Estimation ultra haute fréquence de la volatilité et de la co-volatilité

Christian Y. Robert[1] et Mathieu Rosenbaum[2]

[1] CREST-ENSAE Paris Tech, [2] CMAP-École Polytechnique Paris

Plan

1. Données haute fréquence et bruit de microstructure
2. Modèles de microstructure et estimation de la volatilité
3. Modèle avec zones d'incertitude
4. Estimation de la volatilité
5. Estimation de la co-volatilité

Outline

1. Données haute fréquence et bruit de microstructure
2. Modèles de microstructure et estimation de la volatilité
3. Modèle avec zones d'incertitude
4. Estimation de la volatilité
5. Estimation de la co-volatilité

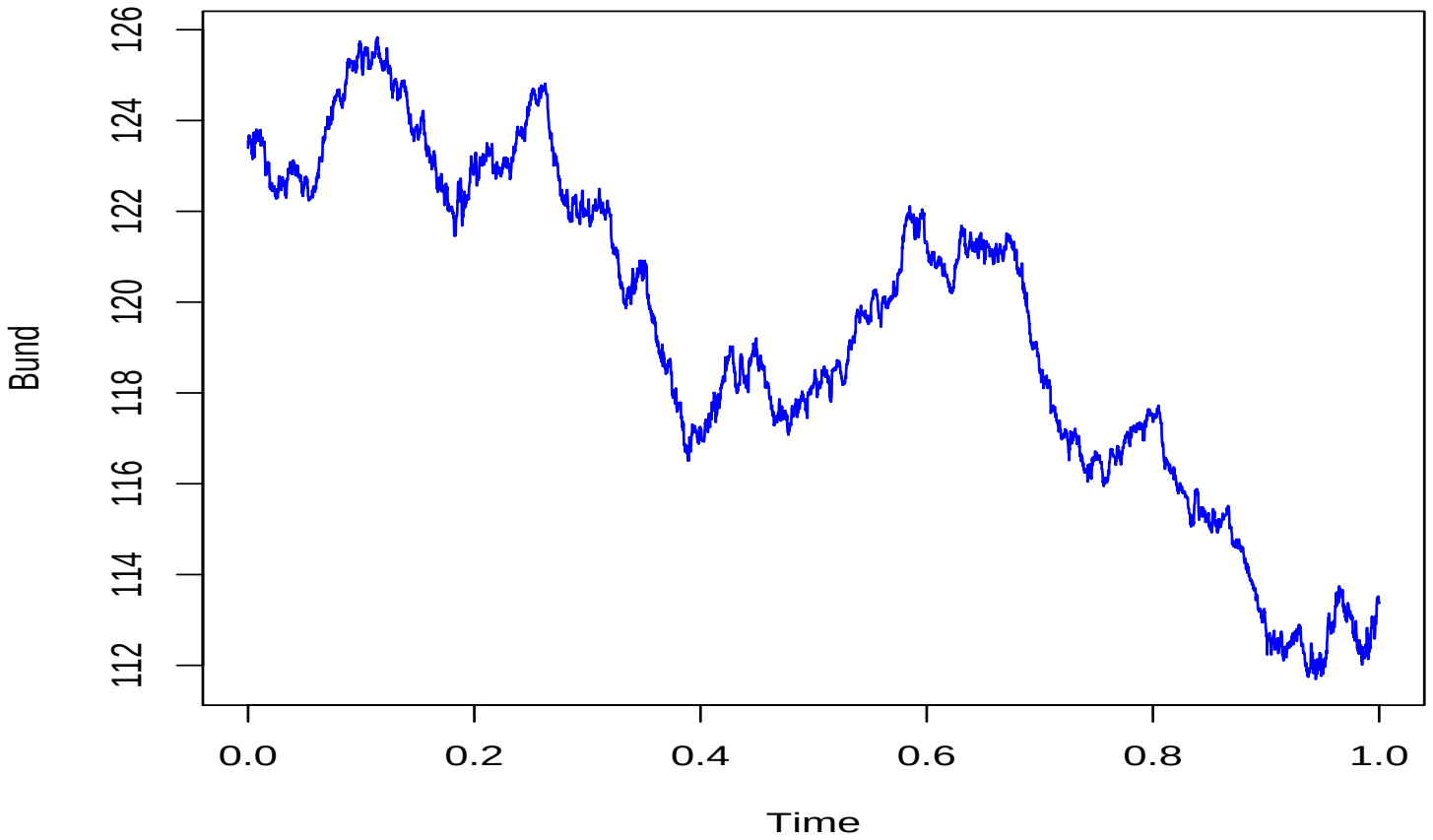
Modèles classiques

- Habituellement, les actifs sont supposés suivre des processus d'Ito de la forme :

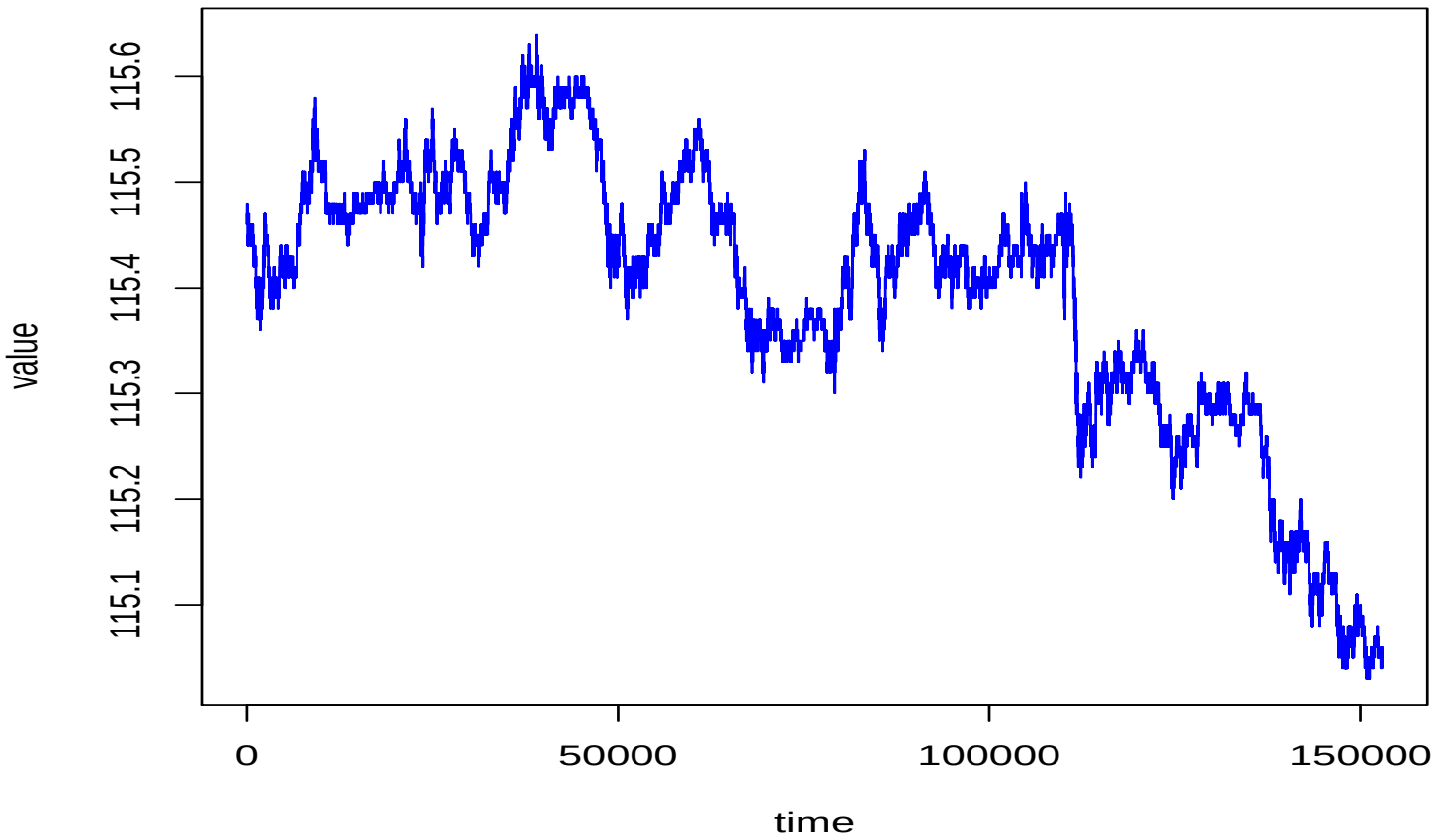
$$\mathrm{d}\log X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}B_t, t \in [0, T].$$

- Ces modèles sont « valides » dans les basses fréquences.

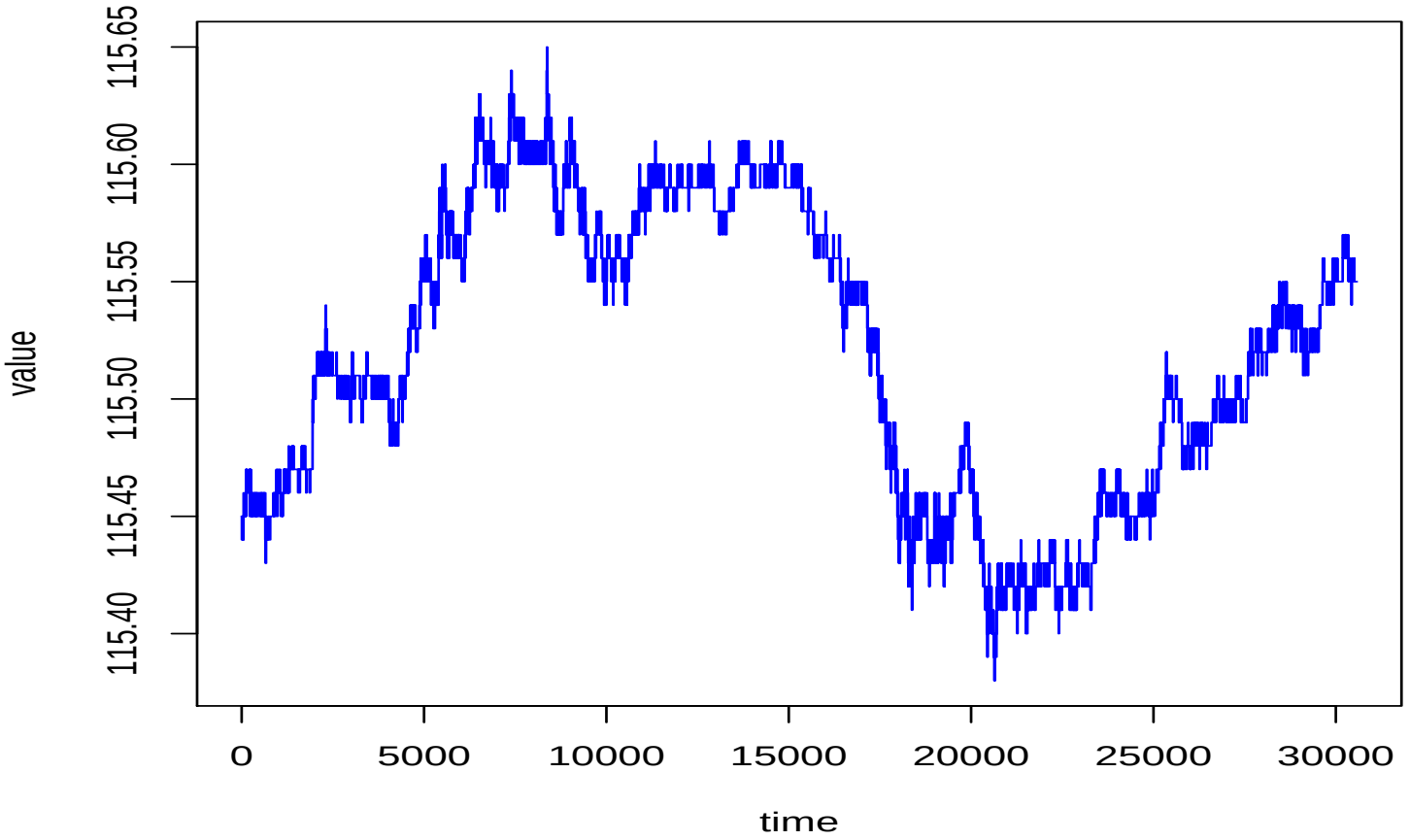
Contrat Bund, du 01-09-2005 au 31-01-2007, une donnée par heure



Contrat Bund, du 05-02-2007 au 09-02-2007, une donnée par seconde

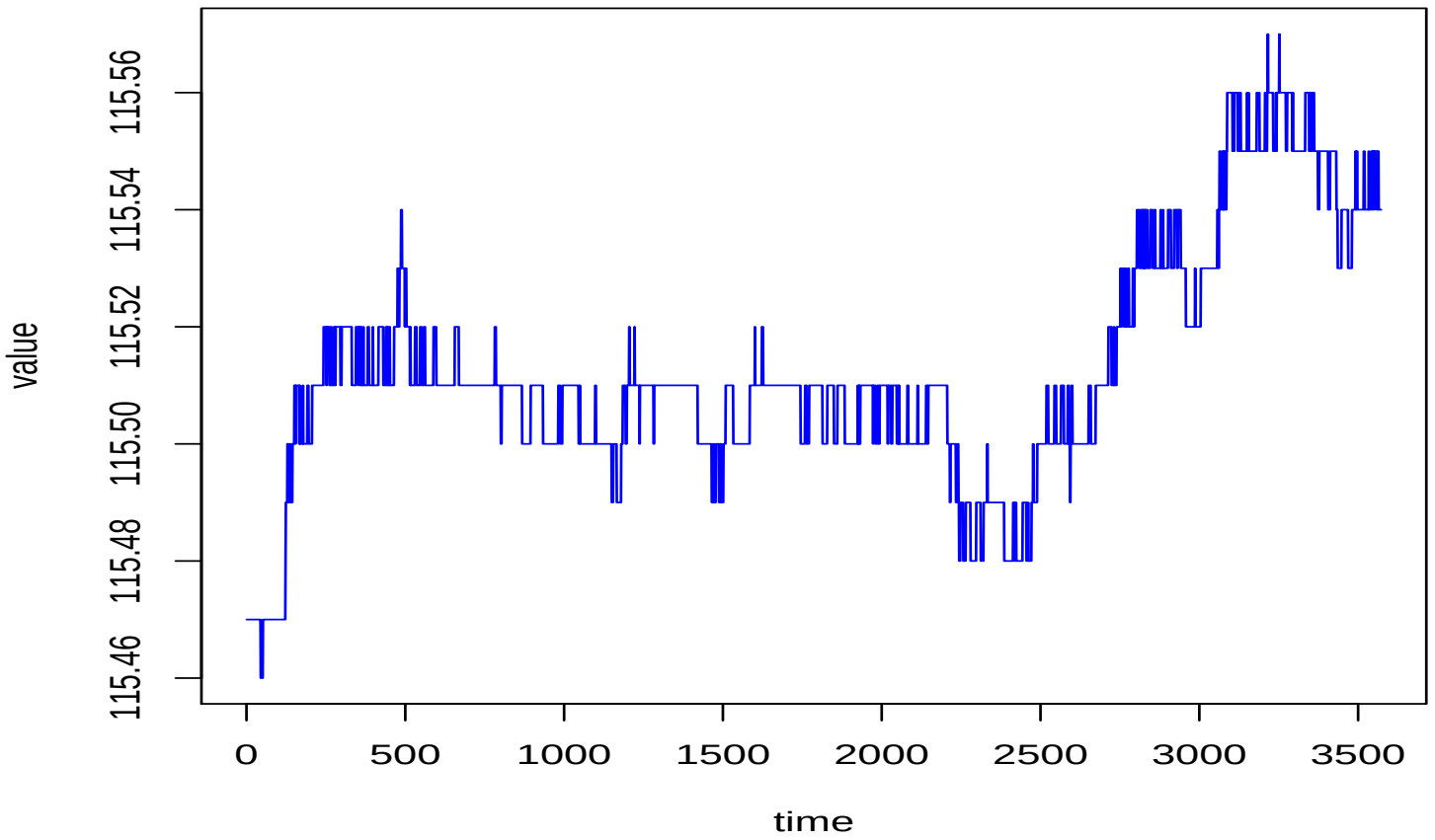


Contrat Bund, 06-05-2007, une donnée par seconde



.

Contrat Bund, 06-05-2007, de 10h à 11h, une donnée par seconde



Données haute fréquence

- Les données haute fréquence ne se comportent pas comme des observations d'une semi-martingale d'Ito.
- En particulier, elles vérifient les propriétés suivantes :
  - données discrètes,
  - dates d'observation aléatoires,
  - saisonnalité intraday.

Bruit de microstructure

- La différence entre le logarithme du prix observé et celui du prix efficient définit le bruit de microstructure.
- Les principales caractéristiques du bruit de microstructure sont :
  - variation quadratique finie,
  - covariation négative avec le log prix efficient,
  - saisonnalité intraday.

Outline

1 Données haute fréquence et bruit de microstructure

2 Modèles de microstructure et estimation de la volatilité

3 Modèle avec zones d'incertitude

4 Estimation de la volatilité

5 Estimation de la co-volatilité

Une quantité pertinente

Il est nécessaire de construire des modèles pour les données haute fréquence afin d'obtenir des estimations de la volatilité.

Quelles utilisations ?

Modèles de market impact, Modèles d'exécution d'ordre limite, Algorithmes d'optimisation, Probabilités d'exécution, Prévision de volatilité, Mesures de risque, Pricing d'option. Voir en particulier Lehalle (2008).

Une quantité pertinente : la volatilité intégrée

$$\int_0^1 \sigma_t^2 \mathrm{d}t.$$

Modèles de bruit de microstructure

Différents types de modèle ont pour but de tenir compte du bruit de microstructure. Des procédures statistiques spécifiques leur sont associées (l'estimateur classique « Realized Volatility » étant extrêmement biaisé dans les hautes fréquences).

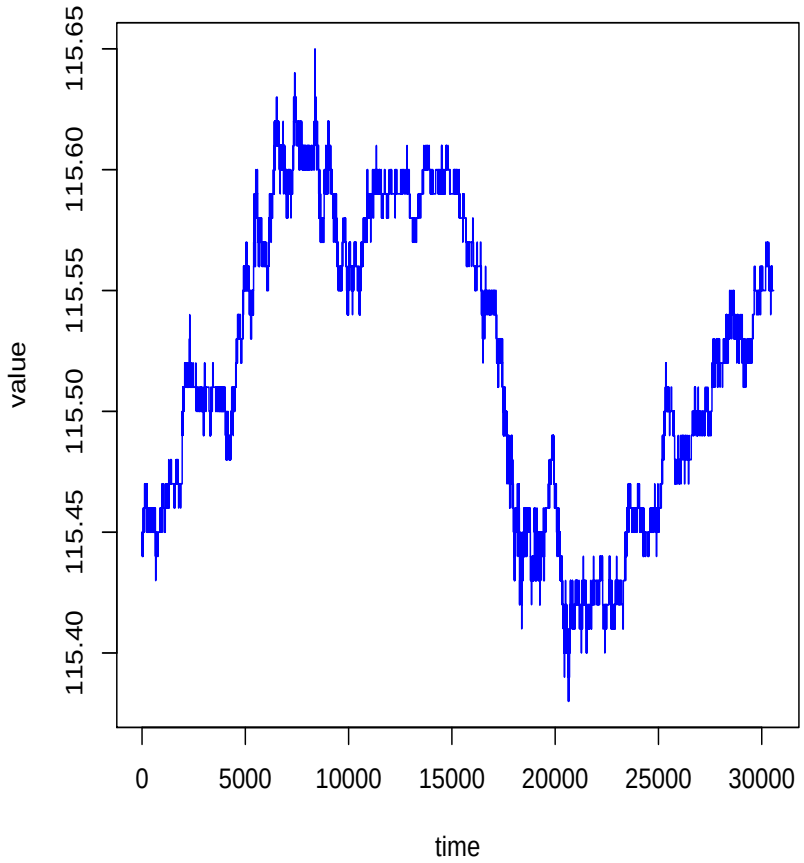
Modélisation du bruit

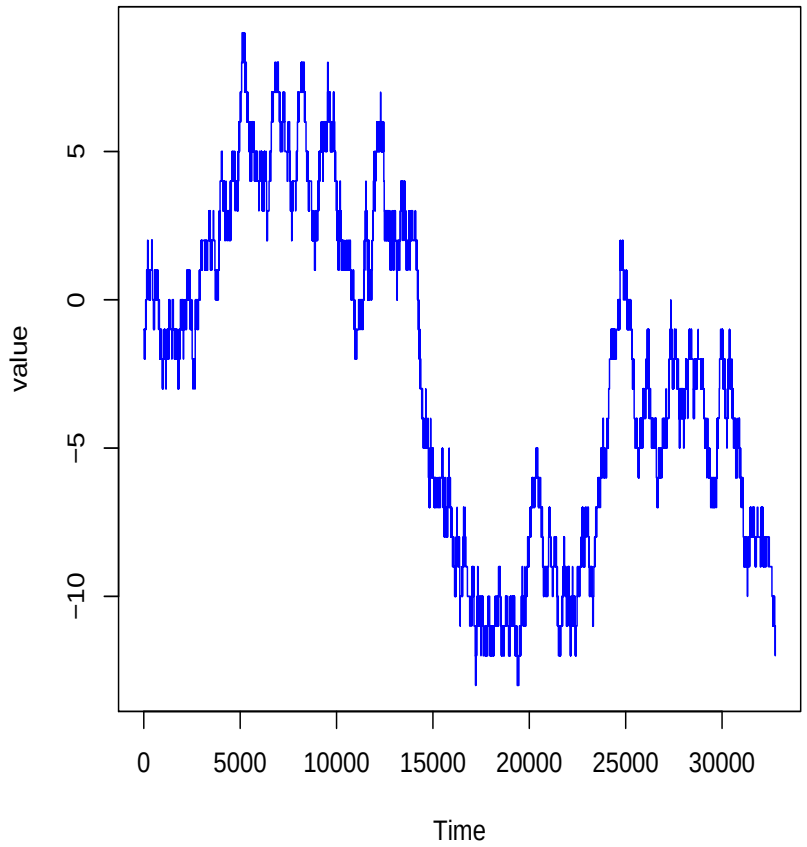
- Bruit additif : Aït-Sahalia, Mykland and Zhang (2006), Zhang (2006), Barndorff-Nielsen *et al.* (2008).
- Contamination du prix efficient à travers un noyau markovien : Li and Mykland (2007), Jacod *et al.* (2007).
- Modèles de prix discrets : Delattre and Jacod (1997), Large (2007), R. (2007), Fukasawa (2008).

Le modèle d'arrondi pur

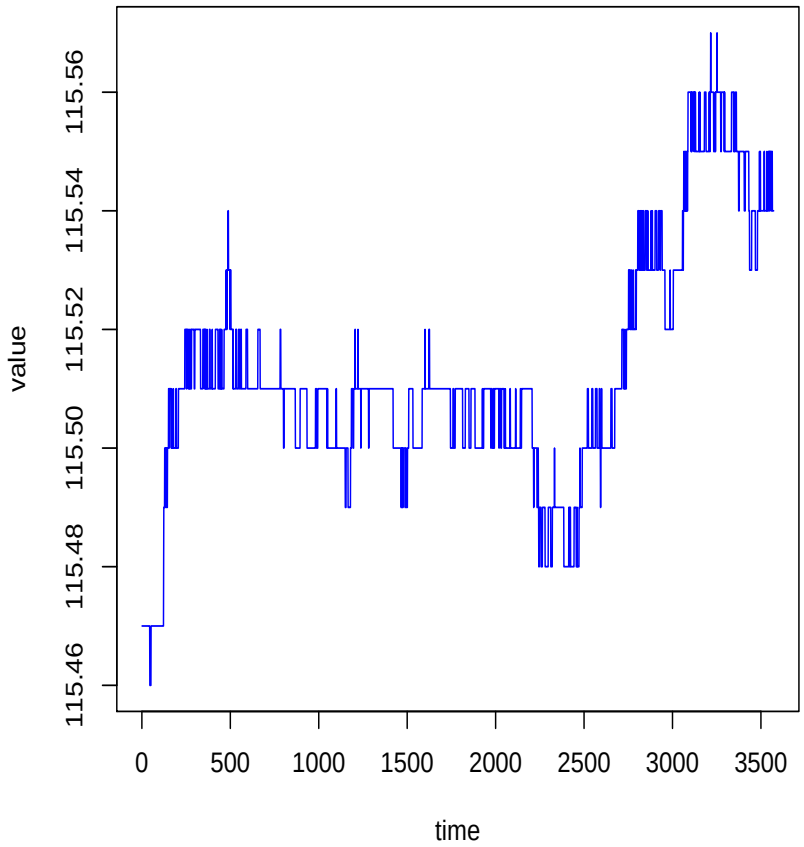
- Certains des modèles précédents ne permettent pas d'obtenir des prix discrets.
- Un modèle de diffusion observé avec erreur d'arrondi est un moyen simple d'obtenir des prix discrets et un comportement diffusif pour un échantillonnage basse fréquence.

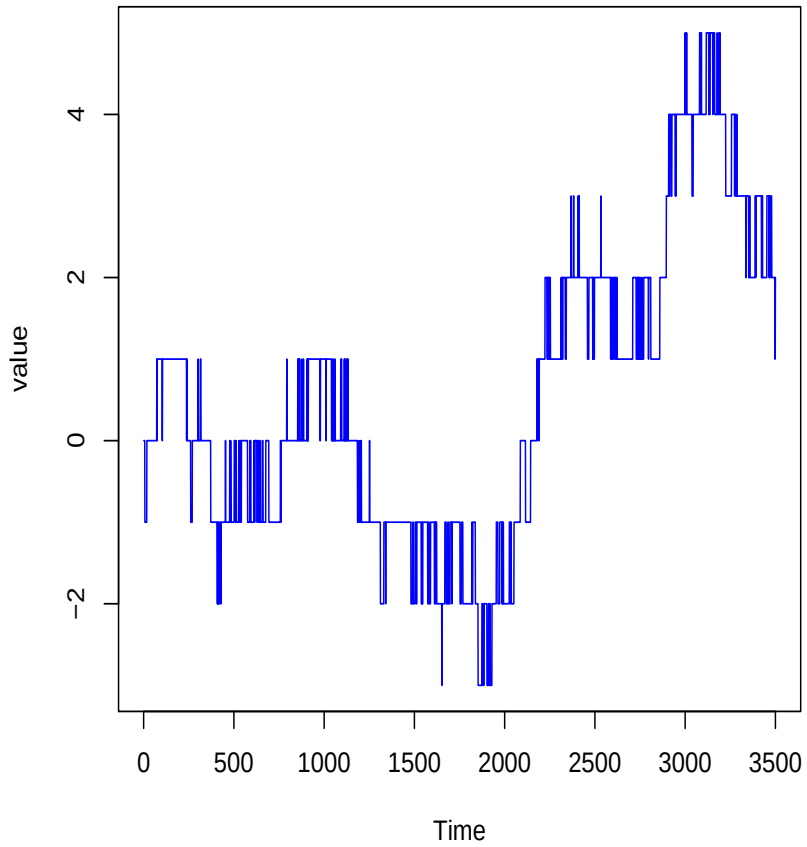
Bund, 06-05-2007, une donnée par seconde et $\lfloor 20 W_t \rfloor$, sur $[0,1]$, $n = 2^{15}$





Bund, 06-05-2007,de 10h à 11h, une donnée par seconde et $\lfloor 20 W_t \rfloor$, sur $[0, 0.1]$





Inconvénients du modèle d'arrondi pur

- Un modèle d'arrondi pur n'est pas « valide » dans les échelles extrêmement fines (trop d'oscillations).
- Comme beaucoup d'autres modèles, il est traité dans le cadre d'un échantillonnage déterministe et régulier.
  - Quelle fréquence utiliser : 1 seconde ? 10 secondes ? 1 minute ?
  - Quel prix utiliser : prix bid ? prix midquote ? prix last traded ?

Outline

1 Données haute fréquence et bruit de microstructure

2 Modèles de microstructure et estimation de la volatilité

3 Modèle avec zones d'incertitude

4 Estimation de la volatilité

5 Estimation de la co-volatilité

Aversion aux changements de prix

Aversion aux changements de prix

- Dans un cadre idéal, les transactions n'auraient lieu que lorsque le prix efficient croise la grille des ticks.
- En pratique, incertitude par rapport au prix efficient et aversion aux changements de prix des participants de marché.
- Le prix change seulement lorsque les participants de marché sont convaincus que le prix efficient est suffisamment loin du dernier prix traité.
- On introduit un paramètre η qui quantifie cette aversion aux changements de prix.

Modèle avec zones d'incertitude

Modèle avec zones d'incertitude : notations

- Prix efficient : X_t.
- t_i : date de la i^e transaction avec changement de prix.
- P_{t_i} : prix de transaction à la date t_i.
- $L_i = |P_{t_{i+1}} - P_{t_i}|$: taille en tick du i^e changement de prix.
- Variables explicatives pour la taille des sauts : χ_t.
- Zones d'incertitude : $U_k = [0, \infty) \times (d_k, u_k)$ avec

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

- τ_i : i^e temps de sortie d'une zone d'incertitude.

Modèle avec zones d'incertitude

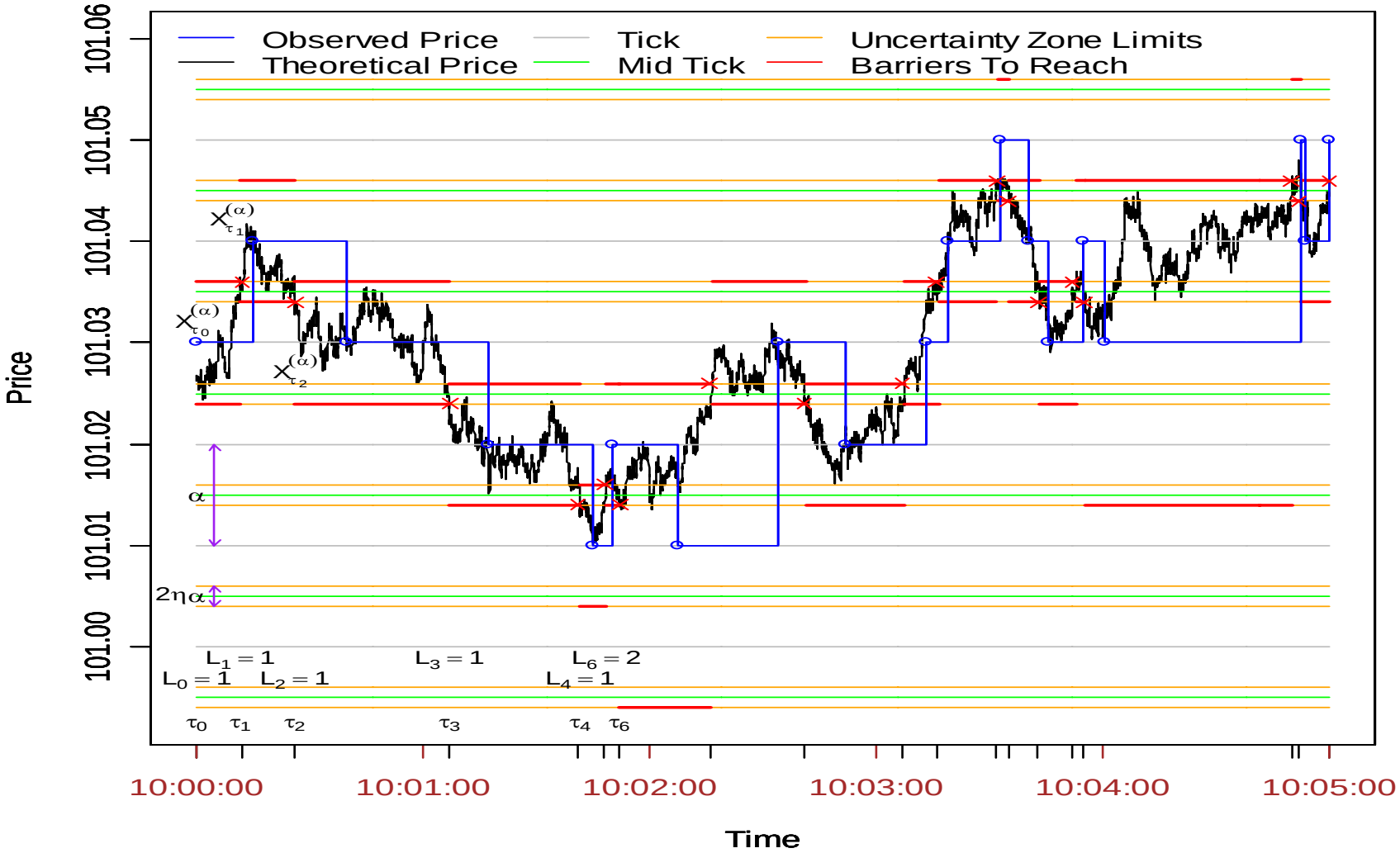
Modèle avec zones d'incertitude : dynamiques

- $$\mathrm{d}\log X_u = a_u \mathrm{d}u + \sigma_{u-}\mathrm{d}W_u.$$
- $$\mathbb{P}[L_i = s|\mathcal{F}_{\tau_i}] = \phi_s(\chi_{\tau_i}).$$
- $$\tau_{i+1} = \inf\left\{t > \tau_i, X_t = X^{(\alpha)}_{\tau_i} \pm \alpha(L_i - \frac{1}{2} + \eta)\right\},$$

avec $X^{(\alpha)}_{\tau_i}$ la valeur de X_{τ_i} arrondie au plus proche multiple de α.

- $$\tau_i \leq t_i < \tau_{i+1} \text{ et } P_{t_i} = X^{(\alpha)}_{\tau_i}.$$

Modèle avec zones d'incertitude



Discussion

Commentaires sur le modèle

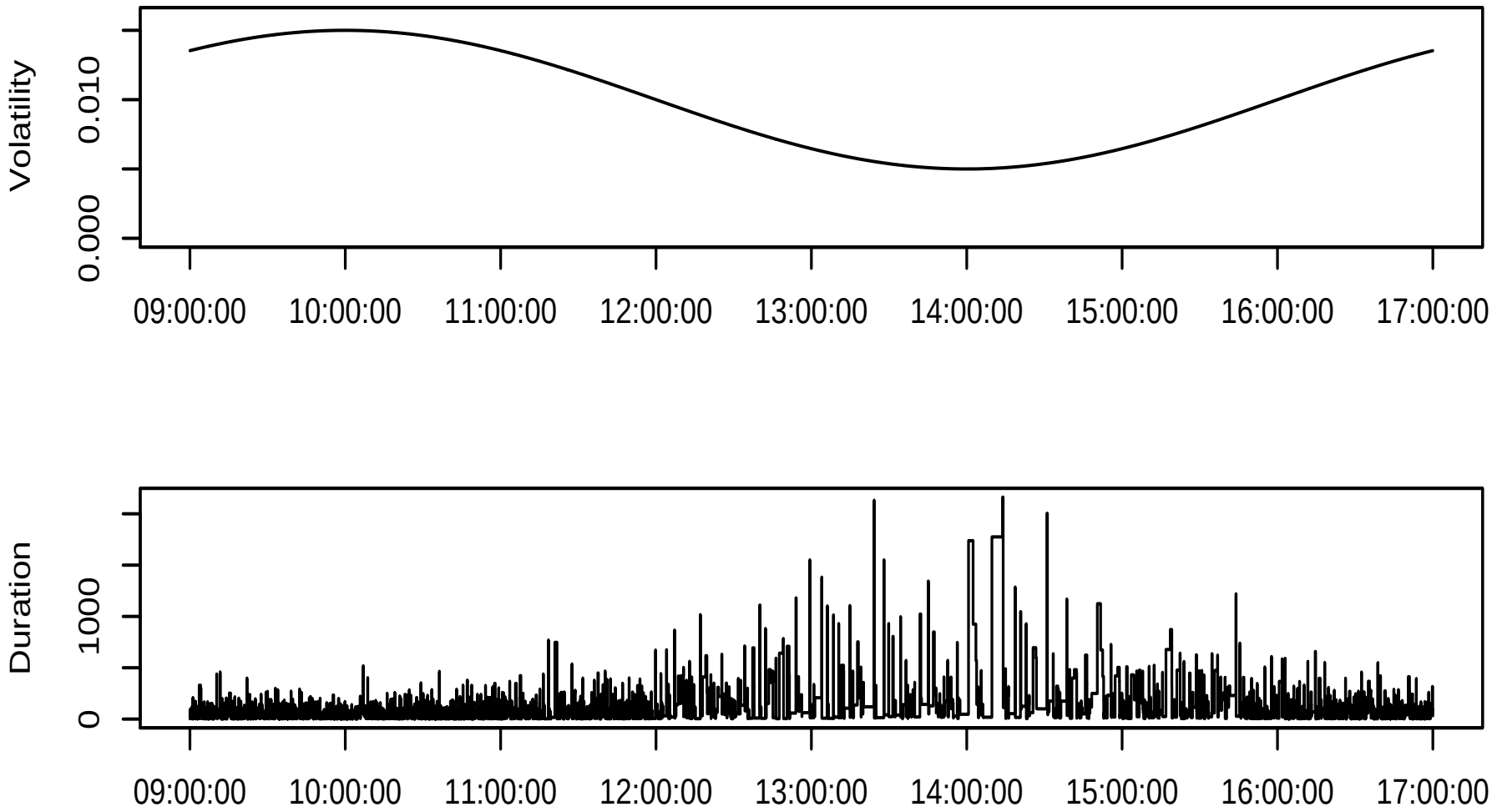
- Modèle Variations/Durations pour les prix last traded.
- Prix efficient latent semi-martingale.
- Dates de transaction aléatoires. Relation inverse entre durations et volatilité.
- Dates de transaction non nécessairement égales aux temps de sortie.
- Sauts de plusieurs ticks. Taille des sauts déterminée par des variables explicatives, liées par exemple au carnet d'ordre.
- Modèle structurel pour le bruit de microstructure.

Le paramètre η

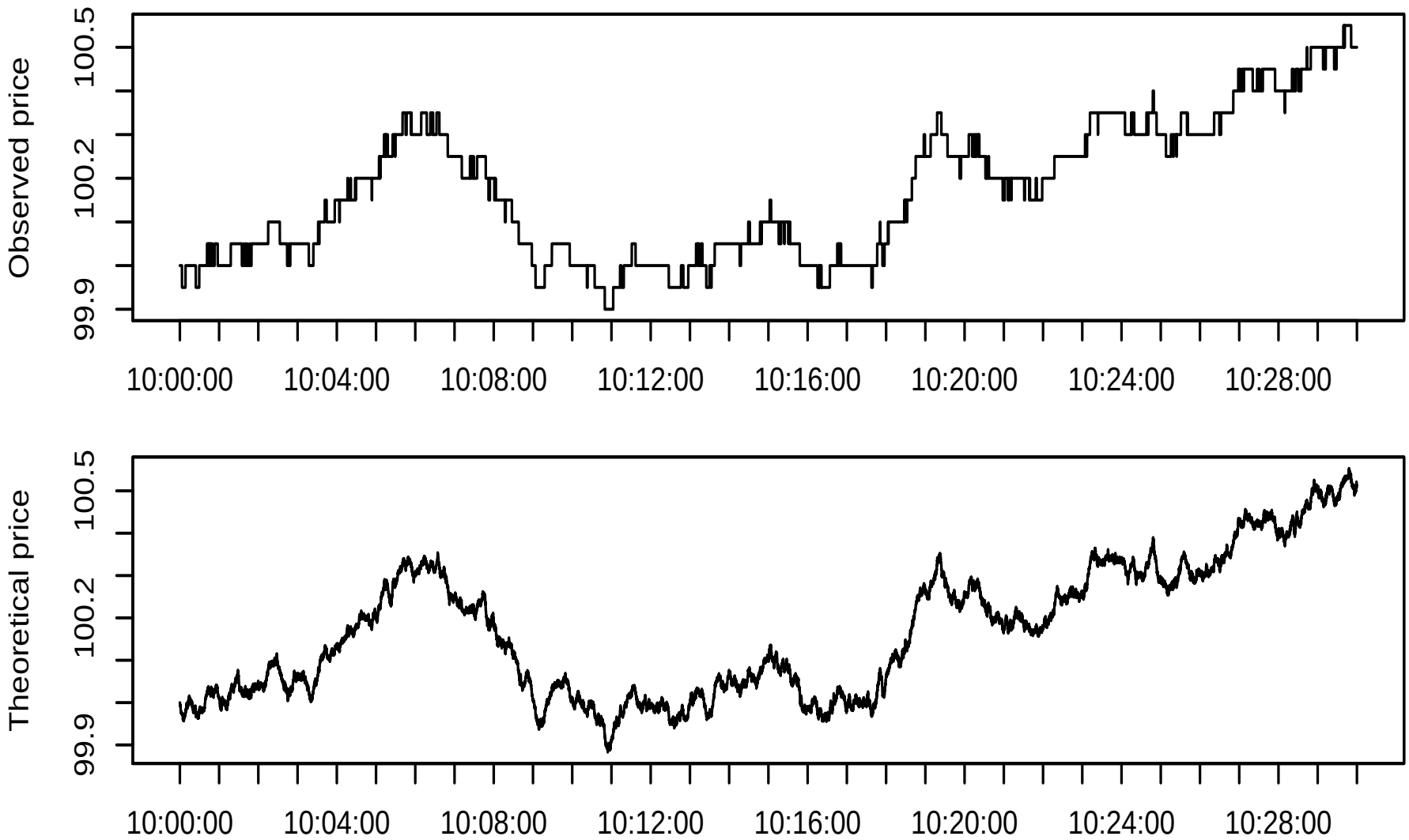
Interprétation de η

- Quantifie l'aversion aux changements de prix (par rapport à la taille du tick) des participants de marché.
- Un η petit ($< 1/2$) signifie que pour les participants de marché, la taille du tick est trop grande et inversement.
- Dans les UHF, le carnet d'ordre ne peut pas suivre le prix efficient et est réticent aux changements de prix. Réticence mesurée par η.
- η comme une mesure de la profondeur de prix moyenne explorée par le volume des transactions.

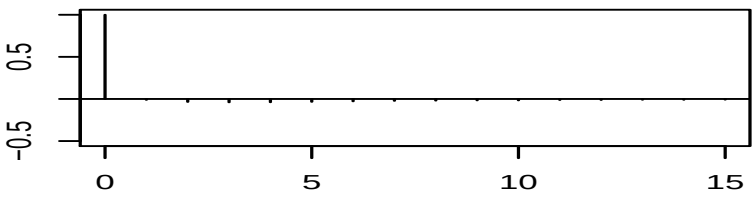
Propriétés : Durations



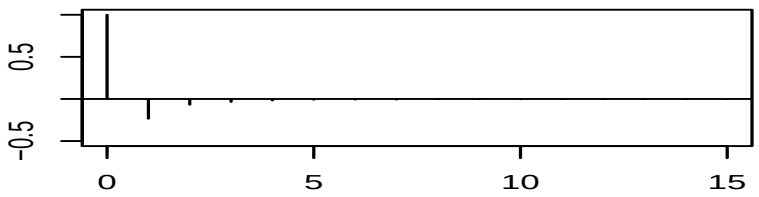
Propriétés : Le prix



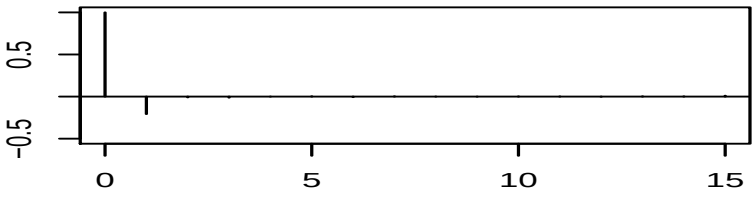
Propriétés : ACF log rendements ($\eta < 1/2$)



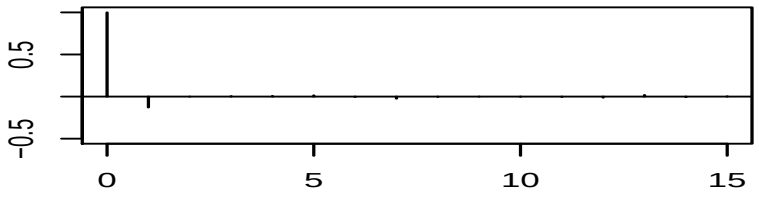
0.1 second sampling period



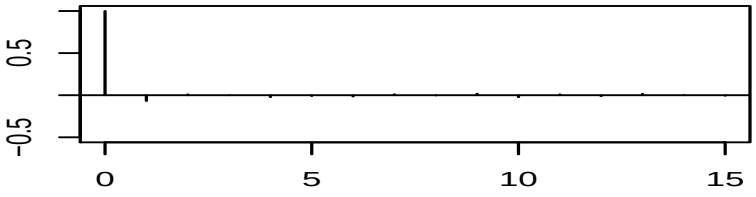
1 second sampling period



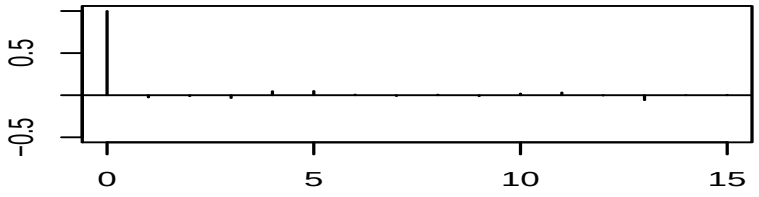
10 seconds sampling period



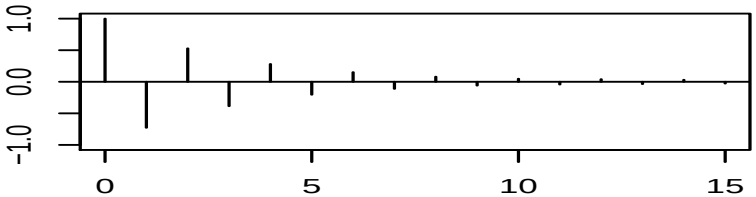
30 seconds sampling period



1 minute sampling period



5 minutes sampling period



Tick time sampling, unit = 1 price change

Tick time sampling, unit = 5 price changes

Outline

1 Données haute fréquence et bruit de microstructure

2 Modèles de microstructure et estimation de la volatilité

3 Modèle avec zones d'incertitude

4 Estimation de la volatilité

5 Estimation de la co-volatilité

Reconstruction du prix efficient

Prix efficient

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

t_i : date d'observation, τ_i : temps de sortie,
P_t : prix observé, X_t : prix efficient.

Estimation de η

Soit $N_{\alpha,t} = \text{card}\{t_i, t_i \leq t\}$. Pour $k = 1, ..., m$,

$$N^{(c)}_{\alpha,t,k} = \sum_{i=1}^{N_{\alpha,t}} \mathbb{I}_{\{|X_{\tau_i} - X_{\tau_{i-1}}| = \alpha k\}},$$

$$N^{(a)}_{\alpha,t,k} = \sum_{i=1}^{N_{\alpha,t}} \mathbb{I}_{\{|X_{\tau_i} - X_{\tau_{i-1}}| = \alpha(k-1+2\eta)\}}.$$

On définit

$$\hat{\eta}_t = \sum_{k=1}^{m} \lambda_{\alpha,t,k} u_{\alpha,t,k},$$

avec

$$\lambda_{\alpha,t,k} = \frac{N^{(a)}_{\alpha,t,k} + N^{(c)}_{\alpha,t,k}}{\sum_{j=1}^{m} \left[N^{(a)}_{\alpha,t,j} + N^{(c)}_{\alpha,t,j}\right]} \text{ et } u_{\alpha,t,k} = \frac{1}{2}\left(k\left(\frac{N^{(c)}_{\alpha,t,k}}{N^{(a)}_{\alpha,t,k}} - 1\right) + 1\right).$$

Estimation dans le modèle avec zones d'incertitude

Théorème

Soit

$$\widehat{RV}_t = \sum_{i=1}^{N_{\alpha,t}} \left(\log(\hat{X}_{\tau_i}) - \log(\hat{X}_{\tau_{i-1}}) \right)^2.$$

On a

$$\alpha^{-1}(\widehat{RV}_t - RV_t) \xrightarrow{\mathcal{L}s} \gamma_t \int_0^t v_u \mathrm{d}W_{\theta_u},$$

où W est un mouvement brownien indépendant de B et θ_u, γ_u et v_u dépendent de X_u, σ_u et de variables explicatives, liées par exemple au carnet d'ordre.

Discussion

Commentaires sur le théorème

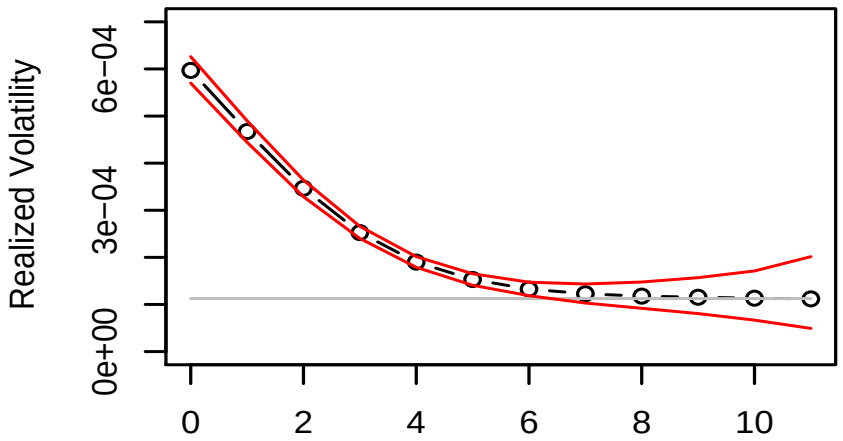
- a_u est seulement progressivement mesurable et localement borné. σ_u est seulement càdlàg adapté. En particulier, σ_u n'est pas nécessairement une semi-martingale d'Ito.
- L'arrondi a lieu sur l'échelle des prix et non des log-prix. Estimer la variation quadratique du log-prix est donc plus difficile qu'estimer celle du prix.
- Les dates d'observation sont aléatoires, endogènes. Ainsi, les théorèmes classiques pour échantillonnage déterministe ne s'appliquent pas.
- L'idée clé est de se placer dans un temps modifié dans lequel les observations sont équidistantes puis d'utiliser des propriétés de stabilité de la convergence dans $\mathbb{D}[0, T]$.

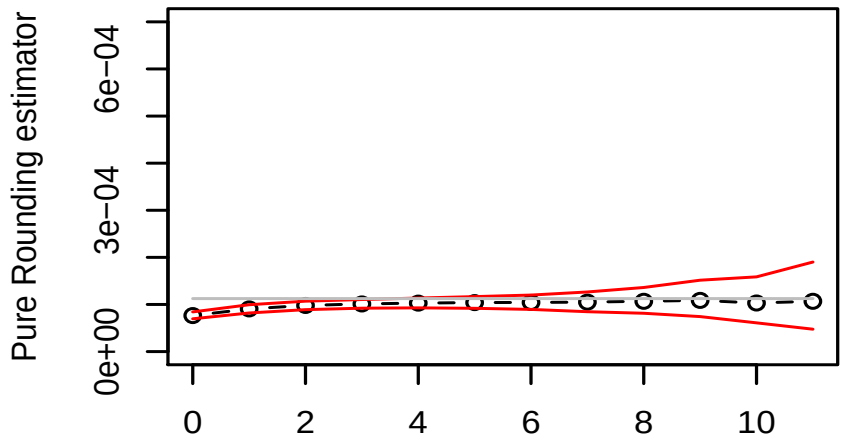
Estimation de η sur données equity

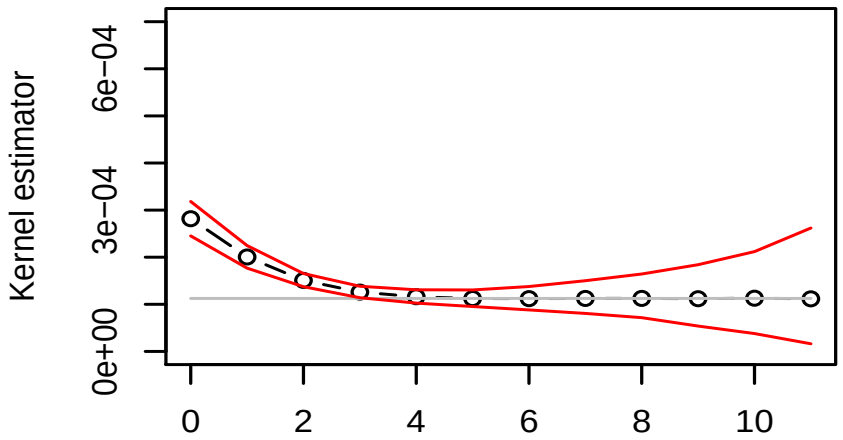
Semaine : du 15-01-2007 au 19-01-2007.

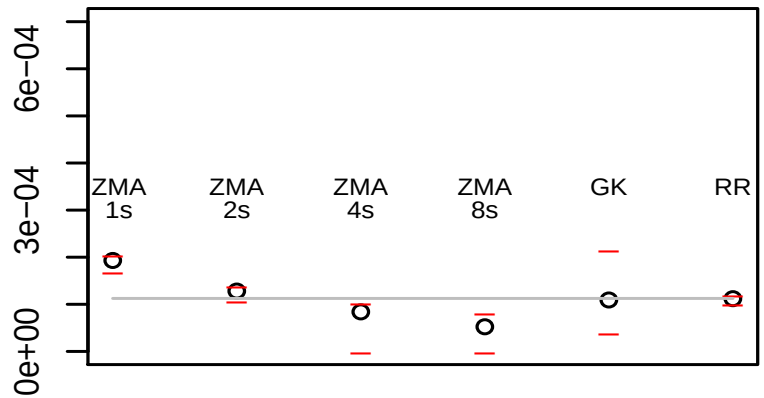
TICKER	$\hat{\eta}$ Jour 1	$\hat{\eta}$ Jour 2	$\hat{\eta}$ Jour 3	$\hat{\eta}$ Jour 4	$\hat{\eta}$ Jour 5
AIRF	0.277	0.273	0.250	0.236	0.250
ALSO	0.189	0.215	0.199	0.199	0.194
BNPP	0.100	0.111	0.146	0.157	0.110
CAGR	0.193	0.242	0.238	0.215	0.209
DANO	0.076	0.110	0.086	0.081	0.110
EAD	0.179	0.243	0.256	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

1000 simulations MC, intervalle de confiance à 90%









Outline

1 Données haute fréquence et bruit de microstructure

2 Modèles de microstructure et estimation de la volatilité

3 Modèle avec zones d'incertitude

4 Estimation de la volatilité

5 Estimation de la co-volatilité

Cadre

On considère maintenant 2 actifs.

Dynamique des log-prix efficients

$$\mathrm{d}\log X_t = \mu_t^X \mathrm{d}t + \sigma_{t-}^X \mathrm{d}W_t,$$

$$\mathrm{d}\log Y_t = \mu_t^Y \mathrm{d}t + \sigma_{t-}^Y \mathrm{d}B_t,$$

avec

$$\mathrm{d}\langle W, B\rangle_t = \rho_t \mathrm{d}t.$$

Quantité d'intérêt

Covariance intégrée

On souhaite estimer

$$\int_0^1 \rho_t \sigma_t^X \sigma_t^Y \mathrm{d}t.$$

Difficultés

Problèmes pratiques

2 difficultés principales :

- Données asynchrones,
- Bruit de microstructure.

Estimateur habituel de la covariance

Covariation réalisée

On observe $(X_{i/n}, Y_{i/n})$, $i = 0, \ldots, n$. Soit

$$\Delta_i^n X = \log X_{i/n} - \log X_{(i-1)/n}.$$

Un estimateur de

$$\int_0^1 \rho_t \sigma_t^X \sigma_t^Y \mathrm{d}t$$

avec précision $n^{-1/2}$ est

$$\sum_{i=1}^{n} \Delta_i^n X \Delta_i^n Y.$$

Méthode de Dacarogna

Interpolation Previous-Tick

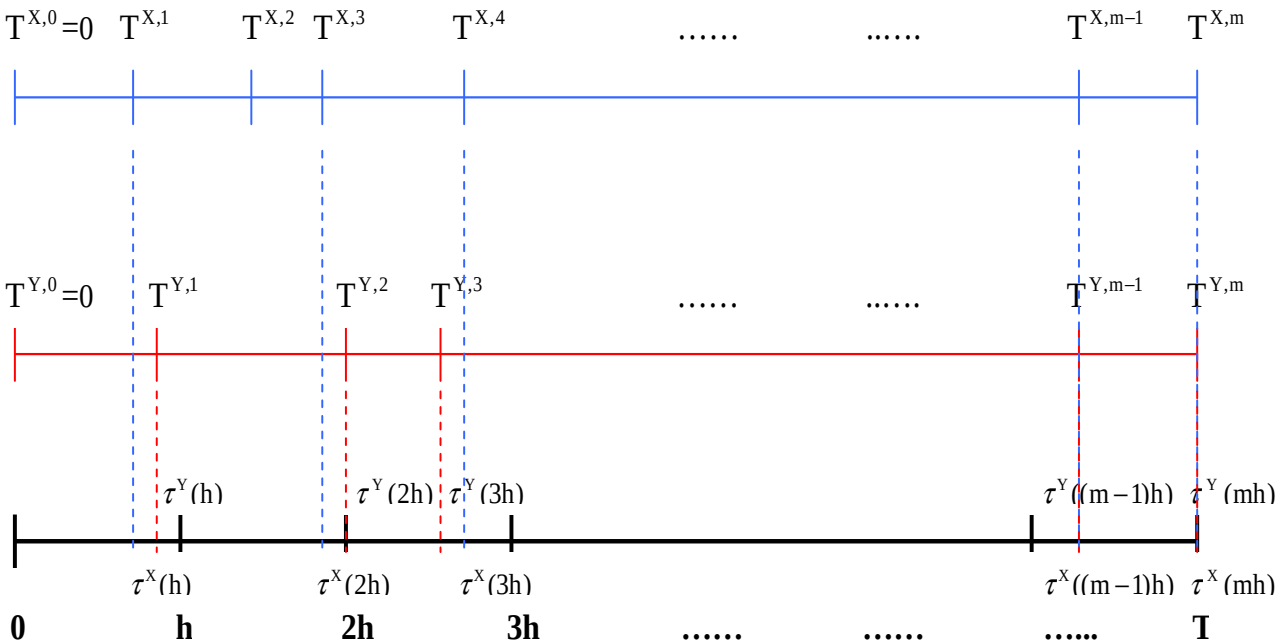
- Supposons maintenant que l'on observe X aux dates $(T^{X,i}), i = 1, \ldots$ et Y aux dates $(T^{Y,i}), i = 1, \ldots$
- On construit

$$\overline{X}_t = X_{T^{X,i}} \text{ for } t \in [T^{X,i}, T^{X,i+1})$$

et

$$\overline{Y}_t = Y_{T^{Y,i}} \text{ for } t \in [T^{Y,i}, T^{Y,i+1}).$$

Estimateur Previous-Tick



Pour h donnée, l'estimateur Previous-Tick de la covariance est

$$V_h = \sum_{i=1}^{m} \left(\log \overline{X}_{ih} - \log \overline{X}_{(i-1)h}\right)\left(\log \overline{Y}_{ih} - \log \overline{Y}_{(i-1)h}\right).$$

Inconvénient de cet estimateur

Effet Epps

- Biais systematique pour cet estimateur.
- Exemple : Supposons que $\log X$ et $\log Y$ sont 2 mouvements browniens de corrélation ρ et que les dates de trading sont les temps d'arrivée de 2 processus de Poisson indépendants. Alors, on peut montrer que

$$\mathbb{E}[V_h] \longrightarrow 0, \text{ as } h \longrightarrow 0.$$

Un estimateur convergent pour données asynchrones

Estimateur de Hayashi et Yoshida

- Soit $I_i^X = (T^{X,i}, T^{X,i+1}]$ et $I_i^Y = (T^{Y,i}, T^{Y,i+1}]$
- l'estimateur Hayashi-Yoshida est

$$U_n = \sum_{i,j} \Delta X(I_i^X) \Delta Y(I_j^Y) \mathbf{1}_{\{I_i^X \cap I_j^Y \neq \varnothing\}}.$$

- Cet estimateur ne requiert pas de choix de h et est convergent si les temps d'arrivée sont indépendants du prix et les volatilités déterministes.
- Cependant, il n'est pas robuste au bruit de microstructure (ce qui entraine aussi l'effet Epps).

Estimation dans le modèle avec zones d'incertitude

Supposons $\tau_1^X \leq \tau_1^Y$. On définit de nouvelles séquences de temps d'arrêt par $\lambda_0^X = \lambda_0^Y = 0$ et

$$\lambda_{i+1}^{(X)} = \min\left\{\tau_j^{(X)} : \tau_j^{(X)} \geq \lambda_i^{(Y)}\right\}, \quad \lambda_{i+1}^{(Y)} = \min\left\{\tau_j^{(Y)} : \tau_j^{(Y)} > \lambda_{i+1}^{(X)}\right\}.$$

Théorème

En se basant sur les temps d'observation définis ci-dessus, l'estimateur de Hayashi et Yoshida pour la co-volatilité est convergent si l'on utilise les valeurs estimées du prix efficient.

Monte Carlo

- 1000 simulations,
- Modèle de Black-Scholes avec volatilité en U,
- Covariation constante $= 4.5 \times 10^{-5}$ ($\rho = 0.4$).

	Hayashi-Yoshida	Nouvel estimateur
Moyenne	8.55×10^{-5}	4.48×10^{-5}
Ecart type	7.8×10^{-6}	5.7×10^{-6}