On Schwarz type model comparison

Shoichi Eguchi   Hiroki Masuda

Kyushu University[1]

Dynstoch 2016 Rennes, June 8–10

1. Preliminaries and objective
2. QBIC: extended Schwarz's reach
3. Examples and simulations
   - Adaptive model-selection consistency in multi-scaling case
4. Summary

[1] Supported by JST CREST.

1. Preliminaries and objective
2. QBIC: extended Schwarz's reach
3. Examples and simulations
   - Adaptive model-selection consistency in multi-scaling case
4. Summary

BIC: Classical Schwarz's model comparison

True $\mathcal{L}(\mathbf{X}_n)$

- Prior density
- (Quasi) log-likelihood

Model I

$$\{\pi_{1,n}(\theta_1),\ \mathbb{H}_{1,n}(\theta_1;\mathbf{X}_n)\}$$

$\cdots$

Model M

$$\{\pi_{M,n}(\theta_1),\ \mathbb{H}_{M,n}(\theta_M;\mathbf{X}_n)\}$$

$$\hat{\theta}_{m,n} \in \underset{\theta_m\in\Theta_m}{\mathrm{argmax}}\ \mathbb{H}_{m,n}(\theta_m;\mathbf{X}_n)$$

- (Quasi) MLE
- (Quasi) marginal log-likelihood

$$\log\left(\int_{\Theta_m} \exp(\mathbb{H}_{m,n}(\theta_1))\pi_{m,n}(\theta_m)d\theta_m\right), \qquad 1 \le m \le M$$

BIC: Classical Schwarz's model comparison

True $\mathcal{L}(\mathbf{X}_n)$

- Prior density
- (Quasi) log-likelihood

Model I

$\{\pi_{1,n}(\theta_1),\ \mathbb{H}_{1,n}(\theta_1;\mathbf{X}_n)\}$

$\cdots$

Model M

$\{\pi_{M,n}(\theta_1),\ \mathbb{H}_{M,n}(\theta_M;\mathbf{X}_n)\}$

$$\hat{\theta}_{m,n} \in \underset{\theta_m \in \Theta_m}{\operatorname{argmax}}\, \mathbb{H}_{m,n}(\theta_m;\mathbf{X}_n)$$

- (Quasi) MLE
- (Quasi) marginal log-likelihood

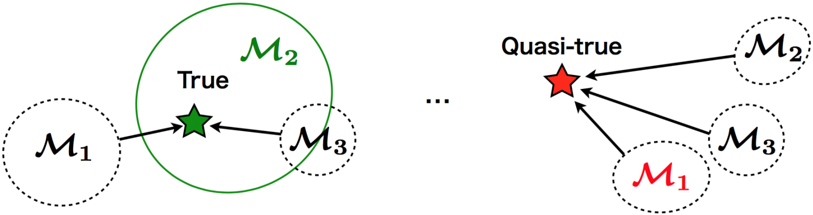
$$\log\left(\int_{\Theta_m} \exp(\mathbb{H}_{m,n}(\theta_1))\pi_{m,n}(\theta_m)d\theta_m\right), \qquad 1 \le m \le M$$

G. Schwarz (1978): Minimizing "$-2\times$(Marginal log-LF)" approximately

$$-2\log\left(\int_{\Theta_m} e^{\mathbb{H}_{m,n}(\theta)}\pi_{m,n}(\theta)d\theta\right) \approx \underbrace{-2\mathbb{H}_{m,n}(\hat{\theta}_{m,n}) + p_m \log n}_{=:\mathrm{BIC}_n^{(m)}} + O(1)$$

- Many variants and extensions of BIC, e.g. Cavanaugh and Neath (1999)

Motivation: Want to provide a BIC-like statistics



- Quasi-MLE $\hat{\theta}_n$ for intractable LF and/or under misspecification
- When $\mathcal{L}(\hat{\theta}_n)$ is approximately mixed-normally distributed?
  e.g. Volatility estimation of Itô semimartingales
- When the components of $\hat{\theta}_n$ converge at different rates?
  e.g. Ergodic diffusions and Lévy processes, observed at high frequency

Objective: a quasi-Bayes model $(\pi_n(\theta),\ \mathbb{H}_n(\theta))_{\theta\in\Theta}$

- **Prior density** $\pi_n(\theta)$**,** $\quad \theta \in \Theta \subset \mathbb{R}^p$
- **Quasi log-LF** $\mathbb{H}_n$**:** $\Omega \times \Theta \to \mathbb{R}$, **Quasi-MLE** $\hat{\theta}_n \in \operatorname{argmax} \mathbb{H}_n$
  - **Locally asymptotically quadratic (LAQ)**

$$\mathbb{H}_n(\theta_0 + A_n(\theta_0)u) - \mathbb{H}_n(\theta_0) = \Delta_n[u] - \frac{1}{2}\Gamma_0[u,u] + r_n(u),$$

$$A_n(\theta) = \mathrm{diag}\left\{a_{1n}(\theta)I_{p_1}, \dots, a_{Kn}(\theta)I_{p_K}\right\}, \quad a_{in}(\theta) \to 0$$

Objective: a quasi-Bayes model $(\pi_n(\theta), \mathbb{H}_n(\theta))_{\theta\in\Theta}$

- **Prior density** $\pi_n(\theta)$, $\theta \in \Theta \subset \mathbb{R}^p$
- **Quasi log-LF** $\mathbb{H}_n$: $\Omega \times \Theta \to \mathbb{R}$, **Quasi-MLE** $\hat{\theta}_n \in \operatorname{argmax} \mathbb{H}_n$
  - **Locally asymptotically quadratic (LAQ)**

$$\mathbb{H}_n(\theta_0 + A_n(\theta_0)u) - \mathbb{H}_n(\theta_0) = \Delta_n[u] - \frac{1}{2}\Gamma_0[u,u] + r_n(u),$$

$$A_n(\theta) = \operatorname{diag}\left\{a_{1n}(\theta)I_{p_1}, \dots, a_{Kn}(\theta)I_{p_K}\right\}, \quad a_{in}(\theta) \to 0$$

Extension of Schwarz's BIC reach

$$-2\log\left\{\int_\Theta \exp\left(\mathbb{H}_n(\theta)\right)\pi(\theta)d\theta\right\} \overset{!}{\approx} -2\mathbb{H}_n(\hat{\theta}_n) + (\textbf{Regularization})$$

- **Unified approach to approximate the marginal quasi-likelihood**
- **Adaptive model-selection consistency w.r.t. quasi-true model**

1. Preliminaries and objective
2. QBIC: extended Schwarz's reach
3. Examples and simulations
   - Adaptive model-selection consistency in multi-scaling case
4. Summary

Assumptions

- **Quasi observed information matrix** $\Gamma_n(\theta) := -A_n \partial_\theta^2 \mathbb{H}_n(\theta) A_n$

[A1] $\mathbb{H}_n \in C^3(\Theta)$ a.s., $\Gamma_n(\theta_0) \xrightarrow{p} \Gamma_0 > 0$ a.s., and for each $q > 0$,

$$\sup_n E\left(|A_n \partial_\theta \mathbb{H}_n(\theta_0)|^q + \sup_\theta |\Gamma_n(\theta)|^q + \sup_\theta |\partial_\theta \Gamma_n(\theta)|^q \right) < \infty$$

[A2] $\sup_{n,\theta} \pi_n(\theta) \vee \pi_n^{-1}(\theta) < \infty$, and $\forall M > 0$, $\sup_{|u| \le M} |\pi_n(\theta_0 + A_n u) - \pi_n(\theta_0)| \to 0$

[A3] $A_n^{-1}(\hat{\theta}_n - \theta_0)$ is $L^r(P)$-bounded for some $r > 3$

[A4] $\forall q > 0$, $\limsup_n E\left(\sup_\theta \lambda_{\min}^{-q}(\Gamma_n(\theta)) \right) < \infty$

[A5] $\limsup_n E\left\{ \left(\int e^{\mathbb{H}_n(\theta_0 + A_n u) - \mathbb{H}_n(\theta_0)} du \right)^\delta \right\} < \infty$ for some $\delta > 0$

- **Could be much weakened for the "stochastic expansion"**
- **"[A1] + [A4] plus alpha" $\Rightarrow$ [A3] (Yoshida (2011))**

Quasi-Bayesian information criterion (QBIC)

Theorem (Approximation of the expected marginal quasi log-LF)

$$
\begin{aligned}
E\Big(-2\log\int_{\Theta}\exp\left(\mathbb{H}_n(\theta)\right)\pi(\theta)d\theta\Big)
&= E\left(-2\mathbb{H}_n(\hat{\theta}_n) + \log\left|-\partial_\theta^2\mathbb{H}_n(\hat{\theta}_n)\right|\right) \\
&\quad - \underbrace{E\left(2\log\pi_n(\hat{\theta}_n) + p\log(2\pi)\right)}_{=O(1)} + o(1)
\end{aligned}
$$

Quasi-Bayesian information criterion (QBIC)

Theorem (Approximation of the expected marginal quasi log-LF)

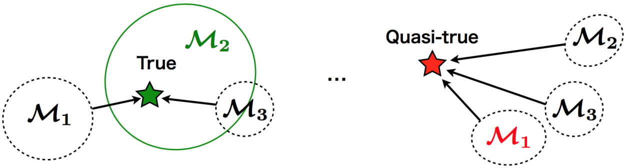
$$E\left(-2\log \int_\Theta \exp\left(\mathbb{H}_n(\theta)\right)\pi(\theta)d\theta \right)$$
$$= E\left(-2\mathbb{H}_n(\hat{\theta}_n) + \log\left| -\partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n)\right| \right)$$
$$- \underbrace{E\left(2\log \pi_n(\hat{\theta}_n) + p\log(2\pi) \right)}_{=O(1)} + o(1)$$

Definition (QBIC: an extended Schwarz's criterion)

$$\hat{\mathcal{S}}_n := -2\mathbb{H}_n(\hat{\theta}_n) + \log\left| -\partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n)\right|$$
$$= -2\mathbb{H}_n(\hat{\theta}_n) + \sum_{k=1}^K p_k \log\left(a_{kn}(\hat{\theta}_n)^{-2} \right) + o_E(1)$$

Approximate unbiased estimator of the relative Kullback-Leibler divergence

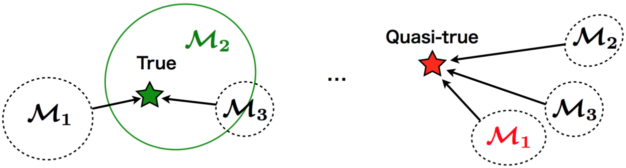
- $\mathrm{KL}_{m,n} := E\{\log g_n^{\mathrm{True}}(X_n)\} - E\{\log f_{m,n}(X_n)\}, \quad 1 \le m \le M$



$$
\begin{aligned}
&E\left(-2\log \int_{\Theta_m} \exp(\mathbb{H}_{m,n}(\theta))\pi_{m,n}(\theta)d\theta \right) \\
&= E\Big\{ -2\mathbb{H}_{m,n}(\hat{\theta}_{m,n}) + \log\left| -\partial_\theta^2 \mathbb{H}_{m,n}(\hat{\theta}_{m,n})\right| \\
&\qquad - \left(2\log \pi_{m,n}(\hat{\theta}_{m,n}) + p_m \log(2\pi) \right) \Big\} + o(1)
\end{aligned}
$$

Approximate unbiased estimator of the relative Kullback-Leibler divergence

- $\mathsf{KL}_{m,n} := E\{\log g_n^{\mathsf{True}}(X_n)\} - E\{\log f_{m,n}(X_n)\}, \quad 1 \le m \le M$



$$
\begin{aligned}
&E\left(-2\log\int_{\Theta_m}\exp(\mathbb{H}_{m,n}(\theta))\pi_{m,n}(\theta)d\theta\right)\\
&= E\Big\{-2\mathbb{H}_{m,n}(\hat\theta_{m,n}) + \log\left|-\partial_\theta^2\mathbb{H}_{m,n}(\hat\theta_{m,n})\right|\\
&\qquad - \left(2\log\pi_{m,n}(\hat\theta_{m,n}) + p_m\log(2\pi)\right)\Big\} + o(1)
\end{aligned}
$$

- $\mathsf{KL}_{m,n} - \mathsf{KL}_{l,n} = E\left(\frac{1}{2}(\hat{\mathcal{S}}^\sharp_{m,n} - \hat{\mathcal{S}}^\sharp_{k,n})\right) + o(1), \quad 1 \le m, l \le M$
  - Equivalently, approximation of the expected log-Bayes factor

1. Preliminaries and objective

2. QBIC: extended Schwarz's reach

3. Examples and simulations
   - Adaptive model-selection consistency in multi-scaling case

4. Summary

(1) QBIC for estimating volatility of continuous process

$$dY_t = \exp\left(\frac{1}{2}\langle X_t, \theta\rangle\right)dw_t, \qquad X_t = (X_{1,t}, \dots, X_{p,t})$$

- Data $(X_{t_j}, Y_{t_j})_{j=0}^n$, $t_j := jT/n$
- e.g. $\sup_{t,\omega} |X_t(\omega)| < \infty$, $P\left\{\lambda_{\min}\left(\frac{1}{T}\int_0^T X_t^{\otimes 2} dt\right) \le r^{-1}\right\} \lesssim r^{-L}$
- Gaussian QLF, (Genon-Catalot and Jacod (1993), Uchida and Yoshida (2013)):

$$\mathbb{H}_n(\theta) = \sum_{j=1}^n \log\phi\left(Y_{t_j};\, Y_{t_{j-1}},\, \frac{T}{n}\exp\left(\langle X_{t_{j-1}}, \theta\rangle\right)\right) \quad \Rightarrow \quad \hat{\theta}_n,\ \sqrt{n}\text{-AMN}$$

(1) QBIC for estimating volatility of continuous process

$$dY_t = \exp\left(\frac{1}{2}\langle X_t, \theta\rangle\right) dw_t, \qquad X_t = (X_{1,t}, \dots, X_{p,t})$$

- Data $(X_{t_j}, Y_{t_j})_{j=0}^n$, $t_j := jT/n$
- e.g. $\sup_{t,\omega} |X_t(\omega)| < \infty$, $P\left\{\lambda_{\min}\left(\frac{1}{T}\int_0^T X_t^{\otimes 2} dt\right) \le r^{-1}\right\} \lesssim r^{-L}$
- Gaussian QLF, (Genon-Catalot and Jacod (1993), Uchida and Yoshida (2013)):

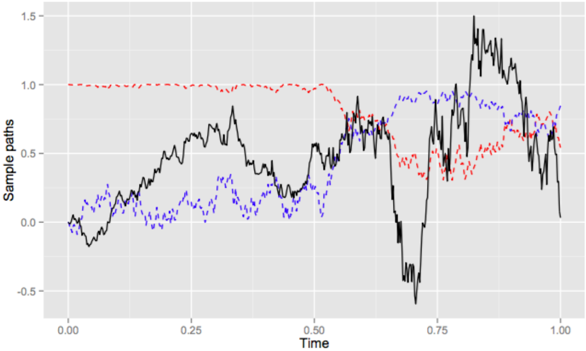
$$\mathbb{H}_n(\theta) = \sum_{j=1}^n \log \phi\left(Y_{t_j};\, Y_{t_{j-1}},\, \frac{T}{n}\exp\left(\langle X_{t_{j-1}}, \theta\rangle\right)\right) \quad \Rightarrow \quad \hat{\theta}_n,\ \sqrt{n}\text{-AMN}$$

Quasi BIC: $\hat{\mathcal{S}}_n = -2\mathbb{H}_n(\hat{\theta}_n) + \log|-\partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n)|$

Formal BIC: $\mathrm{BIC}_n = -2\mathbb{H}_n(\hat{\theta}_n) + p\log n$

Formal AIC(?): $\mathrm{fAIC}_n = -2\mathbb{H}_n(\hat{\theta}_n) + 2p$

- $p = 3; \quad X_{1,t} \equiv 10, \quad X_{2,t} = \cos(B_t), \quad X_{3,t} = \sin(B_t)$
  - B: standard BM independent of w
- $T = 1, \quad \sharp\text{MC} = 1000$



- **Model-selection frequency**

Criterion	$n=200$							$n=500$							$n=1000$						
	1	2	3	4*	5	6	7	1	2	3	4*	5	6	7	1	2	3	4*	5	6	7
QBIC	78	1	1	920	0	0	0	38	0	7	954	1	0	0	27	1	3	969	0	0	0
BIC	6	42	245	703	4	0	0	7	5	161	826	1	0	0	4	1	122	873	0	0	0
fAIC	74	40	236	648	2	0	0	94	2	155	748	1	0	0	119	1	115	765	0	0	0

- **Estimation performance $\theta_0 = (0, -2, 3)$**

		$n=200$			$n=500$			$n=1000$		
		$\hat\theta_1$	$\hat\theta_2$	$\hat\theta_3$	$\hat\theta_1$	$\hat\theta_2$	$\hat\theta_3$	$\hat\theta_1$	$\hat\theta_2$	$\hat\theta_3$
1	mean	-0.0242	-1.7665	2.9373	-0.0091	-1.9145	2.9985	-0.0080	-1.9306	2.9845
	s.d.	0.2290	2.2493	0.8445	0.1366	1.3373	0.5130	0.1028	1.0048	0.3855
2	mean	0.0620	-2.5040	–	0.0870	-2.7296	–	0.1050	-2.8925	–
	s.d.	0.6529	6.3498	–	0.6370	6.1847	–	0.6329	6.1223	–
3	mean	-0.2047	–	2.9181	-0.2029	–	2.9897	-0.2048	–	3.0437
	s.d.	0.0471	–	1.2826	0.0473	–	1.2787	0.0503	–	1.2950
4*	mean	–	-1.9948	2.9628	–	-1.9976	2.9963	–	-2.0038	2.9883
	s.d.	–	0.1712	0.3275	–	0.1053	0.1918	–	0.0751	0.1424
5	mean	-0.1029	–	–	-0.0946	–	–	-0.0871	–	–
	s.d.	0.1484	–	–	0.1527	–	–	0.1518	–	–
6	mean	–	-1.9063	–	–	-1.8867	–	–	-1.8722	–
	s.d.	–	0.4801	–	–	0.4622	–	–	0.4716	–
7	mean	–	–	3.0147	–	–	2.9964	–	–	2.9648
	s.d.	–	–	0.5319	–	–	0.5530	–	–	0.5440

(2) QBIC for estimating ergodic diffusion

$$dX_t = a(X_t, \alpha)dt + b(X_t, \beta)dw_t$$

- **Data** $(X_{t_j})_{j=0}^n$, $t_j = jn^{-2/3}$ $(T_n = n^{1/3} \to \infty)$, $\theta := (\alpha, \beta) \in \mathbb{R}^{p_\alpha} \times \mathbb{R}^{p_\beta}$
- **Efficient Gaussian QLF (Kessler, 1997):**

$$\mathbb{H}_n(\theta) = \sum_{j=1}^n \log \phi\Big(X_{t_j};\ X_{t_{j-1}} + h_n a(X_{t_{j-1}}, \alpha),\ h_n b^2(X_{t_{j-1}}, \beta)\Big)$$

$$\Rightarrow \Big(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0)\Big) \xrightarrow{\mathcal{L}} N\big(0, \mathrm{diag}(\Sigma_\alpha, \Sigma_\beta)\big)$$

(2) QBIC for estimating ergodic diffusion

$$dX_t = a(X_t, \alpha)dt + b(X_t, \beta)dw_t$$

- **Data** $(X_{t_j})_{j=0}^n$, $t_j = jn^{-2/3}$ $(T_n = n^{1/3} \to \infty)$, $\theta := (\alpha, \beta) \in \mathbb{R}^{p_\alpha} \times \mathbb{R}^{p_\beta}$
- **Efficient Gaussian QLF (Kessler, 1997):**

$$\mathbb{H}_n(\theta) = \sum_{j=1}^n \log \phi\Big(X_{t_j};\ X_{t_{j-1}} + h_n a(X_{t_{j-1}}, \alpha),\ h_n b^2(X_{t_{j-1}}, \beta)\Big)$$

$$\Rightarrow \Big(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0)\Big) \xrightarrow{\mathcal{L}} N\big(0, \mathrm{diag}(\Sigma_\alpha, \Sigma_\beta)\big)$$

Quasi BIC: $\hat{S}_n = -2\mathbb{H}_n(\hat{\theta}_n) + \log|-\partial_\alpha^2 \mathbb{H}_n(\hat{\theta}_n)| + \log|-\partial_\beta^2 \mathbb{H}_n(\hat{\theta}_n)|$

Formal BIC: $\mathrm{BIC}_n = -2\mathbb{H}_n(\hat{\theta}_n) + p_\alpha \log T_n + p_\beta \log n$

Formal AIC (CIC, Uchida, 2010): $\mathrm{CIC}_n = -2\mathbb{H}_n(\hat{\theta}_n) + 2p$

Adaptive model-selection consistency in multi-scaling case

$$dX_t = a(X_t, \alpha)dt + b(X_t, \beta)dw_t,$$

- **Avoiding "full-model" search, reducing computation cost [2]:**
  1. **First, make a selection for the diffusion coefficient by the auxiliary quasi-log LF ignoring the drift term:**

$$\hat{\mathcal{S}}_n^1 = -2\mathbb{H}_n^1(\hat{\beta}_n) + \log|-\partial_\beta^2 \mathbb{H}_n^1(\hat{\beta}_n)|$$

  2. **Then, for the drift coefficient by making use of $\hat{\beta}_n$:**

$$\hat{\mathcal{S}}_n^2 = -2\mathbb{H}_n(\hat{\alpha}_n, \hat{\beta}_n) + \log|-\partial_\alpha^2 \mathbb{H}_n(\hat{\alpha}_n, \hat{\beta}_n)|$$

followed by the coefficient-selection consistency.

[2] General multi-scale LAQ case can be treated similarly.

Simulation design

$$dX_t = -X_t dt + \exp\left\{\frac{1}{2}(-2\cos X_t + 1)\right\}dw_t, \quad X_0 = 1$$

$$\text{Diff 1}: \exp\left\{\frac{1}{2}(\theta_{11}\cos X_t + \theta_{12}\sin X_t + \theta_{13})\right\};$$

$$\text{Diff 2}: \exp\left\{\frac{1}{2}(\theta_{11}\cos X_t + \theta_{12}\sin X_t)\right\};$$

$$\text{Diff 3}: \exp\left\{\frac{1}{2}(\theta_{11}\cos X_t + \theta_{13})\right\};$$

$$\text{Diff 4}: \exp\left\{\frac{1}{2}(\theta_{12}\sin X_t + \theta_{13})\right\};$$

$$\text{Diff 5}: \exp\left\{\frac{1}{2}\theta_{11}\cos X_t\right\};\ \text{Diff 6}: \exp\left\{\frac{1}{2}\theta_{12}\sin X_t\right\};\ \text{Diff 7}: \exp\left\{\frac{1}{2}\theta_{13}\right\}$$

$$\text{Drif 1}: \theta_{21}X_t + \theta_{22};$$

$$\text{Drif 2}: \theta_{21}X_t;$$

$$\text{Drif 3}: \theta_{22}$$

Joint QBIC (left) and adaptive QBIC (right)

	Criteria	$n = 1000$						
		Diff 1	Diff 2	Diff 3*	Diff 4	Diff 5	Diff 6	Diff 7
Drif 1	QBIC	5	11	103	0	11	0	1
	BIC	0	24	90	0	50	0	1
	CIC	20	31	121	0	18	0	1
Drif 2*	QBIC	30	17	739	2	77	0	2
	BIC	2	24	542	1	234	0	3
	CIC	96	39	569	3	71	0	3
Drif 3	QBIC	0	0	2	0	0	0	0
	BIC	1	0	28	0	0	0	0
	CIC	9	0	19	0	0	0	0

		$n = 3000$						
		Diff 1	Diff 2	Diff 3*	Diff 4	Diff 5	Diff 6	Diff 7
Drif 1	QBIC	2	0	93	0	0	0	0
	BIC	1	0	110	0	14	0	0
	CIC	22	1	166	0	1	0	0
Drif 2*	QBIC	13	3	876	0	13	0	0
	BIC	1	5	811	0	51	0	0
	CIC	104	7	686	0	6	0	0
Drif 3	QBIC	0	0	0	0	0	0	0
	BIC	0	0	7	0	0	0	0
	CIC	2	0	5	0	0	0	0

		$n = 5000$						
		Diff 1	Diff 2	Diff 3*	Diff 4	Diff 5	Diff 6	Diff 7
Drif 1	QBIC	1	0	72	0	0	0	0
	BIC	0	0	94	0	1	0	0
	CIC	42	0	138	0	0	0	0
Drif 2*	QBIC	15	0	910	0	1	0	0
	BIC	1	1	882	0	17	0	0
	CIC	140	0	675	0	1	0	0
Drif 3	QBIC	0	0	1	0	0	0	0
	BIC	0	0	4	0	0	0	0
	CIC	0	0	4	0	0	0	0

	Criteria	$n = 1000$						
		Diff 1	Diff 2	Diff 3*	Diff 4	Diff 5	Diff 6	Diff 7
Drif 1	QBIC	8	6	108	0	6	0	0
	BIC	0	13	115	0	37	0	0
Drif 2*	QBIC	27	12	774	0	56	0	1
	BIC	2	17	594	0	192	0	1
Drif 3	QBIC	0	0	2	0	0	0	0
	BIC	1	0	28	0	0	0	0

	Criteria	$n = 3000$						
		Diff 1	Diff 2	Diff 3*	Diff 4	Diff 5	Diff 6	Diff 7
Drif 1	QBIC	3	0	92	0	0	0	0
	BIC	1	0	116	0	7	0	0
Drif 2*	QBIC	12	1	888	0	4	0	0
	BIC	1	4	833	0	31	0	0
Drif 3	QBIC	0	0	0	0	0	0	0
	BIC	0	0	7	0	0	0	0

	Criteria	$n = 5000$						
		Diff 1	Diff 2	Diff 3*	Diff 4	Diff 5	Diff 6	Diff 7
Drif 1	QBIC	2	0	71	0	0	0	0
	BIC	0	0	94	0	1	0	0
Drif 2*	QBIC	14	0	912	0	0	0	0
	BIC	1	1	886	0	13	0	0
Drif 3	QBIC	0	0	1	0	0	0	0
	BIC	0	0	4	0	0	0	0

- **Joint QBIC spent nearly double time compared with adaptive QBIC**

1 Preliminaries and objective

2 QBIC: extended Schwarz's reach

3 Examples and simulations
- Adaptive model-selection consistency in multi-scaling case

4 Summary

Summary

$$-2\log\left(\int_\Theta \exp\left(\mathbb{H}_n(\theta)\right)\pi(\theta)d\theta\right) \approx -2\mathbb{H}_n(\hat{\theta}_n)+\log\left|-\partial_\theta^2\mathbb{H}_n(\hat{\theta}_n)\right|-2\log\pi_n(\hat{\theta}_n)-p\log(2\pi)$$

QBIC: Extended Schwarz's statistics for LAQ models under possible multi-scaling

$$\hat{S}_n := -2\mathbb{H}_n(\hat{\theta}_n)+\log\left|-\partial_\theta^2\mathbb{H}_n(\hat{\theta}_n)\right|$$

$$= \underbrace{-2\mathbb{H}_n(\hat{\theta}_n)+\sum_{k=1}^K p_k\log\left(a_{kn}(\hat{\theta}_n)^{-2}\right)}_{\text{Formal BIC}}+\underbrace{\log\left|-A_n\partial_\theta^2\mathbb{H}_n(\hat{\theta}_n)A_n\right|}_{=O_E(1)}$$

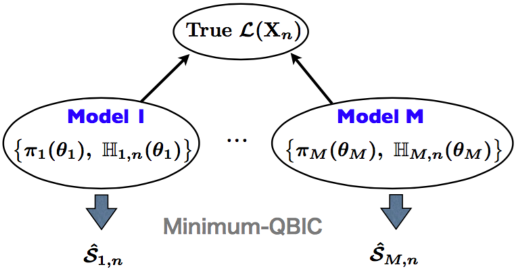
Summary

$$-2\log\left(\int_\Theta \exp\left(\mathbb{H}_n(\theta)\right)\pi(\theta)d\theta\right) \approx -2\mathbb{H}_n(\hat{\theta}_n)+\log\left|-\partial_\theta^2\mathbb{H}_n(\hat{\theta}_n)\right|-2\log \pi_n(\hat{\theta}_n)-p\log(2\pi)$$

QBIC: Extended Schwarz's statistics for LAQ models under possible multi-scaling

$$\hat{\mathcal{S}}_n := -2\mathbb{H}_n(\hat{\theta}_n)+\log\left|-\partial_\theta^2\mathbb{H}_n(\hat{\theta}_n)\right|$$

$$= \underbrace{-2\mathbb{H}_n(\hat{\theta}_n)+\sum_{k=1}^{K}p_k\log\left(a_{kn}(\hat{\theta}_n)^{-2}\right)}_{\text{Formal BIC}} + \underbrace{\log\left|-A_n\partial_\theta^2\mathbb{H}_n(\hat{\theta}_n)A_n\right|}_{=O_E(1)}$$



- $\mathcal{M}_k$ vs $\mathcal{M}_m$, Approximation of
  - $\mathrm{KL}_{m,n} - \mathrm{KL}_{k,n}$
  - $\log \dfrac{P(\mathcal{M}_m|X_n)}{P(\mathcal{M}_k|X_n)}$
- Adaptive consistent selection

References

- Eguchi, S. and Masuda, H. (2016), Schwarz type model comparison for LAQ models. arXiv:1606.01627

[1] Cavanaugh, J. E. and Neath, A. A. (1999), Generalizing the derivation of the Schwarz information criterion. Comm. Statist. Theory Methods, 28, 49–66.

[2] Genon-Catalot, V. and Jacod, J. (1993), On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. Ann. Inst. H. Poincaré Probab. Statist. 29, 119–151.

[3] Kessler, M. (1997), Estimation of an ergodic diffusion from discrete observations. Scand. J. Statist., 24, 211–229.

[4] Lv, J. and Liu, J. S. (2014), Model selection principles in misspecified models, J. R. Stat. Soc. B., 76, 141–167.

[5] Schwarz, G. (1978), Estimating the dimension of a model. Ann. Math. Statist., 6, 461–464.

[6] Uchida, M. and Yoshida, N. (2013), Quasi likelihood analysis of volatility and nondegeneracy of statistical random field. Stochastic Process. Appl. 123, 2851–2876.

[7] Yoshida, N. (2011), Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations, Ann. Inst. Statist. Math., 63, 431–479.