Estimating the efficient price from the order flow

Sylvain Delattre, Christian Robert and Mathieu Rosenbaum

University Paris 7, University Lyon 1, University Paris 6

2 September 2013

Outline

Outline

What is the high frequency price ?

Classical approach in mathematical finance

- Prices of basic products (futures, stocks,. . .) are observed on the market.
- Their values are used in order to price complex derivatives.
- Options traders typically rebalance their portfolio once or a few times a day.
- So, derivatives pricing problems typically occur at the daily scale.

What is the high frequency price ?

High frequency setting

- When working at the ultra high frequency scale, even pricing a basic product, that is assigning a price to it, becomes a challenging issue.
- Indeed, one has access to trades and quotes in the order book.

Example of order book

Carnet d''ordre ML

MICHELIN 3	
FR0000121261 (ML)	
Dernier	**83,65**
Var (%) :	+4,56%
Var (pts) :	3,65
Ouvrir	80,15
Plus haut	85,1
Plus bas	80,15
volume	3 023 872

Heure	Prix 2	volume
13:15:05	83.65	50
13:14:50	83.65	100
13:14:32	83.65	79
13:14:32	83.65	11
13:14:30	83.65	49
13:14:30	83.65	3951
13:13:18	83.65	520
13:13:18	83.65	507
13:13:18	83.65	680
13:13:18	83.65	232

Demandes (Acheteurs - bid) 1

Nb. ordres	Quantités	Cours
1	4 771	83.65
1	225	83.60
3	2 810	83.50
2	1 946	83.45
4	5 955	83.40
11	**15 707**	

Offres (Vendeurs - ask)

Cours	Quantités	Nb. ordres
83.70	365	1
83.75	4 135	3
83.80	1 105	12
83.85	1 275	6
83.90	339	5
	7 219	**27**

What is the high frequency price ?

Different prices

- At a given time, many different notions of price can be defined for the same asset : last traded price, best bid price, best ask price, mid price, volume weighted average price,. . .
- This multiplicity of prices is problematic for many market participants.
- For example, market making strategies or brokers optimal execution algorithms often require single prices of plain assets as inputs.

What is the high frequency price ?

Pricing issues

- Choosing one definition or another for the price can sometimes lead to very significantly different outcomes for the strategies.
- This is for example the case when the tick value (the minimum price increment allowed on the market) is rather large.
- Indeed, this implies that the prices mentioned above differ in a non negligible way.

What is the high frequency price ?

Efficient price

- In practice, high frequency market participants are not looking for the "fair" economic value of the asset.
- What they need is rather a price whose value at some given time summarizes in a suitable way the opinions of market participants at this time.
- This price is called **efficient price**.
- We aim at providing a statistical procedure in order to estimate this efficient price.

Ideas for the estimation strategy

Efficient price

- We focus on large tick assets and assume that the efficient price essentially lies inside the bid-ask spread.
- In order to retrieve the efficient price, the classical approach is to consider the imbalance of the order book.
- Indeed, it is often said by market participants that **the price is where the volume is not**.
- We use a dynamic version of this idea through the order flow.
- We assume that the intensity of arrival of the limit order flow at the best bid level (say) depends on the distance between the efficient price and this level.
- If this distance is large, the intensity should be high and conversely.

Ideas for the estimation strategy

Response function

- We assume the intensity can be written as an increasing deterministic function of this distance.
- This function is called the **order flow response function**.
- A crucial step before estimating the price is to estimate the response function in a non parametric way.
- Then, this functional estimator is used in order to retrieve the efficient price.
- It is also possible to use the buy or sell market order flow. In that case, the intensity of the flow should be high when the distance is small.
- Indeed, in this situation, market takers are not loosing too much money when crossing the spread.

Description of the model

The model

- We assume the bid-ask spread is constant equal to one (tick).
- The efficient price P_t is simply given by $P_0 + \sigma W_t$, with P_0 uniformly distributed on $[p_0, p_0 + 1]$, with p_0 an integer.
- We assume that when a limit order is posted at time t at the best bid level, its price is given by $\lfloor P_t \rfloor$.

Description of the model

The model

- Let N_t be the total number of limit orders posted over $[0, t]$.
- We assume that $(N_t)_{t \geq 0}$ is a Cox process with arrival intensity at time t given by
$$\mu h(Y_t),$$
with
$$Y_t = P_t - \lfloor P_t \rfloor = \{P_t\}$$
and $\int_0^1 h(x)dx = 1$ (identifiability condition).
- The limiting case where h is constant corresponds to orders arriving according to a standard Poisson process.

Observations

Asymptotic setting

- We observe the point process (N_t) on $[0, T]$.
- We let T tend to infinity. It is also necessary to assume that $\mu = \mu_T$ depends on T.
- More precisely
$$T^{5/2+\varepsilon}/\mu_T \to 0.$$

Properties of the process Y_t

Markov process

- Recall that if U is uniformly distributed on $[0,1]$ and X is a real-valued random variable, which is independent of U then $\{U+X\}$ is also uniformly distributed on $[0,1]$.
- We obtain that (Y_t) is a stationary Markov process such that, almost surely,

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T f(Y_s)ds=\int_0^1 f(s)ds.$$

Properties of the process Y_t

Regenerative process

- (Y_t) also enjoys a regenerative property.
- Let $\nu_0 = 0$, $\nu_1 = \inf\{t > 0 : P_t \in \mathbb{N}\}$ and for $n \geq 2$:

$$\begin{aligned}\nu_n &= \inf\{t > \nu_{n-1} : P_t = P_{\nu_{n-1}} \pm 1\}\\ &= \inf\{t > \nu_{n-1} : W_t = W_{\nu_{n-1}} \pm 1/\sigma\}.\end{aligned}$$

- The cycles $(Y_{t+\nu_n})_{0 \leq t < \nu_{n+1} - \nu_n}$ are independent and identically distributed for $n \geq 1$.

Properties of the process Y_t

Limiting behavior

- We get that almost surely

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T f(Y_s)ds = \sigma^2\mathbb{E}\Big(\int_0^{\tau_1} f(\{\sigma W_t\})dt\Big).$$

- In particular, this implies that

$$\sigma^2\mathbb{E}\Big[\int_0^{\tau_1} f(\{\sigma W_t\})dt\Big] = \int_0^1 f(s)ds.$$

- Furthermore,

$$\sqrt{T}\Big(\frac{1}{T}\int_0^T f(Y_t)dt - \int_0^1 f(s)ds\Big]\Big) \xrightarrow{d} N(0,\sigma^2\mathrm{Var}[Z^f]).$$

Outline

Step 1

Estimation of μ_T

- Recall that the intensity of the point process is given by $\mu_T h(Y_t)$ with Y_t the fractional part of P_t.
- Before estimating h, we need to estimate μ_T.
- We have

$$\mathbb{E}\left[\frac{N_T}{\mu_T T}\right] = \mathbb{E}\left[\frac{1}{T}\int_0^T h(Y_t)dt\right] = \frac{1}{T}\int_0^T \mathbb{E}[h(Y_t)]dt = 1.$$

Step 1

Proposition : Estimation of μ_T

- We easily show that

$$\sqrt{T}\left(\frac{\hat{\mu}_T}{\mu_T} - 1\right) \xrightarrow{d} N(0, \sigma^2 Var[Z^h]).$$

Step 2

Estimation of h

- Let k_T be a known deterministic sequence of positive integers. Then define for $j = 1, \dots, k_T$

$$\hat{\theta}_j = k_T \frac{N_{jT/k_T} - N_{(j-1)T/k_T}}{\hat{\mu}_T T} = \frac{k_T}{N_T}\big(N_{jT/k_T} - N_{(j-1)T/k_T}\big).$$

- $\hat{\theta}_j$ is approximately equal to

$$\frac{1}{\mu_T T/k_T} \sum_{i=1}^{\lfloor \mu_T T/k_T \rfloor} \big(N_{(j-1)T/k_T + i/\mu_T} - N_{(j-1)T/k_T + (i-1)/\mu_T}\big).$$

Step 2

Estimation of h

- Conditional on the path of (Y_t), the variables in the sum are independent and if T/k_T is small enough, they approximately follow a Poisson law with parameter $h(Y_{(j-1)T/k_T})$.
- Therefore, if moreover $\mu_T T/k_T$ is sufficiently large, one can expect that $\hat{\theta}_j$ is close to $h(Y_{(j-1)T/k_T})$.
- We assume that k_T is chosen so that for some $p > 0$, as T tends to infinity,

$$T^{p+1/2}/k_T^{p/2} \to 0, \quad k_T T^{1/2}/\mu_T \to 0.$$

Step 2

Estimation of h

- The $\hat{\theta}_j$ introduced above are k_T estimators of quantities of the form $h(u_j)$.
- However, we do not have access to the values of the u_j !
- Nevertheless, we know that they are uniformly distributed on $[0,1]$. We therefore rank the $\hat{\theta}_j$: $\hat{\theta}_{(1)} \leq \hat{\theta}_{(2)} \leq \ldots \leq \hat{\theta}_{(k_T)}$.
- For $u \in [0,1)$, we define the estimator of $h(u)$ the following way :

$$\hat{h}(u) = \hat{\theta}_{(\lfloor uk_T \rfloor + 1)}.$$

Step 3

Estimation of h^{-1}

- Then, the estimator of h^{-1} is naturally defined by the right continuous generalized inverse of $\hat{h}$:

$$\hat{h}^{-1}(t) = \frac{1}{k_T} \sum_{j=1}^{k_T} \mathbb{I}_{\{\hat{\theta}_j \leq t\}}.$$

Response function

Theorem

We have the two following convergences in law in the Skorohod space :

$$\sqrt{T}\big(\hat{h}^{-1}(\cdot) - h^{-1}(\cdot)\big) \xrightarrow{d} \sigma G(\cdot) - \frac{(\cdot)}{h'\big(h^{-1}(\cdot)\big)} \int_0^{h(1^-)} \sigma G(v)dv,$$

$$\sqrt{T}\big(\hat{h}(\cdot) - h(\cdot)\big) \xrightarrow{d} -h'(\cdot)\sigma G\big(h(\cdot)\big) + h(\cdot) \int_0^{h(1^-)} \sigma G(v)dv,$$

where $G(\cdot)$ is a continuous centered Gaussian process with covariance function which is explicitly defined.

Estimation of the efficient price

Theorem

Let

$$\widehat{h(Y_t)} = k_T \frac{N_t - N_{t-T/k_T}}{\hat{\mu}_T T}.$$

and

$$\widehat{Y_t} = \hat{h}^{-1}\big(\widehat{h(Y_t)}\big).$$

We have

$$\sqrt{T}\big(\widehat{Y_t} - Y_t\big) \xrightarrow{d} \sigma G\big(h(Y_t)\big),$$

with G independent of Y_t.

Outline

Notation

Oracle quantities

- Recall that

$$\hat{\theta}_j = k_T \frac{N_{jT/k_T} - N_{(j-1)T/k_T}}{N_T}.$$

- We set

$$\theta_j = k_T \frac{N_{jT/k_T} - N_{(j-1)T/k_T}}{\mu_T T},$$

  and $\hat{h}_e(u) = \theta_{(\lfloor uk_T \rfloor + 1)}$.

- We have

$$\hat{h}_e^{-1}(\theta) = \frac{1}{k_T} \sum_{j=1}^{k_T} \mathbb{I}_{\{\theta_j \leq \theta\}}.$$

A first convergence

The following proposition is a key element for the proof of the theorem.

Proposition

We have

$$\sqrt{T}\big(\hat{h}_e^{-1}(\cdot) - h^{-1}(\cdot)\big) \xrightarrow{d} \sigma^2 G(\cdot) \quad \text{in } D[0, h(1^-)),$$

where $G(\cdot)$ is a centered Gaussian process with explicit covariance function.

Proof of the proposition

Decomposition

- We write $\hat{h}_e^{-1}(t) - h^{-1}(t) = T_1 + T_2 + T_3$, with

$$T_1 = \frac{1}{k_T}\sum_{j=1}^{k_T}\mathbb{I}_{\{\theta_j \leq t\}} - \frac{1}{k_T}\sum_{j=1}^{k_T}\mathbb{I}_{\{\frac{k_T}{T}\int_{(j-1)T/k_T}^{jT/k_T} h(Y_u)du \leq t\}},$$

$$T_2 = \frac{1}{k_T}\sum_{j=1}^{k_T}\mathbb{I}_{\{\frac{k_T}{T}\int_{(j-1)T/k_T}^{jT/k_T} h(Y_u)du \leq t\}} - \frac{1}{T}\sum_{j=1}^{k_T}\int_{(j-1)T/k_T}^{jT/k_T}\mathbb{I}_{\{h(Y_u)\leq t\}}du,$$

$$T_3 = \frac{1}{T}\int_0^T \mathbb{I}_{\{h(Y_u)\leq t\}}du - h^{-1}(t).$$

- The last term is treated thanks to the previous CLT and the two others are shown to be negligible.

Proof of the proposition

CLT

This gives

$$\sqrt{T}\left(\hat{h}_e^{-1}(t) - h^{-1}(t)\right) \xrightarrow{d} N\left(0, \sigma^2 \mathrm{Var}[Z_e(t)]\right)$$

where

$$Z_e(t) = \int_0^{\tau_1} \left(\mathbb{I}_{\{W_s<0, h(1+\sigma W_s)\leq t\}} + \mathbb{I}_{\{W_s>0, h(\sigma W_s)\leq t\}} - h^{-1}(t)\right) ds.$$

Proof of the proposition

Finite dimensional convergence

- We obtain a multidimensional CLT in the same way.
- We have that $Z_e(t)$ is equal to

$$\int_{-1/\sigma}^{1/\sigma}\big(\mathbb{I}_{\{u<0,h(1+\sigma u)\leq t\}}+\mathbb{I}_{\{u>0,h(\sigma u)\leq t\}}-h^{-1}(t)\big)L_{-1/\sigma,1/\sigma}(u)du,$$

  where $L_{-1/\sigma,1/\sigma}(u)$ is the local time stopped at the first exit time from $(-1/\sigma,1/\sigma)$.
- This enables to show that $\mathbb{E}[Z_e(t)]=0$ and to compute explicitly the limiting covariance function $\mathbb{E}[Z_e(t_1)Z_e(t_2)]$.

Proof of the proposition

Tightness

- It remains to prove the tightness of

$$\alpha_T(t) = \sqrt{T}\Big(\frac{1}{T}\int_0^T \mathbb{I}_{\{h(Y_s)\leq t\}}\mathrm{d}s - h^{-1}(t)\Big).$$

- This is done showing that for some $p > 0$ and $p_1 > 1$ and all $0 \leq t_1, t_2 < h(1^-)$:

$$\mathbb{E}\big[|\alpha_T(t_1) - \alpha_T(t_2)|^p\big] \leq c|t_1 - t_2|^{p_1}.$$

Proof of the proposition

Tightness

- We need to consider terms of the form :

$$Y_i(t_1,t_2) = \frac{1}{\sqrt{T}}\Big(\int_{\nu_{i-1}}^{\nu_i} \mathbb{I}_{\{t_1<h(Y_t)\leq t_2\}}dt - \frac{1}{\sigma^2}\big(h^{-1}(t_2)-h^{-1}(t_1)\big)\Big).$$

- Using a local time version of BDG inequality, we show that the following inequality enables to prove tightness

$$\mathbb{E}\big[\big(Y_i(t_1,t_2)\big)^2\big] \leq T^{-1}\mathbb{E}\Big[\Big(\int_{\nu_{i-1}}^{\nu_i} \mathbb{I}_{\{t_1<h(Y_t)\leq t_2\}}dt\Big)^2\Big]$$
$$\leq cT^{-1}|t_2-t_1|^2\mathbb{E}[(L^*)^2],$$

with $L^* = \sup_{u\in[-1/\sigma,1/\sigma]}\big(L_{-1/\sigma,1/\sigma}(u)\big).$

From the proposition to the theorem

Composition and inverse

- The theorems are deduced from the property.
- Indeed, we have

$$\sqrt{T}\big(\hat{h}_e(\cdot) - h(\cdot)\big) \xrightarrow{d} -\sigma h'(\cdot) G\big(h(\cdot)\big),$$

$$\sqrt{T}\Big(\int_0^1 \hat{h}_e(u)du - 1\Big) \xrightarrow{d} -\sigma \int_0^{h(1)} G(v)dv.$$

- Then, remark that

$$\hat{h}^{-1}(t) = \hat{h}_e^{-1}(t\hat{\mu}_T/\mu_T) \text{ and } \hat{\mu}_T/\mu_T = \int_0^1 \hat{h}_e^{-1}(u)du.$$

From the proposition to the theorem

Composition and inverse

- We write $\sqrt{T}\big(\hat{h}^{-1}(\cdot) - h^{-1}(\cdot)\big)$ as

$$= \sqrt{T}\Big(\hat{h}_e^{-1}\big(\cdot\,(\hat{\mu}_T/\mu_T)\big) - h^{-1}\big(\cdot\,(\hat{\mu}_T/\mu_T)\big)\Big)$$
$$+ \sqrt{T}\Big(h^{-1}\big(\cdot\,(\hat{\mu}_T/\mu_T)\big) - h^{-1}(\cdot)\Big),$$

- From the preceding proposition together with the functional delta method and the inverse map theorem, we get the results.

Outline

One experiment on real data

Setting

- Asset : Bund contract on the EUREX market.
- T=5 hours (8 am - 13 am).
- Windows : 30 seconds.
- We compute h^{-1}.

ith estimates of N arrival every second over the past