

MS&E 444 Investment Practice

Exciting Times for Trade Arrivals

Vassil Chatalbashev, Yifan Liang,
Ari Officer and Nikolaos Trichakis

Instructor: Professor Kay Giesecke

Teaching Assistant: Benjamin Armbruster



Department of Management Science and Engineering
Stanford University

June 15, 2007

Table of Contents

Abstract	2
1 Hawkes Process: Background Information	4
1.1 Introduction	4
1.2 Exponential Hawkes: Continuous Time	5
1.3 Exponential Hawkes: Discrete Time	5
1.4 Model Fitting: Exponential Hawkes	6
1.5 Generalized Hawkes Model	7
2 Application of Hawkes Models	8
2.1 Trade Data Collection and Processing	8
2.2 Model Fitting	8
2.3 Trading Strategy and Performance	11

Abstract

It has been empirically observed that buy and sell orders in many financial markets tend to cluster in time. We develop a formal model to account for such a clustering effect. We consider an exponential decay Hawkes model, and a more generalized linear model, assessing the goodness of fit and parameter estimation efficiency of each. Our results indicate that our generalized linear Hawkes model is better suited to modeling high-frequency financial time-series, in which data is abundant, and fast parameter estimation is desired. We also use our model in simple trading strategy, which gives promising results.

Acknowledgments

The authors are grateful to Professor Kay Giesecke, Benjamin Armbruster, Lisa Borland and Jeremy Evtine for stimulating discussions during the project. The authors also wish to thank EvA Funds for sponsoring the access to Trade and Quote (TAQ) database.

Chapter 1: Hawkes Process: Background Information

1.1 Introduction

Empirically, it has been observed that in many financial markets trading activity tends to cluster in time. This is also true when considering “signed” trading activity, i.e. *buy* and *sell* orders. Liquidity providers (“market makers”) in the foreign exchange market are well aware of this clustering, and anecdotal evidence suggests that they pay close attention to the pattern of arrivals of buy and sell orders when setting prices.

Such clustering can be modeled with a multivariate point process [1]. Methods of data analysis for point processes have received much attention [2], [3]. In this work, we focus on a model in which order arrivals are governed by a special class of point process, the Hawkes process [4]. We first consider the case where the buy orders and the sell orders are independent of each other, i.e. there is only a self-exciting effect among orders of the same type. A suitable model for this setting is a univariate Hawkes process. We then consider the cross-exciting effect among orders of different types and the corresponding bi-variate Hawkes process model.

The *propagator function* is introduced in the Hawkes model to characterize the effect of past order arrivals on the future arrival intensities. Exponential functions of a sum of exponential functions are well accepted forms of propagator functions. The advantage of the exponential specification is that when fitting a Hawkes model to empirical market data, the likelihood function can be computed in $O(N)$ steps, whereas for more general propagator functions, $O(N^2)$ steps will be required. However, the model fitting procedure is still computationally intensive and there is no single global optimum due to its non-convexity, even with the exponential functions. In view of this, we introduce a generalized Hawkes process so the model fitting can be formulated as a convex optimization problem, to which many efficient optimization algorithms can be applied.

The Wharton Research Data Services (WRDS) provides access to the NYSE TAQ database where the size and frequency of orders can be extracted. We fit various Hawkes models to these market data and verify the model fitness through statistical measures such as the *QQ-plot*. The fitted models are then used in the next section when designing new trading strategies.

1.2 Exponential Hawkes: Continuous Time

An I -variate Hawkes process focuses on arrival intensities for the counting processes $N_t^{(i)}$, $1 \leq i \leq I$. Arrival intensities $\lambda_t^{(i)}$ conditioned on a filtration \mathcal{F}_t is defined by

$$\lambda_t^{(i)} \Big| \mathcal{F}_t = \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \mathbb{E} \left\{ N_{t+\delta t}^{(i)} - N_t^{(i)} \Big| \mathcal{F}_t \right\}.$$

In the case of purely self-exciting processes, the intensity is a functional of past arrivals. For a linear self-exciting process, we have

$$\lambda_t^{(i)} = \mu^{(i)} + \sum_{j=1}^I \int_{u < t} h_{ij}(t-u) dN_u^{(j)}.$$

Here $\mu^{(i)}$ can be understood as the base intensity of arrivals of type i , i.e. the intensity if there have been no past arrivals of any type, and h_{ij} the propagator of an arrival of type j onto the intensity of arrivals of type i in the future. We first consider parameterized forms for h . In particular we consider the case where h is a sum of exponentials:

$$h_{ij}(s) = \sum_{k=1}^K \alpha_{ijk} e^{-\beta_{ijk}s},$$

so that

$$\lambda_t^{(i)} = \mu^{(i)} + \sum_{k=1}^K \sum_{j=1}^I \int_{u < t} \alpha_{ijk} e^{-\beta_{ijk}(t-u)} dN_u^{(j)}.$$

This specification is labeled a *Hawkes-E(K)* process in [5]. In this work we consider both univariate Hawkes process ($I = 1$), which models the case where orders are of the same type, and the bivariate Hawkes process ($I = 2$), which models the cross-exciting effects among buy and sell orders. For simplicity we assume there is only exponential component in the propagator function h , i.e. $K = 1$. Therefore the univariate case has the form

$$\lambda_t = \mu + \int_{u < t} \alpha e^{-\beta(t-u)} dN_u$$

and the bivariate case has the form

$$\begin{aligned} \lambda_t^{(1)} &= \mu^{(1)} + \int_{u < t} \alpha_{11} e^{-\beta_{11}(t-u)} dN_u^{(1)} + \int_{u < t} \alpha_{12} e^{-\beta_{12}(t-u)} dN_u^{(2)}, \\ \lambda_t^{(2)} &= \mu^{(2)} + \int_{u < t} \alpha_{21} e^{-\beta_{21}(t-u)} dN_u^{(1)} + \int_{u < t} \alpha_{22} e^{-\beta_{22}(t-u)} dN_u^{(2)}. \end{aligned}$$

1.3 Exponential Hawkes: Discrete Time

The classical continuous model implicitly assumes that the probability that more than one event occurs at exactly the same time approaches 0. In the market data we obtained from TAQ data base, this is usually not the case. Since trade data are recorded with time increment of 1 second, we do observe many trades occur at the same time although in reality they may not. Furthermore, even if our data did have a much higher time resolution, due to market microstructure artifacts, it is debatable whether it would be useful to model excitation effects on scales much lower than a second. In view of this, we introduce a Hawkes process model using discrete time.

We denote by $\lambda(i)$ the conditional expected intensity and by $N(i)$ the number of trades during the i th time step. Note that in the continuous model N_t is the counting process up to time t , so we have the approximation

$$N(i) = N_i - N_{i-1}.$$

Note that the continuous process uses subscripts for time indices while the discrete process puts time indices into parenthesis. They are not to be confused. In the bivariate Hawkes model, we use the superscript $(\cdot)^{(i)}$ to distinguish between the buy and sell processes for continuous time. In discrete time we will use the subscript $(\cdot)_i$ for the same purpose. Namely, the notation $\lambda_t^{(i)}$ changes to $\lambda_i(t)$ and similarly $N_t^{(i)}$ changes to $N_i(t)$. With these notations we have for the univariate model

$$\lambda(t) = \mu + \sum_{i < t} \alpha e^{-\beta(t-i)} N(i), \quad (1.1)$$

and for the bivariate model

$$\lambda_1(t) = \mu_1 + \sum_{i < t} \alpha_{11} e^{-\beta_{11}(t-i)} N_1(i) + \sum_{i < t} \alpha_{12} e^{-\beta_{12}(t-i)} N_2(i), \quad (1.2)$$

$$\lambda_2(t) = \mu_2 + \sum_{i < t} \alpha_{21} e^{-\beta_{21}(t-i)} N_1(i) + \sum_{i < t} \alpha_{22} e^{-\beta_{22}(t-i)} N_2(i). \quad (1.3)$$

The time steps are $t = 1, 2, \dots, T$ and $N(t)$ has a Poisson distribution with parameter $\lambda(t)$.

1.4 Model Fitting: Exponential Hawkes

For the univariate case, the log likelihood function of a discrete Hawkes Process is shown to be [6]

$$L = - \int_0^T \lambda_t dt + \int_0^T \log(\lambda_t) dN_t \quad (1.4)$$

$$\approx - \sum_{i=1}^T \lambda(i) + \sum_{i=1}^T N(i) \log(\lambda(i)) \quad (1.5)$$

where $\lambda(i)$ is given by (1.1). When we fit such a model to market data we actually solve the following optimization problem

$$\max \quad L \quad (1.6)$$

$$\text{subject to} \quad \mu > 0, \alpha > 0 \quad (1.7)$$

$$\beta > \alpha \quad (1.8)$$

where L is given by (1.5), the constraint (1.7) is for positivity and the constraint (1.8) is for stability.

For the bivariate case we also observe that

$$L \approx \sum_{i=1}^T \left\{ N_1(i) \log(\lambda_1(i)) + N_2(i) \log(\lambda_2(i)) - \lambda_1(i) - \lambda_2(i) \right\}. \quad (1.9)$$

The corresponding optimization problem has the following form

$$\max \quad L \quad (1.10)$$

$$\text{subject to} \quad \mu_i > 0, \alpha_{ij} > 0 \quad (1.11)$$

$$\beta_{ii} > \alpha_{ii} \quad (1.12)$$

$$\det(I_2 - \Psi) > 0 \quad (1.13)$$

where I_2 is the 2×2 identity matrix, $\Psi = (\psi_{ij}) = (\alpha_{ij}/\beta_{ij})$, L is given by (1.9), the constraint (1.11) is for positivity and the constraints (1.12) and (1.13) are for stability.

1.5 Generalized Hawkes Model

Consider the optimization problem (1.6). This is a simple model with just three variables μ , α and β . Yet the objective function is non-concave in the variables, so solutions are only local maxima and there is no guarantee of global optimality. The model also does not scale complexity of the optimization algorithm grows rapidly with the size of the problem. Furthermore, the optimization problem (1.10) generally involves 10 variables, so it suffers from the same problems to an even higher degree.

The assumption we have made so far is an exponential decay function. Instead we can consider a time-limited, piecewise linear form of $h_{ij}(t)$. If the duration of each constant piece (δt) is less than the time resolution of the data, no information is lost. Moreover, exponentials are themselves “time-limited”, since they die off after 20-30 seconds. Along these lines we develop a generalized Hawkes model where the predicted intensities can be written as

$$\lambda_i(t) = \mu_i + \sum_{k=t-n}^{t-1} w_{i1}(t-k)N_1(k\delta t) + \sum_{k=t-n}^{t-1} w_{i2}(t-k)N_2(k\delta t), \quad (1.14)$$

where $h_{ij}(t)$ consists of n pieces with the t -th having value $w_{ij}(t)$.

The fitting problem can now be cast as a convex optimization program

$$\begin{aligned} & \text{maximize} && L \\ & \text{subject to} && \mu_i \geq \varepsilon, w_{ij} \succeq 0 \\ & && z_{ii} = 1 - \mathbf{1}^T w_{ii}, z_{ij} = \mathbf{1}^T w_{ij} \\ & && z_{ii} \geq \varepsilon, \frac{z_{12}}{z_{11}} \leq z_{22} \\ & && z_{12} = z_{21} + s, s \geq 0 \end{aligned}$$

where the log likelihood function is in the form of (1.9) with $\lambda_i(t)$ now replaced by (1.14). The various constraints correspond to positivity and stationarity. The piecewise linear form of $h_{ij}(t)$ results in many more variables, $w_{ij}(t)$, with a typical size of 200. Still it can now be readily solved as a convex optimization problem using a primal-dual interior point solver for instance. The algorithm runs much, much faster and it comes with a global optimality certificate.

It can be shown that this model corresponds to a linear Poisson regression with parameter constraints where the features ($N_i(t)$) are past data. Naturally, we can consider using a richer set of (possibly non-linear) features that incorporates other potentially useful information such as price, volume, market index and option data. Since we have very large amounts of high-frequency data, it would be possible to avoid the overfitting that would normally accompany the introduction of more parameters. That’s why the ability of our model to handle diverse sources of information is a major advantage it has over the Exponential Hawkes Model.

Chapter 2: Application of Hawkes Models

We develop a model for trade arrival times, fit empirical data to the model by calculating parameters, and test how well the data fits the model. We give preliminary results on a very simple trading strategy which shows promising results.

2.1 Trade Data Collection and Processing

We gathered data on several stocks including YHOO, SNDK, VPHM, XOM, MSFT and TIE from NYSE TAQ database. Our stocks were chosen to span the spectrum of capitalization and volatility. Since the time resolution of the TAQ data is 1 second, we aggregated all our transaction information in one-second intervals by calculating the overall number of buy and sell trades in the interval. In order to classify trades into buy and sell, we employed the Lee-Ready tick test [7]: we compared the price of the trade to the price of the trade 5 ticks ago and classified into buy or sell if the trade price was higher or lower, respectively. In case of a tie, we compared to the next most recent trade (e.g. 4 ticks ago). It would be important to explore other methods of trade classification, including using bid/ask information from the NYSE TAQ database. While recent studies [8] indicate that the tick test performs well for some tasks, it would still be worthwhile to experiment with other classification methods, since our model is very dependent on the qualities of the assigned trade direction.

2.2 Model Fitting

After extracting data from TAQ database and separating trades into buy and sell orders, in Figure 2.1 and 2.2 we plot the most recently frequencies of buy or sell trades conditioned on buy or sell trades, respectively. We easily observe a large self-exciting effect and a small cross-exciting effect, which motivate the use of Hawkes processes to model order arrivals.

Next, we implemented parameter estimation for a bivariate Hawkes process considering both the self-exciting and the cross-exciting effect. We first considered the univariate Hawkes model using exponential decays. The prediction of future intensity involves a weighted-sum of past trades. We significantly accelerated the MATLAB code for model parameter estimation by replacing an iteration loop with a convolution operation. For training data of size up to 20,000 the parameter optimization can be completed in about 35 seconds, which opens the possibility of real-time online parameter update. The bivariate Hawkes process with exponential is given in (1.2) and (1.3). When predicting $\lambda_1(t)$, $N_1(t)$ is the self-exciting process and $N_2(t)$ is the cross-exciting process. A set of typical parameter values is obtained by optimizing over the trading data of MSFT on January 03, 2007.

$$\mu_1 = 0.8769 \quad \alpha_{11} = 0.1689 \quad \beta_{11} = 0.2470 \quad \alpha_{12} = 0.5542 \quad \beta_{12} = 10.7996.$$

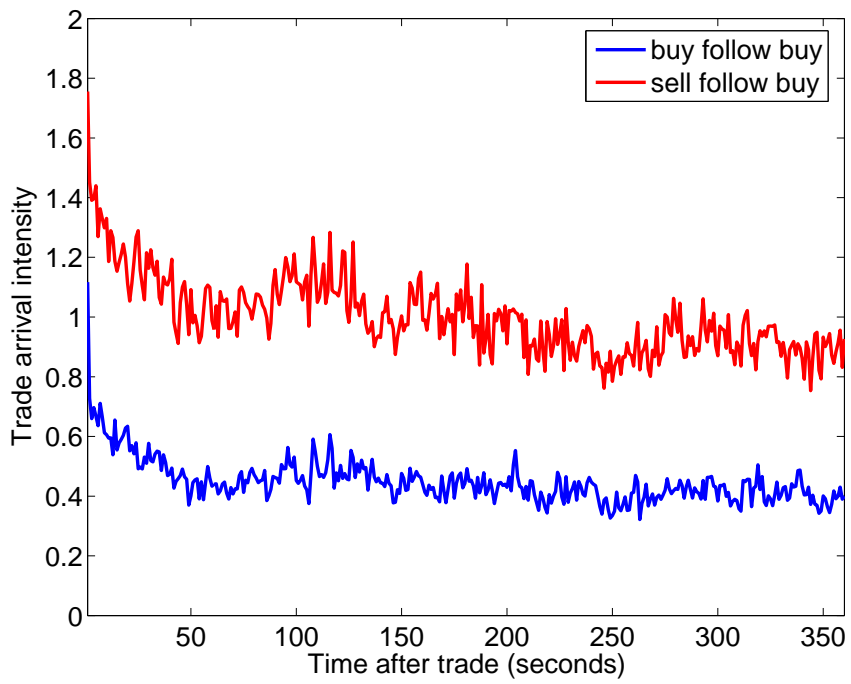


Figure 2.1: Conditional arrival intensities following buy orders, SanDisk (SNDK) 01/02/07

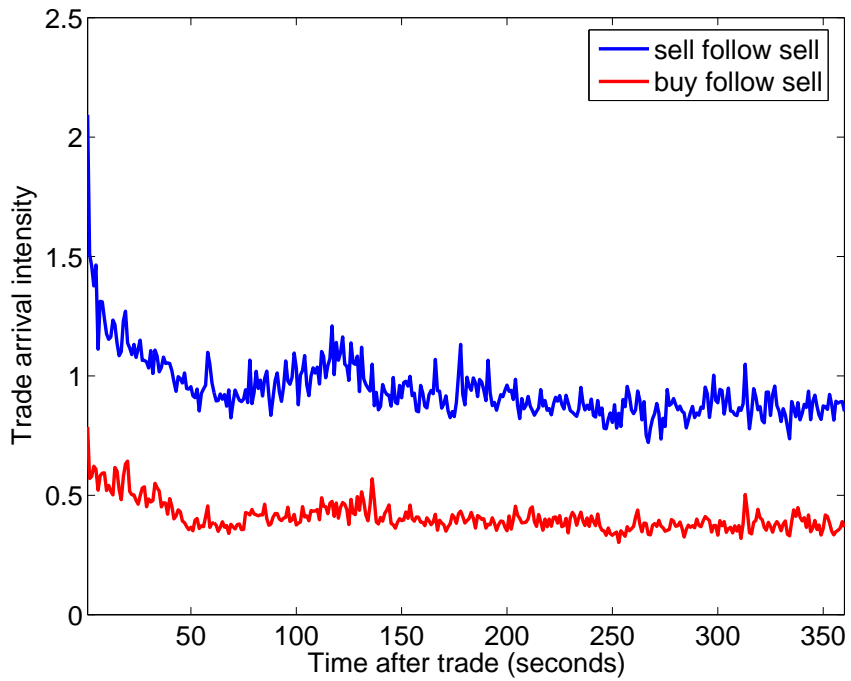


Figure 2.2: Conditional arrival intensities following sellorders, SanDisk (SNDK) 01/02/07

The large exponent β_{12} implies the marginal effect from cross-exciting process as observed in [1]. The relative small exponent β_{11} suggests the clustering effect of trades - a large trade burst is more likely to be followed by trades of the same type. Figure 2.3 illustrates idea. Although the intensity prediction hardly captures the first tic of trade burst, it well predicts the increase of trade intensity that would follow.

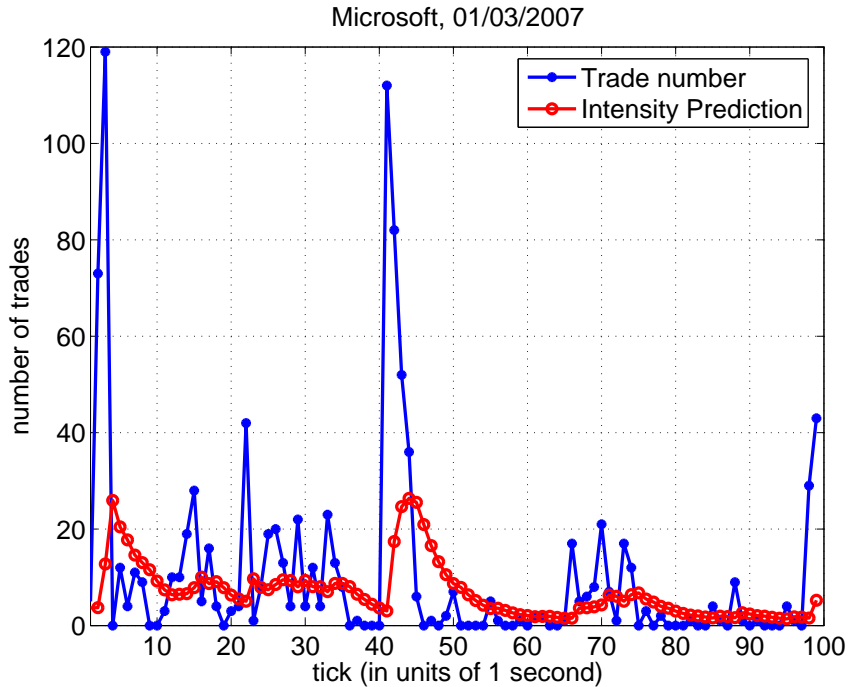


Figure 2.3: Intensity Prediction with exponential Hawkes, MSFT 01/03/2007

We can also work with the generalized Hawkes process model (1.14). Instead of constraining the weighted sum of past trades to have exponentially decaying weights, we assume the weights are free variables themselves. It seems that we have dramatically increased the complexity, since we now have $2n$ parameters instead of just a handful of α 's and β 's. For example, if we want to consider a history of $K = 50$ past arrivals we would have 100 variables. However, as explained before, it is an astonishing fact that this model is superior to the exponential model in both speed and fidelity. The reason is simple: convexity smiles only to this one and turns its back to the exponential. Consequently we can get faster and better results. In Figure 2.4 we plot a sample of the weights estimation from both the exponential model and the generalized model. The maximum likelihood objective value is observed to be 33% higher for the generalized model as compared to the exponential model.

Figures 2.1 and 2.2 show that there is a very small cross-exciting effect between buy and sell orders, and the model fitting picks up this effect. This phenomenon is observed in [1]. Due to the minimal cross-excitation, we will primarily consider the univariate Hawkes model and fit two processes to buy and sell orders separately.

Having calculated the parameters of the Hawkes process, we determine how well those parameters reflect the data through model validation. In Figure 2.5 and 2.6 we show the QQ-plot for both the exponential and the generalized model. It is seen that the Hawkes process models trading data reasonably well. We know that the Hawkes process becomes a standard Poisson process under the stochastic time change

$$t \rightarrow \int_0^t \lambda(s) ds.$$

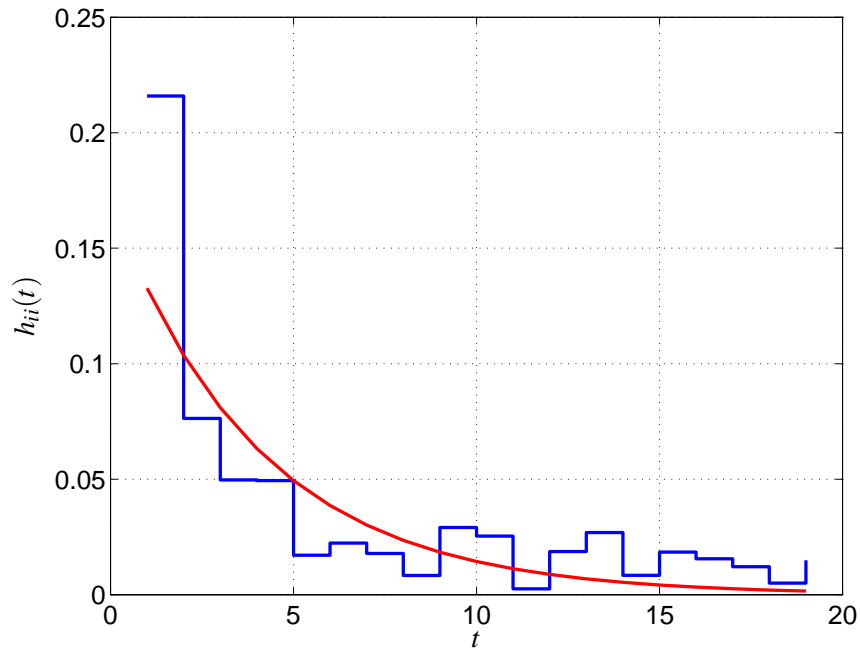


Figure 2.4: YHOO, 07/03/2006

In the discrete case, $\lambda(i)/N(i)$ is approximately the inter-arrival time of the standard Poisson process, which has an exponential distribution. The comparison between the empirical cdf of the inter-arrival time and a standard exponential distribution shows a roughly straight line. This serves as a validation of the Hawkes process model.

2.3 Trading Strategy and Performance

2.3.1 Setup

Given that we can predict buy and sell intensities, we attempt to create some profitable strategies based upon the Hawkes process model for arrival times. Specifically, we employ the exponential and generalized piecewise linear Hawkes processes. Because of the time efficiency and increased accuracy of the generalized process, we use it exclusively unless otherwise stated.

We take our data from the NYSE TAQ Database. Because the data is recorded by second, we discretize time in seconds, which conveniently allows an application of the generalized piecewise linear Hawkes process.

In terms of execution, we assume execution at the worst price recorded in the five seconds following the time of a signal; i.e., we buy for the highest price over the next five seconds and sell for the lowest price over the next five seconds. We do not impose any additional transaction costs nor short selling constraints.

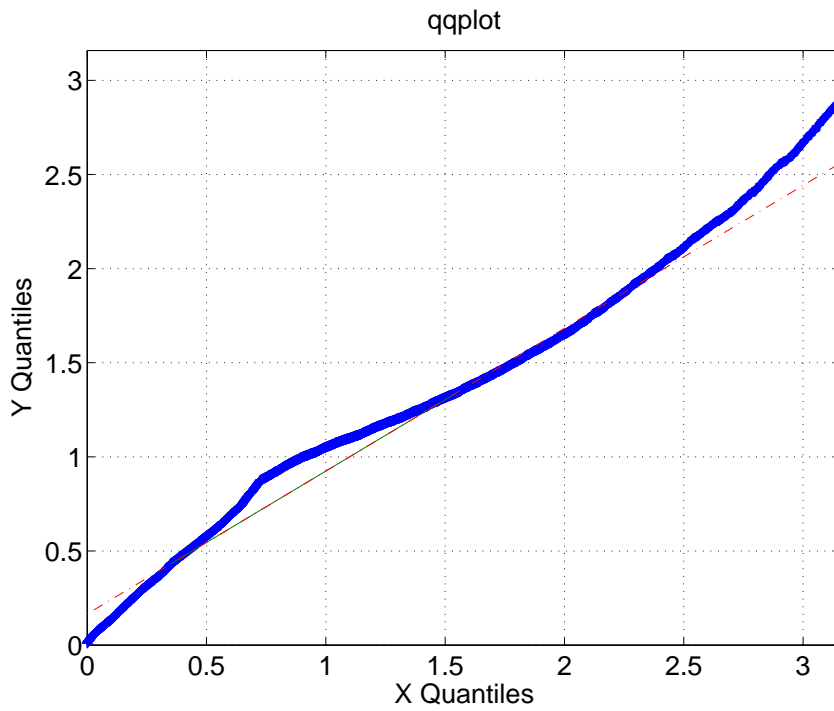


Figure 2.5: QQ-plot, exponential Hawkes process, MSFT 01/03/2007

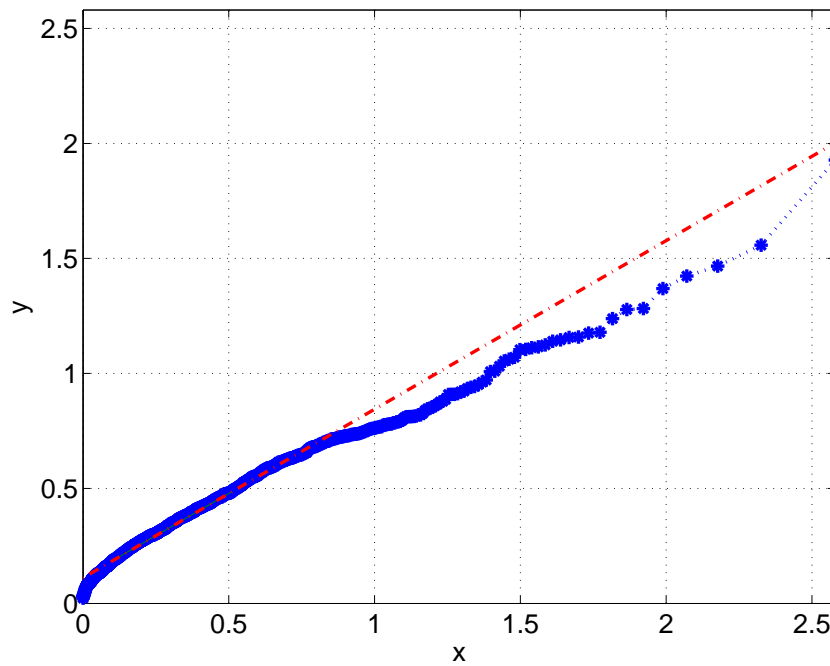


Figure 2.6: QQ-plot, generalized Hawkes process, YHOO, 07/03/2006

Due to the efficient maximum likelihood estimation possible for the generalized piecewise linear model, we can re-estimate parameters frequently; for the purposes of this study, however, we estimate them each day over one day's data, then apply those parameters in trading the next day. We then use those parameters to calculate buy and sell intensities at each second over the trading day, using that information to make trading decisions.

In order to test the model's usefulness, as opposed to that of some complex strategy, we experimented with a very simple trading strategy. We measure performance in dollars earned per single share traded, limiting the position size to one share long or short. We buy the stock when the buy intensity to sell intensity ratio is greater than some fixed constant, ρ , and sell when the sell intensity to buy intensity ratio is greater than ρ . There is no explicit exit strategy, which means that we can only reverse positions when receiving the opposite signal. In order to mark to market, we liquidate positions at the end of the day.

We calibrated ρ using a small portion of the sample data, raising it so as to eliminate most of the noisy trade signals, leaving one or two of the highest intensity ratio opportunities.

2.3.2 Results

Applying this strategy to Microsoft (MSFT) stock using both the exponential and generalized Hawkes processes, we obtain the following results, on average per trade:

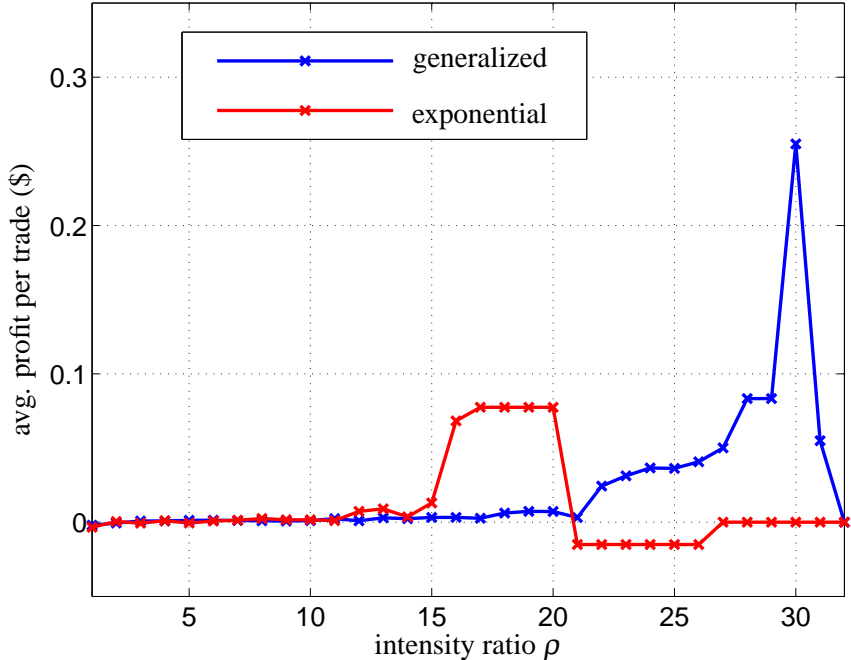


Figure 2.7: Performance for MSFT, 01/03/07-01/11/07

Note how while the generalized process has a similar exponential decay of past trades' effects, it puts much greater weight on those trades in the last second interval as observed in Figure 2.4. Consequently, in times of high activity of one type of trade, that intensity is calculated much higher by the piecewise linear model, resulting in the need for a higher intensity ratio ρ in order to eliminate noise trades. We

then apply the strategy to Titanium Metals Corp. (TIE), achieving exciting results as shown in Figure 2.8. With $\rho = 30$, the average profit per trade is \$.087 over 86 trades (43 roundturns).

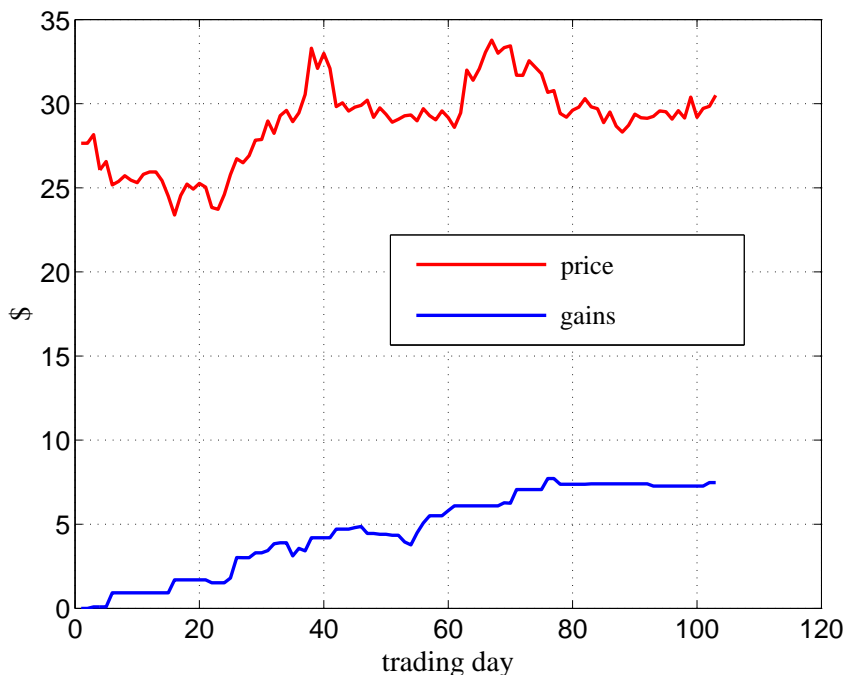


Figure 2.8: Performance for TIE, 09/01/06-01/11/07

Here are the results for some other stocks showing the average profit per share per day over varying lengths of time:

stock	average profit/stock/day	number of trades	ρ	period
YHOO	\$.00565	586	17	07/03/06 – 01/31/07
SNDK	\$.00755	36	15	08/01/06 – 12/29/06
VPHM	\$.01322	14	15	03/08/06 – 03/24/06
XOM	\$.08152	24	7.5	01/03/07 – 01/11/07

Note that Exxon Mobile (XOM) requires a lower threshold ratio, likely resulting from its high daily volume; indeed, setting ρ up to even 10 will eliminate all potential trade signals.

2.3.3 Further Improvements

It is possible that this strategy could be improved numerous ways. We could buy in certain quantities dependent on the intensity ratio and, therefore, its predictive accuracy for the current trend.

Since the current strategy can only reverse positions, it often closes positions when the market has already turned around, which most certainly reduces profits. We must develop an explicit exit strategy depending on intensity, price, time, etc., in order to maximize profitability. Of course, we have the option of placing this information in the strategy itself or in the calculation of the intensities, through choosing different features.

The high intensity ratio should also be relaxed with better exit strategies. The current strategy only captures a handful of opportunities every day, but these seem to actually predict longer term trends, or the strategy could not profit. Probably, an intensity ratio above a certain high threshold implies a flurry of buying or selling that could only occur from some fundamental change in the underlying (e.g., an earnings report). Thus, the technical indicator could be picking up on investors' reaction to a fundamental change. Without an explicit exit strategy, however, the trading simulation simply gives back too much profit between trades, and the resulting signals seem more like noise when viewed under performance.

Given that this naïve strategy produces consistent profits under conservative fill estimates, the application of the Hawkes process to model buy and sell orders in trading strategies seems very promising. Since the strategy has shown profitable results for many different kinds of stocks, future improvements upon it should have reasonable scaling capacity among a variety of securities.

Bibliography

- [1] P. Hewlett, *Clustering of order arrivals, price impact and trade path optimization*. May 2006. Available at [http://www.stanford.edu/~barmbrus/2007msande444/Hewlett2006 price impact.pdf](http://www.stanford.edu/~barmbrus/2007msande444/Hewlett2006%20price%20impact.pdf)
- [2] D. Cox and P. Lewis, *The statistical analysis of series of events*. London: Methuen, 1966.
- [3] P. Lewis, *A branching Poisson process model for the analysis of computer failure patterns*. J. R. Statist. Soc., B 26, 398-456, 1964.
- [4] A. Hawkes, *Spectral of some self-exciting and mutually exciting point processes*. Biometrika, vol. 58, no. 1, pp. 83-90, 1971.
- [5] C. Bowsher, *Modelling security events in continuous time: Intensity based, multivariate point process models*. Nuffield College Economics Discussion Papers, 2005W-26, 2005.
- [6] Y. Ogata, *The asymptotic behavior of maximum likelihood estimators for stationary point processes*. Ann. Inst. Statist. Math., vol. 30, Part A, 243-261, 1978.
- [7] C. Lee and M. Ready, *Inferring trade direction from intraday data*. Journal of Finance, vol. 46, 733-746, 1991.
- [8] T. Finucane, *A direct test of methods for inferring trade direction from intra-day data*. The Journal of Financial and Quantitative Analysis, Vol. 35, No. 4, pp. 553-576, Dec. 2000.