Parameter Estimation of Continuous Processes Using Financial High-Frequency Data

Vladimír Holý
University of Economics, Prague
Faculty of Informatics and Statistics
Department of Econometrics

Robust
Rybník
January 25, 2018

Financial High-Frequency Data

Financial High-Frequency Data

- Ultra-high-frequency data are irregularly spaced time series recorded at highest possible frequency corresponding to each transaction or change in bid/ask prices.
  - *Engle, R. F. (2000). The Econometrics of Ultra-High-Frequency Data. Econometrica, 68(1), 1–22.*
- Financial high-frequency time series include
  - exchange rates,
  - stock prices,
  - commodity prices.
- The price process is often modeled with continuous values and continuous time.
- However, there are some crucial market microstructure specifics such as
  - rounding error (prices have discrete values),
  - discreteness of price changes (transactions can occur only at discrete times),
  - bid-ask spread (transactions can happen either on bid or ask side),
  - informational effects (agents do not behave according to the economic theory).

Market Microstructure Noise

- Market microstructure specifics can be captured by the model with additive noise

$$X_i = P_{T_i} + E_i, \qquad i = 1, \ldots, n,$$

where

  - X_i is the observable price with discrete time,
  - P_{T_i} is the efficient price with continuous time sampled at discrete times T_i,
  - E_i is the market microstructure noise with discrete time.

- We assume the market microstructure noise to be independent white noise and

$$E_i \sim (0, \omega^2).$$

- Generally, the market microstructure noise can be dependent in time and dependent on efficient price.
  - *Hansen, P. R., & Lunde, A. (2006). Realized Variance and Market Microstructure Noise. Journal of Business & Economic Statistics, 24(2), 127–161.*

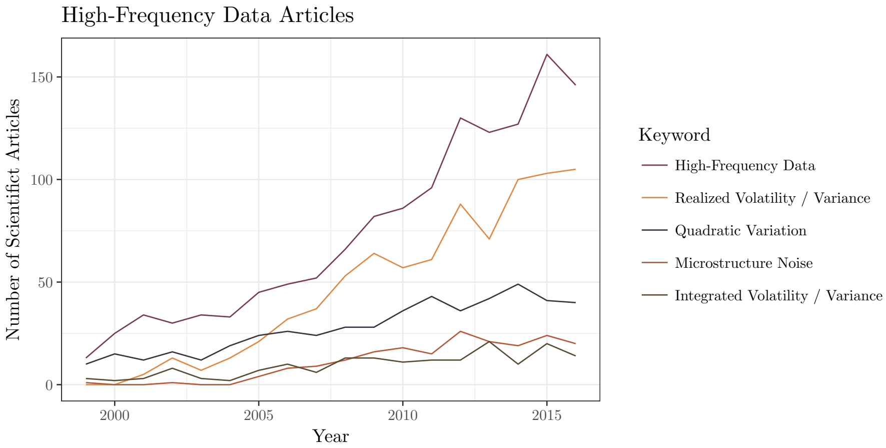
Non-Parametric Approach

- Estimation of the quadratic variation and the integrated variance.
  - *Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. Journal of the American Statistical Association, 100(472), 1394–1411.*
  - *Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2008). Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise. Econometrica, 76(6), 1481–1536.*
  - *Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., & Vetter, M. (2009). Microstructure Noise in the Continuous Case: The Pre-Averaging Approach. Stochastic Processes and Their Applications, 119(7), 2249–2276.*
  - *Nolte, I., & Voev, V. (2012). Least Squares Inference on Integrated Volatility and the Relationship Between Efficient Prices and Noise. Journal of Business & Economic Statistics, 30(1), 94–108.*

Parametric Approach

- Estimation of Wiener process parameters.
  - Discrete process contaminated by the independent white noise follow ARIMA(0,1,1).
  - Parameters are estimated by maximum likelihood of reparametrization.
  - *Aït-Sahalia, Y., Mykland, P. A., & Zhang, L. (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. The Review of Financial Studie, 18(2), 351–416.*
- Estimation of Ornstein-Uhlenbeck process parameters.
  - Discrete process contaminated by the independent white noise follow ARIMA(1,0,1).
  - Parameters are estimated by method of moments, maximum likelihood of reparametrization and direct maximum likelihood.
  - *Holý, V., & Tomanová, P. (2017). Ornstein-Uhlenbeck Process Contaminated by the White Noise: Effects, Estimation and Application. In review.*

Academic Literature



Wiener Process

Wiener Process

- The standard Wiener process is a random variable W_t that satisfies
  - $W_t = 0$,
  - for $0 \leq s < t < u < v$, $W_t - W_s$ and $W_v - W_u$ are independent,
  - for $0 \leq s < t$, $W_t - W_s \sim \mathrm{N}(0, t-s)$.
- We consider the price process given by

$$P_t = \sigma W_t,$$

  where $\sigma > 0$ is the instantaneous volatility.
- When the process is contaminated, the estimate of σ is biased.

ARIMA(0,1,1) Reparametrization

- We consider equidistant sampling $\Delta = T_i - T_{i-1}$, $i = 1, \ldots, n$.
- The first difference of process X_i is

$$X_i - X_{i-1} = P_{T_i} - P_{T_{i-1}} + E_i - E_{i-1}.$$

- We can rewrite the process X_i as

$$X_i = \underbrace{X_{i-1}}_{X_i^{DF}} + \underbrace{R_{T_{i-1},T_i} + E_i - E_{i-1}}_{X_i^{MA}}, \qquad R_{T_{i-1},T_i} = P_{T_i} - P_{T_{i-1}} \sim \mathrm{N}(0, \sigma^2 \Delta).$$

- This is ARIMA(0,1,1) process, which can be reparametrized as

$$X_i = \underbrace{X_{i-1}}_{X_i^{DF}} + \underbrace{\theta V_{i-1} + V_i}_{X_i^{MA}}, \qquad V_i \sim \mathrm{N}(0, \gamma^2).$$

ARIMA(0,1,1) Estimation

- Parameters θ and γ^2 can be estimated by maximizing the likelihood function

$$L(\theta, \gamma^2) = f_{X_0}(x_0) f_{X_1}(x_1 | X_0 = x_0) \cdots f_{X_n}(x_n | X_0 = x_0, \ldots, X_{n-1} = x_{n-1}).$$

- To identify σ^2 and ω^2, we solve the equations

$$\begin{aligned} \gamma^2(1+\theta^2) &= \mathrm{var}[X_i^{MA}] = \sigma^2 \Delta + 2\omega^2, \\ \theta\gamma^2 &= \mathrm{cov}[X_i^{MA}, X_{i-1}^{MA}] = -\omega^2. \end{aligned}$$

- Finally, we get the original estimates as

$$\begin{aligned} \hat{\sigma}^2 &= \Delta^{-1}\hat{\gamma}^2(1+\hat{\theta})^2, \\ \hat{\omega}^2 &= -\hat{\gamma}^2\hat{\theta}. \end{aligned}$$

Ornstein-Uhlenbeck Process

Ornstein-Uhlenbeck Process

- The Ornstein-Uhlenbeck process P_t satisfy

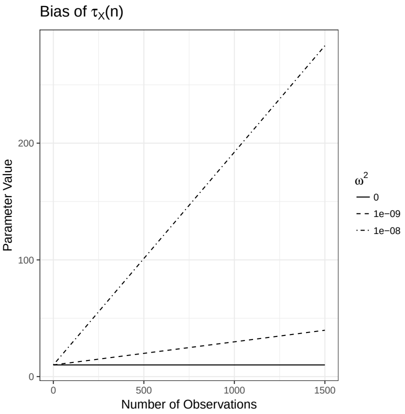
$$\mathrm{d}P_t = \tau(\mu - P_t)\mathrm{d}t + \sigma \mathrm{d}W_t,$$

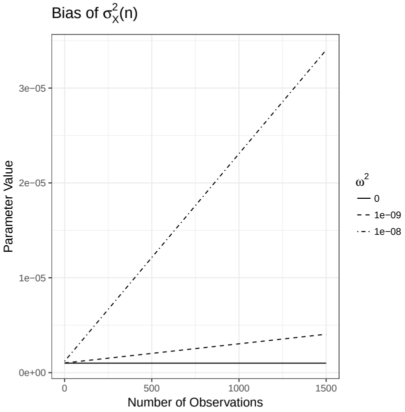
  where
  - W_t is a Wiener process,
  - μ is the long term mean level,
  - $\tau > 0$ is the speed of reversion,
  - $\sigma > 0$ is the instantaneous volatility.
- This stochastic differential equation has solution

$$P_t = P_0 e^{-\tau t} + \mu(1 - e^{-\tau t}) + \sigma \int_0^t e^{-\tau(t-s)} \mathrm{d}W_s.$$

- When the process is contaminated, the estimates of τ and σ are biased.

Bias of Ornstein-Uhlenebeck Parameters





ARIMA(1,0,1) Reparametrization

- When considering equidistant sampling, the process X_i can be decomposed as

$$X_i = P_{T_i} + E_i = P_{T_{i-1}} e^{-\tau\Delta} + \mu(1 - e^{-\tau\Delta}) + \sigma \int_0^\Delta e^{-\tau(\Delta - s)} \mathrm{d}W_s + E_i.$$

- Furthermore, from relation $P_{T_{i-1}} = X_{i-1} - E_{i-1}$ we obtain

$$X_i = \underbrace{\mu(1 - e^{-\tau\Delta})}_{X_i^{IC}} + \underbrace{e^{-\tau\Delta} X_{i-1}}_{X_i^{AR}} + \underbrace{\sigma \int_0^\Delta e^{-\tau(\Delta - s)} \mathrm{d}W_s - e^{-\tau\Delta} E_{i-1} + E_i}_{X_i^{MA}}.$$

- This is ARIMA(1,0,1) process, which can be reparametrized as

$$X_i = \underbrace{\alpha}_{X_i^{IC}} + \underbrace{\varphi X_{i-1}}_{X_i^{AR}} + \underbrace{\theta V_{i-1} + V_i}_{X_i^{MA}}, \qquad V_i \sim \mathrm{N}(0, \gamma^2).$$

ARIMA(1,0,1) Estimation

- Parameters α, φ, θ and γ^2 can be estimated by maximizing the likelihood function

$$L(\alpha, \varphi, \theta, \gamma^2) = f_{X_0}(x_0) f_{X_1}(x_1 | X_0 = x_0) \cdots f_{X_n}(x_n | X_0 = x_0, \ldots, X_{n-1} = x_{n-1}).$$

- Finally, we get the original estimates by solving equations

$$\begin{aligned}
\alpha = X_i^{IC} &= \mu(1 - e^{-\tau\Delta}), \\
\varphi X_{i-1} = X_i^{AR} &= e^{-\tau\Delta} X_{i-1}, \\
\gamma^2(1+\theta^2) = \mathrm{var}[X_i^{MA}] &= \frac{\sigma^2}{2\tau}(1 - e^{-2\tau\Delta}) + \omega^2(1 + e^{-2\tau\Delta}), \\
\theta\gamma^2 = \mathrm{cov}[X_i^{MA}, X_{i-1}^{MA}] &= -\omega^2 e^{-\tau\Delta}.
\end{aligned}$$

Direct Method of Moments

- We use four unconditional moments

$$m_1 = \mathrm{E}[X_i] = \mu, \qquad m_2 = \mathrm{var}[X_i] = \frac{\sigma^2}{2\tau} + \omega^2,$$
$$m_3 = \mathrm{cov}[X_i, X_{i-1}] = \frac{\sigma^2}{2\tau} e^{-\tau\Delta}, \quad m_4 = \mathrm{cov}[X_i, X_{i-2}] = \frac{\sigma^2}{2\tau} e^{-2\tau\Delta}.$$

- By replacing theoretical moments with their estimated counterparts, we get

$$\hat{\mu}_{MOM} = \hat{m}_1, \qquad \hat{\tau}_{MOM} = \frac{1}{\Delta} \log \frac{\hat{m}_3}{\hat{m}_4},$$
$$\hat{\sigma}^2_{MOM} = 2\frac{1}{\Delta}\frac{\hat{m}_3^2}{\hat{m}_4} \log \frac{\hat{m}_3}{\hat{m}_4}, \quad \hat{\omega}^2_{MOM} = \hat{m}_2 - \frac{\hat{m}_3^2}{\hat{m}_4}.$$

Direct Maximum Likelihood Estimation

- Observed variables X_i are normally distributed with conditional density functions $f_{X_{T_i}}(x_{T_i}|X_{T_{i-1}})$ and conditional moments

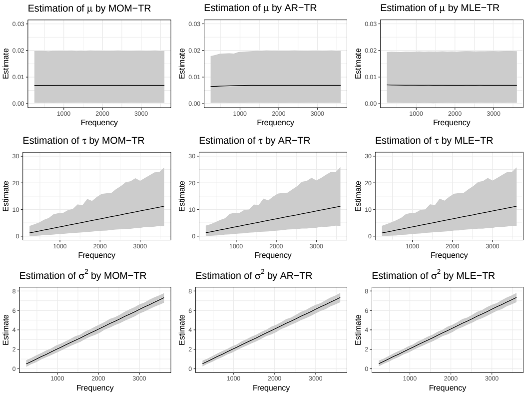
$$\mathrm{E}[X_{T_i}|X_{T_{i-1}} = x_{T_{i-1}}] = \frac{x_{T_{i-1}}\sigma^2 + 2\tau\mu\omega^2}{\sigma^2 + 2\tau\omega^2} e^{-\tau\Delta_i} + \mu\left(1 - e^{-\tau\Delta_i}\right),$$

$$\mathrm{var}[X_{T_i}|X_{T_{i-1}} = x_{T_{i-1}}] = \frac{\sigma^2\omega^2}{\sigma^2 + 2\tau\omega^2} e^{-2\tau\Delta_i} + \frac{\sigma^2}{2\tau}\left(1 - e^{-2\tau\Delta_i}\right).$$

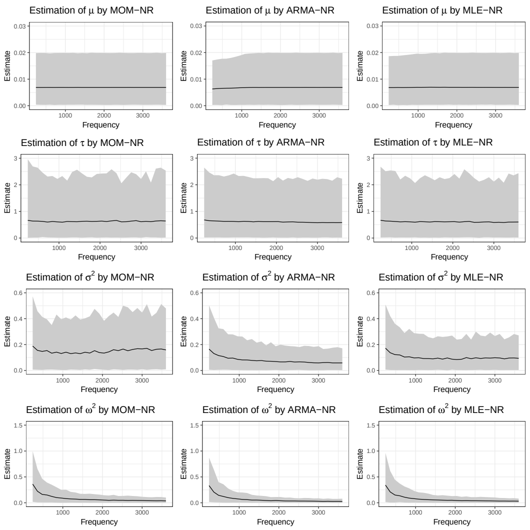
- We get the estimates by maximizing the logarithmic likelihood function

$$l(\mu, \tau, \sigma^2, \omega^2) = \sum_{i=1}^{n} \log f_{X_{T_i}}(x_{T_i}|X_{T_{i-1}} = x_{T_{i-1}}).$$

Bias of Traditional Parameter Estimates



Comparison of Proposed Estimators



Conclusion

Conclusion

- Ignoring the market microstructure noise leads to significant bias and inconsistency of estimates of the process parameters.
- For the Ornstein-Uhlenbeck process, finite-sample simulations show that the maximum likelihood of ARIMA(1,0,1) reparametrization gives the most accurate estimates.
- We assume the standard Ornstein-Uhlenbeck process and the independent Gaussian white noise. However, both of these assumptions are too restrictive for financial data as many papers suggest. Relaxation of Gaussian assumptions is a topic for the future research.
- High-frequency Ornstein-Uhlenbeck process can be utilized e.g. in pairs trading strategies and stochastic volatility models.

Thank you for your attention!

This work was supported by the Internal Grant Agency of University of Economics, Prague under Grant F4/63/2016.