MODELING ULTRA-HIGH-FREQUENCY DATA: THE S&P 500 FUTURE INDEX

Silvia Centanni[1] and Marco Minozzo[1]

[1] Department of Economics, Finance and Statistics
University of Perugia
(e-mail: silvia@stat.unipg.it, minozzo@stat.unipg.it)

ABSTRACT: In recent years, marked point processes have found a natural application in the modeling of ultra-high-frequency financial data. The use of these processes does not require the integration of the data which is usually needed by other modeling approaches. Two main classes of marked point processes have so far been proposed to model financial data: the class of the autoregressive conditional duration models of Engle and Russel and the broad class of doubly stochastic Poisson processes with marks. In this paper we show how to model an ultra-high-frequency data set relative to the prices of the S&P 500 future (SPU01) using a particular class of doubly stochastic Poisson process. Our models allow a natural interpretation of the underlying intensity in terms of news reaching the market and does not require the use of ad-hoc methods to treat the seasonalities. Estimation and filtering is carried out through reversible jump Markov chain Monte Carlo algorithms.

KEYWORDS: Cox process; Marked point process; Monte Carlo Expectation Maximization; Reversible jump Markov chain Monte Carlo; Shot noise process; Tick by tick data.

1 Introduction

Recently, marked point processes (MPP) have found application in the modeling of ultra-high-frequency (UHF) financial data. In UHF databases, for each market event, such as a trade or a quote update by a market maker, the time at which it took place, together with its characteristics, for instance, the price and volume of the transaction, is recorded (Guillaume et al., 1999). Two main classes of models based on MPP have so far been proposed for these data: the class of the autoregressive conditional duration (ACD) models of Engle & Russel (1998) and the class based on doubly stochastic Poisson processes (DSPP) with marks (Rydberg & Shephard, 2000).

In this paper we consider the modeling of an UHF data set relative to the prices of the S&P 500 future (SPU01) using a class of marked DSPP in which the stochastic part of the intensity process has a shot noise form. In particular, as in Centanni & Minozzo (2006), we consider an intensity process which is a given function of a non-explosive MPP with positive jumps that we characterize through the distributions of jump times and sizes. Such an intensity process can be viewed as a generalization of the classical shot noise process (Cox & Isham, 1980).

Let us denote with T_i and Z_i, $i \in \mathbf{N}$, the times and sizes of the ith logreturn change, that is, $Z_i = \ln(S_{T_i}/S_{T_{i-1}})$, where S_t is the price of the future on the S&P 500 index in

t. Then, for a filtered probability space $\left(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}_{t\in[0,T]}\right)$, we assume $(T_i, Z_i)_{i\in\mathbf{N}}$ be a DSPP with marks with intensity process δ. For this process we will assume that $\delta_t = a(t) + b\lambda_t$, where $a(\cdot)$ is an integrable $\mathbb{R}^+$-valued deterministic function, b is a nonnegative parameter, and the process λ has the form $\lambda_t = \sum_{j=0}^{N'_t} X_j\ e^{-k(t-\tau_j)}$, where $(\tau_j, X_j)_{j\in\mathbf{N}}$ is another MPP with $X_j > 0$, and where $N'_t = \#\{j : \tau_j \leq t\}$. For this latter MPP, let us also assume that the conditional distributions $p(\tau_j|\tau_1,\ldots,\tau_{j-1})$ and $p(X_j|\tau_1,\ldots,\tau_j,X_1,\ldots,X_{j-1})$ admit density.

This model allows a natural interpretation for the stochastic changes of the process λ in terms of market perturbations caused by the arrival of relevant news (Kalev et al., 2004). When the jth item of news reaches the market, a sudden increase X_j in trading activity occurs, depending on the importance of the item, followed by a progressive normalization. The random variable τ_j represents the time of the arrival of the jth item of news. The parameter k expresses the speed of absorption of the effect of the news by the market, while $a(\cdot)$ can be interpreted as the activity that the market would have in absence of random perturbations. By adequately choosing the function $a(\cdot)$, it is possible to take into account many of the deterministic features (seasonality, etc.) characterizing intraday price data (see, for example, Guillaume et al., 1999). A particularly simple model belonging to this class is the basic model described in Centanni & Minozzo (2006), in which the logreturn Z_i are assumed independently distributed and the parameters of the process λ are the mean interarrival time ν and the mean size of market perturbations $1/\gamma$.

2 The S&P 500 Future Data Set

We considered all price changes (39,889) from the 9th to the 27th of July 2001 (15 trading days). The market was open from 8.30 in the morning to 15.15 in the afternoon. The data set showed successive logreturns to be not autocorrelated and also not correlated with the waiting times $T_i - T_{i-1}$. Moreover, the time between successive price changes showed to be exponentially distributed. To account for the intraday seasonality in the data, every one of the 15 days considered (each day corresponds to 405 minutes, for a total time horizon of $T = 6{,}075$ minutes) has been sliced into 27 subintervals (15 minutes long). The number of changes of the price of the S&P 500 future within each of the 15 minutes intervals has been calculated. Circles in Figure 1 show, for the 27 intraday intervals, the minimum number of changes of the price of the future (in the corresponding interval) over the 15 days. The full line, obtained from the least square fit (of a second order polynomial) by subtracting a constant equal to 1.2, can be assumed to provide the deterministic part of the intensity process accounting for the intraday seasonality. Thus, to model the data we consider an intensity of the form, for $t \geq 0$, $a_0 + a_1 t + a_2 t^2 + \lambda_t$, where $a_0 = 6.7226$, $a_1 = -0.0306$, $a_2 = 0.0001$. Let us note that the chosen shape for the deterministic part of the intensity is standard in UHF data analysis (see, for instance, Figure 4 of Rydberg & Shephard (2000)).

As usual in financial applications where it is reasonable to assume a partial information setting in which market agents are restricted to observe only the history of

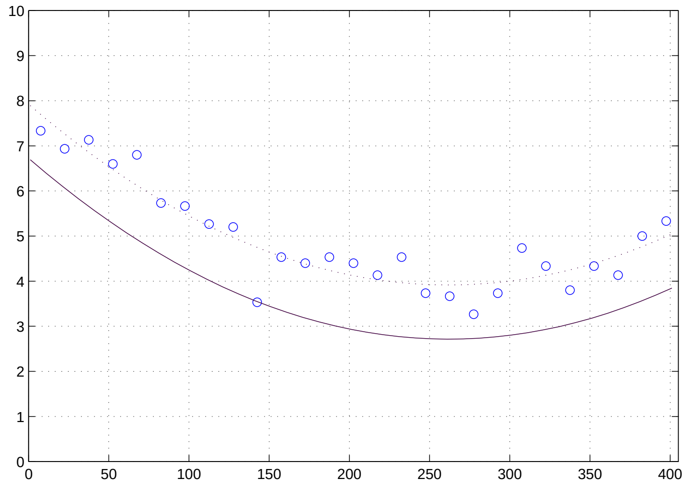

Figura 1. *Detrending of the intraday seasonality in the S&P 500 data set. Every one of the 15 days considered (each day corresponding to 405 minutes, for a total of* $T = 6{,}075$ *minutes) has been sliced into 27 intervals (15 minutes long). The number of changes of the price of the S&P 500 future within each of the 15 minutes intervals has been calculated. Circles in the graph report, for the 27 intraday intervals, the minimum number of changes of the price (in the corresponding interval) over the 15 days. The dotted line is the least square fit of a second order polynomial. The full line is obtained from the least square fit by subtracting a constant value equal to 1.2, and is used as a deterministic 'offset' for the intensity process.*

stock price, we will assume that only the past times and sizes of price changes of the S&P 500 future are observed and not the intensity process. For the following analysis we assume also that the basic model holds.

Estimates of the parameters $\theta = (k, \gamma, \nu)$ were obtained by implementing a stochastic EM (SEM) algorithm as in Centanni & Minozzo (2006). We performed a run of 450 steps of the SEM algorithm (each step involving 6,000 updates of a reversible jump Markov chain Monte Carlo (RJMCMC)). Only the 6,000th sampled intensity is used in the E and M steps of the SEM. This is also used as the initial intensity in the RJMCMC at the subsequent step of the SEM. Initial values for the SEM algorithm are $k = .001$, $\gamma = 25$, $\nu = .1$. Point estimates of k, γ and ν are provided by the sample means over the last 150 steps of the SEM algorithm, and give $\hat{k} = .006$, $\hat{\gamma} = 2.875$ and $\hat{\nu} = .050$.

The filtering of the intensity was performed by running a RJMCMC algorithm assuming the above estimates as the true parameter values. The graph at the top of Figure 2 shows the filtering expectation $\mathrm{E}(\lambda_t | \phi_0^T)$, $t = 0, 5, 10, 15, \ldots, T = 6{,}075$. After a burn in period of 10,000, the RJMCMC algorithm was run for other 10,000 updates, recording a trajectory every 2. The graph at the bottom shows the total (filtered) intensity comprehensive also of the deterministic part shown in Figure 1. The peaks in the top graph show periods of unexpected high trading activity, not accounted for

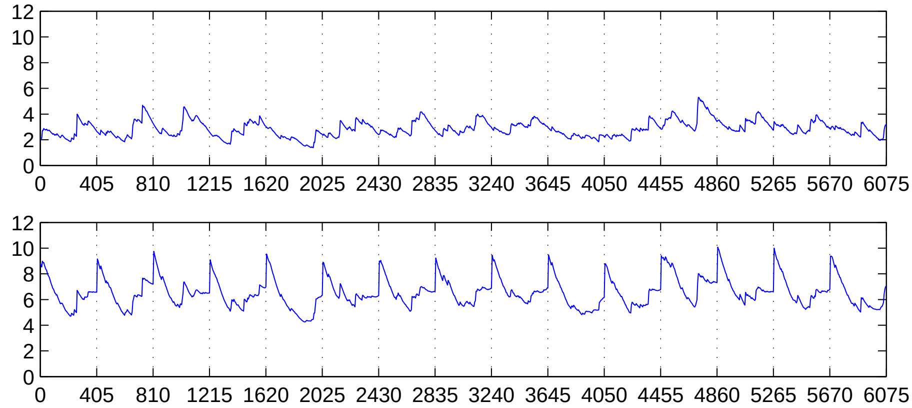

Figura 2. *The graph at the top shows the filtering expectation* $\mathrm{E}(\lambda_t|\phi_0^T)$, $t = 0, 5, 10, 15, \ldots, T =$ 6,075, *under the model described in the text, assuming the parameter values:* $k = .006$, $\gamma = 2.875$, $\nu = .050$. *After a burn in period of* 10,000, *the RJMCMC algorithm was run for other* 10,000 *updates, recording a trajectory every 2. The graph at the bottom shows the total (filtered) intensity comprehensive also of the deterministic part given in Figure 1.*

by the intraday seasonality, which may correspond to the arrival of relevant news. The filtered intensity can be used in the numerical computation of quantities of interest in a number of financial problems, such as price forecasting.

References

CENTANNI, S., & MINOZZO, M. 2006. A Monte Carlo approach to filtering for a class of marked doubly stochastic Poisson processes. *Journal of the American Statistical Association*.

COX, D. R., & ISHAM, V. 1980. *Point Processes*. London: Chapman and Hall.

ENGLE, R. F., & RUSSEL, J. R. 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, **66**, 1127–1162.

GUILLAUME, D. M., DACOROGNA, M. M., DAVÉ, R. R., MÜLLER, U. A., OLSEN, R. B., & PICTET, O. V. 1999. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, **1**, 95–129.

KALEV, P. S., LIU, W. M., PHAM, P. K., & JARNECIC, E. 2004. Public information arrival and volatility of intraday stock returns. *Journal of Banking & Finance*, **28**, 1441–1467.

RYDBERG, T. H., & SHEPHARD, N. 2000. A modelling framework for the prices and times of trades made on the New York stock exchange. In *Nonlinear and Nonstationary Signal Processing*, eds. W. J. Fitzgerald, R. L. Smith, A. T. Walden and P. C. Young, pp. 217–246.