Particle Filtering in High-Frequency data.

A. Platania
Facoltá di Scienze Statistiche ed Economiche
Via Cesare Battisti, 241
35121 Padova Italy
platania@stat.unipd.it

L.C.G. Rogers
Statistical Laboratory
Wilberforce Road
Cambridge CB3 0WB, GB
L.C.G.Rogers@statslab.cam.ac.uk

October 2004

Abstract

We propose a Markov model for high-frequency quote and trade data for equities, to be estimated using particle filtering techniques. Interpreting the state of the Markov chain as the activity of the market, we are able to identify this with considerable precision.

1 Introduction

High frequency data record every quote or transaction registered in a market, so that in principle they should provide more *information* than any subset like commonly used daily data. Nonetheless the usual Gaussian independent increments paradigm for (log) prices completely fails in this settings. Intuitively, with a larger time scale prices result as a sum of many micro-movements and benefit of some Central Limit effect, while on a very short time scale such an approach fails to capture the main features of the situation.

This challenging topic is currently addressed both by academia and pratictioners, but literature is still somehow inadequate and there is no well established tecnique. If researchers focused more on the understanding of market behaviour through a microstructure approach and on the study of the statistical properties of high frequency data, there is a strong market demand for simple models giving trading suggestions in real time (see Dacorogna, Gencay, Muller, Olsen and Pictet for an account).

In what follows we study a stochastic volatility model for market timing. Analogous approaches has already been considered in past literature (see Rogers and Zane and Rydberg and Shephard), but we propose an hybrid particle based tecnique to address the corresponding filtering problem. We will first explore the data, finding some stylized facts that naturally support our main assumptions. We then apply the resulting model to various dataset, obtaining simple and accurate estimates.

This paper is organized as follows: in Section 2 we introduce some terminology and we give a brief description of some basic features of order based markets, in Section 3 we describe our model and discuss the correspondent filtering problem, in Section 4 we propose a filtering algorithm for discrete time and finite space Hidden Markov Models with unknown parameters, in Section 5 we apply our tecnique to simulated and market data, in Section 6 we perform some simple diagnostic

Description of the model...

Filtering problem and related literature...

Applications, results, diagnostic...

Trading rules?

2 Order based markets

2.1 Basic terminology

Most modern financial markets operate continuously, and the typical mismatch between buyers and sellers is solved via an *order based market.* A *market order* is the request to buy or sell immediately at the best available price, while a *limit order* or a *quote* states a limit price corresponding to the worst allowable price for the transaction. Buy limit orders are called *bids*, sell limit orders *offers* or *asks*, and the best quotes, that is the lower ask and the higher bid, are called *inside quotes*, their typical non-zero gap being the *spread.* This quantity is very important in order to understand the dynamics of supply and demand at the trade execution level, matching buyers and sellers, and the process of price negotiation in the market.

As market orders arrive they are matched against quotes of the opposit sign in order of price and arrival time. For example, if bid prices exceed the best ask price or ask prices are less than the best bid, they are called *crossing* or *marketable* limit orders, and during this rare events the bid-ask spread will be negative, creating a concrete arbitrage opportunity.

2.2 Some stylized facts about high frequency data

Let us start with an exploratory analysis of the data to show how some simple observations lead to the specification of a model based on natural assumptions.

First of all, high frequency data are essentially discrete in space and time, as prices have discrete quanta called *ticks* (the minimum interval that price change on is then known as tick size) and transaction times are available at short time intervals. Thus, even if most continuous modelling could account for this with appropriate corrections, it seems convenient to start with a discrete time and space framework.

In this paper we will decouple the price formation problem from the time process, trying to define a model for the instants when quotes are posted. This step is in any case fundamental in the following choice of a suitable price process model. For example, one could observe that an high density of limit orders results in high *liquidity*, and try to link such liquidity effects with price movements. On the other hand, some studies seem to support the idea of prices being driven by a random mechanism independent of the process of times, see Rogers and Zane.

Our issue is to define a proper point process model for the arrival times of market and limit orders, and it is therefore clear that the key variable for our study is market intensity[1]. We are going to define this concept later, for the time being let us look at some plots describing simple measures of market activity.

In Figure 2 we consider a dataset for all the transactions registered in an FX market during the period between Feb and May 2004. Quotes are posted one at the time, each reporting transaction time, price and ask or bid side. We computed the mean number of transactions over the four months for all the twenty seconds periods between 8.00 and 19.00, obtaining a clear intraday profile (IDP). This shape is typical of many other markets, and represents the most important periodic component of the time series. Notice that it is almost impossible to distinguish between ask and bid IDP.

Figure 2 suggest immediately that the intensity of our counting process should not be constant, but it may give the wrong impression that market activity is a deterministic function of time. In fact the deviance between the actual number of observed transactions and their IDP is far from being explained so easily, and many events such as unespected information entering the market can increase suddently market intensity.

If the assumption of intensity being a deterministic function of time was true and the IDP described it closely, we could think that for each twenty second period our observations are generated from a Poisson random variable with intensity equal to the correspondent IDP level.

Using the same dataset as in Figure 2, Figure 3 tests this null hypothesis with a very simple argument. The first subplot shows the density of the observed number of posted quotes over their correspondent IDP (continuous line) togheter with the density of data simulated using deterministic time dependence on the IDP.

As every day included in the dataset we have one (eventually null) observation for all twenty seconds periods, it is enough to simulate an equal number of Poisson samples

[1]We will refer indifferently to market intensity or market activity.

from each of them, using the correspondent IDP as the intensity value and finally normalizing to obtain a density estimate (dashed line). In the second subplot we compute the difference between the two densities. Market activity often tends to be below their null hypothesis IDP value, while sometimes it is far (twice or three times) bigger. This immediately suggests the use of a Markov chain, with the states representing different levels of departure of the observations from the IDP and its transition matrix describing the probability of switching from one to another.

3 A model for market timing

3.1 Poisson process with Markov driven intensity

Our first step strategy is to model the time at which market and limit orders take place. The empirical analysis in Section 2 proves that an assumption of constant or time dependent volatility are inadequate to explain our data. We therefore need to define a counting process with stochastic volatility, and again our previous study suggests that a Markov chain could capture the randomness remaining after accounting for the time dependent part.

Formally, we model the counting process $N(t, \omega)$ of the number of transactions observed in the interval $[0, t]$ as a Poisson process with *random intensity*

$$\lambda_t(\omega) = f(X_t(\omega))\lambda_0(t), \tag{1}$$

where X is an underlying Markov process, $f(\cdot) \geq 0$ is some nonnegative function and $\lambda_0(\cdot)$ is the IDP of activity function. Notice that the only source of randomness in λ comes from the Markov bit: in the following we will assume X to be a finite Markov chain, so that $f(X_t)$ can assume a finite number of values and the resulting filtering problem is more tractable.

It is now easy to define the intensity of trading or market *activity* using a simple moving average such as

$$A_t(\alpha) = \alpha \int_0^t \exp(\alpha(s-t))\mathrm{d}N_s, \tag{2}$$

where $\alpha > 0$ is some positive parameter. Rogers and Zane consider functionals of the share history analogous to the definition in Equation (2), computing analitically their moments under various specifications of the underlying Markov process. The advantage of such quantities is that they are easy to update: in our example $A_t(\alpha)$ decreases exponentially between quotes and jump up by α when a new quote arrives.

3.2 The filtering problem

Even assuming the underlying process to be a simple Markov chain, as soon as we increase the number of states the resulting filtering problem is quite hard. This kind of models are very well known in literature, arising frequently in engineering where they are known as *Hidden Markov Models* (HMM).

In a HMM a Markov chain X is observed indirectly through some noisy observations Y, in our case the counting process. A practical way to define a HMM is through their *state space representation*, characterizing them in two steps via a *system model*, that is a first order Markov state process

$$X_t = g_t(X_{t-1}, \nu_{t-1}), \qquad \nu_t \text{ i.i.d. noise}$$

and their *measurement model*, given by the observation process

$$Y_t = c_t(X_t, w_t), \qquad w_t \text{ i.i.d. noise.}$$

The filtering problem is that of the estimation of the posterior density of the state of the Markov chain given the past observations, and it is particularly challenging when the state space representation depends on some unknown parameters. This is indeed our case, as the transition probabilities of the underlying Markov chain and the intensities of the Poisson process are generally unknown.

BRIEF INTRO TO EXISTING LITERATURE, TO IMPROVE...
In literature, our particular model is known as *Markov Modulated Poisson Process*, and it arises in target tracking, communication systems and many other applications. There are two main approaches for the correspondent filtering problem. The first one (see Elliott and Krishnamurthy) is based on a change of measure argument, but works efficiently only when the underlying parameters are known. In the unknown parameters case it is possible to use EMM style tecniques, but these are both slowly convergent and computationally heavy.

The second one is based on maximum likelihood, but it is clearly impossible to compute the (log) likelihood directly and approximations are needed. Many recent works study the convergence of such estimators, but again the unknown parmeters case is little developed.

4 The algorithm

We now propose a particle filtering algorithm to simulate from the posterior of the state of the Markov chain using an acceptance/rejection tecnique. The usual basic idea (see also the Appendix) is to approximate this unknown density through a Random Measure $\{X_t^i, w_t^i\}_{i=1}^{N_s}$ consisting of a set of support points and associated weights

$$p(X_t|Y_{1:t}) \approx \sum_{i=1}^{N_s} w_t^i \delta(X_t, X_t^i), \tag{3}$$

where N_s is the number of particles, δ the Dirac measure and $Y_{1:t} = \{Y_1, \ldots, Y_t\}$.

Our posterior $p(X_t, \theta|Y_{1:t})$ depends also on additional parameters $\theta = \{\Pi, \mu, \lambda\}$, namely

- the transition probabilities $\Pi = \{\pi_{ij}\} = [\pi_1, \ldots, \pi_s]$ where π_{ij} is the transition probability $p(X_t = j|X_{t-1} = i)$ for i, j in some finite set $S = \{1, \ldots, s\}$ and $\pi_i = [\pi_{i1}, \ldots, \pi_{is}]$ for $i = 1, \ldots, s$
- the initial distribution of the chain $\mu = [\mu_1, \ldots, \mu_s]^T$ with $\mu_i = p(X_1 = i)$
- the Poisson intensities $\lambda = [\lambda_1, \ldots, \lambda_s]$

and we can factorize it as

$$\begin{aligned} p(X_t, \theta|Y_{1:t}) &\propto p(Y_t|X_t, \theta, Y_{1:t-1}) p(X_t|\theta, Y_{1:t-1}) p(\theta|Y_{1:t-1}) \\ &\propto p(Y_t|X_t, \lambda) p(X_t|\Pi, \mu) p(\Pi, \mu) p(\lambda|Y_{1:t-1}), \end{aligned} \tag{4}$$

since Y is Poisson of intensity $\lambda = f(X_t)$, for the Markov property X_t does not depend on $Y_{1:t-1}$ and the parameters are conditionally independent given the past observations.

Notice that the posterior in Equation (4) depends on Π and μ only through two isolated terms, and that we can now integrate them out under mild assumptions using the same calculations of Doucet and Ristic. Assigning a Dirichlet distribution on μ and the rows of Π

$$\mu \sim \mathcal{D}(\alpha_0) \quad \pi_i \sim \mathcal{D}(\alpha_i),$$

where the αs are s-dimensional vectors representing prior knowledge on the parameters, one can infact compute analitically

$$p(X_t) = \int p(X_t|\Pi, \mu)p(\Pi, \mu)\mathrm{d}\Pi\mathrm{d}\mu \tag{5}$$

obtaining the following very intuitive estimate of the transition probabilities[2]

$$p(X_t = j|X_{t-1} = i, X_{1:t-2}) = \frac{n_{ij}(X_{1:t-1}) + \alpha_{ij}}{n_i(X_{1:t-1}) + \sum_{j=1}^{s} \alpha_{ij}} \tag{6}$$

for $n_{ij}(X_{1:t}) = \sum_{k=2}^{t} \delta(X_{t-1}, i)\delta(X_t, j)$ path number of transitions from i to j and $n_i(X_{1:t}) = \sum_{j=1}^{s} n_{ij}(X_{1:t})$ occupation time of state i. This quantity can be estimated recursively if X_t was observed, and the parameters of the Dirichlet distributions can be interpreted as prior observation counts.

Substituting (5) in (4) the posterior collapses to

$$p(X_t, \lambda|Y_{1:t}) \propto p(Y_t|X_t, \lambda)p(X_t)p(\lambda|Y_{1:t-1}), \tag{7}$$

where the density $p(\lambda|Y_{1:t-1})$ can be approximated using kernel smoothing tecniques (see Liu and West).

We can now describe a simple recursive acceptance/rejection method to sample from the density in (7). Suppose we want to update $\{X_{t-1}^i, \lambda_{t-1}^i, w_{t-1}^i\}_{i=1}^{N_s}$, Π_{t-1} and μ_{t-1} from time $t-1$ to time t[3]. Then, for each particle $i = 1, \ldots, N_s$

- sample X_t^i using Π_t
- sample λ_t^i from its smooth kernel density
- compute the Poisson likelihood, a bound M_t for the posterior and draw u from a uniform in $(0, 1)$
- calculate $p(X_t^i, \lambda^i|Y_{1:t})/M_t$ and accept if more than u.

[2]see Doucet and Ristic for the details.

[3]Notice that each particle i carries on its own value for λ, while Π and μ are computed globally.

Given the new $\{X_t^i, \lambda_t^i, w_t^i\}_{i=1}^{N_s}$ it is possible to update the estimates in (6) using the particles transitions between $t-1$ and t.

SOMETHING ELSE TO SAY Possible improvements: resampling, rebalancing, intensity correction. Filter generally viable for any HMM with state space decomposition in probabilistic form.

5 Results

5.1 Simulated data

We first tested our algorithm against various datasets simulated from known MMPP. Convergence is fast and reliable up to a six states chain, depending on its stationary distribution, the number of particles and the starting values.

Figure 4 and Figure 5 present a low dimensional but difficult case. Infact the $s = 3$ states chain has some null transition probabilities, the number of particles $N_s = 100$ is very small and we start our filtering procedure using misplaced initial guesses.

Intensity estimates still converge quickly, and it is easy to control their variance choosing carefully the transition kernel in Equation (7). Here the true values $\lambda = [\lambda_1 \ \ \lambda_2 \ \ \lambda_3] = [10 \ \ 20 \ \ 40]$ are remarcably different, simplifying the correspondent estimation, but later we will discuss the performances of our algorithm on a tight intensities case using true market data.

Figure 5 shows how the combination of poor starting points and a small number of particles result in a poor estimation for some of the parameters. For example the estimates of the third row of transition probabilities $\pi_3 = [0.3 \ \ 0.7 \ \ 0]$ are not satisfactory, the reason being that the third state has the smaller occupation time under the invariant density of the chain. Even if in low dimension we can easily solve for this using bigger samples, the choice of the number of states could be critical in other more challenging settings. This issue is especially relevant when we have only a limited amount of time available to update the estimates, but in general a close evaluationt of this tradeoff is always needed when adding new states with low occupation time.

5.2 Modelling market intensities

To perform our analysis on real market data we studied the set of all bid and ask quotes registered on the LIFFE for the FTSE 100 Future with maturity Dec 2002, and the actual transactions on FX US Dollar/Euro data on the period between Feb and May 2004. We will now present some of the estimation results, also giving some intuition on how to solve some related problems succesfully.

As in Section 2 we divided the period of the day between 8.00 and 19.00 in subintervals of twenty seconds, and we used the filtering procedure of Section 4 using the number of observed quotes in the subperiods to estimate the intensities $\lambda = f\,\lambda_0$. In Figure 6 we display the results for the stochastic part $f(X_t)$, that is the departure of market intensity from the mean intraday profile λ_0.

A nice feature of our approach is that it allows to control the variance of the posterior distribution in many different ways. The most direct one is to specify carefully mean and variance of the kernel density $p(\lambda|Y_{1:t-1})$ selecting an 'artificial evolution' for the parameters, for example

$$p(\lambda|Y_{1:t-1}) \approx \sum_{j=1}^{N_s} w_t^{(j)} \mathcal{N}(\lambda|m_t^{(j)}, h^2 V_t)$$

as described in Liu and West [5], where $\mathcal{N}(\cdot|m, S)$ is multivariate normal of mean m and variance S with $m_t^{(j)} = a\lambda_t^{(j)} + (1-a)\bar{\lambda}_t$ (here $a = \sqrt{1 - h_t^2}$ for h slowly decreasing function and $\bar{\lambda}_t$ is the sample mean) and V_t Monte Carlo posterior variance.

In this framework one could also exploit the fact that each observation Y_t should be 'close' to the intensity λ_i with higher posterior probability. So if

$$\hat{\lambda} = [\hat{\lambda}_1 \ldots \hat{\lambda}_s] = \sum_j w_t^{(j)} \lambda_t^{(j)}$$

and $i \in \{1, \ldots, s\}$ is correspondent to the most likely state of the chain, we could correct $\hat{\lambda}_i$ with some distance $d(Y_t, \hat{\lambda}_i)$. Using this improvement on simulated MMPP data we achieved faster convergence.

Further flexibility is obtained *rebalancing* the posterior probabilities in Equation (6). As we said the prior density parameters α can be interpreted as 'prior observation counts'. One way to make transition probability estimates more volatile is to control the number of prior counts. This is a very powerful tool, and should be used if we believe the Markov chain parameters to evolve in time and to avoid quick converge to a fixed point as the number of particles grows.

Intuitively, higher posterior volatility will improve the calibration of observed data but will also decrease the forecasting power of the estimated Markov structure, and in practice all previous choices should be determined in order to find a good balance for this tradeoff. One easy way to understand if we achieved it is to observe our estimates on a shorter time scale as shown in Figure 7, displaying the same estimates of Figure 6 along a period of about five days. Clearly if in Figure 6 we notice how the model follows market movements with sufficient flexibility, in the smaller time window of Figure 7 we can appreciate its stability.

Bid and ask IDPs were almost undistinguishable, and as expected in both plots ask estimates are sligthly bigger for higher intensities but consistently smaller at lower intensities.

To study absolute differences for the estimates is nonetheless important to look at the final estimates of the transition probabilities and of the invariant distribution. Tipically the invariant distribution assigns much more probability to 'low' and 'intermediate' frequency states, so that the correspondent differences in bid/ask intensities are intuitively more significative.

For example using a five states Markov chain on the FTSE 100 futures dataset we obtained

$$\hat{\pi} = \begin{bmatrix} 0.26 & 0.62 & 0.11 & 0 & 0 \\ 0.06 & 0.67 & 0.25 & 0.01 & 0 \\ 0 & 0.31 & 0.58 & 0.1 & 0 \\ 0 & 0.04 & 0.42 & 0.46 & 0.07 \\ 0 & 0 & 0.03 & 0.38 & 0.59 \end{bmatrix}$$

$$\hat{\lambda} = \begin{bmatrix} 0.2421 & 0.5266 & 1.1716 & 1.9780 & 3.0407 \end{bmatrix}$$

while the occupation time for each state of the chain was, in percentage,

$$\begin{bmatrix} 0.041 & 0.456 & 0.386 & 0.097 & 0.02 \end{bmatrix}.$$

It is worth emphasizing again how some of these probabilities are very small, so that in principle we need a large sample to estimate reliably some of the correspondent parameters. However, notice that the structure of the transition probability matrix is diagonal, and that for this dataset the market is characterized by low and average market activity.

Comparison of differences plot and the four/five states case.

These results were obtained on the same data as in Figure 6 and 7, and as expected adding this state the model selected a new low activity level. Intuitively, estimates were driven sensibly as particles are pushing more flexibility exactly where needed (that is the states of the chain with larger occupation time). This is extremely clear looking at the confidence intervals for the intensity estimate using a three states Markov chain on the first fifty observations in the dataset, as in Figure 8. The plot shows the sequence of posterior means along time plus and minus a fixed multiple of the posterior standard deviation. Notice how as the observed deviance from the IDP is smaller than $\min(\hat{\lambda})$ or bigger than $\max(\hat{\lambda})$ the posterior select with great precision the correspondent smaller or bigger intensity state. On the long run, the presence of a very low state for the intensity helps the model to account for prolonged period of market inactivity[4].

6 Diagnostic

[4]E.g. look at the observations near 70000 in Figure 6.

References

[1] S. ARUMPALAM, S. MASKELL, N. GORDON, and T. CLAPP. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, XX:100–117, 2001.

[2] G.M. COSTANTINIDES. A theory of the nominal term structure of interest rates. *Review of Financial Studies.*, 5:531–552, 1992.

[3] A. DOUCET, N. DE FREITAS, and N. GORDON, editors. *Sequential Monte Carlo Methods in Practice.* Springer-Verlag, New York, 2001.

[4] A. DOUCET and B. RISTIC. Recursive state estimation for multiple switching models with unknown transition probabilities. *IEEE Transactions on Aerospace and Electronic Systems*, 38:1098–1104, 2002.

[5] J. LIU and M. WEST. Combined parameter and state estimation in simulation-based filtering. *in Sequential Monte Carlo Methods in Practice.*, 2001.

[6] L.C.G. ROGERS. One for all. *Risk.*, 10:57–59, 1997.

[7] L.C.G. ROGERS. The potential approach to the term structure of interest rates and foreign exchange rates. *Mathematical Finance.*, 7:157–176, 1997.

[8] L.C.G. ROGERS and D. WILLIAMS. *Diffusions, Markov Processes and Martingales.* Cambridge University Press, 2000.

[9] L.C.G. ROGERS and F.A. YOUSAF. Markov chains and the potential approach to modelling interest rates and exchange rates. *Mathematical Finance. Bachelier Congress.*, 2000.

[10] L.C.G. ROGERS and O. ZANE. Fitting potential models to interest rates and foreign exchange rates. *Vasicek and beyond. Risk Publications.*, pages 327–342, 1997.

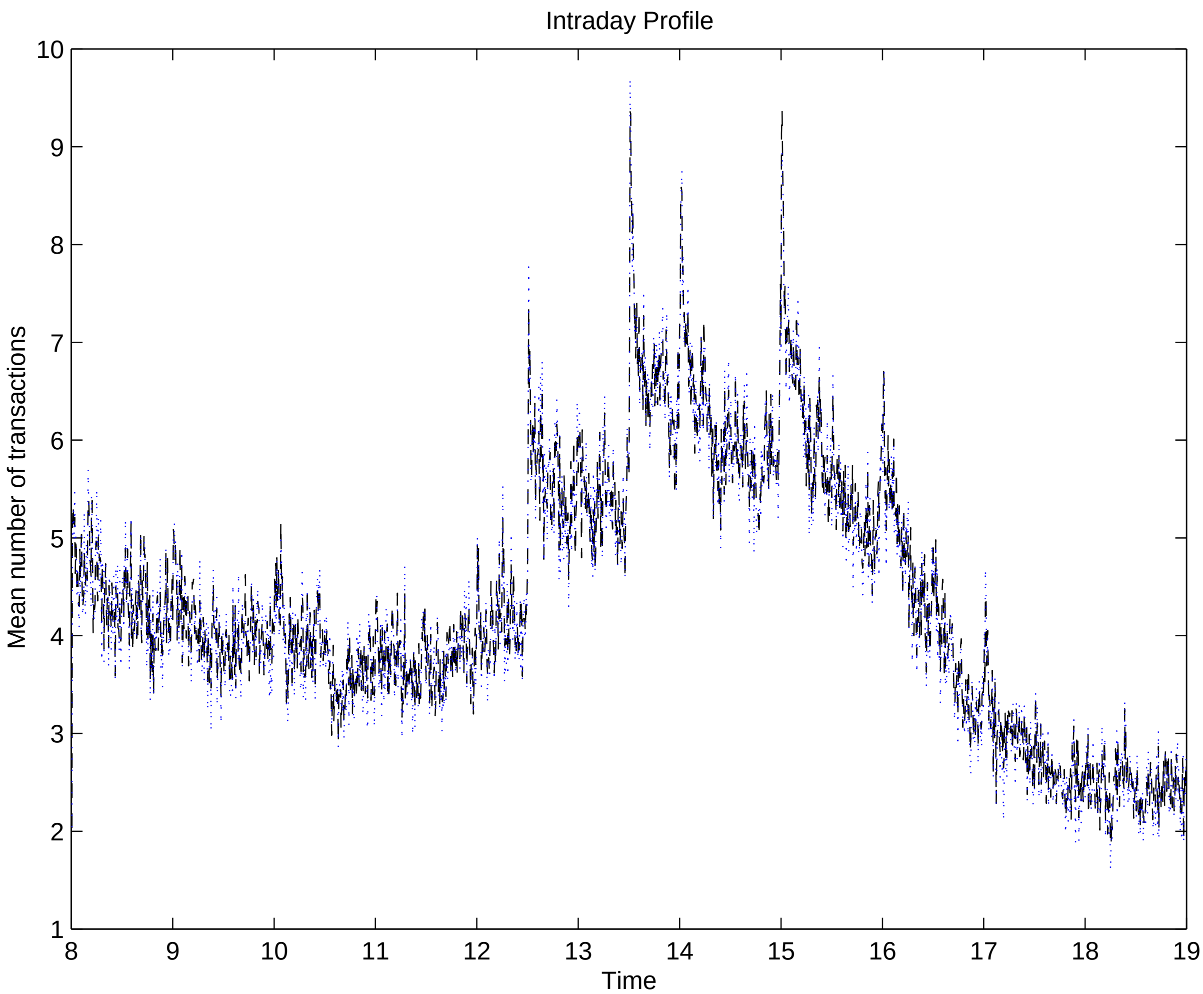

Figure 1: Mean number of ask and bid quotes on all the twenty seconds periods between 8.00 and 19.00. Data includes all the transactions over the four months between Feb/May 2004 for an FX Market.

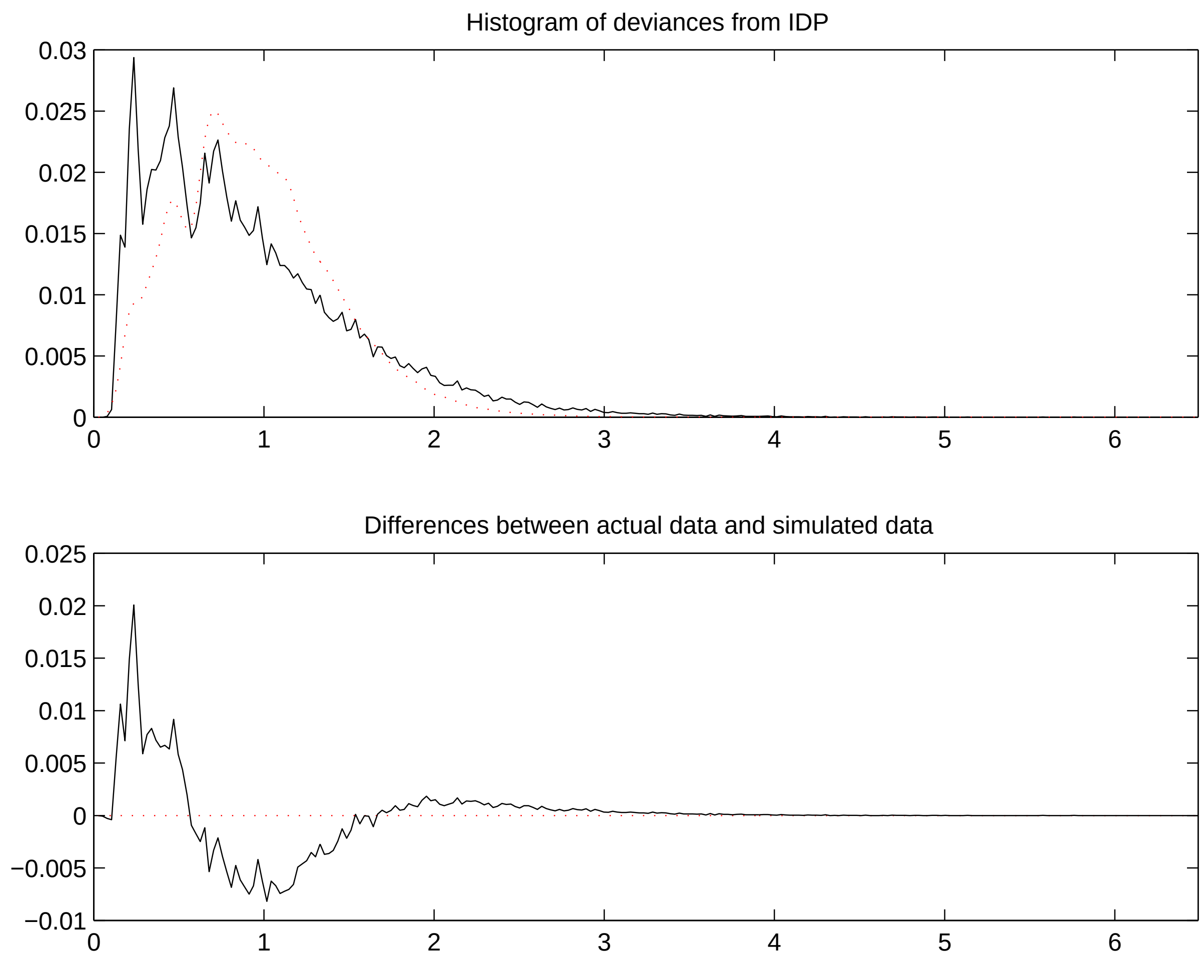

Figure 2: Histogram of the deviations of actual number of ask quotes from their intraday profile, data as in Figure 2.

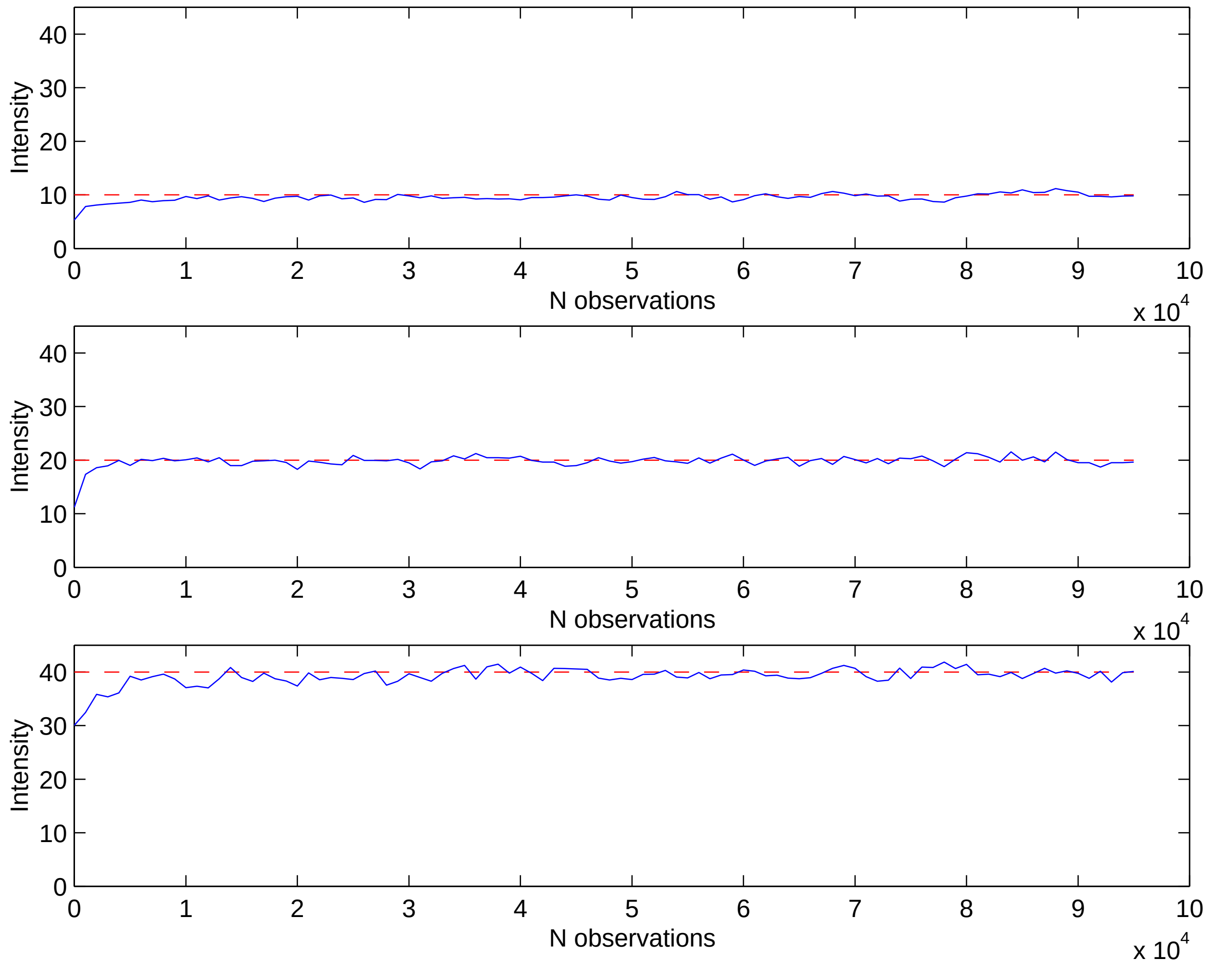

Figure 3: Convergence of intensities estimates on simulated data with $N_s = 100$ particles.

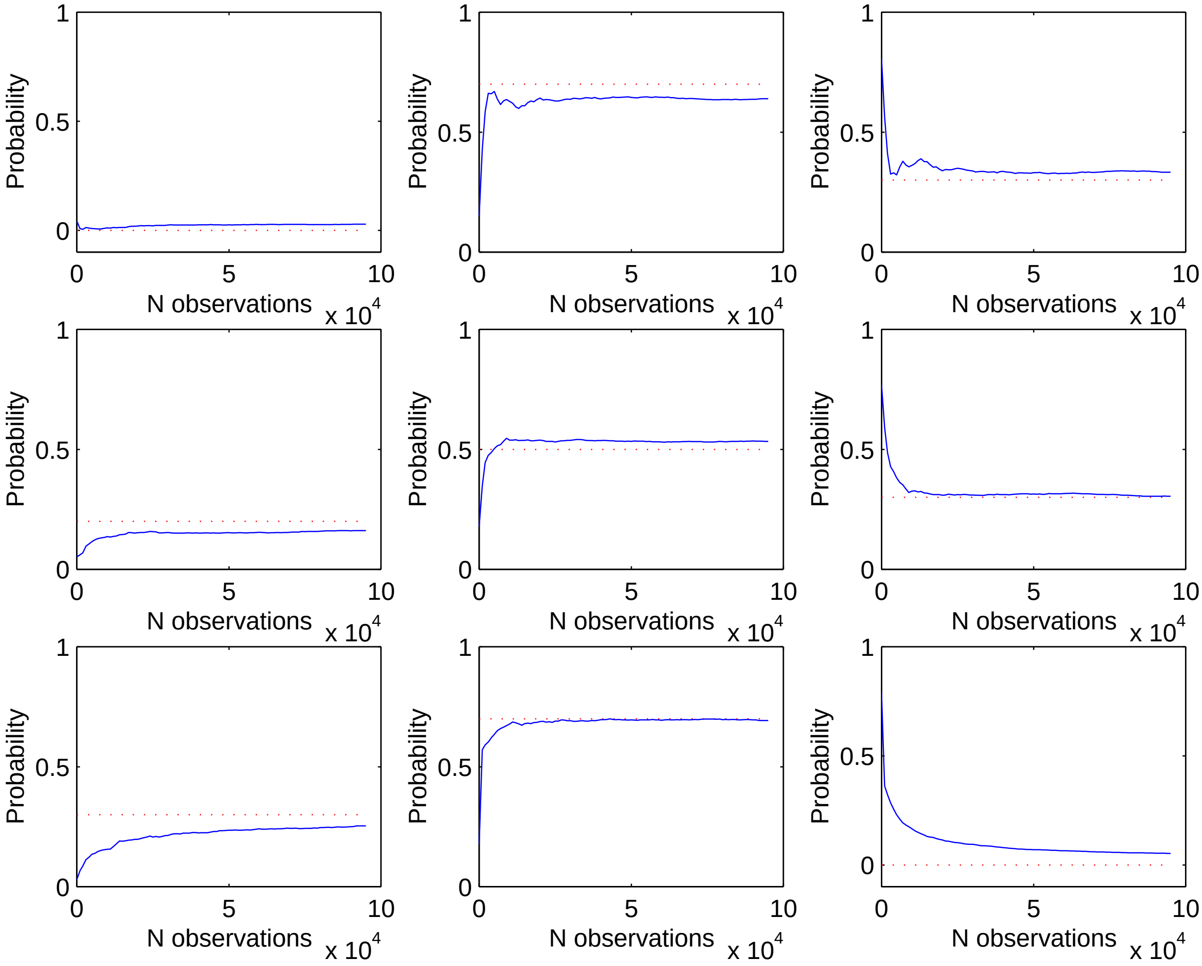

Figure 4: Convergence of transition probabilities estimates on simulated data with $N_s = 100$ particles.

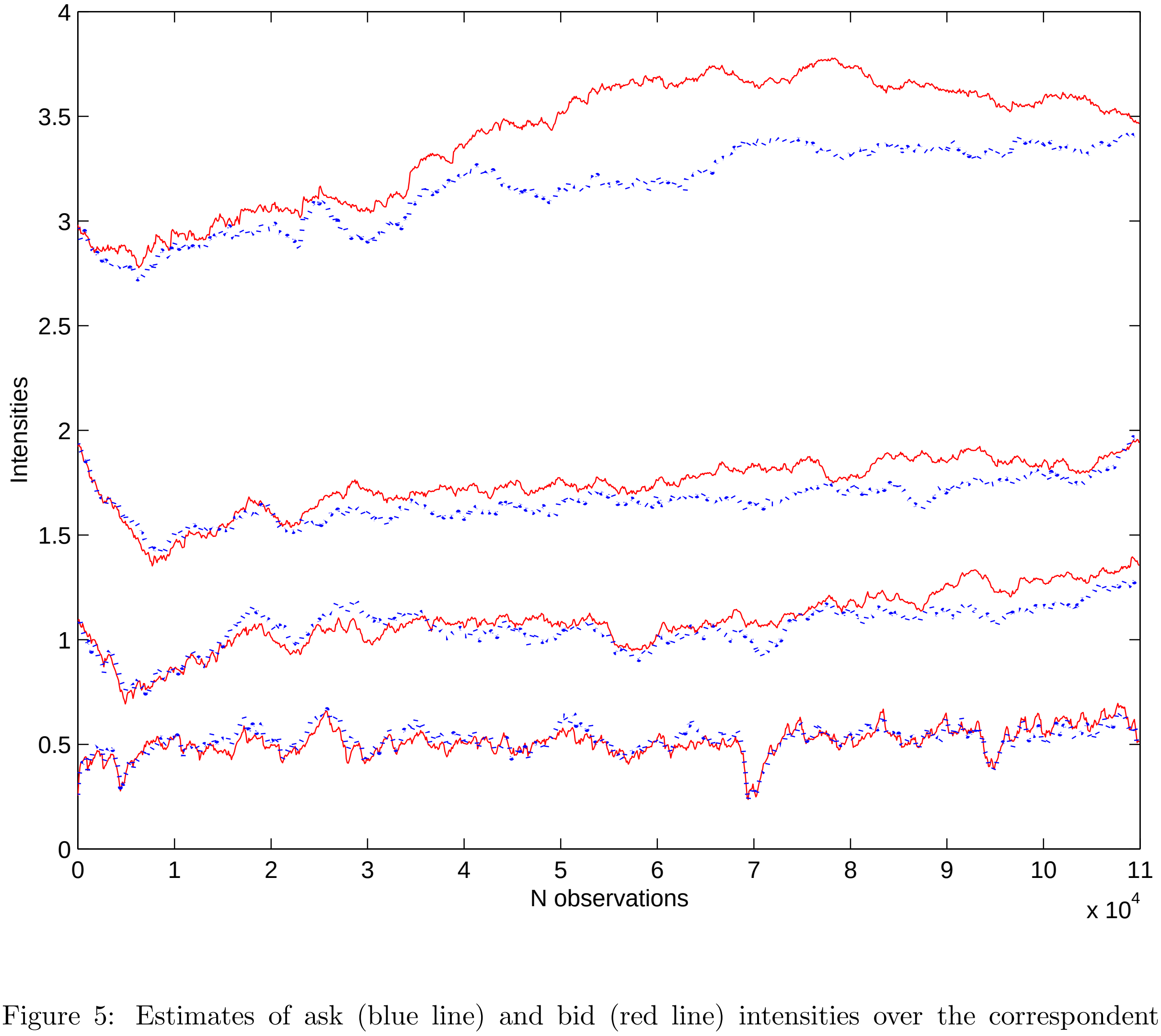

Figure 5: Estimates of ask (blue line) and bid (red line) intensities over the correspondent intraday profile for the FX US Dollar/Euro data.

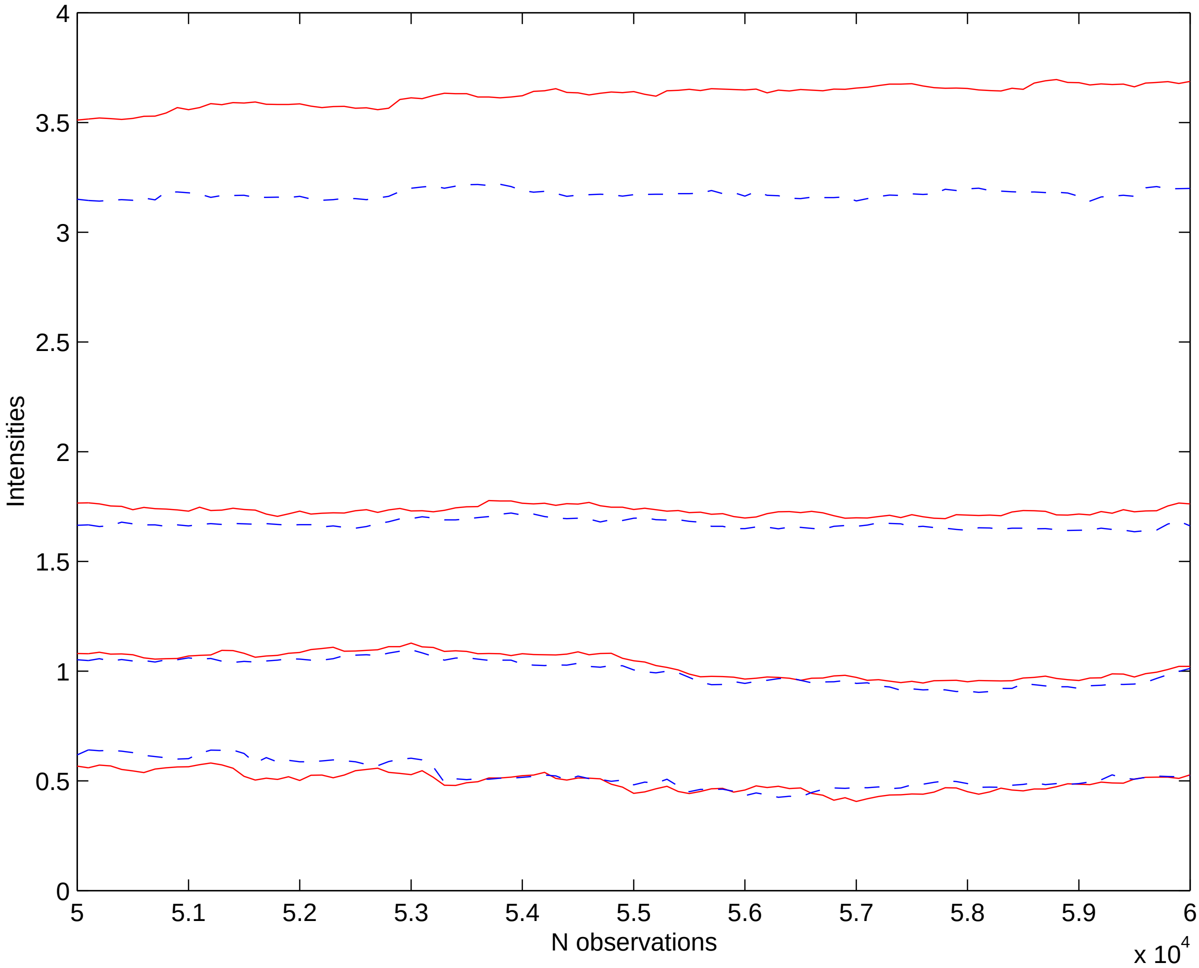

Figure 6: Estimates of ask and bid intensities from Figure 6 on a shorter time scale.

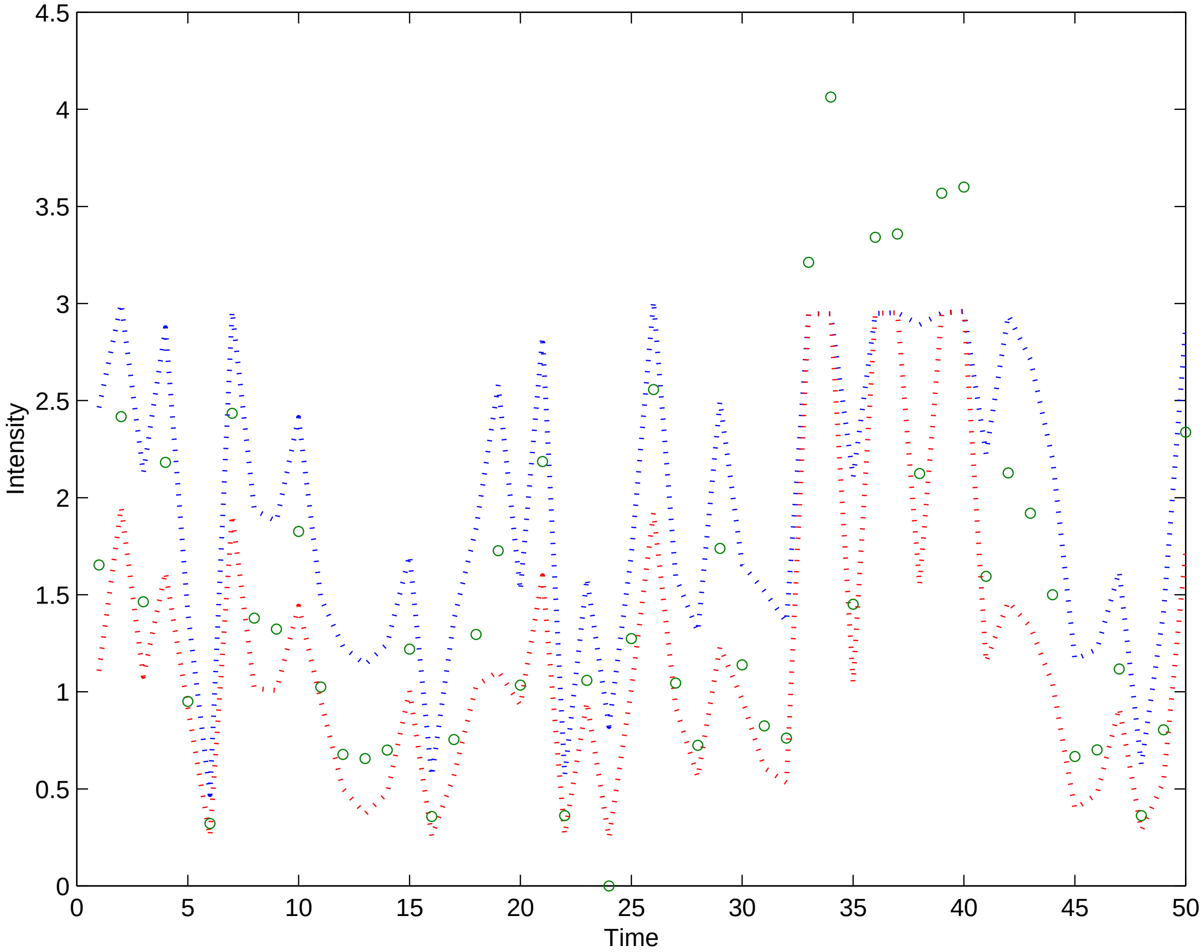

Figure 7: Confidence interval for the posterior means plus (blue dotted line) and minus (red dotted line) a multiple $m_{sde} = 1.95$ of the posterior standard deviation. The circles are the observed number of quotes over the correspondent IDP.

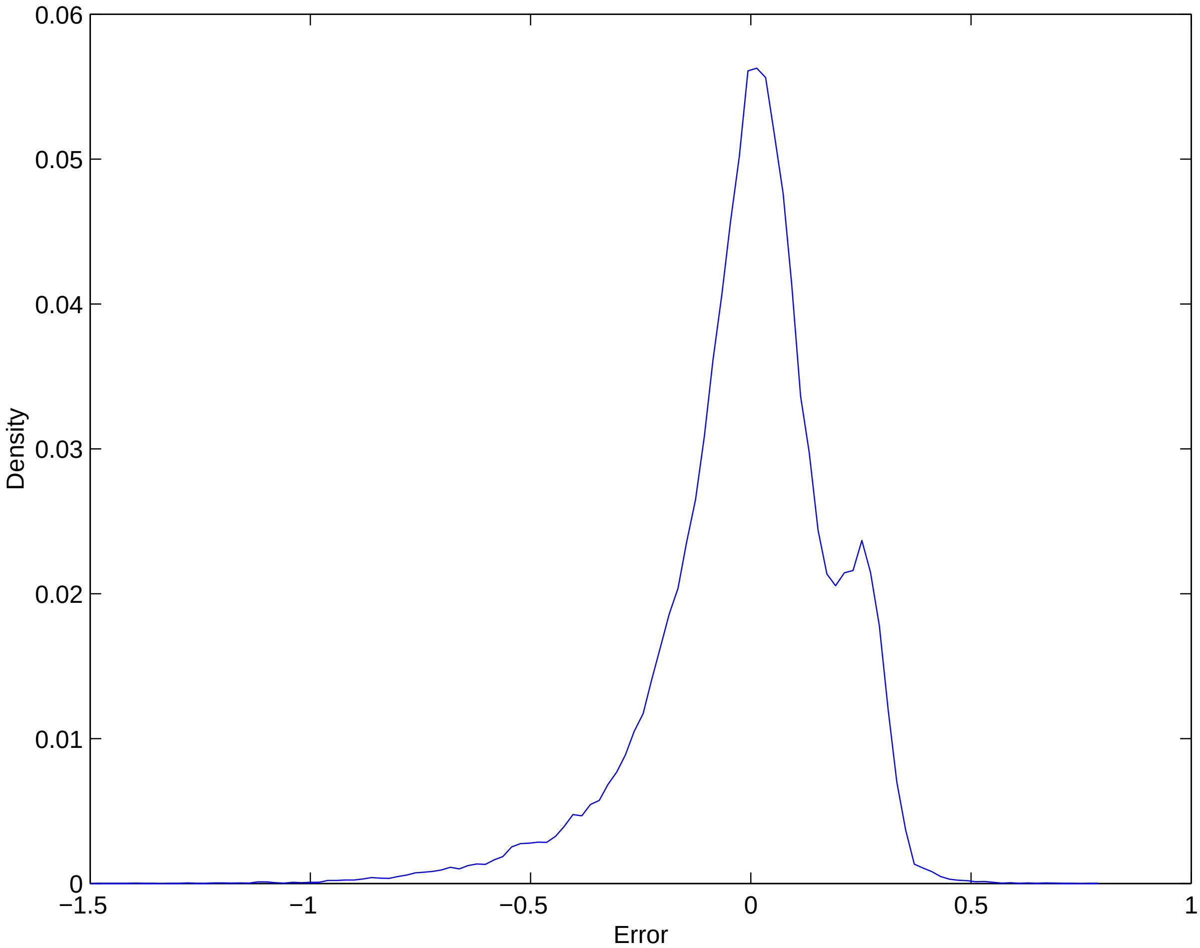

Figure 8: