On the asymptotic structure of Brownian motions with a small lead-lag effect

Yuta Koike

Tokyo Metropolitan University

CREST JST

July 20, 2017

Outline

- Motivation
- Model
- Main results
- Efficient estimation of the lag parameter
- Conclusions and future work

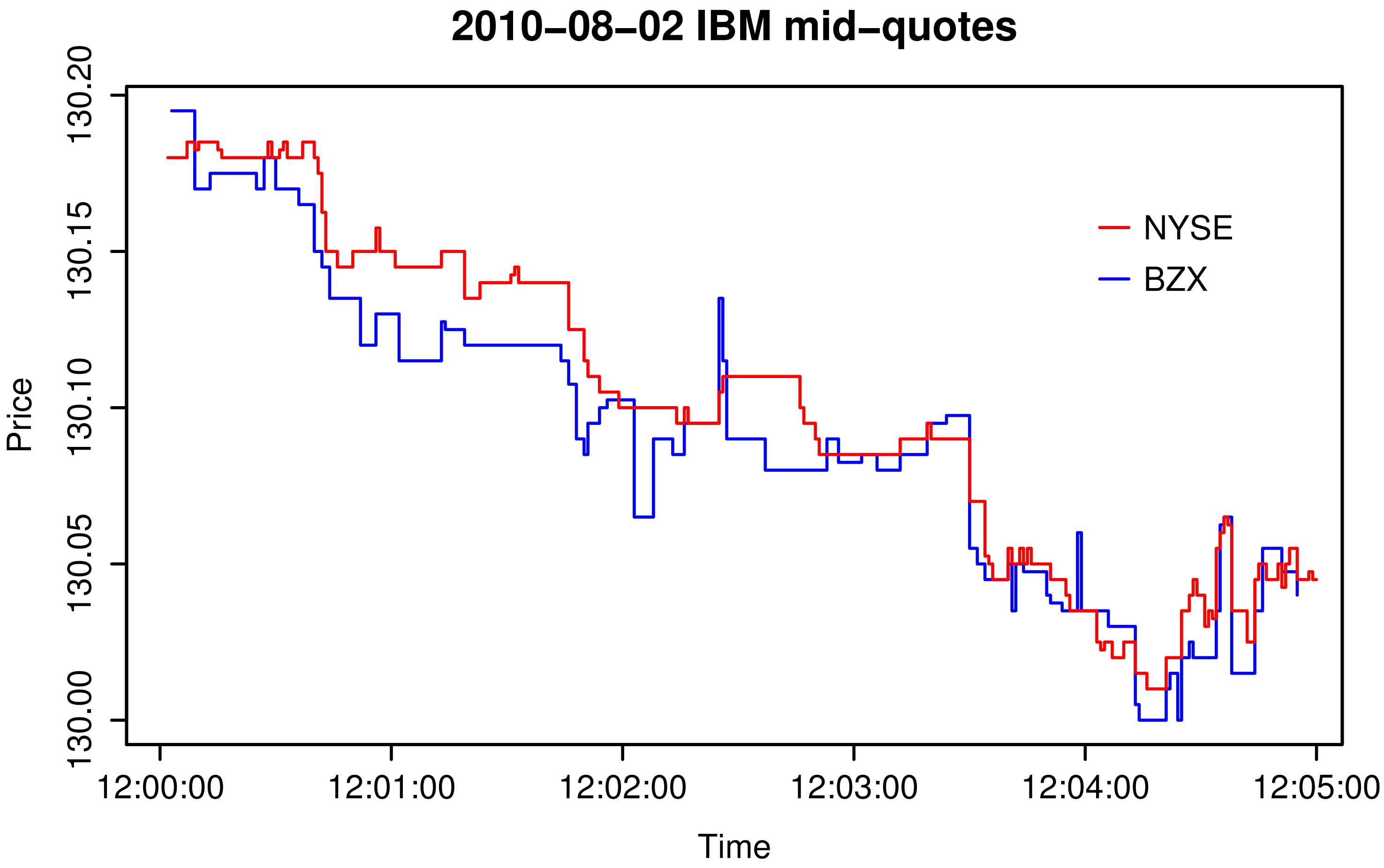
Motivation

- **Lead-lag effect**
  - Two time series are cross-correlated with each other at certain lags; “leader” and “lagger”
- In financial markets, lead-lag effects may occur perhaps because new information is absorbed into each security at different speeds
  - Across different assets
  - Across different trading venues

Motivation

- **Ex1**: Stock index vs index futures (e.g. Kawaller *et al.*, 1987)
  - A stock index consists of many individual stocks; it may be lagging behind the index futures
- **Ex2**: Large stocks vs smaller stocks (e.g. Lo and MacKinlay, 1990)
  - Large stocks are traded more frequently than smaller stocks, so the former may absorb new information faster than the latter

Ex3: One security traded at multiple venues



- Due to the low-latency trading of the recent financial markets, lead-lag effects only appear in a very short period (Tóth and Kertész, 2006)

  $\Longrightarrow$ We need to utilize high-frequency data

- Hoffmann, Rosenbaum and Yoshida (2013) have proposed a model for lead-lag effects in high-frequency financial data ("HRY model")
  - Its practicality in empirical work has recently been established by several authors such as Alsayed and McGroarty (2014), Huth and Abergel (2014), Bollen *et al.* (2017) and Iacus *et al.* (2015)

- These empirical studies show that time lag parameters are typically comparable to the observation frequencies in their scales

- This motivates us to study the HRY model with a “small” lead-lag effect

  - In particular, we are interested in how small lag parameters can be identified in principle

- However, there is few theoretical study for the HRY model and, in particular, nothing has been known about the optimality of statistical inferences for the HRY model

- $\boxed{\text{Aim of this study}}$ fill in this gap

Model

- $B_t = (B_t^1, B_t^2)$ $(t \in \mathbb{R})$: bivariate two-sided Brownian motion (latent process)
  - $B_0 = 0$, $E[(B_1^1)^2] = E[(B_1^2)^2] = 1$ and $E[B_1^1 B_1^2] = \rho \neq 0$
  - $\rho \in (-1, 1)$: correlation parameter
- $\epsilon_1^k, \epsilon_2^k, \ldots$ $(k = 1, 2)$: innovations of observation noise
  - $\epsilon^k \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$
  - ϵ^1 and ϵ^2 are mutually independent

Model

- Observation data $\mathbf{Z}_n = (X_1, \ldots, X_n, Y_1, \ldots, Y_n)^\top$

$$\begin{cases} X_i = B^1_{i/n} + \sqrt{v_n}\epsilon^1_i, & Y_i = B^2_{i/n-\vartheta} + \sqrt{v_n}\epsilon^2_i \quad \text{if } \vartheta \geq 0, \\ X_i = B^1_{i/n-|\vartheta|} + \sqrt{v_n}\epsilon^1_i, & Y_i = B^2_{i/n} + \sqrt{v_n}\epsilon^2_i \quad \text{if } \vartheta < 0 \end{cases}$$

  - $\vartheta \in \mathbb{R}$: time-lag parameter
  - $v_n \geq 0$: variance of the observation noise (assume $\limsup_n v_n < \infty$)

- The case $v_n \equiv 0$ corresponds to (a special case of) the HRY model

Model

- We also consider a noisy observation case due to its importance in high-frequency financial modeling ("market microstructure noise")
- We denote by $\mathbb{P}_{n,\vartheta}$ the distribution of $\mathbf{Z}_n$
  - $\mathbb{P}_{n,\vartheta}$ is defined on $(\mathbb{R}^{2n}, \mathcal{B}^{2n})$
  - $\mathcal{B}^{2n}$ denotes the Borel σ-field of $\mathbb{R}^{2n}$

Model

- $\boxed{\text{Aim of this study}}$ Investigating the asymptotic structure of the experiments $\mathbb{P}_{n,\vartheta}$ when the lag parameter ϑ is small (i.e. when $\vartheta \to 0$ as $n \to \infty$)
- More precisely, for a sequence ϑ_n tending to 0 with the "appropriate" rate, we study the asymptotic behavior of the likelihood ratios $\mathrm{d}\mathbb{P}_{n,\vartheta_n}/\mathrm{d}\mathbb{P}_{n,0}$
- This serves as investigating the asymptotic efficiency of the statistical inference for the lag parameter ϑ when it is small

Main results

- We assume that the existence of the limit

$$\gamma = \lim_{n\to\infty} nv_n \in [0, \infty]$$

- Since $E[(B^k_{i/n} - B^k_{(i-1)/n})^2] = 1/n$,
  - $\gamma = \infty \Longrightarrow$ The observation noise (locally) dominates the latent process
  - $0 < \gamma < \infty \Longrightarrow$ The latent process and the observation noise are balanced
  - $\gamma = 0 \Longrightarrow$ The latent process dominates the observation noise

Main results

- We define the sequence N_n of positive numbers by

$$N_n = \begin{cases} \sqrt{n/v_n} & \text{if } \gamma = \infty, \\ n & \text{otherwise} \end{cases}$$

- N_n could be regarded as an "effective" sample size in the following sense:

Main results

- For a standard Brownian motion $B = (B_t)_{t\in[0,1]}$, the optimal convergence rate to estimate the scale parameter $\sigma > 0$ from the observation data

$$\sigma B_{i/n} + \sqrt{v_n}\epsilon_i, \qquad \epsilon_i \overset{i.i.d.}{\sim} N(0,1) \perp\!\!\!\perp B$$

  is given by $N_n^{-1/2}$ (Gloter and Jacod, 2001)

- The rate $N_n^{-1/2}$ can be seen as the regular parametric rate if we regard N_n as the sample size

Theorem 1

Set $r_n = N_n^{-3/2}$. There are random variables T_n and S_n defined on $(\mathbb{R}^{2n}, \mathcal{B}^{2n})$ and two numbers $I_\gamma > 0$ and $J_\gamma \geq 0$ such that

$$\log \frac{\mathrm{d}\mathbb{P}_{n,r_n u_n}}{\mathrm{d}\mathbb{P}_{n,0}} - \left\{ u_n \mathsf{T}_n + |u_n| \mathsf{S}_n - \frac{u_n^2}{2}(I_\gamma + J_\gamma) \right\} \xrightarrow{p} 0$$

$$\text{under } \mathbb{P}_{n,0} \quad \text{as } n \to \infty$$

for any bounded sequence u_n of real numbers and

$$(\mathsf{T}_n, \mathsf{S}_n) \xrightarrow{d} \mathcal{N}(0, I_\gamma) \otimes \mathcal{N}(0, J_\gamma) \qquad \text{under } \mathbb{P}_{n,0} \quad \text{as } n \to \infty.$$

□ I_γ and J_γ are defined as follows:

$$I_\gamma = \begin{cases} \frac{\rho^2}{1-\rho^2} & \text{if } \gamma = 0, \\ \frac{\rho\left(\sqrt{(1+\rho)(1+\rho+4\gamma)} - \sqrt{(1-\rho)(1-\rho+4\gamma)}\right)}{4\gamma^2} & \text{if } 0 < \gamma < \infty, \\ \frac{\rho^2}{\sqrt{1+\rho} + \sqrt{1-\rho}} & \text{if } \gamma = \infty, \end{cases}$$

$$J_\gamma = \frac{\rho^2}{8}\{J_\gamma^0(1+\rho) + J_\gamma^0(1-\rho)\}$$

Here, for every $a > 0$ we set

$$J_\gamma^0(a) = \begin{cases} \frac{3}{2a^2} & \text{if } \gamma = 0, \\ \frac{1}{8\gamma^2}\left(2 - 3\left(\frac{a}{a+4\gamma}\right)^{1/2} + \left(\frac{a}{a+4\gamma}\right)^{3/2}\right) & \text{if } 0 < \gamma < \infty, \\ 0 & \text{if } \gamma = \infty. \end{cases}$$

- From Theorem 1, when $\gamma = \infty$ (the case that the observation noise is locally dominant), $(\mathbb{P}_{n,\vartheta})_{\vartheta \in \mathbb{R}}$ enjoys the LAN property at $\vartheta = 0$

- Otherwise, our experiment exhibits an asymptotic structure different from the LAN

- This is a typical phenomenon for irregular models, but our situation is different from the “common” irregular models (cf. Chapters 5–7 of Ibragimov and Has’minskii (1981)) in the following sense:
  - The rate $N_n^{-3/2}$ is faster than the “common” rate N_n^{-1}
  - The asymptotic structure seems different from those of any other known irregular models

Efficient estimation of the lag parameter

- We construct efficient estimators for the lag parameter at $\vartheta = 0$ in the case of $\gamma < \infty$
- As stated above, in this case our model does not have the LAN property, so it is not obvious how to define the asymptotically efficient estimators
- Here, following Küchler and Kutoyants (2000), we define the asymptotic efficiency by utilizing the minimax inequality obtained from Theorem I-9.1 of Ibragimov and Has'minskii (1981):

Ibragimov and Has'minskii (1981), Theorem I-9.1

Let Θ be a compact subset of $\mathbb{R}^k$. For each n, let $(\mathbf{P}_{n,\theta})_{\theta\in\Theta}$ be a family of probability distributions dominated by a σ-finite measure. Also, let r_n be a sequence of positive numbers tending to 0 and $L : \Theta \to [0, \infty)$ be a continuous function. Suppose that for any interior point u of Θ and any prior density q on Θ such that $q(u) > 0$, the posterior mean $\tilde{\theta}_n$ with respect to q satisfies

$$\lim_{n\to\infty} \mathbf{E}_{n,u}[|r_n^{-1}(\tilde{\theta}_n - u)|^2] = L(u).$$

Then, for any open set $U \subset \Theta$ and any estimator sequence $\hat{\theta}_n$,

$$\liminf_n \sup_{\theta\in U} \mathbf{E}_{n,\theta}[|r_n^{-1}(\hat{\theta}_n - \theta)|^2] \geq \sup_{u\in U} L(u).$$

Efficient estimation of the lag parameter

- Taking a sequence η_n of positive numbers tending to 0, we apply the above theorem with $\mathbf{P}_{n,\theta} = \mathbb{P}_{n,\theta\eta_n}$ and $\Theta = [-1, 1]$
- We consider a (local) Bayes estimator: taking a prior density q_n on $[-\eta_n, \eta_n]$, we define the Bayes estimator $\tilde{\vartheta}_n$ by

$$\tilde{\vartheta}_n = \int_{-\eta_n}^{\eta_n} \vartheta \frac{\mathrm{d}\mathbb{P}_{n,\vartheta}}{\mathrm{d}\mathbb{P}_{n,0}} q_n(\vartheta)\mathrm{d}\vartheta \Bigg/ \int_{-\eta_n}^{\eta_n} \frac{\mathrm{d}\mathbb{P}_{n,\vartheta}}{\mathrm{d}\mathbb{P}_{n,0}} q_n(\vartheta)\mathrm{d}\vartheta$$

Theorem 2

Suppose that $\eta_n = o(n^{-1})$ and $n^{3/2}\eta_n \to \infty$. Also, suppose that there is a continuous function $q : [-1, 1] \to (0, \infty)$ such that $q_n(\vartheta) = q(\vartheta/\eta_n)$ for $\vartheta \in [-\eta_n, \eta_n]$. Then, there is a random variable $\tilde{u}$ such that $n^{3/2}(\tilde{\vartheta}_n - \vartheta_n)$ converges in distribution to $\tilde{u}$ under $\mathbb{P}_{n,\vartheta_n}$ for any sequence $\vartheta_n \in (-\eta_n, \eta_n)$. Moreover, $\tilde{\vartheta}_n$ is asymptotically efficient at $\vartheta = 0$ in the following sense: for any estimator sequence $\hat{\vartheta}_n$, it holds that

$$\lim_{\delta\to 0} \liminf_n \sup_{|\vartheta|<\delta\eta_n} \mathbb{E}_{n,\vartheta}[n^3|\hat{\vartheta}_n - \vartheta|^2] \geq E[\tilde{u}^2].$$

Efficient estimation of the lag parameter

- $\tilde{u}$ can be explicitly defined as follows:

$$\tilde{u} = \frac{\int_{-\infty}^{\infty} uZ(u)\mathrm{d}u}{\int_{-\infty}^{\infty} Z(u)\mathrm{d}u},$$

where

$$Z(u) = \exp\left(u\zeta_1 + |u|\zeta_2 - \frac{u^2}{2}(I_\gamma + J_\gamma)\right), \qquad u \in \mathbb{R},$$

with ζ_1 and ζ_2 being two mutually independent variables such that $\zeta_1 \sim \mathcal{N}(0, I_\gamma)$ and $\zeta_2 \sim \mathcal{N}(0, J_\gamma)$

Efficient estimation of the lag parameter

- Another natural candidate of the estimators is the MLE
- We define the (local) MLE as

$$\hat{\vartheta}_n = \operatorname*{arg\,max}_{\vartheta \in (-\eta_n, \eta_n)} \frac{\mathrm{d}\mathbb{P}_{n,\vartheta}}{\mathrm{d}\mathbb{P}_{n,0}}$$

- The limiting variable of the MLE can be explicitly written as follows:

$$\hat{u} = \operatorname*{arg\,max}_{u \in \mathbb{R}} Z(u) = \begin{cases} (\zeta_1 + \zeta_2)/(I_\gamma + J_\gamma) & \text{if } \zeta_1 \geq (-\zeta_2) \vee 0, \\ (\zeta_1 - \zeta_2)/(I_\gamma + J_\gamma) & \text{if } \zeta_1 < \zeta_2 \wedge 0, \\ 0 & \text{otherwise} \end{cases}$$

Efficient estimation of the lag parameter

Theorem 3

Suppose that $\eta_n = o(n^{-1})$ and $n^{3/2}\eta_n \to \infty$. Then, $n^{3/2}(\hat{\vartheta}_n - \vartheta_n)$ converges in distribution to $\hat{u}$ under $\mathbb{P}_{n,\vartheta_n}$ for any sequence $\vartheta_n \in (-\eta_n, \eta_n)$.

- We naturally ask the following question: *How is the MLE inefficient compared with the Bayes estimators?*
- To answer this question, we compute their asymptotic variances

Theorem 4

It holds that

$$E\left[\hat{u}^2\right] = \frac{1}{I_\gamma + J_\gamma}\left(1 - \frac{1}{\pi}\arctan\left(\sqrt{\frac{J_\gamma}{I_\gamma}}\right) + \frac{\sqrt{I_\gamma J_\gamma}}{\pi(I_\gamma + J_\gamma)}\right),$$

$$E\left[\tilde{u}^2\right] = \frac{1}{I_\gamma + J_\gamma}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x\Psi(x) - y\Psi(y)}{\Psi(x) + \Psi(y)}\right)^2 \psi_R(x,y)\mathrm{d}x\mathrm{d}y,$$

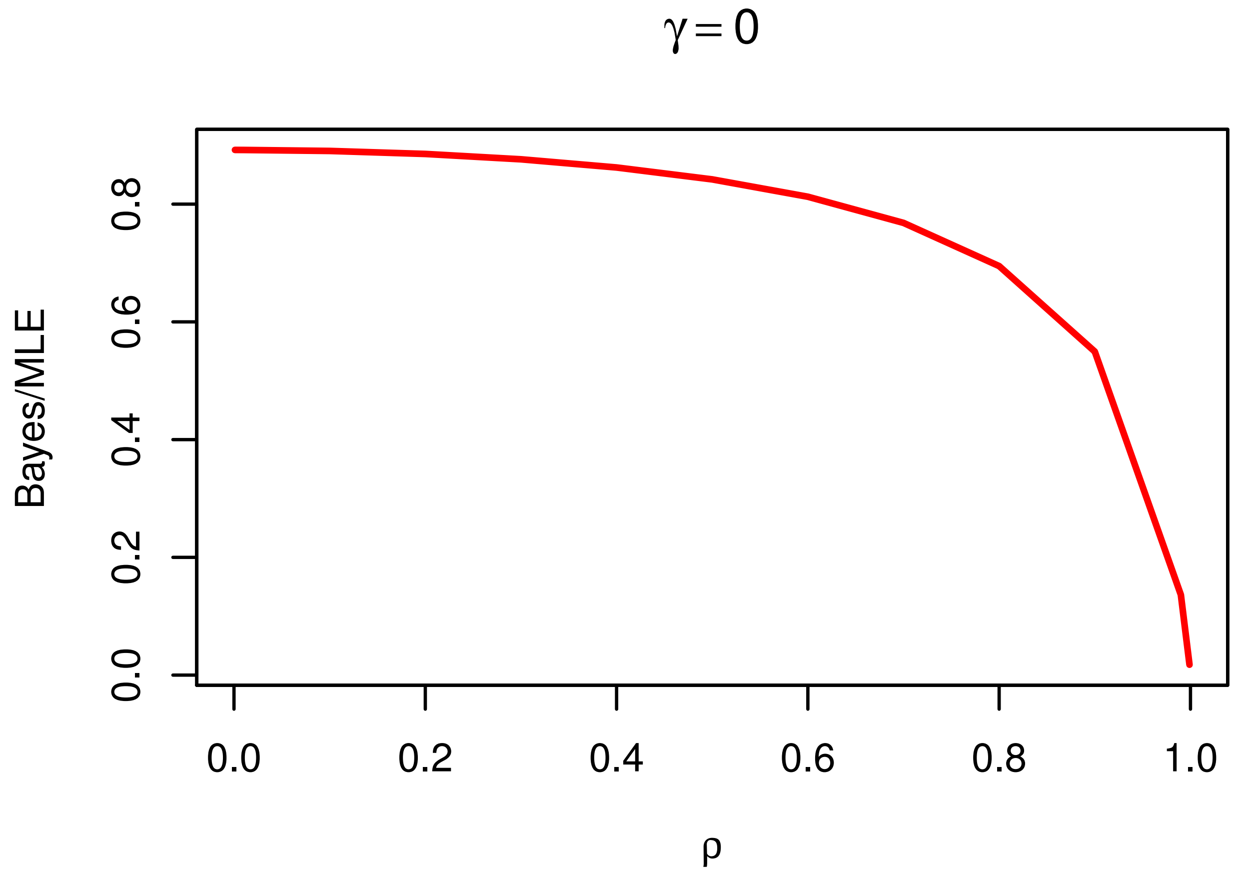
where $\Psi(x) = \int_0^\infty e^{ux-u^2/2}\mathrm{d}u$ and $\psi_R(x,y)$ denotes the bivariate normal density with standard normal marginals and correlation $R = (J_\gamma - I_\gamma)/(J_\gamma + I_\gamma)$.

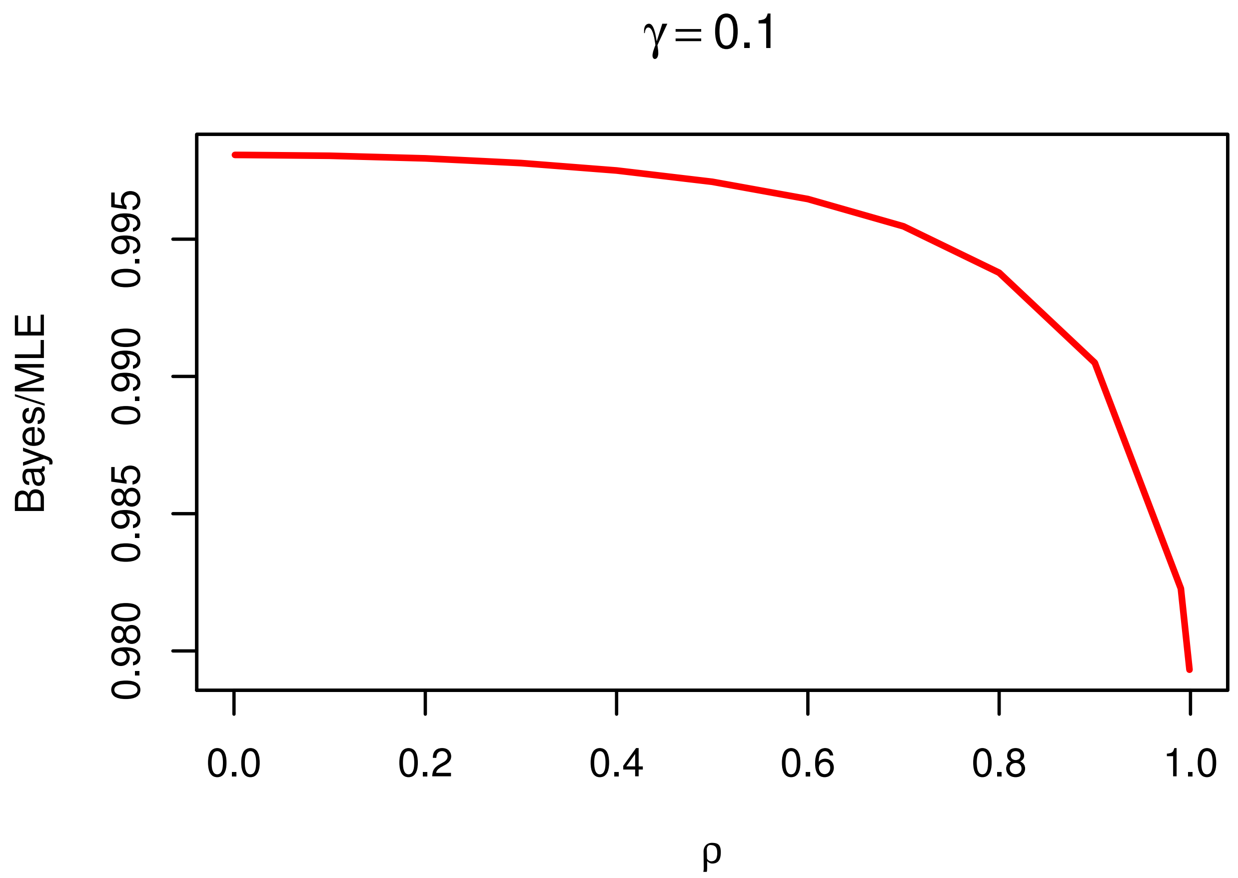
Efficient estimation of the lag parameter

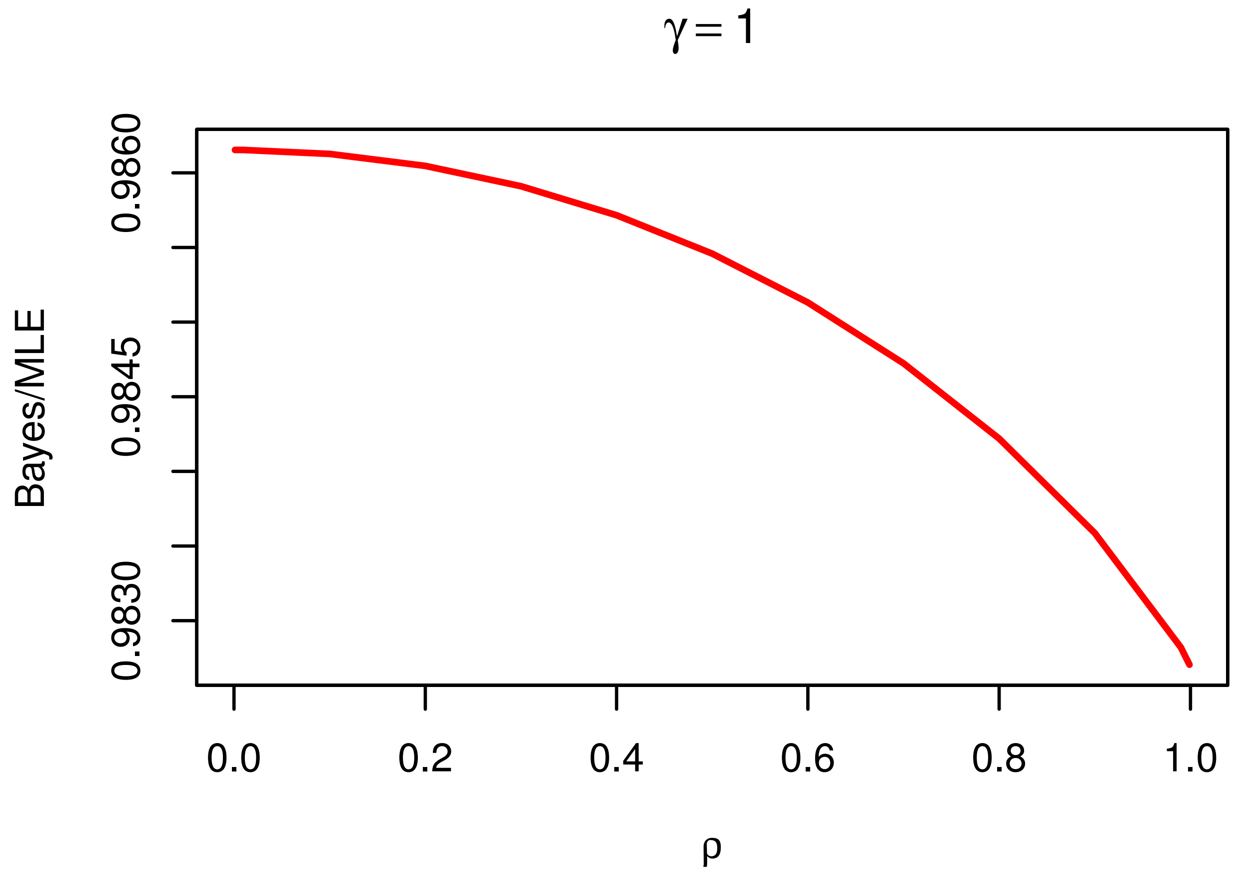
- Although we have analytic expressions of the asymptotic variances of the MLE and the Bayes estimators, it is not easy to quantify how different they are.
- However, if $\gamma = 0$, we can easily check

$$\lim_{|\rho|\to 1} E\left[\tilde{u}^2\right] / E\left[\hat{u}^2\right] = 0,$$

  hence the Bayes estimator can be *much* better than the MLE
- In the following we numerically evaluate the above expressions for $\rho = 0.001, 0.01, 0.1, 0.2, \ldots, 0.8, 0.9, 0.99, 0.999$ and $\gamma = 0, 0.1, 1$


γ = 0
Bayes/MLE
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
ρ


$\gamma = 0.1$
Bayes/MLE
0.980
0.985
0.990
0.995
0.0
0.2
0.4
0.6
0.8
1.0
ρ



$\gamma = 1$
0.9860
0.9845
0.9830
Bayes/MLE
0.0
0.2
0.4
0.6
0.8
1.0
ρ

Conclusions

- We have derived the asymptotic property of the likelihood ratio of Brownian motion model with a small lead-lag structure
  - If the observation noise is dominant, the LAN property holds true
  - Otherwise, a non-standard asymptotic structure appears
- We have shown that the Bayes estimators are asymptotically efficient by utilizing the Ibragimov-Has'minskii theory

Future work

- Non-small lead-lag times
  - When $\gamma < \infty$, the asymptotic structure seems to depend on the limit of the fractional part of $n\vartheta$
- More general situations including stochastic volatilities as well as irregular and non-synchronous sampling times
- Construct more realistic estimators
  - Non-local, estimate the time-lag, correlation and volatilities simultaneously

References

Alsayed, H. and McGroarty, F. (2014). Ultra-high-frequency algorithmic arbitrage across international index futures. *J. Forecast.* **33**, 391–408.

Bollen, N. P., O'Neill, M. J. and Whaley, R. E. (2017). Tail wags dog: Intraday price discovery in VIX markets. *Journal of Futures Markets* **37**, 431–451.

Chan, K. (1993). Imperfect information and cross-autocorrelation among stock prices. *Journal of Finance* **48**, 1211–1230.

Chordia, T., Sarkar, A. and Subrahmanyam, A. (2011). Liquidity dynamics and cross-autocorrelations. *Journal of Financial and Quantitative Analysis* **46**, 709–736.

Glosten, L. R. (1987). Components of the bid-ask spread and the statistical properties of transaction prices. *Journal of Finance* **42**, 1293–1307.

Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Stat.* **5**, 225–242.

Hoffmann, M., Rosenbaum, M. and Yoshida, N. (2013). Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli* **19**, 426–461.

Huth, N. and Abergel, F. (2014). High frequency lead/lag relationships — empirical facts. *Journal of Empirical Finance* **26**, 41–58.

Iacus, S. M., Porro, G., Salini, S. and Siletti, E. (2015). Social networks, happiness and health: from sentiment analysis to a multidimensional indicator of subjective well-being. Working paper, available at arXiv: http://arxiv.org/abs/1512.01569.

Ibragimov, I. and Has'minskii, R. (1981). *Statistical estimation: Asymptotic theory*. Springer.

Kawaller, I. G., Koch, P. D. and Koch, T. W. (1987). The temporal price relationship between S&P 500 futures and the S&P 500 index. *Journal of Finance* **42**, 1309–1329.

Küchler, U. and Kutoyants, Y. A. (2000). Delay estimation for some stationary diffusion-type processes. *Scand. J. Stat.* **27**, 405–414.

Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*. Springer.

Lo, A. W. and MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies* **3**, 175–205.

Tóth, B. and Kertész, J. (2006). Increasing market efficiency: Evolution of cross-correlations of stock returns. *Phys. A* **360**, 505–515.

Key idea of the proof

- Although our model is Gaussian, it is not easy to directly analyze the model because its covariance matrix is a complicated function of ϑ
- For this reason we introduce an auxiliary model which is more tractable than the original model
- For each $n \in \mathbb{N}$, set

$$\begin{aligned}\Theta_n &= \{\vartheta \in \mathbb{R} : v_n - n\vartheta^2 + |\vartheta| \geq 0\} \\ &= \{\vartheta \in \mathbb{R} : |\vartheta| \leq (1 + \sqrt{1 + 4nv_n})/(2n)\}\end{aligned}$$

Key idea of the proof

- For each $\vartheta \in \Theta_n$ we denote by $\widetilde{\mathbb{P}}_{n,\vartheta}$ the law of the vector $\widetilde{\mathbf{Z}}_n = (\widetilde{X}_1, \dots, \widetilde{X}_n, \widetilde{Y}_1, \dots, \widetilde{Y}_n)^\top$ defined by

$$\widetilde{X}_i = B^1_{i/n} + \widetilde{\epsilon}^1_i, \qquad \widetilde{Y}_i = B^2_{i/n} + \widetilde{\epsilon}^2_i,$$

where

$$\begin{cases} \widetilde{\epsilon}^1_{i,n} = \epsilon^1_i, \qquad \widetilde{\epsilon}^2_{i,n} = -n\vartheta(B^2_{i/n} - B^2_{(i-1)/n}) + \sqrt{v_n - n\vartheta^2 + \vartheta}\epsilon^2_i & \text{if } \vartheta \geq 0, \\ \widetilde{\epsilon}^1_{i,n} = -n|\vartheta|(B^1_{i/n} - B^1_{(i-1)/n}) + \sqrt{v_n - n\vartheta^2 + |\vartheta|}\epsilon^1_i, \qquad \widetilde{\epsilon}^2_{i,n} = \epsilon^2_i & \text{if } \vartheta < 0 \end{cases}$$

for $i = 1, \dots, n$

Key idea of the proof

- We can show that the model $\mathbb{P}_{n,\vartheta}$ is well-approximated by $\widetilde{\mathbb{P}}_{n,\vartheta}$ in the Hellinger distance

- The Hellinger distance $H(P, Q)$ between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$ is defined by

$$H(P, Q) = \left(\int_{\mathcal{X}} \left(\sqrt{\frac{\mathrm{d}P}{\mathrm{d}\mu}} - \sqrt{\frac{\mathrm{d}Q}{\mathrm{d}\mu}} \right)^2 \mathrm{d}\mu \right)^{1/2},$$

where μ is a σ-finite measure dominating both P and Q

Proposition 1

If a sequence ϑ_n of positive numbers satisfies $\vartheta_n = o(n^{-1} \vee N_n^{-\frac{4}{3}})$ as $n \to \infty$, then $\sup_{|\vartheta| \leq \vartheta_n} H(\mathbb{P}_{n,\vartheta}, \widetilde{\mathbb{P}}_{n,\vartheta}) \to 0$

Key idea of the proof

- Let us remark the following inequality:

$$||P - Q|| \leq H(P, Q)$$

  for two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$, where we set $||\mu|| = \sup_{f:|f|\leq 1} |\int f \mathrm{d}\mu|$ for a signed measure μ on $(\mathcal{X}, \mathcal{A})$

- The next result shows that the asymptotic structures of $\mathbb{P}_{n,\vartheta}$ and $\widetilde{\mathbb{P}}_{n,\vartheta}$ are identical for sufficiently small ϑ:

Le Cam (1986), Chapter 4, Proposition 2

Let (P_1, P_2) and (Q_1, Q_2) be two pairs of positive measures on the measurable space $(\mathcal{X}, \mathcal{A})$. Then

$$\int \left\{ 1 \wedge \left| \frac{\mathrm{d}P_2}{\mathrm{d}P_1} - \frac{\mathrm{d}Q_2}{\mathrm{d}Q_1} \right| \right\} \mathrm{d}(P_1 + Q_1)$$

$$\leq \|P_1 - Q_1\| + 2\|P_2 - Q_2\| + \sqrt{2\|P_1 - Q_1\|\|P_2 + Q_2\|}.$$

- From an econometric point of view, Proposition 1 is of independent interest since the auxiliary model $\widetilde{\mathbb{P}}_{n,\vartheta}$ has an economic interpretation different from the original model $\mathbb{P}_{n,\vartheta}$
- The model $\widetilde{\mathbb{P}}_{n,\vartheta}$ contains measurement errors correlated to the latent returns $B_{i/n} - B_{(i-1)/n}$
  - In the market microstructure theory, such a correlation is often explained as an effect of asymmetric information (e.g. Glosten, 1987)
  - Some economic arguments suggest that such an information asymmetry would cause a lead-lag effect; see Chan (1993) and Chordia *et al.* (2011) for instance