Estimating the parametric covariation matrix: Equivalence, efficiency and estimation

Sebastian Holtz
Humboldt-Universität zu Berlin

June 10, 2016

Introduction

Consider the d-dimensional observation model

$$Y_i = X_{i/n} + \varepsilon_i,\ i = 1, \ldots, n,$$

where $X_t = \Sigma^{1/2} G_t$ with $G^{(j)} \overset{i.i.d.}{\sim} \Gamma,\ j = 1, \ldots, d$, for Γ Gaussian.

The target of estimation is the quadratic covariation matrix $\Sigma \in \mathbb{R}^{d\times d}$ and belongs to the class

$$\mathfrak{S}_L = \{\Sigma \in \mathbb{R}_+^{d\times d} : \Sigma_0 \leq \Sigma, \|\Sigma\| \leq L\}.$$

The errors ε_i are i.i.d. $\mathcal{N}(0, \eta^2 I_d)$ distributed and independent of G.

Possible application: High-frequency data

Literature concerning microstructure noise includes works by
Aït-Sahalia, Andersen, Barndorff-Nielsen, Bibinger, Christensen, Fan, Gloter, Hansen, Hautsch, Hoffmann, Jacod, Li, Lunde, Mykland, Munk, Podolskij, Reiß, Rosenbaum, Schmidt-Hieber, Shephard, Todorov, Uchida, Vetter, Yoshida, Zhang, Zheng.

Many estimation approaches
realised covariances, quasi Maximum likelihood, realised kernels, preaveraging, scaling, spectral estimators, etc.

What about explicit lower bounds?

$\underline{d = 1}$:

Case $\Gamma = B$

- Gloter and Jacod [2001]: $n^{1/4}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, 8\eta\sigma^3)$

Case $\Gamma = B^H$

- Gloter and Hoffmann [2007]: $r_n = n^{-1/(4H+2)}$
- Sabel and Schmidt-Hieber [2014]: Cramér-Rao bound

$\underline{d > 1}$:

Case $\Gamma = B$

- Bibinger et al. [2014]: Cramér-Rao bound

Here: General approach for a wide class of Γ.

Interlude: Le Cam theory

Definition

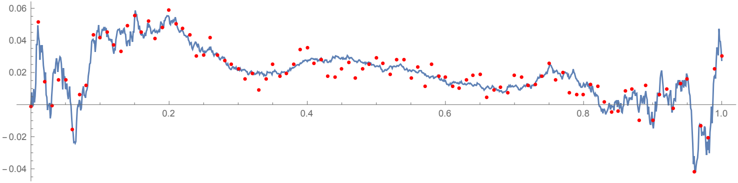
Let $\mathcal{E} = (X, \mathcal{X}, \{P_\theta : \theta \in \Theta\})$ and $\mathcal{F} = (Y, \mathcal{Y}, \{Q_\theta : \theta \in \Theta\})$ be two statistical experiments. Then the Le Cam deficiency between $\mathcal{E}$ and $\mathcal{F}$ is given by

$$\delta(\mathcal{E}, \mathcal{F}) = \inf_K \sup_{\theta \in \Theta} \|K \cdot P_\theta - Q_\theta\|_{TV},$$

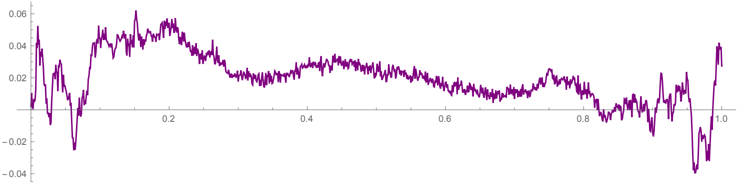
where the infimum is taken over all Markov kernels from $(X, \mathcal{X})$ to $(Y, \mathcal{Y})$.
The Le Cam distance is defined by $\Delta(\mathcal{E}, \mathcal{F}) = \max\{\delta(\mathcal{E}, \mathcal{F}), \delta(\mathcal{F}, \mathcal{E})\}$.

- If $\delta(\mathcal{E}, \mathcal{F}) = 0$ we say that $\mathcal{E}$ *is more informative than* $\mathcal{F}$.
- If $\Delta(\mathcal{E}, \mathcal{F}) = 0$ we say that $\mathcal{E}$ *and* $\mathcal{F}$ *are equivalent.*

Discrete and continuous model



$$\mathcal{D}_n : \; Y_i = X_{i/n} + \varepsilon_i, \; i = 1, \ldots, n.$$



$$\mathcal{C}_n : \; dY_t^n = X_t dt + \eta n^{-1/2} dW_t, \quad t \in [0, 1]$$

Proposition

Denote by $c(s,t)$ *the covariance function of* Γ. *Assume that the variance function* $v(t) = c(t,t)$ *belongs to* $C^\gamma([0,1])$ *for* $\gamma > 1/2$. *Additionally, suppose that for every* $s \in [0,1]$ *the derivative of* $c_s(t) = c(s,t)$ *lies in* $C^\beta([0,s)) \cap C^\beta((s,1])$. *Then it holds*

$$\Delta(\mathcal{D}_n, \mathcal{C}_n) = \mathcal{O}(n^{-(\beta \wedge 1/2 \wedge (\gamma - 1/2))}).$$

Corollary

Let $\Gamma = B^H$ *be a fractional Brownian motion with Hurst exponent* $H > 1/2$. *Then the corresponding discrete and continuous experiment are asymptotically equivalent. More precisely,*

$$\Delta(\mathcal{D}_n, \mathcal{C}_n) = \mathcal{O}(n^{-((2H-1) \wedge 1/2)}).$$

Sequence space model

Some notations

- C_Γ: covariance operator of Γ, given by

$$f(t) \mapsto \int_0^1 c(s,t) f(s) ds$$

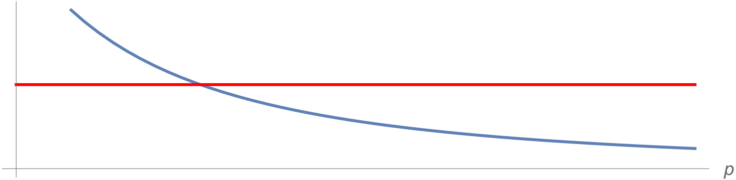
- denote by $\lambda = (\lambda_p)_{p \geq 1}$ and e_p the eigenvalues and eigenfunctions of C_Γ, respectively

Now: Restrict $\mathcal{C}_n : \; dY_t^n = X_t dt + n^{-1/2} dW_t$, to the eigenfunctions e_p and obtain a representation in terms of an independent sequence

$$Y_{np} \sim \mathcal{N}(0, \Sigma\lambda_p + \tfrac{1}{n} I_d), \; p \geq 1. \qquad (\mathcal{S}_n)$$

Equilibrium

Consider the variances $\Sigma\lambda_p + \frac{1}{n}I_d$ and the following plot of λ and $\frac{1}{n}$:



Equilibrium point: $\lambda(P_n) = \frac{1}{n}$:

Assumption

λ is regularly varying at infinity with index $-\alpha,\ \alpha > 1$.

Theorem

For $\lambda(P_n) = \frac{1}{n}$ *the Fisher information satisfies*

$$P_n^{-1} \sum_{p=1}^{\infty} \mathcal{I}_{np}(\Sigma) \to \mathcal{I}(\Sigma),$$

uniformly in $\Sigma \in \mathfrak{S}_L$*. The asymptotic Fisher information matrix* $\mathcal{I}(\Sigma)$ *equals*

$$\frac{\pi C}{4\alpha \sin(\pi/\alpha)} (Q^{\otimes 2})^\top diag\Big\{ diag\Big\{ \frac{s_j^{1/\alpha - 1} - s_i^{1/\alpha - 1}}{s_i - s_j} \Big\}_{1 \leq j \leq d} \Big\}_{1 \leq i \leq d} Q^{\otimes 2},$$

with $Q \in \mathbb{R}^{d \times d}$ *being the orthogonal matrix such that* $\Sigma = Q^\top diag\{s_i\}_{1 \leq i \leq d} Q$ *and* $C = \lim_{n \to \infty} r_n^2 P_n$*.*

Consequences

- the optimal rate for estimating Σ efficiently is given by

$$r_n \sim P_n^{-1/2}$$

- in particular, for some slowly varying L

$$r_n = \mathcal{O}(n^{-1/(2\alpha)} L(n)^{-1/2})$$

- asymptotically sufficient information for estimating Σ efficiently is contained in intervals $(P_n \kappa_n^0, P_n \kappa_n^\infty)$ for $\kappa_n^0 \to 0,\ \kappa_n^\infty \to \infty$

- α as smoothness index: smoothness $\uparrow \Rightarrow$ rate $\downarrow$
(It even holds: $\lambda > \lambda' \Rightarrow \delta(\mathcal{S}_n, \mathcal{S}_n') = 0$)

Theorem

The model $\mathcal{S}_n$ possesses the LAN property, i.e.

$$\log \frac{dP_{\Sigma+r_nH,n}}{dP_{\Sigma,n}}(Y_n) = r_n vec(H)^\top \nabla \ell(Y_n,\Sigma) - r_n^2 \frac{1}{2} vec(H)^\top \mathcal{I}_n(\Sigma) \mathcal{Z} vec(H) + \rho_n,$$

where $r_n \nabla \ell(Y_n,\Sigma) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\Sigma)\mathcal{Z})$ *and* $\rho_n = o_p(1)$.

Corollary

A sequence $vec(\hat{\Sigma}_n)$ *of efficient regular estimators of* $vec(\Sigma)$ *satisfies*

$$r_n^{-1} vec(\hat{\Sigma}_n - (\Sigma + r_n H)) \to \mathcal{N}(0, \tfrac{1}{4}\mathcal{I}(\Sigma)^{-1}\mathcal{Z}),$$

under $P_{\Sigma+r_nH}$ *and for a certain normalisation matrix* $\mathcal{Z} \in \mathbb{R}^{d^2 \times d^2}$.

Common lower bounds

Consider the underlying eigenvalues of the Brownian bridge and the Brownian motion:

$$\lambda_p^{BB} = (\pi p)^{-2} \quad \text{and} \quad \lambda_p^{BM} = \pi^{-2}(p + 1/2)^{-2}.$$

Question: Do the lower bounds coincide?

Proposition

Let $\lambda, \lambda' \in RV(-\alpha)$ induce the models $\mathcal{S}_n$ and $\mathcal{S}'_n$, respectively. Assume $|\lambda_p - \lambda'_p| = \mathcal{O}(p^{-(\alpha+\varepsilon)})$ with $\varepsilon > 1/2$. Then

$$\Delta(\mathcal{S}_n, \mathcal{S}'_n) = \mathcal{O}\left(P_n^{1/2-\varepsilon}\right).$$

What if $\varepsilon \leq 1/2$?

Common lower bounds II

Lemma

Let $\mathcal{S}_n$ and $\mathcal{S}'_n$ be two models induced by λ, λ', respectively. Then

$$\lambda_p/\lambda'_p \to 1 \iff r_n \sim r'_n \text{ and } \mathcal{I}(\Sigma) = \mathcal{I}'(\Sigma).$$

In particular, it holds

$$r_n^2 \sum_{p=1}^{\infty} \|\mathcal{I}_{np}(\Sigma) - \mathcal{I}'_{np}(\Sigma)\| = \mathcal{O}\left(\frac{\lambda_{P_n} - \lambda'_{P_n}}{\lambda_{P_n}}\right).$$

Example

Consider the Brownian motion, the Brownian bridge and the Ornstein-Uhlenbeck process. The underlying eigenvalues are

$$\lambda_p^{BM} = \pi^{-2}(p-1/2)^{-2}, \quad \lambda_p^{BB} = (\pi p)^{-2}, \quad \lambda_p^{OU} = (\pi^2 p^2 + \theta^2)^{-1}.$$

Thus their lower bounds coincide and the corresponding central limit theorem is given by

$$n^{1/4}\mathrm{vec}(\hat{\Sigma}_n - \Sigma) \to \mathcal{N}(0, 2(\Sigma \otimes \Sigma^{1/2} + \Sigma^{1/2} \otimes \Sigma)\mathcal{Z}).$$

Example

Consider the case $\Gamma = B^H,\ 0 < H < 1.$ Due to Bronski [2003] and Chigansky and Kleptsyna [2016] the leading terms of the eigenvalues of C_Γ are known:

$$\lambda_p^{fBM} = \frac{\sin(\pi H)\Gamma(2H+1)}{(p\pi)^{2H+1}} + o(p^{-(2H+1)}).$$

For $d = 1$ the corresponding central limit theorem is given by

$$n^{\frac{1}{(4H+2)}}(\hat{\sigma}_n^2 - \sigma^2) \to \mathcal{N}\left(0, \sigma^{\frac{2(4H+1)}{(2H+1)}} \cdot \frac{(2H+1)^2 \cdot \sin(\pi/(2H+1))}{H(\sin(H\pi)\cdot\Gamma(2H+1))^{1/(2H+1)}}\right).$$

Others: fOU, (fractional) Gaussian sheets, etc.

Oracle estimator

For every observation Y_{np} construct an estimator of $\mathrm{vec}(\Sigma)$ via

$$\mathrm{vec}(\hat{\Sigma}_{np}) = \lambda_p^{-1}\mathrm{vec}\left(Y_{np}Y_{np}^\top - \frac{1}{n}I_d\right).$$

With an optimal choice of (oracle) weights $W_{np}^{Or}(\Sigma)$ set

$$\mathrm{vec}(\hat{\Sigma}_n^{Or}) = \sum_{p=1}^{P_n S_n} W_{np}^{Or}(\Sigma)\mathrm{vec}(\hat{\Sigma}_{np}).$$

Theorem

If $S_n \to \infty$ then, under $P_{\Sigma + r_n H, n}$, the oracle estimator satisfies

$$r_n^{-1} vec(\hat{\Sigma}_n^{Or} - (\Sigma + r_n H)) \to \mathcal{N}(0, \tfrac{1}{4}\mathcal{I}(\Sigma)^{-1}\mathcal{Z}).$$

Adaptive estimator

Assume that a consistent pre-estimate $\text{vec}(\hat{\Sigma}_n^{pre})$ of $\text{vec}(\Sigma)$ can be derived. Plug-in:

$$\text{vec}(\hat{\Sigma}_n^{Ad}) = \sum_{p=1}^{P_n S_n} W_{np}^{Or}(\hat{\Sigma}_n^{pre})\text{vec}(\hat{\Sigma}_{np}).$$

Theorem

Let a pre-estimator vec$(\hat{\Sigma}_n)$ *with* $\|\hat{\Sigma}_n - \Sigma\| = o_p(1)$. *Then under* $P_{\Sigma + r_n H, n}$

$$r_n^{-1} vec(\hat{\Sigma}_n^{Ad} - (\Sigma + r_n H)) \to \mathcal{N}(0, \tfrac{1}{4}\mathcal{I}(\Sigma)^{-1}\mathcal{Z}).$$

Proposition

The MLE is a consistent estimator.

Outlook

- extend the class of model (drifts, time-dependent Σ, etc.)
- find the boundaries of the Le Cam result ($\Gamma = B^H,\ H = 1/4?$)
- sequence space approximations if λ not completely known

References

Markus Bibinger, Nikolaus Hautsch, Peter Malec, and Markus Reiß. Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *The Annals of Statistics*, 42(4):1312–1346, 2014.

Jared C. Bronski. Small ball constants and tight eigenvalue asymptotics for fractional brownian motions. *Journal of Theoretical Probability*, 16(1):87–100, 2003.

Pavel Chigansky and Marina Kleptsyna. Asymptotics of the karhunen-loeve expansion for the fractional brownian motion. *arXiv preprint arXiv:1601.05715*, 2016.

Arnaud Gloter and Marc Hoffmann. Estimation of the hurst parameter from discrete noisy data. *The Annals of Statistics*, 35(5): 1947–1974, 2007.

Arnaud Gloter and Jean Jacod. Diffusions with measurement errors. i. local asymptotic normality. *ESAIM: Probability and Statistics*, 5: 225–242, 2001.

Till Sabel and Johannes Schmidt-Hieber. Asymptotically efficient estimation of a scale parameter in gaussian time series and