Empirical Finance* 40 (2017), p. 174-200. ISSN : 0927-5398. DOI : <https://doi.org/10.1016/j.jempfin.2016.08.004>. URL : <https://www.sciencedirect.com/science/article/pii/S0927539816300810>.
- [5] Emmanuel BACRY, Thibault JAISON et Jean-Francois MUZY. *Estimation of slowly decreasing Hawkes kernels : Application to high frequency order book modelling*. 2014. arXiv : 1412.7096 [q-fin.ST].
- [6] Timothée FABRE et Ioane Muni TOKE. « Neural Hawkes : Non-Parametric Estimation in High Dimension and Causality Analysis in Cryptocurrency Markets ». In : *arXiv preprint arXiv :2401.09361* (2024).