A new approach for the dynamics of ultra high frequency data: the model with uncertainty zones

Christian Y. Robert and Mathieu Rosenbaum

CREST and ENSAE Paris Tech
Timbre J120, 3 Avenue Pierre Larousse, 92245 Malakoff Cedex, France.
chrobert@ensae.fr

CMAP-École Polytechnique Paris
UMR CNRS 7641, 91128 Palaiseau Cedex, France.
mathieu.rosenbaum@polytechnique.edu

2010, February 3

A NEW APPROACH FOR THE DYNAMICS OF ULTRA HIGH FREQUENCY DATA: THE MODEL WITH UNCERTAINTY ZONES

Christian Y. Robert and Mathieu Rosenbaum

In this paper, we provide a model which accommodates the assumption of a continuous efficient price with the inherent properties of ultra high frequency transaction data (price discreteness, irregular temporal spacing, diurnal patterns...). Our approach consists in designing a stochastic mechanism for deriving the transaction prices from the latent efficient price. The main idea behind the model is that, if a transaction occurs at some value on the tick grid and leads to a price change, then the efficient price has been close enough to this value shortly before the transaction. We call uncertainty zones the bands around the mid-tick grid where the efficient price is too far from the tick grid to trigger a price change. In our setting, the width of these uncertainty zones quantifies the aversion to price changes of the market participants. Furthermore, this model enables to derive approximated values of the efficient price at some random times, which is particularly useful for building statistical procedures. Convincing results are obtained through a simulation study and the use of the model over ten representative stocks.

1. INTRODUCTION

Nowadays, a large amount of ultra high frequency financial data is available. Indeed, practitioners are able to accurately record the most relevant market quantities such has transaction prices, bid-ask quotes, bid-ask volumes and all associated timestamps. It is well-known that these data are characterized by irregular temporal spacing, discreteness and diurnal patterns such as the U-shaped volatility over the day, see for example the seminal paper of Engle (2000).

A large body of studies has been centered around the irregular spacing of data in time. Two types of dynamic models on the basis of a point-process representation have been introduced: the duration models like the ACD model by Engle and Russel (1998) or the SVD model by Ghysels *et al.* (1992), and the intensity models, see for example Bowsher (2007). Some papers address both the spacing of the data and the price changes, see Bauwens and Giot (2003), Engle (2000), Ghysels and Jasiak (1998), Russel and Engle (2005).

However, a large side of the mathematical finance theory only assumes that the prices of the assets obey continuous Ito semi-martingales. Although it is reasonable to admit that low frequency financial data behave like observations of a continuous Ito semi-martingale, it is clear that high frequency data do not. Starting from this empirical observation, several recent papers have considered models where market prices are viewed as noisy observations of a semi-martingale efficient price, see among others Andersen *et al.* (2006), Bandi and Russel (2008), Barndorff-Nielsen *et al.* (2009), Ghysels and Sinko (2007), Jacod *et al.* (2009), Kalnina and Linton (2008), Large (2007), Li and Mykland (2007), Rosenbaum (2009), Zhang *et al.* (2005). Nevertheless, these papers aim at estimating or forecasting the integrated volatility from noisy data and do not really focus on reproducing the properties of ultra high frequency data.

We propose a new approach for modeling transaction prices dynamics that accommodates the assumption of a continuous efficient price process with the inherent properties of ultra high frequency data. This approach consists in designing a stochastic mechanism for deriving the transaction prices from the latent efficient price. This kind of idea has been first proposed in Ball (1988), Cho and Frees (1988) and Gottlieb and Kalay (1985).

In an idealistic framework, where the efficient price would be observed, market participants would trade when this price crosses the tick grid. In practice, there is some uncertainty about the efficient price value and market participants are reluctant to price changes. Hence, we consider there is a modification of the transaction price only if some buyers and sellers are truly convinced that the efficient price is sufficiently far from the last traded price. Thus, we assume that, if a transaction occurs at some value on the tick grid and leads to a price change, then the efficient price has been close enough to this value shortly before the transaction. This is formalized in this paper by the model with uncertainty zones, in which a specific parameter quantifies the aversion to price changes (with respect to the tick size) of the market participants.

We describe and discuss our model with uncertainty zones in Section 2. The associated properties of the last traded price, durations and microstructure noise are investigated by Monte Carlo simulations in Section 3. We study some French equity data through the lenses of our model in Section 4.

2. MODEL WITH UNCERTAINTY ZONES

2.1. *Description of the model*

We build in this section our model for the last traded price. Let $(X_t)_{t\geq 0}$ denote the efficient price of the asset. On a rich enough filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, we assume that the logarithm of the efficient price $(Y_t)_{t\geq 0}$ is a $\mathcal{F}_t$-adapted continuous Ito semi-martingale of the form

$$Y_t = \log X_t = \log X_0 + \int_0^t a_u \mathrm{d}u + \int_0^t \sigma_{u-} \mathrm{d}W_u,$$

where $(W_t)_{t\geq 0}$ is a standard $\mathcal{F}$-Brownian motion.

The tick grid, where transaction prices are bound to lie on, is defined as $\{k\alpha; k \in \mathbb{N}\}$, with α the tick size. For $k \in \mathbb{N}$ and $0 < \eta < 1$, we define the zone U_k by $U_k = [0,\infty) \times (d_k, u_k)$ with

$$d_k = (k + 1/2 - \eta)\alpha \text{ and } u_k = (k + 1/2 + \eta)\alpha.$$

Thus, U_k is a band around the mid-tick grid value $(k + 1/2)\alpha$, see Figure 1. Note that when η is smaller than $1/2$, there is no overlap between the zones.

We now explain how transaction prices change in our model. Assume that the current transaction price is $k'\alpha$. If the efficient price was known, a naïve idea would be to say that, as soon as the efficient price becomes closer from $(k'+1)\alpha$ or $(k'-1)\alpha$ than from $k'\alpha$, the transaction price should move to $(k'+1)\alpha$ or $(k'-1)\alpha$. Indeed, at that moment, $k'\alpha$ is not anymore the best approximation of the efficient price by a tick grid value. This approach is unrealistic for the following reasons:

- There is some uncertainty about the efficient price and market participants are reluctant to price changes. Thus, to make a transaction at another price, they have to be really convinced that the efficient price is now sufficiently far from the last traded price. Thus, it is not enough to have the efficient price closer to a new value of the tick grid than the current transaction value. It has to be "significantly" closer. Consequently, we assume that the transaction price may jump from price $k'\alpha$ to price $k\alpha$ with $k' \neq k$ only once the efficient price exited down the zone U_k or exited up the zone U_{k-1}.

- Even if market participants believe the efficient price is now really different from the last traded price, they are, of course, not obliged to trade. We will introduce a sequence $(L_i)_{i\geq 0}$ of discrete random variables, each L_i representing the absolute value in number of ticks of the price jump between the i-th and the $(i+1)$-th transaction leading to a price change. As explained in Section 4, the distribution of the variable L_i will depend on the value of some market quantities at the time of the i-th transaction. Indeed, one may for example think that jumps of several ticks are more likely to occur if the volatility is high or under particular configurations of the order book[1].

We now build the sequence $(\tau_i)_{i\geq 0}$ of the exit times from the uncertainty zones which will lead to a change in the transaction price. Let $\tau_0 = 0$ and assume without loss of generality that τ_1 is the exit time of $(X_t)_{t\geq 0}$ from the set (d_{k_0-1}, u_{k_0}) where $k_0 = X_0^{(\alpha)}/\alpha$, with $X_0^{(\alpha)}$ the value of X_0 rounded to the nearest multiple of α. Moreover, suppose the $(L_i)_{i\geq 1}$ are $\mathcal{F}_{\tau_i}$-measurable[2] and define recursively τ_{i+1} as the exit time of $(X_t)_{t>\tau_i}$ from the set $(d_{k_i - L_i}, u_{k_i + L_i - 1})$, where $k_i = X_{\tau_i}^{(\alpha)}/\alpha$, *i.e.*

$$\tau_{i+1} = \inf\Big\{t : t > \tau_i, X_t = X_{\tau_i}^{(\alpha)} - \alpha(L_i - \frac{1}{2} + \eta) \text{ or } X_t = X_{\tau_i}^{(\alpha)} + \alpha(L_i - \frac{1}{2} + \eta)\Big\}.$$

[1]Furthermore, even if they are quite rare on liquid stocks, some market orders have such a volume that they imply jumps of more than one tick, see Section 4.

[2]$\mathcal{F}$ is a larger filtration than those of the market participants, see Section 4.

In particular, if $X_{\tau_i} = d_j$ for some $j \in \mathbb{N}$, τ_{i+1} is the exit time of $(X_t)_{t>\tau_i}$ from the set (d_{j-L_i}, u_{j+L_i-1}), and if $X_{\tau_i} = u_j$ for some $j \in \mathbb{N}$, τ_{i+1} is the exit time of $(X_t)_{t>\tau_i}$ from the set (d_{j-L_i+1}, u_{j+L_i}), see Figure 1.

The last traded price process is characterized by the couples of transaction times and transaction prices with price changes $(t_i, P_{t_i})_{i\geq 0}$. Let $t_0 = 0$ and $P_0 = X_0^{(\alpha)}$. It would be probably unrealistic to assume that the τ_i precisely give us the times of the transactions with price change (that is $t_i = \tau_i$). Indeed, even if one wants to trade because one is convinced that the efficient price is now far from the last traded price, its reaction time and delays due to the trading process have still to be taken into account. Moreover, this assumption is not necessary in the main applications of our model such as retrieving the efficient price or estimating the volatility, see Section 2.3. We only assume that between τ_i and τ_{i+1}, at least one transaction has occurred at price P_{t_i} and t_i is the time of the first of these transactions. So, for $i \geq 1$, we assume that $\tau_i \leq t_i < \tau_{i+1}$ and $P_{t_i} = X_{\tau_i}^{(\alpha)}$.

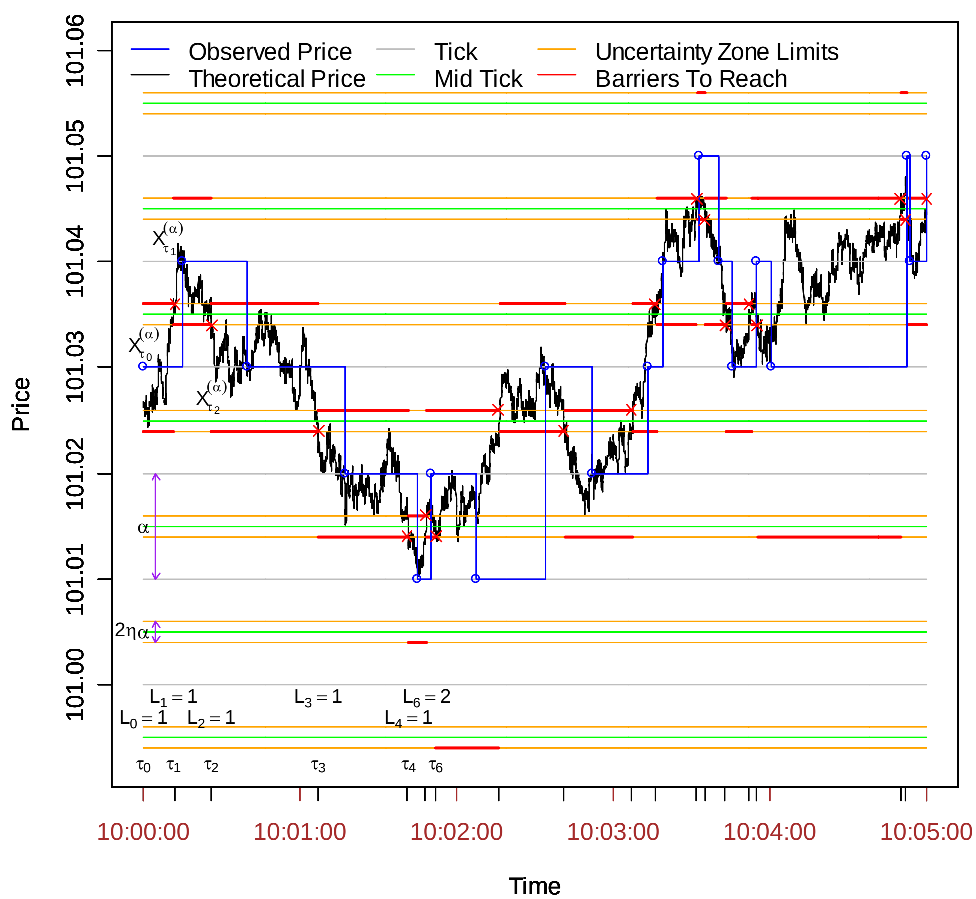

FIGURE 1.— *Example of trajectories of the latent price and of the observed price. The red crosses denote the exit points associated to the τ_i.*

2.2. *Aversion to price changes and optimal tick size*

The idea behind this model is that, in some sense, market participants feel more comfortable when the asset price is constant than when it is moving. However, there are times when the transaction price changes because they estimate that the last traded price value is not reasonable anymore. In our model, the parameter η quantifies this aversion to price changes (with respect to the tick size). It controls the width of the uncertainty zones and therefore, in tick unit, the larger η, the farther from the last traded price the efficient price has to be so that a price change occurs.

One can also see the parameter η as a measure of the relevance of the tick size on the market. Indeed, if $\eta < 1/2$, market participants are convinced they have to trade at a new price before the efficient price crosses this new price on the tick grid. So, it means that the tick size appears too large to them. Conversely, a large η ($\eta > 1/2$) means that the tick size appears too small. From the tick size perspective, an ideal market is consequently a market where η is equal to $1/2$.

There are several other ways to interpret the parameter η. One can for example consider the efficient price as a kind of random walk in a random environment given by the order book. This environment is more or less reluctant when the efficient price is going through it. This reluctancy could be characterized by η. Another possibility is to view η as a measure of the usual prices depth explored by the transaction volumes.

A natural estimation procedure for the parameter η is given in Robert and Rosenbaum (2009a). We define an alternation (resp. continuation) of one tick as a price jump of one tick whose direction is opposite to (resp. the same as) the one of the preceding price jump. Let $N_{\alpha,t}^{(a)}$ and $N_{\alpha,t}^{(c)}$ be respectively the number of alternations and continuations of one tick over the period $[0,t]$. An estimator of η over $[0,t]$ is given by

$$\hat{\eta}_{\alpha,t} = \frac{N_{\alpha,t}^{(c)}}{2N_{\alpha,t}^{(a)}}.$$

In the asymptotics where the tick size goes to zero, a central limit theorem with normalizing speed α^{-1} is proved in Robert and Rosenbaum (2009a). A slightly more complicated estimator, using all the price jumps, is given together with its asymptotic theory in the same paper.

2.3. *Retrieving the efficient price and statistical procedures*

Our model enables to retrieve the values of the efficient price at time τ_i by the simple relation

$$X_{\tau_i} = P_{t_i} - sign(P_{t_i} - P_{t_{i-1}})(1/2 - \eta)\alpha.$$

Hence, since we can estimate η, we can approximately recover X_{τ_i} from $P_{t_{i-1}}$ and P_{t_i}. As shown by the following examples, this is very convenient for building statistical procedures relative to the efficient price.

Example 1 : Estimation of the integrated volatility

An estimator of the integrated volatility of the efficient price over $[0,t]$, $\int_0^t \sigma_s^2 \mathrm{d}s$, is simply given by a realized volatility measure computed over the estimated values of the efficient price

$$\sum_{\tau_i \leq t} \Big(\frac{\hat{X}_{\tau_i} - \hat{X}_{\tau_{i-1}}}{\hat{X}_{\tau_{i-1}}}\Big)^2,$$

where $\hat{X}_{\tau_i} = P_{t_i} - sign(P_{t_i} - P_{t_{i-1}})(1/2 - \hat{\eta}_{\alpha,t})\alpha$. The accuracy of this estimator is α and its asymptotic theory is available in Robert and Rosenbaum (2009a). It is also shown in the same paper that in the case where two assets are observed, the same kind of ideas can be used in order to build a consistent estimator of the integrated co-volatility.

Example 2 : Test procedure

The preceding relation can also be applied to test the model. Assume for simplicity that the drift process is null and that the variables L_i are independent of the efficient price. Then, given that $L_i = 1$, for any α, the random variables $|X_{\tau_{i+1}} - X_{\tau_i}|$ are independent and identically distributed, see Robert and Rosenbaum (2009a) for details. Consequently, a natural idea for assessing the model is to draw the autocorrelogram of the $(|\hat{X}_{\tau_{i+1}} - \hat{X}_{\tau_i}|)_{i \in I_1}$, where $I_1 = \{i : \ i \geq 0, \ L_i = 1\}$ and to compute the associated Ljung-Box statistics. Indeed, this autocorrelogram is an approximation of the one of the $(|X_{\tau_{i+1}} - X_{\tau_i}|)_{i \in I_1}$.

Note that the model is also very useful if one is interested in mathematical finance issues in the high frequencies. For example, this model can be used to compute the hedging error when hedging derivatives in the high frequencies, see Robert and Rosenbaum (2009b).

2.4. *Discussion on our approach*

The first novelty of our model is that it reproduces all the main stylized facts of both transaction data and durations, together with an efficient latent price which is a continuous time semi martingale. To our knowledge, this is the first model filling this gap. This is what we show in this paper through some Monte Carlo experiments (see Section 3) and a large empirical study (see Section 4). Also, this model enables to solve several statistical and financial problems in the high frequencies, see Robert and Rosenbaum (2009a,b).

This model is not an economic microstructure model. It explains neither the behavior of market participants nor the formation of the price. Hence this paper is not about how microstructure noise is created from the interactions between the agents and the features of the market, as for example in Diebold and Strasser (2008).

An important feature of our approach is that a continuous time semi martingale is used to derive a continuous time model for the price. This is the cornerstone of our work and is crucial for us, in order to be able to solve mathematical finance issues. Using a continuous time semi martingale to build a price model is of course arguable and perhaps not really satisfying from an economic point of view. However, such an approach is powerful when one just aims at describing (and not explaining) the prices.

Since our model is a continuous time model, the dynamics can be investigated at any time scale, in tick time or traded time. It is a quite new approach for describing ultra high frequency data, very different from those of models aiming at deriving the dynamics of prices at a given subsampling frequency, see for example Hasbrouck (2000, 2006).

3. SIMULATION STUDY

We now explore the properties of our model through the simulation of a sample path of the efficient price process. We consider the following model

$$\mathrm{d}X_t = \sigma_t X_t \mathrm{d}W_t,\ x_0 = 100,\ t \in [0,1],$$

where $(\sigma_t)_{t\geq 0}$ is a deterministic function of time given in Figure 2, $\alpha = 0.05$, $\eta = 0.05$, and, for $i \geq 1$, $L_i = 1$ and $t_i = \tau_i$. Our simulation accuracy is 0.1 second[3].

3.1. *Properties of the last traded price and its durations*

Figure 2 represents the volatility function used in our simulation together with the durations between price changes.

[3]More precisely, the interval [0,1] corresponds to one trading day of eight hours and the discretization mesh is $(3600 \times 8 \times 10)^{-1}$ on $[0,1]$.

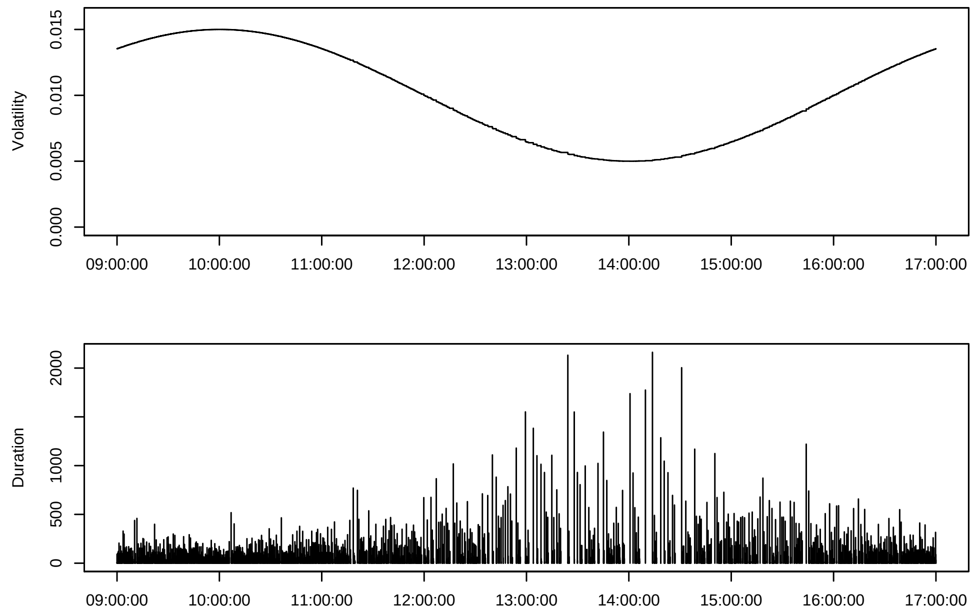

FIGURE 2.— *Volatility trajectory and durations.*

The chosen U-shaped form for the volatility is classical and the durations have a behavior which is in inverse relation to those of the volatility. This reproduces a usual empirical characteristic of high frequency financial data, see for example Engle (2000), Gourieroux and Jasiak (2001). See also Renault and Werker (2009) for the relationship between durations and volatility. This is due to the fact that the exit times of the uncertainty zones are longer when the volatility is low.

On Figure 3, we show the sample paths of the last traded and efficient prices over half an hour.

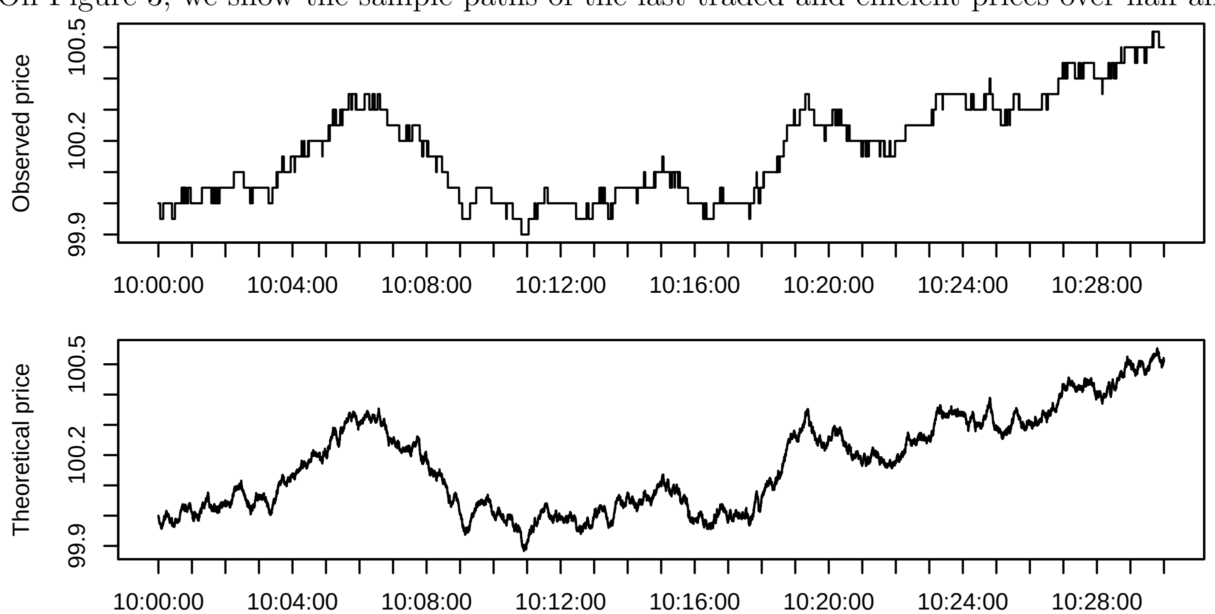

FIGURE 3.— *Sample paths of the last traded and efficient prices*

The commonly observed numerous and quick oscillations of one tick of the last traded price (bid-ask bounce) are reproduced thanks to the behavior of the semi-martingale efficient price around the uncertainty zones ($\eta < 1/2$).

We finally draw some autocorrelation functions for the logarithmic returns on Figure 4. Note that the aim of such graphs is just to show that our model reproduces the stylized facts of real data, see Section 4. However, one has to be cautious with their interpretation because of stationary issues in trading times.

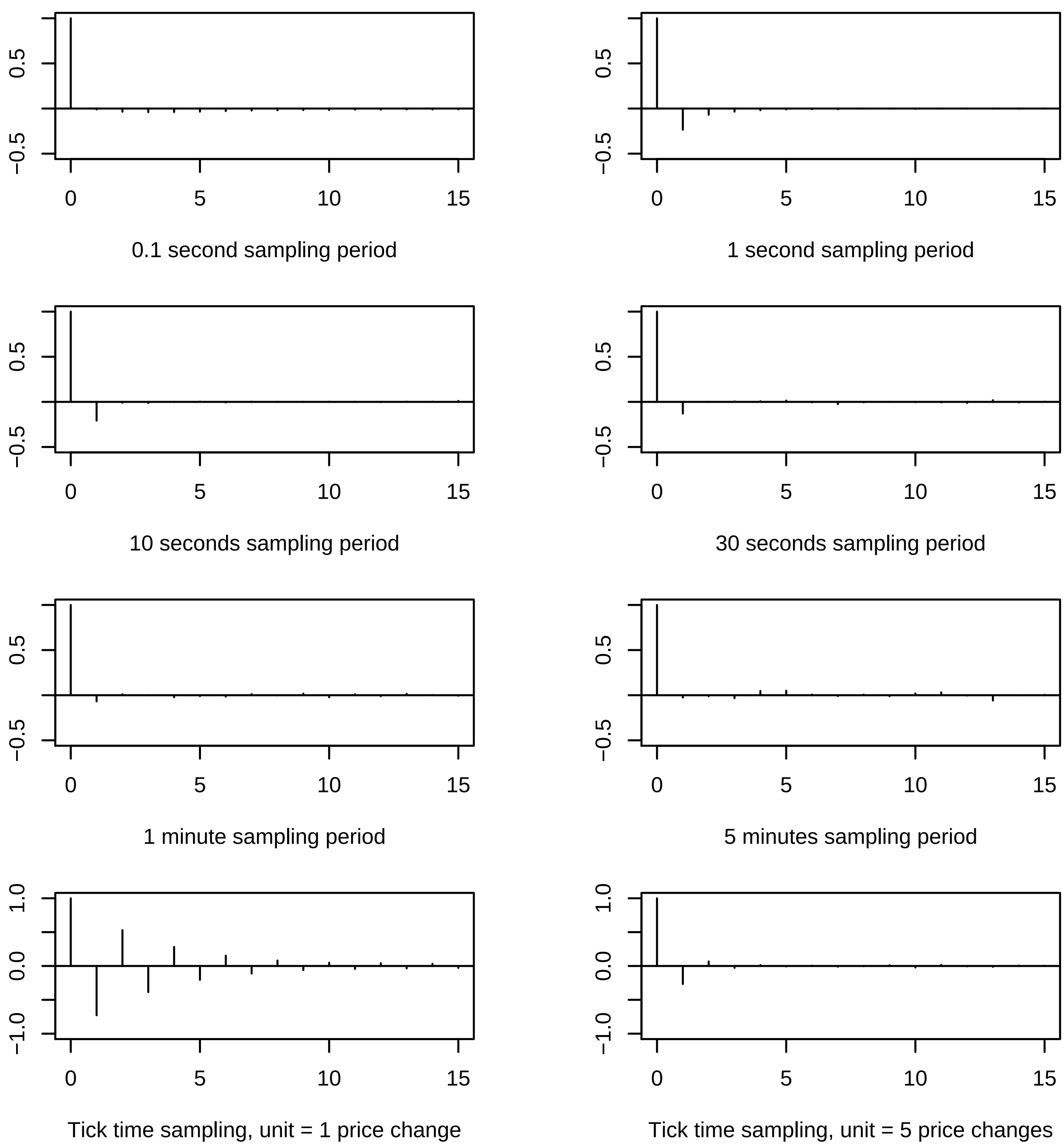

FIGURE 4.— *Autocorrelation functions of the log returns.*

Thanks to the uncertainty zones, we observe in our model the stylized fact of a significative negative first order correlation between the returns for sampling frequencies between one and thirty seconds. Moreover, in tick time, many of the higher order autocorrelations are significant and systematically alternate sign. This phenomenon was empirically brought to light by Aït-Sahalia *et al.* (2005) and Griffin and Oomen (2008).

3.2. *Properties of the microstructure noise*

In the financial econometrics literature, the difference between the logarithms of the last traded price and the efficient price is referred to as the microstructure noise. A first empirical study of the properties of the microstructure noise is presented in Hansen and Lunde (2006). In our approach, contrary to the usual statistical ways of modeling microstructure noise, the noise is derived from the efficient price. Thus, we are able to underline some reasonable features of the microstructure noise (see also Diebold and Strasser (2008) for an economic analysis). For illustration, we give different graphs linked to the noise.

The following graph shows the sample path of the microstructure noise process over half an hour and the autocorrelation function of this process.

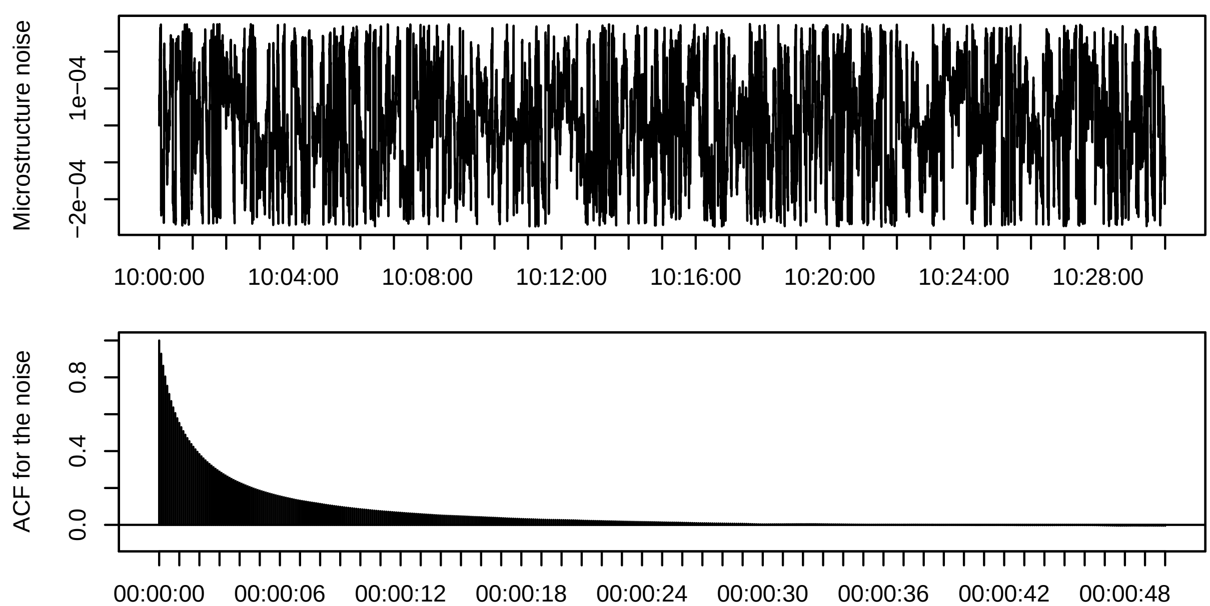

FIGURE 5.— *Sample path and autocorrelation function of the noise.*

At a first glance, the noise seems quite stationary and strongly positively correlated in the ultra high frequencies (below ten seconds). This agrees in particular with the results of Rosenbaum (2007).

We now treat the increments of the noise. These increments appear naturally in the study of the estimation of the integrated volatility under the assumption of an additive microstructure noise, see for example Zhang *et al.* (2005). We first draw on Figure 6 some autocorrelation functions.

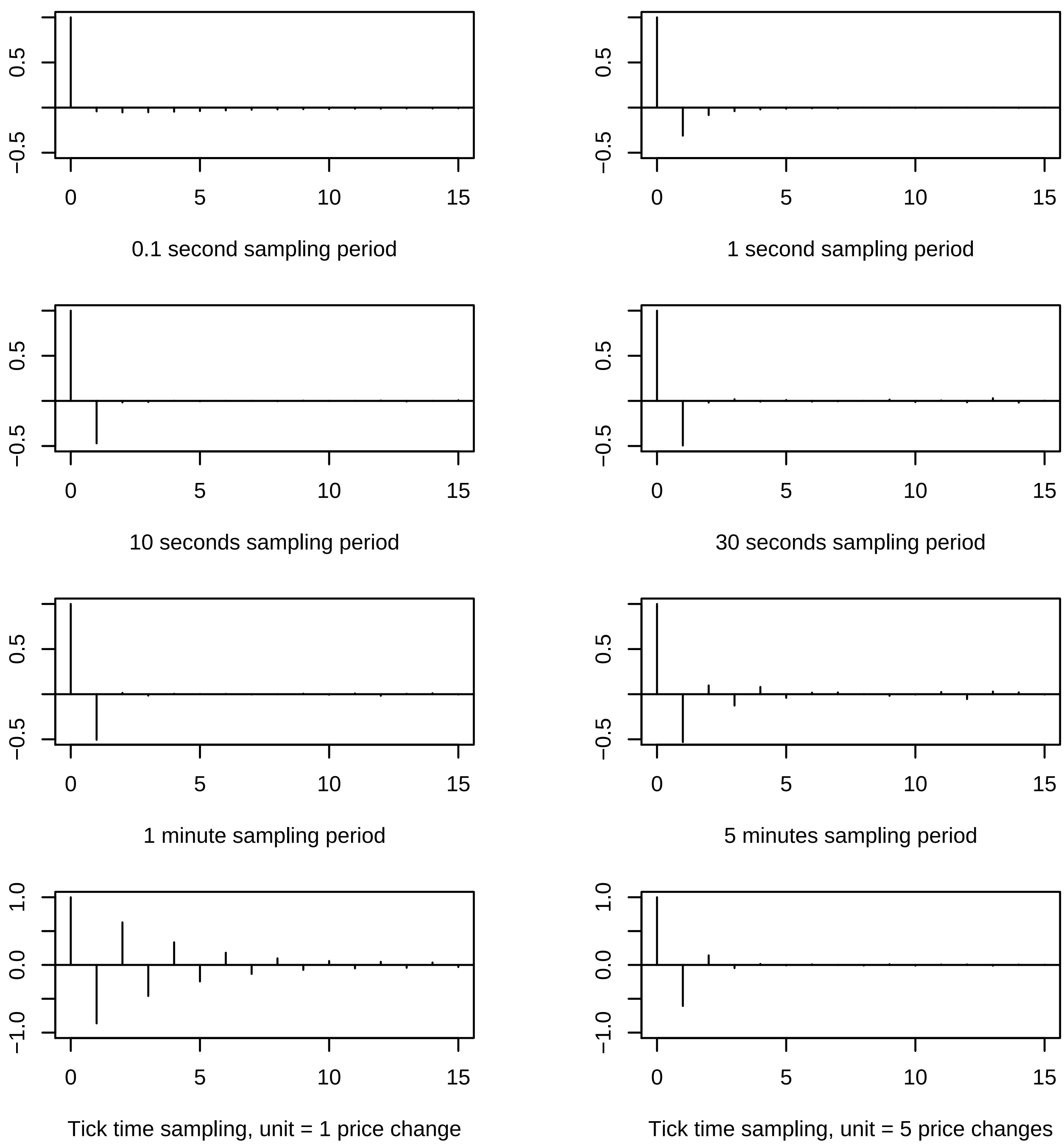

FIGURE 6.— *Autocorrelation functions of the increments of the noise.*

The autocorrelation functions of the increments of the noise in calendar time suggest a $MA(1)$ structure for a sampling frequency smaller than ten seconds. This is no longer true in tick time, where an autoregressive model is probably more convenient. Remark that, while the microstructure noise seems not so far from a white noise in calendar time, it is highly dependent in tick time. Such kind of results is also found by Griffin and Oomen (2008). We end with some cross-correlation functions on Figure 7.

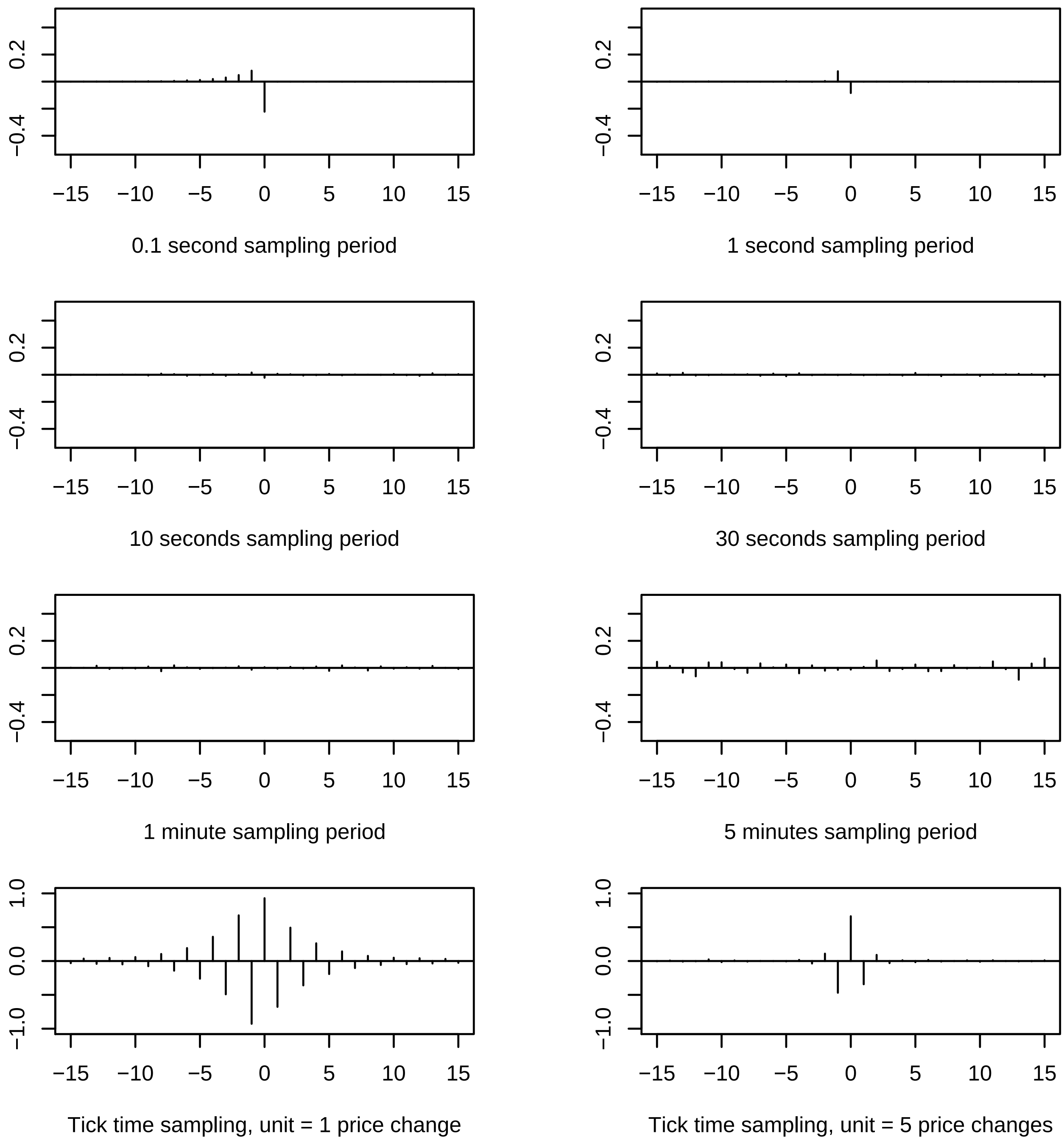

FIGURE 7.— *Cross-correlation functions between the increments of the noise and the log returns.*

Figure 7 shows that the hypothesis of zero-correlation between the increments of the noise and the increments of the log returns in calendar time seems reasonable for a sampling frequency smaller than ten seconds. In tick time, the cross-correlations structure is much more complex and shows an intricate dependence between both components of the last traded price.

Eventually, we have shown that the model with uncertainty zones enables to accommodate the inherent properties of prices, durations and microstructure noise, together with a semi-martingale efficient price. In particular, this model allows for discrete prices, a bid-ask bounce, an inverse relation between durations and volatility and jumps in the price of several ticks, the size of the jumps determined by explanatory variables, involving for example the order book (see Section 4). Finally, relevant behaviors for the autocorrelograms and cross-correlograms of the returns and microstructure noise are obtained, both in calendar and tick time.

4. REAL DATA ANALYSIS

We now study some French Equity data through the lenses of our model. The sample period is from 2007, January 15 to 2007, January 19. We consider all trades from 8:30 AM to 4:25 PM (GMT). We pick the ten following stocks on Euronext: Air France (AIRF), Alsthom (ALSO), BNP-Paribas (BNPP), Crédit Agricole (CAGR), Danone (DANO), EADS (EAD), France Telecom (FTE), Renault (RENA), Saint-Gobain (SGOB), Total-Fina (TOTF).

4.1. *Summary statistics*

Table I reports some summary statistics for prices.

TICKER	Tick size	Min price	Max price	Mean number of trades per day	Average spread	Max spread
AIRF	0.01	34.03	35.10	2604	0.0156	0.21
ALSO	0.05	93.80	98.00	5002	0.0658	0.50
BNPP	0.05	83.00	85.95	5431	0.0528	0.15
CAGR	0.01	32.07	33.20	6016	0.0137	0.10
DANO	0.10	112.50	116.80	2940	0.1027	0.30
EAD	0.01	23.80	25.94	5800	0.0156	0.12
FTE	0.01	21.55	22.34	6187	0.0122	0.06
RENA	0.05	90.15	94.5	3231	0.0588	0.25
SGOB	0.05	68.05	71.55	4847	0.0558	0.20
TOTF	0.05	50.80	53.15	7771	0.0506	0.10

TABLE I

Summary statistics for prices (euros) and spreads.

The ten considered stocks are representative in term of tick sizes. Remark that the average spreads are quite close to the tick sizes and that the number of trades is larger than 2500 per day. This means that we are dealing with fairly liquid assets.

We give in Table II and Table III summary statistics for the transactions and durations.

TICKER	Mean volume per trade	Median volume per trade	Stand. Dev. Volume per trade	Transactions at the Bid price (%)
AIRF	344	180	866	0.45
ALSO	215	98	497	0.47
BNPP	467	220	763	0.45
CAGR	867	370	1887	0.49
DANO	244	122	404	0.47
EAD	725	400	1346	0.49
FTE	1174	601	1972	0.44
RENA	337	185	617	0.50
SGOB	350	185	586	0.54
TOTF	1175	457	2287	0.46

TABLE II

Summary statistics for transactions.

TICKER	Average duration between moves	Median duration between moves	Max duration between moves	Stand. Dev. duration between moves
AIRF	43.2	23	651	59.5
ALSO	25.5	14	406	32.8
BNPP	27.5	15	450	36.9
CAGR	19.7	11	418	27.5
DANO	55.8	28	653	74.1
EAD	22.9	11	449	33.0
FTE	21.6	11	364	29.8
RENA	45.6	22	774	66.9
SGOB	30.7	16	1033	45.1
TOTF	22.6	13	452	28.5

TABLE III

Summary statistics for durations between transactions leading to a price change.

The mean of the volume per trade is always larger than the associated median value and smaller than the standard deviation, suggesting an heavy-tailed distribution. The same relative behavior of the mean, median and standard deviation is observed for the durations, suggesting over-dispersion relative to an exponential distribution. Nevertheless, such a conclusion is to take with care because of the diurnal effects in the durations.

We now consider the last traded price. It is built from the observation of the transaction prices with an accuracy of one second[4]. Table IV and V give information about the number and the size of the jumps in the last traded price.

TICKER	Number of jumps over the week	Average size of the jumps$\times 10^{-5}$	Min. size of the jumps	Max. size of the jumps	Stand. Dev. size of the jumps
AIRF	3290	-9.726	-0.10	0.10	0.022
ALSO	5576	-26.004	-0.45	0.70	0.075
BNPP	5164	-27.110	-0.15	0.15	0.054
CAGR	7199	-9.306	-0.09	0.07	0.016
DANO	2548	-58.869	-0.30	0.30	0.104
EAD	6208	-1.127	-0.15	0.12	0.019
FTE	6588	2.125	-0.06	0.08	0.013
RENA	3121	41.653	-0.35	0.35	0.069
SGOB	4634	-8.631	-0.20	0.20	0.059
TOTF	6298	-12.702	-0.15	0.15	0.050

TABLE IV

Summary statistics for jumps in the last traded price (1).

TICKER	Prop. jumps of size 1	Prop. jumps of size -1	Prop. jumps of size 2	Prop. jumps of size -2	Prop. jumps of size 3	Prop. jumps of size -3
AIRF	0.333	0.336	0.085	0.077	0.028	0.036
ALSO	0.385	0.391	0.083	0.078	0.019	0.021
BNPP	0.465	0.479	0.029	0.025	0.001	0.001
CAGR	0.351	0.359	0.097	0.095	0.032	0.029
DANO	0.479	0.490	0.015	0.012	0.001	0.000
EAD	0.339	0.327	0.095	0.098	0.039	0.041
FTE	0.406	0.408	0.075	0.071	0.014	0.015
RENA	0.401	0.408	0.073	0.073	0.014	0.014
SGOB	0.435	0.454	0.052	0.045	0.005	0.004
TOTF	0.495	0.496	0.004	0.004	0.000	0.000

TABLE V

Summary statistics for jumps in the last traded price (2).

[4]We take the last value if several prices occur inside a second.

Although the considered assets are fairly liquid, the number of jumps of more than one tick is far from being negligible. Note also that the distributions of the jumps are astonishingly symmetric.

Finally, Figure 8 and Figure 9 provide the autocorrelation functions of the log returns in tick time and the durations of the last traded price on 2007, January 16 (we take the second day of the sample to avoid any Monday effect).

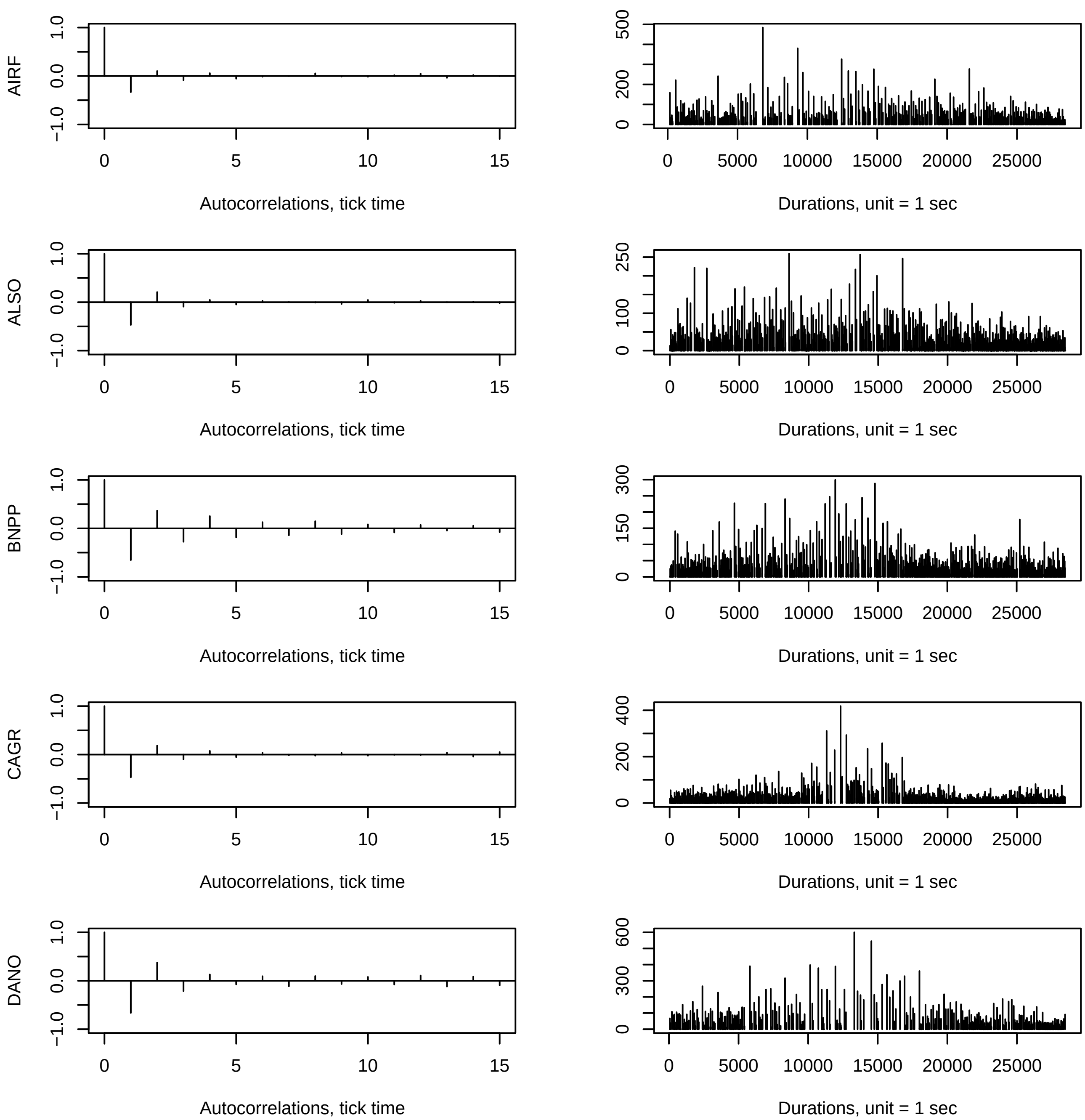

FIGURE 8.— *Autocorrelation functions of the log returns and durations - Part 1*

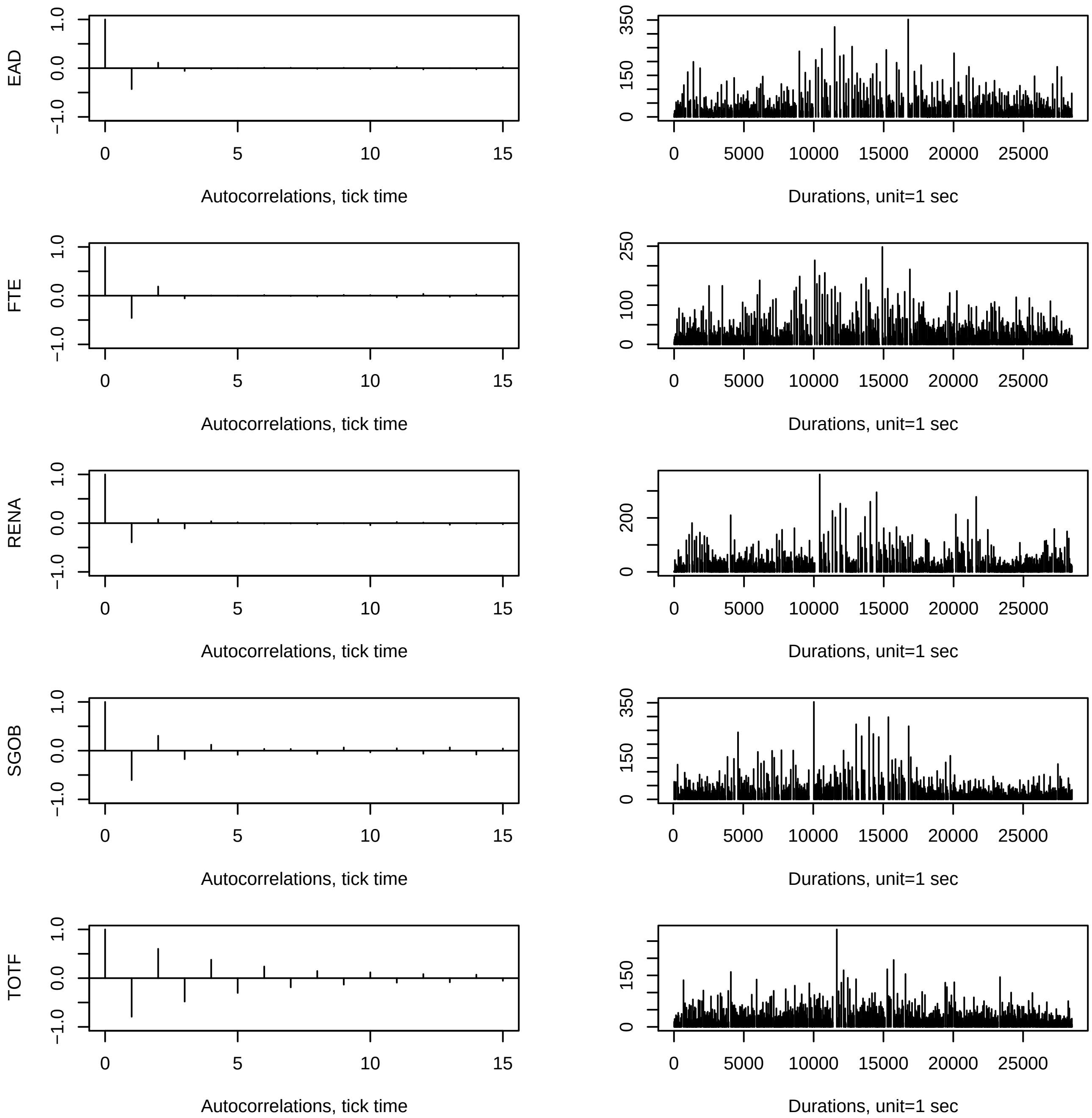

Figure 9.— *Autocorrelation functions of the log returns and durations - Part 2*

These graphs agree with the usual stylized facts of these kind of data and the phenomena they show (negative autocorrelation of the log returns, diurnal effects in the durations. . .) are clearly reproduced by the model with uncertainty zones, see Section 3.

4.2. *Estimation and test of the model*

Estimation of η

In Table VI, we give for the ten assets the estimated values of η, the parameter of aversion to price changes.

TICKER	$\hat{\eta}$ 2007-01-15	$\hat{\eta}$ 2007-01-16	$\hat{\eta}$ 2007-01-17	$\hat{\eta}$ 2007-01-18	$\hat{\eta}$ 2007-01-19
AIRF	0.277	0.273	0.250	0.236	0.250
ALSO	0.189	0.215	0.199	0.199	0.194
BNPP	0.100	0.111	0.146	0.157	0.110
CAGR	0.193	0.242	0.238	0.215	0.209
DANO	0.076	0.110	0.086	0.081	0.110
EAD	0.179	0.243	0.256	0.234	0.227
FTE	0.192	0.221	0.246	0.274	0.192
RENA	0.167	0.274	0.190	0.294	0.279
SGOB	0.088	0.136	0.131	0.129	0.118
TOTF	0.048	0.058	0.083	0.065	0.069

TABLE VI

Estimation of η.

The estimated values for η are remarkably stable, what is favorable for our model. The smaller values observed on Monday could be due to the classical “Monday” effect on the market. It is interesting to note that for a given tick size, one can both observe a large and a small value of η, see for example Renault (RENA) and Total (TOTF).

Determinants of the size of the jumps

We now present a simple study of the conditional distribution of the L_i, see Hausman *et al.* (1992) for a complementary study. We assume that the jump sizes are bounded (which is empirically not restrictive) and denote by m their maximal value. For simplicity, we also assume $t_i = \tau_i$. We consider the following multinomial LOGIT type model

$$\mathbb{P}_{\chi_{\tau_i}}[L_i = k] = p_k(\chi_{\tau_i};\theta),\ 1 \leq k \leq m,$$

where θ is an unknown parameter in $\mathbb{R}^{3(m-1)}$, χ_{τ_i} is a vector of explanatory variables known by the market participants at time τ_i and $\mathbb{P}_{\chi_{\tau_i}}$ is the conditional probability given these explanatory variables. We set

$$p_1(\chi_{\tau_i};\theta) = \frac{1}{1+\sum_{j=2}^{m}\exp(b_j^\top \breve{\chi}_{\tau_i})}$$

and for $2 \leq k \leq m$,

$$p_k(\chi_{\tau_i};\theta) = \frac{\exp(b_k^\top \breve{\chi}_{\tau_i})}{1+\sum_{j=2}^{m}\exp(b_j^\top \breve{\chi}_{\tau_i})},$$

with $\breve{\chi}_{\tau_i} = (\chi_{\tau_i}, 1)^\top \in \mathbb{R}^3$, $b_k \in \mathbb{R}^3$ for $k = 2, \ldots, m$ and $\theta = (b_2^\top, \ldots, b_m^\top)$. The components of χ_{τ_i} are respectively an "instantaneous volatility proxy" and a mean spread measure. More precisely, $10^{-8} \times \chi_{\tau_i,1}$ is equal to the sum of the squares of the one second log-returns over ten seconds before time τ_i and $\chi_{\tau_i,2}$ is equal to the average of the spreads observed at transaction times over 1 minute before time τ_i. The constant m is chosen as the smallest value such that the number of jumps larger than or equal to this value is smaller than 2%. Finally, all jumps whose size is bigger than m are treated as jumps of size m. The maximum likelihood estimated coefficients for the different assets are given in table VII.

TICKER	k	$b_{1,k}$	$b_{2,k}$	$b_{3,k}$
AIRF	2	0.0194 (0.0078)	40.2326 (7.0183)	−2.0951 (0.1124)
	3	0.0475 (0.0074)	56.1089 (8.1826)	−3.5840 (0.1591)
	4	0.0474 (0.0081)	65.9922 (9.2036)	−4.3299 (0.1979)
	5	0.0625 (0.0071)	88.1013 (7.7364)	−4.7260 (0.1881)
ALSO	2	0.0182 (0.0030)	23.8741 (1.7020)	−3.4288 (0.1252)
	3	0.0276 (0.0033)	31.95067 (2.2005)	−5.3454 (0.1892)
BNPP	2	0.0491 (0.0081)	70.0345 (5.5041)	−7.1428 (0.3184)
CAGR	2	0.0358 (0.0063)	87.1074 (7.0635)	−2.7149 (0.1021)
	3	0.0724 (0.0066)	139.95 (8.4273)	−4.7173 (0.1431)
DANO	2	0.0245 (0.0107)	35.3407 (4.8191)	−7.7316 (0.5506)
EAD	2	−0.0014* (0.0016)	88.5505 (6.0260)	−2.6112 (0.0980)
	3	−0.0004* (0.0008)	135.1651 (6.9225)	−4.3796 (0.1366)
	4	0.0000* (0.0002)	179.0233 (7.5008)	−5.7583 (0.1720)
FTE	2	0.0309 (0.0043)	129.01 (10.256)	−3.6411 (0.1384)
	3	0.0457 (0.0049)	157.83 (16.537)	−5.6923 (0.2455)
RENA	2	0.0048* (0.0046)	26.460 (2.7421)	−3.4175 (0.1798)
	3	0.0244 (0.0043)	43.331 (3.5935)	−6.3562 (0.3020)
SGOB	2	0.0367 (0.0042)	49.1837 (3.5182)	−5.6428 (0.2198)
TOTF	2	0.0184 (0.0049)	116.44 (16.7018)	−10.924 (0.9041)

TABLE VII

*Maximum likelihood estimation of the coefficients of the LOGIT regression. The symbol * indicates* 5% *non significative values.*

Everything else constant, the probability of the size of the jumps decreases with the size, according to Table V. Moreover this probability increases with the "instantaneous volatility" and the "mean spread", which confirms the intuition.

Test of the model

As suggested in Section 2.3, we eventually draw the autocorrelograms of $(|\hat{X}_{\tau_{i+1}} - \hat{X}_{\tau_i}|)_{i \in I_1}$ and give the associated p-values for the Ljung-Box statistics.

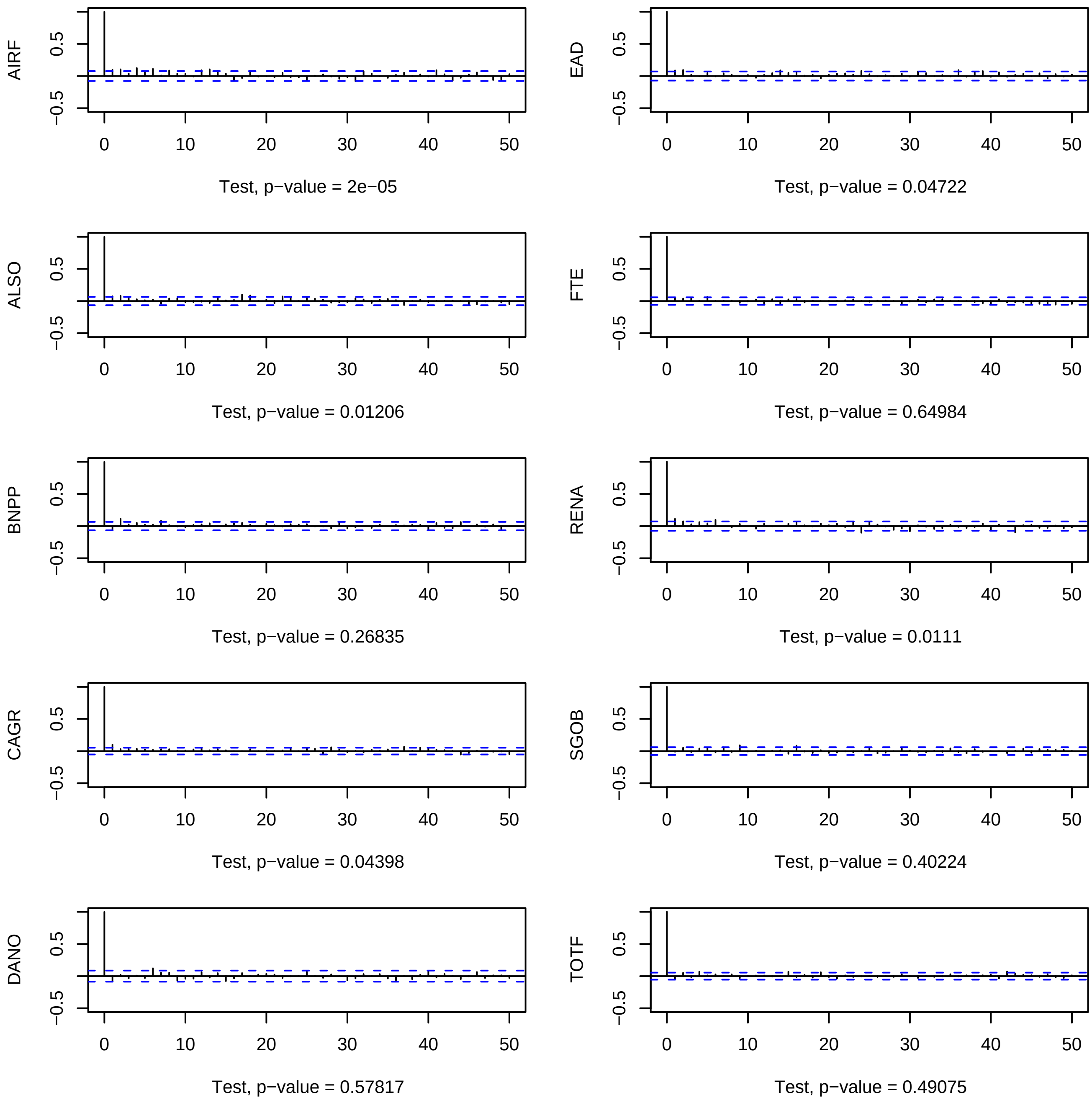

FIGURE 10.— *Test procedure*

The results are very satisfactory. The autocorrelograms of the corrected values $(|\hat{X}_{\tau_{i+1}} - \hat{X}_{\tau_i}|)_{i \in I_1}$ are systematically quite flat. Recall that these results are obtained with only one estimated parameter (η).

4.3. *Application: Estimation of the integrated volatility*

In Table VIII, we give for the ten assets the estimated values of the integrated volatility.

TICKER	Volatility (%) 2007-01-15	Volatility (%) 2007-01-16	Volatility (%) 2007-01-17	Volatility (%) 2007-01-18	Volatility (%) 2007-01-19
AIRF	20.5	21.0	17.6	20.7	33.8
ALSO	32.3	24.8	33.8	35.6	33.6
BNPP	15.3	15.0	17.5	21.1	18.6
CAGR	19.0	25.5	29.4	19.7	27.6
DANO	11.5	16.0	13.0	13.4	20.1
EAD	20.3	30.0	45.0	41.6	42.7
FTE	19.4	22.9	28.6	34.5	27.9
RENA	12.2	37.0	15.6	23.3	23.3
SGOB	16.6	25.6	22.0	26.8	25.9
TOTF	16.1	19.2	23.7	20.4	23.3

TABLE VIII

Estimation of the annualized volatility (square root of our estimator $\times 252^{1/2}$*).*

It is interesting to note that for stable values of η, we can have very varying values for the volatility, see for example Air France (AIRF). This is another element in favor of our model.

ACKNOWLEDGEMENTS

We are grateful to Charles-Albert Lehalle from Crédit Agricole Cheuvreux, Groupe CALYON, for providing the data and for fruitful discussions. We also thank Nour Meddahi for helpful comments.

REFERENCES

Aït-Sahalia, Y., Mykland, P., and L. Zhang (2005): "Ultra High Frequency Estimation with Dependent Microstructure Noise," To appear in *Journal of Econometrics*.

Andersen, T.G., Bollerslev, T., and N. Meddahi (2006): "Market Microstructure Noise and Realized Volatility Forecasting," To appear in *Journal of Econometrics*.

Barndorff-Nielsen, O., Hansen, P.R., Lunde, A., and N. Shephard (2009): "Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," *Econometrica* 76, 1481-1536.

Ball, C. (1988): "Estimation Bias Induced by Discrete Security Prices," *Journal of Finance*, 43, 841-865.

Bandi, F., and J. Russel (2008): " Microstructure Noise, Realized Variance and Optimal Sampling," *Review of Economic Studies*,75, 339-369.

Bauwens, L., and P. Giot (2003): "Asymmetric ACD Models: Introducing Price Information in ACD Models with a Two State Transition Model," *Empirical Economics*, 28, 123.

Bowsher, C.G. (2007): "Modelling Security Markets in Continuous Time: Intensity based, Multivariate Point Process Models," *Journal of Econometrics*, 141, 876-912.

Cho, D., and E. Frees (1988): “Estimating the Volatility of Discrete Stock Prices,” *Journal of Finance*, 43, 451-466.

Diebold, F.X., and G.H. Strasser (2008): “On the Correlation Structure of Microstructure Noise in Theory and Practice,” Working paper.

Engle, R. (2000): “The Econometrics of Ultra High Frequency Data,” *Econometrica*, 68, 1-22.

Engle, R., and J.R. Russell (1998): “Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data,” *Econometrica*, 66, 11271162.

Ghysels, E., Gourieroux, C., and J. Jasiak (2004): “Stochastic Volatility Duration Models,” *Journal of Econometrics*, 119, 413433.

Ghysels, E., and J. Jasiak (1998): “GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model,” *Studies in Nonlinear Dynamics and Econometrics*, 2, 133-149.

Ghysels, E., and A. Sinko (2007): “Volatility Prediction and Microstructure Noise,” To appear in *Journal of Econometrics*.

Gottlieb, G., and A. Kalay (1985): “Implications of the Discreteness of Observed Stock Prices,” *Journal of Finance*, 40, 135-154.

Gourieroux, C., and J. Jasiak (2001): “Financial Econometrics: Problems, Models, and Methods,” Princeton University Press.

Griffin, J.E., and R.C.A. Oomen (2008): “Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time ?,” *Econometric Reviews*, 27, 230-253.

Hansen, P.R., and A. Lunde (2006): “Realized Variance and Market Microstructure Noise,” *Journal of Business and Economics Statistics*, 24, 127-161.

Hasbrouck, J. (2000): “The Dynamics of Discrete Bid and Ask Quotes,” *Journal of Finance*, 54, 2109-2142.

Hasbrouck, J. (2006): “Empirical Market Microstructure,” *Oxford University Press*.

Hausman, J., Lo, A., and C. MacKinlay (1992): “An Ordered Probit Analysis of Transaction Stock Prices,” *Journal of Financial Economics*, 31, 319-330.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., and M. Vetter (2009): “Microstructure Noise in the Continuous Case: The Pre-Averaging Approach,” *Stochastic Processes and their Applications*, 119, 2249-2276.

Kalnina, I., and O.B. Linton (2008): “Estimating Quadratic Variation Consistently in the Presence of Correlated Measurement Error,” *Journal of Econometrics*, 147, 47-59.

Large, J. (2007): “Estimating Quadratic Variation when Quoted Prices Change by a Constant Increment,” To appear in *Journal of Econometrics*.

Li, Y., and P.A. Mykland (2007): “Are Volatility Estimators Robust with Respect to Modeling Assumptions?,” *Bernoulli*, 13, 601-622.

Renault, E., and B.J.M. Werker (2009): “Causality Effects in Return Volatility Measures with Random Times,” To appear in *Journal of Econometrics*.

Robert, C.Y., and M. Rosenbaum (2009a): “Volatility and Covariation Estimation when Microstructure Noise and Trading Times are Endogenous,” To appear in *Mathematical Finance*.

Robert, C.Y., and M. Rosenbaum (2009b): “On the Microstructural Hedging Error,” Working paper.

Rosenbaum, M. (2007): “A New Microstructure Noise Index,” To appear in *Quantitative Finance*.

Rosenbaum, M. (2009): “Integrated Volatility and Round-Off Error,” *Bernoulli*, 15, 687-720.

Russell, J.R., and R.F. Engle (2005): "A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The ACM-ACD Model," *Journal of Business and Economic Statistics*, 23, 166-180.

Zhang, L., Mykland, P.A., and Y. Aït-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394-1411.