Empirical properties and asset modelling at high frequency

Riadh Zaatour

Numéro d'ordre : 2014ECAP0027

Thèse de doctorat de l'École Centrale Paris

Spécialité :
Mathématiques Appliquées

Laboratoire d'accueil :
Mathématiques Appliquées aux Systèmes

présentée par

Riadh Zaatour

pour l'obtention du grade de
Docteur de l'École Centrale Paris

Propriétés empiriques et modélisation d'actifs en haute fréquence

Empirical properties and asset modelling at high frequency

---

dirigée par Frédéric Abergel

Soutenue publiquement le 10 Mars 2014 devant le jury composé de :

Frédéric Abergel	(École Centrale Paris),	*Directeur de thèse*
Emmanuel Bacry	(Ecole Polytechnique),	*Examinateur*
Fulvio Baldovin	(Università degli Studi di Padova),	*Rapporteur*
Bernard Lapeyre	(École des Ponts ParisTech),	*Rapporteur*
Stéphane Tyč	(McKay-Brothers),	*Examinateur*

To my family
To my wife
To my former teachers

Remerciements

Je tiens d'abord à remercier Frédéric Abergel, sans qui ce manuscrit, et bien d'autres choses, n'auraient pas eu lieu. J'ai eu la chance de le côtoyer dans le monde de l'entreprise ainsi que dans le monde académique, et ai toujours trouvé en lui une personne professionnellement, scientifiquement et humainement très riche et très enrichissante. Ses conseils avisés et sa confiance sans bornes ont permis à ce travail de commencer, progresser et aboutir.

Je remercie aussi chaleureusement Bernard Lapeyre et Fulvio Baldovin qui ont aimablement accepté d'être rapporteurs de cette thèse.

Emmanuel Bacry, dont les travaux sont un fil conducteur de la plus grande partie de ce manuscrit, ainsi que Stéphane Tyč, par ailleurs créateur et chef de l'entreprise pour laquelle je travaille, me font un grand honneur en étant dans mon jury.

Cette thèse a été aussi l'occasion de collaboration avec José Da Fonseca. Au delà, une amitié est née et j'espère qu'elle s'approfondira, et que la collaboration ne s'arrêtera pas après la thèse.

L'équipe de recherche de BNP Paribas m'a aimablement accueilli dans de très bonnes conditions pendant quelques mois, et cet accueil a été décisif pour l'avancement de cette thèse. Je remercie Boris Leblanc, Casimir Bernand Mantel et tous les autres membres de l'quipe pour leur amabilité et leur disponibilité.

Laurent Guerby et Cédric Joulain ont été très patients avec mes questions répétées concernant les serveurs et les données. Je leur suis profondément reconnaissant.

Je remercie aussi les amis de Fiquant, ceux qui ont terminé l'aventure comme ceux qui y sont encore, qui ont été pour moi, par leurs parcours très différents, une source de richesse sur les plans scientifique et humain.

La bonne humeur d'Annie, de Sylvie et de Catherine a réussi à être communicative même durant les moments de doute. Qu'elles en soient ici chaleureusement remerciées.

Je remercie aussi mes anciens collègues à Natixis, qui m'ont introduit à la finance et m'ont permis, en me faisant confiance, de progresser au fil des années passées dans cette banque.

Mes amis à McKay-Brothers m'offrent tous les jours un cadre de travail stimulant et agréable. Qu'ils en soient ici remerciés.

Ce travail est dédié à ma famille, ma femme, et à mes anciens professeurs. J'espère qu'il est digne de ce qu'ils se sont efforcés de me transmettre.

Résumé

Cette thèse explore théoriquement et empiriquement certains aspects de la formation et de l'évolution des prix des actifs financiers observés en haute fréquence. Nous commençons par l'étude de la dynamique jointe de l'option et de son sous-jacent. Les données haute fréquence rendant observable le processus de volatilité réalisée du sous-jacent, nous cherchons à savoir si cette information est utilisée pour fixer les prix des options. Nous trouvons que le marché ne l'exploite pas. Les modèles de volatilité stochastique sont donc à considérer comme des modèles à forme réduite. Cette étude permet néanmoins de tester la pertinence d'une mesure de couverture empirique que nous appelons *delta effectif*. C'est la pente de la régression des rendements des prix de l'option sur ceux du sous-jacent. Elle fournit un indicateur de couverture assez satisfaisant et indépendant de toute modélisation. Pour la dynamique des prix, nous nous tournons dans les chapitres suivants vers des modèles plus explicites de la microstructure du marché. L'une des caractéristiques de l'activité de marché est son regroupement, ou clustering. Les processus de Hawkes, processus ponctuels présentant cette caractéristique, fournissent donc un cadre mathématique adéquat pour l'étude de cette activité. La représentation Markovienne de ces processus, ainsi que leur caractère affine quand le noyau est exponentiel, permettent de recourir aux puissants outils analytiques que sont le générateur infinitésimal et la formule de Dynkin pour calculer différentes quantités qui leur sont reliées, telles que les moments ou autocovariances du nombre d'évènements sur un intervalle donné. Nous commençons par un cadre monodimensionnel, assez simple pour éclairer la démarche, mais suffisamment riche pour permettre des applications telles que le groupement des instants d'arrivée d'ordres de marché, la prévision de l'activité de marché à venir sachant l'activité passée, ou la caractérisation de formes inhabituelles, mais néanmoins observées, de signature plot où la volatilité mesurée décroît quand la fréquence d'échantillonnage augmente. Nos calculs nous permettent aussi de rendre la calibration des processus de Hawkes instantanée en recourant à la méthode des moments. La généralisation au cas multidimensionnel nous permet ensuite de capturer, avec le clustering, le phénomène de retour à la moyenne qui caractérise aussi l'activité de marché observée en haute fréquence. Des formules générales pour le signature plot sont alors obtenues et permettent de relier la forme de celui-ci à l'importance relative du clustering ou du retour à la moyenne. Nos calculs permettent aussi d'obtenir la forme explicite de la volatilité associée à la limite diffusive, connectant la dynamique de niveau microscopique à la volatilité observée macroscopiquement, par exemple à l'échelle journalière. En outre, la modélisation des activités d'achat et de vente par des processus de Hawkes permet de calculer l'impact d'un méta ordre sur le prix de l'actif. On retrouve et on explique alors la forme concave de cet impact ainsi que sa relaxation temporelle. Les résultats analytiques obtenus dans le cas multidimensionnel fournissent ensuite le cadre adéquat à l'étude de la corrélation. On présente alors des résultats généraux sur l'effet Epps, ainsi que sur la formation de la corrélation et du lead lag.

Mots-clés: dynamique jointe, processus de Hawkes, calibration, microstructure, volatilité, signature plot, impact d'un méta ordre, corrélation, effet Epps, lead-lag.

Abstract

This thesis explores theoretical and empirical aspects of price formation and evolution at high frequency. We begin with the study of the joint dynamics of an option and its underlying. The high frequency data making observable the realized volatility process of the underlying, we want to know if this information is used to price options. We find that the market does not process this information to fix option prices. The stochastic volatility models are then to be considered as reduced form models. Nevertheless, this study tests the relevance of an empirical hedging parameter that we call *effective delta*. This is the slope of the regression of option price increments on those of the underlying. It proves to be a satisfactory model-independent hedging parameter. For the price dynamics, we turn our attention in the following chapters to more explicit models of market microstructure. One of the characteristics of the market activity is its clustering. Hawkes processes are point processes with this characteristic, therefore providing an adequate mathematical framework for the study of this activity. Moreover, the Markov property associated to these processes when the kernel is exponential allows to use powerful analytical tools such as the infinitesimal generator and the Dynkin formula to calculate various quantities related to them, such as moments or autocovariances of the number of events on a given interval. We begin with a monovariate framework, simple enough to illustrate the method, but rich enough to enable applications such as the clustering of arrival times of market orders, prediction of future market activity knowing past activity, or characterization of unusual shapes, but nevertheless observed, of signature plot, where the measured volatility decreases when the sampling frequency increases. Our calculations also allow us to make instantaneous calibration of the process by relying on the method of moments. The generalization to the multidimensional case then allow us to capture, besides the clustering, the phenomenon of mean reversion, which also characterizes the market activity observed in high frequency. General formulas for the signature plot are then obtained and used to connect its shape to the relative importance of clustering or mean reversion. Our calculations also allow to obtain the explicit form of the volatility associated with the diffusive limit, therefore connecting the dynamics at microscopic level to the macroscopic volatility, for example on a daily scale. Additionally, modelling buy and sell activity by Hawkes processes allows to calculate the market impact of a meta order on the asset price. We retrieve and explain the usual concave form of this impact as well as its relaxation with time. The analytical results obtained in the multivariate case provide the adequate framework for the study of the correlation. We then present generic results on the Epps effect as well as on the formation of the correlation and the lead lag.

Mots-clés: joint dynamics, Hawkes processes, calibration, microstructure, volatility, signature plot, market impact, correlation, Epps effect, lead-lag.

Contents

Introduction (In French)

Cette thèse explore, tant sur le plan empirique que celui de la modélisation, la formation et l'évolution des prix des actifs financiers observés en haute fréquence. L'observation fine du processus des prix est rendue possible grâce à la disponibilité de bases de données financières de type tick à tick, c'est à dire relatant toute évolution du carnet d'ordre. Cela ouvre des perspectives pour la compréhension du mécanisme de formation des prix, mais arrive aussi avec son lot de nouvelles questions.

Nous exposons dans cette introduction les différentes questions auxquelles nous nous sommes intéressés et résumons notre démarche ainsi que les résultats auxquels nous sommes arrivés, en motivant les différentes directions prises dans nos recherches.

0.1 Dynamique jointe option/action entre théorie et pratique

0.1.1 Étude intraday et delta effectif

L'observation des prix en haute fréquence permet de rendre observables des quantités qui ne l'étaient pas à une résolution plus grossière. Par exemple, le processus de volatilité réalisée pendant la journée de trading est maintenant observable grâce à la disponibilité des données d'une part, et à une multitude d'estimateurs d'autre part.

D'un autre côté, les modèles à volatilité stochastique utilisés dans le calcul des prix d'options semblent donner de bonnes performances aussi bien statiques (génération du smile de volatilité) que dynamiques (persistence du smile, meilleure qualité de couverture ...). Ces modèles sont généralement utilisés comme outils d'extrapolation entre les prix d'options liquides (les options vanilles) et celles exotiques, généralement non cotées. Mais ils présentent aussi un scénario de formation du prix de l'option. Ce prix est, d'après ces modèles, le résultat du traitement par le marché des innovations sur le spot du sous jacent et de sa variance. Cette dernière étant rendue observable grâce aux données haute fréquence, il est alors légitime de demander si les données racontent la même histoire de formation des prix des options.

Dans un modèle où le sous-jacent S_t suit une volatilité stochastique, nous pouvons écrire sous la probabilité historique:

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^1 \tag{1}$$

$$dV_t = \nu dt + \zeta dW_t^2 \tag{2}$$

$$d < W^1, W^2 >_t = \rho dt, \tag{3}$$

l'application de la formule d'Itô au prix de l'option $C_t = C(t, S_t, V_t)$ permet d'établir une relation de régression de celui-ci sur le spot du sous-jacent, isolant le risque 'spot' du risque de pure volatilité :

$$dC_t = \underbrace{\left(\frac{\partial C}{\partial S} + \frac{\zeta\rho}{\sqrt{V_t}S_t}\frac{\partial C}{\partial V}\right)}_{\text{delta spot effectif}} dS_t + \underbrace{\zeta\sqrt{1-\rho^2}\frac{\partial C}{\partial V}}_{\text{delta volatilité}} dZ_t. \tag{4}$$

Cette régression est parfaitement observée sur les données optionnelles comme le montre la figure 1 ci-dessous, où l'on régresse les incréments du prix d'option sur ceux du spot pour des options sur Kospi 200 qui sont les plus liquides au monde. Remarquons que la fréquence d'échantillonnage est importante, car une fréquence très élevée enlève toute corrélation entre accroissements. Ce phénomène est connu sous le nom d'effet de EPPS (Epps, 1979), et se traduit par la décroissance de la corrélation estimée quand la fréquence d'échantillonnage augmente. Il fait partie de ces faits stylisés qui apparaissent quand on étudie les données à haute fréquence. Cet effet disparaît dès que la période d'échantillonnage est de quelques minutes. Cela est à comparer au monde action/action, où l'on doit attendre quelques dizaines de minutes pour le voir disparaître. L'effet *delta* de l'action sur l'option est certainement une explication de la décroissance de ce temps.

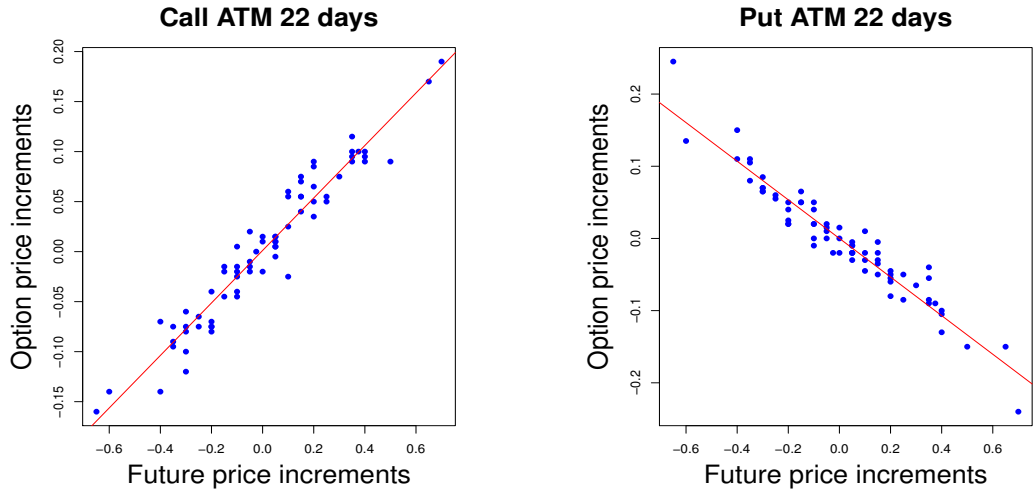

Figure 1: Illustration de la relation de régression des incréments du prix de l'option par rapport aux incréments de son sous jacent. Comme intuitivement attendu, la pente est positive pour le Call et négative pour le Put.

Nous effectuons ces régressions de manière journalière sur notre ensemble de données (options sur Eurostoxx, Kospi 200 et Dax), et trouvons généralement des R^2 avoisinant les 90%. Ainsi, la plus grande part de variance du prix de l'option s'explique par la dynamique du spot. Ensuite, nous calculons la variance réalisée en cours de journée sur des fenêtres de temps glissantes, et par différents estimateurs. Nous essayons de relier le résidu de la première régression à la dynamique de variance ainsi calculée mais les résultats s'avèrent insignifiants, contredisant l'intuition apportée par le modèle à volatilité stochastique.

Cela nous pousse à remettre cette intuition en cause et à nous poser la question de la part de risque que ces modèles projettent sur le spot, et la part de risque qu'ils attribuent à la dynamique de la variance. Pour cela, nous devons calculer les ordres de grandeurs relatifs de ce que nous avons appelé "delta spot effectif" et "delta volatlité" dans l'équation (4). Cette dernière, écrite sous la probabilité historique, dépend de paramètres inobservables tels que la volatilité de la variance, ou la corrélation entre spot et variance. Mais ces paramètres restent inchangés par changement de probabilité. Nous pouvons donc les rendre observables par calibration d'un

modèle spécifié sous la probabilité risque neutre. Nous calibrons donc tous les jours le modèle de Heston aux prix des options liquides observées, et retrouvons les poids relatifs du risque spot et du risque de volatilité. Cela nous permet de calculer un R^2 implicite de ce modèle, et nous arrivons globalement au même résultat observé empiriquement, à savoir que notre modèle de volatilité stochastique attribue la majeure partie de la variance du prix de l'option à la variance du spot du sous-jacent.

De plus, le *delta effectif* calculé empiriquement, défini comme la pente de la régression des rendements des prix de l'option sur ceux du sous-jacent est très proche du delta total calculé par le modèle de Heston. Ainsi, malgré un résultat négatif de la recherche de l'effet de la variance réalisée dans le prix des options, nous obtenons une quantification du risque spot en accord entre l'étude empirique et le modèle à volatilité stochastique. La sensibilité totale au spot, capturée dans le coefficient de régression des incréments de l'option sur ceux du sous-jacent s'avère être un bon paramètre de couverture, d'autant plus robuste qu'il est parfaitement empirique et est indépendant de tout modèle et de toute hypothèse. Sa robustesse est confirmée par une simulation de portefeuille contenant une option couverte en delta sur des données historiques.

En guise d'application, remarquons qu'une pratique répandue en salles des marchés consiste à postuler une dynamique de la volatilité implicite quand le spot du sous-jacent varie. La nappe de volatilité implicite est paramètrée sous la forme $\Sigma = \Sigma(S, m)$ où m est la moneyness $m = \frac{K}{S}$. Le delta de l'option est alors corrigé pour rendre compte de cette dynamique et s'écrit :

$$\frac{dC}{dS} = \Delta_{BS} - \underbrace{\text{Vega}_{\text{BS}} \times \text{Skew} \frac{K}{S^2}}_{\text{correction statique}} + \underbrace{\text{Vega}_{\text{BS}} \times \frac{\partial \Sigma}{\partial S}}_{\text{correction dynamique}} \tag{5}$$

où $\text{Skew} = \frac{\partial \Sigma}{\partial m}$.

La correction statique provient du changement de moneyness de l'option quand le spot change, tandis que la correction dynamique est due au mouvement imposé à la nappe de volatilité par la dynamique du spot.

L'étude faite précédemment fournit alors une manière de tester la pertinence de cette dynamique. À la fin de la journée de trading, la comparaison de ce delta corrigé au delta effectif calculé par régression et qui est complètement indépendant de toute hypothèse de modèlisation peut fournir un outil de test des dynamiques de volatilité implicites utilisées.

0.1.2 Augmentation de l'échelle de temps et dynamique de volatilité implicite

Une critique peut cependant être faite à l'étude effectuée jusqu'à présent : c'est celle de l'échelle de temps. En effet, les variances réalisées sont calculées sur des fenêtres de temps glissantes assez petites, ne dépassant pas la demi-heure. Cela peut donner lieu à beaucoup de bruit. On aimerait donc généraliser l'étude à une échelle de temps plus grande où l'on aurait plus confiance dans nos estimateurs de variance réalisée, en l'estimant par exemple sur une journée entière et en essayant de voir l'impact de ses variations sur l'évolution du prix de l'option.

Cette approche est problématique. En effet, l'option change de jour en jour. Ce n'est plus le même instrument entre deux journées de trading. La prise en compte des effets de passage du temps et de changement de moneyness peut appeler à faire des hypothèses qui viendraient polluer l'étude. L'idée est donc de chercher une quantité indépendante de la maturité et de la moneyness et que l'on traiterait comme la même quantité financière de jour en jour.

L'étude précédente, en établissant l'adéquation du delta effectif obtenu empiriquement avec celui de Heston, ainsi que du R^2 empirique de la régression de l'option sur l'action avec celui implicite dans le modèle de Heston, augmente notre confiance dans l'intuition apportée par la volatilité stochastique. Nous développons alors les prédictions théoriques de ce modèle. Parmi elles, l'espérance de la variance réalisée, qui s'écrit :

$$\mathbb{E}\left[\frac{1}{T}\int_0^T V_t dt\right] = V_\infty + (V_0 - V_\infty)\,\frac{1-e^{-\kappa T}}{\kappa T}, \tag{6}$$

où V_0 désigne la variance instantanée.

Cette équation donne une paramétisation possible de la volatilité à la monnaie, où l'on voit que le niveau de volatilité dépend de la variance instantanée V_0. Différentes études comme (Gatheral, 2006) prédisent que cette variance instantanée est au premier ordre linéaire en fonction du spot, ce que nous arrivons à vérifier empiriquement, en calibrant des nappes de volatilité implicite à la paramétrisation ci-dessus. Grâce à (Durrleman, 2008), la variance instantanée V_0 est interprètée comme étant la volatilité implicite à la monnaie quand la maturité tend vers 0, et désigne donc la même quantité "financière" de jour en jour, ce qui résoud notre problème mentionné plus haut.

La régression linéaire de V_0 sur le spot est très significative et nous permet comme précédemment d'isoler la dépendance au spot de cette quantité. Nous essayons ensuite d'expliquer la variance résiduelle de V_0 entre l'ouverture et la clôture du marché par la variance réalisée sur la journée, mais les résultats sont, cette fois encore, non significatifs.

En conclusion, le marché ne semble pas traiter l'information de volatilité réalisée pour fixer le prix des options. Les modèles à volatilité stochastiques, bien qu'en adéquation sur beaucoup d'aspects avec ce qui est réellement observé dans les données, ne sont pas à prendre comme modèles structurels, c'est à dire donnant un mécanisme de formation des prix, mais plutôt comme des modèles à forme réduite. Pour emprunter au langage de la physique, ces modèles décrivent la cinématique des prix et ne donnent pas de renseignement sur la dynamique, à savoir les forces et interactions qui donnent naissance à ces prix.

Pour étudier cette dynamique, nous nous tournons donc vers une vision plus microscopique, au niveau des évènements de carnet d'ordre.

0.2 Processus de Hawkes et clustering

0.2.1 Étude analytique

L'activité de trading observée au niveau du carnet d'ordre est une suite d'évènements ponctuels irréguliers tels que le placement d'un ordre limite, l'exécution d'un ordre marché ou l'annulation d'un ordre limite. Les processus ponctuels sont donc l'outil de modèlisation idéal à ce degré de granularité.

Malheureusement, le processus ponctuel le plus simple, à savoir le processus de Poisson, est complètement inadéquat pour modèliser ce genre d'évènements. En effet, étant sans mémoire, ce processus ne capture pas le principal fait stylisé de l'activité de trading, à savoir son regroupement dans le temps, ou *clustering*.

Une généralisation consiste à faire dépendre l'intensité du processus ponctuel de ses réalisations passées. Cela donne les processus de Hawkes, introduits dans (Hawkes, 1971). Quand la dépendance au passé est portée par un noyau exponentiel, le couple formé du processus de

saut et de son intensité est un processus de Markov. En régime stationnaire, l'intensité de ce processus s'écrit alors :

$$\lambda_t = \lambda_\infty + \int_0^t \alpha e^{-\beta s} dN_s.$$

Additionnellement, ce processus a une structure affine, ce qui permet de caractériser facilement sa fonction génératrice des moments comme fonction exponentiellement affine de ses variables d'état.

Cette propriété nous permet de calculer différents moments du processus de Hawkes, à savoir par exemple la moyenne ou la variance du nombre d'évènements sur un intervalle donné quand le processus a atteint son régime stationnaire. Ceci est faisable en dérivant la fonction génératrice des moments et en résolvant les équations différentielles qui en découlent. Cette méthode directe et classique pour les processus affines donne lieu à des calculs fastidieux et surtout difficilement généralisables en grande dimension. Nous lui préfèrerons une technique plus simple basée sur la formule de Dynkin.

En effet, pour notre processus de Markov $X_t = (N_t, \lambda_t)$, et pour toute fonction f dans le domaine de son générateur infinitésimal $\mathcal{L}$, nous avons pour $s > t$ la formule de Dynkin:

$$\mathbb{E}_t \left[f\left(X_s\right) \right] = f\left(X_t\right) + \mathbb{E}_t \left[\int_t^s \mathcal{L} f\left(X_u\right) du \right] \tag{7}$$

qui permet de calculer des espérances conditionnelles de fonctions du processus de Markov $X_t = (\lambda_t, N_t)$, en particulier des moments, et qui s'avère très pratique lorsque le terme de droite est facilement calculable.

C'est par application de cette formule à N_t, N_t^2, N_t^3, N_t^4 que nous calculerons la moyenne, variance, skew et kurtosis du nombre de sauts d'un processus de Hawkes sur un intervalle de temps de longueur donnée, quand le processus a atteint son régime stationnaire.

La formule la plus importante est certainement l'autocovariance du nombre de sauts sur des intervalles consécutifs, à savoir $\mathbb{E}_t \left[\left(N_{t_1} - N_t\right) \left(N_{t_3} - N_{t_2}\right) \right]$, où $t < t_1 < t_2 < t_3$, que nous obtenons par conditionnements successifs, et qui tire son importance du fait que les processus de Hawkes sont utilisés justement parce qu'ils permettent ce genre d'autocorrélation.

Centrée et réduite grâce aux moments calculés précédemment, nous trouvons alors la fonction d'autocorrélation du nombre de sauts sur des intervalles consécutifs de longeur τ :

Proposition. *Pour un processus de Hawkes $X_t = (\lambda_t, N_t)$ suivant la dynamique présentée plus haut, la fonction d'autocorrélation du nombre de sauts ayant lieu sur des intervalles de temps de longueur τ séparés par un lag δ s'écrit :*

$$Acf(\tau, \delta) = \frac{e^{-2\beta\tau} \left(e^{\alpha\tau} - e^{\beta\tau}\right)^2 \alpha(\alpha - 2\beta)}{2\left(\alpha(\alpha - 2\beta)\left(e^{(\alpha-\beta)\tau} - 1\right) + \beta^2\tau(\alpha - \beta)\right)} e^{(\alpha-\beta)\delta}. \tag{8}$$

Cette autocorrélation est donc positive et décroit exponentiellement avec le lag. La rapidité de sa décroissance dépend de la différence $\alpha - \beta$, ou de manière équivalente du rapport $\frac{\alpha}{\beta}$ qui est la norme L_1 du noyau, et qui doit être strictement infèrieur à 1 pour garantir la stabilité du processus. Plus cette norme est élevée, plus la mémoire du processus est importante comme on s'y attend intuitivement.

0.2.2 Applications

Empiriquement, on observe le phénomène de clustering dans l'activité de trading. Le nombre de trades ayant lieu sur des intervalles de temps consécutifs (qui ne se chevauchent pas) est

positivement autocorrélé et cette autocorrélation décroit avec le lag. Cette propriété n'aurait pas été observée si l'activité de trading obéissait à un processus de Poissson, et ces observations empiriques suggèrent donc le processus de Hawkes comme candidat adéquat à la modélisation de l'activité de trading au niveau des temps d'arivée des ordres.

Classiquement, les processus de Hawkes sont calibrés par maximisation de vraisemblance. Cette méthode peut être très lente car elle suppose de boucler plusieurs fois sur les évènements à chaque calcul de la fonction objectif. Les moments calculés jusqu'à présent suggèrent d'utiliser la méthode des moments ou la méthode généralisée des moments pour l'estimation des paramètres du processus de Hawkes. Cette méthode a l'avantage d'être instantanée, et indépendante de la taille de l'échantillon, vu que les évènements sont parcourus une seule fois pour le calcul des moments empiriques.

Dans un premier temps, nous validons donc la méthode des moments comme bon outil d'estimation des paramètres du processus de Hawkes en recourant à une expérience Monte Carlo, où nous simulons un processus de Hawkes, puis essayons de retrouver ses paramètres par Maximum de vraisemblance et par méthode des moments, en partant du même ensemble de paramètres initiaux pour les deux (évidemments différents des paramètres objectifs). Cela valide la méthode et surtout le choix des moments sur lesquels on calibre, et l'on trouve qu'il est important de calibrer sur la fonction d'autocorrélation pour obtenir de bons résultats.

Une fois cette méthode validée, nous l'utilisons pour calibrer le processus de Hawkes sur des données réelles de temps d'arrivée de trades. Nous arrivons à reproduire la forme de l'autocorrélation du nombre de trades observée comme le montre la figure 2. En calibrant tous les jours, sur un ensemble de données couvrant deux ans, nous remarquons que les paramètres sont assez stables, et notons aussi que le paramètre λ_∞ est de l'ordre du dixième du paramètre λ dans le cadre Poisson ce qui prouve que le processus de Hawkes explique une large part des évènements par sa structure de branchement.[1]

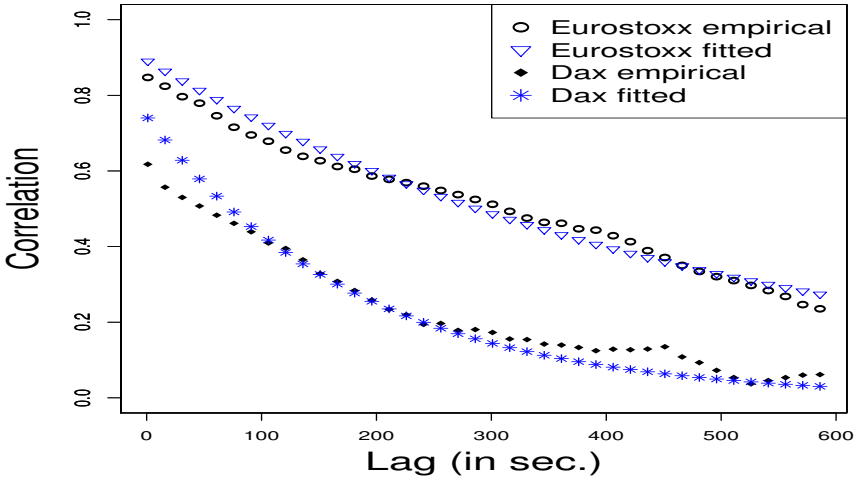

Figure 2: Autocorrélation empirique du nombre de trades ayant lieu sur des intervalles de temps de longueur $\tau = 60$ secondes, comparée à l'autocorrélation générée par un processus de Hawkes calibré sur la journée de trading.

[1] Le λ Poisson est plutôt proche de l'intensité moyenne induite par le processus de Hawkes, à savoir $\frac{\lambda_\infty \beta}{\beta - \alpha}$, mais dans notre comparaison nous voulons mettre en valeur la faible contribution du bruit de fond du processus de Hawkes, porté par λ_∞, au total des évènements, ceux-ci étant en majorité générés grâce au branchement.

Pour mieux quantifier cette structure de branchement, nous calculons une réponse impulsionnelle qui la caractérise, que nous définissons comme l'espérence du nombre d'évènements déclenchés par un évènement qui a lieu en t=0 si le processus est indéfiniment observé. Cette réponse impulsionnelle s'écrit

$$N_{response} = \frac{\alpha}{\beta - \alpha}. \tag{9}$$

Estimée sur des données réelles, nous remarquons qu'elle caractérise bien la liquidité : plus les actifs sont liquides, plus cette quantité est grande.

Ensuite, après calibration du modèle, il est naturel de se poser la question sur sa capacité de prévision. On veut prévoir l'activité de trading à un horizon de temps donné sachant l'activité ayant eu lieu dans le passé récent. Nous caractérisons cette activité par le nombre de trades par intervalles de longueur τ en secondes, divisé par τ. Cela donne donc une quantité y_t homogène à une intensité. L'autocorrélation du nombre de trades sur des intervalles de temps consécutifs étant positive et décroissant exponentiellement, il est alors naturel de modèliser l'activité de trading par un processus AR(p). Pour clarifier, l'arrivée des trades (individuellement) étant un processus de Hawkes, on dit que le nombre de trades par intervalle est un AR(p):

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + u_t.$$

Quand un processus de Hawkes est calibré sur les temps d'arrivée des trades, les coefficients $(a_1, \ldots, a_p)^\top$ sont solutions du système linéaire:

$$\begin{pmatrix} 1 & \text{Acf}[\tau,0] & ... & ... & \text{Acf}[\tau,(p-2)\tau] \\ \text{Acf}[\tau,0] & 1 & \text{Acf}[\tau,0] & & \text{Acf}[\tau,(p-3)\tau] \\ ... & & & & \\ ... & & & & \\ \text{Acf}[\tau,(p-2)\tau] & & & \text{Acf}[\tau,0] & 1 \end{pmatrix} \times \begin{pmatrix} a_1 \\ ... \\ ... \\ ... \\ a_p \end{pmatrix} = \begin{pmatrix} \text{Acf}[\tau,0] \\ ... \\ ... \\ ... \\ \text{Acf}[\tau,(p-1)\tau] \end{pmatrix}. \tag{10}$$

La résolution de cette équation linéaire pour un p suffisemment grand permet de calculer les coefficients a_i et de décider, selon la valeur relative de ces coefficients à quel ordre s'arrêter pour avoir une bonne précision.

Des tests de prévision "in sample" effectués sur des données réelles nous permettent alors de montrer la supèriorité de ce procédé de prévision par rapport à une méthode plus simple consistant à supposer que le processus générateurs des trades est un Poisson, et donc que la meilleure prévision du nombre de trades sur l'intervalle de temps qui vient est la moyenne de ce processus de Poisson. Notons toutefois que si l'horizon de prévision est assez grand (de l'ordre de la demi-heure), les performances de ces deux méthodes de prévision coïncident, ce qui était attendu vu que l'autocorrélation du nombre de trades s'éteint à cet horizon de temps.

La partie suivante est fortement influencée par l'article fondateur (Bacry et al., 2013a). Nous cherchons à y connecter l'aspect microscopique de l'activité de trading, qui se manifeste dans l'aspect discret du mouvement des prix, à son aspect macroscopique, où le prix ressemble à une diffusion continue. Inspirés par (Bacry et al., 2013a), nous modèlisons les sauts positifs et négatifs du processus de prix par deux processus de Hawkes indépendants:

$$S_t = S_0 + \left(N_t^{up} - N_t^{down}\right) \frac{\delta}{2} \tag{11}$$

où δ est la valeur du tick (le facteur $\frac{1}{2}$ s'expliquant par le fait qu'un saut positif du bid ou de l'ask d'un tick, s'accomagne d'un saut positif du mid-price d'un demi tick).

Les processus N_t^{up} et N_t^{down} sont des processus de Hawkes capturant les sauts positifs et négatifs du mid-price. Nous les considérons indépendants et ayant les mêmes paramètre pour éviter une explosion du processus de prix. Leurs intensités respectives s'écrivent donc :

$$\lambda_t^{up} = \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s^{up} \tag{12}$$

$$\lambda_t^{down} = \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s^{down}. \tag{13}$$

Ainsi, nous ne considérons qu'un aspect clustering, contrairement à l'article fondateur (Bacry et al., 2013a) où seulement un aspect d'excitation mutuelle est considéré.

L'objectif est ici de connecter les paramètres gouvernant l'activité observée en haute fréquence, aux caractéristiques macroscopiques du processus de prix, comme sa volatilité. Ce modèle nous permet par exemple d'expliquer le phénomène de *signature plot* observé sur certains actifs financiers. Le *signature plot* est la forme que prend la fonction reliant la volatilité réalisée estimée par sommation des carrés des rendements, à la fréquence d'échantillonnage de ces derniers. Rappelons que si les rendements résultaient d'une diffusion Brownienne, alors la volatilité estimée serait indépendante de la fréquence d'échantillonnage et le signature plot serait plat.

Dans la pratique, on observe une forme biaisée. Généralement, l'estimation de la volatilité réalisée augmente avec la fréquence d'échantillonnage. Ceci est dû à l'autocorrélation négative des rendements à petite échelle de temps, et a été expliqué dans (Bacry et al., 2013a) en introduisant un modèle de microstructure basé sur des processus de Hawkes s'excitant mutuellement. Il arrive cependant que l'on observe une forme de signature plot inversée, où la volatilité estimée diminue quand on augmente la fréquence d'échantillonnage. Ceci est dû à une autocorrélation positive des rendements à petite échelle, et est observé pour certains actifs peu liquides ou même certains jours pour des actions liquides. Notre modèle capte bien ce phénomène. Nous calculons en effet le signature plot induit par notre modèle qui s'avère être positif et croissant avec la période d'échantillonnage.

Nous obtenons par ailleurs la convergence de ce modèle vers une diffusion Brownienne. Dans un autre article fondateur de (Bacry et al., 2013b), les auteurs obtiennent des résultats de convergence plus généraux, notemment avec des noyaux non exponentiels du processus de Hawkes, en recourant à des méthodes martingales. Nous donnons quand même l'idée d'une preuve plus simple basée sur une variante du théorème de Donsker pour les séries aléatoires présentant une dépendance assez faible, et qui s'applique bien dans le cas de processus de Hawkes avec un noyau exponentiel.

Par ailleurs, nous obtenons grâce à nos calculs faits précédemment l'expression de la volatilité asymptotique :

$$\sigma^2 = \frac{\delta^2}{2} \frac{\lambda_\infty}{\left(1 - \frac{\alpha}{\beta}\right)^3}, \tag{14}$$

où l'on voit clairement que plus l'effet clustering est important (i.e $\frac{\alpha}{\beta}$ est proche de 1), plus la volatilité asymptotique est grande, comme on peut s'y attendre intuitivement.

Nous terminons par une comparaison des prévisions de notre modèle et de celles du modèle de (Bacry et al., 2013a). Nous calibrons dans les deux cas le modèle de microstructure aux données empiriques, et calculons la volatilité asymptotique induite par ce modèle et la comparons à la volatilité réalisée observée.

Les résultats montrent que notre modèle surestime systématiquement cette volatilité, alors que celui de (Bacry et al., 2013a) la sous estime systématiquement. En effet, en ne modèlisant que le clustering, notre modèle exagère les mouvements haussiers ou baissiers de l'actif financier, alors que (Bacry et al., 2013a) qui modèlise seulement le retour à la moyenne, les modère. Cela indique clairement que la vérité est certainement entre les deux et fait le sujet du prochain chapitre.

Table 1: Volatilités Asymptotiques

Symbol	λ_∞	α	β	σ_{BDHM}	Empirical σ	$\sigma_{\text{Clustring}}$
Eurostoxx	0.0184	0.0160	0.0219	12.79%	40.77%	80.64%
Dax	0.0429	0.0226	0.0259	4.80%	25.75%	71.36%
Bund	0.0267	0.0180	0.0261	2.70%	8.25%	14.47%
Bobl	0.0228	0.0187	0.0288	2.45%	6.92%	11.19%

Note. Exemple des différences entre volatilité asymptotique induite par un modèle à la (Bacry et al., 2013a), et une volatilité asymptotique induite par notre modèle qui ne capte que le clustering contre la volatilité asymptotique mesurée. La vérité se trouve clairement entre les deux.

0.3 Clustering et retour à la moyenne

0.3.1 Signature plot, corrélation et volatilité asymptotique

L'étude faite précédemment nous pousse donc à considèrer le cadre général. En effet, outre le clustering précédemment modèlisé, et qui se révèle dans l'autocorrélation des occurrences d'évènements du même genre, telle que les sauts positifs ou négatifs du prix, ou l'arrivé d'ordres d'achat ou de vente, on assiste à un phénomène de retour à la moyenne, qui se révèle dans l'autocorrélation d'évènements contradictoires (achat/vente, hausse/baisse). En empruntant le langage de (Bouchaud et al., 2004), on voit que là où le clustering mène à une super-diffusion du prix, le retour à la moyenne mène à une sous-diffusion, et c'est la subtile interaction entre ces deux phénomènes qui donne le caractère diffusif des prix.

Pour capturer ces deux phénomènes, nous considérons dans ce chapitre des processus de Hawkes multidimensionnels, présentant de l'auto-excitation ainsi que de l'excitation mutuelle entre leurs composantes. La formule de Dynkin appliquée précédemment pour le calcul des moments s'applique ici aussi et donne des formules compactes pour les différents moments du processus.

Cette structure riche du processus nous permet de considérer un modèle unifiant le modèle de (Bacry et al., 2013a) et notre modèle simple du précédent chapitre, où les intensités s'écrivent maintenant :

$$\lambda_t^u = \lambda_\infty + \int_0^t \alpha_s e^{-\bar{\beta}(t-v)} dN_v^u + \int_0^t \alpha_m e^{-\bar{\beta}(t-v)} dN_v^d \tag{15}$$

$$\lambda_t^d = \lambda_\infty + \int_0^t \alpha_m e^{-\bar{\beta}(t-v)} dN_v^u + \int_0^t \alpha_s e^{-\bar{\beta}(t-v)} dN_v^d. \tag{16}$$

Ce modèle englobe celui présenté dans le précédent chapitre, et englobe aussi le modèle de retour à la moyenne présenté dans l'article fondateur (Bacry et al., 2013a). Ainsi, la calibration de ce modèle complet, et la calibration de ses deux cas particuliers nous permettent d'effectuer des "likelihood ratio tests". En calibrant sur les données que nous avons sur une large palette d'actifs, nous arrivons à la conclusion qu'il est généralement impossible de se contenter de la composante clustering seule, ou de la composante retour à la moyenne seule pour modèliser les données, chose que nous avons pressentie précédemment.

Dans ce modèle plus riche, on aboutit à la formule générale du signature plot :

Proposition. *Le signature plot* $C(\tau) = \mathbb{E}[\hat{C}(\tau)]$, *où* $\hat{C}(\tau) = \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(S_{(n+1)\tau} - S_{n\tau}\right)^2$, *s'écrit:*

$$C(\tau) = \frac{\nu^2}{2}\Lambda\left(\kappa^2 + \left(1-\kappa^2\right)\frac{\left(1-e^{-\tau\gamma}\right)}{\gamma\tau}\right) \tag{17}$$

avec

$$\Lambda = \frac{\bar{\beta}\lambda_\infty}{\bar{\beta} - \alpha_s - \alpha_m}\ ,\ \kappa = \frac{\bar{\beta}}{\bar{\beta} + \alpha_m - \alpha_s}\ \textit{and}\ \gamma = \bar{\beta} + \alpha_m - \alpha_s.$$

On voit alors facilement que la forme du signature plot dépend de la différence de l'effet Clustering par rapport à celui de retour à la moyenne. Si le retour à la moyenne est plus fort, alors la volatilité estimée augmente avec la fréquence d'échantillonnage, puisque ce fort retour à la moyenne induit une autocorrélation négative des rendements à petite échelle de temps, ce qui augmente la volatilité estimée qui capte ces oscillations. Si, par contre, le clustering l'emporte, la volatilité estimée décroitra avec la période τ d'échantillonnage, cela étant dû à l'autocorrélation positive des rendements dans ce cas.

En effet, l'autocorrélation des rendements s'écrit dans ce modèle :

$$\text{CorrStock}(\tau,\delta) = -\frac{e^{-(\delta+2\tau)(\bar{\beta}+\alpha_m-\alpha_s)}\left(e^{\tau\left(\bar{\beta}+\alpha_m-\alpha_s\right)}-1\right)^2\left(\alpha_m-\alpha_s\right)\left(2\bar{\beta}+\alpha_m-\alpha_s\right)}{2\bar{\beta}^2\left(\bar{\beta}+\alpha_m-\alpha_s\right)}. \tag{18}$$

D'un autre côté, le même raisonnement que dans le chapitre précédent permet d'obtenir la limite diffusive du modèle, et nous aboutissons à une volatilité limite qui s'écrit :

Proposition. *Pour le modèle général décrit ci-dessus, la volatilité asymptotique s'écrit :*

$$\sigma^2 = \frac{\nu^2}{2}\frac{\bar{\beta}^3\lambda_\infty}{(\bar{\beta}-\alpha_s-\alpha_m)(\bar{\beta}+\alpha_m-\alpha_s)^2} \tag{19}$$

La même expérince effectuée précédemment, à savoir la comparaison de la volatilité limite induite par ce modèle de microstructure contre la volatilité réalisée mesurée empiriquement sur les données, permet de confirmer la nécessité de prendre en compte simultanément le clustering et le retour à la moyenne comme le prédisaient nos "likelihood ratio tests".

Table 2: Volatilités Asymptotiques

Symbol	λ_∞	α_s	α_m	β	Empirical σ	Toy model σ
Eurostoxx	0.0119	0.0474	0.0401	0.1074	24.64%	19.08%
Dax	0.0581	0.0341	0.0491	0.1050	25.75%	23.45%
Bund	0.0136	0.0471	0.0410	0.1081	5.35%	4.62%
Bobl	0.0125	0.0465	0.0396	0.1062	4.45%	4.44%

Note. Exemple de volatilités asymptotiques induites par le modèle général contre la volatilité réalisée observée, à comparer aux volatilités asymptotiques induites par la modélisation du retour à la moyenne seul ou du clustering seul.

0.3.2 Market impact dans un modèle simple

Nous terminons ce chapitre par une autre application possible des processus de Hawkes. Dans le chapitre précédent, nous nous sommes intéressés au clustering des temps des arrivées des ordres de marché. Ce clustering est aussi observé si nous signons ces ordres, à savoir si nous les différencions en ordre d'achat et ordre de vente.

Même si les ordres ne sont pas signés dans les fichiers de données que nous avons, nous recourons à un algorithme de signature simple, documenté dans (Lee and Ready, 1991). Cela nous permet d'observer de l'autocorrélation des ordres de même signe (achat ou vente), ainsi que la corrélation d'ordres de signes différents (les achats provoquant des ventes et inversement). Les processus de Hawkes sont donc de bons candidats pour la modélisation de ces dynamiques.

En posant N_t^{Buy} and N_t^{Sell} le nombre cumulé d'ordres d'achat et de vente respectivement, nous écrivons leurs intensités respectives comme suit :

$$\lambda_t^{Buy} = \lambda_\infty + \int_0^t \alpha_s e^{-\bar{\beta}(t-u)} dN_u^{Buy} + \int_0^t \alpha_m e^{-\bar{\beta}(t-u)} dN_u^{Sell}$$
$$\lambda_t^{Sell} = \lambda_\infty + \int_0^t \alpha_m e^{-\bar{\beta}(t-u)} dN_u^{Buy} + \int_0^t \alpha_s e^{-\bar{\beta}(t-u)} dN_u^{Sell}.$$

Suivant (Hewlett, 2006), en considérant comme dans (Bouchaud et al., 2004), que le prix en t résulte de la superposition des impacts de tous les trades qui ont eu lieu jusqu'à t, nous arrivons à mesurer l'impact vu en $t > 0$ d'un trade qui a lieu en $t = 0$, dans un modèle simple d'arrivée d'ordre cité plus haut, où tous les trades sont considérés portant sur des quantités unitaires et où le prix se forme grâce au déséquilibre entre achat et vente dans le modèle simple de Kyle (on notera λ_{K} le facteur d'impact de ce modèle défini dans (Kyle, 1985)). Cet impact s'écrit :

$$I(t) = \lambda_{\text{K}} \left(1 + \frac{(\alpha_s - \alpha_m) e^{-\bar{\beta} t}}{\bar{\beta} + \alpha_m - \alpha_s} \right), \tag{20}$$

où l'on voit qu'il se décompose en impact permanent et impact transitoire.

L'impact d'un méta-ordre résultant de l'exécution d'un gros ordre en le décomposant donne alors la figure suivante familière dans la littérature sur le market impact :

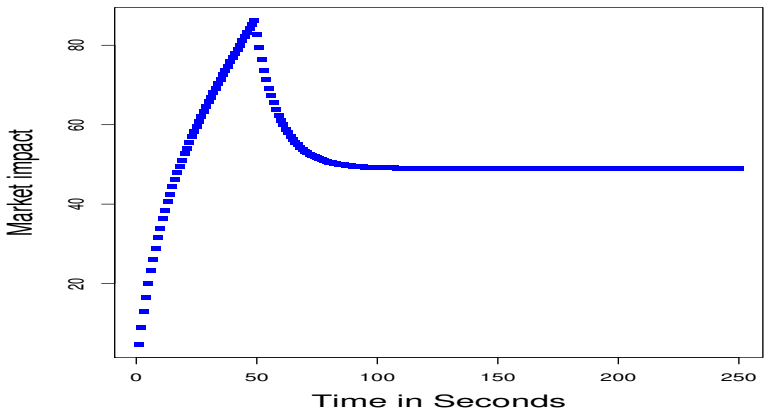

Figure 3: Impact d'un méta ordre d'achat de 50 actions, exécuté durant 50 secondes, à la vitesse d'un ordre par seconde. Nous prenons le paramètre d'échelle $\lambda_{\text{K}} = 1$.

Notons que l'obtention de la partie concave de la figure suppose que $\alpha_s > \alpha_m$, ce qui est cohérent avec la décomposition d'un gros ordre, générant un regroupement des exécutions, et cohérent aussi avec les résultats numériques des calibrations à l'activité d'achat et de vente où l'on trouve que l'effet clustering domine l'effet retour à la moyenne.

0.4 Corrélation et lead-lag

0.4.1 Modèle

Les résultats développés précédemment pour des processus de Hawkes multidimensionels fournissent un cadre adéquat pour l'étude des phénomènes de corrélation entre deux actifs. Pour ce faire, nous modèlisons les sauts positifs et négatifs de deux actifs par des processus de Hawkes

$$S_t^1 = S_0^1 + \left(N_t^{1,u} - N_t^{1,d}\right)\frac{\nu_1}{2},$$
$$S_t^2 = S_0^2 + \left(N_t^{2,u} - N_t^{2,d}\right)\frac{\nu_2}{2}$$

où ν_1 et ν_2 désignent la valeur du tick pour chacun des actifs respectivement. Le processus de Hawkes à 4 dimensions $N_t = (N_t^{1,u}, N_t^{1,d}, N_t^{2,u}, N_t^{2,d})^\top$ a pour vecteur intensité $\lambda_t = (\lambda_t^{1,u}, \lambda_t^{1,d}, \lambda_t^{2,u}, \lambda_t^{2,d})^\top$ qui suit la dynamique:

$$d\lambda_t = \beta(\lambda_\infty - \lambda_t)dt + \alpha dN_t$$

avec:

$$\alpha = \begin{pmatrix} \alpha_s^1 & \alpha_m^1 & x & 0 \\ \alpha_m^1 & \alpha_s^1 & 0 & x \\ y & 0 & \alpha_s^2 & \alpha_m^2 \\ 0 & y & \alpha_m^2 & \alpha_s^2 \end{pmatrix}; \beta = \begin{pmatrix} \bar{\beta}_1 & 0 & 0 & 0 \\ 0 & \bar{\beta}_1 & 0 & 0 \\ 0 & 0 & \bar{\beta}_2 & 0 \\ 0 & 0 & 0 & \bar{\beta}_2 \end{pmatrix} \tag{21}$$

et $\lambda_\infty = (\lambda_{1,\infty}, \lambda_{1,\infty}, \lambda_{2,\infty}, \lambda_{2,\infty})^\top \in \mathbb{R}_+^4$.

Nous pouvons ainsi contrôler la connexion des deux actifs à travers les paramètres de couplage dans les sous matrices carrées 2×2 en haut à droite et en bas à gauche de la matrice α. Dans la paramétrisation ci-haut, nous avons supposé une corrélation positive. C'est ainsi que les sauts positifs du deuxième actif provoquent des sauts positifs pour le premier à travers le paramètre x, et que les sauts positifs du premier actif provoquent des sauts positifs du deuxième via y.

Dans le cadre de ce modèle, nous reconstruisons l'effet de Epps évoqué plus haut, en calculant en toute généralité la formule donnant la corrélation estimée en fonction des paramètres du modèle et de la période d'échantillonnage.

Plus spécifiquement, nous remarquons que le modèle fabrique par construction une corrélation avec un retard temporel, qui peut être mise à jour si l'évolution des actifs est observée à l'échelle de temps adéquate, à savoir en haute fréquence.

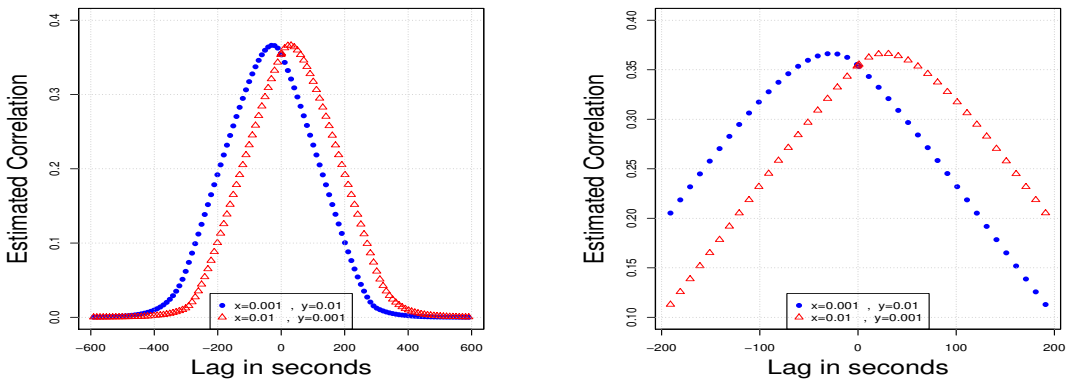

Figure 4: Illustration du lead lag dans la formation de la corrélation en haute fréquence. La figure de droite est un zoom autour de 0 de la figure de gauche.

Ce temps de retard nécessaire à l'obtention de la corrélation maximale constitue alors un candidat à la quantification du lead-lag entre deux actifs qui présentent cet aspect.

Les théorèmes limites de (Bacry et al., 2013b) permettent en outre d'obtenir la matrice de variances/covariances limite du modèle en fonction de ses paramètres:

Proposition. *En posant la matrice de variance covariance asymptotique du modèle :*

$$\sigma\sigma^\top = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix}$$

on a:

$$\sigma_{11}^2 = \frac{\nu_1^2}{4} \frac{2\lambda_{1,\infty}(\bar{\beta}_1\bar{\beta}_2^2x^2y + \bar{\beta}_1^3\theta_2\gamma_2^2) + \lambda_{2,\infty}\bar{\beta}_2x(\bar{\beta}_2^2x\theta_1 + \bar{\beta}_1^2\gamma_2^2)}{(\gamma_1\gamma_2 - xy)^2(\theta_1\theta_2 - xy)},$$
$$\sigma_{12} = \frac{\nu_1\nu_2}{4} \frac{2\bar{\beta}_1y\lambda_{1,\infty}(\bar{\beta}_2^2x\gamma_1 + \bar{\beta}_1^2\theta_2\gamma_2) + 2\bar{\beta}_2x\lambda_{2,\infty}(\bar{\beta}_1^2y\gamma_2 + \bar{\beta}_2^2\theta_1\gamma_1)}{(\gamma_1\gamma_2 - xy)^2(\theta_1\theta_2 - xy)},$$
$$\sigma_{22}^2 = \frac{\nu_2^2}{4} \frac{2\lambda_{2,\infty}(\bar{\beta}_2\bar{\beta}_1^2y^2x + \bar{\beta}_2^3\theta_1\gamma_1^2) + \lambda_{1,\infty}\bar{\beta}_1y(\bar{\beta}_1^2y\theta_2 + \bar{\beta}_2^2\gamma_1^2)}{(\gamma_1\gamma_2 - xy)^2(\theta_1\theta_2 - xy)}$$

avec $\gamma_i = \bar{\beta}_i + \alpha_m^i - \alpha_s^i$ *et* $\theta_i = \bar{\beta}_i - \alpha_m^i - \alpha_s^i \qquad i \in \{1, 2\}.$

où l'on retrouve pour les variances les formules du chapitre précédent si le couplage est annulé.

0.4.2 Étude empirique

La calibration du modèle sur des données réelles permet de rendre compte de sa capacité à retrouver certains faits stylisés et intuitions, à savoir par exemple qu'un indice est généralement leader par rapport aux actions comme le montre la figure 5. En outre, l'utilisation de la quantification du lead lag présentée plus haut permet de mieux interpréter cette relation en lui attribuant une valeur numérique.

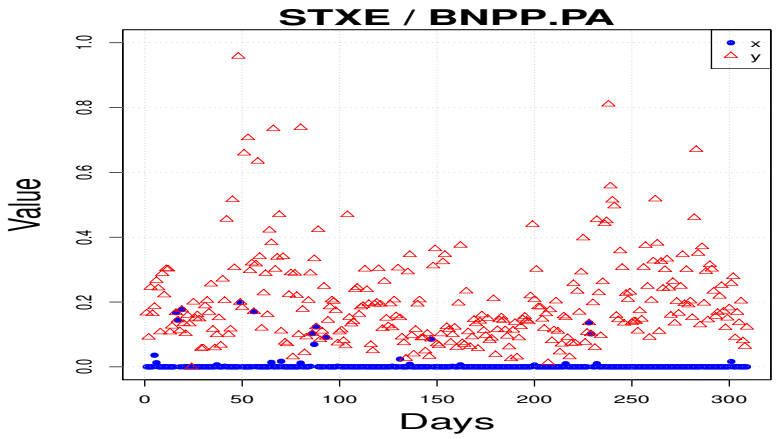

Figure 5: Séries temporelles des paramètres x et y pour la paire constituée de l'indice Eurostoxx et de l'action BNP Paribas, où l'on voit clairement que l'indice influence grandement l'action, tandis que l'impact réciproque est faible.

Chapter 1

What Drives Option Prices?

Note : A part of this chapter is published in The Journal Of Trading, Volume 7, number 3, pp. 12-28, 2012, DOI: 10.3905/jot.2012.7.3.012.

Abstract

We rely on high frequency data to explore the joint dynamics of underlying and option markets. In particular, high frequency data make observable the realized variance process of the underlying, so its effects on option price dynamics are tested. Empirical results are confronted with the predictions of stochastic volatility models. The study reveals that while the modelling of stochastic volatility gives more robust models, the market does not process information on the realized variance to update option prices.

Contents

Introduction

A huge effort has been made to adapt the Black and Scholes model (Black and Scholes, 1973) in order to explain the observed implied volatility smile phenomenon. Alternative pricing models have been proposed that correct some biases of Black and Scholes model and relax its restrictive assumptions, leading for instance to heavy tails and skewness in log returns. Thanks to nice model properties (such as affinity, as, for example, in Heston Model (Heston, 1993)), and efficient numerical methods (for instance the Fast Fourier Transform algorithm (Carr and Madan, 1999)), or closed formula for local volatility model (Dupire, 1994), many models can be perfectly calibrated to a smooth implied volatility surface. Therefore, they can equally well capture the marginal distributions of the spot, while predicting different dynamical behavior for the underlying as well as for the smile itself.

Many empirical studies have established the need for multifactor models to capture the smile dynamics. Stochastic volatility models or jump diffusion models are from this perspective judged to be superior to the local volatility one *a la* Dupire. One of the major drawbacks of this model, is that it predicts a flattening of the forward smile whereas the formers generate a persisting smile, as observed in option markets.

Moreover, among the features that one can add to the Black and Scholes framework (jumps, stochastic volatility, stochastic interest rates ..), stochastic volatility modelling yields the best dynamical model performance as was established in (Cao et al., 1997).

Consequently, stochastic volatility modelling is the corner stone of option pricing theory. This modifies the replication paradigm, as, under stochastic volatility models, perfect replication by dynamic trading in the stock is no longer possible. Another option has to be introduced to complete the market and the replication now relies on dynamic trading in both the stock and this option, to the extent that researchers now speak of a *volatility delta*. A nice exposition of these ideas can be found in (Wilmott, 2006) and (Rebonato, 2004).

As usual when one deals with option pricing, these stochastic volatility models take an "implied route" : calibrate model parameters in order to replicate the observed market vanilla option prices, then use the model with these calibrated parameters to price more exotic options.

What interests us in this chapter is that a stochastic volatility model, besides being an extrapolation tool between liquid vanilla prices and non quoted exotic option prices, presents also a scenario of option price formation and evolution. According to stochastic volatility models, option prices are the result of the processing, by the market, of the information on spot and variance stochastic processes. We want to know if the data tells the same story.

Fortunately, the availability of high frequency data, and especially of the hardware and software capabilities allowing practitioners and researchers to deal with them, enhanced the accuracy of realized variance estimation. One is now able to build very accurate time series of daily realized variance and even estimate it intra daily, making it a market observable if not a tradable asset. Econometric literature dealing with this subject is huge. See for example (Andersen and Benzoni, 2009) and references therein.

A legitimate question is then to reconsider the scenario of option price formation in the light of this newly available data. Does the scenario presented by stochastic volatility models actually occur in real markets? As the spot process is revealed, option prices adjust to be consistent with it, but does the same occur with the variance process that high frequency data allow to reveal? In other words, are option prices adjusted to be consistent with the realized variance process?

We attempt here to answer these questions. We explore the relationship between realized variance time series and option prices, or equivalently, implied volatility. The followed strategy is to explore the joint dynamics of option and underlying markets. The quadratic variation process of the underlying being observable, we seek, ex-post, its effects on option prices.

The chapter is organised as a back and forth between theoretical predictions of stochastic volatility models and data observation.

In section 1, we describe the TRTH (Thomson Reuters Tick History) data we use, as well as its processing in BNP Paribas. We also present the three stock indices that the study deals with, and some characteristics of their respective markets.

Section 2 presents the stochastic volatility framework and its implications on the dynamics of option prices.

In section 3, we conduct an empirical study of the intraday evolution of option prices. Inspired by the framework presented in the previous section, we isolate the spot component in option price dynamics and try to explain the residual by the realized variance. We conclude that the changing intraday underlying variance does not impact the market view of implied volatility, which can be almost fully explained by the spot dynamics.

In section 4, we compare empirical observations to stochastic volatility predictions, quantified thanks to the calibration of Heston stochastic volatility model to market data. This reveals a comparable weight given to the spot component in the option dynamics by the model and data, but a divergence in the projection of the residual : while the model attributes it to stochastic variance, the data does not confirm that. Moreover, the intraday study gives a measure of option price dependence on the spot, which we call the effective delta. This turns out to be a good hedging parameter, well approximated by the Heston total delta, confirming our method of assessing the spot part in the option price dynamics, and confirming other studies' results on the hedging performance of stochastic volatility models, such as (Cao et al., 1997).

In section 5, armed with the confidence gained through the two last sections that a stochastic volatility model describes well the cinematics we are observing, even if it does not identify the same causes, we generalize the study to a daily time scale. Even if an option changes from day to day (as its maturity is diminished by one day), stochastic volatility models give us some more tools to identify a maturity and strike independent quantity characterizing the option price. We study the influence of daily realized variance on the evolution of that quantity. The result remains unchanged: while the stochastic volatility framework gives a satisfactory phenomenological description of option markets, its structural interpretation is not observable in the data. The market does not process information revealed by realized spot variance to fix option prices. Observation of the realized variance does not improve our predictive power on option price or implied volatility dynamics, as compared to the observation of the spot process alone.

Section 6 summarizes the study.

1.1 The data

We rely on tick by tick data from TRTH (Thomson Reuters Tick History). The data consists of trade and quote files timestamped in milliseconds. These files are processed by BNP-Paribas Equities & Derivatives Quantitative R&D Histo team. Among other things, this processing consists in storing the data in a special format that saves the necessary disk space,

and especially makes data requests very efficient.

For this study, we consider three world indices: Eurostoxx 50, Dax 30 and Kospi 200. These indices have in common a very liquid derivatives market, and are, by other aspects, very different, as discussed in the sequel.

Kospi 200 options are the most liquid exchange listed derivatives in the world as far as trading volume is concerned. In 2010, Kospi 200 options traded in Korea accounted for 70% of the index options volume traded worldwide with more than 3.5 billions contracts[1]. This is certainly due to the fact that these contracts are much smaller than those traded on other exchanges. Most traded options on Kospi 200 are those of next month expiry. Almost all the trading activity takes place on the exchange. The most actively traded options can register more than 100000 trade events in one day.

Options on Dax and Eurostoxx are also very actively traded. Their turnover is greater than that of Kospi 200 options, but due to a bigger contract value for Dax and Eurostoxx options on the one hand, and to a greater number of traded maturities on the other hand, the number of trades for any particular option is much smaller than in the Kospi 200 case. In fact, next month and quarterly maturities up to 18 months are actively traded. Even if a non negligible proportion of the trades takes place on the OTC markets, the option order books are very active and quotes are regularly updated (the time scale here is of the order of the second).

By other aspects, these three markets are very different. For example, Dax index is a total return index, which means that dividends are reinvested, contrary to both other indices. Additionally, while trading activity on Eurostoxx and Dax derivatives is essentially done by finance professionals, Korean market presents the curious property that approximately the third of trading activity on Kospi 200 options is done by individual householders, a fact unique to this market. Moreover, in the KRX market, options are more liquid than futures. In fact, in 2010, while the number of traded contracts on options exceeded 3.5 billions, the number of traded futures contracts is below 100 millions. See http://eng.krx.co.kr for details.

Although the three markets are very different, our investigation gives the same results for all of them, allowing us to interpret these findings as stylized facts of the joint dynamics of option and underlying markets.

1.2 Theoretical framework

Under the real world probability measure, we assume that stock price dynamics are given by a stochastic volatility diffusion as in the equations below :

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^1 \tag{1.1}$$

$$dV_t = \nu dt + \zeta dW_t^2 \tag{1.2}$$

$$d < W^1, W^2 >_t = \rho dt, \tag{1.3}$$

where we make the usual (but not at all trivial) assumption that the price C_t of an option on the underlying S is a smooth function of time, spot and underlying variance, i.e. $C_t = C(t, S_t, V_t)$.

Therefore, the Itô formula gives us the dynamics of this option price :

[1] source : World Federation of Exchanges, 2010 Derivatives Markets Survey Report, http://www.world-exchanges.org

$$
\begin{aligned}
dC_t = &\left[\frac{\partial C}{\partial t} + \mu S_t \frac{\partial C}{\partial S} + \frac{1}{2} V_t S_t^2 \frac{\partial^2 C}{\partial S^2} + \nu \frac{\partial C}{\partial V} + \frac{1}{2}\zeta^2 \frac{\partial^2 C}{\partial V^2} + \sqrt{V_t}\zeta\rho S_t \frac{\partial^2 C}{\partial S \partial V}\right] dt \\
&+ \sqrt{V_t} S_t \frac{\partial C}{\partial S} dW_t^1 \\
&+ \zeta \frac{\partial C}{\partial V} dW_t^2.
\end{aligned} \tag{1.4}
$$

Stochastic terms dominate drift terms as their order of magnitude is $\sqrt{dt}$ whereas it is dt for the latter. Additionally, we will work with detrended data and assume that this enable us to ignore the drift term in the above equation, leaving us with the stochastic part only. So we end with the detrended option price process :

$$
dC_t = \sqrt{V_t} S_t \frac{\partial C}{\partial S} dW_t^1 + \zeta \frac{\partial C}{\partial V} dW_t^2. \tag{1.5}
$$

Option price is then, as expected, governed by two sources of randomness : the spot, and its stochastic volatility. Notice however that, as this stochastic volatility is itself correlated with the spot, part of its own randomness can be attributed to the noise governing the spot price itself. To isolate the pure spot noise one can write :

$$
dC_t = \underbrace{\left(\sqrt{V_t} S_t \frac{\partial C}{\partial S} + \zeta\rho \frac{\partial C}{\partial V}\right)}_{\text{pure spot risk}} dW_t^1 + \underbrace{\zeta\sqrt{1-\rho^2}\frac{\partial C}{\partial V}}_{\text{pure volatility risk}} dZ_t, \tag{1.6}
$$

where now W_t^1 and Z_t are independent brownian motions.

Recalling that the detrended spot price is $dS_t = \sqrt{V_t} S_t dW_t^1$, we rewrite the above equation in terms of dS_t so as to emphasize the option and spot price dynamics relationship :

$$
dC_t = \underbrace{\left(\frac{\partial C}{\partial S} + \frac{\zeta\rho}{\sqrt{V_t} S_t}\frac{\partial C}{\partial V}\right)}_{\text{Effective delta}} dS_t + \underbrace{\zeta\sqrt{1-\rho^2}\frac{\partial C}{\partial V}}_{\text{pure volatility delta}} dZ_t. \tag{1.7}
$$

The so-called *effective delta* in the above equation carries the total first order sensitivity of the option price to spot movements in this framework. This is the classical delta $\frac{\partial C}{\partial S}$ corrected by an ajustment factor accounting for the correlation between the spot and its volatility. The orthogonal term carries the volatility specific noise, hence the name *pure volatility delta*.

To summarize, we end with a regression relationship between option price and spot price increments :

$$
dC_t = \beta dS_t + \alpha dZ_t. \tag{1.8}
$$

Obviously, the coefficients of the above regression relationship are time dependent and may be stochastic. In (Mykland and Zhang, 2008) the authors study the problem of inference of such an instantaneous β in a generalized one factor framework of implied volatility dynamics and its implication on option hedging. For our purpose, we assume that the regression coefficients do not vary a lot within one day, which is the time window that we consider to begin our study. The regression of option and underlying price increments can then be conducted, quantifying the underlying spot influence on the option price dynamics. The residual of this regression carries the *pure volatility risk*, for which we will explore, if any, the relationship with the realized variance of the underlying. This is done in the next section.

1.3 Intraday joint dynamics of option and underlying prices

When dealing with statistical time series analysis, a trade off has to be made between the stationarity assumption on the one hand, which is necessary for the estimation of different quantities of interest such as correlations, volatilities, regressions and so on.. and the length of the data set on the other hand, as this has to be sufficiently long in order to assume that the estimators have converged.

This problem is harder when we deal with option price time series because an option fundamentally changes from day to day. For instance, today's ATM 30-day maturity option becomes a 15-day maturity option in two weeks, with a non negligible theta effect that may be hard to quantify without many modelling assumptions. Moreover, the option may no longer be considered ATM due to the spot movements, which completely changes its characteristics. So, comparing option prices spanning several days may be misleading as we are not talking about the same financial product (unlike a stock price which is generally the same from day to day). Consequently, some care have to be taken in the study of the joint dynamics of option and underlying prices.

In order to study raw option prices without relying on any modelling assumptions, we look at intraday dynamics. In the case of intraday price changes, we can consider that the maturity of the option does not change, and generally, spot movements leave the option in the same moneyness zone. With high frequency data, we still have long enough price series to get reliable estimations.

However, there is another issue with high frequency data, not specific to option prices. If we want to make a regression study of two or more price time series, we begin by sampling them at a given frequency to synchronize them, and then conduct our regression. In an intraday study with high frequency data, we are then tempted to sample the data at the highest possible frequency in order to obtain the longest possible time series and ensure a better convergence rate of our statistical estimators. Unfortunately, the ideal sampling frequency is not necessarily the greatest possible, because of, among other phenomena, the Epps effect.

1.3.1 Epps effect

Epps effect (Epps, 1979) is the observed decrease of estimated correlation of price time series as the sampling frequency increases. It is mainly due to the asynchronicity of trading, as well as discretization effects (Münnix et al., 2010). It is well documented for stock/stock price time series. We observe the same effect for the asset and its derivative product.

Table 1.1 presents an illustration of this effect. We take trades and quotes data for a 22-day maturity call option on Kospi 200 and the future[2] on the same index, both maturing the same day on March 2008. We consider mid prices for both securities. Estimation of the daily correlation of price returns, sampled at different time interval lengths, is shown in Table 1.1.

As shown by the confidence intervals of Table 1.1, asymptotic correlation is reached in few minutes for pairs of index future and option considered. The order of magnitude of the phenomenon in stock/stock case is of 20 to 30 minutes. This shortening of the time scale can be explained by delta effect of future's price on option's price: this mechanical effect leads to a

[2]For all our studies in the sequel of the joint dynamics of option and underlying prices, we take the future price as a proxy for the underlying index. This is a good proxy for many reasons : the future is tradable and is the main hedging instrument for options on the index. As we are not interested in trends but in the noise governing the dynamics, we can assume it is the same for the future and for the spot itself. Even if the future has in general a stochstic interest rate component, this can be ignored as we consider short maturities, enabling to assume that interest rates are deterministic. In fact, future/index correlations for the maturities we consider are always greater than 99%.

fairly fast adjustment of option quotes when future prices move.

Table 1.1: Epps Effect

Sampling period in seconds	Estimated correlation	95% Conf interval	99% Conf interval
300	97.10%	[95.5% , 98.1%]	[94.8% , 98.4%]
180	96.48%	[95.0% , 97.5%]	[94.5% , 97.7%]
150	95.83%	[94.3% , 96.9%]	[93.7% , 97.2%]
120	95.19%	[93.7% , 96.3%]	[93.1% , 96.6%]
90	93.41%	[91.6% , 94.8%]	[91.0% , 95.2%]
60	93.50%	[92.1% , 94.6%]	[91.6% , 95.0%]
45	91.56%	[90.0% , 92.8%]	[89.5% , 93.2%]
30	90.84%	[89.5% , 92.0%]	[89.1% , 92.3%]
20	87.63%	[86.2% , 88.9%]	[85.7% , 89.3%]
10	80.57%	[79.1% , 81.9%]	[78.6% , 82.4%]
5	69.94%	[68.4% , 71.4%]	[67.9% , 71.8%]
2	52.63%	[51.3% , 53.9%]	[50.9% , 54.3%]

*Note.*Correlation versus Sampling period for a 22-day Call option on Kospi 200 and the index future. Confidence intervals construction relies on the Fisher Transformation of the sample correlation coefficient, as well as a bivariate normality assumption.

There exist some techniques to deal with asynchronicity of data in covariance estimation. We can cite (Hayashi and Yoshida, 2005) who propose an unbiased and consistent estimator of realized covariance of two diffusions observed asynchronously and in discrete time, that does not require any data pre-processing. Another variance/covariance estimator using all available data without extra manipulation is proposed in (Mancino and Malliavin, 2002). Tests of the first estimator with our data give satisfactory results, in accordance with a few minutes data sampling period. As we will need to conduct regression studies (and even multidimensional regressions), we rely on time series sampling. Doing that with a period of 5 minutes seems to be a good trade off between the time needed for the Epps effect to disappear, and the need to retain enough data for our estimators to give reliable results, as we work in an intraday setting. This 5-minute sampling allow us also to assume that any other microstructure noise has disappeared.

1.3.2 Regression analysis

As usual in high frequency data, the regression analysis is conducted on price increments of the option and the future. The regression relationship established in section 1.2 is well observed in the data. Typical option price versus future price increment plots look like the Figure 1.1.

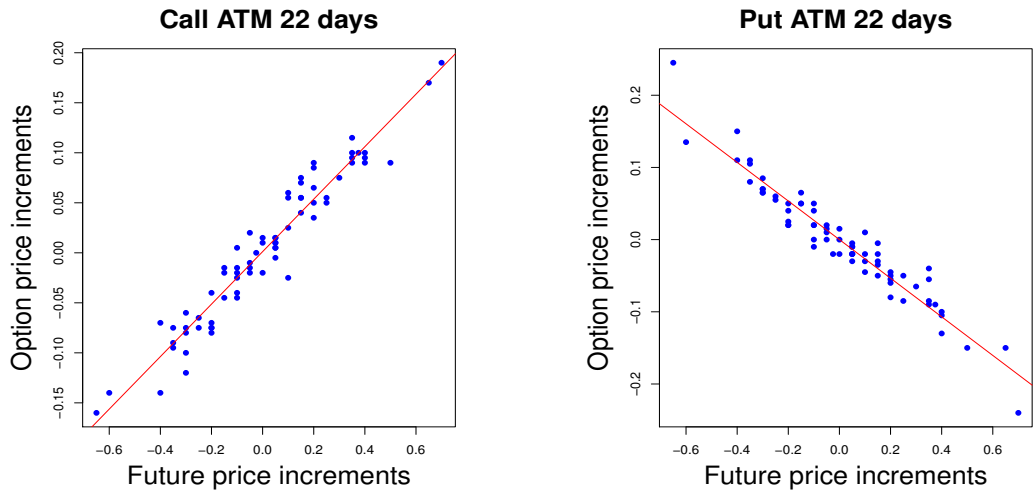

Figure 1.1: Evidence of the linear regression relationship of option price increments and future price increments as observed in the data (here Kospi 200). As expected, the slope is positive for the Call and negative for the Put.

We perform this regression on our universe of options and futures. Every day, the actively traded (for kospi 200) or quoted (for Dax and Eurostoxx) options are considered, as well as the front maturing future. Even if the three markets considered are very different by multiple aspects -as noticed in the data description- the results are roughly the same: a very good linear regression relationship (nearly 0 p-values) with a very high coefficient of determination R^2.

R^2 of option / future increments regression

The observed coefficient of determination R^2 is very high, and is most of the time very close to 1. In Figure 1.2, we plot as an example the calculated R^2 for all the options on Eurostoxx and Dax that we consider. Every point in the graph is the R^2 obtained for a 1-day regression of option price increments on future price increments.

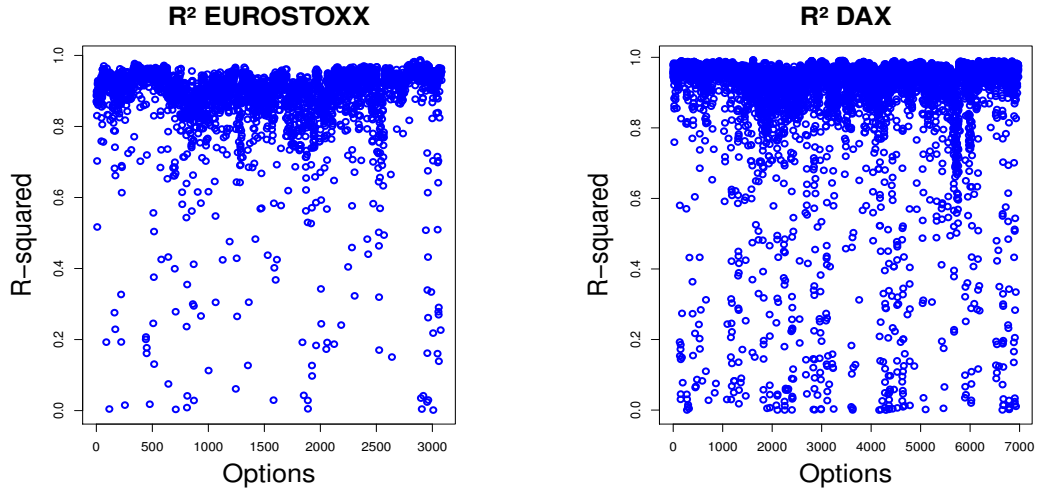

Figure 1.2: Observed R^2 of option/future increments regression for Eurostoxx and Dax.

Even if some of the regressions in the Figure 1.2 reveal a bad quality of fit as shows the

small R^2 value, further investigation allows to discard them. Indeed, we inspect individually the days and options giving an R^2 which is less than 0.5. In all these cases, the individual investigation reveals the presence of an outlier that highly impacts the regression. In Figure 1.3, such a situation is presented for a Dax option: in this case, at two different times, errors in the fed best ask quote were responsible for a large move in the option mid price which feeds our regression analysis. When these outliers are taken away (and they are generally corrected in the market data), the R^2 takes a very high value, which is clear on the graph.

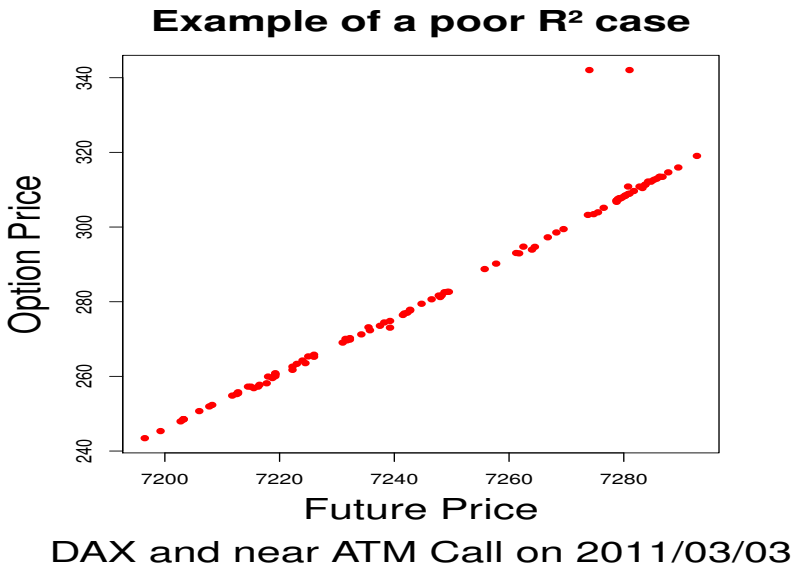

Figure 1.3: Example of a poor R^2 case : presence of the two outliers largely accounts for the bad quality of fit denoted by a small R^2.

As a conclusion, most of option price variance is explained by the underlying variance, but an orthogonal factor does exist, hence the strictly less than one R^2 value. Motivated by the stochastic volatility model intuition, we seek this orthogonal factor in the realized variance time series.

Multiple linear regression on the future and its realized variance

Recall that the purpose of the regression we are conducting is to relate the option price dynamics to the underlying dynamics, as well as the variance of underlying dynamics, as suggested by stochastic volatility models. The second step of our intraday study is then to calculate the intraday realized variance time series. To estimate these variance time series, we sample the futures price data for the three indices every second, and then use the classical variance estimator on different time windows, namely 15, 20 and 25 minutes:

$$V_t = A \times \sum_{i=1}^{N} \ln(\frac{S_{i+1}}{S_i})^2, \tag{1.9}$$

where A is the adequate annualization factor for each time window.

This gives us three time series of realized variance, one for each time window considered. Motivated by the above noted considerations on the Epps effect, we then construct 5-minute spaced time series of realized variance. For example, for variance time series constructed with a 25 minute window, begin at 9:30, take the future prices between 9:05 and 9:30 sampled every second and calculate the above estimator, obtaining V_1, then move to 9:35, take the future

prices between 9:10 and 9:35 sampled every second and compute the same estimation obtaining V_2 and so on and so forth... We pick also the future prices and option prices at the times of variance calculation to do our regressions.

As an example, in Figure 1.4, we plot the spot of Eurostoxx against its realized volatility (defined as the square root of variance, and calculated in a 20 minute window here). Notice the decreasing volatility during the middle of the day due to the decreasing market activity during lunch time. Notice also the overall negative correlation of volatility movements and price movements, but near the end of the day, where intense activity before the closing raises the volatility even if the price goes up.

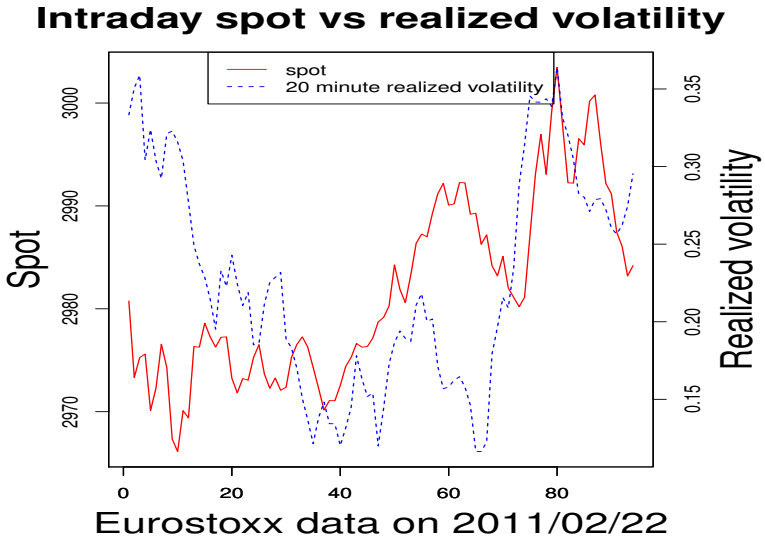

Figure 1.4: This figure ensures the plausibility of the realized volatility estimated : observe the lunch time effect in the middle of the day, and the overall U-shape of daily volatility.

We then make a multiple linear regression of option price increments on spot and variance increments. Over-all R^2 of the total regression remains unchanged compared to the one factor regression of the preceding subsection, as well as the slope of the spot factor. Moreover, the regression hypothesis on the realized variance is clearly rejected with p-values exceeding 0.5.

Even if Figure 1.4 advocates the reliability of our realized variance estimator, it can be argued that our high sampling frequency of future prices used to estimate the realized variance time series induces spurious autocorrelations of returns that can be due for example to order splitting, and that inflates the realized variance estimator.

To avoid such pitfalls, we perform the same analysis with another realized variance estimator that is less sensitive to the market microstructure noise, namely the Garman-Klass variance estimator (Garman and Klass, 1980):

$$V_t = \frac{1}{2}\left(h_t - l_t\right)^2 - \left(2\ln\left(2\right) - 1\right)\left(c_t - o_t\right)^2, \tag{1.10}$$

where h_t, l_t are respectively the highest and lowest log-price for the considered time window. And o_t (respectively c_t) is the opening (respectively the closing) log-price for this time window.

Using this estimator leads to the same results. The intraday realized variance does not have any effect on the option price dynamics. Option price variance is almost entirely explained by

the underlying price variance. On some days, their correlation seems to be perfet. Observe for instance Figure 1.5 graph of a Eurostoxx Put option versus the future:

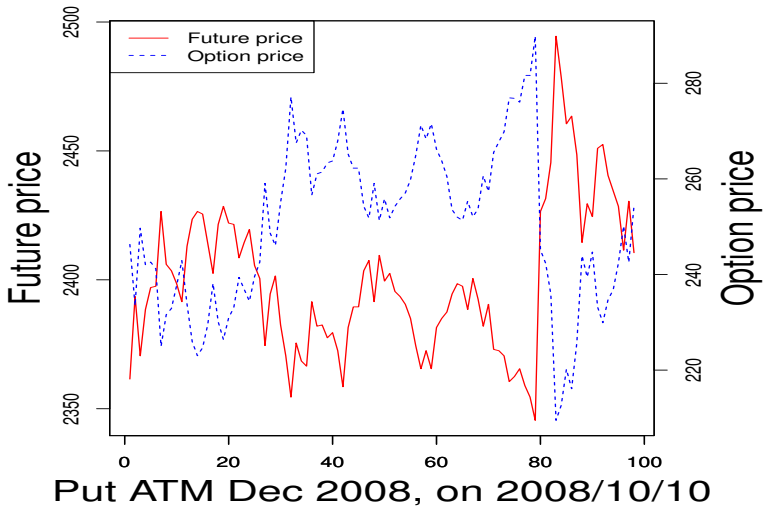

Figure 1.5: The anticorrelation of intraday put price and future price seems perfect.

These empirical observations are in total contradiction with the intuition behind stochastic volatility models. In fact, when one sees Figure 1.5, one wonder if the stochastic volatility is worth modelling at all! Does an R^2 of 0.9 let any place for another factor for modelling the option price dynamics? How does this quantitatively compare to stochastic volatility model predictions?

1.4 Stochastic volatility model predictions

1.4.1 implied R^2

We recalled in section 1.2 that the option price dynamics in the stochastic volatility framework are driven by two factors: one coming from the stock and the second being what we identified as a pure volatility risk, namely:

$$dC_t = \underbrace{\left(\sqrt{V_t}S_t\frac{\partial C}{\partial S} + \zeta\rho\frac{\partial C}{\partial \sigma}\right)}_{\text{A}} dW_t^1 + \underbrace{\zeta\sqrt{1-\rho^2}\frac{\partial C}{\partial V}}_{\text{B}} dZ_t. \tag{1.11}$$

Empirical investigation reveals a very high weight of the spot price in the option price dynamics, which explains the very high observed R^2. But what do stochastic volatility models quantitatively predict? Do they equally spread option price variance over the underlying and its variance, or do they project the risk on one factor preferably to the other?

To answer this question, we have to quantify the weights (A and B) in Equation (1.11) relatively to each other. These weights depend on unobservable quantities, such as the volatility of variance ζ or the underlying/variance correlation ρ. Fortunately, these quantities are not altered by an absolutely continuous change of probability. So, the calibration of a pricing model specified under the risk neutral measure should make them observable, giving us orders of magnitude for the weights governing the spot process and the pure variance process in the

option price evolution, and allowing us to compare model predictions to the data.

In order to design a benchmark, we choose to calibrate Heston model to the observed option prices. Under the risk neutral measure, this model writes

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^1 \tag{1.12}$$

$$dV_t = \kappa(\theta - V_t)\, dt + \zeta \sqrt{V_t} dW_t^2 \tag{1.13}$$

$$d < W^1, W^2 >_t = \rho dt. \tag{1.14}$$

This choice is naturally explained by the analytical tractability of Heston model. The model being affine, its charcteristic function is known in closed form and the pricing by FFT as described in (Carr and Madan, 1999) is straightforward, allowing a fast calibration, pricing and greeks calculation.

To design the calibration set, we rely on historical data for rate curves (EUR and KRW), and collect liquid option mid prices for our three indices. This leaves us with the next month maturity options for the Kospi 200[3], and with maturities less than 18 months for the Dax and Eurostoxx. At the end of every trading day of our three indices (precisely half an hour before the close), mid-prices of actively traded options are observed for every traded maturity. Put-Call parity relation enables us to determine the forward price of each index. These markets being very liquid, this is a reliable method to calculate the forward. Indeed, the values of the forward we find by applying the relation to different option strikes generally differ by less than 0.05%, this difference being explained by asynchronicity of price contributions and market microstructure noise. We then take the median forward given by the Put-Call parity relation applied to near the money strikes.

We define our objective function as the sum of squared differences between model and market prices as in (Cao et al., 1997), and conduct the minimisation procedure thanks to (Lourakis, Jul. 2004).

In Table 1.2, we give mean calibration results and their standard deviations for the three indices for the period of March to April 2011. Globally, orders of magnitude of the parameters correspond to what is generally reported in the literature, see for example values in (Gatheral, 2006). The only strange value concerns the κ of Kospi 200: as we have only one smile, the speed of mean reversion is really huge.

Table 1.2: Heston Model Calibration Results

Index	σ_0	κ	θ	ζ	ρ
Eurostoxx	0.2255	1.8893	0.0778	0.5739	-0.9072
	(0.0289)	(0.6072)	(0.0066)	(0.1653)	(0.0847)
Dax	0.2027	1.9887	0.0677	0.68311	-0.7698
	(0.0274)	(0.6128)	(0.0402)	(0.2164)	(0.0455)
Kospi 200	0.1737	9.5877	0.1990	0.9892	-0.5813
	(0.0303)	(6.7695)	(1.0397)	(0.3067)	(0.0854)

Note. Mean and standard deviation of the calibrated Heston parameters on the set of actively contributed options for the months of March and April 2011.

[3]liquidity is concentrated in contracts with the nearest expiration in the Kospi 200 options market. This can be explained by the huge proportion of individual investors, which is due to many reasons like the small size of the contracts and individuals' appetite and enthusiasm for trading. An analysis done in 2001 but still relevant by many aspects, can be found here : http://www.futuresindustry.org/fi-magazine-home.asp?a=994

Recall that our aim is to have orders of magnitude of the weights in Equation (1.11) relative to each other, so after the calibration of the models, we have to calculate the derivatives of the price with respect to the spot and to the volatility, to obtain the quantities A and B of Equation (1.11). These quantities are easily calculated within the Heston model by the same Fourier transform method leading to the option price, where some algebra is applied to calculate derivatives of the characteristic function with respect to the spot and to the volatility. Notice that even a calculation by finite differences is fast enough. The values obtained for A and B enable us to calculate the ratio $\frac{A^2}{A^2+B^2}$ which is the ratio of the variance carried by the underlying dynamics to the total variance of option price increments. It is then the R^2 of a regression of option prices on spot prices implied by our stochastic volatility model.

Figure 1.6 shows a plot of this implied R^2. To our surprise, the results are not far from the really observed ones: the implied R^2 in a stochastic volatility framework is still very high (around 0.9). Individual investigation of poor values reveal poorly calibrated options by the model and can be safely ignored. The result is therefore in accordance with the empirical finding. Stochastic volatility model attributes the major part of option price variance to the underlying price movements, in accordance with observation in the data. Option price variance due to the underlying volatility dynamics is far smaller in these models.

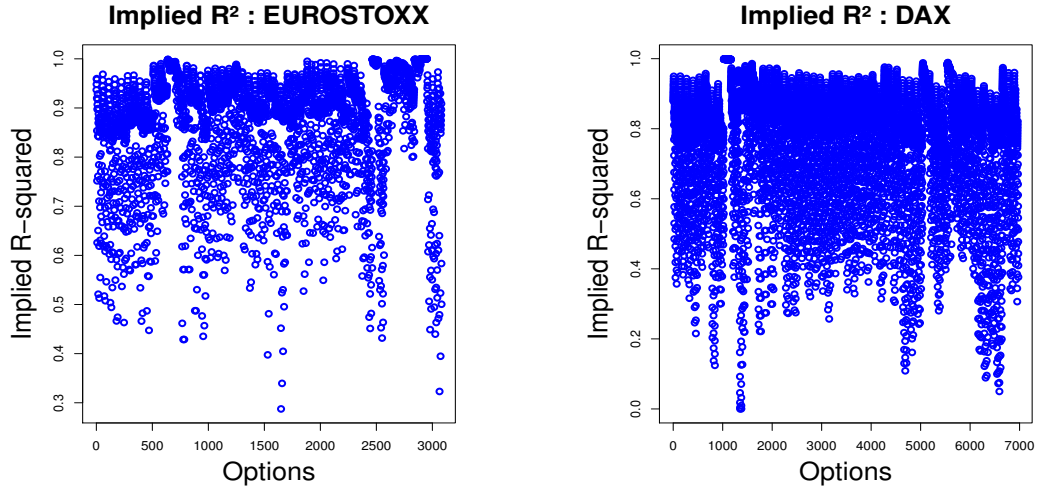

Figure 1.6: implied R^2 corresponding to the ratio of the variance explained by the spot process to the total variance of the option price process in a stochastic volatility framework, where we calibrate a Heston model to the data to calculate values of the different parameters in equation (1.11).

To push the investigation further, we see in the next section if the model predicts the same slope as seen in the data in the option/underlying regression relationship.

1.4.2 Regression slope versus model delta

Slopes of the regressions in the previous section are always less than 1 in absolute value and, as expected, positive for Calls and negative for Puts. Recall that these slopes represent the total spot sensitivity of the option in our stochastic volatility framework, and were interpreted as being the *effective delta* of the option in the regression analysis. If our intuition is correct, these slopes give a model-free hedge ratio of the option allowing to hedge all the spot dependence. A legitimate question is then if this *model-free* delta is a robust hedge ratio, and how does it compare to the Heston model total delta, namely :

$$\Delta_{Heston} = \left(\frac{\partial C}{\partial S} + \frac{\zeta \rho}{\sqrt{V_t} S_t} \frac{\partial C}{\partial V} \right) \tag{1.15}$$

Hedging performance of the regression slope

In order to quantify the robustness of this model-free hedge ratio, we perform the following experiment:

- For every index, we collect the options that have been actively traded for at least 20 days since their inception. Even if we do not require these days to be consecutive, it is actually the case in practice, as the option is actively traded when it is relatively near the money, and virtually disappears when it is deeply IN our OUT of the money, or actually disappears when it matures, which are the two retained cases for closing the position as explained later. That leaves us with 975 options on Kospi 200, 527 options on Eurostoxx, 748 options on Dax, for the period between January 2007 and April 2011.

- For every option, we consider that the trader takes a short position on the first day of data availability. The retained price C_0 is the median of best BID quotes of the last 15 minutes of the trading day sampled every second. At the same time, the trader performs the regression $dC_t = \beta dS_t + dZ_t$ of option price increments on future price increments during the trading day and uses the calculated *effective delta* to enter a future position (on the front maturing future). The retained price of the futures is the median mid price of the last 15 minutes of the trading day sampled every second. As the spread on future prices for such liquid indices is very tight, working with mid price instead of bid or ask price depending on the sign of the position simplifies the study without impacting the results. As no cash is needed to enter the futures position, the cash position at inception of the portfolio consists in C_0. Notice also that we take median prices instead of mean prices as taking the median automatically gets rid of outliers and errors in the data.

- At the end of every trading day, the cash position earns the risk free rate, that we calculate based on historical Euribor rate curves for Eurostoxx and Dax, and Koribor rate curves for the Kospi 200. The trader reperforms the regression and updates the delta position, earning $\Delta_{t-1} \times (S_t - S_{t-1})$ that he adds to his cash position.

- Close the option position if the option goes out the forward moneyness region $[0.85, 1.15]$ as it will no longer be actively traded, or on the eve of the maturity day, by buying back the option. The retained price C_T for the option is the median of best ASK quotes of the last 15 minutes of the trading day sampled every second. Closing the option at the eve of its maturity day avoids us to enter in the subtleties of the regulations of the last trading day of options and futures. That simplifies the analysis without impacting the conclusions.

- For each option, we obtain then a final PnL, that we transform into a dimensionless number, by scaling it by the initial investment, i.e. relative pnL $= \frac{\text{PnL}}{C_0}$. This will make comparison between different options and even between indices easier and allow aggregation of data.

In Table 1.3, we report means and standard deviations of calculated relative PnL for the three indices under consideration. The three of them are positive, gaining around 20% of the initial investment. This confirms that the *effective delta* is actually a good hedging ratio. The strictly positive sign of the final PnL in the three cases is not at all surprising : banks take margins! Notice that these margins seem to be greater in Kospi 200 option market. This can be explained by the great number of non professional agents in this market.

Table 1.3: Effective Delta Hedging Results

Index	mean relative PnL	std dev
KOSPI 200	0.2489727	0.4992516
EUROSTOXX 50	0.2057907	0.3349121
DAX 30	0.2141580	0.2456583

Note. Means and standard deviations of the relative PnL of the model-free hedging strategy.

For completeness, we also plot in Figure 1.7 the density of relative realized PnL of this model-free hedging strategy.

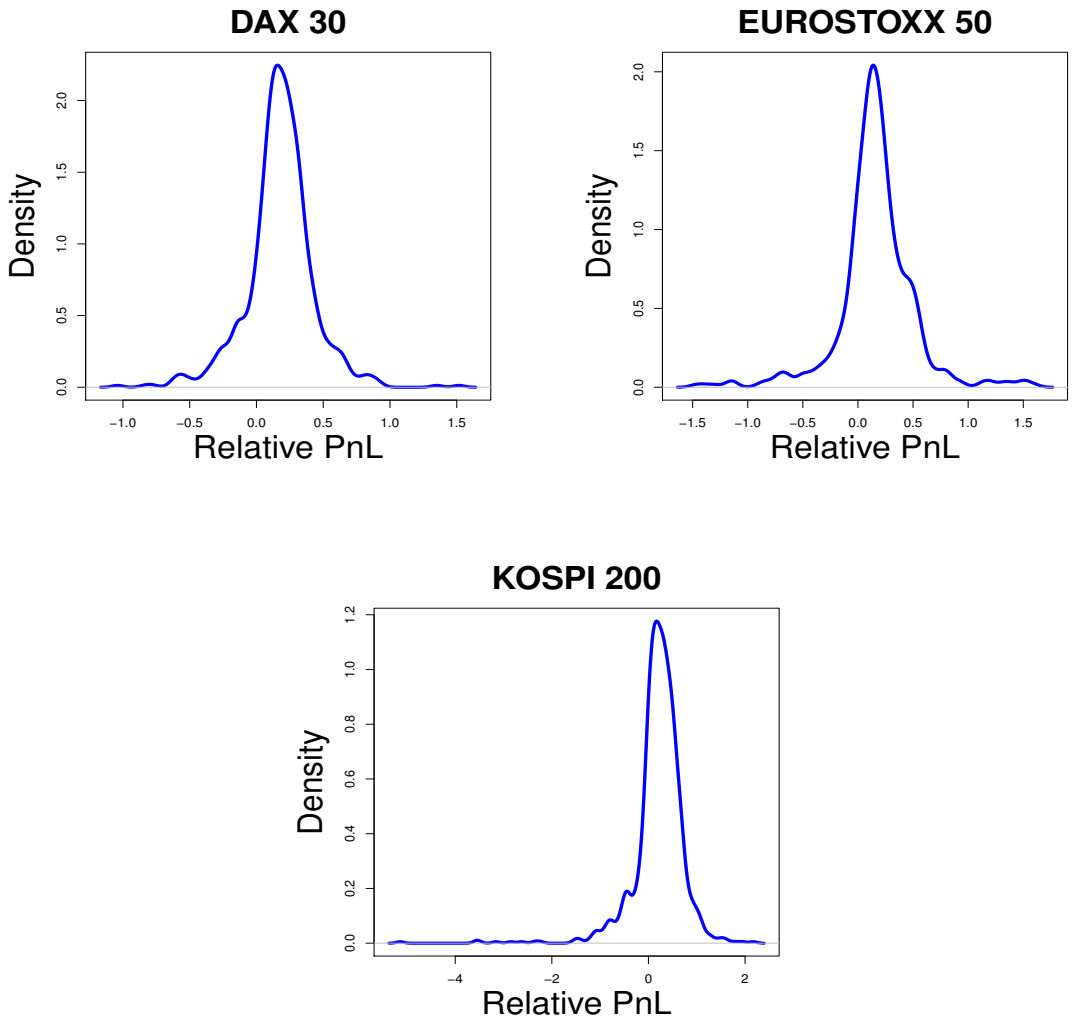

Figure 1.7: Density of the relative PnL of a model-free-delta-hedged option position.

We conclude that this model-free hedge ratio is robust enough. We compare it in the sequel to other hedge ratios obtained by pricing models.

model free vs model deltas

The regression analysis of the option price dynamics revealed the option "effective delta". We stated in section 1.4.2 that this is actually a robust model-free hedge ratio. In this section, we

compare this hedge ratio to other model-based deltas. We obtain typically figures looking like Figure 1.8 :

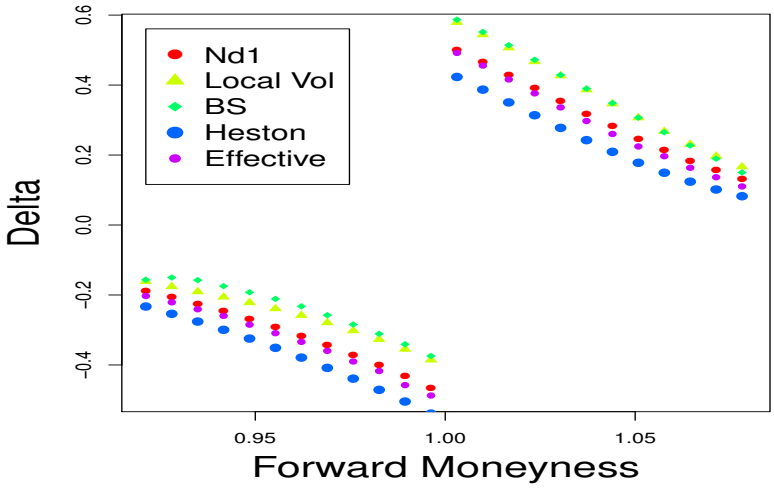

Figure 1.8: Comparison of different delta measures on one month Dax options.

Candidate hedge ratios are listed below:

- **Nd1** is the delta obtained by simply inverting the Black Scholes formula to obtain the implied volatility Σ of the option, and then calculating the Black and Scholes delta: $\Delta = N(d_1)$ where N is the standard normal cumulative distribution function, and $d_1 = \frac{\ln\left(\frac{S}{K}\right)+\left(r+\Sigma^2\right)T}{\Sigma\sqrt{T}}$.
- **Local vol** is the delta obtained for a local volatility model *a la Dupire* that we calibrate to the observed Dax implied volatility surface.
- **BS Delta** is the total Black Scholes delta, which takes into account the implied volatility smile: $\Delta = \frac{\partial C}{\partial S} + Vega\frac{\partial \Sigma}{\partial S}$.
- **Heston** is the local minimum variance hedge ratio in the Heston model : $\Delta = \frac{\partial C}{\partial S} + \frac{\zeta\rho}{\sigma S}\frac{\partial C}{\partial \sigma}$.
- **Effective** is the regression coefficient of option price increments on future price increments as in section 1.4.2.

Observe that **Local vol** delta and **BS Delta** are close to each other, as both of them take into account the smile without adding a dynamical part (dependence of the smile itself on the spot). Heston model yields the closest results to the model-independent effective delta, which proved to be a robust hedge ratio in the study of section 1.4.2. This is conform to the findings of (Cao et al., 1997) that modelling stochastic volatility yields the best dynamical model behavior, i.e. the best hedge performance.

Notice however that practitioners, when using local volatility or Black Scholes volatility, generally add a dynamical term in the smile, letting the volatility surface move as the spot moves. In this case, implied volatility surface is parametrized as $\Sigma = \Sigma(S, m)$ where m is the moneyness, say $m = \frac{K}{S}$.

The first dependence in S isolates the dynamics of the implied volatility, whereas the second reflects the mechanical volatility change when the spot moves, changing the moneyness of the option. We then have:

$$d\Sigma = \frac{\partial \Sigma}{\partial S} dS + \frac{\partial \Sigma}{\partial m} dm \tag{1.16}$$

where $\frac{\partial \Sigma}{\partial m}$ is the skew of the smile, and $dm = \frac{\partial m}{\partial S} dS = -\frac{K}{S^2} dS$.

So finally :

$$C = C\Big(S, \Sigma\big(S, m\,(S)\,\big)\Big) \tag{1.17}$$

And the total delta writes :

$$\frac{dC}{dS} = \Delta_{BS} - \underbrace{\text{Vega}_{\text{BS}} \times \text{Skew} \frac{K}{S^2}}_{\text{static correction}} + \underbrace{\text{Vega}_{\text{BS}} \times \frac{\partial \Sigma}{\partial S}}_{\text{dynamic correction}} \tag{1.18}$$

Comparison, at the end of the trading day, of the total delta with the effective delta calculated by regression can give an idea of the quality of the dynamical correction. We will return to these considerations in the sequel.

We conclude from this section that stochastic volatility models do well in capturing dynamical option price properties. This answers positively the question raised in the previous section of whether this feature is worth modelling. The answer is fortunately yes, and the implied R^2 study as well as the delta study of this section reinforce the well established results of empirical research dealing with robust dynamical model performances of stochastic volatility models.

However, stochastic volatility models tell a different story of option price formation and evolution than that found in the data in section 1.3. In the model, the underlying stochastic variance is a factor driving option prices, which could not be observed in the data.

A criticism that can be made to the study in section 1.3 is that of time scales. In fact, small time scales considered can result in a noisy, or non convergent realized volatility estimator. Larger time scales, however, will result in a fewer data points within a trading day, making any intraday statistical study worthless.

Recall that one problem when dealing with options is that the nature of a given option changes from day to day. Its maturity as well as its moneyness are not the same, and hence, so are its dynamical properties. In order to be able to enlarge the time scale of the study, we must rely on a maturity- and moneyness-independent quantity, enabling us to treat it as being the same financial quantity from day to day. This will be done in the next section.

1.5 Several day analysis : implied volatility perspective

The previous section made us more confident in the stochastic volatility paradigm as a reduced form description of the option price dynamics. With this regained confidence in this framework, we push its analysis further in order to come with a quantity that characterizes option price dynamics, without depending on a maturity nor a moneyness measure.

Recall that in Heston model the instantaneous variance writes

$$dV_t = \kappa\,(V_\infty - V_t)\,dt + \zeta \sqrt{V_t} dW_t, \tag{1.19}$$

and the expected instantaneous variance in this model obeys a simple first order ODE, whose solution is

$$\mathbb{E}\left[V_T\right] = V_\infty + (V_0 - V_\infty)\, e^{-\kappa T}. \tag{1.20}$$

The expected total variance from time 0 to time T, which in a deterministic time dependent volatility model would be the Black Scholes implied variance, as established for example in (Lee, 2005) is:

$$\mathbb{E}\left[\frac{1}{T}\int_0^T V_t dt\right] = V_\infty + (V_0 - V_\infty)\,\frac{1 - e^{-\kappa T}}{\kappa T}. \tag{1.21}$$

This functional form can then be applied to fit the term structure of the implied volatility. Implied volatility expansions as in (Lewis, 2000), give us a second order polynomial form of the smile as a function of the moneyness. We then apply these functional form as well as the term structure parametrization to fit option smiles.

We end with

$$\Sigma_{K,T} = \Sigma_{ATM} + Smile \times \ln\left(\frac{K}{S}\right) + Curve \times \ln\left(\frac{K}{S}\right)^2 \tag{1.22}$$

$$\Sigma_{ATM} = \sqrt{V_\infty + (V_0 - V_\infty)\,\frac{1 - e^{-\kappa T}}{\kappa T}} \tag{1.23}$$

Notice that in practice, some care is taken to avoid volatility explosions on the wings and the polynomial parametrization is tweaked for very high and low strikes in order to obey non arbitrage conditions established in (Lee, 2004).

This is a very pleasant parametrization of the smile that results from stochastic volatility dynamics. Not only does it describe the smile in a simple polynomial form, but also the presence of three parameters is coherent with empirical studies of dynamical behavior of volatility surfaces. For instance, (Cont and da Fonseca, 2002) find that daily variations of implied volatility surfaces can be satisfactorily explained by three factors. The first of these factors presents the overall shift in the volatility surface, and accounts for more than 90% of the variability. This can be read in the quantity V_0 of the suggested parametrization. Indeed, V_0 characterizes the overall level of the smile surface. Moreover, this is a maturity and moneyness independent quantity. It is therefore a good candidate for our study.

In (Durrleman, 2008), the author establishes the convergence of at the money implied volatility to the spot (continuous part) volatility in a general model with a Brownian component and a jump component of finite variation. In our continuous diffusion setting, this gives a nice financial interpretation of V_0 which is the zero-maturity limit of at the money implied variance, and then, according to (Durrleman, 2008), the spot variance.

We then translate the question on the dynamics of option prices in terms of this quantity : what drives V_0 dynamics? More precisely, we expect to see a spot component in the dynamics of V_0, and we will seek for a stochastic volatility component as done previously with raw prices. Independence of V_0 with respect to maturity and moneyness will enable us to aggregate several days of data in our observation. We will not be constrained to estimate realized volatility in small time windows, circumventing the problem raised about the time scale of the study of section 1.3.

The first task is to isolate the spot component in the V_0 dynamics.

1.5.1 Spot dependence

V_0 is defined in a Heston stochastic volatility setting as the limit of at the money implied variance for short maturities. As such, it is equivalent to the spot instantaneous variance according to (Durrleman, 2008). It is naturally a stochastic quantity.

In the Heston model framework, (Gatheral, 2006) gives a formula for the conditional expectation of the instantaneous variance :

$$\mathbb{E}\left[V_t|x_T\right] \approx \hat{V}_T + \rho\zeta \frac{x_T}{W_T} \int_0^T \hat{V}_s e^{-\lambda(T-s)} ds, \tag{1.24}$$

where :

$$x_t = \ln \frac{S_t}{K} \tag{1.25}$$

$$W_T = V_\infty T + (V_0 - V_\infty) \frac{1 - e^{-\kappa T}}{\kappa} \tag{1.26}$$

$$\lambda = \kappa - \frac{1}{2}\rho\zeta \tag{1.27}$$

$$\hat{V}_s = \left(V_0 - V_\infty \frac{\kappa}{\lambda}\right) e^{-\lambda s} + V_\infty \frac{\kappa}{\lambda}. \tag{1.28}$$

The interesting conclusion of these equations is that when V_0 is calibrated for $K = S_0$, it evolves linearly with respect to $x = \ln \frac{S}{S_0} \approx \frac{S}{S_0} - 1$, when S is not far from S_0.

In order to check in the data if this linear approximation is plausible, we fit every five minutes our parametrized implied volatility model to option mid quotes, then we plot the so obtained pairs (S, V_0). The Figure 1.9 shows a typical kind of graph, here obtained for Eurostoxx. Modelling linearly the response of V_0 to the spot movements is clearly a good approximation. Unsurprisingly, the slope of the regression is negative, reflecting the leverage effect of spot movements in implied volatility dynamics. That was clear in the equations above, as the slope is proportional to the spot/volatility correlation ρ of Heston model, which is usually a negative parameter (as found in our estimations).

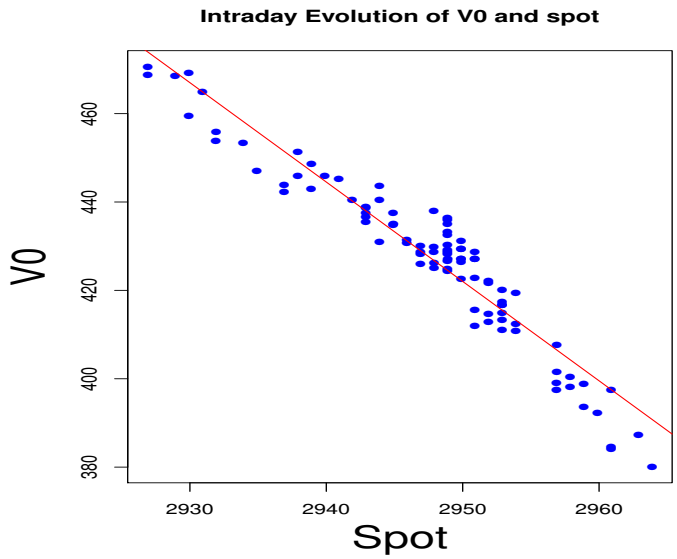

Figure 1.9: Graph of intradaily calibrated V_0 versus spot (for Eurostoxx, on April 12th 2011) : Evidence of the linear relationship as a good approximation.

Therefore :

$$V_0 = V_0^0 + \beta \times \ln\frac{S}{S_0} \text{ where } V_0^0 = V_0\left(S_0\right)$$

Thus, the fixed strike volatility parametrization (near the money, to avoid arbitrage problems on the wings) becomes :

$$\Sigma_{K,T} = \Sigma_{ATM}\left(S\right) + Smile \times \ln\left(\frac{K}{S}\right) + Curve \times \ln\left(\frac{K}{S}\right)^2 \tag{1.29}$$

$$\Sigma_{ATM}\left(S\right) = \sqrt{V_\infty + \left(V_0^0 + \beta \times \ln\frac{S}{S_0} - V_\infty\right)\frac{1-e^{-\kappa T}}{\kappa T}}. \tag{1.30}$$

Notice also that now, the option delta writes

$$\Delta = \frac{dC}{dS} \tag{1.31}$$

$$= \frac{\partial C}{\partial S} + \frac{\partial C}{\partial \Sigma}\frac{\partial \Sigma}{\partial S} \tag{1.32}$$

$$= \Delta_{BS} + \text{Vega}_{\text{BS}} \times \left(\frac{\partial \Sigma_{ATM}\left(S\right)}{\partial S} - \frac{Smile}{S} - Curve \times \frac{2}{S} \times \ln\left(\frac{K}{S}\right)\right), \tag{1.33}$$

where $\text{Vega}_{\text{BS}} \times \frac{\partial \Sigma_{ATM}(S)}{\partial S}$ accounts for the before mentioned dynamical correction of the delta.

1.5.2 Pure volatility effect on V_0

In order to isolate the spot effect in V_0, we have to come with an estimate of β.

β is itself stochastic and depends on market conditions. A careful look at the implied volatility parametrization reached so far tells us that a zero β indicates a pure sticky delta regime. We can also calculate a value $\beta\left(S\right)$ for which the smile would be sticky strike[4]. For our parametrization this can not be fully achieved and these are rather good news, as a fully sticky strike smile is arbitrable as established in (Daglish et al., 2007). So finally β characterizes the regime of volatility which depend on market conditions, as described in (Derman, 1999).

An exhaustive analysis is then conducted :

- The smile is calibrated to a market snapshot every 5 minutes, giving us series of V_0.
- The β process is given by a daily β estimation.

All this preparation enable us to study the spot orthogonal component of V_0. Recall that our purpose is to seek if realized volatility has an effect on the dynamics of option prices that we capture here thanks to V_0.

In order to seek this effect, we calculate the daily realized volatility. We estimate it thanks to the quadratic variation estimator as in section 1.3. We then estimate the following regression :

$$V_0^{close} - V_0^{open} - \underbrace{Slope \times \frac{\Delta S}{S}}_{\text{spot effect}} = \alpha \times \underbrace{\left(realizedVol - V_0^{open}\right)}_{\text{orthogonal movement}}$$

On the right left hand side of the above equation is the daily evolution of V_0 from which we substracted the spot effect as explained by the β. In the right hand side, the algebraic quantity

[4] it is the value of β for which $\frac{\partial \Sigma(K,T)}{\partial S} = 0$

$(realizedVol - V_0^{open})$ captures the realized volatility effect. It is written as a difference between actually realized volatility, and V_0 at the open of the market. The intuition behind that is that V_0 being an implied volatility, it is the guess of market agents of future realized volatility. We are therefore asking the question if market agents, at the end of the day, update their new volatility guess (V_0^{close}) based on the volatility that was actually realized, raising it when realized volatility raises, and decreasing it in the opposed case.

Figure 1.10 and in Figure 1.11 are scatter plots illustrating this regression relationship. The graphics do not exhibit a clear linear dependence. Even if we plotted a regression line, its slope is very close to 0.

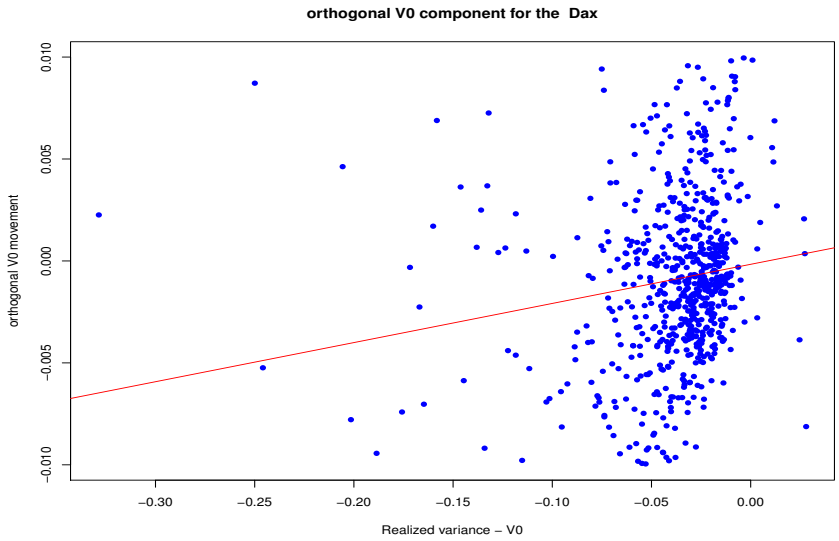

Figure 1.10: Effect of Realized Volatility on V_0 dynamics for the DAX.

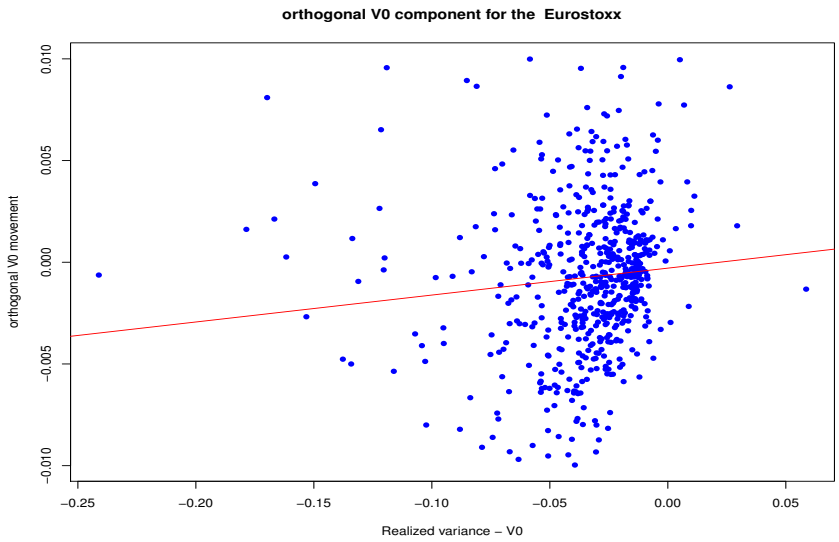

Figure 1.11: Effect of Realized Volatility on V_0 dynamics for the EUROSTOXX.

The criticism made to the intraday study that realized variance estimator may not have converged cannot be made here, as we estimate realized variance on the whole day. There is no dependence of option prices on realized quadratic variation in the data. All is in the spot, and even if a second factor is needed to capture dynamical properties of option prices, the realized quadratic variation does not help to predict it.

1.6 Conclusion

As a conclusion, we see that while there is a great temptation to see the stochastic volatility models as structural ones, deducing option prices and their dynamics from fundamental market quantities (which are the spot and its instantaneous variance, with the latter made observable thanks to the availability of high frequency data and estimation models), the reality of the option markets completely denies that.

Stochastic volatility models are just reduced form models. They meet a lot of success in the description of the overall market behavior, but do not give the structural mechanism governing option price formation and its evolution.

Price and volatility time series analysis conducted so far does not give satisfactory results when it comes to price formation mechanisms. In order to study these latter, we need a greater zoom on the picture in the microscopic level. We will then deal with order book events.

Electronic order books feed intra day databases with detailed information on every single event affecting them, therefore revealing the trading process. At this granularity level, the trading process consists of a collection of irregular point occurrences of events along the time axis, such as limit order arrivals, market order arrivals and cancellations. This naturally leads to the use of point processes as principal modelling tool to deal with such phenomena.

The simplest model for a point process is the homogeneous Poisson Process. Unfortunately, it is inadequate for modelling of order book events. The reason is that one striking characteristic of these events is their clustering in time, a feature that the completely random and memoryless homogeneous Poisson Process is unable to capture.

Therefore, in the next chapter, we turn our attention to Hawkes processes. They present an adequate and versatile framework for order book events modelling.

Chapter 2

Hawkes Process : Fast Calibration, Application to Trade Clustering and Diffusive Limit

Note : A part of this chapter is published in Journal of Futures Markets, Volume 34, Issue 6, pages 497-606, June 2014.

Abstract

In this chapter, we provide explicit formulas for the moments and the autocorrelation function of the number of jumps over a given interval for a self-excited Hawkes process. These computations are possible thanks to the affine property of this process. Using these quantities, an implementation of the method of moments for parameter estimation that leads to a fast optimization algorithm is developed. The estimation strategy is applied to trade arrival times for major stocks that show a clustering behaviour, a feature the Hawkes process can effectively handle. As the calibration is fast, the estimation is rolled to determine the stability of the estimated parameters. A forecasting test underlining the advantages of the Hawkes process as a modelling framework is performed. Lastly, the analytical results enable the computation of the diffusive limit in a simple model for the price evolution based on the Hawkes process. This determines the connection between the parameters driving the high frequency activity and the daily volatility.

Contents

Introduction

Trading activity leads to time series of irregularly spaced points that show a clustering behaviour. This stylized property suggests the use of the Hawkes process, a point process mathematically defined by (Hawkes, 1971), which is an extension of the classical Poisson process that possesses this clustering property. It explains the large number of works on trading activity and more generally high-frequency econometrics based on this process as a modelling framework. To name only a few let us quote (Hewlett, 2006), (Bowsher, 2007), (Large, 2007), (Bacry et al., 2013a) or (Muni Toke and Pomponio, 2012)[1].

There are other stochastic processes possessing this clustering property, they are often more sophisticated than the Hawkes process in the sense that their dynamic involves several lags and strong nonlinearities. They are actively studied in the econometrics literature, see (Hautsch, 2012) for a general overview. As opposed to these sophisticated stochastic processes the Hawkes process has a likelihood function which is known in closed-form. As a consequence, most of the existing literature focuses on the estimation of the dynamics. However, despite or because of its simplicity the Hawkes process has several advantages from an analytical point of view, that we will develop in this chapter, that allow for very interesting applications.

First, we show how to compute in closed-form the moments of any order of the number of jumps over a given time interval. This analytical tractability even extends to the autocorrelation function of the number of jumps. As such, we can develop an estimation strategy based on these quantities which, compared with the likelihood estimation strategy, is extremely fast. This aspect is crucial when it comes to applications such as high frequency trading activity which requires a fast estimation procedure. The maximization of the likelihood function, that can take several minutes, cannot be used in real applications. What is more, with the estimation being immediate, we can roll the estimation procedure and study the parameter stability which is an essential aspect in practice.

Second, thanks to its analytical tractability we can explicitly compute the impulse function of the Hawkes process and therefore the trading activity, modelized by this process, can easily be analysed.

Third, we can easily perform some forecast analysis and specify the horizon beyond which the Hawkes process does not perform better than a very simple model.

Fourth, within a simple toy model for a stock based on the Hawkes process we compute the diffusive limit for the asset and therefore make the link between the microscopic activity (i.e. the trading activity at high frequency) to the macroscopic activity (i.e. the daily volatility as used in the Black-Scholes model) explicit. To perform such analysis the analytical tractability of the Hawkes process turns out to be essential and underlines the advantages related to the simplicity of this process. This work falls in a new trend of the literature developed by (Cont et al., 2010), (Cont and De Larrard, 2011), (Cont and De Larrard, 2012), (Bacry et al., 2013a), (Abergel and Jedidi, 2013a), (Bacry et al., 2013b), (Kirilenko et al., 2013), (Abergel and Jedidi, 2013b) aiming at connecting these two scales (the high frequency quantities and the daily quantities).

The structure of this chapter is as follows. In the first section, we describe the analytical framework which comprises the basic properties of the Hawkes process as well as the Dynkin formula that will be our main mathematical tool. Using these results, the computation of the moments and the autocorrelation function of the number of jumps over a given time interval

[1] For non financial applications see (Vere-Jones, 1970), (Veen and Schoenberg, 2008), (Lewis and Mohler, 2011) and (Mohler et al., 2011).

is provided. This section also contains the usual optimization algorithms used in the literature and the method of moments based on the analytical results. In a second part, we present the data, various estimation results and an impulse response analysis allowed by the model, as well as a forecasting experiment. This second part is completed with a toy model for a stock for which we derive the limit properties. Finally, we conclude and provide some useful results in the appendix.

2.1 The Analytical Framework

2.1.1 Dynamics and affine structure of the moment-generating function

The Hawkes process was defined in (Hawkes, 1971) and is a self-excited point process whose intensity depends on the path followed by the point process. More precisely, the point process is determined by the intensity process $(\lambda_t)_{t\geq 0}$ through the relations:

$$\mathbb{P}\left[N_{t+h} - N_t = 1 | \mathcal{F}_t\right] = \lambda_t h + o\left(h\right) \tag{2.1}$$
$$\mathbb{P}\left[N_{t+h} - N_t > 1 | \mathcal{F}_t\right] = o\left(h\right) \tag{2.2}$$
$$\mathbb{P}\left[N_{t+h} - N_t = 0 | \mathcal{F}_t\right] = 1 - \lambda_t h + o\left(h\right), \tag{2.3}$$

where $(\mathcal{F}_t)_{t\geq 0}$ is a filtration on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ containing the filtration generated by $(N_t)_{t\geq 0}$.

The intensity follows the dynamic:

$$d\lambda_t = \beta\left(\lambda_\infty - \lambda_t\right) dt + \alpha dN_t. \tag{2.4}$$

A jump of N_t at a given time will increase the intensity which increases the probability of another jump thanks to equation (2.1) and justifies the use of the term "self-exciting" to qualify this process. The jumps tend to cluster but the process does not blow up because the drift becomes negative whenever the intensity is above $\lambda_\infty > 0$ (β is by hypothesis positive) and prevents any explosion. Furthermore, applying Ito's lemma to $e^{\beta t}\lambda_t$ yields:

$$\lambda_t = e^{-\beta t}\left(\lambda_0 - \lambda_\infty\right) + \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s. \tag{2.5}$$

From (2.5) we also observe that the impact of a jump on the intensity dies out exponentially as time passes. For the existence and uniqueness results we refer to chapter 14 of (Daley and Jones, 2008) and references therein, of particular interest is (Brémaud and Massoulié, 1994).

As t gets larger the impact of λ_0, the initial value for the intensity, vanishes leaving us with:

$$\lambda_t \sim \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s.$$

Our presentation differs slightly from the usual one found in the literature where the Hawkes intensity is written as:

$$\lambda_t = \lambda_\infty + \int_{-\infty}^t \alpha e^{-\beta(t-s)} dN_s. \tag{2.6}$$

The equation (2.6) leads to a stochastic differential equation similar to (2.4), the process starts infinitely in the past and is at its stationary regime. In our case we have a dependency with respect to the initial position λ_0 in equation (2.5) but, as mentioned above, for t large enough its impact will vanish.

Our presentation for the Hawkes process follows closely (Errais et al., 2010) and is motivated by the fact that we want to perform stochastic differential calculus.

The process $X_t = (\lambda_t, N_t)$ is a Markov process in the state space $D = \mathbb{R}_+ \times \mathbb{N}$. This is a key property that will give us very powerful tools to investigate the distributional properties of the process. Among these tools is the infinitesimal generator.

Consider a sufficiently regular function $f : D \to \mathbb{R}$, the infinitesimal generator of the process, denoted $\mathcal{L}$, is the operator acting on f such that:

$$\mathcal{L}f(x) = \lim_{h \to 0} \frac{\mathbb{E}_t^x [f(X_{t+h})] - f(x)}{h}$$

with $\mathbb{E}_t^x [.] = \mathbb{E}^x [.|\mathcal{F}_t]$ and $X_t = x$ ($\mathbb{E}_0 [.] = \mathbb{E} [.]$).

In the case of a Hawkes process, this writes:

$$\mathcal{L}f(x) = \beta(\lambda_\infty - \lambda_t) \frac{\partial f}{\partial \lambda}(x) + \lambda_t \Big[f(\lambda_t + \alpha, N_t + 1) - f(x) \Big]. \tag{2.7}$$

For every function f in the domain of the infinitesimal generator, the process:

$$M_t = f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_u)\, du$$

is a martingale relative to its natural filtration (see for example Proposition 1.6 of chapter VII in (Revuz and Yor, 1999)), thus for $s > t$ we have:

$$\mathbb{E}_t \left[f(X_s) - \int_0^s \mathcal{L}f(X_u)\, du \right] = f(X_t) - \int_0^t \mathcal{L}f(X_u)\, du$$

by the martingale property and we finally obtain the important Dynkin formula:

$$\mathbb{E}_t [f(X_s)] = f(X_t) + \mathbb{E}_t \left[\int_t^s \mathcal{L}f(X_u)\, du \right]. \tag{2.8}$$

This formula allows the computation of conditional expectations of functions of the Markov process (λ_t, N_t) which turns out to be very useful when the expectation on the right hand side can be easily calculated. In the following subsection we will rely heavily on this formula to compute some distributional properties of Hawkes process.

As underlined in (Errais et al., 2010) the process $X_t = (\lambda_t, N_t)$ is a Markov process which is *affine*. A Markov process is said to be affine if its drift, covariance matrix and jump intensities are affine functions of the state vector. The Markov process $X_t = (\lambda_t, N_t)$ that we consider here obviously satisfies this definition. A thorough treatment of affine processes is given in (Duffie et al., 2003), and a more detailed and accessible treatment can be found in (Duffie, 2007).

The main ingredient of the analytical tractability of affine Markov processes is that their conditional Laplace transform is exponentially affine with respect to the present state variables of the process. This can be shown to be an equivalent definition of affinity, see for instance (Duffie, 2007). Therefore, a closed-form solution for the moment-generating function of our process is available.

Let $u = (u_1, u_2)^\top \in \mathbb{R}^2$, the conditional moment-generating function of $X_T = (\lambda_T, N_T)$ is defined as $f(t, X_t) = \mathbb{E}_t^x \left[e^{u^\top X_T} \right] = \mathbb{E}_t^x \left[e^{u_1 \lambda_T + u_2 N_T} \right]$. Clearly, $f(t, X_t)$ must be a martingale and the function f satisfies:

$$\frac{\partial f}{\partial t}(t, X_t) + \mathcal{L} f(t, X_t) = 0 \tag{2.9}$$

with boundary condition $f(T, X_T) = e^{u^\top X_T}$. As $X_t = (\lambda_t, N_t)$ is a Markov affine point process we guess the solution of (2.9) is an exponential affine form of the state variable, that is to say:

$$f(t, X_t) = e^{a(t) + b(t)\lambda_t + c(t)N_t}. \tag{2.10}$$

Setting this guess into equation (2.9) we obtain the system of ordinary differential equations:

$$\frac{\partial a}{\partial t} = -\beta \lambda_\infty b(t) \tag{2.11}$$

$$\frac{\partial b}{\partial t} = \beta b(t) + 1 - e^{\alpha b(t) + c(t)} \tag{2.12}$$

$$\frac{\partial c}{\partial t} = 0 \tag{2.13}$$

with terminal conditions $a(T) = 0$, $b(T) = u_1$ and $c(T) = u_2$.

The above system of ODE fully characterizes the moment-generating function and the Laplace transform of the process which completely determines its distribution. However, an explicit solution for the equation (2.12) is usually not available. From the moment-generating function (2.10) we can retrieve the moments of the process *after* differentiating it with respect to u_1 or u_2, depending on which variable is considered, and evaluating the resulting function for $(u_1, u_2) = 0$. This computation leads to the differentiation of the system of ODE with respect to u_1 or u_2. Therefore, we can compute the different moments of the number of jumps by successively differentiating the system of ordinary differential equations. In our particular case (i.e. for a one-dimensional Hawkes process) the computations are feasible and allow to calculate all the moments of the process.

Lastly, the computation of the autocovariance function of the number of jumps increments, that is to say:

$$\mathbb{E}_t^x \left[(N_{t_4} - N_{t_3})(N_{t_2} - N_{t_1}) \right] \tag{2.14}$$

with $t < t_1 < t_2 < t_3 < t_4$ can be obtained from (2.10) by performing successive conditioning. In that case it will introduce the intensity process, which appears on the right hand side of (2.10), and implies that the joint moment-generating function (i.e. (λ_t, N_t)) has to be evaluated. The resulting system of ODE and its differentiation become more complicated. Because the quantity (2.14) is essential, it carries the clustering property of the Hawkes process, we need to develop a simpler approach to perform the computation.

In (Errais et al., 2010), the authors rely on this approach to compute the expected number of jumps, as well as the expected intensity and its variance in the stationary regime. Computations become rapidly tedious for higher moments. The computational difficulties obviously increase with the dimension of the process. Having this generalization in mind, we prefer then to rely on an alternative approach based on the Dynkin formula, that we present here with the one-dimensional Hawkes process to illustrate the method on this simple case. We will then take advantage from its power in the next chapter when we will come to multivariate Hawkes processes. For completeness, we provide in the appendix the detailed calculations based on the moment generating function, showing the difficulties raised by this method in our case.

2.1.2 Computing the moments and the autocovariance function

Our aim in this section is to compute the moments of the process $X_t = (\lambda_t, N_t)$ and also the autocovariance of the number of jumps over a period τ. To achieve this, we rely on the

infinitesimal generator of the process given by (2.7) and Dynkin's formula (2.8). In order to obtain the expected number of jumps and the expected intensity we use the following lemma:

Lemma 2.1. *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic given by (2.4) then the expected number of jumps $\mathbb{E}[N_t]$ and the expected intensity $\mathbb{E}[\lambda_t]$ satisfy the set of ODE:*

$$\begin{aligned} d\mathbb{E}\left[N_t\right] &= \mathbb{E}\left[\lambda_t\right] dt & (2.15)\\ d\mathbb{E}\left[\lambda_t\right] &= \left(\beta\lambda_\infty + (\alpha-\beta)\,\mathbb{E}\left[\lambda_t\right]\right) dt. & (2.16)\end{aligned}$$

Proof. We apply Dynkin's formula (2.8) to $f \equiv N_t$ and taking into account the fact that:

$$\mathcal{L}f\left(X_t\right) = \lambda_t$$

we obtain:

$$\mathbb{E}\left[N_t\right] = N_0 + \mathbb{E}\left[\int_0^t \lambda_s ds\right].$$

Using Fubini-Tonnelli's theorem we have:

$$\mathbb{E}\left[N_t\right] = N_0 + \int_0^t \mathbb{E}\left[\lambda_s\right] ds. \tag{2.17}$$

Differentiating this integral equation gives (2.15).

This equation could have been obtained by recalling that $N_t - \int_0^t \lambda_s ds$ is a martingale, by definition of the intensity of a point process, as explained in (Brémaud, 1981). We nevertheless quote the Dynkin formula method as the same reasoning will prove useful for other functions as well.

To obtain the ODE (2.16) we rely again on Dynkin's formula. Following (Errais et al., 2010), let $f \equiv \lambda_t$ in (2.7) then as we have:

$$\mathcal{L}f\left(X_t\right) = \beta\left(\lambda_\infty - \lambda_t\right) + \alpha\lambda_t.$$

Dynkin's formula leads to:

$$\begin{aligned}\mathbb{E}\left[\lambda_t\right] &= \lambda_0 + \mathbb{E}\left[\int_0^t \left(\beta\left(\lambda_\infty - \lambda_s\right) + \alpha\lambda_s\right) ds\right]\\ &= \lambda_0 + \beta\lambda_\infty t + (\alpha-\beta)\int_0^t \mathbb{E}\left[\lambda_s\right] ds\end{aligned}$$

where, as before, we used Fubini-Tonnelli's theorem to swap the integration and expectation operators. Taking the differential with respect to t yields the ordinary differential equation satisfied by the expected intensity (2.16). ■

Note that (2.16) can be explicitly computed and once obtained another integration leads to the expression for $\mathbb{E}[N_t]$. If these quantities are known, a similar procedure will give higher order moments as we have:

Lemma 2.2. *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamics given by (2.4) then $\mathbb{E}[\lambda_t^2]$, $\mathbb{E}[\lambda_t N_t]$ and $\mathbb{E}[N_t^2]$ satisfy the set of ODE:*

$$\begin{aligned} d\mathbb{E}[N_t^2] &= 2\mathbb{E}[\lambda_t N_t]dt + \mathbb{E}[\lambda_t]dt & (2.18)\\ d\mathbb{E}[\lambda_t N_t] &= \beta\lambda_\infty\mathbb{E}[N_t]dt + (\alpha-\beta)\mathbb{E}[\lambda_t N_t]dt + \mathbb{E}[\lambda_t^2]dt + \alpha\mathbb{E}[\lambda_t]dt & (2.19)\\ d\mathbb{E}[\lambda_t^2] &= (\alpha^2 + 2\beta\lambda_\infty)\mathbb{E}[\lambda_t]dt + 2(\alpha-\beta)\mathbb{E}[\lambda_t^2]dt. & (2.20)\end{aligned}$$

Proof. For function $f \equiv N^2$ the equation (2.7) gives:

$$\mathcal{L}f\left(X_t\right) = 2\lambda_t N_t + \lambda_t$$

and Dynkin's formula (2.8) results in:

$$\mathbb{E}[N_t^2] = N_0^2 + 2\int_0^t \mathbb{E}[\lambda_u N_u]du + \int_0^t \mathbb{E}[\lambda_u]du \tag{2.21}$$

and differentiating this equation with respect to t leads to (2.18).

Following the same procedure for $f \equiv \lambda N$ and $f \equiv \lambda^2$ give the ODE (2.19) and (2.20), respectively. ■

These two lemmas allow the computation of different quantities useful to perform the estimation of the process. In fact, we have:

Proposition 2.3. *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamics given by (2.4) then we have the following equalities:*

$$\lim_{t\to\infty} \mathbb{E}\left[N_{t+\tau} - N_t\right] = \frac{\beta\lambda_\infty}{\beta-\alpha}\tau = \Lambda\tau \tag{2.22}$$

with $\Lambda = \frac{\lambda_\infty}{1-\alpha/\beta}$ (the stationary regime expected intensity) gives the long run expected value of the number of jumps during a time interval of length τ. The variance is:

$$\begin{aligned} V(\tau) &= \lim_{t\to\infty} \mathbb{E}\left[(N_{t+\tau} - N_t)^2\right] - \mathbb{E}\left[N_{t+\tau} - N_t\right]^2 \\ &= \Lambda\left(\tau\kappa_-^2 + \left(1-\kappa_-^2\right)\frac{\left(1-e^{-\tau\gamma_-}\right)}{\gamma_-}\right) \end{aligned} \tag{2.23}$$

where:

$$\Lambda = \frac{\lambda_\infty}{1-\alpha/\beta}\ ,\ \kappa_- = \frac{1}{1-\alpha/\beta}\ \text{and}\ \gamma_- = \beta - \alpha.$$

The covariance is given by:

$$\begin{aligned} \mathrm{Cov}\left(\tau,\delta\right) &= \lim_{t\to\infty} \mathbb{E}\left[(N_{t+\tau} - N_t)(N_{t+2\tau+\delta} - N_{t+\tau+\delta})\right] - \mathbb{E}\left[(N_{t+\tau} - N_t)\right]\mathbb{E}\left[(N_{t+2\tau+\delta} - N_{t+\tau+\delta})\right] \\ &= \frac{\lambda_\infty\beta\alpha(2\beta-\alpha)\left(e^{(\alpha-\beta)\tau}-1\right)^2}{2(\alpha-\beta)^4}e^{(\alpha-\beta)\delta} \end{aligned} \tag{2.24}$$

for $\delta > 0$.

Proof. Taking into account the initial condition $\mathbb{E}\left[\lambda_0\right] = \lambda_0$, a constant variation technique allows to find the solution of (2.16) which writes:

$$\mathbb{E}\left[\lambda_t\right] = \lambda_\infty\beta\frac{e^{(\alpha-\beta)t}-1}{\alpha-\beta} + e^{(\alpha-\beta)t}\lambda_0. \tag{2.25}$$

Notice that we easily recognize in the above equation the stability condition of our self-exciting Hawkes process, namely $\frac{\alpha}{\beta} < 1$. Using the above result in equation (2.17) yields the expression for the mean number of jumps:

$$\mathbb{E}\left[N_t\right] = N_0 + \frac{\lambda_\infty\beta\left(-1+e^{(\alpha-\beta)t}-(\alpha-\beta)t\right)}{(\alpha-\beta)^2} + \frac{\left(-1+e^{(\alpha-\beta)t}\right)}{\alpha-\beta}\lambda_0.$$

We are interested in the expected number of jumps during an interval of length τ. Using the previous computation we conclude that it is given by:

$$\mathbb{E}\left[N_{t+\tau}-N_t\right]=\frac{-\lambda_\infty\beta\tau}{\alpha-\beta}+e^{t(\alpha-\beta)}\frac{\left(-\lambda_\infty\beta+e^{(\alpha-\beta)\tau}\lambda_\infty\beta-\alpha\lambda_0+e^{(\alpha-\beta)\tau}\alpha\lambda_0+\beta\lambda_0-e^{(\alpha-\beta)\tau}\beta\lambda_0\right)}{(\alpha-\beta)^2}. \tag{2.26}$$

The equation (2.26) depends on λ_0 the initial value for the intensity which is unknown. To eliminate this value we take the limit $t\to\infty$ and under the stability condition $\frac{\alpha}{\beta}<1$ we obtain (2.22).

We need to compute the second moment of the number of jumps during a given interval, namely:

$$I=\mathbb{E}\left[(N_{t_2}-N_{t_1})^2\right]=\mathbb{E}\left[\mathbb{E}_{t_1}\left[N_{t_2}^2\right]-2N_{t_1}\mathbb{E}_{t_1}\left[N_{t_2}\right]+N_{t_1}^2\right]. \tag{2.27}$$

Using the ODE (2.18) it results that:

$$\mathbb{E}_{t_1}[N_{t_2}^2]=N_{t_1}^2+2\int_{t_1}^{t_2}\mathbb{E}_{t_1}[\lambda_u N_u]du+\int_{t_1}^{t_2}\mathbb{E}_{t_1}[\lambda_u]du \tag{2.28}$$

and when inserted in the previous equation leads to:

$$I=2\int_{t_1}^{t_2}\mathbb{E}[\lambda_u N_u]du+\int_{t_1}^{t_2}\mathbb{E}[\lambda_u]du-2\mathbb{E}\left[N_{t_1}\int_{t_1}^{t_2}\mathbb{E}_{t_1}[\lambda_u]du\right]. \tag{2.29}$$

The first integral of (2.29) can be computed thanks to (2.19) and it gives:

$$\begin{aligned}I_1&=\int_{t_1}^{t_2}\mathbb{E}[\lambda_u N_u]du\\&=\int_{t_1}^{t_2}e^{(\alpha-\beta)(u-t_1)}\mathbb{E}[\lambda_{t_1}N_{t_1}]du+\int_{t_1}^{t_2}\int_{t_1}^{u}e^{(\alpha-\beta)(u-s)}\{\beta\lambda_\infty\mathbb{E}[N_s]+\mathbb{E}[\lambda_s^2]+\alpha\mathbb{E}[\lambda_s]\}dsdu\end{aligned}$$

whilst for the third term of (2.29) is:

$$\begin{aligned}I_2&=\mathbb{E}\left[N_{t_1}\int_{t_1}^{t_2}\mathbb{E}_{t_1}[\lambda_u]du\right]\\&=\mathbb{E}\left[N_{t_1}\left(\int_{t_1}^{t_2}e^{(\alpha-\beta)(u-t_1)}\lambda_{t_1}du+\int_{t_1}^{t_2}\int_{t_1}^{u}e^{(\alpha-\beta)(u-r)}\beta\lambda_\infty drdu\right)\right]\\&=\int_{t_1}^{t_2}e^{(\alpha-\beta)(u-t_1)}du\mathbb{E}[N_{t_1}\lambda_{t_1}]+\int_{t_1}^{t_2}\int_{t_1}^{u}e^{(\alpha-\beta)(u-r)}drdu\beta\lambda_\infty\mathbb{E}\left[N_{t_1}\right].\end{aligned}$$

As we have $\mathbb{E}[N_s]=\mathbb{E}[N_{t_1}]+\int_{t_1}^{s}\mathbb{E}[\lambda_r]dr$ we arrive after substitution and simplification to:

$$I=\int_{t_1}^{t_2}\mathbb{E}[\lambda_u]du+2\int_{t_1}^{t_2}\int_{t_1}^{u}e^{(\alpha-\beta)(u-s)}\left\{\beta\lambda_\infty\int_{t_1}^{s}\mathbb{E}[\lambda_r]dr+\mathbb{E}[\lambda_s^2]+\alpha\mathbb{E}[\lambda_s]\right\}dsdu.$$

We can therefore calculate the time t expected second moment of the number of jumps occurring during an interval of length τ, by first conditioning on $\mathcal{F}_t$, obtaining $\mathbb{E}_t\left[(N_{t+\tau}-N_t)^2\right]$, which is an expression depending only on the expectations $\mathbb{E}[\lambda_t]$ and $\mathbb{E}[\lambda_t^2]$. These last two terms depend on λ_0 but by letting $t\to\infty$ we obtain their stationary regime value and get an expression independent of the initial intensity. As a result we have the second moment of the number of jumps over a time interval of length τ.

To specify further the result, we note $\lim_{t\to\infty}\mathbb{E}[\lambda_t]=\Lambda$, $\lim_{t\to\infty}\mathbb{E}[\lambda_t^2]=\Lambda_2$ and taking the limit on the expression for I we reach:

$$\lim_{t\to+\infty}\mathbb{E}\left[(N_{t+\tau}-N_t)^2\right]=\lim_{t\to+\infty}\tau\Lambda+2\int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-s)}\int_t^s drdsdu\beta\lambda_\infty\Lambda \\ +2\int_t^{t+\tau}\int^u e^{(\alpha-\beta)(u-s)}dsdu\{\Lambda_2+\alpha\Lambda\}$$

where the integral expressions are given by:

$$\int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-s)}\int_t^s drdsdu=-(\alpha-\beta)^{-1}\frac{\tau^2}{2}-(\alpha-\beta)^{-2}\tau+(\alpha-\beta)^{-3}(e^{(\alpha-\beta)\tau}-1)$$
$$\int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-s)}dsdu=-(\alpha-\beta)^{-1}\tau+(\alpha-\beta)^{-2}(e^{(\alpha-\beta)\tau}-1).$$

The equation (2.23) is deduced using the above equation and (2.22).

To compute the autocovariance function of the number of jumps during different time intervals we need to determine $\mathbb{E}_t\left[(N_{t_1}-N_t)\left(N_{t_3}-N_{t_2}\right)\right]$, where $t<t_1<t_2<t_3$. In order to simplify notations we consider the variables $\Delta_1=t_1-t$, $\Delta_2=t_3-t_2$ and $\delta=t_2-t_1$. By performing successive conditionings we get:

$$\mathbb{E}\left[(N_{t_1}-N_t)\left(N_{t_3}-N_{t_2}\right)\right]=\mathbb{E}\left[\mathbb{E}_t\left[\mathbb{E}_{t_1}\left[\mathbb{E}_{t_2}\left[(N_{t_1}-N_t)\left(N_{t_3}-N_{t_2}\right)\right]\right]\right]\right].$$

The innermost conditional expectation is:

$$\mathbb{E}_{t_2}\left[(N_{t_1}-N_t)\left(N_{t_3}-N_{t_2}\right)\right]=(N_{t_1}-N_t)\times\left[\frac{\lambda_\infty\beta\left(-1+e^{(\alpha-\beta)\Delta_2}-(\alpha-\beta)\Delta_2\right)}{(\alpha-\beta)^2}+\frac{\left(-1+e^{(\alpha-\beta)\Delta_2}\right)}{\alpha-\beta}\lambda_{t_2}\right]$$

thanks to calculations done for the first moment. Then, conditioning down by $\mathcal{F}_{t_1}$, one has to compute:

$$\mathbb{E}_{t_1}\left[\lambda_{t_2}\right]=\lambda_\infty\beta\frac{e^{(\alpha-\beta)\delta}-1}{\alpha-\beta}+e^{(\alpha-\beta)\delta}\lambda_{t_1}. \tag{2.30}$$

This results in an expression depending on $N_{t_1}\lambda_{t_1}$ and $N_t\lambda_{t_1}$. Further conditioning down with respect to $\mathcal{F}_t$, one has to calculate :

$$\mathbb{E}_t\left[\lambda_{t_1}\right]=\lambda_\infty\beta\frac{e^{(\alpha-\beta)\Delta_1}-1}{\alpha-\beta}+e^{(\alpha-\beta)\Delta_1}\lambda_t.$$

Lastly, the quantity $\mathbb{E}_t\left[\lambda_{t_1}N_{t_1}\right]$ is already known from the previous computations. Collecting all results together we determine the autocovariance of the process and to simplify the final expression we suppose $\Delta_1=\Delta_2=\tau$, so that we obtain:

$$\lim_{t\to\infty}\mathbb{E}\left[(N_{t+\tau}-N_t)(N_{t+2\tau+\delta}-N_{t+\tau+\delta})\right]=\frac{\lambda_\infty\beta\alpha(2\beta-\alpha)\left(e^{(\alpha-\beta)\tau}-1\right)^2}{2(\alpha-\beta)^4}e^{(\alpha-\beta)\delta}+\frac{\lambda_\infty^2\beta^2}{(\alpha-\beta)^2}\tau^2. \tag{2.31}$$

If we subtract the mean value we obtain the expression (2.24). ■

The strategy of taking the limit to simplify the dependency of the results with respect to the initial value of the process, which is unknown, is borrowed from (Aït-Sahalia et al., 2010) who used the Hawkes process for modelling contagion effects between stocks. These results require the stability condition $\frac{\alpha}{\beta}<1$ which allows us to put the process in its long run stationary regime.

The autocovariance function contains information regarding the self-exciting or clustering property of the Hawkes process but it is more convenient to derive the autocorrelation function. We do not provide a proof for the following result as it is straightforward to obtain from the previous proposition.

Proposition 2.4. *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamics given by (2.4) then the autocorrelation function of the number of jumps over a given interval τ is:*

$$\begin{aligned} Acf(\tau,\delta) &= \lim_{t\to\infty} \frac{\mathbb{E}[(N_{t+\tau}-N_t)(N_{t+2\tau+\delta}-N_{t+\tau+\delta})] - \mathbb{E}[(N_{t+\tau}-N_t)]\mathbb{E}[(N_{t+2\tau+\delta}-N_{t+2\tau+\delta})]}{\sqrt{var(N_{t+\tau}-N_t)var(N_{t+2\tau+\delta}-N_{t+\tau+\delta})}} \\ &= \frac{e^{-2\beta\tau}\left(e^{\alpha\tau}-e^{\beta\tau}\right)^2\alpha(\alpha-2\beta)}{2\left(\alpha(\alpha-2\beta)\left(e^{(\alpha-\beta)\tau}-1\right)+\beta^2\tau(\alpha-\beta)\right)}e^{(\alpha-\beta)\delta} \end{aligned} \tag{2.32}$$

where the lag is δ.

The above expression is always positive when $\alpha < \beta$, which is the stability condition of the process, and decays exponentially with the lag δ. The half-life depends on the difference $\alpha-\beta$, or, said differently, on $\frac{\alpha}{\beta}$, which is the L^1-norm of the kernel. As intuitively expected, the greater is the L^1-norm of the kernel (providing it remains less than unity), the greater is the process memory, and hence the autocorrelation function.

Notice also that the background intensity λ_∞ is not involved in the autocorrelation, a property that could have been expected.

For completeness, we provide in the appendix the complete expressions for the third and fourth moments.

The computation performed above allows us to determine the moments up to the second order of $(X_t)_{t\geq 0}$ as well as the autocorrelation function for the number of jumps over an interval τ. Following this approach we can compute higher order moments. The key ingredient underlying the computations is the stability of the polynomial functions with respect to the infinitesimal generator of the Hawkes process. More precisely, the expected value of a polynomial function of the process $(X_t)_{t\geq 0}$ (i.e. $\sum_{i\leq nj\leq m} a_{ij}x^i x^j$) can be expressed as a function of polynomial functions of same or lower degree. This property is a consequence of the affine structure of the Hawkes process and has been used for the classical standard affine model of (Duffie and Kan, 1996) in (Cuchiero et al., 2012) and (Filipović et al., 2013). To be more precise, if we denote $Z_t = \left(\mathbb{E}[\lambda_t], \mathbb{E}[N_t], \mathbb{E}[\lambda_t^2], \mathbb{E}[\lambda_t N_t], \mathbb{E}[N_t^2]\right)^\top$ then using the ODE obtained above the vector Z_t satisfies the ODE:

$$\frac{dZ_t}{dt} = AZ_t + B \tag{2.33}$$

with:

$$A = \begin{pmatrix} \alpha-\beta & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \alpha^2+2\beta\lambda_\infty & 0 & 2(\alpha-\beta) & 0 & 0 \\ \alpha & \beta\lambda_\infty & 1 & \alpha-\beta & 0 \\ 1 & 0 & 0 & 2 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} \beta\lambda_\infty \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The solution of this ODE is given by:

$$Z_t = e^{At}Z_0 + \int_0^t e^{A(t-s)}Bds \tag{2.34}$$

where the exponential of the matrix is computed using classical algorithms, see (Golub and Van Loan, 1996).

2.1.3 Inference strategies

In this subsection, we first present the classical Maximum Likelihood approach usually used to calibrate the Hawkes process and underline the numerical difficulties. Then, using the explicit expression for the moments and the autocorrelation function computed in the previous subsection, we develop a method of moment estimation strategy whose computational speed appears to be very fast compared to the existing alternatives.

Maximum likelihood estimation

Let $(X_t)_{t\geq 0}$ be a simple point process on $[0,T]$ and $t_1 ... t_{N_T}$ denote a realization of $(N_t)_{t\geq 0}$ over $[0,T]$, then, as established in proposition 7.2.III of (Daley and Jones, 2002), the log-likelihood of $(X_t)_{t\geq 0}$ is of the form:

$$\begin{aligned} L &= \int_0^T (1-\lambda_s)\, ds + \int_0^T \ln(\lambda_s)\, dN_s \\ &= \int_0^T (1-\lambda_s)\, ds + \sum_{i=1}^{N_T} \ln(\lambda_{t_i}). \end{aligned}$$

In the case of a Hawkes process we have:

$$\begin{aligned} L &= \int_0^T (1-\lambda_t)\, dt + \int_0^T \ln(\lambda_t)\, dN_t \\ &= \int_0^T \left(1-\lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s\right) dt + \int_0^T \ln\left(\lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s\right) dN_t \\ &= T - T\lambda_\infty + \sum_{i=1}^{N_T} \int_0^T \alpha e^{-\beta(t-t_i)} \times 1_{\{t\geq t_i\}} dt + \sum_{i=1}^{N_T} \ln\left(\lambda_\infty + \sum_{t_j \leq t_i} \alpha e^{-\beta(t_i - t_j)}\right) \end{aligned}$$

and simplifying the above integral

$$\begin{aligned} \int_0^T \alpha e^{-\beta(t-t_i)} \times 1_{\{t\geq t_i\}} dt &= \left[-\frac{\alpha}{\beta} e^{-\beta(t-t_i)} \times 1_{\{t\geq t_i\}}\right]_0^T - \int_0^T -\frac{\alpha}{\beta} e^{-\beta(t-t_i)} \times \delta_{\{t=t_i\}} dt \\ &= \frac{\alpha}{\beta} - \frac{\alpha}{\beta} e^{-\beta(T-t_i)} \end{aligned}$$

we end with:

$$L = T - T\lambda_\infty - \sum_{i=1}^{N_T} \frac{\alpha}{\beta}\left(1 - e^{-\beta(T-t_i)}\right) + \sum_{i=1}^{N_T} \ln(\lambda_\infty + \alpha A(i)) \tag{2.35}$$

where $A(i) = \sum_{t_j \leq t_i} e^{-\beta(t_i - t_j)}$.

The estimation requires a non linear optimization algorithm such as Nelder-Mead to find the maximum of this function. We stress that for each set of parameters the evaluation of this function requires a loop over the observations which for the problem at hand, trade clustering, is very large.

Some authors, such as (Ozaki, 1979), pointed out that $A(i)$ used in (2.35) satisfies a recursive relation. Indeed, defining $A(1) = 0$ then for $i \geq 1$, $A(i+1) = e^{-\beta(t_{i+1} - t_i)} \times (1 + A(i))$, simplifies the calculation of the likelihood function and speeds up evaluation. However, the calibration still takes a few minutes and a large number of function calls are performed. Any simplification of the calibration procedure is therefore of interest.

Fast Hawkes process calibration

Even with the improvement previously presented the parameter estimation procedure based on the maximum likelihood function is still very time consuming. Having computed explicitly the moments as well as the autocorrelation function for the Hawkes process a natural estimation strategy is the generalized method of moments, see (Hall, 2004) for an exhaustive treatment, and (Bollerslev and Zhou, 2002) and (Bollerslev and Zhou, 2004) for an application to finance.

The inference problem now writes:

$$\hat{\theta} = \text{argmin} \left\{ (M - f(\theta))^\top W (M - f(\theta)) \right\} \tag{2.36}$$

where M is the vector of empirical moments (eventually related to the autocorrelation function), $f(\theta)$ is the vector of corresponding theoretical moments and W is a symmetric positive definite weighting matrix.

This method of estimation is known to be consistent and asymptotically normal. It can also be shown that with a suitable choice of the weighting matrix W, the estimator is asymptotically efficient in the sense that it has the smallest covariance matrix. Intuitively, the optimal weighting matrix will attribute less weight to the noisier moments, and more weight to moments that are easiest to estimate with better accuracy. It turns out indeed that the optimal weight matrix is the inverse variance-covariance matrix of the error terms, so in our notation $W_{opt} = \mathbb{E}\left[\left(M - f\left(\hat{\theta}\right) \right) \times \left(M - f\left(\hat{\theta}\right) \right)^\top \right]^{-1}$, with $\hat{\theta}$ being the true parameter value. As this true parameter value is unknown we rely on Hansen's two-step method (Hansen, 1982) to obtain the optimal weighting matrix:

- Let W be the identity matrix and estimate the parameters. This gives a consistent estimate θ_0.
- Update the weight matrix as : $W_{opt} = \mathbb{E}\left[(M - f(\theta_0)) \times (M - f(\theta_0))^\top \right]^{-1}$, finding θ_1, and iterate this step until convergence, i.e $\theta_{k+1} \approx \theta_k$.

As we are primarily interested in speed we fix ex-ante this matrix W to reduce the discrepancies between the components of the vector involved in the objective function, sacrificing optimality of the estimator for the estimation speed[2]. In that case the optimization problem turns out to be the simple least squares method (LS in the sequel) that can be solved very quickly by Levenberg-Marquardt algorithm (we use the implementation provided by (Lourakis, Jul. 2004)). Some numerical experiments lead us to the following conclusions: the optimization problem (2.36) based on the mean and variance of number jumps during an interval τ (i.e. equations (2.22) and (2.23)), and autocorrelation function (2.32) gives good results if calibration quality and speed are taken into account. Computation time remains negligible and is faster than the MLE. If speed is not of concern, this fast computation can be done to feed the MLE with a good first guess of the parameters.

From a numerical point of view we found it simpler and more robust to work with normalized quantities, in which case the optimization problem is:

$$\hat{\theta} = \text{argmin} \left\{ \left(1 - \frac{f(\theta)}{M} \right)^\top W \left(1 - \frac{f(\theta)}{M} \right) \right\} \tag{2.37}$$

where the components of the vector $(1 - \frac{f(\theta)}{M})$ are $(1 - \frac{f_i(\theta)}{M_i})$ and involve the relative ratio of the theoretical moment to its empirically estimated counterpart. This is made possible because

[2] A similar scaling appears in (Aït-Sahalia et al., 2010) where the authors use the matrix W to control the discrepancies between the moments.

all the considered moments are different from zero. In that case we can choose for W the identity, thus simplifying the specification of this matrix tremendously. Let us emphasize that the evaluation of the empirical moments is only made *once* during the optimization procedure which explains why this estimation strategy is intrinsically faster than the MLE.

This gives us a very appealing estimation procedure. Not only is it instantaneous and this point is crucial if the objective is to apply a model to a high-frequency problem (optimal execution, price impact analysis of a trade), but it also has the advantage of being robust against data pollution, an aspect which is very common in such data and this point was already underlined in (Bacry et al., 2013a).

For instance, many data vendors round timestamps of market events to a certain precision (millisecond in the case of our data, or even up to one second for other data providers). In this case, only one timestamp will be considered in the likelihood procedure, whereas a least squares procedure will consider the actual number of events, even if they have the same timestamp.

Robustness check of the fast calibrations

To assess the quality of our estimation procedure we perform the following Monte Carlo experiment. We randomly generate the parameters λ_∞, α and β, with $\beta > \alpha$. We choose them in the interval $[0\ ,1]$ because these are the values obtained in our empirical study. We simulate a Hawkes process for a duration of $T = 8$ hours, corresponding to a trading day, using the thinning algorithm as presented in (Ogata, 1981). We carry out a calibration using both the MLE and the LS estimators. For the MLE, calibration is conducted using Nelder-Mead algorithm as implemented in the open-source library NL-opt[3]. The LS estimations are conducted using the Levenberg-Marquardt algorithm provided by (Lourakis, Jul. 2004). The initial guess for the parameters is their true value multiplied by a uniform random variable on the interval [0.5 1.5] (both algorithms have the same starting point). The estimation test is performed 25000 times for each algorithm.

Table 2.1 contains the results. The mean relative error value expressed in percent is reported for each parameter. For the standard deviation we compute the root mean squared relative error value that we also express in percent.

Table 2.1: Estimation Strategies Comparison

	MLE			Fast Calibration I			Fast Calibration II		
	λ_∞	α	β	λ_∞	α	β	λ_∞	α	β
mean	0.33	0.16	0.23	0.68	-0.26	-0.23	0.18	0.07	0.25
Std. dev.	5.74	7.13	6.69	8.39	10.66	11.04	4.41	7.47	7.69

Note. A Monte Carlo experiment comparing the MLE and fast calibration estimation strategies. Fast calibration I is an LS estimation based on the first, second moments and the autocorrelation function computed for several lags. Fast calibration II is a LS estimation based on the autocorrelation function computed for several lags (λ_∞ is then deduced using (2.22)). For each method and each parameters we report the mean relative error value and the corresponding standard deviation (both expressed in percent).

We perform two fast calibrations. The first is based on the least squares optimization problem (2.37) with W the identity, the first two moments given by (2.22) and (2.23) computed with $\tau = 60$ seconds as well as the autocorrelation function (2.32) with $\tau = 60$ seconds and lags ranging from 0 to 600 seconds (by step of 60 seconds). For the second fast calibration, the objective function (2.37) only depends on the autocorrelation function (2.32) with $\tau = 60$ seconds and lags ranging from 0 to 600 seconds (by step of 60 seconds). To obtain λ_∞, because

[3]See `http://ab-initio.mit.edu/wiki/index.php/NLopt`

the autocorrelation function does not depend on it, we use (2.22) with $\tau = 60$ seconds. That is, we suppose this equation is satisfied without error.

The MLE leads to a relative error of 0.33% for λ_∞ and similar magnitudes for α and β as we have 0.16% and 0.23%, respectively. What we call relative error is the difference between the estimated parameter and the true parameter, rescaled by the value of the true parameter. For example, the difference between the estimated λ_∞ and the true λ_∞ is 0.33% of the true value of λ_∞ that we used to generate the process.

Overall, the estimation procedure is satisfactory as mean values and standard deviations are small. For the first fast calibration, the results are reported in column titled "Fast calibration I", the mean values are also small, they are comparable to those obtained with the MLE, and the standard deviation values are roughly 1.5 times those for the MLE but remain acceptable. In conclusion, the fast calibration algorithm performs well and the computational cost is much lower. For the second fast calibration (i.e. column "Fast calibration II") similar conclusions are obtained. Namely, the mean values and the standard deviations are small. Notice also that for this second method we obtain the smallest fit errors for the background noise carried by λ_∞ and the branching structure carried by α. The fit error on β is similar for the three methods.

2.2 Applications

2.2.1 Data

We rely on tick-by-tick data from TRTH (Thomson Reuters Tick History). The data consist of trades and quotes files timestamped in milliseconds. We study two stocks in particular: BNP Paribas and Sanofi, as well as the futures on the Eurostoxx and the Dax. For each studied day we consider the futures with the shortest maturity date. The data cover the period between 2010/01/01 to 2011/12/31.

All the trading days begin at 9:00 and end at 17:30. We neglect the first and last 15 minutes in order to avoid the open and close auctions. We end with 8 trading hours per day, between 9:15 and 17:15.

Most of our study deals with trade time arrivals and statistics on the number of trades occurring on intervals of fixed length. As stated before, the fact that the timestamps have a bounded precision, the millisecond in our case, is another appealing feature of our method of estimation. Indeed, many trades will have the same time to the nearest millisecond even if they did not take place at the same time. This millisecond will count as a unique entry in the ML estimation procedure, whereas in the LS based inference all the trades will be taken into account when computing the moments. This, among other reasons, will make the method of moments more robust to data imprecision, a fact typical to high frequency data.

2.2.2 Trade clustering

As outlined in the introduction, trading activity is not a completely random and memoryless process. Consequently, the Poisson process is not suitable for modelling trade arrival times. As shown in Figure 2.1, a qqplot of inter arrival times of trades against an exponential distribution clearly rejects the Poisson process as the data generating process for the order flow.

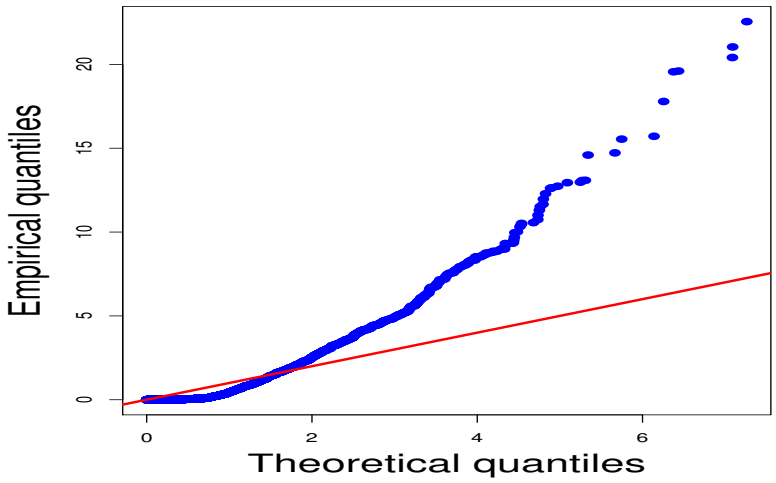

Figure 2.1: QQplot of inter-trade durations against exponential distribution. Inter arrival trade times are clearly not exponential. Graph for Eurostoxx futures trades on 2011/03/03 for the first trading hour.

In fact, trades tend to cluster and an illustration is given in Figure 2.2 where we plot an histogram of the number of trades occurring every minute during a trading day for the Eurostoxx. The clustering is graphically clear.

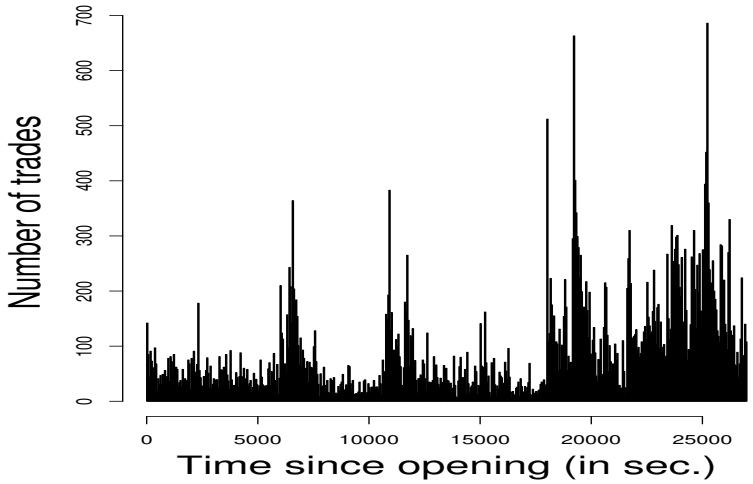

Figure 2.2: Number of trades by time intervals of 1 minute: presence of clusters is graphically apparent. Graph for Eurostoxx futures trades on 2011/03/03.

Numerous reasons can explain this clustering of trade arrival times, among them liquidity takers splitting their orders so as to minimize their market impact, or insider traders reacting rapidly to take advantage from information they have before it is widespread in the market: these justify a one sided trade clustering (i.e. either buy or sell initiated trades). On the other hand, heterogeneity of market participants is responsible for the two-sided trade clustering. A complete study can be found in (Sarkar and Schwartz, 2006).

To quantify this clustering in time we compute the correlation of the number of trades occurring during different time intervals of fixed length separated by a time lag. A plot of this

autocorrelation as a function of the lag gives information about the degree of clustering. If the data were generated by a Poisson process this autocorrelation would be equal to zero.

We want to compute:

$$\mathrm{C}(\tau,\delta)=\frac{\mathbb{E}[(N_{t+\tau}-N_t)(N_{t+2\tau+\delta}-N_{t+\tau+\delta})]-\mathbb{E}[(N_{t+\tau}-N_t)]\mathbb{E}[(N_{t+2\tau+\delta}-N_{t+\tau+\delta})]}{\sqrt{\mathrm{var}(N_{t+\tau}-N_t)\mathrm{var}(N_{t+2\tau+\delta}-N_{t+\tau+\delta})}}. \quad (2.38)$$

To this end, we count the number of trades occurring during two sliding non overlapping intervals of length $\tau = 60$ seconds separated by a certain time lag δ. We change this time lag δ from 1 second to 20 minutes by a step of 1 second. This gives 1200 number of trades autocorrelation points for the function $\delta \to \mathrm{C}(\tau,\delta)$ that we report in Figure 2.3 for different time intervals τ for a given symbol (in that case the Eurostoxx). We clearly see that the autocorrelation is positive and significant and that it decreases with the time lag. It is also remarkable that this is true for all the time interval lengths considered, and that independently from τ this memory effect seems to take around 10 minutes to become negligible.

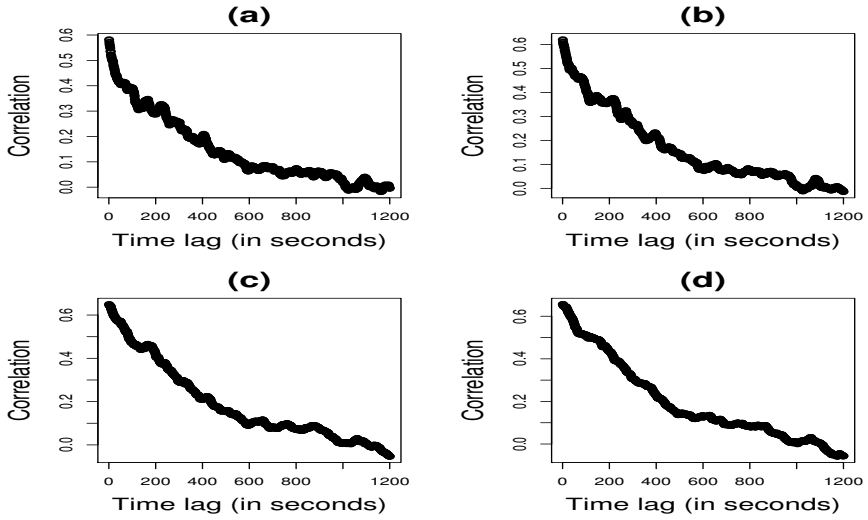

Figure 2.3: Empirical autocorrelation function $\delta \to C(\tau,\delta)$, given by (2.38), of the number of trades occurring on consecutive intervals of length τ separated by a time lag δ as a function of the lag δ. In the four plots, we change the length of the time interval τ considered. (a) corresponds to $\tau = 20$ seconds, (b) $\tau = 30$ seconds, (c) $\tau = 60$ seconds and (d) $\tau = 90$ seconds. The shape of the function remains identical, even if the plot is noisier when the time interval length decreases. The data are Eurostoxx futures trades on 2010/01/07.

Figure 2.4 confirms that the same phenomenon is observed for all the tested symbols. The absolute value of the autocorrelation is higher for the two futures, which are far more liquid than the stocks, but the same decreasing shape is observed and the time life of this autocorrelation seems to be very close for all the symbols.

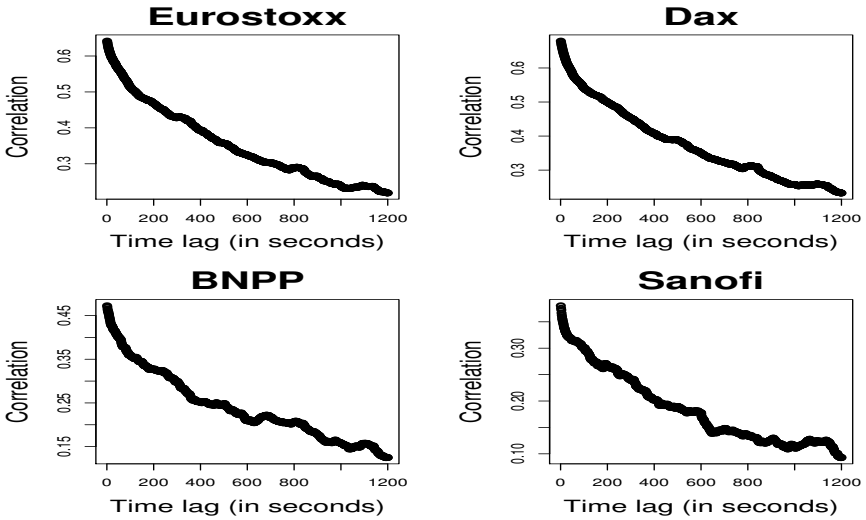

Figure 2.4: Empirical autocorrelation function $\delta \to C(\tau, \delta)$ of the number of trades with $\tau = 60$ seconds for Eurostoxx, Dax, BNPP and Sanofi, averaged every day for the month January 2010.

Overall, these stylized facts justify the use of the Hawkes process as modelling framework.

Thanks to the closed-form solutions of the previous section we estimate the parameters using the fast calibration algorithm presented in the inference subsection. Precisely, we rely on the method called "Fast calibration II" in the Monte Carlo experiment of the previous section and that proved to be efficient. Based on Figure 2.5 we choose to fit the autocorrelation function (2.32) for $\tau = 60$ seconds and δ ranging from 0 to 600 seconds by step of 60 seconds because Figure 2.4 convinces us that this choice is suitable for all the analyzed stocks. For each symbol we perform a daily calibration using the two-year sample described in the data subsection. In Table 2.2 we report the mean and median estimated values as well as the standard deviations. The table also contains in column "Poisson-λ" the intensity that we obtain if a Poisson process is calibrated.

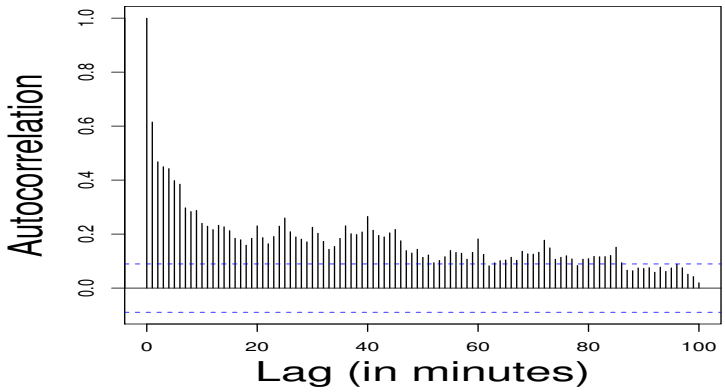

Figure 2.5: Number of trades empirical autocorrelation. The number of trades are computed for an interval $\tau = 60$ seconds and the lag δ is now measured in *minutes*. The symbol is Eurostoxx on 2010/10/26.

Table 2.2: Calibration Results

Symbol	Measure	Poisson-λ	λ_∞	α	β
Eurostoxx	Mean	1.4480	0.0625	0.0869	0.0911
	Std. dev.	0.6283	0.0209	0.0229	0.0237
	Median	1.3343	0.0593	0.0843	0.0882
Dax	Mean	1.7814	0.0664	0.0993	0.1034
	Std. dev.	0.7322	0.0249	0.0218	0.0226
	Median	1.6214	0.0609	0.0988	0.1028
BNPP	Mean	0.8627	0.0556	0.0760	0.0819
	Std. dev.	0.3923	0.0231	0.0192	0.0219
	Median	0.7438	0.0508	0.0724	0.0772
Sanofi	Mean	0.6704	0.0453	0.0747	0.0806
	Std. dev.	0.1873	0.0213	0.0212	0.0240
	Median	0.6087	0.0414	0.0700	0.0758

Note. Calibration results for two years of data. We calibrate daily a Hawkes process to the trade arrival times for each symbol. For comparison, we put the Poisson equivalent λ, defined as the mean number of trades per second for every day. For every measure, we put mean, standard deviation and median values for the calibrated parameters.

A striking fact is that λ_∞ is much smaller than the Poisson-λ as the former amounts to only 5% to 10% of the latter. Even if the Poisson-λ have to be compared (and is indeed very close) to the long term expected intensity $\Lambda = \frac{\lambda_\infty}{1-\alpha/\beta}$, as stated in Proposition 2.3, we find it interesting to compare the Hawkes process background intensity to a Poisson equivalent intensity, as this gives information on the part of the process that is attributed to the background noise, compared to that attributed to the branching structure. Our results suggest that the fraction of events due to the background process activity is small in comparison of the part due to the branching structure of the process (i.e. the self-exciting property captured by α). Note also that the stability condition ($\alpha < \beta$) is satisfied although the optimization algorithm is performed without constraints.

The quality of the fit can be assessed graphically as reported in Figure 2.6. It suggests that the Hawkes process captures well the empirical property of trade arrival times.

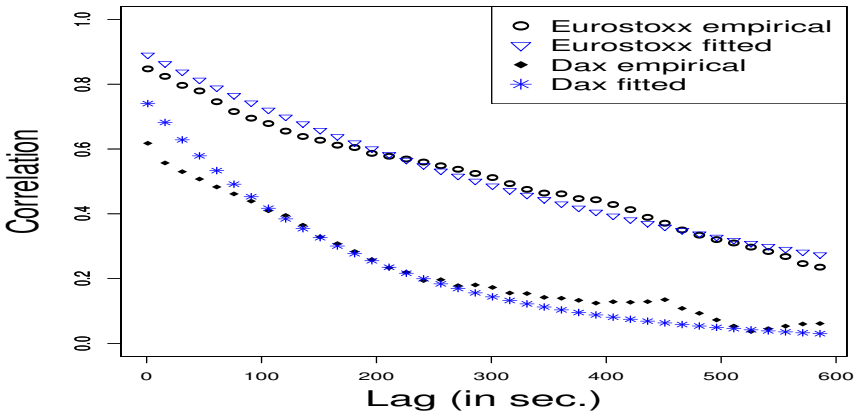

Figure 2.6: Empirical autocorrelation of the time series of the number of trades occurring during $\tau = 60$ seconds versus the theoretically fitted one. For the Dax, estimated parameters are: $\lambda_\infty = 0.0326806$, $\alpha = 0.0431643$ and $\beta = 0.0486235$ and for Eurostoxx, estimated parameters are: $\lambda_\infty = 0.033282$, $\alpha = 0.04259$ and $\beta = 0.0446049$.

For robustness, we also conduct calibrations of the model on a variety of traded assets ranging from interest rate futures to commodity, energy and foreign exchange futures. Results are reported in Table 2.3.

Table 2.3: Calibration Results for Other Assets

Symbol	Measure	Poisson λ	λ_∞	α	β	$N_{response}$
Bund	Mean	1.2742	0.0671	0.0956	0.1013	19.683
	Std. dev.	0.4434	0.0244	0.0218	0.0232	8.359
	Median	1.2063	0.0645	0.0934	0.0983	18.256
Bobl	Mean	0.6699	0.0546	0.0816	0.0894	13.236
	Std. dev.	0.1705	0.0216	0.0225	0.0257	6.455
	Median	0.6173	0.0498	0.0790	0.0858	11.329
Schatz	Mean	0.6245	0.0473	0.0877	0.0952	14.896
	Std. dev.	0.1397	0.0209	0.0211	0.0236	7.452
	Median	0.5712	0.0440	0.0835	0.0903	13.556
JPY	Mean	1.6023	0.0536	0.1130	0.1173	29.737
	Std. dev.	0.7144	0.0172	0.0207	0.0215	11.123
	Median	1.5510	0.0518	0.1133	0.1178	29.139
EURO	Mean	4.1955	0.0788	0.1220	0.1245	53.526
	Std. dev.	1.7419	0.0271	0.0192	0.0194	18.999
	Median	4.1956	0.0770	0.1258	0.1282	52.045
GOLD	Mean	2.3191	0.0852	0.1104	0.1149	27.716
	Std. dev.	0.8824	0.0290	0.0237	0.0244	10.472
	Median	2.1555	0.0815	0.1175	0.1212	25.743
Crude Oil Brent	Mean	2.0453	0.0550	0.1243	0.1279	37.528
	Std. dev.	0.7018	0.0154	0.0143	0.0147	12.877
	Median	1.9787	0.0535	0.1255	0.1302	36.860
Natural GAS	Mean	1.4524	0.0688	0.1177	0.1241	21.170
	Std. dev.	0.4362	0.0181	0.0156	0.0166	8.465
	Median	1.3653	0.0680	0.1168	0.1246	18.532
Sugar	Mean	0.8082	0.0434	0.1213	0.1289	19.964
	Std. dev.	0.3214	0.0196	0.0174	0.0190	11.930
	Median	0.6869	0.0382	0.1272	0.1351	18.539
CORN	Mean	1.0338	0.0626	0.1069	0.1146	17.348
	Std. dev.	0.4563	0.0213	0.0332	0.0334	9.400
	Median	0.9563	0.0537	0.1226	0.1312	17.333
WHEAT	Mean	1.1562	0.0639	0.1119	0.1193	18.807
	Std. dev.	0.4334	0.0242	0.0215	0.0227	8.951
	Median	1.0926	0.0594	0.1182	0.1244	17.372

Note. Calibration results for two years of data (2010 and 2011) and for different asset classes (interest rates, foreign exchange, metal commodities) to assess the robustness of the model. Schatz, Bobl and Bund are respectively the 2-year, 5-year and 10-year futures on German government bonds. For each asset class we take for daily data the front maturing futures to perform the calibration. The last column $N_{response}$ will be defined in subsection 2.2.3

For all the assets, the standard deviation values are small compared to the mean parameter values and the similarity between the mean and median values, denoting the relative stability of the calibration and absence of outliers, confirm the ability of the Hawkes process to fit the data. As for the four previous assets all the calibrations lead to solutions that satisfy the stability constraint $\alpha < \beta$, albeit no constraints were used during the optimization procedure. Lastly, let us underline the fact that $\alpha - \beta$ controls the decreasing shape of the autocorrelation function as it is clear from equation (2.32) and quantifies to which extent the data generating process

departs from a Poisson process.

2.2.3 Branching structure of trading activity

An interesting property of the Hawkes process is its branching structure. Indeed, the occurrence of a jump increases the intensity of the process, thereby the probability to observe another jump. As pointed out by (Hewlett, 2006), this results in a direct and indirect impulse response of the process intensity to a jump event. Denoting the expected increase of the process intensity at time t as a response to a jump occurring at time 0 by $f(t)$, we have the following decomposition:

- Direct response: an increase of the intensity by α that will decay exponentially as time passes, thus leading to an increase of the intensity at time s of $\alpha e^{-\beta s}$.
- Indirect response: at any time s between 0 and t, the direct increase of the intensity by $\alpha e^{-\beta s}$ leads to an indirect increase of the expected number of jumps at time t which equals to $\alpha e^{-\beta s} ds f(t-s)$; we then need to integrate over the range $[0;t]$ to obtain the total indirect effect.

As a consequence, the expected direct and indirect increase of the intensity at time t caused by a jump at time 0 writes as:

$$f(t) = \alpha e^{-\beta t} + \int_0^t \alpha e^{-\beta s} f(t-s) ds.$$

The solution of this integral equation is given by:

$$f(t) = \alpha e^{-(\beta-\alpha)t}.$$

Therefore, the $N_{response}$ which is the expected number of jumps triggered by one jump occurring at time 0 if the process is observed indefinitely (the impulse response) is:

$$N_{response} = \int_0^\infty f(s) ds = \frac{\alpha}{\beta - \alpha}. \tag{2.39}$$

We use our daily calibrations on real data of the Hawkes process to measure an average $N_{response}$ for the studied assets and report the results in Table 2.4.

Table 2.4: Market Liquidity Indicator

Symbol	Average $N_{response}$
Dax	26
Eurostoxx	22
BNPP	14
Sanofi	10

Note. $N_{response}$ as a characteristic of market liquidity. Symbols are ranked from the most liquid to the less liquid.

We clearly observe a difference between futures, given by the symbols Dax and Eurostoxx, and the stocks represented by BNPP and Sanofi. This points towards considering this number as an indicator of liquidity and trading activity. Formula (2.39) suggests the ratio $\frac{\alpha}{\beta}$ as the key quantity to evaluate the impulse response value and the numbers are consistent with the fact that futures are more actively traded than the stocks due to a stronger branching structure (controlled by α and β).

A robustness check was also performed for other assets in Table 2.3. According to this measure a trade on the Bund triggers more other trades than does a trade on the Bobl. The Euro currency, and to a lesser extent the JPY, seem to be more reactive markets than the others. Among the commodities, the Crude Oil Brent dominates the Natural Gas, Sugar, Corn and the Wheat.

2.2.4 Forecasting

Once the model is calibrated, one is naturally interested in its forecasting power. In our case, as stressed in the preceding section, the model essentially captures the autocorrelation of the number of trades occurring in consecutive time intervals. The shape of this function, exponentially decreasing function to zero, naturally suggests to model the number of trades occurring in fixed time intervals as an autoregressive process.

Given a set of moments $t_0 < t_1 < t_2 < \ldots < t_N$ with $t_i - t_{i-1} = \tau$, we define $y_{t_i} = \frac{1}{\tau}(N_{t_i} - N_{t_{i-1}})$ and denote y_t with $t \in \{t_1, \ldots, t_N\}$. For simplicity, we work with the centered version of y_t and consider the autoregressive model:

$$y_t = a_1 y_{t-1} + u_t$$

where u_t is a random variable, independent of y_t with 0 mean. Then, in the case of a Hawkes model, the parameters of the AR model for the number of trades can be computed. Indeed, from the previous results we have

$$\mathbb{E}_{t-1}[y_t^2] = a_1 \mathbb{E}_{t-1}[y_t y_{t-1}] + \mathbb{E}_{t-1}[y_t u_t]$$

leading to

$$a_1 = \frac{\mathbb{E}_{t-1}[y_t y_{t-1}]}{\mathbb{E}_{t-1}[y_t^2]} = \frac{\mathrm{Cov}(\tau, 0)}{\mathrm{V}(\tau)} = \mathrm{Acf}(\tau, 0)$$

where the autocovariance and variance functions are known.

The same reasoning can be done if one considers an AR(p) model :

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + u_t$$

then for every $i \in \{1, .., p\}$, multiplying the above equation by y_{t-i} and taking the expectation yields p linear equations with coefficients a_i :

$$C(\tau, \delta = (i-1) \times \tau) = V(\tau) a_i + \sum_{j=1, j \neq i}^{p} a_j C(\tau, \delta = (|i-j| - 1) \times \tau)$$

therefore the $(a_1, \ldots, a_p)^\top$ is the solution of the linear system:

$$\begin{pmatrix} 1 & \mathrm{Acf}[\tau,0] & ... & ... & \mathrm{Acf}[\tau,(p-2)\tau] \\ \mathrm{Acf}[\tau,0] & 1 & \mathrm{Acf}[\tau,0] & & \mathrm{Acf}[\tau,(p-3)\tau] \\ ... & & & & \\ ... & & & & \\ \mathrm{Acf}[\tau,(p-2)\tau] & & & \mathrm{Acf}[\tau,0] & 1 \end{pmatrix} \times \begin{pmatrix} a_1 \\ ... \\ ... \\ ... \\ a_p \end{pmatrix} = \begin{pmatrix} \mathrm{Acf}[\tau,0] \\ ... \\ ... \\ ... \\ \mathrm{Acf}[\tau,(p-1)\tau] \end{pmatrix}. \tag{2.40}$$

Solving this equation for a sufficiently large first guess of p allows the calculation of the a_i coefficients and to decide the length of the autoregressive process by discarding the small coefficients.

We conduct the forecast experiment by daily calibrating the Hawkes process and use the parameters of each calibrated day as forecast parameters for the following day. In order to assess the quality of this prediction, we calculate the residues of our forecast and compare them to the residues of a forecast that relies on a Poisson process. Results are summarized in Table 2.5. They show a superior forecasting power for the Hawkes process.

Table 2.5: Hawkes vs Poisson Forecast Performance

		Symbol							
		Eurostoxx		Dax		BNPP		Sanofi	
Horizon	Measure	Hawkes	Poisson	Hawkes	Poisson	Hawkes	Poisson	Hawkes	Poisson
1	Mean	-0.0083	0.0523	-0.0207	0.0499	0.0087	0.0330	0.0084	0.0234
	Std. dev.	1.1836	1.4243	1.3865	1.7028	0.6252	0.6833	0.4031	0.4278
	Skew	-0.3341	-0.9205	-0.3237	-0.7918	-0.4882	-1.4757	-0.9462	-2.2286
5	Mean	0.0322	0.0498	0.0312	0.0489	0.0295	0.0330	0.0195	0.0214
	Std. dev.	1.3200	1.4262	1.5742	1.7028	0.6402	0.6812	0.4111	0.4301
	Skew	-0.9645	-0.9184	-0.8519	-0.7904	-1.4394	-1.4712	-2.1188	-2.2276
10	Mean	0.0442	0.0472	0.0467	0.0492	0.0358	0.0321	0.0205	0.0186
	Std. dev.	1.3742	1.4280	1.6382	1.7009	0.6632	0.6813	0.4275	0.4352
	Skew	-0.9915	-0.9152	-0.8625	-0.7884	-1.5231	-1.4626	-2.2051	-2.2064
30	Mean	0.0348	0.0300	0.0499	0.0448	0.0299	0.0225	0.0085	0.0046
	Std. dev.	1.4297	1.4331	1.6861	1.6896	0.6839	0.6853	0.4482	0.4493
	Skew	-0.8972	-0.8929	-0.7794	-0.7784	-1.4160	-1.4238	-2.1542	-2.1466

Note. For every symbol, mean, standard deviation and skew of the forecast residues are given for Hawkes process and Poisson process prediction method. The superiority of a Hawkes process forecast over a Poisson process forecast disappears when the forecast horizon increases.

Nevertheless, both models give exactly the same results when the forecast horizon is sufficiently large (half an hour in our case). As an illustration, notice that in Figure 2.7, the standard deviation of Hawkes forecast residues converges to that of the Poisson residues when the forecast horizon increases. Moreover, as illustrated by Figure 2.8, both distributions of residues converge when the forecast horizon is sufficiently large. This is an expected result as the Hawkes forecast method relies on the autocorrelation of the number of trades falling in consecutive intervals, which is significant for small time lags, but disappears after twenty to thirty minutes as previously shown in the paper.

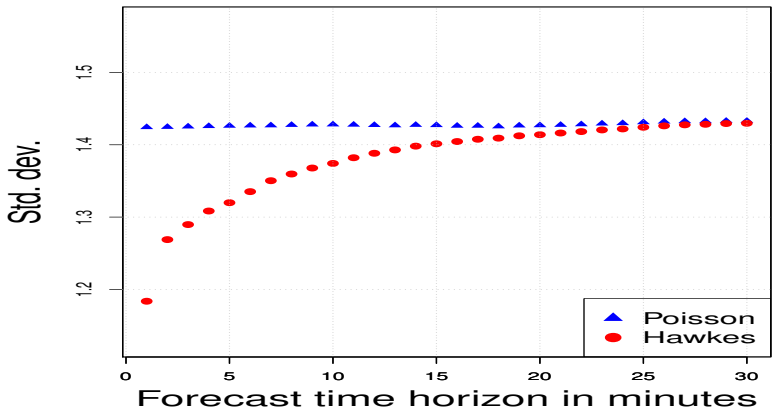

Figure 2.7: Standard deviation of residues of Hawkes and Poisson forecasts versus forecast time horizon.

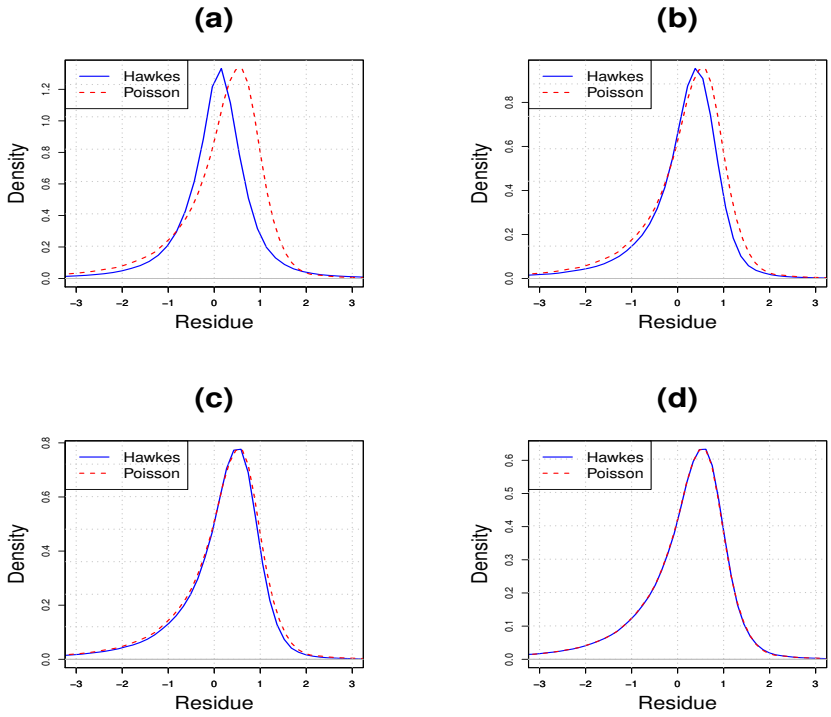

Figure 2.8: Densities of residues collapsing when forecast horizon increases. (a), (b), (c) and (d) correspond to forecast horizons of 1 minute, 5 minutes, 10 minutes and 30 minutes, respectively.

We also notice that the most accurate forecasts are obtained with a number of lags around 4 or 5 as is clear in Table 2.6 presenting the distribution of the number of lags necessary for the prediction. This is in accordance with the fact that the memory effect does not last for a very long time. Hence, the system (2.40) has to be solved for moderate values of p, which can be done very efficiently.

Table 2.6: Frequency Of Optimal Number Of Lags

Symbol	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
Eurostoxx	0.0020	0.0122	0.0446	0.3681	0.5691	0.0040
Dax	0	0.0020	0.0100	0.1896	0.7944	0.0040
BNPP	0	0.0123	0.1284	0.6403	0.2190	0
Sanofi	0.0021	0.0576	0.1725	0.6350	0.1328	0

*Note.*For every symbol, we present the frequency of the number of lags that were necessary to make the forecast in the AR(p) model of the number of trades occurring over a time interval. The system (2.40) is solved for a sufficiently large dimension. We consider p to be the rank from which the autoregressive coefficients become non significant. We considered a cut off absolute value of 0.001.

2.2.5 Diffusive limit and signature plot

In all the preceding sections we dealt with the trading process from a microscopic point of view, that is to say at transaction level. In the classical High-Frequency literature, mainly developed by econometricians, most of the studies are carried out at this time scale and aim at explaining the price formation process. Many models encompass the subtle interactions of the many components of the trading process such as order flow, order signs, volumes and other quantities to achieve a microscopic foundation of the price process. For instance, the ACD models, as proposed in (Engle and Russell, 1998), fall in this trend of research. For an overview of models and techniques we refer to (Hautsch, 2012) and references therein.

In this chapter, we consider a simpler framework focusing only on the order flow. Obviously, this simplification comes at the cost of neglecting important aspects of the price formation process, such as the volume for example, but it allows us to address the important question of connecting the microscopic price formation process observed at transaction level to its macroscopic properties at a coarser time scale. In other words, we connect the stochastic differential equations used to model an asset price evolution at a daily frequency, such as in the Black-Scholes model which relies mainly on the continuous Brownian motion, to the discontinuous point process describing individual transactions. The Hawkes process, thanks to its strong analytical tractability, enables us to relate these two time scales.

In recent years, many authors developed this bottom-up point of view in price modelling, establishing connections between order-book level price formation mechanisms and statistical macroscopic price properties. Among other references let us mention (Abergel and Jedidi, 2013a) whose model for the order book is based on a multidimensional Markov chain with independent Poissonian order arrival times. Within their framework they prove the convergence of the price process to a Brownian motion, thus bridging the gap between high frequency and low frequency quantities. They generalize this approach to Hawkes process arrivals in (Abergel and Jedidi, 2013b). In (Cont and De Larrard, 2011) and (Cont and De Larrard, 2012), the order book is described as a Markovian queuing system for which the authors establish a diffusive limit and calculate some quantities of interest such as volatility. In (Kirilenko et al., 2013), the authors use this same idea of different time scales and relate microscopic causes to macroscopic effects, they study the influence of high frequency traders on asset volatility.

Our work is greatly influenced by the seminal article of (Bacry et al., 2013a) in which the

authors introduce a model for microstructure price evolution based on mutually exciting Hawkes processes. They connect the signature plot of volatility and Epps effect of asset correlations to the model parameters driving the price process. Also of great interest is (Bacry et al., 2013b) where the authors establish diffusive limits for such kind of models.

In this subsection, we consider the modelling framework proposed by (Bacry et al., 2013a)[4] and develop a toy model for the movements of the *mid* price of a traded asset using the Hawkes processes presented in the analytical part. The *mid* price is the mean of the best ask price and the best bid price in the order book. As the best ask price or the best bid price moves up (down) by one tick the mid price will move up (down) by *half* a tick. Despite its simplicity, the model captures the essential features of the price process.[5]. The model writes:

$$S_t = S_0 + \left(N_t^{up} - N_t^{down}\right)\frac{\nu}{2} \tag{2.41}$$

where ν is the tick value. The N_t^{up} and N_t^{down} are Hawkes processes capturing the up and down jumps of the mid price. Both of them follow a dynamic of the form (2.4). We consider them independent and with the same parameters in order to avoid price explosion. In the stationary regime their intensities are given by:

$$\lambda_t^{up} = \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s^{up} \tag{2.42}$$

$$\lambda_t^{down} = \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s^{down}. \tag{2.43}$$

Later, we provide the specification of (Bacry et al., 2013a)'s model and explain the differences. In order to relate this high frequency description for the price to its low frequency description behaviour we need a limit theorem. We know thanks to (Bacry et al., 2013b) that the above specified model converges to a Brownian diffusion. Moreover, the authors rely on the martingale theory and limit theorems for semi-martingales to prove stability and convergence results for a general model with mutually exciting processes and a generic kernel[6].

With an exponential kernel, our case is simpler. We then sketch an alternative proof of the convergence result[7]. Moreover, the quantities calculated before will allow to establish easily the limit volatility of the diffusion.

With an exponential kernel for the considered Hawkes processes, the process $X_t = \left(N_t^{up}, \lambda_t^{up}, N_t^{down}, \lambda_t^{down}\right)$ is a Markov process. Its infinitesimal generator writes:

$$\begin{aligned}\mathcal{L}f(x) = {} & \beta\left(\lambda_\infty - \lambda_t^{up}\right)\frac{\partial f}{\partial \lambda^{up}}(x) + \beta\left(\lambda_\infty - \lambda_t^{down}\right)\frac{\partial f}{\partial \lambda^{down}}(x) \\ & + \lambda_t^{up}\left[f\left(N_t^{up}+1, \lambda_t^{up}+\alpha, N_t^{down}, \lambda_t^{down}\right) - f(x)\right] \\ & + \lambda_t^{down}\left[f\left(N_t^{up}, \lambda_t^{up}, N_t^{down}+1, \lambda_t^{down}+\alpha\right) - f(x)\right].\end{aligned}$$

[4]We might use the symbol "BDHM" which stands for Bacry, Delattre, Hoffmann, and Muzy.

[5]For instance, the so-called trade-throughs (Pomponio and Abergel, 2013) (i.e. trades consuming many successive limits and then moving the best quote by more than one tick) can be regarded in this model as successive one-tick movements occurring very closely in time.

[6]The function $g(t) = \alpha e^{-\beta t}$ is called the kernel of the Hawkes process. Other forms are possible but this choice leads to the most tractable Hawkes process.

[7]Recall that the result is known in full generality thanks to (Bacry et al., 2013b). We simply present other ideas that can be applied in the simple exponential kernel case, and that may shed some light on the result as we consider it from another perspective.

The explicit form of the infinitesimal generator allows us to apply Foster-Lyapounov techniques in order to establish stability results. We refer to (Meyn and Tweedie, 2009) for a detailed exposition on such techniques as well as stochastic stability concepts. For instance, ergodicity of the process X_t, that is to say its convergence to a stationary regime, can be easily established thanks to the Foster-Lyapounov test function criterion. In our case, we define the function $V(x) = \frac{\lambda^{up}+\lambda^{down}}{2\lambda_\infty}$, then a simple calculation yields the *geometric drift condition*:

$$\mathcal{L}V(x) \leq (\alpha - \beta)V(x) + \beta. \tag{2.44}$$

As $\alpha < \beta$ this condition ensures the mean reversion of the process and thanks to Theorems 6.1 and 7.1 in (Meyn and Tweedie, 1993) (and especially (CD3)), the V-uniform ergodicity of the process X_t.

Let us then write unit-time price increments:

$$\eta_i = \left[(N_i^{up} - N_{i-1}^{up}) - (N_i^{down} - N_{i-1}^{down})\right] \times \frac{\nu}{2}$$

and consider the random sums:

$$S_n = \sum_{i=1}^{n} \eta_i$$

with $\{\eta_i; i = 1 \dots n\}$ being the price increments (note that $\mathbb{E}[\eta_i] = 0$). We focus on the asymptotic behaviour of the rescaled price process:

$$\bar{S_t^n} = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}}.$$

The V-uniform ergodicity and Theorem 16.1.5 in (Meyn and Tweedie, 2009) allows us to conclude that the increments are geometrically mixing and Theorem 19.3 of (Billingsley, 1999) proves that $\bar{S_t^n}$ converges to a Brownian motion in the sense of Skorokhod topology:

$$\bar{S_t^n} \Rightarrow \sigma W_t.$$

The same proof technique was used in (Abergel and Jedidi, 2013a) with a more feature-rich model. The underlying problem is the same: one considers the rescaled sums of random variables and wants to apply the Functional Central Limit Theorem, known as Donsker's theorem that establishes the convergence of the rescaled Random Walk to the Brownian Motion. This theorem cannot be used here as the increments are correlated, (for the Random Walk the increments are independent). However, it can be extended to the case of "weak dependence". The "degree" of dependence has to be controlled in order for the convergence to hold. A manner to control this degree of dependence is to show that the random variables are geometrically mixing. The relationships between geometric mixing and V-uniform ergodicity is thoroughly treated in (Meyn and Tweedie, 2009).

Moreover, calculations done before for the moments of the Hawkes process increments lead to a very simple expression for the volatility. In fact, we have:

$$\begin{aligned}
\sigma^2 &= \lim_{n\to\infty} \frac{\mathrm{Var}(S_n)}{n} \\
&= 2\frac{\nu^2}{4}\left(\mathbb{E}[(N_1^{up} - N_0^{up})^2] - \mathbb{E}[N_1^{up} - N_0^{up}]^2\right) \\
&+ 4\frac{\nu^2}{4}\sum_{i=1}^{\infty} \mathbb{E}[(N_1^{up} - N_0^{up})(N_{1+i}^{up} - N_i^{up})] - \mathbb{E}[N_1^{up} - N_0^{up}]\mathbb{E}[N_{1+i}^{up} - N_i^{up}] \\
&= \frac{\nu^2}{2}\left(V(1) + 2\sum_{k=0}^{\infty} \mathrm{Cov}(1,k)\right)
\end{aligned}$$

where thanks to (2.24) $\mathrm{Cov}(1,k)$ writes:

$$\mathrm{Cov}(1,k) = \frac{\lambda_\infty \beta\alpha(2\beta-\alpha)\left(e^{(\alpha-\beta)}-1\right)^2}{2(\alpha-\beta)^4} e^{(\alpha-\beta)k}.$$

Then, summing up with the expression of the variance and after some simplifications we obtain:

Proposition 2.5. *The asymptotic volatility of the toy model 2.41 writes:*

$$\sigma^2 = \frac{\nu^2}{2}\frac{\lambda_\infty}{\left(1-\frac{\alpha}{\beta}\right)^3}. \tag{2.45}$$

Notice that the stationarity and ergodicity of the increments η_i, together with the convergence of the series $\sigma^2 = \mathbb{E}[\eta_0^2] + 2\sum_{n=0}^{\infty}\mathbb{E}[\eta_0\eta_n]$ established in the above proposition are sufficient to conclude that $\bar{S}_t^n$ converges to a Brownian motion in the sense of Skorokhod topology thanks to theorem 19.1 in (Billingsley, 1999).

Note the dependence of the asymptotic volatility on the ratio $\frac{\alpha}{\beta}$. The larger this ratio is, the larger the volatility is (under the hypothesis of stability $\alpha < \beta$). An upward (downward) shock is more likely to trigger another upward (downward) chock if this ratio is large and therefore it induces a positive autocorrelation for the mid price and a more persistent path with the effect of increasing asset's volatility.

Besides giving a framework that allows the connection of the microscopic price formation mechanism to its macroscopic behaviour, as shown above, the Hawkes process can reproduce some stylized facts across time scales. Among these stylized facts is the volatility signature plot which depends on the realized variance over a period T calculated by sampling the data by time intervals of length τ. Within the toy model (2.41) we have:

$$\begin{aligned}
\hat{C}(\tau) &= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(S_{(n+1)\tau}-S_{n\tau}\right)^2 \\
&= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(\left(N^{up}_{(n+1)\tau}-N^{up}_{n\tau}\right)-\left(N^{down}_{(n+1)\tau}-N^{down}_{n\tau}\right)\right)^2\frac{\nu^2}{4} \\
&= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(N^{up}_{(n+1)\tau}-N^{up}_{n\tau}\right)^2\frac{\nu^2}{4} + \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(N^{down}_{(n+1)\tau}-N^{down}_{n\tau}\right)^2\frac{\nu^2}{4} \\
&- 2\frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(N^{up}_{(n+1)\tau}-N^{up}_{n\tau}\right)\left(N^{down}_{(n+1)\tau}-N^{down}_{n\tau}\right)\frac{\nu^2}{4}.
\end{aligned}$$

By definition, the mean signature plot is the expectation of the above quantity and can be computed explicitly as we have (we omit the proof as it is straightforward):

Proposition 2.6. *The mean signature plot, defined by $C(\tau) = \mathbb{E}[\hat{C}(\tau)]$, is given by:*

$$\begin{aligned}
C(\tau) &= \mathbb{E}[\hat{C}(\tau)] \\
&= \frac{\nu^2}{2\tau}V(\tau) \\
&= \frac{\nu^2}{2}\Lambda\left(\kappa_-^2 + \left(1-\kappa_-^2\right)\frac{\left(1-e^{-\tau\gamma_-}\right)}{\tau\gamma_-}\right)
\end{aligned}$$

where:

$$\Lambda = \frac{\lambda_\infty}{1-\alpha/\beta}\ ,\ \kappa_- = \frac{1}{1-\alpha/\beta}\ \textit{and}\ \gamma_- = \beta-\alpha.$$

We eventually refer to the signature plot instead of the mean signature plot. Notice that when τ becomes larger the above expression converges to the asymptotic diffusive variance of the model calculated in (2.45). The mean signature plot is an increasing function with respect to τ (or equivalently the signature plot is decreasing with the sampling frequency) and this is due to the positive serial autocorrelation of the returns. This captures situations as the one observed in Figure 2.9.

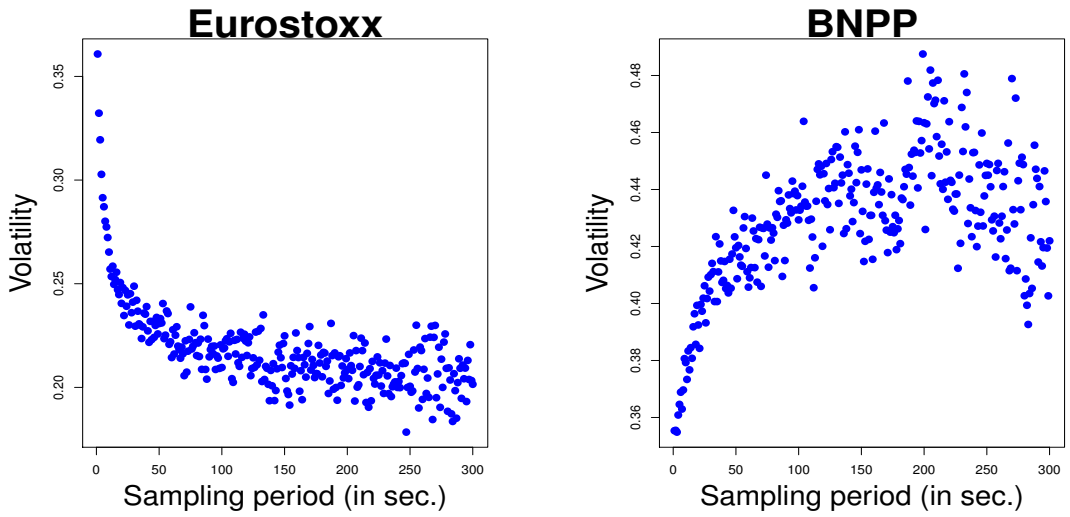

Figure 2.9: Signature Plot for Eurostoxx futures and BNPP stock on 2011/04/01, computed on mid prices to eliminate bid-ask bounce.

Lastly, within our simple toy model we can determine the autocorrelation function of the price increments computed over intervals of size τ and lagged by δ, it is given by:

$$CorrStock(\tau,\delta) = \frac{\mathbb{E}\left[\left(S_{t+\tau}-S_t\right)\left(S_{t+2\tau+\delta}-S_{t+\tau+\delta}\right)\right]}{\sqrt{\mathbb{E}\left[\left(S_{t+\tau}-S_t\right)\right]\mathbb{E}\left[\left(S_{t+2\tau+\delta}-S_{t+\tau+\delta}\right)\right]}} \tag{2.46}$$

which thanks to the closed-form formula obtained in the analytical section leads to:

$$\frac{e^{-2\beta\tau+\delta(\alpha-\beta)}\left(e^{\tau\beta}-e^{\tau\alpha}\right)^2\alpha\left(2\beta-\alpha\right)}{2\beta^2\left(\beta-\alpha\right)}. \tag{2.47}$$

The stability condition (i.e. $\alpha < \beta$) ensures the positiveness of this serial autocorrelation and it confirms the intuition developed above. Moreover, the decay of the autocorrelation as a function of the lag depends on the parameter α which tends to reduce the decaying behaviour of this function.

In order to compare our results to (Bacry et al., 2013a) let us briefly recapitulate their main findings. In (Bacry et al., 2013a), the authors propose a model similar to (2.41) but with Hawkes processes that are *mutually-excited* and not *self-excited* as in our case. To be more precise the dynamics for the intensities are given by:

$$\lambda_t^{up} = \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s^{down} \tag{2.48}$$

$$\lambda_t^{down} = \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s^{up}. \tag{2.49}$$

Notice that an up jump increases the down intensity that increases the probability of a down jump and if this one occurs it will increase the up jump intensity. The process is purely mutually excited and possesses a mean reversion behaviour. These dynamics are significantly different from those driving the toy model. The diffusive limit for this model is:

$$\sigma^2_{\text{BDHM}} = \frac{\nu^2}{2} \frac{\lambda_\infty}{\left(1 - \frac{\alpha}{\beta}\right)\left(1 + \frac{\alpha}{\beta}\right)^2} \tag{2.50}$$

and the mean signature plot is:

$$C(\tau) = \frac{\nu^2}{2} \Lambda \left(\kappa_+^2 + (1 - \kappa_+^2) \frac{1 - e^{-\tau\gamma_+}}{\tau\gamma_+} \right)$$

with:

$$\Lambda = \frac{\lambda_\infty}{1 - \alpha/\beta} \ , \ \kappa_+ = \frac{1}{1 + \alpha/\beta} \text{ and } \gamma_+ = \alpha + \beta.$$

In (Bacry et al., 2013a)'s model an upward shock will increase the down intensity and trigger a downward shock on the mid price, thereby leading to a mean reverting behaviour for the mid price. Therefore, as a function of the sampling period the signature plot is decreasing with respect to τ (or equivalently the signature plot is increasing with the sampling frequency) because of this negative serial autocorrelation of the returns. In (Andersen et al., 1999) the authors define the signature plot as "the patterns of bias injected in realized volatility as underlying returns are sampled progressively more frequently". This bias can lead to a decreasing or increasing volatility as a function of the sampling period. The decreasing pattern is the most frequent but inverted pattern is also possible and can be found in data, even for liquid stocks, as illustrated in Figure 2.9. BDHM model is compatible with a decreasing pattern whilst the toy model is compatible with an increasing pattern[8]. Due to the positive (negative) autocorrelation of the returns in the toy (BDHM) model we have, for a given pair (α, β), the inequality $\sigma > \sigma_{\text{BDHM}}$.

Within the specification (2.48) and (2.49) the autocorrelation function (2.46) can be computed and is given by:

$$-\frac{e^{-(\delta+2\tau)(\beta+\alpha)} \left(e^{\tau(\beta+\alpha)} - 1\right)^2 \alpha \left(2\beta + \alpha\right)}{2\beta^2 \left(\beta + \alpha\right)}. \tag{2.51}$$

This quantity is negative and this confirms the intuition. Also, the decay of the autocorrelation as a function of the lag depends on parameter α, now controlling the mutual excitation property of the process, and this parameter increases the decaying behaviour of the function.

Notice that both models produce different asymptotic volatility formulas. In order to assess their plausibility, we calibrate a Hawkes process to the mid price up-jumps and calculate the asymptotic volatilities for the two models; the toy model given by (2.45) and (Bacry et al., 2013a)'s specification (2.50). For the toy model the parameter estimation is performed as in the previous subsection. We use only the autocorrelation function with $\tau = 60$ seconds and δ ranging from 0 to 600 seconds by step of 60 seconds and λ_∞ is deduced thanks to (2.22) with $\tau = 60$ seconds (it corresponds to the "Fast calibration II"). For (Bacry et al., 2013a)'s model we use the MLE algorithm[9].

[8] Let us stress that these figures are based on mid price sampling so that any noise due to bid-ask bounce is eliminated. Also, whether the signature plot is an increasing or decreasing function of the sampling frequency depends on the market conditions. Some liquid stocks (or indexes) can display both, BNPP is an example, and understanding the determinants of this functional dependency remains an open question.

[9] For this model we only report the volatility value σ_{BDHM}.

Results are reported in Table 3.2 (apart from the asymptotic volatilities we only report the estimated parameters for the toy model). We rescaled the obtained volatilities by the spot value in order to obtain the more usual Black-Scholes volatility corresponding to a lognormal model. One can clearly see that our model systematically overestimates volatility whereas (Bacry et al., 2013a)'s model systematically underestimates it. Qualitatively, this is because our model magnifies up and down movements, thus inflating the realized volatility, whereas (Bacry et al., 2013a)'s model moderates up and down movements thanks to its mean reversion mechanism, thus underestimating the volatility.

Table 2.7: Asymptotic Volatility Values

Symbol	λ_∞	α	β	σ_{BDHM}	Empirical σ	Toy model σ
Eurostoxx	0.0184	0.0160	0.0219	12.79%	40.77%	80.64%
Dax	0.0429	0.0226	0.0259	4.80%	25.75%	71.36%
BNPP	0.0379	0.0481	0.0569	5.36%	41.86%	72.24%
Sanofi	0.0279	0.0488	0.0587	4.00%	30.30%	45.37%
Bund	0.0267	0.0180	0.0261	2.70%	8.25%	14.47%
Bobl	0.0228	0.0187	0.0288	2.45%	6.92%	11.19%
Schatz	0.0257	0.0223	0.0372	1.38%	2.63%	5.39%
JPY	0.0313	0.0659	0.0764	1.98%	9.89%	29.88%
EURO	0.0474	0.0648	0.0725	4.87%	16.76%	112.25%
GOLD	0.0728	0.0775	0.0868	3.77%	24.48%	71.64%
Crude Oil Brent	0.0474	0.0472	0.0528	6.22%	41.25%	126.29%
Natural GAS	0.0548	0.0931	0.1090	11.24%	58.65%	150.30%
Sugar	0.0410	0.0556	0.0758	10.58%	51.37%	73.30%
CORN	0.0419	0.0552	0.0694	9.34%	43.73%	75.94%
WHEAT	0.0451	0.0626	0.0763	10.59%	57.17%	94.53%

Note. Values for the asymptotic volatilities as given by our toy model (2.45) and the (Bacry et al., 2013a) model (2.50) as well as the realized volatility of the day. We also put median estimated values for the Hawkes processes in the toy model (2.41), (2.42) and (2.43). Both models give plausible values. σ_{BDHM} volatility underestimates systematically the realized volatility, whereas our toy model systematically overestimates it.

Our results suggest a more general specification allowing for both self and mutual excitations. The particularly simple approach we adopted in our calculations based on the infinitesimal generator and Dynkin's formula may be generalized to a multidimensional setting and may lead to tractable results. It would allow for both effects, that were underlined in the empirical literature (see (Hautsch, 2012)), on the price dynamic and provide a deeper understanding of how the volatility measured at a macroscopic level (at a daily or low frequency) depends on the trading activity observed at a microscopic level (at high frequency).

2.3 Conclusion

In this chapter we explicitly compute the moments and the autocorrelation function of the number of jumps over an interval for the Hawkes process. Using these quantities we develop a method of moments estimation strategy which is extremely fast compared with the usual maximum likelihood estimation strategy. This aspect is essential as we are interested in the trade clustering activity observed in high frequency data or if we wish to apply the model in real time. We use our estimation framework to calibrate the Hawkes process on trades for four stocks over a two-year sample. The Hawkes process can cope with the trade clustering effect thanks to its autocorrelation structure. As our calibration is fast we roll the daily estimation over two years to analyze the parameters stability, and they are found

to be reasonably stable. We perform a robustness check on other assets and obtain similar results.

Thanks to the analytical tractability of the Hawkes process we explicitly compute the impulse response associated with the process which determines the market impact of a trade. We use this as a measure of liquidity and the estimated parameters lead to reasonable conclusions.

Then, a forecast experiment allows to assess the superiority to the Hawkes framework with respect to a simpler model.

Lastly, within a simple model based on the Hawkes process we explicitly compute the diffusive limit for the price process. This allows us to connect the microscopic dynamic, that is to say the high frequency dynamic, to the macroscopic dynamic, the volatility computed at a daily frequency (with the Black-Scholes volatility being the most well-known quantity).

Our work points towards several extensions. First, we computed the diffusive limit under the restrictive hypothesis that the Hawkes processes are only self excited whereas in (Bacry et al., 2013a), on which we heavily rely, the Hawkes processes are only mutually excited. The reality should lie between the two and requires Hawkes processes that are both self and mutually excited. To this end we would need to perform the computations in the multidimensional case. The next chapter will deal with this more general problem.

2.4 Appendix

Computing the moments by derivation of the moment-generating function

Let $u = (u_1, u_2)^\top \in \mathbb{R}^2$, the conditional moment-generating function of $X_T = (\lambda_T, N_T)$ is defined as $f(t, X_t) = \mathbb{E}_t^x \left[e^{u^\top X_T} \right] = \mathbb{E}_t^x \left[e^{u_1 \lambda_T + u_2 N_T} \right]$. Clearly, $f(t, X_t)$ must be a martingale and the function f satisfies:

$$\frac{\partial f}{\partial t}(t, X_t) + \mathcal{L} f(t, X_t) = 0 \tag{2.52}$$

with boundary condition $f(T, X_T) = e^{u^\top X_T}$. As $X_t = (\lambda_t, N_t)$ is a Markov affine point process we guess the solution of (2.52) is an exponential affine form of the state variable, that is to say:

$$f(t, X_t) = e^{a(t) + b(t)\lambda_t + c(t) N_t}. \tag{2.53}$$

Setting this guess into equation (2.52) we obtain the system of ordinary differential equations:

$$\frac{\partial a}{\partial t} = -\beta \lambda_\infty b(t) \tag{2.54}$$

$$\frac{\partial b}{\partial t} = \beta b(t) + 1 - e^{\alpha b(t) + c(t)} \tag{2.55}$$

$$\frac{\partial c}{\partial t} = 0 \tag{2.56}$$

with terminal conditions $a(T) = 0$, $b(T) = u_1$ and $c(T) = u_2$.

Choosing $u_1 = 0$ in our Laplace transform, the moment-generating function for N_t which writes if we take $u_2 = u$:

$$\mathbb{E}_t^x \left[e^{u N_T} \right] = e^{a(t) + b(t)\lambda_t + u N_t}.$$

Expressed in terms of $\tau = T - t$ the expectation and the system of ordinary differential equations are given by:

$$\mathbb{E}_t^x \left[e^{u(N_{t+\tau} - N_t)} \right] = e^{a(\tau) + \lambda_t b(\tau)} \tag{2.57}$$

with:

$$\frac{\partial a}{\partial \tau} = \beta \lambda_\infty b(\tau), \tag{2.58}$$

$$\frac{\partial b}{\partial \tau} = -\beta b(\tau) - 1 + e^{\alpha b(\tau) + u} \tag{2.59}$$

and $a(0) = b(0) = 0$.

Expected number of jumps

We start by the expression of the Laplace transform:

$$\mathbb{E} \left[e^{u(N_{t+\tau} - N_t)} | \mathcal{F}_t \right] = e^{a(\tau) + \lambda_t b(\tau)} \tag{2.60}$$

Taking the derivatives at $u = 0$ yields :

$$\begin{aligned} \mathbb{E}\left[N_{t+\tau} - N_t | \mathcal{F}_t \right] &= \frac{\partial}{\partial u} e^{a(\tau) + \lambda_t b(\tau)} |_{u=0} \\ &= \left(a_u(\tau) + \lambda_t b_u(\tau) \right) e^{a(\tau) + \lambda_t b(\tau)} |_{u=0}, \end{aligned} \tag{2.61}$$

where $b_u = \partial_u b$ and $a_u = \partial_u a$, and with initial condition $a_u(0) = b_u(0) = 0$.

Let us now calculate the derivatives of (2.58) and (2.59) with respect to u taken in $u = 0$.

$$\partial_\tau b_u|_{u=0} = -\beta b_u(\tau) + (\alpha b_u(\tau) + 1)\, e^{\alpha b(\tau)}|_{u=0} \tag{2.62}$$
$$\partial_\tau a_u|_{u=0} = \beta\lambda_\infty b_u(\tau)|_{u=0} \tag{2.63}$$

For $u = 0$, one solution to the system of equations (2.58) and (2.59) is $a(\tau) = b(\tau) = 0$ and it turns out to be unique. Putting that in the preceding equations gives :

$$\partial_\tau b_u = -(\beta - \alpha)\, b_u(\tau) + 1 \tag{2.64}$$
$$\partial_\tau b_u = \beta\lambda_\infty b_u(\tau) \tag{2.65}$$

So finally, for $u = 0$:

$$b_u(\tau) = \frac{-1}{\beta - \alpha}\left(e^{-(\beta-\alpha)\tau} - 1\right) \tag{2.66}$$
$$a_u(\tau) = \frac{\beta\lambda_\infty}{(\beta - \alpha)^2}\left(e^{-(\beta-\alpha)\tau} - 1\right) + \frac{\beta\lambda_\infty}{\beta - \alpha}\tau \tag{2.67}$$

Recalling from (2.61) that :

$$\mathbb{E}\left[N_{t+\tau} - N_t|\mathcal{F}_t\right] = a_u(\tau) + \lambda_t b_u(\tau) \tag{2.68}$$

we end with :

$$\mathbb{E}\left[N_{t+\tau} - N_t|\mathcal{F}_t\right] = \frac{\beta\lambda_\infty}{(\beta - \alpha)^2}\left(e^{-(\beta-\alpha)\tau} - 1\right) + \frac{\beta\lambda_\infty}{\beta - \alpha}\tau - \frac{\lambda_t}{\beta - \alpha}\left(e^{-(\beta-\alpha)\tau} - 1\right). \tag{2.69}$$

Notice that the above conditional expectation still depends on the unobservable variable λ_t and is therefore unusable in practice as it is. To circumvent this dependence on a latent variable, we further take the expectation of the above expression. This will remove the conditioning at the left hand side of the equation, and replace λ_t on the right hand side by its previously calculated expectation.

Finally, letting $t \to \infty$ gives the stationary regime value of the expected number of jumps during an interval of length τ :

$$\mathbb{E}\left[N_{t+\tau} - N_t\right] = \frac{\beta\lambda_\infty}{\beta - \alpha}\tau, \tag{2.70}$$

Notice that for $\alpha = 0$ we have $\mathbb{E}\left[N_\tau\right] = \lambda_\infty\tau$ as expected.

Variance of number of jumps over a time interval

The same method as previously allows to calculate the second moment of the number of jumps during a time interval of length τ, namely $\mathbb{E}[(N_{t+\tau} - N_t)^2|\mathcal{F}_t]$.

Taking the second derivative of (2.57) at $u = 0$, we get

$$\mathbb{E}\left[(N_{t+\tau} - N_t)^2|\mathcal{F}_t\right] = \left[(a_u(\tau) + \lambda_t b_u(\tau))^2 + a_{u^2}(\tau) + \lambda_t b_{u^2}(\tau)\right] e^{a(\tau)+\lambda_t b(\tau)}|_{u=0}, \tag{2.71}$$

where $b_{u^2} = \partial_{u^2} b$ and $a_{u^2} = \partial_{u^2} a$.

Recalling that for $u = 0$, $a = 0$ and $b = 0$ are the unique solutions of equations (2.59) and (2.58), we have :

$$\mathbb{E}\left[(N_{t+\tau} - N_t)^2|\mathcal{F}_t\right] = \left[(a_u(\tau) + \lambda_t b_u(\tau))^2 + a_{u^2}(\tau) + \lambda_t b_{u^2}(\tau)\right]|_{u=0}, \tag{2.72}$$

where the only unknowns are b_{u^2} and a_{u^2}.

Then, taking derivatives of equations (2.59) and (2.58) we get the EDO's for b_{u^2} and a_{u^2} evaluated at $u = 0$

$$\partial_\tau b_{u^2} = -(\beta - \alpha) b_{u^2} + (\alpha b_u + 1)^2 \tag{2.73}$$
$$\partial_\tau a_{u^2} = \beta \lambda_\infty b_{u^2} \tag{2.74}$$

with boundary condition $b_{u^2}(0) = a_{u^2}(0) = 0$.

Recalling that b_u was previously calculated in equation (2.66), the first equation gives

$$b_{u^2}(\tau) = \int_0^\tau e^{-(\beta-\alpha)(\tau-s)} (\alpha b_u(s) + 1)^2 ds \tag{2.75}$$
$$= \int_0^\tau e^{-(\beta-\alpha)(\tau-s)} \left(\alpha \frac{1 - e^{-(\beta-\alpha)s}}{\beta - \alpha} + 1 \right)^2 ds \tag{2.76}$$
$$= -\frac{-e^{2\tau(\alpha-\beta)}\alpha^2 + \beta^2 + e^{\tau(\alpha-\beta)}(\alpha - \beta)(\alpha + \beta + 2\tau\alpha\beta)}{(\alpha - \beta)^3} \tag{2.77}$$

A direct integration gives $a_{u^2}(\tau)$

$$a_u^2 = \frac{\lambda_\infty \beta \left(\alpha^2 + e^{2\tau(\alpha-\beta)}\alpha^2 - 4\alpha\beta - 2\beta^2 - 2\tau\alpha\beta^2 + 2\tau\beta^3 + 2e^{\tau(\alpha-\beta)} \left(\beta^2 + 2\alpha\beta(1 + \tau\beta) - \alpha^2(1 + 2\tau\beta) \right) \right)}{2(\alpha - \beta)^4} \tag{2.78}$$

Putting all that together in equation (2.72) gives the expression of the conditional second moment.

As noticed in the case of the first moment, the obtained expression still depends on the unobservable variable λ_t. Getting the expected value of the conditional second moment allows to get rid of this dependence, as λ_t is then replaced by its expectation which has been previously calculated in the text.

That done, calculation of the variance of the number of jumps during a time interval is straightforward by application of the formula :

$$V = \mathbb{E}\left[(N_{t+\tau} - N_t)^2 \right] - \mathbb{E}\left[(N_{t+\tau} - N_t) \right]^2 \tag{2.79}$$

As before, to obtain the long run, stationary regime value of this variance, we let $t \to \infty$ in the so obtained expression, ending with :

$$V = \Lambda \left(\tau \kappa_-^2 + \left(1 - \kappa_-^2 \right) \frac{\left(1 - e^{-\tau\gamma_-} \right)}{\gamma_-} \right) \tag{2.80}$$

where:

$$\Lambda = \frac{\lambda_\infty}{1 - \alpha/\beta} \ , \ \kappa_- = \frac{1}{1 - \alpha/\beta} \text{ and } \gamma_- = \beta - \alpha.$$

Autocorrelation function

Our aim in this paragraph is to calculate the autocorrelation function of the number of jumps during different time intervals. The key quantity to be calculated here is the uncentered covariation : $\mathbb{E}[(N_{t_1} - N_t)(N_{t_3} - N_{t_2})|\mathcal{F}_t]$, where $t < t_1 < t_2 < t_3$.

In order to simplify notations for the sequel, consider the variables $\Delta_1 = t_1 - t$, $\Delta_2 = t_3 - t_2$ and $\delta = t_2 - t_1$.

The main idea for this calculation, is to start from the expression :

$$\mathbb{E}[e^{u(N_{t_1}-N_t)+v(N_{t_3}-N_{t_2})}|\mathcal{F}_t]. \tag{2.81}$$

It is then clear that taking the mixed derivative of the above function of the dummy variables u and v with respect to both of them in the point $u = v = 0$ yields the desired covariation.

The calculation will be conducted by successive conditioning with respect to t_2, t_1 and finally t.

To begin with, calculate :

$$\mathbb{E}_{t_2}[e^{v(N_{t_3}-N_{t_2})}] = e^{a(\Delta_2)+b(\Delta_2)\lambda_{t_2}}, \tag{2.82}$$

where

$$\begin{aligned} \partial_\tau b &= -\beta b - 1 + e^{\alpha b+v} \\ \partial_\tau a &= \beta\lambda_\infty b \\ b(0) &= 0 \\ a(0) &= 0, \end{aligned} \tag{2.83}$$

thanks to the study made in the preceding paragraphs.

Then, consider the Laplace transform of λ_{t_2} conditionally to $\mathcal{F}_{t_1}$. The same reasoning as precedently, relying on the affine structure of the Markov process $X_t = (\lambda_t, N_t)$, gives the following expression for the Laplace transform of λ_{t_2} :

$$\mathbb{E}_{t_1}[e^{b(\Delta_2)\lambda_{t_2}}] = e^{a_1(\delta)+b_1(\delta)\lambda_{t_1}}, \tag{2.84}$$

where

$$\begin{aligned} \partial_\tau b_1 &= -\beta b_1 - 1 + e^{\alpha b_1} \\ \partial_\tau a_1 &= \beta\lambda_\infty b_1 \\ b_1(0) &= b(\Delta_2) \\ a_1(0) &= 0, \end{aligned} \tag{2.85}$$

and finally

$$\mathbb{E}_t[e^{u(N_{t_1}-N_t)+b_1(\delta)\lambda_{t_1}}] = e^{a_2(\Delta_1)+b_2(\Delta_1)\lambda_t} \tag{2.86}$$

where

$$\begin{aligned} \partial_\tau b_2 &= -\beta b_2 - 1 + e^{\alpha b_2+u} \\ \partial_\tau a_2 &= \beta\lambda_\infty b_2 \\ b_2(0) &= b_1(\delta) \\ a_2(0) &= 0. \end{aligned} \tag{2.87}$$

The final solution is then

$$\mathbb{E}[(N_{t_1} - N_t)(N_{t_3} - N_{t_2})|\mathcal{F}_t] = e^{a(\Delta_2)+a_1(\delta)+a_2(\Delta_1)+b_2(\Delta_1)\lambda_t}. \tag{2.88}$$

Now, take the derivatives with respect to u then v at the point $u = v = 0$.

It is clear that $\partial_u a = \partial_u b = \partial_u a_1 = \partial_u b_1 = 0$, so we have to calculate

$$(\partial_{uv} a_2(\Delta_1)+\partial_{uv} b_2(\Delta_1)\lambda_t)+(\partial_u a_2(\Delta_1)+\partial_u b_2(\Delta_1)\lambda_t)(\partial_v a(\Delta_2)+\partial_v a_1(\delta)+\partial_v a_2(\Delta_1)+\partial_v b_2(\Delta_1)\lambda_t) \tag{2.89}$$

As was done in the preceding sections, these derivatives with respect to u and v will be calculated simply by taking the derivatives of the preceding systems of ODE's with respect to those variables. Indeed, taking the derivatives of the system (2.83) with respect to v yields :

$$\begin{aligned} \partial_\tau b_v &= -\beta b_v + (\alpha b_v + 1)e^{\alpha b} \\ \partial_\tau a_v &= \beta\lambda_\infty b_v \\ b_v(0) &= 0 \\ a_v(0) &= 0. \end{aligned} \tag{2.90}$$

Knowing that for $v = 0$ we have $b = 0$ that is the unique solution of the system (2.83), we conclude that $a = 0$ and $b_v(\tau) = \int_0^\tau e^{(\alpha-\beta)(\tau-s)}ds = I_1(\tau)$ so $a_v(\tau) = \beta\lambda_\infty \int_0^\tau I_1(s)ds = \beta\lambda_\infty I_2(\tau)$.

Then, take the derivatives of the system (2.85) with respect to v, knowing that for $v = 0$ we have $b = 0$ and $b_1 = 0$ that are the unique solutions of, respectively, systems (2.83) and (2.85):

$$\begin{aligned} \partial_\tau b_{1v} &= (\alpha - \beta)b_{1v} \\ \partial_\tau a_{1v} &= \beta\lambda_\infty b_{1v} \\ b_{1v}(0) &= b_v(\Delta_2) \\ a_{1v}(0) &= 0, \end{aligned} \tag{2.91}$$

therefore $b_{1v}(\tau) = b_v(\Delta_2)e^{(\alpha-\beta)\tau}$ et $a_{1v}(\tau) = \beta\lambda_\infty b_v(\Delta_2)\int_0^\tau e^{(\alpha-\beta)s} = \beta\lambda_\infty b_v(\Delta_2)I_0(\tau)$.

Then, taking further the derivative of the system (2.91) with respect to v yields :

$$\begin{aligned} \partial_\tau b_{2v} &= (\alpha - \beta)b_{2v} \\ \partial_\tau a_{2v} &= \beta\lambda_\infty b_{2v} \\ b_{2v}(0) &= b_{1v}(\delta) \\ a_{2v}(0) &= 0. \end{aligned} \tag{2.92}$$

Consequently $b_{2v}(\tau) = b_{1v}(\delta)e^{(\alpha-\beta)\tau}$ et $a_{2v}(\tau) = b_{1v}(\delta)I_0(\tau)$.

Additionally, second derivative of system (2.87), with the now usual trick that for $u = v = 0$ we have $b_1 = 0$ that is the unique solution of (2.85) and then $b_2 = 0$ is the unique solution of (2.87), we have :

$$\begin{aligned} \partial_\tau b_{2u} &= (\alpha - \beta)b_{2u} + 1 \\ \partial_\tau a_{2u} &= \beta\lambda_\infty b_{2u} \\ b_{2u}(0) &= 0 \\ a_{2u}(0) &= 0, \end{aligned} \tag{2.93}$$

thus $b_{2u}(\tau) = I_1(\tau)$ et $a_{2u}(\tau) = \beta\lambda_\infty I_2(\tau)$.

Now, taking the mixed derivative of (2.87) with respect to u and v in the point $u = v = 0$ yields :

$$\begin{aligned}
\partial_\tau b_{2uv} &= (\alpha-\beta)b_{2uv} + (\alpha b_{2u}+1)(\alpha b_{2v}+1) \\
\partial_\tau a_{2uv} &= \beta\lambda_\infty b_{2uv} \\
b_{2uv}(0) &= 0 \\
a_{2uv}(0) &= 0,
\end{aligned} \tag{2.94}$$

which solutions are :

$$\begin{aligned}
b_{2uv}(\tau) &= \int_0^\tau e^{(\alpha-\beta)(\tau-s)}(\alpha b_{2u}(s)+1)(\alpha b_{2v}(s)+1)ds \\
a_{2uv}(\tau) &= \beta\lambda_\infty \int_0^\tau b_{2uv}(s)ds
\end{aligned}$$

Now, putting all the calculation results together, we are able to calculate the conditional uncentered covariance of the number of jumps during time intervals of length $\Delta_1 = \Delta_2 = \tau$ with a time lag δ.

$$\mathbb{E}[(N_{t+\tau} - N_t)(N_{t+2\tau+\delta} - N_{t+2\tau+\delta})|\mathcal{F}_t] \tag{2.95}$$

As with the calculation of the variance, this quantity still depends on t. We take further its expectation to get rid of the conditioning and we let $t \to \infty$ to get its stationary regime value, yielding :

$$\begin{aligned}
&\frac{1}{2(\alpha-\beta)^4} \times \lambda_\infty\beta \\
&\quad\Bigg(2\alpha^2 - e^{(\alpha-\beta)\delta}\alpha^2 - 2e^{(\alpha-\beta)\tau}\alpha^2 + 2e^{(\alpha-\beta)(\delta+\tau)}\alpha^2 - e^{(\alpha-\beta)(\delta+2\tau)}\alpha^2 \\
&\quad - 2\alpha\beta + 2e^{(\alpha-\beta)\delta}\alpha\beta + 2e^{(\alpha-\beta)\tau}\alpha\beta - 4e^{(\alpha-\beta)(\delta+\tau)}\alpha\beta \\
&\quad + 2e^{(\alpha-\beta)(\delta+2\tau)}\alpha\beta + 2\alpha^2\beta\tau - 4\alpha\beta^2\tau + 2\beta^3\tau\Bigg) \\
&\quad \frac{1}{2(\alpha-\beta)^4}\lambda_\infty^2\beta\Big(2\alpha^2\beta\tau^2 - 4\alpha\beta^2\tau^2 + 2\beta^3\tau^2\Big).
\end{aligned}$$

Further centering and norming this quantity with the long run mean and variance of jump numbers during intervals of length τ yields the autocorrelation function of jump number between intervals of length τ with a lag δ :

$$\begin{aligned}
Acf\,(\tau,\delta) &= lim_{t\to\infty}\frac{\mathbb{E}[(N_{t+\tau} - N_t)(N_{t+2\tau+\delta} - N_{t+2\tau+\delta})] - \mathbb{E}[(N_{t+\tau} - N_t)]\mathbb{E}[(N_{t+2\tau+\delta} - N_{t+2\tau+\delta})]}{\sqrt{var(N_{t+\tau} - N_t)var(N_{t+2\tau+\delta} - N_{t+2\tau+\delta})}} \\
&= \frac{e^{-2\beta\tau}\left(e^{\alpha\tau} - e^{\beta\tau}\right)^2\alpha(\alpha-2\beta)}{2\left(\alpha(\alpha-2\beta)\left(e^{(\alpha-\beta)\tau}-1\right) + \beta^2\tau(\alpha-\beta)\right)}e^{(\alpha-\beta)\delta}
\end{aligned} \tag{2.96}$$

As a conclusion, calculations are still feasable by relying on the moment generating function method, but they become very tedious. We do not see any manner to easily generalize them for the multidimensional case. Hence our preference for the Dynkin formula method exposed in the text.

Expression for the skewness

It is more easily obtained by the Dynkin formula method. By application of the infinitesimal generator operator to adequate functions, one has the following ordinary differential equations:

$$
\begin{aligned}
d\mathbb{E}[N^3] &= \mathbb{E}[\lambda_t]dt + 3\mathbb{E}[\lambda_t N_t]dt + 3\mathbb{E}[\lambda_t N_t^2]dt \\
d\mathbb{E}[\lambda_t N_t^2] &= \mathbb{E}[\lambda_t^2]dt + 2\mathbb{E}[\lambda_t^2 N_t]dt + \alpha\mathbb{E}[\lambda_t]dt + 2\alpha\mathbb{E}[\lambda_t N_t]dt + (\alpha-\beta)\mathbb{E}[\lambda_t N_t^2]dt + \beta\lambda_\infty\mathbb{E}[N_t^2]dt \\
d\mathbb{E}[\lambda_t^2 N_t] &= \mathbb{E}[\lambda_t^3]dt + 2\alpha\mathbb{E}[\lambda_t^2]dt + 2(\alpha-\beta)\mathbb{E}[\lambda_t^2 N_t]dt + \alpha^2\mathbb{E}[\lambda_t]dt + (\alpha^2+2\lambda_\infty\beta)\mathbb{E}[\lambda_t N_t]dt \\
d\mathbb{E}[\lambda_t^3] &= 3(\alpha-\beta)\mathbb{E}[\lambda_t^3]dt + 3\left(\alpha^2+\lambda_\infty\beta\right)\mathbb{E}[\lambda_t^2]dt + \alpha^3\mathbb{E}[\lambda_t]dt.
\end{aligned}
$$

The stationary regime third moment then writes:

$$
\begin{aligned}
\lim_{t\to\infty}\mathbb{E}\left[(N_{t+\tau}-N_t)^3\right] = \frac{1}{2(\alpha-\beta)^6}\lambda_\infty\beta & \\
\Bigg[& -e^{2(\alpha-\beta)\tau}\alpha^2(2\alpha-3\beta)(\alpha-\beta) \\
& + 2e^{(\alpha-\beta)\tau}\alpha\left(\alpha^3-4\alpha^2\beta+3\alpha\beta^2+6\beta^3+3(\lambda_\infty+\alpha)(\alpha-2\beta)(\alpha-\beta)\beta\tau\right) \\
& + \beta\Bigg(3\alpha\left(\alpha^2-\alpha\beta-4\beta^2\right) \\
& + 2(-\alpha+\beta)\left(3\lambda_\infty\alpha(\alpha-2\beta)+\beta^2(2\alpha+\beta)\right)\tau \\
& + 6\lambda_\infty(\alpha-\beta)^2\beta^2\tau^2 + 2\lambda_\infty^2\beta(-\alpha+\beta)^3\tau^3\Bigg)\Bigg].
\end{aligned}
$$

Expression for the kurtosis

Similarly to the preceding paragraph, we have the following ordinary differential equations:

$$
\begin{aligned}
d\mathbb{E}[N^4] &= \mathbb{E}[\lambda_t]dt + 4\mathbb{E}[\lambda_t N_t]dt + 6\mathbb{E}[\lambda_t N_t^2]dt + 4\mathbb{E}[\lambda_t N_t^3]dt \\
d\mathbb{E}[\lambda_t N_t^3] &= \mathbb{E}[\lambda_t^2]dt + 3\mathbb{E}[\lambda_t^2 N_t]dt + 3\mathbb{E}[\lambda_t^2 N_t^2]dt + \alpha\mathbb{E}[\lambda_t]dt + 3\alpha\mathbb{E}[\lambda_t N_t]dt + 3\alpha\mathbb{E}[\lambda_t N_t^2]dt \\
&\quad + (\alpha-\beta)\mathbb{E}[\lambda_t N_t^3]dt + \lambda_\infty\beta\mathbb{E}[N_t^3]dt \\
d\mathbb{E}[\lambda_t^2 N_t^2] &= \mathbb{E}[\lambda_t^3]dt + 2\mathbb{E}[\lambda_t^3 N_t]dt + 2\alpha\mathbb{E}[\lambda_t^2] + 4\alpha\mathbb{E}[\lambda_t^2 N_t]dt + 2\alpha\mathbb{E}[\lambda_t^2 N_t^2]dt + \alpha^2\mathbb{E}[\lambda_t]dt \\
&\quad + 2\alpha^2\mathbb{E}[\lambda_t N_t]dt + (\alpha^2+2\lambda_\infty\beta)\mathbb{E}\lambda_t N_t^2]dt - 2\beta\mathbb{E}\lambda_t^2 N_t^2]dt \\
d\mathbb{E}[\lambda_t^3 N_t] &= \mathbb{E}[\lambda_t^4]dt + 3\alpha\mathbb{E}[\lambda_t^3]dt + 3(\alpha-\beta)\mathbb{E}[\lambda_t^3 N_t]dt + 3\alpha^2\mathbb{E}[\lambda_t^2]dt + 3(\alpha^2+\lambda_\infty\beta)\mathbb{E}[\lambda_t^2 N_t]dt \\
&\quad + \alpha^3\mathbb{E}[\lambda_t]dt + \alpha^3\mathbb{E}[\lambda_t N_t]dt \\
d\mathbb{E}[\lambda_t^4] &= (4\alpha-4\beta)\mathbb{E}[\lambda_t^4]dt + \left(6\alpha^2+4\lambda_\infty\beta\right)\mathbb{E}[\lambda_t^3]dt + 4\alpha^3\mathbb{E}[\lambda_t^2]dt + \alpha^4\mathbb{E}[\lambda_t]dt.
\end{aligned}
$$

The stationary regime fourth moment of the process writes:

$$\lim_{t\to\infty}\mathbb{E}\left[(N_{t+\tau}-N_t)^4\right]=\frac{1}{6(\alpha-\beta)^8}\times$$
$$\lambda_\infty\beta\Bigg[-2e^{3(\alpha-\beta)\tau}\alpha^3(3\alpha-4\beta)(\alpha-\beta)(2\alpha-\beta)$$
$$+3e^{2(\alpha-\beta)\tau}\alpha^2\Bigg(6\alpha^4+6(\lambda_\infty-3\alpha)\alpha^2\beta+3\alpha(-8\lambda_\infty+\alpha)\beta^2$$
$$+6(4\lambda_\infty+5\alpha)\beta^3-18\beta^4+4(\lambda_\infty+2\alpha)(2\alpha-3\beta)(\alpha-\beta)^2\beta\tau\Bigg)$$
$$-6e^{(\alpha-\beta)\tau}\alpha\Bigg(\alpha^5+6\lambda_\infty\alpha^3\beta-3\alpha^4\beta-24\lambda_\infty\alpha^2\beta^2-\alpha^3\beta^2$$
$$+24\lambda_\infty\alpha\beta^3+20\alpha^2\beta^3-45\alpha\beta^4-14\beta^5$$
$$+2(\lambda_\infty+\alpha)(\alpha-\beta)\beta\left(2\alpha^3-8\alpha^2\beta+3\alpha\beta^2+18\beta^3\right)\tau$$
$$+6(\lambda_\infty+\alpha)^2(\alpha-2\beta)(\alpha-\beta)^2\beta^2\tau^2\Bigg)$$
$$+\beta\Bigg(\alpha\left(2\alpha^4+18\lambda_\infty\alpha(\alpha-2\beta)^2+15\alpha^3\beta+22\alpha^2\beta^2-216\alpha\beta^3-84\beta^4\right)$$
$$+6\beta(-\alpha+\beta)\left(6\lambda_\infty\alpha^3+6\alpha(-6\lambda_\infty+\alpha)\beta^2+8\alpha\beta^3+\beta^4\right)\tau$$
$$+6\lambda_\infty(\alpha-\beta)^2\beta\left(6\lambda_\infty\alpha(\alpha-2\beta)+\beta^2(8\alpha+7\beta)\right)\tau^2+36\lambda_\infty^2\beta^3(-\alpha+\beta)^3\tau^3+6\lambda_\infty^3(\alpha-\beta)^4\beta^2\tau^4\Bigg)\Bigg]$$

Chapter 3

Clustering and Mean Reversion in a Hawkes Microstructure Model

Note : A part of this chapter will be published in Journal of Futures Markets, DOI: 10.1002/fut.21676.

Abstract

This chapter generalizes the preceding one. We provide explicit formulas for the first and second moments and the autocorrelation function of the number of jumps over a given interval for the multivariate Hawkes process. These computations are possible thanks to the affine property of this process. We unify the stock price models of the preceding chapter, both of them based on the Hawkes process, one having a mean reverting behaviour whilst the other has a clustering behaviour, and build a model having these two properties. We compute various statistics as well as the diffusive limit for the stock price that determines the connection between the parameters driving the high frequency activity to the daily volatility.
Additionally, we differentiate market activity by signing the market orders (buy or sell). Modelling this activity as a multivariate Hawkes process, we compute the market impact of a trade as well as a collection of trades. We obtain the usual concave-form function documented in the empirical literature.

Contents

Introduction

Our study of the self-exciting Hawkes process in the preceding chapter allowed us to experiment with models capturing the clustering property of the market microstructure and compare them to models taking into account only the mean reversion behaviour. The end of the preceding chapter strongly suggests to take into account both self and cross correlation effects in the market microstructure, a property that is handled by multivariate Hawkes processes presenting simultaneously self and mutual excitation.

The purpose of this chapter is then to generalize the results obtained in the preceding one to the multivariate case, and calculate analytically some moments of the n-dimension Hawkes process in closed form. The Dynkin formula approach used previously will then prove to be very efficient.

We use these analytical results to understand the statistical properties of a stock price that is expressed as a function of a Hawkes process with both self and mutual excitation, unifying the models presented in the previous chapter. For instance, we compute explicitly the general form of the signature plot and explain the different patterns that it can take. It turns out that depending on the parameter values, the microscopic stock returns have a dominant mean reverting or clustering behaviour, resulting in a decreasing or increasing signature plot pattern. Also, we compute the diffusive limit associated with the stock dynamics as well as its asymptotic volatility and connect these to its microscopic dynamics.

Additionally, whereas in the preceding chapter we considered market activity, i.e. market order arrival times as a whole, uncovering and studying its clustering property, we differentiate in this chapter between buy and sell activities. Using a model for buy and sell order arrival times based on the multivariate Hawkes process, we compute the market impact of a trade. Extension of the result to a set of orders allows to recover the usual concave shape function of meta order impact as presented in the literature.

The structure of this chapter is as follows. In the first section, we show empirical evidences of mean reverting and clustering behaviours in high frequency data. In the second section, we describe the analytical framework which comprises the basic properties of the Hawkes process as well as the Dynkin formula that will be our main mathematical tool. Using these results, the computation of the moments and the autocorrelation function of the number of jumps over a given time interval is provided. In the third section, we develop two applications. The first application is a toy model for the stock price based on a multivariate Hawkes process for which we compute different statistical properties and the diffusive limit for the stock. The second application is a toy model for buy and sell order arrival times enabling us to compute in closed form the market impact of a meta order.

3.1 Empirical Evidences

We define the clustering as the strong autocorrelation in the occurrence frequency of a certain event. The mean reversion is the strong correlation in the occurrence of two opposite events, such as buy and sell order arrivals or up and down price movements. Both clustering and mean reversion are important characteristics of many aspects of the trading activity. The random walk nature of the price process observed on a coarse time scale is the result of their subtle interaction on a microscopic time scale. Borrowing the language of (Bouchaud et al., 2004), we can say that clustering (*persistence*), leads to *super diffusion* whereas mean reversion (*anti-persistence*) leads to *sub-diffusion*, the overall effect being the diffusive aspect of the price process.

For example, time series of the number of up mid price jumps during consecutive time intervals of length τ are highly autocorrelated. This correlation persists for several lags as is clear in Figure 3.1. The same is true for down mid price jumps. The same phenomenon is observed daily and for any choice of τ.

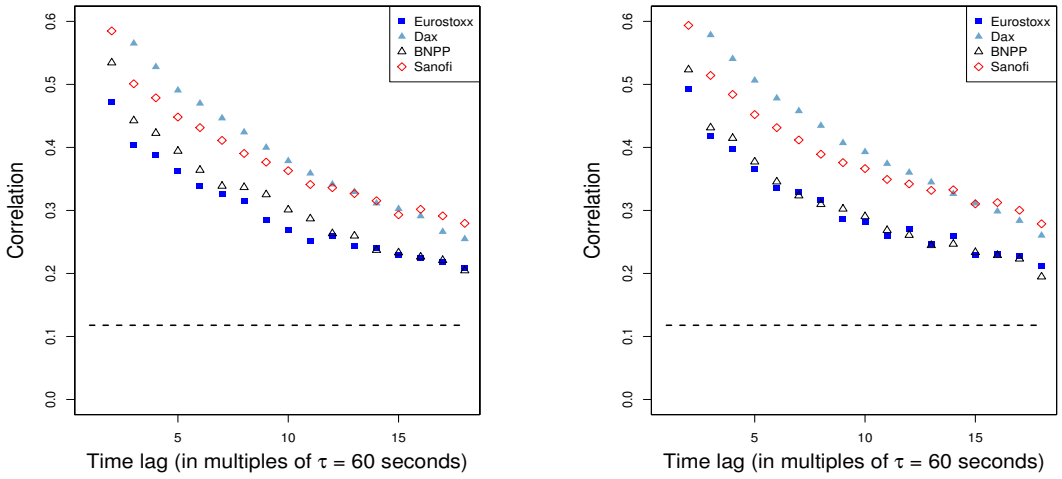

Figure 3.1: Autocorrelation of the up (left) and down (right) mid price jumps occurring on consecutive, non overlapping time intervals of length $\tau = 60$ seconds. We take average values of September 2011. The same plots can be observed for other values of τ. The dashed line presents the significance threshold of the estimated correlations at a 99% confidence level.

Additionally, if one considers time series of buy and sell order arrivals[1], the same phenomenon is observed as is clear in Figure 3.2.

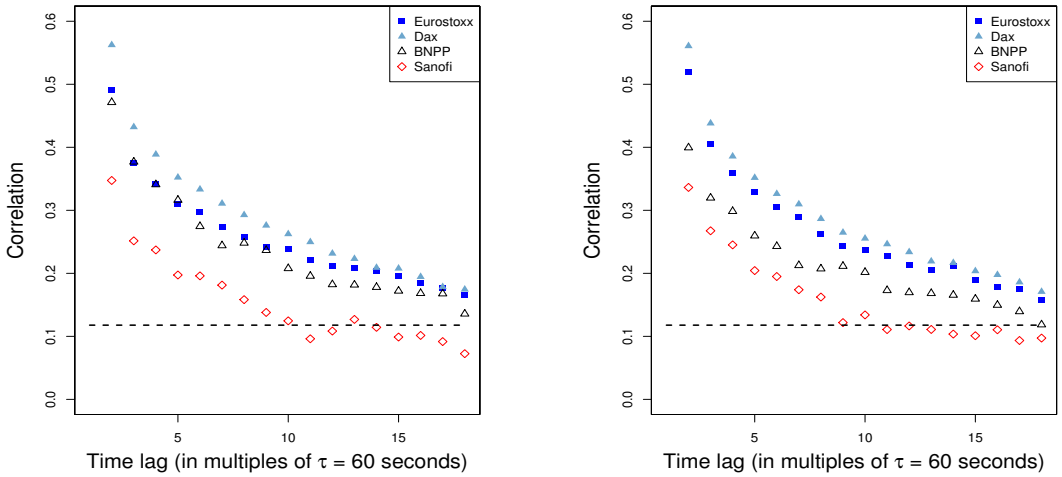

Figure 3.2: Autocorrelation of the number of buy trades (left) and sell trades (right) occurring on consecutive, non overlapping time intervals of length $\tau = 60$ seconds. We take average values of September 2011. The same plots can be observed for other values of τ. The dashed line presents the significance threshold of the estimated correlations at a 99% confidence level.

[1]Adequate algorithms have to be applied to raw trade data to sign the trades, i.e to differentiate buy and sell trades. This will be explained later.

Conversely, cross correlation functions between the number of occurrences of opposite phenomena during a time interval also present the same property. Up mid price jumps seem then to trigger down jumps inducing a mean reversion of price returns. The correlation is also significant and persists for several lags as can be seen in Figure 3.3.

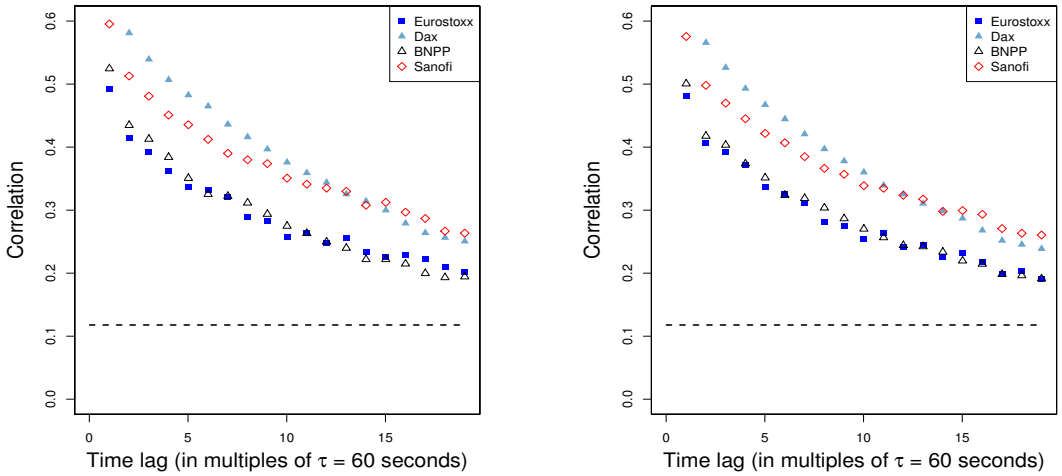

Figure 3.3: Cross correlation of up and down jumps of the mid price, where in the left figure down jumps are lagged, whereas in the right figure up jumps are lagged. Lags are measured in seconds and the intervals can overlap. Time interval length is $\tau = 60$ seconds. We take average values of September 2011.

Buy orders seem also to trigger sell orders, which in turn trigger other buy orders, as is clear in Figure 3.4[2].

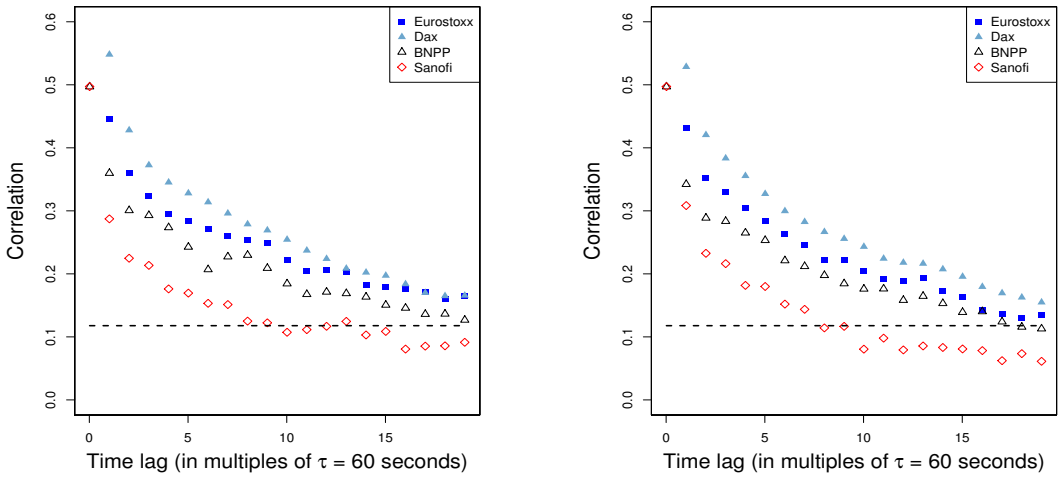

Figure 3.4: Cross correlation of the number of buy and sell trades, where in the left figure sell trades are lagged, whereas in the right figure buy trades are lagged. Lags are measured in seconds and the intervals can overlap. Time interval length is $\tau = 60$ seconds. We take average values of September 2011.

[2]Notice that persistence of this phenomenon is less for Sanofi which is the less liquid among the considered assets.

Therefore, the clustering effect is compensated by a mean reversion phenomenon as illustrated in the cross correlation functions. This rich interaction in the micro structure level results macroscopically in a price trajectory for which the efficient market hypothesis as well as the Brownian diffusion approximation seem to be reasonable.

Clearly, Hawkes processes present a versatile mathematical framework allowing us to deal with such phenomena. Precisely, we need to consider multivariate Hawkes process presenting self excitation as well as mutual excitation between their components. We analytically study these processes in the next section and calculate some of their distributional properties.

3.2 Mathematical Framework

3.2.1 The multivariate Hawkes process

In the multivariate case, the Hawkes process can also be defined thanks to the stochastic differential equation satisfied by its intensity which writes in the matrix form :

$$d\lambda_t = \beta(\lambda_\infty - \lambda_t)dt + \alpha dN_t \tag{3.1}$$

with β, α two $n \times n$ real matrices and λ_∞ a vector of $\mathbb{R}_+^n$. Applying Ito's lemma to $e^{\beta t}\lambda_t$ yields:

$$\lambda_t = e^{-\beta t}\left(\lambda_0 - \lambda_\infty\right) + \lambda_\infty + \int_0^t e^{-\beta(t-v)}\alpha dN_v. \tag{3.2}$$

From (3.2) and under the hypothesis that β has positive eigenvalues we observe that the impact on the intensity of a jump dies out exponentially as time passes. Also, as the intensities must be positive notice that the matrix α has to be component-wise positive. For the existence and uniqueness results we refer to Chapter 14 of (Daley and Jones, 2008) and references therein, of particular interest is (Brémaud and Massoulié, 1994).

As t gets larger the impact of λ_0, the initial value for the intensity, vanishes leaving us with:

$$\lambda_t \sim \lambda_\infty + \int_0^t e^{-\beta(t-v)}\alpha dN_v.$$

As in the monovariate case, $X_t = (\lambda_t, N_t)$ is a Markov process in the state space $D = \mathbb{R}_+^n \times \mathbb{N}^n$.

The infinitesimal generator of the diffusion is given by

$$\mathcal{L}f = \left(\beta(\lambda_\infty - \lambda)\right)^\top \nabla^\top f + \lambda^\top \begin{pmatrix} f(\lambda + \alpha e_1, N_t + e_1) - f \\ \cdots \\ \cdots \\ f(\lambda + \alpha e_n, N_t + e_n) - f \end{pmatrix} \tag{3.3}$$

for $f : D \to \mathbb{R}$, where $\nabla f = (\partial_{\lambda_1} f, ..., \partial_{\lambda_n} f)$ is a $1 \times n$ vector and $(e_i)_{i=1..n}$ is the canonical basis of $\mathbb{R}^n$ ($\top$ stands for the matrix transpose).

And we have the Dynkin formula:

$$\mathbb{E}\left[f\left(X_s\right)|\mathcal{F}_t\right] = f\left(X_t\right) + \mathbb{E}\left[\int_t^s \mathcal{L}f\left(X_v\right)dv|\mathcal{F}_t\right]. \tag{3.4}$$

This gives a very convenient way to calculate conditional expectations of functions of the Markov process $X_t = (\lambda_t, N_t)$ when the expectation of the right hand side of the preceding equation can be easily computed.

The infinitesimal generator of the diffusion leads, thanks to Feynman-Kac's formula, to the computation of the moment-generating function. Denoted by $\phi(t,z,u) = \mathbb{E}\left[e^{z^\top \lambda_t + u^\top N_t}\right]$ for $z \in \mathbb{R}^n$ and $u \in \mathbb{N}^n$, this function solves the partial differential equation with initial condition:

$$\begin{cases} \partial_t \phi = \mathcal{L}\phi \\ \phi(0,z,u) = e^{z^\top \lambda_0 + u^\top N_0}. \end{cases}$$

The model being affine we look for a solution of the form $e^{a_t + b_t^\top \lambda + u^\top N}$ with $a_t \in \mathbb{R}$ and $b_t \in \mathbb{R}^n$. It leads to a set of ordinary differential equations:

$$\begin{cases} \partial_t a = b^\top \beta \lambda_\infty \\ \partial_t b = -\beta^\top b + h - \mathbf{1} \end{cases}$$

with initial conditions $a_0 = 0$, $b_0 = z$, and $\mathbf{1} = (1,\dots,1)^\top$ whilst the function h is defined as:

$$h = \begin{pmatrix} e^{b^\top \alpha e_1 + u^\top e_1} \\ \dots \\ e^{b^\top \alpha e_n + u^\top e_n}. \end{pmatrix}.$$

From a numerical point of view it is always possible to simulate the ODE but explicitly computing the solution is difficult. Also, if we are interested in the moments of the process then we need to derive the solution with respect to the parameter z which in turn leads to the computation of the derivative with respect to this parameter of the ODE. Although the first moment can be easily computed, higher moments remain a challenge. To the extent that we are interested only in the moments we rely on the simpler computation strategy based on Dynkin's formula (3.4).

3.2.2 Computing the moments and autocorrelation function

Our aim in this section is to compute the moments of the process $X_t = (\lambda_t, N_t)$ and also the autocovariance of the number of jumps over a period τ . To achieve this we rely on the infinitesimal generator of the process given by (3.3) and Dynkin's formula (3.4). In order to obtain the expected number of jumps and the expected intensity we use the following lemma.

Lemma 3.1. *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic given by (3.1) then the expected number of jumps $\mathbb{E}[N_t]$ and the expected intensity $\mathbb{E}[\lambda_t]$ satisfy the set of ODE:*

$$d\mathbb{E}[\lambda_t] = \beta(\lambda_\infty - \mathbb{E}[\lambda_t])dt + \alpha\mathbb{E}[\lambda_t]dt \tag{3.5}$$
$$d\mathbb{E}[N_t] = \mathbb{E}[\lambda_t]dt. \tag{3.6}$$

These equations can be integrated explicitly as we have:

$$\begin{aligned} \mathbb{E}[\lambda_t] &= (\alpha-\beta)^{-1}\left(e^{(\alpha-\beta)t} - I\right)\beta\lambda_\infty + e^{(\alpha-\beta)t}\lambda_0 \\ &= c_0(t)\lambda_0 + c_1(t) \end{aligned} \tag{3.7}$$

and

$$\begin{aligned} \mathbb{E}[N_t] &= N_0 + (\alpha-\beta)^{-1}\left(e^{(\alpha-\beta)t} - I\right)\lambda_0 + \left(\left((\alpha-\beta)^{-1}\right)^2\left(e^{(\alpha-\beta)t} - I\right)\beta\lambda_\infty - (\alpha-\beta)^{-1}\beta\lambda_\infty t\right) \\ &= N_0 + c_2(t)\lambda_0 + c_3(t). \end{aligned} \tag{3.8}$$

Proof. The formulas (3.3) and (3.4) can be extended to a function f with values in $\mathbb{R}^n$ by applying these equations to each component of f.

Then, as in the preceding chapter, applying the Dynkin formula to $f \equiv N_t$, and then differentiating the obtained integral equation yields (3.6), and the same reasoning applied to $f \equiv \lambda_t$ yields the ordinary differential equation satisfied by the expected intensity (3.5). ■

Thanks to the expressions (3.7) and (3.8) we can compute the following asymptotic expectations:

Lemma 3.2. *Given a Hawkes process with dynamic given by (3.1) then long term expected intensity is given by:*

$$\lim_{t\to+\infty} \mathbb{E}\left[\lambda_t\right] = \bar{\lambda}_\infty = -(\alpha-\beta)^{-1}\beta\lambda_\infty \tag{3.9}$$

whilst the long term expected number of jumps over an interval τ is:

$$\begin{aligned}\lim_{t\to+\infty} \mathbb{E}\left[N_{t+\tau} - N_t\right] &= -(\alpha-\beta)^{-1}\beta\lambda_\infty\tau \\ &= \bar{\lambda}_\infty\tau.\end{aligned} \tag{3.10}$$

Establishing (3.9) requires that $\alpha-\beta$ has negative eigenvalues, which is the classical stability condition of the multivariate Hawkes Process as stated in (Hawkes, 1971). Therefore, from now on we will suppose this property satisfied. The computation of the second order moments leads to the following lemma:

Lemma 3.3. *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic given by (3.1) then the functions $\mathbb{E}\left[N_t N_t^\top\right]$, $\mathbb{E}\left[\lambda_t N_t^\top\right]$ and $\mathbb{E}\left[\lambda_t \lambda_t^\top\right]$ solve the set of ODE:*

$$\frac{d}{dt}\mathbb{E}\left[N_t N_t^\top\right] = \mathbb{E}\left[\lambda_t N_t^\top\right] + \mathbb{E}\left[N_t \lambda_t^\top\right] + \mathrm{diag}(\mathbb{E}[\lambda_t]) \tag{3.11}$$

$$\frac{d}{dt}\mathbb{E}\left[\lambda_t N_t^\top\right] = \beta\lambda_\infty\mathbb{E}\left[N_t^\top\right] + (\alpha-\beta)\mathbb{E}\left[\lambda_t N_t^\top\right] + \mathbb{E}\left[\lambda_t\lambda_t^\top\right] + \alpha\mathrm{diag}(\mathbb{E}[\lambda_t]) \tag{3.12}$$

$$\frac{d}{dt}\mathbb{E}\left[\lambda_t \lambda_t^\top\right] = \beta\lambda_\infty\mathbb{E}\left[\lambda_t^\top\right] + \mathbb{E}[\lambda_t]\lambda_\infty^\top\beta^\top + (\alpha-\beta)\mathbb{E}\left[\lambda_t\lambda_t^\top\right] + \mathbb{E}\left[\lambda_t\lambda_t^\top\right](\alpha-\beta)^\top + \alpha\mathrm{diag}(\mathbb{E}[\lambda_t])\alpha^\top \tag{3.13}$$

and the long term covariance matrix for the intensity $\Lambda_\infty = \lim_{t\to+\infty}\mathbb{E}\left[\lambda_t\lambda_t^\top\right]$ solves the algebraic matrix equation:

$$(\alpha-\beta)\bar{\Lambda}_\infty + \bar{\Lambda}_\infty(\alpha-\beta)^\top + \alpha\mathrm{diag}(\bar{\lambda}_\infty)\alpha^\top = 0 \tag{3.14}$$

with $\bar{\Lambda}_\infty = \Lambda_\infty - \bar{\lambda}_\infty\bar{\lambda}_\infty^\top$ where $\bar{\lambda}_\infty$ is given by (3.9).

Proof. The proof follows the same steps at the proof of Lemma 3.1 with computations slightly more involved. ■

Using the previous lemmas we can compute second order moment as well as the auto-covariance function of the number of jumps over a given time interval. We have the following lemma

Lemma 3.4. *The long term second order moment of the number of jumps over a given interval $\tau > 0$ is:*

$$\begin{aligned}\mathrm{Cov}(\tau) &= \lim_{t\to+\infty}\mathbb{E}\left[(N_{t+\tau}-N_t)(N_{t+\tau}-N_t)^\top\right] - \mathbb{E}\left[(N_{t+\tau}-N_t)\right]\mathbb{E}\left[(N_{t+\tau}-N_t)^\top\right] \\ &= J_1 + J_1^\top + \tau\mathrm{diag}(\bar{\lambda}_\infty)\end{aligned} \tag{3.15}$$

with $J_1 = c_5(\tau)(\bar{\Lambda}_\infty + \alpha\mathrm{diag}(\bar{\lambda}_\infty))$ and

$$c_5(\tau) = -(\alpha-\beta)^{-1}\tau + (\alpha-\beta)^{-2}(e^{(\alpha-\beta)\tau} - I).$$

Proof. We start with:

$$
\begin{aligned}
I_1 &= \mathbb{E}\left[(N_{t+\tau}-N_t)(N_{t+\tau}-N_t)^\top\right] \\
&= \mathbb{E}\left[N_{t+\tau}N_{t+\tau}^\top\right] - \mathbb{E}\left[N_{t+\tau}N_t^\top\right] - \mathbb{E}\left[N_tN_{t+\tau}^\top\right] + \mathbb{E}\left[N_tN_t^\top\right] && (3.16)\\
&= 2\mathbb{E}\left[N_tN_t^\top\right] + \int_t^{t+\tau}\mathbb{E}\left[\lambda_sN_s^\top\right] + \mathbb{E}\left[N_s\lambda_s^\top\right] + \mathrm{diag}(\mathbb{E}\left[\lambda_s\right])ds - \mathbb{E}\left[N_{t+\tau}N_t^\top\right] - \mathbb{E}\left[N_tN_{t+\tau}^\top\right] && (3.17)
\end{aligned}
$$

where from (3.16) to (3.17) we used (3.11). Moreover, we have:

$$
\begin{aligned}
I_2 &= \int_t^{t+\tau}\mathbb{E}\left[\lambda_sN_s^\top\right]ds \\
&= \int_t^{t+\tau}e^{(\alpha-\beta)(s-t)}\mathbb{E}\left[\lambda_tN_t^\top\right]ds + \int_t^{t+\tau}\int_t^s e^{(\alpha-\beta)(s-u)}\left\{\beta\lambda_\infty\mathbb{E}\left[N_u^\top\right] + \mathbb{E}\left[\lambda_u\lambda_u^\top\right] + \alpha\mathrm{diag}(\mathbb{E}\left[\lambda_u\right])\right\}duds \\
&= \int_t^{t+\tau}e^{(\alpha-\beta)(s-t)}\mathbb{E}\left[\lambda_tN_t^\top\right]ds + \int_t^{t+\tau}\int_t^s e^{(\alpha-\beta)(s-u)}\left\{\beta\lambda_\infty\mathbb{E}\left[N_t^\top\right] + \beta\lambda_\infty\int_t^u\mathbb{E}\left[\lambda_r^\top\right]dr\right\}duds && (3.18)\\
&+ \int_t^{t+\tau}\int_t^s e^{(\alpha-\beta)(s-u)}\left\{\mathbb{E}\left[\lambda_u\lambda_u^\top\right] + \alpha\mathrm{diag}(\mathbb{E}\left[\lambda_u\right])\right\}duds && (3.19)
\end{aligned}
$$

where we used successively (3.12) and (3.8). The fifth term of (3.17) is, after using the ODE for $\mathbb{E}\left[N_t\right]$ and conveniently conditioning, equal to:

$$
\begin{aligned}
I_3 &= \mathbb{E}\left[N_{t+\tau}N_t^\top\right] \\
&= \mathbb{E}\left[\left(N_t + \int_t^{t+\tau}(c_0(s-t)\lambda_t + c_1(s-t))ds\right)N_t^\top\right]. && (3.20)
\end{aligned}
$$

The first term of (3.20) will cancel with the first term of (3.17), the second term of (3.20) with will cancel with the first term of (3.18) whilst the last term of (3.20) with the second term of (3.17) when $t\to+\infty$. Therefore, for t large we have

$$
\begin{aligned}
I_1 &= \underbrace{\int_t^{t+\tau}\int_t^s e^{(\alpha-\beta)(s-u)}\left\{\beta\lambda_\infty\int_t^u\mathbb{E}\left[\lambda_r^\top\right]dr + \mathbb{E}\left[\lambda_u\lambda_u^\top\right] + \alpha\mathrm{diag}(\mathbb{E}\left[\lambda_u\right])\right\}duds}_{K_1} && (3.21)\\
&+ K_1^\top + \int_t^{t+\tau}\mathrm{diag}(\mathbb{E}\left[\lambda_s\right])ds. && (3.22)
\end{aligned}
$$

Replacing in K_1 the expectations involving λ_t by their long term values we obtain:

$$
\begin{aligned}
K_1 &= c_4(\tau)\beta\lambda_\infty\bar{\lambda}_\infty^\top + c_5(\tau)(\Lambda_\infty + \alpha\mathrm{diag}(\bar{\lambda}_\infty)) && (3.23)\\
c_4(\tau) &= -(\alpha-\beta)^{-1}\frac{\tau^2}{2} - (\alpha-\beta)^{-2}\tau + (\alpha-\beta)^{-3}\left(e^{(\alpha-\beta)\tau} - I\right) && (3.24)\\
c_5(\tau) &= -(\alpha-\beta)^{-1}\tau + (\alpha-\beta)^{-2}(e^{(\alpha-\beta)\tau} - I). && (3.25)
\end{aligned}
$$

As $c_4(t)$ is related to $c_5(\tau)$ through:

$$
c_4(\tau) = \left(\frac{\tau^2}{2} - c_5(\tau)\right)(-(\alpha-\beta)^{-1}), \tag{3.26}
$$

K_1 can be rewritten as:

$$K_1 = \frac{\tau^2}{2}\bar{\lambda}_\infty\bar{\lambda}_\infty^\top + c_5(\tau)(\bar{\Lambda}_\infty + \alpha\text{diag}(\bar{\lambda}_\infty)).$$

Taking into account:

$$\lim_{t\to\infty}\mathbb{E}\left[N_{t+\tau} - N_t\right]\mathbb{E}\left[\left(N_{t+\tau} - N_t\right)^\top\right] = \tau^2\bar{\lambda}_\infty\bar{\lambda}_\infty^\top$$

we deduce the result. ■

As we are interested in the autocorrelation structure of the process the following quantity proves to be essential

Lemma 3.5. *Given $t_1 < t_2 \leq t_3 < t_4$ with $t_2 - t_1 = \tau_1$, $t_4 - t_3 = \tau_2$ and $t_3 - t_2 = \delta$ we have:*

$$\begin{aligned}\text{Cov}_1(\tau_1,\tau_2,\delta) &= \lim_{t_1\to+\infty}\mathbb{E}\left[(N_{t_4} - N_{t_3})(N_{t_2} - N_{t_1})^\top\right] - \mathbb{E}\left[(N_{t_4} - N_{t_3})\right]\mathbb{E}\left[(N_{t_2} - N_{t_1})^\top\right] \\ &= c_2(\tau_2)c_0(\delta)c_2(\tau_1)\left(\bar{\Lambda}_\infty + \alpha\text{diag}(\bar{\lambda}_\infty)\right) \qquad (3.27)\end{aligned}$$

with $\bar{\Lambda}_\infty$ given by (3.14) and $\bar{\lambda}_\infty$ by (3.9).

Proof. We need to determine:

$$\begin{aligned}I_4 &= \mathbb{E}\left[(N_{t_4} - N_{t_3})(N_{t_2} - N_{t_1})^\top\right] \\ &= \mathbb{E}\left[\mathbb{E}_{t_3}\left[(N_{t_4} - N_{t_3})\right](N_{t_2} - N_{t_1})^\top\right] \qquad (3.28) \\ &= \mathbb{E}\left[(c_2(\tau_2)\lambda_{t_3} + c_3(\tau_2))(N_{t_2} - N_{t_1})^\top\right] \qquad (3.29) \\ &= c_2(\tau_2)c_0(\delta)\mathbb{E}\left[\lambda_{t_2}(N_{t_2} - N_{t_1})^\top\right] + c_2(\tau_2)c_1(\delta)\mathbb{E}\left[(N_{t_2} - N_{t_1})^\top\right] + c_3(\tau_2)\tau_1\bar{\lambda}_\infty^\top \qquad (3.30)\end{aligned}$$

where from (3.28) to (3.29) we used (3.8), and from (3.29) to (3.30) we used (3.7) as well as (3.10). Taking into account that:

$$\begin{aligned}\mathbb{E}\left[\lambda_{t_2}(N_{t_2} - N_{t_1})^\top\right] &= \mathbb{E}\left[\lambda_{t_2}N_{t_2}^\top\right] - \mathbb{E}\left[\lambda_{t_2}N_{t_1}^\top\right] \\ &= e^{(\alpha-\beta)\tau_1}\mathbb{E}\left[\lambda_{t_1}N_{t_1}^\top\right] + \int_{t_1}^{t_2} e^{(\alpha-\beta)(t_2-s)}\left\{\beta\lambda_\infty\mathbb{E}\left[N_s^\top\right] + \mathbb{E}\left[\lambda_s\lambda_s^\top\right] + \alpha\text{diag}(\mathbb{E}\left[\lambda_s\right])\right\}ds \\ &\quad - \left(c_0(\tau_2)\mathbb{E}\left[\lambda_{t_1}N_{t_1}^\top\right] + c_1(\tau_2)\mathbb{E}\left[N_{t_1}^\top\right]\right).\end{aligned}$$

The first term of the last equation simplifies with last-but-one term. Replacing $\mathbb{E}\left[N_s^\top\right]$ by its integral given by (3.8) allows us to simply the last term of the equation and we are left with

$$\mathbb{E}\left[\lambda_{t_2}(N_{t_2} - N_{t_1})^\top\right] = \int_{t_1}^{t_2} e^{(\alpha-\beta)(t_2-s)}\left\{\beta\lambda_\infty\int_{t_1}^{s}\mathbb{E}\left[\lambda_u^\top\right]du + \mathbb{E}\left[\lambda_s\lambda_s^\top\right] + \alpha\text{diag}(\mathbb{E}\left[\lambda_s\right])\right\}ds.$$

Taking the long term values for the expectations (involving only the process λ_t) we get:

$$\mathbb{E}\left[\lambda_{t_2}(N_{t_2} - N_{t_1})^\top\right] = c_5(\tau_1)\beta\lambda_\infty\bar{\lambda}_\infty^\top + c_2(\tau_1)\left(\Lambda_\infty + \alpha\text{diag}(\bar{\lambda}_\infty)\right). \qquad (3.31)$$

As $c_5(\tau)$ is related to $c_2(\tau)$ through:

$$c_5(\tau) = (\tau - c_2(\tau))\left(-(\alpha-\beta)^{-1}\right) \qquad (3.32)$$

when used in conjunction with (3.31) in (3.30) leads to, after taking into account (3.10) for the second term and the definition for $\bar{\lambda}_\infty$ given by (3.9), to the following expression for I_4:

$$\begin{aligned} I_4 &= c_2(\tau_2)c_0(\delta)\left\{\tau_1\bar{\lambda}_\infty\bar{\lambda}_\infty^\top + c_2(\tau_1)\left(\bar{\Lambda}_\infty + \alpha\text{diag}(\bar{\lambda}_\infty)\right)\right\} \\ &+ c_2(\tau_2)c_1(\delta)\bar{\lambda}_\infty^\top\tau_1 + c_3(\tau_2)\tau_1\bar{\lambda}_\infty^\top. \end{aligned}$$

Taking into account the equalities $c_3(\tau) = (\tau - c_2(\tau))\bar{\lambda}_\infty$ and $c_1(\delta) = (I - c_0(\delta))\bar{\lambda}_\infty$ then if we subtract to I_4 the following quantity:

$$\lim_{t_1\to\infty}\mathbb{E}\left[N_{t_4} - N_{t_3}\right]\mathbb{E}\left[\left(N_{t_2} - N_{t_1}\right)^\top\right] = \tau_2\tau_1\bar{\lambda}_\infty\bar{\lambda}_\infty^\top$$

we obtain the result. ■

It is possible to relax the assumption of overlapping intervals made in the previous lemma. In fact, we have

Lemma 3.6. *Given $t_1 < t_3 \leq t_2 < t_4$ with $t_2 - t_1 = \tau_1$, $t_4 - t_3 = \tau_2$ and $t_3 - t_1 = \delta$ (note the difference with the previous lemma), then:*

$$\begin{aligned} \text{Cov}_2(\tau_1,\tau_2,\delta) &= \lim_{t_1\to+\infty}\mathbb{E}\left[(N_{t_4} - N_{t_3})(N_{t_2} - N_{t_1})^\top\right] - \mathbb{E}\left[N_{t_4} - N_{t_3}\right]\mathbb{E}\left[\left(N_{t_2} - N_{t_1}\right)^\top\right] \\ &= \text{Cov}_1(\tau_1,\tau_2 - (\tau_1 - \delta),0) + \text{Cov}(\tau_1 - \delta) + \text{Cov}_1(\delta,\tau_1 - \delta,0). \end{aligned} \tag{3.33}$$

Proof. Under the hypothesis of the lemma we can decompose the expectation as:

$$\begin{aligned} I_5 &= \mathbb{E}[(N_{t_4} - N_{t_3})(N_{t_2} - N_{t_1})^\top] \\ &= \mathbb{E}[((N_{t_4} - N_{t_2}) + (N_{t_2} - N_{t_3}))\,(N_{t_2} - N_{t_1})^\top] \\ &= \mathbb{E}[(N_{t_4} - N_{t_2})(N_{t_2} - N_{t_1})^\top] + \mathbb{E}[(N_{t_2} - N_{t_3})(N_{t_2} - N_{t_1})^\top] \\ &= \mathbb{E}[(N_{t_4} - N_{t_2})(N_{t_2} - N_{t_1})^\top] + \mathbb{E}[(N_{t_2} - N_{t_3})(N_{t_2} - N_{t_3})^\top] \\ &+ \mathbb{E}[(N_{t_2} - N_{t_3})(N_{t_3} - N_{t_1})^\top]. \end{aligned}$$

Similarly, if we decompose the product $\mathbb{E}[N_{t_4} - N_{t_3}]\mathbb{E}[(N_{t_2} - N_{t_1})^\top]$ then using Lemma 3.5 we obtain the announced result. ■

The expression for the autocovariance function leads naturally to the autocorrelation function of the number of jumps over a given time interval denoted as $\text{Corr}(\tau,\delta)$ which is a function of:

$$\lim_{t\to+\infty}\mathbb{E}\left[(N_{t+\tau} - N_t)(N_{t+\tau+\delta} - N_{t+\delta})^\top\right] - \mathbb{E}\left[N_{t+\tau} - N_t\right]\mathbb{E}\left[\left(N_{t+\tau+\delta} - N_{t+\delta}\right)^\top\right] \tag{3.34}$$

where, depending whether $\delta \geq \tau$ or not, we use either (3.27) or (3.33) for the centered autocovariance, and the square root of the diagonal terms of the matrix $\text{Cov}(\tau)$ in order to rescale by the variances.

The lemma 3.5 leads to the following useful result whose proof is straightforward

Lemma 3.7. *Given the covariance matrix $\text{Cov}_1(\tau_1,\tau_2,\delta)$ defined by (3.27) then we have:*

$$\begin{aligned} \bar{\Sigma} &= \sum_{j=0}^{\infty}\text{Cov}_1(1,1,j) \\ &= (\alpha - \beta)^{-2}(I - e^{\alpha-\beta})(\bar{\Lambda}_\infty + \alpha\text{diag}(\bar{\lambda}_\infty)). \end{aligned} \tag{3.35}$$

A last lemma which eases significantly the computation is given by:

Lemma 3.8. *We suppose a two-dimensional Hawkes process $X_t = (\lambda_t, N_t)$ with values in $D = \mathbb{R}_+^2 \times \mathbb{N}^2$. Given the matrix $\bar{\Sigma}$ of (3.35), the covariance matrix $\mathrm{Cov}(1)$ defined by (3.15) and define $M = \mathrm{Cov}(1) + 2\bar{\Sigma}$ then:*

$$M_{11} + M_{22} - M_{12} - M_{21} = \tilde{M}_{11} + \tilde{M}_{22} - \tilde{M}_{12} - \tilde{M}_{21} \tag{3.36}$$

with:

$$\tilde{M} = \tilde{J}_1 + \tilde{J}_1^\top + \mathrm{diag}(\bar{\lambda}_\infty) \tag{3.37}$$

$$\tilde{J}_1 = -(\alpha - \beta)^{-1}(\bar{\Lambda}_\infty + \alpha \mathrm{diag}(\bar{\lambda}_\infty)). \tag{3.38}$$

The numerical consequences of the previous lemma are important because the left hand side of (3.36) involves exponential of matrices, through c_5 of Lemma 3.4, whereas the right hand side involves no exponentiation.

These lemmas provide the main equations that will be involved in the applications developed in this paper.

3.3 Applications

3.3.1 Data Description and Estimation Algorithm

We rely on tick-by-tick data from TRTH (Thomson Reuters Tick History). We deal with futures on indices such as Dax and Eurostoxx, as well as some other commodity, interest rates and forex futures. The data covers the period between 2010/01/01 to 2011/12/31. It consists of quote files recording quote changes (bid/ask prices and quantities) timestamped up to the millisecond, as well as trade files recording the transactions (prices and quantities) timestamped up to the millisecond.

For some of our empirical investigations, we need to infer the sign of the trades and to this purpose we rely on the Lee and Ready algorithm as introduced in (Lee and Ready, 1991).

The estimation algorithm relies on maximum likelihood method[3]. From Proposition 7.2.III of (Daley and Jones, 2002), the log-likelihood of a point process $(N_t)_{t\geq 0}$ writes up to an additive constant:

$$\begin{aligned} L &= -\int_0^T \mathbf{1}^\top \lambda_t dt + \int_0^T \ln(\lambda_t)^\top dN_t \\ &= -\int_0^T \mathbf{1}^\top \left(\lambda_\infty + \int_0^t e^{-\beta(t-s)} \alpha dN_s\right) dt + \sum_{i=1}^n \sum_{j=1}^{N_T^j} \ln\left(\lambda_{t_j}^i\right) \end{aligned}$$

So that, the log-likelihood of the multidimensional Hawkes process is the sum of the log-likelihoods generated by the observation of each coordinate process:

$$L = \sum_{m=1}^n L_m.$$

[3]Note that as was done in the monovariate case, another estimation algorithm can be developed using the moments and the autocorrelation function. We implemented it and made a Monte Carlo experiment as was done in the preceding chapter in order to assess its quality. It turns out to be very efficient especially in terms of computation speed. We will be interested here in likelihood ratio calculations, therefore we rely on MLE algorithm only.

This takes a particularly simple form if we consider a diagonal structure for the β matrix, that is $\beta = diag(\bar{\beta}_1, ..., \bar{\beta}_n)$, we obtain thanks to (Ogata, 1981):

$$L_m = -\lambda_\infty^m T - \sum_{i=1}^{n}\sum_{j=1}^{N} \frac{\alpha_{mi}}{\bar{\beta}_m}\left(1 - e^{\bar{\beta}_i(T-t_j)}\right) + \sum_{j=1}^{N_T} \ln\left[\lambda_\infty^m + \sum_{i=1}^{n} \alpha_{mi} R^{mi}(j)\right],$$

where:

$$\begin{aligned}
R^{mi}(1) &= 0, \\
R^{mm}(j) &= e^{-\bar{\beta}_m(t_j^m - t_{j-1}^m)}(1 + R^{mm}(j-1)), \\
R^{mi}(j) &= e^{-\bar{\beta}_m(t_i^m - t_{j-1}^m)} R^{mi}(j-1) + \sum_{k: t_{j-1}^m \leq t_k^i < t_j^m} e^{-\bar{\beta}_m(t_k^m - t_j^i)} \text{ for } i \neq m.
\end{aligned}$$

These recursive equations enable a very efficient calculation of the likelihood function.

3.3.2 Generalized Bacry-Delattre-Hoffmann-Muzy model

Model specification

In this section, we closely follow the spirit of Bacry, Delattre, Hoffmann, and Muzy (2013a). Our purpose is to build a model for the evolution of the *mid* price with both clustering and mean reversion in the asset microscopic returns. The model writes:

$$S_t = S_0 + \left(N_t^u - N_t^d\right)\frac{\nu}{2}, \tag{3.39}$$

where ν is the tick value. The N_t^u and N_t^d are Hawkes processes capturing the *up* and *down* jumps of the mid price. We consider that both processes are self exciting as well as mutually exciting. In the stationary regime, the intensities write:

$$\lambda_t^u = \lambda_\infty + \int_0^t \alpha_s e^{-\bar{\beta}(t-v)} dN_v^u + \int_0^t \alpha_m e^{-\bar{\beta}(t-v)} dN_v^d \tag{3.40}$$

$$\lambda_t^d = \lambda_\infty + \int_0^t \alpha_m e^{-\bar{\beta}(t-v)} dN_v^u + \int_0^t \alpha_s e^{-\bar{\beta}(t-v)} dN_v^d, \tag{3.41}$$

and within our original notations this translates to:

$$\alpha = \begin{pmatrix} \alpha_s & \alpha_m \\ \alpha_m & \alpha_s \end{pmatrix}; \beta = \begin{pmatrix} \bar{\beta} & 0 \\ 0 & \bar{\beta} \end{pmatrix}$$

where α_s stands for the self excitation parameter, α_m stands for the mutual excitation parameter, and the two intensities lead to a two-dimensional vector $\lambda_t = (\lambda_t^u, \lambda_t^d)^\top$ whilst the two jump processes give the vector $N_t = (N_t^u, N_t^d)^\top$.

Notice that α is component-wise positive, that is to say $\alpha_s > 0$ and $\alpha_m > 0$. As a result, we exclude any inhibitory effects (i.e. a jump in one component of the process decreasing the probability of jump for the other component)[4].

A stability condition for the bivariate Hawkes process used in the above model is that the matrix $\alpha - \beta$ has eigenvalues with negative real parts. An intuitive manner to see that is that

[4]Strictly speaking, in order to implement inhibitory effects with negative α components, one has to ensure that the overall intensity remains positive. This can be done by taking the exponential of the parametrized intensity, or by implementing a threshold to its possible values (Bowsher, 2007).

time t expected intensity calculated in (3.7) involves the expression $e^{(\alpha-\beta)t}$.

The eigenvalues of the matrix $\alpha - \beta$ are:

$$x_1 = \alpha_s - \bar{\beta} - \alpha_m \tag{3.42}$$
$$x_2 = \alpha_s - \bar{\beta} + \alpha_m. \tag{3.43}$$

so that the stability condition translates to:

$$x_1 < 0 \Leftrightarrow \bar{\beta} - \alpha_s + \alpha_m > 0 \tag{3.44}$$
$$x_2 < 0 \Leftrightarrow \bar{\beta} - \alpha_s - \alpha_m > 0. \tag{3.45}$$

We choose to consider perfectly symmetric processes so that the resultant price is balanced (i.e. it possesses the martingale property). This property will be necessary in order to derive the asymptotic volatility. Note that this asymptotic result can be obtained for more general matrices α and β so long as some constraints are satisfied. For example, if we consider a non symmetric setting we can still ensure the martingale property of the price. For instance, considering:

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}; \beta = \begin{pmatrix} \bar{\beta}_1 & 0 \\ 0 & \bar{\beta}_2 \end{pmatrix},$$

the formula for the first moment allow us to show that for the price to be a martingale, that is in order for the up jumps to completely offset the down jumps on average, the above variables have to satisfy:

$$\frac{\alpha_{11} + \alpha_{12}}{\bar{\beta}_1} = \frac{\alpha_{21} + \alpha_{22}}{\bar{\beta}_2}.$$

We nevertheless stick with our simpler setting as the aim here is to highlight clustering and mean reversion interactions. Considering a non-symmetric setting uncovers asymmetries in such interactions but we leave that interesting problem for a future research.

Calibration to empirical data

Using the data described in the data subsection we build the best bid and ask changes timestamped up to the millisecond. This allows us to produce time records of up and down mid price jumps corresponding to the bivariate Hawkes process (N_t) of our framework. Parameter estimation is conducted using the maximum likelihood estimator with the likelihood function computed as described in the previous subsection. The optimization is performed with the Nelder-Mead algorithm[5]. We estimate the model daily over a 2-year sample. After the first estimation, we use the obtained parameters as a first guess for the next calibration day. Results are reported in Table 3.1, where they appear to be very stable.

The model encompasses the (Bacry et al., 2013a) model as well as our toy model of the preceding chapter as particular cases. It is then of interest to determine whether modelling both the clustering and mean reversion in really needed. To that end, we conduct a likelihood ratio test and report the results in Table 3.1[6]. The values suggest that both clustering and mean reversion are needed to produce an accurate description of mid-price dynamics. Note that for some days for the Eurostoxx and the German bonds (Bobl, Bund and Schatz) the mean reversion effect alone might be sufficient.

[5]We used the open source library NL-opt, see http://ab-initio.mit.edu/wiki/index.php/NLopt

[6]When we compare the different models, with and without self excitation/mutual excitation, the starting points for the optimization algorithm have the same values for the common parameters. Let us underline the fact that for the Eurostoxx we had ten days for which the calibration was problematic so that these days were excluded.

Table 3.1: Calibration Results

Symbol	Measure	λ_∞	α_s	α_m	β	Clustering	Mean Reversion
Eurostoxx	Mean	0.0175	0.0062	0.0698	0.0958	1364.29	2.07
	Std. dev.	0.0065	0.0082	0.0242	0.0252	415.88	37.51
	Median	0.0156	0.0044	0.0624	0.0902	1270.00	0.00
Dax	Mean	0.0661	0.0347	0.0500	0.1019	673.07	404.00
	Std. dev.	0.0286	0.0102	0.0158	0.0221	319.35	299.83
	Median	0.0581	0.0341	0.0491	0.1050	708.50	406.00
BNPP	Mean	0.0495	0.2685	0.1792	0.4660	3394.05	12359.29
	Std. dev.	0.0220	0.1241	0.0841	0.1919	2612.05	13150.40
	Median	0.0439	0.2451	0.1648	0.4335	2944.00	8542.50
Sanofi	Mean	0.0345	0.1941	0.1572	0.3691	2969.66	7086.07
	Std. dev.	0.0188	0.1182	0.0902	0.1968	2123.99	6553.23
	Median	0.0301	0.1583	0.1363	0.3199	2230.00	4505.00
Bund	Mean	0.0250	0.0189	0.0792	0.1180	1179.34	68.45
	Std. dev.	0.0123	0.0115	0.0226	0.0275	245.24	82.22
	Median	0.0213	0.0181	0.0767	0.1154	1187.50	59.00
Bobl	Mean	0.0172	0.0138	0.0673	0.1009	1259.34	5.82
	Std. dev.	0.0073	0.0108	0.0226	0.0215	277.01	81.47
	Median	0.0147	0.0126	0.0607	0.0979	1207.00	18.50
Schatz	Mean	0.0153	0.0196	0.0662	0.1056	1480.55	27.34
	Std. dev.	0.0047	0.0124	0.0210	0.0224	385.31	79.29
	Median	0.0147	0.0172	0.0621	0.1003	1423.00	31.00
JPY	Mean	0.0449	0.2516	0.2420	0.5089	4410.78	4780.38
	Std. dev.	0.0174	0.1029	0.0941	0.1800	2085.32	2892.44
	Median	0.0434	0.2375	0.2294	0.4863	3973.00	4271.50
EURO	Mean	0.0622	0.3733	0.3550	0.7360	8956.69	10563.86
	Std. dev.	0.0240	0.1255	0.1145	0.1922	4041.59	5433.36
	Median	0.0596	0.3878	0.3682	0.7593	8669.00	9646.00
GOLD	Mean	0.0902	0.4080	0.3109	0.7300	3486.06	9290.28
	Std. dev.	0.0266	0.1481	0.1248	0.2288	2418.24	7088.23
	Median	0.0913	0.4243	0.3487	0.7866	3237.00	7333.00
Crude Oil Brent	Mean	0.0865	0.4506	0.2411	0.7068	7224.19	13651.24
	Std. dev.	0.0265	0.1407	0.0755	0.1838	5580.90	8525.66
	Median	0.0872	0.4628	0.2437	0.7267	4468.00	11025.00
Natural GAS	Mean	0.0614	0.3503	0.3044	0.6683	3766.40	6345.64
	Std. dev.	0.0194	0.1120	0.1121	0.1943	1563.27	2623.62
	Median	0.0608	0.3553	0.3038	0.6914	3707.00	6101.00
Sugar	Mean	0.0373	0.1998	0.1766	0.3932	1440.07	1868.66
	Std. dev.	0.0118	0.0877	0.0861	0.1625	1353.06	1546.63
	Median	0.0342	0.1689	0.1471	0.3379	938.50	1389.50
CORN	Mean	0.0337	0.1879	0.1803	0.3861	1521.93	1704.87
	Std. dev.	0.0130	0.0609	0.0618	0.1166	762.05	781.92
	Median	0.0317	0.1831	0.1744	0.3681	1374.00	1563.00
WHEAT	Mean	0.0473	0.2370	0.2104	0.4659	1698.37	2221.75
	Std. dev.	0.0168	0.0952	0.0843	0.1644	1220.72	1486.22
	Median	0.0448	0.2196	0.2002	0.4365	1417.00	1808.00

Note. Calibration results for two years of data. We calibrate daily a bivariate Hawkes process to the up and down mid price jump times for each symbol. The column Clustering presents the likelihood ratio statistic when we consider only clustering. The column Mean reversion presents the likelihood ratio statistic when we consider only mean reversion. For the likelihood ratio test as the difference between the models is one parameter the ratio has a $\chi^2(1)$ distribution, if $X \sim \chi^2(1)$ then we have $\mathbb{P}(X > 3.84) = 0.05$.

Statistical properties

Having specified the dynamics for the up and down price jumps, we can analyze the statistical properties of the asset returns in this toy model. Thanks to the computations carried out in the analytical section many of these properties can be explicitly expressed in terms of the parameters driving the Hawkes processes.

Within this simple model we can compute the autocovariance function of the price increments. To this end we consider the expected covariance of price increments over two non-overlapping time intervals of length τ and with lag δ, it is defined as:

$$\begin{aligned}\text{CovStock}(\tau,\delta) &= \mathbb{E}\left[(S_{t+\tau}-S_t)\left(S_{t+2\tau+\delta}-S_{t+\tau+\delta}\right)\right] \\ &= \mathbb{E}\left[\left((N^u_{t+\tau}-N^u_t)-(N^d_{t+\tau}-N^d_t)\right)\left((N^u_{t+2\tau+\delta}-N^u_{t+\tau+\delta})-(N^d_{t+2\tau+\delta}-N^d_{t+\tau+\delta})\right)\right]\frac{\nu^2}{4}.\end{aligned}$$

For the model considered this quantity can be explicitly computed as we have:

Proposition 3.9. *The autocovariance function CovStock(τ,δ) is given by:*

$$\text{CovStock}(\tau,\delta) = \frac{\nu^2}{4}\left(M_{11}+M_{22}-M_{12}-M_{21}\right) \tag{3.46}$$

where $M = \text{Cov}_1(\tau,\tau,\delta)$ is given by (3.27) in Lemma 3.5.

For intensities following the dynamics (3.40) and (3.41). The above equation leads to the expression for the autocorrelation function of price increments:

$$\text{CorrStock}(\tau,\delta) = -\frac{e^{-(\delta+2\tau)(\beta+\alpha_m-\alpha_s)}\left(e^{\tau(\beta+\alpha_m-\alpha_s)}-1\right)^2(\alpha_m-\alpha_s)(2\beta+\alpha_m-\alpha_s)}{2\beta^2(\beta+\alpha_m-\alpha_s)}. \tag{3.47}$$

We can then see that if $\alpha_m = 0$ and $\alpha_s > 0$ then the above correlation is positive. An up jump increases the mid price and the up intensity (without affecting the down intensity) which increases the likelihood of another up jump. As a result, there is a positive autocorrelation of the stock returns and a clustering of jumps in the same direction. On the contrary, if $\alpha_s = 0$ and $\alpha_m > 0$ the above correlation is negative. An up jump increases the mid price and the down intensity which increases the likelihood of a down jump and a decrease of the mid price. This leads to a mean reverting behaviour of the mid price. Whenever $\alpha_s = \alpha_m$ the above autocorrelation is equal to zero because the two opposite effects offset each other. In all cases, the autocorrelation of returns vanishes as the time lag increases because we have $\beta+\alpha_m-\alpha_s > 0$ which is a stability condition of the model as stated before. Moreover, the decay of the autocorrelation as a function of the lag increases with α_m whilst α_s has the opposite effect.

Having a better understanding of the clustering and mean reverting behaviour in this model we can analyze its impact on the signature plot. The use of high-frequency data leads to the estimation of the volatility from returns sampled at possibly different frequencies. The dependency of the resulting volatility on the sampling frequency is called the signature plot and is of tremendous importance in practice. Within the toy model this effect can be explicitly analyzed. Indeed, the realized variance over a period T calculated by sampling the data with time intervals of length τ can be written as:

$$\begin{aligned}\hat{C}(\tau) &= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(S_{(n+1)\tau}-S_{n\tau}\right)^2 = \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(\left(N^u_{(n+1)\tau}-N^u_{n\tau}\right)-\left(N^d_{(n+1)\tau}-N^d_{n\tau}\right)\right)^2\frac{\nu^2}{4} \\ &= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(N^u_{(n+1)\tau}-N^u_{n\tau}\right)^2\frac{\nu^2}{4}+\frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(N^d_{(n+1)\tau}-N^d_{n\tau}\right)^2\frac{\nu^2}{4} \\ &\quad -2\frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(N^u_{(n+1)\tau}-N^u_{n\tau}\right)\left(N^d_{(n+1)\tau}-N^d_{n\tau}\right)\frac{\nu^2}{4}.\end{aligned}$$

The mean signature plot, or more simply signature plot, is the expectation of the above quantity and is explicitly given by:

Proposition 3.10. *The signature plot* $C(\tau) = \mathbb{E}[\hat{C}(\tau)]$ *is:*

$$C(\tau) = \frac{\nu^2}{4\tau}\left(M_{11} + M_{22} - M_{12} - M_{21}\right) \tag{3.48}$$

$$= \frac{\nu^2}{2}\Lambda\left(\kappa^2 + \left(1-\kappa^2\right)\frac{\left(1-e^{-\tau\gamma}\right)}{\gamma\tau}\right) \tag{3.49}$$

where $M = \mathrm{Cov}(\tau)$ *in* (3.48) *is the second moment matrix* (3.15) *in Lemma 3.4 whilst* (3.49) *is the expression when the intensities follow the dynamics* (3.40) *and* (3.41) *with the parameters:*

$$\Lambda = \frac{\bar{\beta}\lambda_\infty}{\bar{\beta} - \alpha_s - \alpha_m}\ ,\ \kappa = \frac{\bar{\beta}}{\bar{\beta} + \alpha_m - \alpha_s} \text{ and } \gamma = \bar{\beta} + \alpha_m - \alpha_s.$$

Proof.

The computations are analytically tractable as we have explicit formulas for the inverse and exponential of a 2×2 symmetric matrix:

$$M = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \tag{3.50}$$

$$e^M = \frac{e^{a-b}}{2}\begin{pmatrix} 1+e^{2b} & -1+e^{2b} \\ -1+e^{2b} & 1+e^{2b} \end{pmatrix} \tag{3.51}$$

and the inverse of M writes:

$$M^{-1} = \frac{1}{a^2-b^2}\begin{pmatrix} a & -b \\ -b & a \end{pmatrix} \tag{3.52}$$

yielding:

$$M_{11} + M_{22} - M_{12} - M_{21} = 2\Lambda\left(\kappa^2\tau + \left(1-\kappa^2\right)\frac{\left(1-e^{-\tau\gamma}\right)}{\gamma}\right)$$

so that the expected signature plot is given by:

$$\mathbb{E}[\hat{C}(\tau)] = \frac{\nu^2}{4\tau}(M_{11} + M_{22} - M_{12} - M_{21})$$

which leads to the announced result. ■

Notice that $\alpha_s = 0$ enables us to retrieve (Bacry et al., 2013a)'s results, that is to say the case with intensities which are only mutually excited. If $\alpha_m = 0$ then we recover the result of the preceding chapter addressing the case of self-excitation only. Also, whenever the clustering and mean reversion effects are equal (i.e. $\alpha_m = \alpha_s$) the signature plot is flat and the volatility estimated at every time resolution is equal to $\frac{\nu^2}{2}\lambda_\infty$, there is no bias due to mean reversion or clustering because these two effects perfectly offset each other. Lastly, taking the limit τ to infinity leads to the asymptotic volatility as it will be proven in the following propositions.

The shape of the expected signature plot depends on the imbalance between clustering and mean reversion effects. If the mean reversion aspect is stronger, then the estimated volatility will increase with the sampling frequency, or said differently, the estimated volatility will be a decreasing function of τ which is the sampling period. Qualitatively, a strong mean reversion effect induces a negative autocorrelation of the returns at a smaller time scale and, therefore, a higher estimated volatility. When the estimation is done with a larger sampling period, returns are aggregated, and their oscillations tend to offset each other resulting in a smaller estimated volatility. On the contrary, if the clustering effect dominates, the autocorrelation

of the returns is positive and therefore smaller time scales do not introduce oscillations which results in a smaller estimated volatility, and larger time scales aggregate positively correlated returns inflating the estimated volatility.

Let us now focus on the computation of the diffusive limit associated with the stock dynamic. It allows the determination of the connection between the microscopic price formation process observed at transaction level to its macroscopic properties at a coarser time scale. In other words, we connect the stochastic differential equations used to model an asset price evolution at a daily frequency, such as in the Black-Scholes model which relies mainly on the continuous Brownian motion, to the discontinuous point process describing individual transactions. The Hawkes process, thanks to its strong analytical tractability, enables us to relate these two time scales.

To fulfil this objective a limit theorem is needed. In (Bacry et al., 2013b), the authors rely on martingale theory and limit theorems for semi-martingales to prove stability and convergence results for a general model with mutually exciting processes and a generic kernel. As we have done in last chapter, we sketch a simpler proof which may shed some light on the process dynamics in this simpler case with exponential kernels. The process $X_t = \left(N_t^u, \lambda_t^u, N_t^d, \lambda_t^d\right)$ is a Markov process and its infinitesimal generator writes:

$$\begin{aligned}\mathcal{L}f\left(x\right) = \bar{\beta}\left(\lambda_\infty - \lambda_t^u\right)\frac{\partial f}{\partial \lambda^u}\left(x\right) + \bar{\beta}\left(\lambda_\infty - \lambda_t^d\right)\frac{\partial f}{\partial \lambda^d}\left(x\right) \\ + \lambda_t^u\Big[f\left(N_t^u+1, \lambda_t^u+\alpha_s, N_t^d, \lambda_t^d+\alpha_m\right) - f\left(x\right)\Big] \\ + \lambda_t^d\Big[f\left(N_t^u, \lambda_t^u+\alpha_m, N_t^d+1, \lambda_t^d+\alpha_s\right) - f\left(x\right)\Big].\end{aligned}$$

Ergodicity of the process X_t, that is to say its convergence to a stationary regime, can be easily established thanks to the Foster-Lyapounov test function criterion. In our case, define the function $V(x) = \frac{\lambda^u+\lambda^d}{2\lambda_\infty}$, then a simple calculation yields the *geometric drift condition*:

$$\mathcal{L}V(x) \leq -(\bar{\beta} - \alpha_m - \alpha_s)V(x) + \bar{\beta}, \tag{3.53}$$

which grants, thanks to Theorems 6.1 and 7.1 in (Meyn and Tweedie, 2009) (and especially (CD3)), the V-uniform ergodicity of the process X_t.

Let us then write unit-time price increments:

$$\eta_i = \Big[(N_i^u - N_{i-1}^u) - (N_i^d - N_{i-1}^d)\Big] \times \frac{\nu}{2},$$

and consider the random sums

$$S_n = \sum_{i=1}^{n} \eta_i,$$

where $\{\eta_i; i = 1 \dots n\}$ denotes a set of price increments (note that $\mathbb{E}[\eta_i] = 0$). We are interested in the asymptotic behaviour of the price process

$$\bar{S_t^n} = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}}.$$

The V-uniform ergodicity and Theorem 16.1.5 in (Meyn and Tweedie, 2009) allows us to conclude that the increments are geometrically mixing and Theorem 19.3 of (Billingsley, 1999) proves that $\bar{S_t^n}$ converges to a Brownian motion in the sense of Skorokhod topology:

$$\bar{S_t^n} \Rightarrow \sigma W_t$$

with the volatility is given by:

$$\sigma^2 = \lim_{n\to\infty} \frac{\mathrm{Var}(S_n)}{n}.$$

The closed-form formulas obtained in the analytical part enable the explicit computation of this volatility as we have the following proposition:

Proposition 3.11. *The volatility σ^2 is given:*

$$\sigma^2 = \frac{\nu^2}{4}\left(M_{11} + M_{22} - M_{12} - M_{21}\right) \tag{3.54}$$

with

$$M = \mathrm{Cov}(1) + 2\bar{\Sigma} \tag{3.55}$$

where the two matrices in (3.55) are given by (3.15) and (3.35), respectively.

Proof.

The volatility is given by:

$$\begin{aligned}
\sigma^2 &= \lim_{n\to\infty} \frac{\mathrm{Var}(S_n)}{n} \\
&= \frac{\nu^2}{4}\mathbb{E}[\left((N_1^u - N_0^u) - \left(N_1^d - N_0^d\right)\right)^2] \\
&+ 2\frac{\nu^2}{4}\sum_{n=1}^{\infty}\mathbb{E}[\left((N_1^u - N_0^u) - \left(N_1^d - N_0^d\right)\right)\left((N_{1+n}^u - N_n^u) - \left(N_{1+n}^d - N_n^d\right)\right)] \\
&= \frac{\nu^2}{4}\left(\mathrm{Cov}(1)_{11} + \mathrm{Cov}(1)_{22} - \mathrm{Cov}(1)_{12} - \mathrm{Cov}(1)_{21}\right) \\
&+ 2\frac{\nu^2}{4}\left(\bar{\Sigma}_{11} + \bar{\Sigma}_{22} - \bar{\Sigma}_{12} - \bar{\Sigma}_{21}\right).
\end{aligned}$$

Define $M = \mathrm{Cov}(1) + 2\bar{\Sigma}$ then we obtain the result. ■

Notice that the stationarity and ergodicity of the increments η_i, together with the convergence of the series $\sigma^2 = \mathbb{E}[\eta_0^2] + 2\sum_{n=0}^{\infty}\mathbb{E}[\eta_0\eta_n]$ established in the above proposition are sufficient to conclude that $\bar{S}_t^n$ converges to a Brownian motion in the sense of Skorokhod topology thanks to theorem 19.1 in (Billingsley, 1999).

To evaluate this asymptotic volatility, the Lemma 3.8 is useful. This expression is valid for a general dynamic but for the particular choice made in this work the volatility turns out to have a very simple expression as the next proposition shows:

Proposition 3.12. *If intensities follow the dynamics (3.40) and (3.41) the volatility (3.54) has the expression:*

$$\begin{aligned}
\sigma^2 &= \frac{\nu^2}{2}\Lambda\kappa^2 \\
&= \frac{\nu^2}{2}\frac{\bar{\beta}^3\lambda_\infty}{(\bar{\beta} - \alpha_s - \alpha_m)(\bar{\beta} + \alpha_m - \alpha_s)^2}
\end{aligned}$$

with:

$$\Lambda = \frac{\bar{\beta}\lambda_\infty}{\bar{\beta} - \alpha_s - \alpha_m}, \quad \kappa = \frac{\bar{\beta}}{\bar{\beta} + \alpha_m - \alpha_s}.$$

As expected, increasing the self excitation parameter α_s, and hence the clustering effect, increases the volatility as this leads to positive autocorrelation of the returns. On the contrary, increasing the mutual excitation parameter α_m decreases the overall volatility because of the negative autocorrelation of the returns. Obviously, these parameters have to remain in the region $\bar{\beta} - \alpha_s - \alpha_m > 0$ in order for the stability condition to be satisfied. Also, taking the limit $\tau \to +\infty$ in (3.49) leads to an equality between the signature plot and the asymptotic volatility.

In (Bacry et al., 2013a) the authors consider a mean reversion effect only, and obtain the same formula as above with $\alpha_s = 0$, whereas in the previous chapter we consider clustering alone and obtain the same formula as above with $\alpha_m = 0$. In the previous chapter, calibration was done on real data for both models (clustering only or mean reversion only), and empirical results showed that considering only one effect leads to a systematically underestimated or systematically overestimated realized volatility.

We conduct here the same realized variance experiment for the generalized model. Results are reported in Table 3.2. They clearly show that calibrating the clustering as well as mean reversion effects result in a better fit of the realized volatility.

Table 3.2: Asymptotic Volatilities

Symbol	λ_∞	α_s	α_m	β	Empirical σ	Toy model σ
Eurostoxx	0.0119	0.0474	0.0401	0.1074	24.64%	19.08%
Dax	0.0581	0.0341	0.0491	0.1050	25.75%	23.45%
BNPP	0.0739	0.5287	0.2311	0.7798	56.43%	55.31%
Sanofi	0.0658	0.3503	0.3543	0.7246	31.42%	31.47%
Bund	0.0136	0.0471	0.0410	0.1081	5.35%	4.62%
Bobl	0.0125	0.0465	0.0396	0.1062	4.45%	4.44%
Schatz	0.0108	0.0360	0.0484	0.1044	2.01%	1.86%
JPY	0.0520	0.3320	0.3295	0.6814	10.74%	10.67%
EURO	0.0169	0.1007	0.1017	0.2224	5.49%	5.65%
GOLD	0.0677	0.3168	0.3561	0.6929	14.28%	14.35%
Crude Oil Brent	0.0359	0.1394	0.1394	0.2989	14.04%	14.09%
Natural GAS	0.0495	0.2752	0.2801	0.5744	48.84%	48.70%
Sugar	0.0315	0.1702	0.1642	0.3544	61.24%	61.21%
CORN	0.0591	0.3160	0.2855	0.6215	58.16%	58.08%
WHEAT	0.0642	0.2819	0.2884	0.5903	59.49%	59.46%

Note. Median values of asymptotic volatilities as calculated by the general toy model, compared to the realized volatilities of the day. We also put median values of Hawkes model parameters.

The analytical results enable us to further develop the model properties. In fact, wa can analyze the price impact of a quote-moving trade, i.e. a trade resulting in an up or down price jump. We illustrate the impact of an "up" move but similar results apply to a "down" move.

Proposition 3.13. *Given τ and δ two positive real numbers, the impact of an "up" move on the stock returns computed over a time interval of length τ is:*

$$\begin{aligned}
\text{IMP}(\tau, \delta) &= \lim_{t \to +\infty} \mathbb{E}[S_{t+\tau+\delta} - S_{t+\delta} | dN_t^u = 1] && (3.56)\\
&= \frac{\nu}{2\bar{\lambda}_\infty^u}(M_{11} - M_{12}) && (3.57)\\
&= -\nu \frac{e^{-\tau(\beta+\alpha_m-\alpha_s)} e^{-\delta(\beta+\alpha_m-\alpha_s)} \left(e^{\tau(\beta+\alpha_m-\alpha_s)} - 1\right)(\alpha_m - \alpha_s)(2\beta + \alpha_m - \alpha_s)}{4\left(\beta + \alpha_m - \alpha_s\right)^2} && (3.58)
\end{aligned}$$

with $M = c_2(\tau)c_0(\delta)(\bar{\Lambda}_\infty + \alpha \text{diag}(\bar{\lambda}_\infty))$.

Proof.

Given τ_0, τ and δ the following quantity needs to be computed:

$$\mathbb{E}[(S_{t+\tau+\delta} - S_{t+\delta})(N^u_{t+\tau_0} - N^u_t)] = \frac{\nu}{2}\mathbb{E}[(N^u_{t+\tau+\delta} - N^u_{t+\delta})(N^u_{t+\tau_0} - N^u_t)] \tag{3.59}$$

$$- \frac{\nu}{2}\mathbb{E}[(N^d_{t+\tau+\delta} - N^d_{t+\delta})(N^u_{t+\tau_0} - N^u_t)]. \tag{3.60}$$

As we have:

$$\begin{aligned} I_1 &= \mathbb{E}[(N^u_{t+\tau+\delta} - N^u_{t+\delta})(N^u_{t+\tau_0} - N^u_t)] \\ &= \sum_{i=0}^{\infty} \mathbb{E}[N^u_{t+\tau+\delta} - N^u_{t+\delta} | N^u_{t+\tau_0} - N^u_t = i] \times \mathbb{P}[N^u_{t+\tau_0} - N^u_t = i] \times i. \end{aligned}$$

As we have:

$$\begin{aligned} \mathbb{P}[N^u_{t+\tau_0} - N^u_t = 1 | \mathcal{F}_t] &= \lambda_t \tau_0 + o(\tau_0) \\ \mathbb{P}[N^u_{t+\tau_0} - N^u_t > 1 | \mathcal{F}_t] &= o(\tau_0) \\ \mathbb{P}[N^u_{t+\tau_0} - N^u_t = 0 | \mathcal{F}_t] &= 1 - \lambda_t \tau_0 + o(\tau_0), \end{aligned}$$

then taking the limit $\tau_0 \to 0$ and for t large we deduce that:

$$\lim_{\tau_0 \to 0} \frac{I_1}{\tau_0} \sim \mathbb{E}[N^u_{t+\tau+\delta} - N^u_{t+\delta} | dN^u_t = 1] \bar{\lambda}^u_\infty.$$

Moreover, as we have:

$$\mathbb{E}[(N^u_{t+\tau+\delta} - N^u_{t+\delta})(N^u_{t+\tau_0} - N^u_t)] = (\mathrm{Cov}_1(\tau_0, \tau, \delta))_{11},$$

we obtain:

$$\mathbb{E}[N^u_{t+\tau+\delta} - N^u_{t+\delta} | dN^u_t = 1] = \frac{1}{\bar{\lambda}^u_\infty}(c_2(\tau)c_0(\delta)(\bar{\Lambda}_\infty + \alpha \mathrm{diag}(\bar{\lambda}_\infty)))_{11}.$$

Similar computations can be carried out for (3.60) and the announced result is obtained. ■

It might be convenient to consider the stock evolution over an infinitesimal interval. From the previous proposition it is straightforward to deduce:

Proposition 3.14. *Given the price impact function (3.56) then the infinitesimal price impact function is given by:*

$$\mathrm{IMP}(\delta) = \lim_{\tau \to 0} \frac{\mathrm{IMP}(\tau, \delta)}{\tau} \tag{3.61}$$

$$= \frac{\nu}{2\bar{\lambda}^u_\infty}(M_{11} - M_{12}) \tag{3.62}$$

$$= -\nu \frac{e^{-\delta(\beta + \alpha_m - \alpha_s)}\,(\alpha_m - \alpha_s)\,(2\beta + \alpha_m - \alpha_s)}{4\,(\beta + \alpha_m - \alpha_s)} \tag{3.63}$$

with $M = c_0(\delta)(\bar{\Lambda}_\infty + \alpha \mathrm{diag}(\bar{\lambda}_\infty))$.

It is often meaningful to consider the *cumulative* impact of a trade[7] on the stock return which to integrate with respect to δ the above quantity. Simple computations give:

[7] Notice that we call *trade* here an up or down jump of the mid price.

Proposition 3.15. *Given the function* IMP(δ) *of Proposition 3.14 then the cumulative price impact of trades up to a given time t is:*

$$\text{CIMP}(t) = \int_0^t \text{IMP}(\delta)d\delta \tag{3.64}$$

$$= \frac{\nu}{2\bar{\lambda}_\infty^u}(M_{11}(t) - M_{12}(t)) \tag{3.65}$$

$$= -\nu\frac{\left(1 - e^{-t(\beta+\alpha_m-\alpha_s)}\right)(\alpha_m - \alpha_s)(2\beta + \alpha_m - \alpha_s)}{4(\beta + \alpha_m - \alpha_s)^2} \tag{3.66}$$

with $M(t) = \int_0^t c_0(\delta)d\delta(\bar{\Lambda}_\infty + \alpha\text{diag}(\bar{\lambda}_\infty))$.

Note that if $\alpha_s > \alpha_m$ then the function is increasing and concave and in that case the shape is similar to the one obtained in Figure 1 of (Dufour and Engle, 2000). This condition denotes a clustering of trades coming from the split of a large order.

Clearly, a limitation of the above formulas is that we consider only quote-moving trades, which is rarely the case compared to the whole market activity as can be seen in (Pomponio and Abergel, 2013). Market impact has then to be studied in a more general setting and we do this in next section.

3.3.3 Market Impact in a Buy-Sell Toy Model

As stated before, another aspect of the trading activity presenting both clustering and mean reversion effects is the sell and buy order arrivals. Time arrivals of buy and sell orders can then be modelled by a bivariate Hawkes process. Let N_t^{Buy} and N_t^{Sell} be the cumulated number of buy and sell orders respectively, and let their respective intensities obey:

$$\lambda_t^{Buy} = \lambda_\infty + \int_0^t \alpha_s e^{-\bar{\beta}(t-u)}dN_u^{Buy} + \int_0^t \alpha_m e^{-\bar{\beta}(t-u)}dN_u^{Sell}$$

$$\lambda_t^{Sell} = \lambda_\infty + \int_0^t \alpha_m e^{-\bar{\beta}(t-u)}dN_u^{Buy} + \int_0^t \alpha_s e^{-\bar{\beta}(t-u)}dN_u^{Sell}.$$

The model is mathematically the same as before in the sense that the matrices α and β in (3.1) have very particular forms. Namely, they are given by:

$$\alpha = \begin{pmatrix} \alpha_s & \alpha_m \\ \alpha_m & \alpha_s \end{pmatrix}; \beta = \begin{pmatrix} \bar{\beta} & 0 \\ 0 & \bar{\beta} \end{pmatrix}$$

where α_s stands for the self excitation parameter and α_m stands for the mutual excitation parameter. The two intensities lead to a two-dimensional vector $\lambda_t = (\lambda_t^{Buy}, \lambda_t^{Sell})^\top$ whilst the two jump processes give the vector $N_t = (N_t^{Buy}, N_t^{Sell})^\top$.

Although the model is similar to the previous one we do not focus on the same phenomena. Here, we work at a more *microscopic* level than the mid price up and down jumps of the previous subsection because we are now interested in market orders (that might not lead to a mid-price move).

Notice that we choose symmetric processes in order for the price buy and sell pressures to be equal. As stated previously, this balance can be achieved even with some asymmetry in the parameters but we prefer to remain in the simplest possible setting.

In order to estimate the parameters of our model with real data we need to preprocess the TRTH files to build records of Buy and Sell transaction time arrivals. In fact, transaction files

of TRTH do not contain any flag specifying the sign of the transaction. We then take trade and quote files for the data as well as transaction files and apply the Lee and Ready algorithm (Lee and Ready, 1991) to sign trades, thus obtaining records of buy and sell order arrival times corresponding to our bivariate Hawkes process $N_t = (N_t^{Buy}, N_t^{Sell})$. Calibration is done using the MLE as before. We conduct a daily calibration over a 2-year data sample and we restrict our study to futures on Eurostoxx and Dax and two stocks: BNPP and Sanofi.

Results are reported in Table 3.3, where they seem to be stable and where likelihood ratio tests clearly reject the hypothesis of clustering only or mean reversion only.

Table 3.3: Calibration Results for Buy and Sell Orders

Symbol	Measure	λ_∞	α_s	α_m	β	Clustering	Mean Reversion
Eurostoxx	Mean	0.0907	0.5674	0.1395	0.7233	3367.51	52672.29
	Std. dev.	0.0229	0.1433	0.0598	0.1738	2047.06	22999.03
	Median	0.0908	0.5807	0.1410	0.7484	3396.00	49070.00
Dax	Mean	0.0938	0.5929	0.1588	0.7663	3496.19	47232.60
	Std. dev.	0.0235	0.1408	0.0613	0.1708	3054.18	20459.51
	Median	0.0952	0.6155	0.1695	0.8087	3666.00	43388.00
BNPP	Mean	0.0427	0.3383	0.0707	0.4263	2006.10	34381.42
	Std. dev.	0.0172	0.1358	0.0456	0.1672	1826.07	19204.96
	Median	0.0388	0.3129	0.0593	0.3888	1746.50	29710.00
Sanofi	Mean	0.0264	0.2063	0.0484	0.2738	1318.24	19197.74
	Std. dev.	0.0096	0.0814	0.0368	0.1099	1133.54	8615.12
	Median	0.0244	0.1918	0.0407	0.2494	1024.00	17409.50

Note. Calibration results for two years of data. We calibrate daily a bivariate Hawkes process to the time arrivals of buy and sell orders. The column Clustering presents the likelihood ratio statistic when we consider only clustering. The column Mean reversion presents the likelihood ratio statistic when we consider only mean reversion. For the likelihood ratio test as the difference between the models is one parameter the ratio has a $\chi^2(1)$ distribution, if $X \sim \chi^2(1)$ then we have $\mathbb{P}(X > 3.84) = 0.05$.

As the primary interest of order book events modelling is the price dynamics, let us consider a model of price formation based on buy and sell order arrival times. To remain simple, we suppose that every trade has a unit quantity. Following (Kyle, 1985), the long run price return is assumed to depend linearly on the imbalance between buy and sell trades. More precisely, we denote by $F_t = N_t^{Buy} - N_t^{Sell}$ the cumulated buy and sell orders imbalance at time t, the price at t writes:

$$P_t = \mathbb{E}_t[P_\infty] = P_0 + \lambda_{\text{K}} \mathbb{E}_t[F_\infty], \tag{3.67}$$

where we noted λ_{K} the coefficient of linear dependence of price return on buy-sell order imbalance. This coefficient is known as the *Kyle's Lambda*. We closely follow (Hewlett, 2006), where the author uses two independent univariate Hawkes processes for N_t^{Buy} and N_t^{Sell} and obtains a simple expression for the market impact function which specifies how a buy/sell order affects the stock price.

In order to calculate this market impact, and following (Bouchaud et al., 2004), we consider that the price at time t results from the superposition of market impacts of all the trades that took place prior to t. The market impact of a trade is attenuated as time passes. Formally, we have:

$$P_t = P_0 + \int_0^t I(t-u) dF_u. \tag{3.68}$$

The function $I(t)$, called the propagator, quantifies how buy and sell order imbalances impound into price and is the key element to determine the stock dynamic. Equating (3.67) and (3.68) we obtain:

$$\lambda_{\text{K}} F_t + \lambda_{\text{K}} \mathbb{E}_t[F_\infty - F_t] = \int_0^t I(t-u) dF_u. \tag{3.69}$$

Additionally, we have:

$$\mathbb{E}_t[F_\infty - F_t] = \int_t^\infty \mathbb{E}_t[\lambda_u^{Buy} - \lambda_u^{Sell}] du.$$

From (3.7) the expected intensity is equal to:

$$\mathbb{E}_t[\lambda_u] = (\alpha - \beta)^{-1}(e^{(\alpha-\beta)(u-t)} - I)\beta\lambda_\infty + e^{(\alpha-\beta)(u-t)}\lambda_t$$

where λ_t is the vector of buy and sell intensities (i.e. $\lambda_t = (\lambda_t^{Buy}, \lambda_t^{Sell})^\top$). Using this equality as well as the closed-form formulas for the matrix exponential and matrix inverse for the model considered we arrive at[8]

$$\mathbb{E}_t[F_\infty - F_t] = \int_t^\infty e^{-(\bar{\beta}+\alpha_m-\alpha_s)(u-t)}(\lambda_t^{Buy} - \lambda_t^{Sell}) du \tag{3.70}$$

$$= \frac{\lambda_t^{Buy} - \lambda_t^{Sell}}{\bar{\beta} + \alpha_m - \alpha_s}. \tag{3.71}$$

Combining (3.69) and (3.71) we deduce an integral representation for the propagator given by:

$$\begin{aligned}
\int_0^t I(t-u) dF_u &= \lambda_{\text{K}} F_t + \lambda_{\text{K}} \mathbb{E}_t[F_\infty - F_t] \\
&= \lambda_{\text{K}} F_t + \lambda_{\text{K}} \frac{\lambda_t^{Buy} - \lambda_t^{Sell}}{\bar{\beta} + \alpha_m - \alpha_s} \\
&= \lambda_{\text{K}}(N_t^{Buy} - N_t^{Sell}) + \lambda_{\text{K}} \frac{\lambda_t^{Buy} - \lambda_t^{Sell}}{\bar{\beta} + \alpha_m - \alpha_s}.
\end{aligned}$$

Knowing that:

$$\lambda_t^{Buy} = \lambda_\infty + \int_0^t \alpha_s e^{-\bar{\beta}(t-u)} dN_u^{Buy} + \int_0^t \alpha_m e^{-\bar{\beta}(t-u)} dN_u^{Sell}$$

and

$$N_t^{Buy} = \int_0^t dN_u^{Buy},$$

as well as similar computations for the sell side we are led to:

$$\begin{aligned}
\int_0^t I(t-u) dF_u &= \lambda_{\text{K}}\Big(\int_0^t 1 + \frac{(\alpha_s - \alpha_m)e^{-\bar{\beta}(t-u)}}{\bar{\beta} + \alpha_m - \alpha_s}(dN_u^{Buy} - dN_u^{Sell})\Big) \\
&= \lambda_{\text{K}}\Big(\int_0^t 1 + \frac{(\alpha_s - \alpha_m)e^{-\bar{\beta}(t-u)}}{\bar{\beta} + \alpha_m - \alpha_s} dF_u\Big).
\end{aligned}$$

Therefore, the propagator which gives the impact of a single trade executed at time 0 as seen at time t is:

$$I(t) = \lambda_{\text{K}}\left(1 + \frac{(\alpha_s - \alpha_m)e^{-\bar{\beta}t}}{\bar{\beta} + \alpha_m - \alpha_s}\right), \tag{3.72}$$

[8] It is indeed easy to establish that for the matrix $\alpha - \beta$ considered we have:

$$(\alpha - \beta)^{-1} = \left(\alpha_m^2 - (\bar{\beta} - \alpha_s)^2\right)\begin{pmatrix} \bar{\beta} - \alpha_s & \alpha_m \\ \alpha_m & \bar{\beta} - \alpha_s \end{pmatrix}$$

$$e^{(\alpha-\beta)t} = \frac{1}{2} e^{-t(\beta+\alpha_m-\alpha_s)}\begin{pmatrix} 1 + e^{2t\alpha_m} & -1 + e^{2t\alpha_m} \\ -1 + e^{2t\alpha_m} & 1 + e^{2t\alpha_m} \end{pmatrix}$$

where one can see that this impact decomposes into a permanent impact effect and a transient effect.

This result enables us to study the impact of a large order that is executed by splitting it into consecutive single-share executions. We have the following result whose proof is omitted:

Proposition 3.16. *Denote by $\mathcal{I}(T)$ the price impact at time T of a continuum of orders executed over the interval $[0;t]$ then by definition this function is given by:*

$$\mathcal{I}(T) = \int_0^t I(T-s)ds \tag{3.73}$$

where the function I is (3.72). For the model considered the integral leads to:

$$\mathcal{I}(T) = \begin{cases} \lambda_{\text{K}} \left(T + \frac{(\alpha_s - \alpha_m)}{(\beta + \alpha_m - \alpha_s)\beta} (1 - e^{-\bar{\beta}T}) \right) & T \leq t \\ \lambda_{\text{K}} \left(t + \frac{(\alpha_s - \alpha_m)}{(\beta + \alpha_m - \alpha_s)\beta} (e^{-\bar{\beta}(T-t)} - e^{-\bar{\beta}T}) \right) & T \geq t. \end{cases}$$

We graphically illustrate this function, Figure 3.5 presents a plot of the market impact of an order to buy 50 shares of stock, executed during 50 seconds at the rate of 1 stock per second. One clearly sees that the impact increases during the execution time of the order, and then decreases steadily to attain the permanent impact.

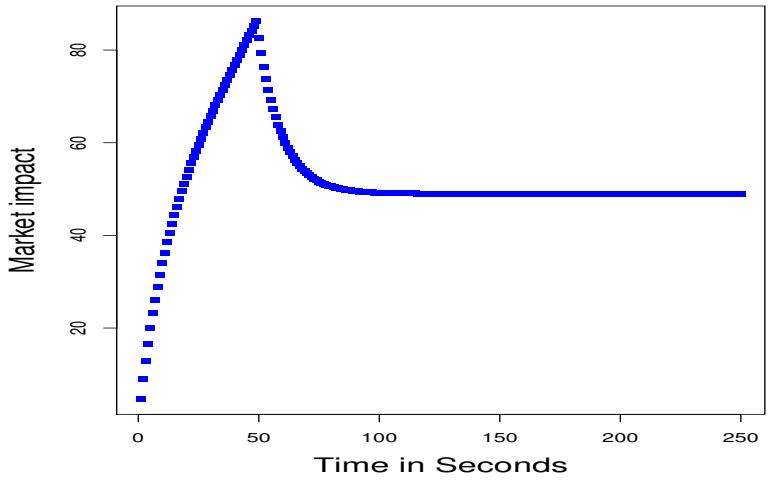

Figure 3.5: Market impact of an order to buy 50 shares of stock, executed during 50 seconds at the rate of 1 stock per second. We took the scaling parameter $\lambda_{\text{K}} = 1$.

Notice that in order to obtain the concave shape during the execution phase, we must have $\alpha_s > \alpha_m$. Not only is this coherent by a clustering of orders during this phase, due to the executing agent, but it is also coherent with market activity in general where clustering systematically dominates mean reversion when it comes to buy/sell activity modelling as is clear from Table 3.3.

The assumption of unit share individual executions can be relaxed by introducing the notion of *execution profile* as defined for example in (Moro et al., 2009).

We define the *execution profile* as a function of time f with support on the interval $[0;t]$ with $\int_0^t f(s)ds = Q$ which is the total quantity to be executed. With this definition, the impact of the meta order writes in integral notation:

$$\mathcal{I}(T) = \int_0^T I(T-s)f(s)ds$$

which is just the convolution of the execution profile and market impact.

Finally, let us notice that after developing models for market microstructure (first application in this section) and market impact (second application in this section), one wants to unify these dynamics in one model. This was recently done in a seminal article by (Bacry and Muzy, 2013). What is more, the authors rely on Hawkes processes with generic kernels and identify, among other things, feedback effects of price changes on trading activity. We think that this new point of view will greatly influence our (and other's) research in the subject.

3.4 Conclusion

In this chapter we explicitly compute the first and second moments and the autocovariance function of the number of jumps over a given time interval of a multivariate Hawkes process. These computations are possible thanks to the affine property of the process and the use of the Dynkin formula.

Using these quantities we compute several statistical properties for a stock dynamics based on the Hawkes process. It allows us to unify the pioneering model of (Bacry et al., 2013a) with the model developed in the previous chapter. The first one possesses a mean reverting behaviour whilst the second one has a clustering behaviour. We explicitly compute the signature plot and analyze the impact of the parameters on the shape of this function. Furthermore, we compute the diffusive limit which enables the connection of the parameters driving the price at high-frequency with the parameter driving the stock price a low frequency (i.e. the daily Black-Scholes volatility). Lastly, we explicitly compute the impulse response function which quantifies the impact of a quote-moving trade on the stock price. For all these results we can analyze the impact of the mean reverting and clustering parameters and find results consistent with intuition.

We further exploit the results by analyzing the price impact of a trade in a Buy and Sell toy model based on the Hawkes process. We provide an explicit expression for this function and extend the results by computing the impact of a collection of trades. This allows us to recover the typical concave-shape/relaxation pattern of the execution impact function of a meta order.

Although some of the formulas are derived for a specific choice of the matrix parameters most of the results are valid for general matrices (to the extent that some stability conditions are satisfied). Within this framework it is possible to consider other important market microstructure problems. As an example, we investigate in the next chapter the correlation and the lead-lag relationship between two assets and analyze them through the lenses of Hawkes processes.

Chapter 4

Correlation and Lead-Lag Relationships in a Hawkes Microstructure Model

Abstract

Our aim in this chapter is to develop a multi-asset model based on the Hawkes process describing the evolution of the assets at high frequency and to study the correlation as well as the lead-lad relationships between the stocks within this framework. Thanks to the strong analytical tractability allowed by the use of Hawkes process, several statistical quantities are explicitly computed and some insight is given on the impact of the model parameters on these quantities. Furthermore, we compute the covariance matrix associated with the diffusive limit of the model so that the relation between the parameters driving the assets at high and low frequencies is explicit. We illustrate our results using index futures and stocks quoted in the Eurex market, for which the model proves efficient in capturing the existing lead-lag relationships between the assets.

Contents

Introduction

The interaction between stocks is an important aspect of financial theory. From optimal portfolio choice to basket option pricing, one of the key ingredients is the modelling of the dependence between stocks. The correlation appears to be the natural mathematical concept to handle the interaction and in fact underlies many financial models. The correlation provides information on contemporaneous evolutions but in the markets this simultaneousness can be too stringent. A concept that relaxes this hypothesis is the lead-lag relationship which has also been extensively studied in the literature, among many others let us mention (Herbst et al., 1987). More recently, the availability of high-frequency data also triggered research on this subject as the works of (de Jong and Nijman, 1997) and (Huth and Abergel, 2012) attest.

The purpose of this work is to develop a multi asset model for stocks based on the Hawkes process and to study the correlation and lead-lag relationships within this framework. The model specifies the high-frequency dynamics for the stocks. Our work heavily relies on the dynamics proposed by (Bacry et al., 2013a). In the particular case of exponential kernel for the Hawkes process, which allows for very explicit computations, we develop a model for which many of the statistical properties of the stocks can be computed. Using the theoretical results of (Bacry et al., 2013b) we compute the diffusive limit for the stocks thereby connecting the model driving the assets at high frequency with the covariance matrix driving the assets at low frequency (i.e., daily). Within this framework we analyze the correlation as well as the lead-lag relationships between the stocks and provide some expansions to better understand the impact of the model parameters on different financial quantities. We then perform an empirical analysis on two index futures and four major stocks quoted on the Eurex market to illustrate the model. Lastly, within our framework extracting a lead-lag relationship between stocks at low frequency is problematic and confirms the results of (Huth and Abergel, 2012).

The structure of this chapter is as follows. In the first section, we describe the analytical framework giving a dynamic for two stocks based on the Hawkes process and derive various statistical quantities as well as the expression for the diffusive limit associated with the two stocks. In the second section, an empirical analysis is performed to illustrate the capabilities of the model.

4.1 The Bacry-Delattre-Hoffmann-Muzy Model

4.1.1 The stock dynamics

As in the previous chapters, we adopt the modeling framework proposed by (Bacry et al., 2013a) to describe the evolution of the *mid* price of two traded assets:

$$S_t^1 = S_0^1 + \left(N_t^{1,u} - N_t^{1,d}\right)\frac{\nu_1}{2},$$
$$S_t^2 = S_0^2 + \left(N_t^{2,u} - N_t^{2,d}\right)\frac{\nu_2}{2}$$

where ν_1 and ν_2 are the tick values for the first stock and second stock, respectively. Let, $N_t^{1,u}$ and $N_t^{1,d}$ be the Hawkes processes capturing the up and down jumps of the mid price for the first stock and $N_t^{2,u}$ and $N_t^{2,d}$ the corresponding equivalent for the second stock.

In this work the four dimensional Hawkes process $N_t = (N_t^{1,u}, N_t^{1,d}, N_t^{2,u}, N_t^{2,d})^\top$ and $\lambda_t = (\lambda_t^{1,u}, \lambda_t^{1,d}, \lambda_t^{2,u}, \lambda_t^{2,d})^\top$ follow a dynamic of the form (3.1) with:

$$\alpha = \begin{pmatrix} \alpha_s^1 & \alpha_m^1 & x & 0 \\ \alpha_m^1 & \alpha_s^1 & 0 & x \\ y & 0 & \alpha_s^2 & \alpha_m^2 \\ 0 & y & \alpha_m^2 & \alpha_s^2 \end{pmatrix}; \beta = \begin{pmatrix} \bar{\beta}_1 & 0 & 0 & 0 \\ 0 & \bar{\beta}_1 & 0 & 0 \\ 0 & 0 & \bar{\beta}_2 & 0 \\ 0 & 0 & 0 & \bar{\beta}_2 \end{pmatrix} \tag{4.1}$$

and $\lambda_\infty = (\lambda_{1,\infty}, \lambda_{1,\infty}, \lambda_{2,\infty}, \lambda_{2,\infty})^\top \in \mathbb{R}_+^4$.

The connection between the two stocks is controlled through the 2×2 upper-right and lower-left sub-matrices (called coupling submatrices in the sequel). Here, we supposed that the two stocks have an overall positive correlation, hence our parametrization. Notice that if $x = y = 0$ then the two stocks move independently.

For $x > 0$, an up move of the second stock, through a jump of $N_t^{2,u}$, induces an increase of the intensity $\lambda_t^{1,u}$ which increases the probability of a jump of $N_t^{1,u}$ over the next time period and therefore an up move of the first stock. Similar reasoning applies to a down move. As a consequence, the movements of the second stock will be reproduced by the first stock. By construction these related evolutions are *not simultaneous*, and will lead to lead lag relationship when observed at the adequate time scale. In the specific case where $x > 0$ and $y = 0$, the second asset leads the first asset and we can qualify as a positive lead-lag relationship, as the stocks will move in the same direction and that their overall correlation will be positive.

We restrict ourselves to a very particular form for the coupling submatrices matrices in order to make as much as possible explicit the dependency between the stocks. For example, the 2×2 upper-right matrix of α could be replaced with:

$$\begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}, \tag{4.2}$$

then the impact of an up move of the second stock on the first one will be different from the impact of a down move (of the second asset on the first asset). A priori this is an appealing feature because linkages between stocks are indeed different in bear and bull markets. However, to keep the analytical expressions simple we restrict this study to the symmetric model (i.e., $x_1 = x_2$).

If we wish to consider negative lead-lag relationship of the second stock on the first one, that is to say a resulting negative correlation, then the 2×2 upper-right sub-matrix of α should be:

$$\begin{pmatrix} 0 & x \\ x & 0 \end{pmatrix}, \tag{4.3}$$

also with the possibility to differentiate up and down moves.

Similar considerations apply to 2×2 lower-left sub-matrix of α that controls the transmission of the shocks affecting the first stock to the second stock. Also, it is worth underlining the possibility of disymmetric effects between the two stocks in the sense that the impact of the first stock on the second one need not to be equal to, nor in the same direction of the impact of the first stock on the second. This strongly contrasts with the usual correlation which by definition is a reflexive relation. Although the model allows for a disymmetric relationship between the stocks it seems to us natural to choose the upper-right matrix consistently with the lower-left matrix. For the choice made for the 2×2 upper-right matrix of α in (4.1) the natural choice is the one made for the lower-left sub-matrix whereas in the case of (4.3) then the natural choice is:

$$\begin{pmatrix} 0 & y \\ y & 0 \end{pmatrix}. \tag{4.4}$$

It would be even tempting to choose $x = y$ so that the lead-lag relationship would be reflexive. However, there are some practical cases where allowing for some disymmetry, that is to say different values for x and y, is particularly relevant; the most well-known example being the relation between a futures on an index and a stock.

Following (Bacry et al., 2013a) we will be interested in the diffusive limit for $S_t = (S_t^1, S_t^2)^\top$ associated with the dynamics (4.1). In order to perform such a limit computation the matrices α and β in (4.1) are such that $\mathbb{E}[N_t^{1,u}] = \mathbb{E}[N_t^{1,d}] = \mathbb{E}[N_t^{2,u}] = \mathbb{E}[N_t^{2,d}]$, this ensures the martingale property for the stocks. This martingale property can be obtained for more general matrices α and β but we restrict this study these particular forms to keep the expressions simple.

Thanks to the computations made in the previous chapter, we know that the matrix $\alpha - \beta$ must have negative eigenvalues for the multivariate Hawkes process to be stable, which translates to:

$$\gamma_1 + \gamma_2 \pm \left((\gamma_1 + \gamma_2)^2 - 4(\gamma_1\gamma_2 - xy)\right)^{\frac{1}{2}} > 0, \tag{4.5}$$

$$\theta_1 + \theta_2 \pm \left((\theta_1 + \theta_2)^2 - 4(\theta_1\theta_2 - xy)\right)^{\frac{1}{2}} > 0 \tag{4.6}$$

with $\gamma_i = \bar{\beta}_i + \alpha_m^i - \alpha_s^i$, $\theta_i = \bar{\beta}_i - \alpha_m^i - \alpha_s^i$ for $i \in \{1, 2\}$. Conditions ensuring these inequalities are $\gamma_1\gamma_2 > xy$ and $\theta_1\theta_2 > xy$ that, from now on, we suppose satisfied.

4.1.2 Statistical properties

Having sprecified the dynamics for the stocks we focus on the computation of various statistical properties associated with the assets. The use of high-frequency data enables the computation of the realized volatility and the estimator, for data sampled using time intervals of length τ, is written as:

$$\begin{aligned}
\hat{C}_1(\tau) &= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(S^1_{(n+1)\tau} - S^1_{n\tau}\right)^2 = \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(\left(N^{1,u}_{(n+1)\tau} - N^{1,u}_{n\tau}\right) - \left(N^{1,d}_{(n+1)\tau} - N^{1,d}_{n\tau}\right)\right)^2\frac{\nu_1^2}{4} \\
&= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(N^{1,u}_{(n+1)\tau} - N^{1,u}_{n\tau}\right)^2\frac{\nu_1^2}{4} + \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(N^{1,d}_{(n+1)\tau} - N^{1,d}_{n\tau}\right)^2\frac{\nu_1^2}{4} \\
&- 2\frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(N^{1,u}_{(n+1)\tau} - N^{1,u}_{n\tau}\right)\left(N^{1,d}_{(n+1)\tau} - N^{1,d}_{n\tau}\right)\frac{\nu_1^2}{4}.
\end{aligned}$$

The mean signature plot, or more simply signature plot, is the expectation of the above quantity and is explicitly given by:

Proposition 4.1. *The signature plot for the first asset $C_1(\tau) = \mathbb{E}[\hat{C}_1(\tau)]$ is:*

$$C_1(\tau) = \frac{\nu_1^2}{4\tau}\left(M_{11} + M_{22} - M_{12} - M_{21}\right) \tag{4.7}$$

where $M = \mathrm{Cov}(\tau)$ is the second moment matrix calculated in Lemma 3.4.

The volatility of the second stock is obtained by the same calculations:

$$C_2(\tau) = \frac{\nu_2^2}{4\tau}\left(M_{33} + M_{44} - M_{34} - M_{43}\right). \tag{4.8}$$

Such a quantity was calculated in the previous chapter when a stock was modeled alone. Nevertheless, it remains interesting to calculate it here, within a multi-asset model. For instance, let S_t^1 be a stock and S_t^2 be an index future that captures the overall market evolution. It can then be interesting to assess the part of stock volatility that can be explained by market activity through the coupling parameters x and y. We will make this more clear in the sequel.

As we deal with a multi-asset model, the covariance between the two stocks is the quantity of interest and can be obtained by the usual estimator:

$$\begin{aligned}\widehat{\mathrm{CovS}}(\tau) &= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(S^1_{(n+1)\tau}-S^1_{n\tau}\right)\left(S^2_{(n+1)\tau}-S^2_{n\tau}\right)\\ &= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\Big[\left(N^{1,u}_{(n+1)\tau}-N^{1,u}_{n\tau}\right)\left(N^{2,u}_{(n+1)\tau}-N^{2,u}_{n\tau}\right)-\left(N^{1,u}_{(n+1)\tau}-N^{1,u}_{n\tau}\right)\left(N^{2,d}_{(n+1)\tau}-N^{2,d}_{n\tau}\right)\\ &- \left(N^{1,d}_{(n+1)\tau}-N^{1,d}_{n\tau}\right)\left(N^{2,u}_{(n+1)\tau}-N^{2,u}_{n\tau}\right)+\left(N^{1,d}_{(n+1)\tau}-N^{1,d}_{n\tau}\right)\left(N^{2,d}_{(n+1)\tau}-N^{2,d}_{n\tau}\right)\Big]\frac{\nu_1\nu_2}{4}.\end{aligned}$$

This estimator along with the analytical results, mainly the second moment of Lemma 3.4, lead to an explicit expression for the covariance as the following proposition shows:

Proposition 4.2.

$$\mathrm{CovS}(\tau) = \frac{\nu_1\nu_2}{4\tau}\Big[M_{13}-M_{14}-M_{23}+M_{24}\Big] = \frac{\nu_1\nu_2}{4\tau}\Big[M_{31}-M_{41}-M_{32}+M_{42}\Big] \tag{4.9}$$

with the matrix $M = \mathrm{Cov}(\tau)$ *given by* (3.15) *of Lemma* 3.4. *The last equality in* (4.9) *stands from the symmetry of the matrix* M.

Rescaling by the above calculated volatilities of the stocks, one obtains an estimator of the correlation as a function of the sampling period τ, and retrieves the well known Epps effect. Recall that this effect materializes in the decrease of the estimated correlation when the sampling frequency increases. An illustration in this model is given in Figure 4.1.

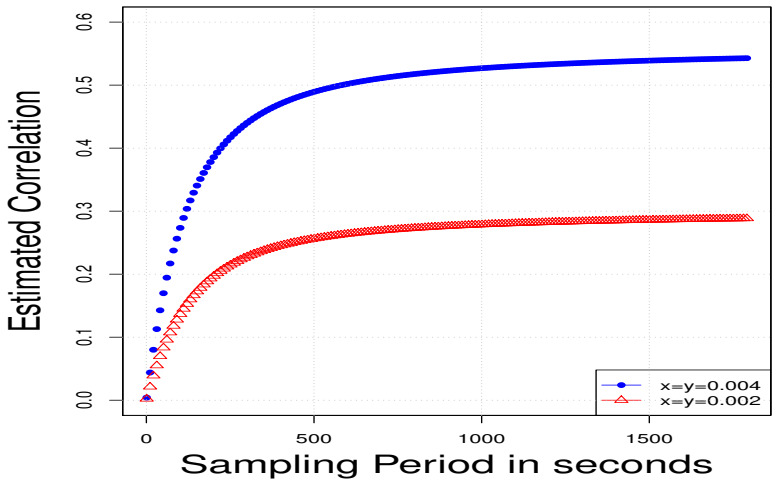

Figure 4.1: Illustration of the Epps effect reconstruction. We considered perfectly symmetric stocks, with $\alpha_s = \alpha_m = 0.004$ and considered two values for the coupling coefficients x and y. Notice that when these latter are halved, the asymptotic correlation is always roughly halved. Notice also that the time taken to attain the asymptotic correlation does not depend on the coupling coefficients.

Beyond the correlation between the two stocks is the lead-lag relationship. The estimator of the lagged covariance between the two stocks if we consider S^2 as the leader is:

$$\begin{aligned}\widehat{\mathrm{L}}_{2\to1}(\tau,\delta) &= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(S^1_{(n+1)\tau+\delta}-S^1_{n\tau+\delta}\right)\left(S^2_{(n+1)\tau}-S^2_{n\tau}\right)\\ &= \frac{1}{T}\sum_{n=0}^{T/\tau-1}\Big[\left(N^{1,u}_{(n+1)\tau+\delta}-N^{1,u}_{n\tau+\delta}\right)\left(N^{2,u}_{(n+1)\tau}-N^{2,u}_{n\tau}\right)-\left(N^{1,u}_{(n+1)\tau+\delta}-N^{1,u}_{n\tau+\delta}\right)\left(N^{2,d}_{(n+1)\tau}-N^{2,d}_{n\tau}\right)\\ &- \left(N^{1,d}_{(n+1)\tau+\delta}-N^{1,d}_{n\tau+\delta}\right)\left(N^{2,u}_{(n+1)\tau}-N^{2,u}_{n\tau}\right)+\left(N^{1,d}_{(n+1)\tau+\delta}-N^{1,d}_{n\tau+\delta}\right)\left(N^{2,d}_{(n+1)\tau}-N^{2,d}_{n\tau}\right)\Big]\frac{\nu_1\nu_2}{4}.\end{aligned}$$

for $\delta > 0$. Conversely, if the first stock is taken as the leader the estimator becomes:

$$\widehat{\mathrm{L}}_{1\to 2}(\tau,\delta) = \frac{1}{T}\sum_{n=0}^{T/\tau-1}\left(S^2_{(n+1)\tau+\delta} - S^2_{n\tau+\delta}\right)\left(S^1_{(n+1)\tau} - S^1_{n\tau}\right).$$

These two estimators combined with the results developed in the analytical part of previous chapter give the following proposition:

Proposition 4.3. *Given the two estimators defined above then estimated lagged covariances, depend on which asset is chosen as leader, write as:*

$$\mathrm{L}_{2\to 1}(\tau,\delta) = \frac{\nu_1\nu_2}{4\tau}\Big[\mathrm{Cov}(\tau,\delta)_{13} - \mathrm{Cov}(\tau,\delta)_{14} - \mathrm{Cov}(\tau,\delta)_{23} + \mathrm{Cov}(\tau,\delta)_{24}\Big], \tag{4.10}$$

$$\mathrm{L}_{1\to 2}(\tau,\delta) = \frac{\nu_1\nu_2}{4\tau}\Big[\mathrm{Cov}(\tau,\delta)_{31} - \mathrm{Cov}(\tau,\delta)_{32} - \mathrm{Cov}(\tau,\delta)_{41} + \mathrm{Cov}(\tau,\delta)_{42}\Big] \tag{4.11}$$

for $\tau > 0$ and $\delta > 0$ whilst $\mathrm{Cov}(\tau,\delta)$ stands for $\mathrm{Cov}_1(\tau,\tau,\delta)$ or $\mathrm{Cov}_2(\tau,\tau,\delta)$ (these two quantities given by (3.27) and (3.33)) depending on how δ compares with τ.

From these two quantities we can define a lead-lag correlation between the two stocks for $(\tau,\delta) \in \mathbb{R}_+ \times \mathbb{R}$ as:

$$\mathcal{C}(\tau,\delta) = \begin{cases} \frac{\mathrm{L}_{2\to 1}(\tau,\delta)}{\sqrt{C_1(\tau)C_2(\tau)}} & \text{if} \quad \delta > 0, \\ \frac{\mathrm{L}_{1\to 2}(\tau,-\delta)}{\sqrt{C_1(\tau)C_2(\tau)}} & \text{if} \quad \delta < 0. \end{cases}$$

An illustration of the lagged correlation is given in Figure 4.2.

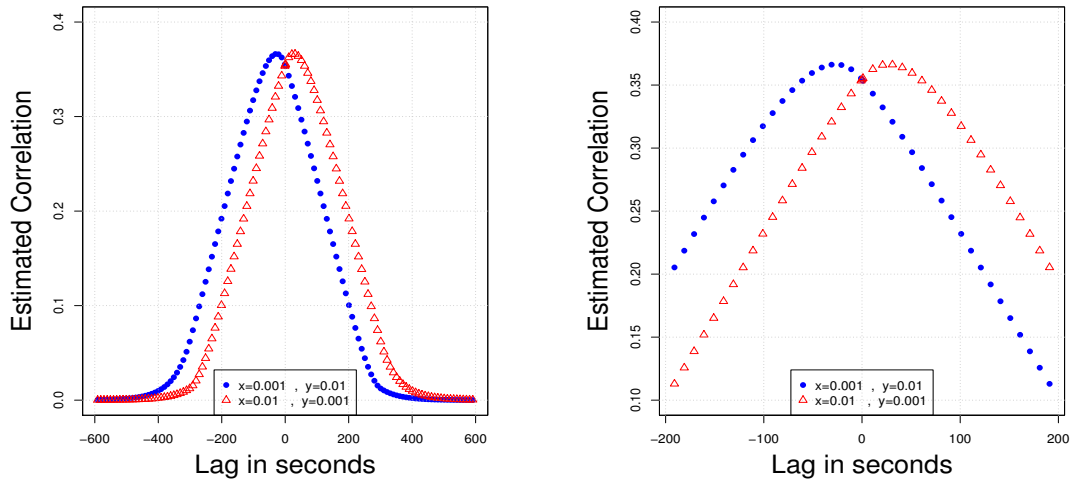

Figure 4.2: Illustration of the asymmetry of the correlation induced by a Hawkes process. We considered perfectly symmetric stocks, with $\alpha^1_s = \alpha^2_s = \alpha^1_m = \alpha^2_m = 0.004$ and considered two values for the coupling coefficients x and y. The figures report $\delta \to \mathcal{C}(\tau,\delta)$ for $\delta \in \mathbb{R}$ with the function $\mathcal{C}$ defined in Proposition 4.3 for two pairs of values for (x,y); the first one is $(0.001, 0.01)$ and gives the blue-bullet curve, the second one is $(0.01, 0.001)$ and gives the red-triangle curve. In the right figure, a zoom around zero lag is performed to help to visualize the fact that maximum correlation is achieved for a certain lag.

Suppose that $x > 0$ and $y = 0$ then an up jump of the second asset will induce an increase of $\lambda^{1,u}_t$ which increases the probability of an up jump of the first asset that will occur, if it occurs, with a delay. As a result, the function $\delta \to \mathcal{C}(\tau,\delta)$ should be increasing at the vicinity of 0_+.

Conversely, if $x = 0$ and $y > 0$ then an up move of the first asset will imply an increase of $\lambda_t^{2,u}$ which increases the probability of an up jump of the second asset that will occur, if it occurs, with a delay. In that case the function $\delta \to \mathcal{C}(\tau, \delta)$ will be decreasing at 0_-. Clearly, the function $\delta \to \mathcal{C}(\tau, \delta)$ should converge to zero for $|\delta|$ large and we expect a humped shape. If the effect provoked by x dominates then the maximum of $\delta \to \mathcal{C}(\tau, \delta)$ should be for $\delta > 0$, and in that case the second asset leads the first one, whereas if the y effect dominates we expect the opposite, that is to say, a maximum for δ negative and in that case the first asset is the leader. Lastly, the two effects can cancel each other and it leads to a centred function, none of the assets leads the other.

These equations enable to numerically compute the different statistics but if we wish to develop a better understanding of the model then an expansion with respect to the parameters can be performed. This task is carried out in the next subsection.

Expansion with respect to coupling parameters

As we are primarily interested in the interaction between the two stocks the expansion will be the done with respect to x and y, that we qualify as coupling or lead-lag parameters. We first focus on the impact of the coupling parameters on the signature plot of the assets, the following proposition, which proof is postponed to the Appendix, explains the dependency with respect to these parameters:

Proposition 4.4. *Given the signature plot for the first asset of Proposition 4.1 then we have the following first order Taylor expansion:*

$$C(\tau) = C_0(\tau) + xC_x(\tau) + yC_y(\tau) \tag{4.12}$$

where:

$$\begin{aligned} C_0(\tau) &= \frac{\nu_1^2}{2}\Lambda_1 \left(\kappa_1^2 + \left(1 - \kappa_1^2\right) \frac{\left(1 - e^{-\tau\gamma_1}\right)}{\gamma_1 \tau}\right), \\ C_x(\tau) &= \frac{\nu_1^2}{2}\frac{\Lambda_2}{\theta_1} \left(\kappa_1^2 + \left(1 - \kappa_1^2\right) \frac{\left(1 - e^{-\tau\gamma_1}\right)}{\gamma_1 \tau}\right), \\ C_y(\tau) &= 0 \end{aligned}$$

and

$$\Lambda_i = \frac{\bar{\beta}_i \lambda_{i,\infty}}{\theta_i}, \quad \kappa_i = \frac{\bar{\beta}_i}{\gamma_i}, \tag{4.13}$$

with θ_i and γ_i defined for $i \in \{1, 2\}$ were previously defined.

The expression of $C_0(\tau)$ appears in the previous chapter. Note that $C_y(\tau) = 0$. Of importance is the fact that the functions $\tau \to C_0(\tau)$ and $\tau \to C_x(\tau)$ have the same shape which implies that the coupling parameter does not alter the shape of the signature plot. It will be decreasing in the purely mutually excited case, i.e. $\alpha_s^1 = 0$ as in Bacry et al. (2013a), or increasing in the pure clustering case, i.e $\alpha_m^1 = 0$ as in our second chapter, or could be even flat if $\alpha_s^1 = \alpha_m^1$.

Having computed the expansion of the signature plot similar computations give the expression for the expansion of the covariance as we have:

Proposition 4.5.

$$\mathrm{Cov}(\tau) = \frac{\nu_1 \nu_2}{4} \left(\mathrm{Cov}_0(\tau) + x\mathrm{Cov}_x(\tau) + y\mathrm{Cov}_y(\tau)\right) \tag{4.14}$$

where:

$$
\begin{aligned}
\mathrm{Cov}_0(\tau) &= 0 \\
\mathrm{Cov}_x(\tau) &= \frac{2\Lambda_2}{\tau\gamma_1^2\gamma_2^3(\gamma_1+\gamma_2)}\Big((\alpha_m^2-\alpha_s^2)(\bar{\beta}_2+\gamma_2)\gamma_1^2-(1-\tau\gamma_1)\bar{\beta}_2^2\gamma_2(\gamma_1+\gamma_2)\Big) \\
&+ \frac{\Lambda_2}{\tau\gamma_1^2\gamma_2(\gamma_1+\gamma_2)}\Big(2\gamma_1\gamma_2+2\bar{\beta}_2\gamma_2+(\alpha_m^2-\alpha_s^2)^2\Big)e^{-\gamma_1\tau} \\
&- \frac{\Lambda_2(\bar{\beta}_2+\gamma_2)(\alpha_m^2-\alpha_s^2)}{\tau\gamma_1^2\gamma_2^3(\gamma_1+\gamma_2)}\Big(\gamma_2^2+2\gamma_1(\gamma_1+\gamma_2)+\tau\gamma_2^2(\gamma_1+\gamma_2)\Big)e^{-\gamma_2\tau} \\
\mathrm{Cov}_y(\tau) &= \frac{2\Lambda_1}{\tau\gamma_1^3\gamma_2^2(\gamma_1+\gamma_2)}\Big((\alpha_m^1-\alpha_s^1)(\bar{\beta}_1+\gamma_1)\gamma_2^2-(1-\tau\gamma_2)\bar{\beta}_1^2\gamma_1(\gamma_1+\gamma_2)\Big) \\
&+ \frac{\Lambda_1}{\tau\gamma_1\gamma_2^2(\gamma_1+\gamma_2)}\Big(2\gamma_1\gamma_2+2\bar{\beta}_1\gamma_1+(\alpha_m^1-\alpha_s^1)^2\Big)e^{-\gamma_2\tau} \\
&- \frac{\Lambda_1(\bar{\beta}_1+\gamma_1)(\alpha_m^1-\alpha_s^1)}{\tau\gamma_1^3\gamma_2^2(\gamma_1+\gamma_2)}\Big(\gamma_1^2+2\gamma_2(\gamma_1+\gamma_2)+\tau\gamma_1^2(\gamma_1+\gamma_2)\Big)e^{-\gamma_1\tau}.
\end{aligned}
$$

Remark 4.6. *It is of interest to expand the lead-lag relation given by Proposition 4.3 but it leads to equations far too large.*

4.1.3 The diffusive limit behaviour

In order to compute the diffusive limit of the model we use the important Corollary 1 of Bacry et al. (2013b) which gives a central limit theorem:

Proposition 4.7. *Let N_t the four-dimensional Hawkes process then for $t \in [0;1]$*

$$\frac{N_{nt}}{\sqrt{n}} - \sqrt{n}t\bar{\lambda}_\infty \tag{4.15}$$

converge in law for the Skorohod topology to

$$(\beta-\alpha)^{-1}\beta\Sigma^{\frac{1}{2}}W_t \tag{4.16}$$

with $\{W_t; t \geq 0\}$ a Brownian motion and Σ the diagonal matrix with the i^{th} element given by $((\beta-\alpha)^{-1}\beta\lambda_\infty)_i$.

Let us then write unit-time price increments for the first asset as similar results apply to the second asset:

$$\eta_i^1 = \Big[(N_i^{1,u}-N_{i-1}^{1,u})-(N_i^{1,d}-N_{i-1}^{1,d})\Big]\times\frac{\nu_1}{2},$$

and consider the random sums

$$S_n^1 = \sum_{i=1}^{n}\eta_i^1,$$

Denote by

$$\bar{S}_t^{1,n} = \frac{S_{\lfloor nt\rfloor}^1}{\sqrt{n}}.$$

and let $\bar{S}_t^n = (\bar{S}_t^{1,n}, \bar{S}_t^{2,n})^\top$ the vector of the two stocks. Thanks to Proposition 4.7 we obtain the diffusive limit for the stocks as we have:

Proposition 4.8. *The vector $\bar{S}_t^n$ converge in law to the vector $\tilde{S}_t = (\tilde{S}_t^1, \tilde{S}_t^2)^\top$ whose dynamic is given by:*

$$
\begin{aligned}
d\tilde{S}_t^1 &= \frac{\nu_1}{2}\sum_{j=1}^{4}(m_{1j}-m_{2j})dW_t^j, \\
d\tilde{S}_t^2 &= \frac{\nu_2}{2}\sum_{j=1}^{4}(m_{3j}-m_{4j})dW_t^j
\end{aligned}
$$

with m the matrix appearing in equation (4.16).

The stock price increments follow a Gaussian distribution, in the literature this model is usually referred to as the Bachelier model. From this result we deduce the covariance matrix of the assets

Proposition 4.9. *Let the 2×2 covariance matrix of the assets such that:*

$$\sigma\sigma^\top = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix}$$

then we have:

$$\sigma_{11}^2 = \frac{\nu_1^2}{4} \sum_{j=1}^{4} (m_{1j} - m_{2j})^2, \tag{4.17}$$

$$\sigma_{12} = \frac{\nu_1 \nu_2}{4} \sum_{j=1}^{4} (m_{1j} - m_{2j})(m_{3j} - m_{4j}), \tag{4.18}$$

$$\sigma_{22}^2 = \frac{\nu_2^2}{4} \sum_{j=1}^{4} (m_{3j} - m_{4j})^2 \tag{4.19}$$

that are explicitly given by

$$\sigma_{11}^2 = \frac{\nu_1^2}{4} \frac{2\lambda_{1,\infty}(\bar{\beta}_1 \bar{\beta}_2^2 x^2 y + \bar{\beta}_1^3 \theta_2 \gamma_2^2) + \lambda_{2,\infty} \bar{\beta}_2 x (\bar{\beta}_2^2 x \theta_1 + \bar{\beta}_1^2 \gamma_2^2)}{(\gamma_1 \gamma_2 - xy)^2 (\theta_1 \theta_2 - xy)}, \tag{4.20}$$

$$\sigma_{12} = \frac{\nu_1 \nu_2}{4} \frac{2\bar{\beta}_1 y \lambda_{1,\infty} (\bar{\beta}_2^2 x \gamma_1 + \bar{\beta}_1^2 \theta_2 \gamma_2) + 2\bar{\beta}_2 x \lambda_{2,\infty} (\bar{\beta}_1^2 y \gamma_2 + \bar{\beta}_2^2 \theta_1 \gamma_1)}{(\gamma_1 \gamma_2 - xy)^2 (\theta_1 \theta_2 - xy)}, \tag{4.21}$$

$$\sigma_{22}^2 = \frac{\nu_2^2}{4} \frac{2\lambda_{2,\infty}(\bar{\beta}_2 \bar{\beta}_1^2 y^2 x + \bar{\beta}_2^3 \theta_1 \gamma_1^2) + \lambda_{1,\infty} \bar{\beta}_1 y (\bar{\beta}_1^2 y \theta_2 + \bar{\beta}_2^2 \gamma_1^2)}{(\gamma_1 \gamma_2 - xy)^2 (\theta_1 \theta_2 - xy)} \tag{4.22}$$

with $\gamma_i = \bar{\beta}_i + \alpha_m^i - \alpha_s^i$ and $\theta_i = \bar{\beta}_i - \alpha_m^i - \alpha_s^i \qquad i \in \{1, 2\}$.

Remark 4.10. *Note that when $x = y = 0$ then $\sigma_{12} = 0$ and $\sigma_{11}^2 = \frac{\nu_1^2}{2} \Lambda_1 \kappa_1^2$ as was proved in the preceding chapter.*

Remark 4.11. *From Proposition 4.1 we can retrieve the diffusive limit for the volatility by considering the limit $\lim_{\tau \to +\infty} C(\tau)$ in equation (4.7), it leads to the computation of $\lim_{\tau \to +\infty} \frac{1}{\tau} \mathrm{Cov}(\tau)$ with $\mathrm{Cov}(\tau)$ given by equation (3.15). Taking into account the form of $c_5(\tau)$ this limit is given by:*

$$\lim_{\tau \to +\infty} \frac{1}{\tau} \mathrm{Cov}(\tau) = \tilde{J}_1 + \tilde{J}_1^\top + \mathrm{diag}(\bar{\lambda}_\infty) \tag{4.23}$$

with $\tilde{J}_1 = -(\alpha - \beta)^{-1}(\bar{\Lambda}_\infty + \alpha \mathrm{diag}(\bar{\lambda}_\infty))$.

Similar remark applies to the covariance between the stocks given in Proposition 4.2, considering $\lim_{\tau \to +\infty} \mathrm{CovS}(\tau)$ leads to the same limit equation.

Remark 4.12. *Taking the limit with respect to τ in (4.10) and (4.11) gives $\lim_{\tau \to +\infty} \mathrm{L}_{1,2}(\tau, \delta) = \lim_{\tau \to +\infty} \mathrm{L}_{2,1}(\tau, \delta) = 0$. Hence, we cannot extract lead-lag relationship in the diffusive case or equivalently at low frequency (i.e., daily) a fact already underlined by Huth and Abergel (2012) and Bacry et al. (2013b).*

From equation (4.21) we conclude that if $x = y = 0$ then $\sigma_{12} = 0$ which is expected. Due to sign constraints on the parameters if $x > 0$ and/or $y > 0$ we also conclude that $\sigma_{12} > 0$ which is also consistent with the dynamic for the stocks. If in the matrix α of (4.1) the upper-right matrix is given by (4.3) and lower-left matrix is given by (4.4) then the volatilities σ_{11}^2 and σ_{22}^2 will still be given by (4.20) and (4.22) whilst the covariance σ_{12} will be minus the term (4.21). Hence, a negative correlation will be achieved and it is consistent with the dynamic in that case.

4.2 Empirical Analysis

4.2.1 Data description and estimation methodology

We rely on tick-by-tick data from TRTH (Thomson Reuters Tick History). We deal with futures on indices such as Dax and Eurostoxx (respectively noted FDX and STXE in the tables), as well as some other stocks: Renault, Peugeot, Société Générale and BNP Paribas (respectively RENA.PA, PEUP.PA, SOGN.PA and BNPP.PA). The data covers the period between 2010/01/01 to 2011/12/31. It consists of quote files recording quote changes (bid/ask prices and quantities) timestamped up to the millisecond. For every considered day, we took the front maturing future for the indices, i.e the future with nearest maturity, which is generally the most traded one. For every considered asset, we neglect the first and last 15 minutes in order to avoid the open and close auctions.

We estimate the model as described in (4.1) thanks to maximum likelihood method. For MLE maximisations, we rely on the Nelder Mead algorithm as implemented in the open source library NL-opt[1]. Notice that at every optimisation step, we check the stability condition of the priocess as specified in (4.5) and (4.6).

4.2.2 Estimation results

Results are gathered in Table 4.1. The table gives means, median and standard deviations of the parameters of the process, calibrated daily to the data.

We are naturally primarily interested in the coupling parameters x and y. We give them a closer look by plotting the time series of these pairs. In Figure 4.3, one can see that for the pair (SOGN.PA,BNPP.PA), x and y take heterogeneous values. There are some periods where the one clearly dominates the other, and some other periods where these values are not so far. The leader and lagger roles in this pair of large French banks seem to change their role from period to period.

As for the pair formed by the (STXE, BNPP.PA), different roles are well determined. The future on the index clearly leads the stock.

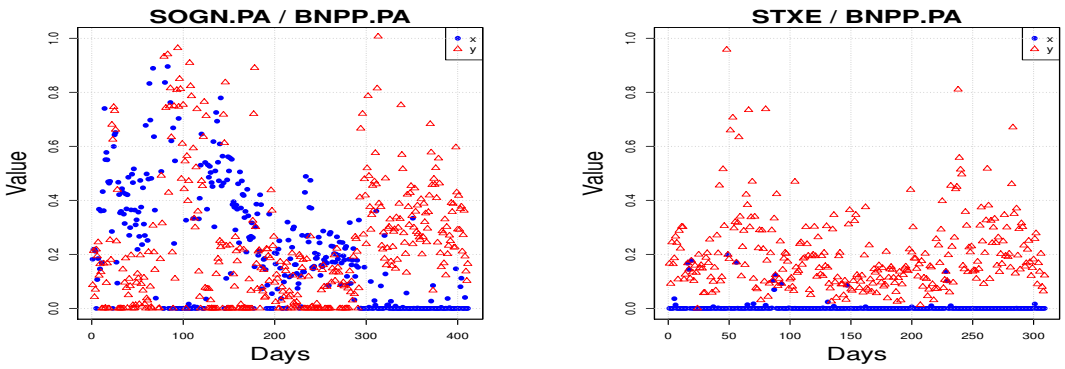

Figure 4.3: We plot time series of the daily calibrated parameters x (blue-bullet points) and y (red-triangle points) for the pairs (SOGN.PA, BNPP.PA) on the left and (STXE,BNPP.PA) on the right, illustrating the evolution of the respective lead lag relationships in time. Notice that the index systematically leads the stock as is clear on the right figure. The relationship between Société Générale and BNP Paribas is more mitigated, even that beginning April 2011, the two stocks began a bearish period, where Société Générale is clearly the leader.

[1] http://ab-initio.mit.edu/wiki/index.php/NLopt

Table 4.1: Calibration Results

Pair	Measure	$\lambda_{1,\infty}$	α_s^1	α_m^1	β^1	$\lambda_{2,\infty}$	α_s^2	α_m^2	β^2	x	y
STXE,SOGN.PA	Mean	0.0111	0.0049	0.0601	0.1215	0.0556	0.1797	0.0928	0.6028	0.0059	0.2488
	Median	0.0075	0.0011	0.0528	0.0899	0.0352	0.0329	0.0169	0.3465	0.0000	0.1891
	Std. dev.	0.0133	0.0085	0.0358	0.1109	0.0717	0.2601	0.1522	0.6379	0.0271	0.1920
STXE,BNPP.PA	Mean	0.0114	0.0039	0.0553	0.1028	0.0580	0.1041	0.0712	0.4260	0.0057	0.2156
	Median	0.0079	7e-04	0.0487	0.0772	0.0363	1e-04	0.0072	0.2100	0.0000	0.1650
	Std. dev.	0.0119	0.0074	0.0341	0.0943	0.0622	0.2245	0.1587	0.5926	0.0265	0.2038
STXE,RENA.PA	Mean	0.0081	0.0039	0.0636	0.0960	0.0324	0.1169	0.0833	0.4630	0.0018	0.2028
	Median	0.0055	7e-04	0.0572	0.0793	0.0249	0.0358	0.0208	0.2537	0.0000	0.1683
	Std. dev.	0.0097	0.0075	0.0343	0.0711	0.0313	0.1680	0.1306	0.5037	0.0150	0.1446
STXE,PEUP.PA	Mean	0.0051	0.0029	0.0615	0.0816	0.0130	0.0383	0.0535	0.2178	0.0058	0.1223
	Median	0.0035	9e-04	0.0559	0.0719	0.0110	0.0080	0.0380	0.1531	0.0000	0.1117
	Std. dev.	0.0058	0.0049	0.0292	0.0528	0.0103	0.0592	0.0619	0.1936	0.0221	0.0670
STXE,FDX	Mean	0.0086	0.0032	0.0604	0.0860	0.0510	2e-04	0.0066	0.0710	1e-04	0.0706
	Median	0.0072	0.0016	0.0517	0.0714	0.0496	0.0000	0.0000	0.0619	0.0000	0.0619
	Std. dev.	0.0082	0.0045	0.0373	0.0678	0.0294	8e-04	0.0131	0.0357	4e-04	0.0354
SOGN.PA,BNPP.PA	Mean	0.0263	0.0845	0.2012	0.5604	0.0382	0.1103	0.1907	0.6488	0.2271	0.2575
	Median	0.0184	0.0031	0.1515	0.3416	0.0283	0.0172	0.1440	0.4296	0.1610	0.1845
	Std. dev.	0.0276	0.1727	0.2213	0.6180	0.0343	0.2280	0.2302	0.7615	0.2844	0.2846
RENA.PA,PEUP.PA	Mean	0.0185	0.0830	0.0460	0.3353	0.0150	0.0734	0.1147	0.3992	0.2173	0.0638
	Median	0.0132	0.0014	1e-04	0.1643	0.0113	0.0511	0.0943	0.2649	0.1840	0.0378
	Std. dev.	0.0205	0.1348	0.0855	0.4060	0.0185	0.0855	0.0823	0.4514	0.1547	0.0847

Note. Mean, median and standard deviation of calibration results for the different pairs. In order to read the table, and in accordance with our previous notations, notice that for every pair (A,B), the value of x carries the influence of B on A, and the value of y carries the influence of A on B.

In Figure 4.4, one can see that for the pair of stocks (RENA.PA, PEUP.PA), things are not as for the pair of banks. In fact, x clearly dominates y, and then Peugeot seems to influence Renault more than the inverse. This can be explained by two facts: first, Peugeot stock is cheaper than Renault (average price in the considered period: 24 Eur for Peugeot versus 35 Eur for Renault). Therefore, bets on the automotive industry will be taken preferably in Peugeot rather than Renault, as Peugeot will ensure a better leverage. The french state being one of the principal stock holders in Renault tends also to attract automotive industry speculators to Peugeot in the first place.

And finally, the Eurostoxx, the index of Euro zone stocks, naturally leads the DAX German index.

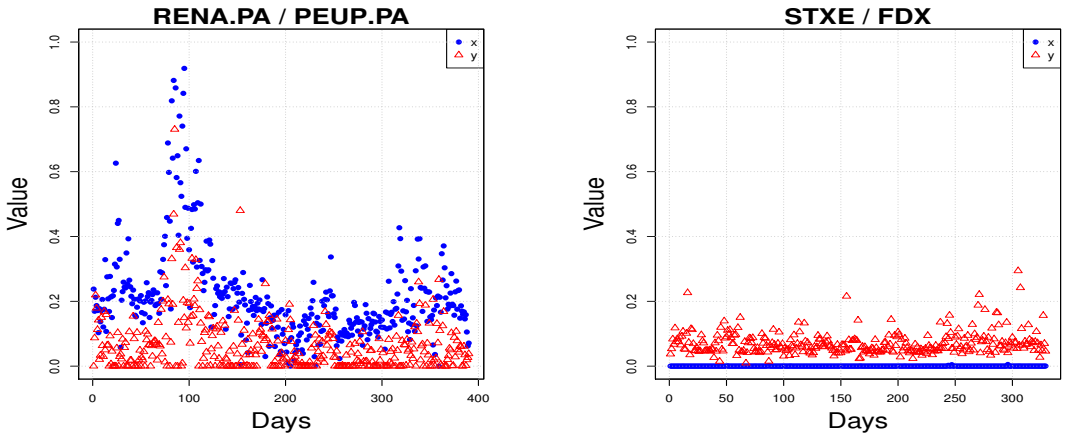

Figure 4.4: We plot time series of the daily calibrated parameters x (blue-bullet points) and y (red-triangle points) for the pairs (RENA.PA, PEUP.PA.PA) on the left and (STXE,FDX) on the right, illustrating the evolution of the respective lead lag relationships in time. The European index systematically leads the German index, while Peugeot systematically leads Renault, even if the feedback effect is here more pronounced than in the case of the indices.

These lead-lag relationships can also be expressed in terms of time. Indeed, we have seen in Figure 4.2 that the correlation induced by a Hawkes framework, when x and y are different, reaches its maximum value when returns are calculated during two (overlapping) periods with a certain lag. This lag can be used to characterise the lead relationship. Indeed, for a pair (A,B), if $x > y$, and then B influences A, the maximum correlation is achieved with a positive lag, and vice versa. Once our models are calibrated, we can calculate this optimal lag daily, which gives the Table 4.2.

Table 4.2: Lead Times

Pair	Mean	Median	Std. dev.
STXE,SOGN.PA	-2.39	-2	3.06
STXE,BNPP.PA	-3.38	-3	3.53
STXE,RENA.PA	-3.04	-3	2.90
STXE,PEUP.PA	-3.89	-3	4.05
STXE,FDX	-9.56	-9	4.55
SOGN.PA,BNPP.PA	0.14	0	2.12
RENA.PA,PEUP.PA	4.23	3	4.66

Note. Mean, median and standard deviation of lead times. A negative value $\boldsymbol{t}$ indicates that maximum correlation between the components of the pair is attained if we measure the increments of the second pair component $\boldsymbol{t}$ seconds after those of the first componnet.

4.3 Conclusion

In this chapter we develop a multi-asset model using Hawkes process in the spirit of Bacry et al. (2013a). Within that framework we compute various statistical properties such as the signature plot, the covariance and the lagged covariance between the stocks. This enable us to precisely analyze the lead-lag relationship between the stocks, this is an important quantity that was, and still is, well studied within other frameworks (e.g. de Jong and Nijman (1997)). We perform some Taylor expansions for these quantities in order to gain a better understanding of the impact of the parameters on these key quantities. We also compute the diffusive limit associated with the model thereby connecting the parameters driving the high-frequency dynamic with the low frequency (i.e., daily) evolution of the stocks. We find that it is not possible to capture this lead-lag relationship at low frequency, a result that confirms previous works (e.g., Huth and Abergel (2012)).

We estimate the model using a two-year sample of high-frequency data for two index futures and four major stocks quoted on the Eurex market. We find that the Eurostoxx leads all the stocks which is a somewhat expected result as it is the largest European index. For the same reasons we find that the Eurostoxx also leads the DAX. For the two stocks in the banking sector when averaged over all the sample none of them clearly leads the other. However, when the lead-lag relationship is analysed at a daily frequency we find some periods during which one stock clearly leads the other. For the two stocks of the automotive industry one of the stock systematically leads the other and it is consistent with the shareholder structure of these companies (one of them is partially state owned). Overall, the results are consistent with financial intuition.

4.4 Appendix

Proof. of Proposition 4.4
To perform the expansion we need to differentiate the function 3.15.

Let us define $\alpha_x = \frac{d\alpha}{dx}$, $\alpha_y = \frac{d\alpha}{dy}$ $\alpha^0 = \alpha_{x=0,y=0}$. Starting from $I = \alpha^{-1}\alpha$ and taking into account the fact that β does not depend on x and y we deduce:

$$(\alpha^{-1})_x = -\alpha^{-1}\alpha_x\alpha^{-1}$$
$$((\alpha-\beta)^{-1})_x = -(\alpha-\beta)^{-1}\alpha_x(\alpha-\beta)^{-1}$$

(these equations will be evaluated for $\alpha = \alpha^0$) and similarly:

$$\begin{aligned}(\alpha^{-2})_x &= -\alpha^{-2}(\alpha_x\alpha + \alpha\alpha_x)(\alpha^2)^{-1}\\ &= -\alpha^{-2}\alpha_x\alpha^{-1} - \alpha^{-1}\alpha_x\alpha^{-2}.\end{aligned}$$

For example, from the previous equation we deduce that:

$$((\alpha-\beta)^{-2})_x = -(\alpha-\beta)^{-2}\alpha_x(\alpha-\beta)^{-1} - (\alpha-\beta)^{-1}\alpha_x(\alpha-\beta)^{-2}.$$

From the equation $\bar\lambda_\infty = -(\alpha-\beta)^{-1}\beta\lambda_\infty$ we define:

$$\bar\lambda^0_\infty = -(\alpha^0-\beta)^{-1}\beta\lambda_\infty$$
$$\bar\lambda^0_{x,\infty} = \left(\frac{d\bar\lambda_\infty}{dx}\right)_{x=0} = \left((\alpha^0-\beta)^{-1}\alpha_x(\alpha^0-\beta)^{-1}\right)\beta\lambda_\infty$$

with similar equation for $\bar\lambda^0_{y,\infty}$. We expand $\bar\Lambda_\infty$ solution of:

$$(\alpha-\beta)\bar\Lambda_\infty + \bar\Lambda_\infty(\alpha-\beta)^\top + \alpha\mathrm{diag}(\bar\lambda_\infty)\alpha^\top = 0$$

in the form $\bar\Lambda_\infty = \bar\Lambda^0_\infty + x\bar\Lambda^0_{x,\infty} + y\bar\Lambda^0_{y,\infty}$ with:

$$(\alpha^0-\beta)\bar\Lambda^0_\infty + \bar\Lambda^0_\infty(\alpha^0-\beta)^\top + \alpha^0\mathrm{diag}(\bar\lambda^0_\infty)(\alpha^0)^\top = 0$$

and

$$\begin{aligned}(\alpha^0-\beta)\bar\Lambda^0_{x,\infty} + \bar\Lambda^0_{x,\infty}(\alpha^0-\beta)^\top = &-\alpha_x\bar\Lambda^0_\infty - \bar\Lambda^0_\infty{\alpha_x}^\top - \alpha_x\mathrm{diag}(\bar\lambda^0_\infty)(\alpha^0)^\top\\ &- \alpha^0\mathrm{diag}(\bar\lambda^0_{x,\infty})(\alpha^0)^\top - \alpha^0\mathrm{diag}(\bar\lambda^0_\infty)\alpha_x^\top\end{aligned}$$

with a similar equation for $\bar\Lambda^0_{y,\infty}$.

We need to differentiate $\mathrm{Cov}(\tau)$ given by (3.15) which leads to differentiate $c_5(\tau)$ (3.25). We denote $c_5^0(\tau)$ the function evaluated for $x = y = 0$ and the derivative is just:

$$\begin{aligned}\partial_x c_5(\tau)_{|_{x=y=0}} &= \partial_x c_5^0(\tau)\\ &= (\alpha^0-\beta)^{-1}\alpha_x(\alpha^0-\beta)^{-1}\tau + \tau(\alpha^0-\beta)^{-2}\alpha_x e^{(\alpha^0-\beta)\tau}\\ &- \left((\alpha^0-\beta)^{-2}\alpha_x(\alpha^0-\beta)^{-1} + (\alpha^0-\beta)^{-1}\alpha_x(\alpha^0-\beta)^{-2}\right)(e^{(\alpha^0-\beta)\tau} - I).\end{aligned}$$

Lastly, the expansion of J_1 is:

$$\begin{aligned}J_1 &= J_1^0 + xJ^0_{1,x},\\ J_1^0 &= c_5^0(\tau)(\bar\Lambda^0_\infty + \alpha\mathrm{diag}(\bar\lambda^0_\infty)),\\ J^0_{1,x} &= \partial_x c_5^0(\tau)(\bar\Lambda^0_\infty + \alpha^0\mathrm{diag}(\bar\lambda^0_\infty)) + c_5^0(\tau)(\bar\Lambda_{x,\infty} + \alpha_x\mathrm{diag}(\bar\lambda^0_\infty) + \alpha^0\mathrm{diag}(\bar\lambda^0_{x,\infty})).\end{aligned}$$

Therefore, we have the expansion for $\mathrm{Cov}(\tau)$ given by:

$$\begin{aligned}\mathrm{Cov}(\tau) = &J_1^0 + (J_1^0)^\top + \tau\mathrm{diag}(\bar\lambda^0_\infty)\\ &+ x\left(J^0_{1,x} + (J^0_{1,x})^\top + \tau\mathrm{diag}(\bar\lambda^0_{x,\infty})\right) + y\left(J^0_{1,y} + (J^0_{1,y})^\top + \tau\mathrm{diag}(\bar\lambda^0_{y,\infty})\right)\end{aligned}$$

from which we deduce the expansion of the signature plot. ■

General Conclusion and Outlooks

In this thesis, we have been interested in some price formation and market microstructure mechanisms. We have dealt with these issues mainly by relying on Hawkes processes which proved to be a powerful and versatile tool. A lot remains to be done and the Hawkes process framework constitutes certainly the build block to many other applications beyond the ones that we develop in this thesis.

One first direction of research is to relate the first and last chapters of the thesis. In fact, the joint dynamics of an option and its underlying investigated in the first chapter can be modeled in the 4-dimensional Hawkes framework developed in the last chapter.

$$S_t = S_0 + \left(N_t^{1,u} - N_t^{1,d}\right)\frac{\nu_1}{2},$$
$$C_t = C_0 + \left(N_t^{2,u} - N_t^{2,d}\right)\frac{\nu_2}{2}$$

where ν_1 and ν_2 are the tick values for the first stock (S_t) and the option (C_t) respectively. If we deal with a Call option, we take the usual parametrisation of α and β matrices driving the 4-dimensional Hawkes process as following:

$$\alpha = \begin{pmatrix} \alpha_s^1 & \alpha_m^1 & x & 0 \\ \alpha_m^1 & \alpha_s^1 & 0 & x \\ y & 0 & \alpha_s^2 & \alpha_m^2 \\ 0 & y & \alpha_m^2 & \alpha_s^2 \end{pmatrix}; \beta = \begin{pmatrix} \bar{\beta}_1 & 0 & 0 & 0 \\ 0 & \bar{\beta}_1 & 0 & 0 \\ 0 & 0 & \bar{\beta}_2 & 0 \\ 0 & 0 & 0 & \bar{\beta}_2 \end{pmatrix}$$

and $\lambda_\infty = (\lambda_{1,\infty}, \lambda_{1,\infty}, \lambda_{2,\infty}, \lambda_{2,\infty})^\top \in \mathbb{R}_+^4$.

The coupling parameters x and y of the Hawkes dynamics will then carry the delta effect of the underlying on the option, and a feed back effect of the option dynamics on the underlying.

Notice that without any further developments, and thanks to the computations done in the last chapter, we can calculate the *effective delta* defined in the first chapter as the slope of the regression of option price increments on those of the underlying. In this model, asymptotic values of the stock variance and the stock/option correlation write:

$$\sigma_S^2 = \frac{\nu_1^2}{4}\frac{2\lambda_{1,\infty}(\bar{\beta}_1\bar{\beta}_2^2x^2y + \bar{\beta}_1^3\theta_2\gamma_2^2) + \lambda_{2,\infty}\bar{\beta}_2x(\bar{\beta}_2^2x\theta_1 + \bar{\beta}_1^2\gamma_2^2)}{(\gamma_1\gamma_2 - xy)^2(\theta_1\theta_2 - xy)}$$
$$Cov(S,C) = \frac{\nu_1\nu_2}{4}\frac{2\bar{\beta}_1y\lambda_{1,\infty}(\bar{\beta}_2^2x\gamma_1 + \bar{\beta}_1^2\theta_2\gamma_2) + 2\bar{\beta}_2x\lambda_{2,\infty}(\bar{\beta}_1^2y\gamma_2 + \bar{\beta}_2^2\theta_1\gamma_1)}{(\gamma_1\gamma_2 - xy)^2(\theta_1\theta_2 - xy)}$$

with $\gamma_i = \bar{\beta}_i + \alpha_m^i - \alpha_s^i$ and $\theta_i = \bar{\beta}_i - \alpha_m^i - \alpha_s^i \qquad i \in \{1,2\}$.

Then the effective delta is:

$$\begin{aligned}\Delta_{effective} &= \frac{Cov(S,C)}{\sigma_S^2} \\ &= \frac{\nu_2}{\nu_1}\frac{2\bar{\beta}_1 y\lambda_{1,\infty}(\bar{\beta}_2^2 x\gamma_1 + \bar{\beta}_1^2\theta_2\gamma_2) + 2\bar{\beta}_2 x\lambda_{2,\infty}(\bar{\beta}_1^2 y\gamma_2 + \bar{\beta}_2^2\theta_1\gamma_1)}{2\lambda_{1,\infty}(\bar{\beta}_1\bar{\beta}_2^2 x^2 y + \bar{\beta}_1^3\theta_2\gamma_2^2) + \lambda_{2,\infty}\bar{\beta}_2 x(\bar{\beta}_2^2 x\theta_1 + \bar{\beta}_1^2\gamma_2^2)}\end{aligned}$$

When dealing with options, one wants to take the option strike into account, which complicates the problem by increasing its dimension. One then wants ideally to come with a sufficiently rich model taking into account this strike dimension, and remaining analytically tractable.

Another research direction is to take into account the volume aspect of the trading process. In fact, the Hawkes model used in this thesis deals only with arrival time of orders, and not with their volumes. This is clearly a restrictive assumption as volumes are heterogeneous in real order books. A solution would be to rely on marked Hawkes processes, where the marks would account for the order quantities. Moreover, it seems reasonable to give more weight to trades with larger volumes and vice versa. An effort in this direction was made by (Shek, 2010). A lot remains to be done on the analytical side as well as asymptotic results.

To end with, notice that computations made in this thesis relyed heavily on the Markov property of the Hawkes process, which results from the exponential form of the kernel. In (Bacry et al., 2012), the authors estimate a non-parametric kernel and find that it is slowly decaying, suggesting a stronger memory effect than that induced by an exponential kernel. A trade off between these two approaches would be to parametrise the kernel with a power law decaying function instead of the rapidly decaying exponential function, in the hope of being closer to reality, while remaining analytically tractable.

Bibliography

Frederic Abergel and Aymen Jedidi. A Mathematical Approach to Order Book Modeling. *International Journal of Theoretical and Applied Finance*, 16(5), 2013a.

Frederic Abergel and Aymen Jedidi. On the Stability and Price Scaling Limit of a Hawkes Process-Based Order Book Model. *working paper*, 2013b.

Yacine Aït-Sahalia, Julio Cacho-Diaz, and Roger J.A. Laeven. Modeling financial contagion using mutually exciting jump processes. Working paper, National Bureau of Economic Research, 2010.

Torben G. Andersen and Luca Benzoni. *Handbook of Financial Time Series*, chapter Realized Volatility. New York, NY: Springer Verlag, 2009.

Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. (understanding, optimizing, using and forecasting) realized volatility and correlation. New York University, Leonard N. Stern School Finance Department Working Paper Seires 99-061, New York University, Leonard N. Stern School of Business-, Oct 1999.

Emmanuel Bacry and Jean-François Muzy. Hawkes model for price and trades high-frequency dynamics. *Working Paper*, 2013.

Emmanuel Bacry, Khalil Dayri, and Jean-François Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B - Condensed Matter and Complex Systems*, 85(5):1–12, May 2012.

Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13:65–77, January 2013a.

Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. Scaling limits for hawkes processes and application to financial statistics. *Stochastic Processes and Applications*, 123:2475–2499, 2013b.

Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. Wiley, New York, 1999.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, May-June 1973.

T. Bollerslev and H. Zhou. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109:33–65, 2002.

T. Bollerslev and H. Zhou. Corrigendum to "Estimating stochastic volatility diffusion using conditional moments of integrated volatility". *Journal of Econometrics*, 119:221–222, 2004.

Jean-Philippe Bouchaud, Yuval Gefen, Marc Potters, and Matthieu Wyart. Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quantitative Finance*, 10:176–190, April 2004.

Clive G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141:876–912, 2007.

Pierre Brémaud. *Point Processes and Queues, Martingale Dynamics*. Springer, Berlin Heidelberg New York, 1981.

Pierre Brémaud and Laurent Massoulié. Imbedded construction of stationary sequences and point processes with a random memory. *Queueing Systems*, 17:213–234, 1994.

Charles Quanwei Cao, Gurdip S. Bakshi, and Zhiwu Chen. Empirical performance of alternative option pricing models. *The Journal of Finance*, 52(5), December 1997.

Peter Carr and Dilip B. Madan. Option valuation using fast fourier transform. *The Journal of Computational Finance*, 2(4):61–73, 1999.

Rama Cont and José da Fonseca. Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60, 2002.

Rama Cont and Adrien De Larrard. Price Dynamics in a Markovian Limit Order Market. *Social Science Research Network Working Paper Series*, January 2011.

Rama Cont and Adrien De Larrard. Order Book Dynamics in Liquid Markets: Limit Theorems and Diffusion Approximations. *Social Science Research Network Working Paper Series*, February 2012.

Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563, 2010.

C. Cuchiero, M. Keller-Ressel, and J. Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16:711–740, 2012.

José Da Fonseca and Riadh Zaatour. Hawkes process: Fast calibration, application to trade clustering and diffusive limit. *Working Paper SSRN-2294112, forthcoming Journal of Futures Markets*, 2013a.

José Da Fonseca and Riadh Zaatour. Clustering and mean reversion in a Hawkes microstructure model. *Working Paper*, 2013b.

Toby Daglish, John Hull, and Wulin Suo. Volatility surfaces: theory, rules of thumb, and empirical evidence. *Quantitative Finance*, 7(5):507–524, 2007.

Daryl J. Daley and David Vere Jones. *An introduction to the theory of point processes, Vol 1, Elementary theory and methods*. Springer, Berlin Heidelberg New York, 2. ed. edition, 2002.

Daryl J. Daley and David Vere Jones. *An introduction to the theory of point processes, Vol 2, General theory and structure*. Springer, Berlin Heidelberg New York, 2. ed. edition, 2008.

F. de Jong and T. Nijman. High frequency analysis of Lead-Lag relationships between financial markets. *Journal of Empirical Finance*, 4:259–277, 1997.

Emmanuel Derman. Regimes of volatility. 1999. Quantitative Strategies Research Notes, Goldman Sachs,New York, NY. Also in Risk Magazine, April, 1999.

D. Duffie, D. Filipović, and W. Schachermayer. Affine processes and applications in finance. *The Annals of Applied Probability*, 13(3):984–1053, 2003.

Darrel Duffie and Rui Kan. A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6 (4):379–406, 1996.

Darrell Duffie. Financial modeling with affine processes. A draft course available on the internet, 2007.

Alfonse Dufour and Robert F. Engle. Time and the Price Impact of a Trade. *The Journal of Finance*, 55(6), 2000.

Bruno Dupire. Pricing with a smile. *Risk*, (7):18–20, 1994.

Valdo Durrleman. Convergence of at-the-money implied volatilities to the spot volatility. *Journal Of Applied Probability*, 45(2):542–550, 2008.

Robert F. Engle and Jeffrey R. Russell. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, 66(5):1127–1162, September 1998.

Thomas W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366):291–298, Jun 1979.

Eymen Errais, Kay Giesecke, and Lisa R. Goldberg. Affine point processes and portfolio credit risk. *SIAM J. Financial Math.*, 1(1):642–665, 2010.

Damir Filipović, Eberhard Mayerhofer, and Paul Schneider. Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics*, 176(2):93–111, 2013.

Mark B. Garman and Michael J. Klass. On the estimation of security price volatilities from historical data. *The Journal of Business*, 53(1):67–78, January 1980.

Jim Gatheral. *The Volatility Surface : A Practitioner's Guide*. Wiley Finance, 2006.

Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in Mathematical Sciences. The Johns Hopkins University Press, Baltimore, 1996.

Alastair R. Hall. *Generalized Method of Moments*. Advanced Texts in Econometrics. Oxford University Oxford, Oxford, 2004. ISBN 9780198775201.

Anthony D. Hall and Nikolaus Hautsch. Order aggressiveness and order book dynamics. *Empirical Economics*, 30(4):973–1005, 2006.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–54, July 1982.

Nikolaus Hautsch. *Econometrics of Financial High-frequency Data*. Springer, Berlin Heidelberg New York, 2012.

Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

Takaki Hayashi and Nakahiro Yoshida. On covariance estimation of non-synchronously observed diffusion processe. *Bernoulli*, 11(2):359–379, 2005.

Anthony F. Herbst, Joseph P. McCormack, and Elizabeth N. West. Investigation of a lead-lag relationship between spot stock indices and their futures contracts. *Journal of Futures Markets*, 7(4):373–381, 1987.

Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–43, 1993.

Patrick Hewlett. Clustering of order arrivals, price impact and trade path optimisation. 2006.

M. Hoffmann, M. Rosenbaum, and N. Yoshida. Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli*, 19(2):426–461, 2013.

Nicolas Huth and Fréd'eric Abergel. High Frequency Lead/lag Relationships Empirical facts. *working paper*, 2012.

Andrei Kirilenko, Richard B. Sowers, and Xiangqian Meng. A multiscale model of high-frequency trading. *Algorithmic Finance*, 2(1):59–98, February 2013.

Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

Jeremy Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, February 2007.

Charles M. C. Lee and Mark J. Ready. Inferring Trade Direction from Intraday Data. *The Journal of Finance*, 46(2), June 1991.

Roger W. Lee. The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, 14(3):469–480, 2004.

Roger W. Lee. Implied volatility: Statics, dynamics, and probabilistic interpretation. *Recent Advances in Applied Probability*, 2005.

Alan L. Lewis. *Option Valuation under Stochastic Volatility.* Finance Press, 2000.

E. Lewis and G. Mohler. A nonparametric em algorithm for multiscale hawkes processes. *Journal of Nonparametric Statistics*, 1, 2011.

M.I.A. Lourakis. levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++. [web page] `http://www.ics.forth.gr/~lourakis/levmar/`, Jul. 2004. [Accessed on 31 Jan. 2005.].

Maria Elvira Mancino and Paul Malliavin. Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics*, 6(1):49–61, 2002.

Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability.* Cambridge University Press, Cambridge, 2nd edition, 2009.

Sean P. Meyn and R. L. Tweedie. Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes. *Advances in Applied Probability*, 25(3):518–548, 1993.

G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493): 100–108, 2011.

Esteban Moro, Javier Vicente, Luis G. Moyano, Austin Gerig, J. Doyne Farmer, Gabriella Vaglica, Fabrizio Lillo, and Rosario N. Mantegna. Market impact and trading profile of hidden orders in stock markets. *Physical Revue E*, 80(6), December 2009.

Ioane Muni Toke and Fabrizio Pomponio. Modelling trades-through in a limit order book using hawkes processes. *Economics: The Open-Access, Open-Assessment E-Journal*, 6(2012-22), 2012. doi: 10.5018/economics-ejournal.ja.2012-22. URL http://dx.doi.org/10.5018/economics-ejournal.ja.2012-22.

Michael C. Münnix, Rudi Schäfer, and Thomas Guhr. Impact of the tick-size on financial returns and correlations. *Physica A*, 389(21):4828–4843, 2010.

Per Aslak Mykland and Lan Zhang. Inference for volatility-type objects and implications for hedging. *Statistics and Its Interface*, 1:255–278, 2008.

Yosihiko Ogata. On Lewis simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

T Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.

Fabrizio Pomponio and Frederic Abergel. Multiple-limit trades: empirical facts and application to lead-lag measures. *Quantitative Finance*, 13(5):783–793, 2013.

Riccardo Rebonato. *Volatility and Correlation, The Perfect Hedger and The Fox*, chapter Stochastic-Volatility Processes. John Wiley & Sons Ltd, 2. ed. edition, 2004.

D. Revuz and M. Yor. *Continuous martingales and Brownian motion*. 3rd ed, Springer, Berlin Heidelberg New York, 1999.

Asani Sarkar and Robert A. Schwartz. Two-sided markets and intertemporal trade clustering: insights into trading motives. Staff Reports 246, Federal Reserve Bank of New York, 2006.

Howard Shek. Modeling High Frequency Market Order Dynamics Using Self-Excited Point Process. *SSRN eLibrary*, 2010.

Alejandro Veen and Frederic P. Schoenberg. Estimation of spacetime branching process models in seismology using an emtype algorithm. *Journal of the American Statistical Association*, 103:614–624, June 2008.

Daley Vere-Jones. Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society Series B*, 32(1):1–62, 1970.

Paul Wilmott. *Paul Wilmott On Quantitative Finance*, chapter Stochastic Volatility. John Wiley & Sons Ltd, 2. ed. edition, 2006.