Asymptotic expansion

Yan LIU

1. Model Class in Time Series Analysis

1. Let D_1 be

$$D_1 = \{f; f(\lambda) = \sum_{r=-\infty}^{\infty} a(r)\exp(-ir\lambda),\ a(r) = a(-r),\ \sum_{r=-\infty}^{\infty} |r||a(r)| < \infty\}.$$

2. Let D_{ARMA} be

$$\begin{aligned} D_{\text{ARMA}} =&\{f; f(\lambda) = \frac{\sigma^2}{2\pi}\frac{|\sum_{j=0}^{q} a_j e^{ij\lambda}|^2}{|\sum_{j=0}^{p} b_j e^{ij\lambda}|^2}, \sigma^2 > 0, p, q \,\text{integers} \\ &\sum_{j=0}^{q} a_j z^j \text{ and } \sum_{j=0}^{q} a_j z^j \text{ are both bounded away from zero for } |z| \le 1\}. \end{aligned}$$

2. Formulae

Lemma 2.1 (Taniguchi (1983))**.** *Suppose* $f_1(\lambda), \ldots, f_s(\lambda) \in D_1$ *and* $g_1(\lambda), \ldots, g_s(\lambda) \in D_{ARMA}$. *Suppose the Toeplitz matrices are defined by*

$$\begin{aligned} (\Gamma_j)_{ml} &= \left\{\int_{-\pi}^{\pi} e^{i(m-l)\lambda} f_j(\lambda) d\lambda\right\}, \\ (\Lambda_j)_{ml} &= \left\{\int_{-\pi}^{\pi} e^{i(m-l)\lambda} g_j(\lambda) d\lambda\right\}. \end{aligned}$$

Then

$$\frac{1}{n}\text{tr}(\Gamma_1\Lambda_1^{-1}\Gamma_2\Lambda_2^{-1}\cdots\Gamma_s\Lambda_s^{-1}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\prod_{j=1}^{s}\{f_j(\lambda)g_j(\lambda)^{-1}\}d\lambda + O(n^{-1}).$$

Lemma 2.2 (TK or TA (1979))**.**

$$\sqrt{n}(\hat\theta_n^i - \theta^i) = \bar I^{ii'} Z_{i'} + n^{-1/2}\Big[\bar I^{ii'}\bar I^{j'k'} Z_{i'j'} Z_{k'} + \frac{1}{2}\bar I^{ii'}\bar I^{j'm'}\bar I^{k'l'}\bar\Gamma_{i'j'k'} Z_{m'} Z_{l'}\Big] + o_p(n^{-1/2}), \quad i = 1, \ldots, p.$$

Suppose

$$\mathcal{I}_{ij} = \frac{1}{4\pi}\int_{-\pi}^{\pi}(\partial_i f_\theta(\omega)\partial_j f_\theta(\omega)) f_\theta(\omega)^{-2} d\omega.$$

$$\mathcal{J}_{ijk} = -\frac{1}{2\pi}\int_{-\pi}^{\pi}(\partial_i f_\theta(\omega)\partial_j f_\theta(\omega)\partial_k f_\theta(\omega))f_\theta(\omega)^{-3}d\omega + \frac{1}{4\pi}\int_{-\pi}^{\pi}(\partial_i f_\theta(\omega)\partial_j\partial_k f_\theta(\omega))f_\theta(\omega)^{-2}d\omega.$$

$$\mathcal{K}_{ijk} = \frac{1}{2\pi}\int_{-\pi}^{\pi}(\partial_i f_\theta(\omega)\partial_j f_\theta(\omega)\partial_k f_\theta(\omega))f_\theta(\omega)^{-3}d\omega.$$

3. some lemmas

Lemma 3.1 (Petrov's lemma)**.** *Consider X is a d-dimensional random vector with distribution function F. Suppose there are a constant $c \in (0,1]$ and another constant $c_1 \in (0,\infty)$ such that,*

$$|E\, e^{iu\cdot X}| \leq 1 - c \tag{3.1}$$

for all $u \in \mathbb{R}^d$ satisfying $\|u\| \geq c_1$. Then for any $u \in \mathbb{R}^d$ and $\|u\| \leq c_1$,

$$|E\, e^{iu\cdot X}| \leq \exp(-\frac{c}{8c_1^2}\|u\|^2). \tag{3.2}$$

Proof. Denote $g(u) = |Ee^{iu\cdot X}|^2$. Then

$$g(u) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\cos(u\cdot(x-y))F(dx)F(dy), \tag{3.3}$$

since F(dx)F(dy) is an even function. Note $1-\cos^2(x) \leq 2(1-\cos(x))$, we have

$$1 - g(2^m x) \leq 4^m(1-g(x)) \quad \text{for all } x \in \mathbb{R}^d. \tag{3.4}$$

The remainder of proof is left for readers. □

Assumption 3.2 (Cramër's condition)**.**

$$\limsup_{|u|\to\infty}|E\, e^{iu\cdot Z_1}| < 1. \tag{3.5}$$

Denote

$$p_{p-2}(z;X) = d\Psi_{p-2}(X)/dz \tag{3.6}$$

$$w(f;r,F) = \int_{\mathbb{R}^d}\sup_{z:|z|\leq r}|f(y+z)-f(y)|F(dy). \tag{3.7}$$

Theorem 3.3. *Suppose $L_0 > 0$, and a symmetric matrix Σ_0 satisfies $\Sigma_0 > \Sigma$. Under assumption (**??**), for some constants C and $\delta > 0$, we have*

$$|E\, f(S_n) - \int_{\mathbb{R}^d} f(z)p_{p-2}(z;S_n)\, dz| \leq C\omega(f;n^{-L_0},N_d(0,\Sigma_0)) + \overline{o}(n^{-(p-2)/2-\delta}). \tag{3.8}$$

Proof. See Yoshida(2006). □

4. Information Geometry

Define

$$g_{ij} = \langle \partial_i, \partial_j \rangle = E_p[\, \partial_i l \partial_j l \,]. \tag{4.1}$$

Then the class of probability function constitute a Rieman manifold.

We have α-connection defined as follows:

$$\Gamma^{(\alpha)}_{ijk} = E_p\left[\left(\partial_i \partial_j l + \frac{1-\alpha}{2} \partial_i l \partial_j l \right) \partial_k l \right] \tag{4.2}$$

ANDERSON(1971)–SOME THEOREMS

GEN RYU

Part 1. For Kreiss(1987)

[Lemma 5.5.2 –Theorem 5.5.7]

1. THE MODEL

$$\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{y}_{t-1} = \boldsymbol{u}_t. \tag{1.1}$$

or

$$y_t + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} = u_t. \tag{1.2}$$

Under the assumption (A.1) below, we can write the process as

$$y_t = \sum_{s=0}^{\infty} (-\boldsymbol{B})^s \boldsymbol{u}_{t-s}, \quad t = \ldots, -1, 0, 1, \ldots.$$

Let $\boldsymbol{F}$ be the sample variance, then it can be written in the form like

$$\boldsymbol{F} = E\boldsymbol{y}_t\boldsymbol{y}_t' = \sum_{s=0}^{\infty} (-B)^s \Sigma (-B')^s.$$

Also define $\tilde{\boldsymbol{F}}$ as

$$\tilde{\boldsymbol{F}} = \sum_{s=0}^{\infty} \tilde{\boldsymbol{B}}^s \tilde{\Sigma} \tilde{\boldsymbol{B}}'^s,$$

where

$$\tilde{\Sigma} = \begin{pmatrix} \sigma^2 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{O} \end{pmatrix}, \quad \tilde{\boldsymbol{B}} = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_p \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Date: May 15, 2012.

2. ASSUMPTIONS

2.1. Set 1.

(A.1) $\{\boldsymbol{y}_t\}$ is a sequence of random vectors satisfying (1.1) with $\{\boldsymbol{u}_t\}$ independently and identically distributed with $E\boldsymbol{u}_t = \mathbf{0}$ and $E\boldsymbol{u}_t\boldsymbol{u}_t' = \boldsymbol{\Sigma}$.

(A.2) $-\boldsymbol{B}$ has all characteristic roots less than 1 in absolute value.

(A.3) $\boldsymbol{F}$ is positive definite.

2.2. Set 2.

(A.1)' $\{y_t\}$ is a sequence of random vectors satisfying (1.2) with $\{u_t\}$ independently and identically distributed with $Eu_t = 0$ and $Eu_t^2 = \sigma^2$.

(A.2)' The roots of the associated polynomial equation are less than 1 in absolutely value.

3. THEOREMS

Theorem 3.1. *Under the set 1 of assumptions,* $\sqrt{T}(\hat{\boldsymbol{B}}' - \boldsymbol{B}')$ *has a limiting normal distribution with mean* $\mathbf{0}$ *and covariance matrix* $\boldsymbol{F}^{-1} \otimes \boldsymbol{\Sigma}$*.*

Theorem 3.2. *Under the set 2 of assumptions,* $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ *has a limiting normal distribution with mean* $\mathbf{0}$ *and covariance matrix* $\sigma^2\tilde{\boldsymbol{F}}^{-1}$*.*

Instead of (A.1) or (A.1)', we have the result if we have the assumption
(A.1)" $E|\, u_{it}\,|^{2+\epsilon} < m, \quad i = 1, 2, \ldots, p, t = 1, 2, \ldots,$ for some $\epsilon > 0$ and some m.

4. AUXILIARY RESULTS

4.1. Equation.

It is true that

$$\boldsymbol{F} = \Sigma + \boldsymbol{B}\boldsymbol{F}\boldsymbol{B'},$$

which shows that if Σ is positive definite, then $\boldsymbol{F}$ is also positive definite.

Complex Analysis; 留数計算など

Yan LIU

1. 基本

1.1. 心得

複素解析のポイント

1. 積分経路を決める；
2. 正則であるため、コーシー積分をする；
3. もし特異点がある場合、留数を求める。

1.2. Example

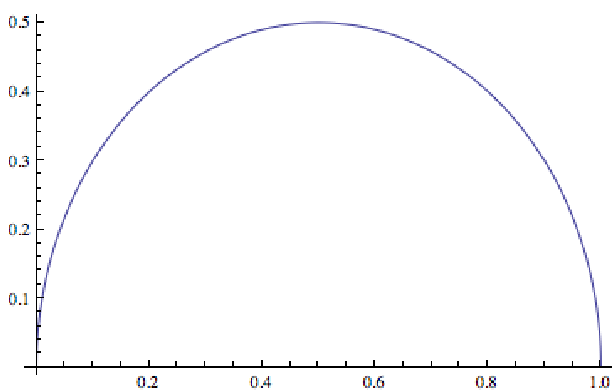
Calculation

$$\int_0^1 \frac{1}{x+1}\,dx = \quad ? \tag{1.1}$$

As what we are taught in the calculus course, you can easily answer this question. The answer is

$$\int_0^1 \frac{1}{x+1}\,dx = [\,\log(x+1)\,]_0^1 = \log 2. \tag{1.2}$$

What we will do here is to apply the residual theorem to it! Let us consider the integral is on the complex plane, and we choose the integral path as follows.



Then,

$$\int_0^1 \frac{1}{x+1}\,dx = -\int_0^\pi \frac{1/2ie^{i\theta}}{1+1/2+1/2e^{i\theta}}\,d\theta = -\int_0^\pi \frac{ie^{i\theta}}{3+1e^{i\theta}}\,d\theta \tag{1.3}$$

$$= \int_{-1}^1 \frac{1}{3+y}\,dy = [\log(3+y)]_{-1}^1 = \log 2. \tag{1.4}$$

2. Integration with Residue Theorem

Formula 1.1

$$\int_0^{2\pi} \frac{d\theta}{a+b\cos\theta} = \frac{2\pi}{\sqrt{a^2-b^2}}. \tag{2.1}$$

To derive the formula, you only have to think about the singularity in it.

2.1. Time Series Model

We will give some examples of integration of spectrums to look at how powerful residue theorem is in time series.

2.1.1. MA(1)

$$\int_{-\pi}^{\pi}(1+\theta e^{i\lambda})(1+\theta e^{-i\lambda})\,d\lambda = \int_{|z|=1}(1-\theta z)(1-\theta/z)\,\frac{dz}{iz} \tag{2.2}$$

$$= \int_{|z|=1}\frac{(1-\theta z)(z-\theta)}{iz^2}\,dz \tag{2.3}$$

$$= 2\pi i\cdot \mathrm{Res}(f_{\mathrm{MA}(1)},0) = 2\pi(1+\theta^2). \tag{2.4}$$

Formula 2.1

$$\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}(1+\theta e^{i\lambda})(1+\theta e^{-i\lambda})\,d\lambda = \sigma^2(1+\theta^2) \tag{2.5}$$

2.2. AR(1)

$$\int_{-\pi}^{\pi}\frac{1}{(1-\theta e^{i\lambda})(1-\theta e^{-i\lambda})}\,d\lambda = \int_{|z|=1}\frac{1}{(1+\theta z)(1+\theta/z)}\,\frac{dz}{iz} \tag{2.6}$$

$$= \int_{|z|=1}\frac{1}{(1+\theta z)(z+\theta)}\,\frac{dz}{i} \tag{2.7}$$

$$= 2\pi i\cdot \mathrm{Res}(f_{\mathrm{AR}(1)},-\theta) = \frac{2\pi}{1-\theta^2}. \tag{2.8}$$

Formula 2.2

$$\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{1}{(1-\theta e^{i\lambda})(1-\theta e^{-i\lambda})}\,d\lambda = \frac{\sigma^2}{1-\theta^2} \tag{2.9}$$

2.3. ARMA(1,1)

$$\begin{aligned}
\int_{-\pi}^{\pi}\frac{(1+\theta e^{i\lambda})(1+\theta e^{-i\lambda})}{(1-\phi e^{i\lambda})(1-\phi e^{-i\lambda})}\,d\lambda &= \int_{|z|=1}\frac{(1-\theta z)(1-\theta/z)}{(1+\phi z)(1+\phi/z)}\,\frac{dz}{iz} &(2.10)\\
&= \int_{|z|=1}\frac{(1-\theta z)(z-\theta)}{(1+\phi z)(z+\phi)z}\,\frac{dz}{i} &(2.11)\\
&= 2\pi i(\mathrm{Res}(f_{\mathrm{ARMA}(1)},-\phi)+\mathrm{Res}(f_{\mathrm{ARMA}(1)},0)) &(2.12)\\
&= 2\pi\frac{1+2\theta\phi+\theta^2}{1-\phi^2}. &(2.13)
\end{aligned}$$

Formula 2.3

$$\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{(1+\theta e^{i\lambda})(1+\theta e^{-i\lambda})}{(1-\phi e^{i\lambda})(1-\phi e^{-i\lambda})}\,d\lambda = \sigma^2\frac{1+2\theta\phi+\theta^2}{1-\phi^2}. \tag{2.14}$$

2.4. AR(2)

$$\begin{aligned}
&\int_{-\pi}^{\pi}\frac{1}{(1-\theta_1 e^{i\lambda}-\theta_2 e^{i2\lambda})(1-\theta_1 e^{-i\lambda}-\theta_2 e^{-i2\lambda})}\,d\lambda &(2.15)\\
= &\int_{|z|=1}\frac{z}{(1+\theta_1 z-\theta_2 z^2)(z^2+\theta_1 z-\theta_2)}\,\frac{dz}{i} &(2.16)
\end{aligned}$$

Note that the roots $z_\pm$ of $z^2+\theta_1 z-\theta_2=0$ lies in the unit circle, and $z_\pm^2=-\theta_1 z_\pm+\theta_2$, then

$$1+\theta_1 z_- -\theta_2 z_-^2 = (1-\theta_2^2)+(1+\theta_2)\theta_1 z_- = (1+\theta_2)(1-\theta_2+\theta_1 z_-). \tag{2.17}$$

Note again that $z_+ + z_- = -\theta_1$ and $z_+ z_- = -\theta_2$,

$$\begin{aligned}
\text{(equation above)} &= 2\pi(\mathrm{Res}(f_{\mathrm{AR}(2)},z_+)+\mathrm{Res}(f_{\mathrm{AR}(2)},z_-)) &(2.18)\\
&= 2\pi\Big(\frac{z_+}{(1-\theta_2)(1-\theta_2+\theta_1 z_+)(z_+-z_-)}+\frac{z_-}{(1-\theta_2)(1-\theta_2+\theta_1 z_-)(z_--z_+)}\Big)\\
&= 2\pi\frac{1-\theta_2}{(1+\theta_2)[\,(1-\theta_2)^2-\theta_1^2\,]}
\end{aligned}$$

Formula 2.4

$$\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{1}{(1-\theta_1 e^{i\lambda}-\theta_2 e^{i2\lambda})(1-\theta_1 e^{-i\lambda}-\theta_2 e^{-i2\lambda})}\,d\lambda = \frac{(1-\theta_2)\sigma^2}{(1+\theta_2)[\,(1-\theta_2)^2-\theta_1^2\,]} \tag{2.19}$$

3. Estimation for model

A way to estimate the coefficients of models is to use the spectrum. Let us think the AR(1) model fitted by AR spectrum $f(\lambda;\theta)$:

$$\frac{\partial}{\partial\theta}\int_{-\pi}^{\pi}\frac{(1+\theta e^{i\lambda})(1+\theta e^{-i\lambda})}{(1+1/2e^{i\lambda})(1+1/2e^{-i\lambda})}\,d\lambda = 0. \tag{3.1}$$

The integrand can be looked as an ARMA model, i.e. the result will be

$$\frac{\partial}{\partial\theta}\frac{1-\theta-\theta^2}{3/4} = 0, \tag{3.2}$$

and then the true value is

$$\theta_0 = \frac{1}{2}. \tag{3.3}$$

4. Some interesting transformation

It is easy to see that

$$\frac{1}{2}(e^{i\theta}+e^{-i\theta}) = \cos\theta. \tag{4.1}$$

Formula 2.1

Let $\theta = iz$, then

$$\cosh z = \frac{1}{2}(e^{-z}+e^{z}) = \cos iz. \tag{4.2}$$

Similarly,

$$\sinh z = -i\sin iz. \tag{4.3}$$

Game

SUMMARY

YAN LIU

References

[1] Chan, N. H., L. Peng, and Y. Qi (2006) "Quantile inference for near-integrated autoregressive time series with infinite variance," *Statistica Sinica*, Vol. 16, No. 1, p. 15.

1. Preliminaries

1.1. notations.

1. $p(x)$	the density of η_1
2. $f(x)$	the density of ϵ
3. $\theta_0(\tau) = (\beta_0(\tau), \gamma_n)^T$	the true value of $\theta(\tau)$

1.2. Model.

- Model

$$Y_i = \gamma_n Y_{i-1} + \epsilon_i,$$

where for a real number γ,

$$\gamma_n = 1 - \gamma/n$$

- Innovations

$$\epsilon_i = \sum_{j=0}^{\infty} c_j \eta_{i-j},$$

where

$$c_k = \begin{cases} 0, & \text{when } k = 0, \\ k^{-\beta} l(k) & \text{when } k \geq 1, \end{cases}$$

and $l(\cdot)$ is a slowly varying function and

$$\beta > 1/\alpha,$$

within the J_1 topology,

$$a_n^{-1} \sum_{i=1}^{[ns]} \eta_i \xrightarrow{J_1} Z_a(s),$$

Date: October 21, 2014.

where for a slowly varying function $L(x)$,

$$a_n = \inf\{x : P(|\eta_0| > x) \leq 1/n\} = n^{1/\alpha} L(n).$$

1.3. Assumptions.

- $|p'(x)| \leq C_1 (1 + |x|)^{-(1+\delta)}$ for some $\delta > \max\{0, \alpha - 1\}$ and all $x \in \mathbb{R}$
- $|p'(x) - p'(y)| \leq C_2 |x - y| (1 + |x|)^{-(1+\delta)}$ for all $x, y \in \mathbb{R}$ with $|x - y| < 1$

2. Main Results

Theorem 2.1. *Assume assumptions above. If $\beta > 2/\alpha$, then*

$$D_n(\hat{\theta}(\tau) - \theta_0(\tau)) \xrightarrow{\mathcal{L}} \frac{\sigma}{f(\beta_0(\tau))} (A(s))^{-1} \big(W(\tau, 1), \int_0^1 S(s) dW(\tau, s) \big)^T,$$

where

$$A(x) = \int_0^1 (1, x(s))^T (1, x(s)) ds.$$

In particular,

$$a_n \sqrt{n} (\hat{\alpha}(\tau) - \gamma_n) \xrightarrow{\mathcal{L}} \frac{\sigma}{f(\beta_0(\tau))} \frac{\int_0^1 S(s) dW(\tau, s) - W(\tau, 1) \int_0^1 S(s) ds}{\int_0^1 S^2(s) ds - \big(\int_0^1 S(s) ds \big)^2}$$

and

$$\Big(\sum_{t=1}^n Y_{t-1}^2 - \big(\sum_{t=1}^n Y_{t-1} \big)^2 \Big)^{1/2} (\hat{\alpha}(\tau) - \gamma_n) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \frac{\sigma^2}{f^2(\beta_0(\tau))}\big),$$

where

$$D_n = \mathrm{diag}(\sqrt{n}, a_n \sqrt{n}),$$

and

$$\sigma^2 = E\psi_\tau^2(\epsilon_0) + 2 \sum_{j=1}^\infty E\psi_\tau(\epsilon_0) \psi_\tau(\epsilon_j),$$

a standard Brownian motion $W(\tau, \cdot)$ of

$$S(s) = \lambda (Z_\alpha(s) - \gamma \int_0^s e^{-\gamma(s-t)} dZ_\alpha(t)),$$

$$\lambda = \sum_{j=0}^\infty c_j.$$

3. Quantile Regression

3.1. **Quantile Regression.**

- **?**
  MLE by assuming that the $v_t's$ are i.i.d. skewed-Laplace distributed with unit scale

- p.d.f.

$$f(v;\delta,\tau)=\frac{\tau(1-\tau)}{\delta}\exp\Big\{-\frac{v}{\delta}(\tau-\mathbb{1}(v<0))\Big\}$$

- In the case of heteroscedasticity, **?** suppose

$$h_t=\mathrm{Var}(v_t|\mathcal{F}_{t-1}).$$

  Assume

$$\epsilon_t=v_t/\sqrt{h_t},$$

  its pdf is

$$g(\epsilon_t;\tau)=\sqrt{1-2\tau+2\tau^2}\exp\Big\{\epsilon_t\frac{\sqrt{1-2\tau+2\tau^2}}{\tau-\mathbb{1}(\epsilon_t\geq 0)}\Big\}.$$

4. SUMMARY

(1) Bayesian inference setting

(2) Model selection with DIC

(3) Two-regime threshold quantile CAPM-GARCH model

$$q_\tau(r_t)=\begin{cases}\phi_0^{(1)}(\tau)+\phi_1^{(1)}(\tau)r_{t-1}+\beta^{(1)}(\tau)r_{m,t}, & \text{if } r_{m,t-d}\leq c(\tau)\\ \phi_0^{(2)}(\tau)+\phi_1^{(2)}(\tau)r_{t-1}+\beta^{(2)}(\tau)r_{m,t}, & \text{if } r_{m,t-d}> c(\tau)\end{cases}$$

$$h_t=\begin{cases}\alpha_0^{(1)}(\tau)+\alpha_1^{(1)}(\tau)a_{t-1}^2+\lambda^{(1)}(\tau)h_{t-1}, & \text{if } r_{m,t-d}\leq c(\tau)\\ \alpha_0^{(2)}(\tau)+\alpha_1^{(2)}(\tau)a_{t-1}^2+\lambda^{(2)}(\tau)h_{t-1}, & \text{if } r_{m,t-d}> c(\tau)\end{cases},$$

where

$$a_{t-1}=r_{t-1}-q_\tau(r_{t-1}).$$

(4) Parameter estimation

SUMMARY

YAN LIU

References

[1] Chen, C. W., R. Gerlach, and D. Wei (2009) "Bayesian causal effects in quantiles: Accounting for heteroscedasticity," *Computational Statistics & Data Analysis*, Vol. 53, No. 6, pp. 1993–2007.
[2] Chen, C. W., S. Lin, and L. Philip (2012) "Smooth transition quantile capital asset pricing models with heteroscedasticity," *Computational Economics*, Vol. 40, No. 1, pp. 19–48.
[3] Koenker, R. and J. A. Machado (1999) "Goodness of fit and related inference processes for quantile regression," *Journal of the american statistical association*, Vol. 94, No. 448, pp. 1296–1310.
[4] Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002) "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 64, No. 4, pp. 583–639.

1. DIC

1.1. notations.

1. $\boldsymbol{r}$ — the set of data
2. θ — the unknown parameters
3. $p(\boldsymbol{r}|\theta)$ — the likelihood of the data $\boldsymbol{r}$

1.2. DIC.

- DIC (Spiegelhalter et al. (2002))

$$D(\theta) = -2 \log p(\boldsymbol{r}|\theta).$$

- Concept
  "goodness of fit" + "model complexity"

- Goodness of fit

$$\bar{D} = E_{\theta|\boldsymbol{r}}[D(\theta)]$$

Date: October 21, 2014.

- Model complexity

$$\begin{aligned} P_D &= E_{\theta|\boldsymbol{r}}[D(\theta)] - D[E_{\theta|\boldsymbol{r}}(\theta)] \\ &= \bar{D} - D(\bar{\theta}) \end{aligned}$$

- パラメータの数が分からないから、それをリスクで評価する。MCMC は使いやすいです。by 伊庭幸人先生

2. CAPM

1. R_t the expected return
2. $R_{m,t}$ the expected market portfolio return
3. $r_{f,t}$ the risk free rate

2.1. CAPM.

- the sensitivity of the expected excess returns on security to expected market risk premium

- Description

$$ER_t - r_{f,t} = \beta(ER_{m,t} - r_{f,t})$$

- Determination of β

$$\beta = \frac{\mathrm{Cov}(R_t - r_{f,t}, R_{m,t} - r_{f,t})}{\mathrm{Var}(R_{m,t} - r_{f,t})}$$

- β_t should change over time
$\rightarrow$ a smooth transition regime switching CAPM with heteroscedasticity

3. Quantile Regression

3.1. Quantile Regression.

- Koenker and Machado (1999)
MLE by assuming that the $v_t's$ are i.i.d. skewed-Laplace distributed with unit scale

- p.d.f.

$$f(v; \delta, \tau) = \frac{\tau(1-\tau)}{\delta} \exp\Big\{-\frac{v}{\delta}(\tau - \mathbb{1}(v < 0))\Big\}$$

- In the case of heteroscedasticity, Chen et al. (2009) suppose
$$h_t = \mathrm{Var}(v_t|\mathcal{F}_{t-1}).$$
Assume
$$\epsilon_t = v_t/\sqrt{h_t},$$
its pdf is
$$g(\epsilon_t;\tau) = \sqrt{1-2\tau+2\tau^2}\exp\Big\{\epsilon_t\frac{\sqrt{1-2\tau+2\tau^2}}{\tau-\mathbb{1}(\epsilon_t \geq 0)}\Big\}.$$

4. SUMMARY

(1) Bayesian inference setting

(2) Model selection with DIC

(3) Two-regime threshold quantile CAPM-GARCH model

$$q_\tau(r_t) = \begin{cases} \phi_0^{(1)}(\tau) + \phi_1^{(1)}(\tau)r_{t-1} + \beta^{(1)}(\tau)r_{m,t}, & \text{if } r_{m,t-d} \leq c(\tau) \\ \phi_0^{(2)}(\tau) + \phi_1^{(2)}(\tau)r_{t-1} + \beta^{(2)}(\tau)r_{m,t}, & \text{if } r_{m,t-d} > c(\tau) \end{cases}$$
$$h_t = \begin{cases} \alpha_0^{(1)}(\tau) + \alpha_1^{(1)}(\tau)a_{t-1}^2 + \lambda^{(1)}(\tau)h_{t-1}, & \text{if } r_{m,t-d} \leq c(\tau) \\ \alpha_0^{(2)}(\tau) + \alpha_1^{(2)}(\tau)a_{t-1}^2 + \lambda^{(2)}(\tau)h_{t-1}, & \text{if } r_{m,t-d} > c(\tau) \end{cases},$$

where
$$a_{t-1} = r_{t-1} - q_\tau(r_{t-1}).$$

(4) Parameter estimation

Yan LIU

Asymptotic Theory in Statistics

0.1 Convolution

Define the convolution $f * g$ as

$$f * g(x) = \int_{-\infty}^{\infty} f(y)g(x-y)\,dy. \tag{0.1.1}$$

Theorem 0.1.1. *Suppose $f(x) \in \mathscr{L}^1(x)$ and $g(x) \in \mathscr{L}^1(x)$. Then*

$$h(x) := (f * g)(x) \in \mathscr{L}^1(x). \tag{0.1.2}$$

Theorem 0.1.2. *Let X and Y be two independent random variables, whose density function is f and g. Then the density function of $X+Y$ is $f * g$.*

note. The distribution of the sum of two random variables can be written in the form of convolution if the two random variables are independent. Conversely, if two random variables are dependent, then joint distribution of two is required to derive the distribution of sum.

There are many algebraic properties for convolution, which is listed below:

1. commutativity
$$f * g = g * f; \tag{0.1.3}$$
2. Associativity
$$f * (g * h) = (f * g) * h; \tag{0.1.4}$$
3. Distributivity
$$f * (g + h) = (f * g) * h; \tag{0.1.5}$$
4. scalar multiplication
$$a(f * g) = (af) * g = f * (ag); \tag{0.1.6}$$
5. identity element
$$f * \delta = f. \tag{0.1.7}$$

The calculations around convolution are:

1. integration
$$\int_{\mathbb{R}^d} f * g(x)\,dx = (\int_{\mathbb{R}^d} f(x)\,dx)(\int_{\mathbb{R}^d} g(x)\,dx); \tag{0.1.8}$$
2. differentiation
$$(f * g)' = f' * g = f * g'. \tag{0.1.9}$$

0.2 Stieltjes Integral

Let $f(x)$ and $g(x)$ be real-valued bounded function defined on a closed interval $[a, b]$. Take a partition of the interval

$$a = x_0 < x_1 < \cdots < x_n = b. \tag{0.2.1}$$

Then the Stieltjes integral is defined as

$$\sum_{i=0}^{n-1} f(\xi_i)(g(x_{i+1}) - g(x_i)) \tag{0.2.2}$$

with $\xi \in [x_i, x_{i+1}]$. If the sum exists uniquely as $\max|x_{i+1} - x_i| \to 0$, then the integral is denoted as

$$\int f(x)\, dg(x). \tag{0.2.3}$$

If f and g have a common point of discontinuity, then the integral does no exist. However, if f is continuous and g' is Riemann integrable over the specified interval, then

$$\int f(x)\, dg(x) = \int f(x) g'(x)\, dx. \tag{0.2.4}$$

In fact, if F_1 and F_2 are distribution functions, then the function F defined by

$$F(x) := \int F_1(x - y)\, dF_2(y) \tag{0.2.5}$$

is called the convolution of distribution functions F_1 and F_2. This is also denoted as $F = F_1 * F_2$.

LONG RANGE DEPENDENCE

YAN LIU

1. Reference

Dahlhaus (1996), Stoch. Proc. their Appl.

2. notations

2.1. Notations.

1. X_j, $j \geq 1$	a stationary Gaussian sequence
2. μ	mean
3. $\sigma^2 f(x, \theta)$	spectral density
4. $E \subset \mathbb{R}^p$	compact
5. $\bar{X}_N$	$= (1/N) \sum_{j=1}^{n} X_j$
6. $\boldsymbol{Z}$	$= (X_1 - \bar{X}_N, \ldots, X_N - \bar{X}_N)'$
7. $A_N(\theta)$	$N \times N$ matrix with entries $[A_N(\theta)]_{jk} = a_{j-k}(\theta)$ below
8. $W(\theta)$	the $p \times p$ matrix with j, kth entry $w_{jk}(\theta)$
9. ξ	$= (\xi_0 \ldots, \xi_r)$
10. ϕ	$= (\phi_0, \ldots, \phi_q)$
11. $\dot{G}$	the derivative of G

3. Fundamental Setting

3.1. Locally Stationary Processes.

Definition 3.1. $\{X_{t,T}\}$ is called *locally stationary*

$$X_{t,T} = \mu\Big(\frac{t}{T}\Big) + \int_{-\pi}^{\pi} \exp(i\lambda t) A^0_{t,T}(\lambda) d\xi(\lambda).$$

3.2. Wigner-Ville spectrum.

$$f_T(u, \lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \mathrm{Cov}(X_{[uT-s/2],T}, X_{[uT+s/2],T}) \exp(-i\lambda s).$$

Date: July 9, 2015.

The spectrum is defined by

$$f(u, \lambda) = |A(u, \lambda)|^2.$$

4. Main Results

Theorem 4.1. *If $X_{t,T}$ is locally stationary and $A(u, \lambda)$ is uniformly Lipschitz continuous in both components with index $\alpha > 1/2$ then we have for all $u \in (0, 1)$,*

$$\int_{-\pi}^{\pi} |f_T(u, \lambda) - f(u, \lambda)|^2 d\lambda = o(1).$$

note. Usually, $f_T(u, \lambda)$ does not converge pointwise to $f(u, \lambda)$.

5. words

1. rather the exception than the rule どちらかといえば例外的で

6. New knowledge

- The compensation effect in the Whittle estimator appears when the observations X_t are pure Gaussian or linear is rather the exception than the rule!!

DISTANCE CORRELATION

YAN LIU

1. Reference

Székely, Rizzo and Bakirov (2007), AS.

2. Definition

2.1. notations.

1. $\boldsymbol{X}$ $n \times p$ of X

2.2. Population version of OGA.

- for $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, define

$$a_{kl} = |X_k - X_l|_p,$$

$$\bar{a}_{k\cdot} = \frac{1}{n}\sum_{l=1}^{n} a_{kl},$$

$$\bar{a}_{\cdot l} = \frac{1}{n}\sum_{k=1}^{n} a_{kl},$$

$$\bar{a}_{\cdot\cdot} = \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl},$$

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot},$$

  $k, l = 1, \ldots, n$.

- the empirical distance covariance $V_n(\boldsymbol{X}, \boldsymbol{Y})$

$$V_n^2(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{n^2}\sum_{k,l=1}^{n} A_{kl}B_{kl}.$$

Date: July 21, 2015.

- The distance covariance

$$V^2(X,Y) = \|f_{X,Y}(t,s) - f_X(t)f_Y(s)\|^2 = \frac{1}{c_p c_q}\int_{\mathbb{R}^{p+q}} \frac{|f_{X,Y}(t,s) - f_X(t)f_Y(s)|^2}{|t|_p^{1+p}|s|_q^{1+q}} dtds.$$

- the empirical distance correlation $R_n(\boldsymbol{X},\boldsymbol{Y})$

$$R_n^2(\boldsymbol{X},\boldsymbol{Y}) = \begin{cases} \frac{V_n^2(\boldsymbol{X},\boldsymbol{Y})}{\sqrt{V_n^2(\boldsymbol{X})V_n^2(\boldsymbol{Y})}}, & V_n^2(\boldsymbol{X})V_n^2(\boldsymbol{Y}) > 0, \\ 0, & V_n^2(\boldsymbol{X})V_n^2(\boldsymbol{Y}) = 0. \end{cases}$$

- the empirical characteristic functions

$$f_{X,Y}^n(t,s) = \frac{1}{n}\sum_{k=1}^{n} \exp\{i\langle t, X_k\rangle + i\langle s, Y_k\rangle\}$$

$$f_X^n(t) = \frac{1}{n}\sum_{k=1}^{n} \exp\{i\langle t, X_k\rangle\},$$

$$f_Y^n(s) = \frac{1}{n}\sum_{k=1}^{n} \exp\{i\langle s, Y_k\rangle\}.$$

3. Main Results

Theorem 3.1. *If $(\boldsymbol{X},\boldsymbol{Y})$ is a sample from the joint distribution of (X,Y), then*

$$V_n^2(\boldsymbol{X},\boldsymbol{Y}) = \|f_{X,Y}^n(t,s) - f_X^n(t)f_Y^n(s)\|^2.$$

Theorem 3.2. *If $E|X|_p < \infty$ and $E|Y|_q < \infty$, then almost surely*

$$\lim_{n\to\infty} V_n(\boldsymbol{X},\boldsymbol{Y}) = V(\boldsymbol{X},\boldsymbol{Y}).$$

Theorem 3.3. (i) *If $E(|X|_p + |Y|_q) < \infty$, then*

$$0 \le R \le 1,$$

and $R(X,Y) = 0$ if and only if X and Y are independent.

Theorem 3.4. *Suppose that the random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ have the joint probability density function*

$$\psi_{X,Y}(x,y) = \psi_X(x)\psi_Y(y) \sum_{n \in C} \rho_n P_n(x) Q_n(y),$$

where C denote a countable index set with a zero element. Then

$$V^2(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{\gamma_p \gamma_q} \sum_{j \in C, j \neq 0} \sum_{k \in C, k \neq 0} \rho_j \bar{\rho}_k \mathcal{A}_{jk} \mathcal{B}_{jk},$$

whenever the sum converges absolutely.

INTRODUCTION TO FOURIER ANALYSIS

GEN RYU

1. Fourier series on the circle

Consider

$$f(\omega) = \sum_{n=0}^{\infty} (A_n \cos n\omega + B_n \sin n\omega),$$

where $\sum_{n=0}^{\infty}(|A_n| + |B_n|) < \infty$. The values of f determined on any interval of length 2π. A standard choice is the interval $\mathbb{T} = (-\pi, \pi]$, where we identify 2π-periodic functions on $\mathbb{R}$ with functions on $\mathbb{T}$. The alternative way to representate it is to rewrite it in the complex form

$$f(\omega) = \sum_{n=-\infty}^{\infty} C_n e^{in\omega}. \tag{1.1}$$

Theorem 1.1. *Suppose that $\sum_{n\in\mathbb{Z}} |C_n| < \infty$. Then f defined by* (1.1) *is a continuous function on $\mathbb{T}$. The coefficients are obtained as*

$$C_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) e^{-in\omega} d\omega, \quad n \in \mathbb{Z}. \tag{1.2}$$

If g is any other L^1 function on $\mathbb{T}$, we have the Fourier reciprocity formula

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) g(\omega) d\omega = \sum_{n\in\mathbb{Z}} C_n D_{-n}$$

where D_n is the Fourier coefficient of g, defined by (1.2) *with f replaced by g. In particular we have Parseval's identity*

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\omega)|^2 d\omega = \sum_{n\in\mathbb{Z}} |C_n|^2.$$

Proposition 1.2. *Suppose that $\sum_{n\in\mathbb{Z}} |n^k C_n| < \infty$ for some $k = 2, 3, \ldots$. Then $f(\omega) := \sum_{-\infty}^{\infty} C_n e^{in\omega}$ is a k-times defferentiable function with $f^{(k)}(\omega) = \sum_{n\in Z} (in)^k C_n e^{in\omega}$ a continuous function.*

Date: June 23, 2012.

Corollary 1.3. *The convolution of an absolutely convergent trigonometric series f with an arbitrary L^1 function g has the representation*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)g(\lambda-\omega)d\omega = \sum_{n\in\mathbb{Z}} C_n D_n e^{in\lambda}.$$

2. Factorial and Bessel Functions

Let $C_n = 0$ for $n \le 0$ and $C_n = r^n/n!$ where $r \ge 0$ and $n = 1, 2, \ldots$. Then we have

$$f(\omega) = \sum_{n=0}^{\infty} \frac{r^n}{n!} e^{in\theta} = \sum_{n=0}^{\infty} \frac{(re^{i\theta})^n}{n!} = \exp(re^{i\theta}),$$

and then

$$\frac{r^n}{n!} = \frac{1}{2\pi}\int_{-\pi}^{\pi} exp(re^{\theta})\exp(-in\omega)d\omega, \quad r \ge 0,\ n = 0, 1, \ldots.$$

Here we define $I(2r)$ as

$$I(2r) = \sum_{n=0}^{\infty}\left(\frac{r^n}{n!}\right)^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} exp(2r\cos\omega)d\omega, \quad r \ge 0.$$

3. integration

Whether $m = n$ or $m \neq n$

$$\int_{-\pi}^{\pi} \cos(m\omega)\sin(n\omega)\,d\omega = 0 \tag{3.1}$$

When $m \neq n$ then

$$\int_{-\pi}^{\pi} \cos(m\omega)\cos(n\omega)\,d\omega = 0 \tag{3.2}$$

$$\int_{-\pi}^{\pi} \sin(m\omega)\sin(n\omega)\,d\omega = 0 \tag{3.3}$$

$$\int_{0}^{\pi} \sin m\omega \sin n\omega\,d\omega = 0 \tag{3.4}$$

When $m = n (\neq 0)$ then

$$\int_{-\pi}^{\pi} \cos(m\omega)\cos(n\omega)\,d\omega = \pi \tag{3.5}$$

$$\int_{-\pi}^{\pi} \sin(m\omega)\sin(n\omega)\,d\omega = \pi \tag{3.6}$$

When $m = n = 0$ then

$$\int_{-\pi}^{\pi} \cos(m\omega)\cos(n\omega)\,d\omega = 2\pi \tag{3.7}$$

$$\int_{-\pi}^{\pi} \sin(m\omega)\sin(n\omega)\,d\omega = 0 \tag{3.8}$$

Proposition 3.1.

Proposition 3.2.

Theorem 3.3.

Theorem 3.4.

Theorem 3.5.

Definition 3.6.

Definition 3.7.

Corollary 3.8.

SUMMARY

YAN LIU

References

[1] Fan, J., C. Zhang, and J. Zhang (2001) "Generalized likelihood ratio statistics and Wilks phenomenon," *Annals of statistics*, pp. 153–193.

1. preliminaries

1.1. notations.

1. $\mathbb{Z}^3$	the 3-dimensional integer lattice
2. $I_n = \{\boldsymbol{i} : \boldsymbol{i} \in \mathbb{Z}^3, 1 \le i_k \le n_k, k = 1, 2, 3\}$	a rectangular region on $\mathbb{Z}^3$
3. $\boldsymbol{i} = (i_1, i_2, i_3)$	a site
4. $(Y_{\boldsymbol{j}}, X_{\boldsymbol{j}}) \in \mathbb{R}^3$, where $\boldsymbol{j} \in I_n$	a random field indexed by $\mathbb{Z}^3$
5. $X_{\boldsymbol{i}} = (X_{1\boldsymbol{i}}, X_{2\boldsymbol{i}})$	observation
6. $g : \boldsymbol{x} \mapsto g(\boldsymbol{x}) := E[\, Y_{\boldsymbol{i}} \vert \boldsymbol{X}_{\boldsymbol{i}} = \boldsymbol{x}\,]$	the spatial regression function $g : \mathbb{R}^2 \to \mathbb{R}$
7. $\phi(x)$	father wavelets (scaling functions)
8. $\psi(x)$	mother wavelets
9. $\boldsymbol{n}_\pi = n_1 n_2 n_3$	
10. $f_{\boldsymbol{X}}(x)$	the spatial marginal density function
11. $2^{j_0} \simeq \ln \boldsymbol{n}_\pi$	
12. $2^{j_1} \simeq \boldsymbol{n}_\pi^{1/2} / (\ln \boldsymbol{n}_\pi)^7$	

1.2. Translations and dilations of wavelets and related results.

For $j_0, j, k \in \mathbb{Z}$,

(i) $\phi_{j_0 k}(x) = 2^{j_0/2} \phi(2^{j_0} x - k)$

(ii) $\psi_{jk}(x) = 2^{j/2} \phi(2^j x - k)$

The bases for $g(x_1, x_2) \in L^2(\mathbb{R}^2)$ are given by

1. $\Phi_{j_0 \boldsymbol{k}}(x_1, x_2) = \phi_{j_0 k_1}(x_1) \phi_{j_0 k_2}(x_2)$
2. $\Psi^{(1)}_{j\boldsymbol{k}}(x_1, x_2) = \phi_{j k_1}(x_1) \psi_{j k_2}(x_2)$
3. $\Psi^{(2)}_{j\boldsymbol{k}}(x_1, x_2) = \psi_{j k_1}(x_1) \phi_{j k_2}(x_2)$
4. $\Psi^{(3)}_{j\boldsymbol{k}}(x_1, x_2) = \psi_{j k_1}(x_1) \psi_{j k_2}(x_2)$

Date: October 16, 2014.

1.3. **2-variate spatial regression function** $g(x_1, x_2)$**.** Assume g belong to a subset of Besov space $B^s_{p,q}$. For $s > 0$ and $1 \le p, q \le \infty$,

$$F^s_{p,q}(M) = \Big\{ g; g \in B^s_{p,q}, \|g\|_{B^s_{p,q}} \le M, s > 2/p, \operatorname{supp} g \subseteq [0,1]^2 \Big\}$$

and for $\boldsymbol{k} = (k_1, k_2)$, $k_1, k_2 \in \mathscr{K} = \{0, 1, \dots, 2^{j_0} - 1\}$,

$$g(x_1, x_2) = \sum_{\boldsymbol{k} \in \mathscr{K}^2} \alpha_{j_0 \boldsymbol{k}} \Phi_{j_0 \boldsymbol{k}}(x_1, x_2) + \sum_{j \ge j_0} \sum_{\boldsymbol{k} \in \mathscr{K}^2} \sum_{l=1}^{3} \beta^{(l)}_{j\boldsymbol{k}} \Psi^{(l)}_{j\boldsymbol{k}}(x_1, x_2)$$

As a result,

$$\begin{aligned} \alpha_{j_0 \boldsymbol{k}} &= \int_{[0,1]^2} g(\boldsymbol{x}) \Phi_{j_0 \boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} \\ \beta^{(l)}_{j\boldsymbol{k}} &= \int_{[0,1]^2} g(\boldsymbol{x}) \Psi^{(l)}_{j_0 \boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} \end{aligned}$$

1.4. **Wavelet-based estimator.** For δ satisfying $\delta^2 = K_0 \ln \boldsymbol{n}_\pi / \boldsymbol{n}_\pi$ with $K_0 > 2C_6$,

$$\hat{g}(x_1, x_2) = \sum_{\boldsymbol{k} \in \mathscr{K}^2} \hat{\alpha}_{j_0 \boldsymbol{k}} \Phi_{j_0 \boldsymbol{k}}(x_1, x_2) + \sum_{j \ge j_0}^{j_1 - 1} \sum_{\boldsymbol{k} \in \mathscr{K}^2} \sum_{l=1}^{3} \hat{\beta}^{(l)}_{j\boldsymbol{k}} \mathbb{1}(|\hat{\beta}^{(l)}_{j\boldsymbol{k}}| > \delta) \Psi^{(l)}_{j\boldsymbol{k}}(x_1, x_2),$$

where

$$\begin{aligned} \hat{\alpha}_{j_0 \boldsymbol{k}} &= \frac{1}{\boldsymbol{n}_\pi} \sum_{\boldsymbol{i} \in I_{\boldsymbol{n}}} \frac{Y_{\boldsymbol{i}} \Phi_{j_0 \boldsymbol{k}}(\boldsymbol{X}_{\boldsymbol{i}})}{f_{\boldsymbol{X}}(\boldsymbol{X}_{\boldsymbol{i}})} \\ \hat{\beta}^{(l)}_{j\boldsymbol{k}} &= \frac{1}{\boldsymbol{n}_\pi} \sum_{\boldsymbol{i} \in I_{\boldsymbol{n}}} \frac{Y_{\boldsymbol{i}} \Psi^{(l)}_{j\boldsymbol{k}}(\boldsymbol{X}_{\boldsymbol{i}})}{f_{\boldsymbol{X}}(\boldsymbol{X}_{\boldsymbol{i}})} \end{aligned}$$

As a result,

$$\begin{aligned} E\hat{\alpha}_{j_0 \boldsymbol{k}} &= \alpha_{j_0 \boldsymbol{k}} \\ E\hat{\beta}^{(l)}_{j\boldsymbol{k}} &= \beta^{(l)}_{j\boldsymbol{k}}. \end{aligned}$$

Remark 1.1. C_6 is a constant from Bernstein's inequality.

2. Main Results

Lemma 2.1 (**?**, Lemma 2.1)**.**

Theorem 2.2. *Let $\hat{g}$ be the wavelet estimator given as above. Then, for all $M \in (0, \infty)$, $s < r$ and $p, q \in [1, \infty]$, there exists a constant C, which does not depend on s, p, q and $\boldsymbol{n}_\pi$, such that*

$$\sup_{g \in F^s_{p,q}(M)} E \int_{[0,1]^2} \Big(\hat{g}(\boldsymbol{x}) - g(\boldsymbol{x}) \Big)^2 d\boldsymbol{x} \le C \ln \boldsymbol{n}_\pi \cdot \boldsymbol{n}_\pi^{-2s/(2s+2)}.$$

2.1. **Notations in the proof.**

1. $j_s = j_s(\boldsymbol{n})$ $\qquad 2^{j_s} \simeq \boldsymbol{n}_\pi^{1/(2s+2)}$
2. $\xi_{\boldsymbol{i}} = \frac{Y_{\boldsymbol{i}}\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X_i})}{f_{\boldsymbol{X}}(\boldsymbol{X_i})} - E\frac{Y_{\boldsymbol{i}}\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X_i})}{f_{\boldsymbol{X}}(\boldsymbol{X_i})}$
3. $\eta_{\boldsymbol{i}}^{(l)} = \frac{Y_{\boldsymbol{i}}\Psi_{j\boldsymbol{k}}^{(l)}(\boldsymbol{X_i})}{f_{\boldsymbol{X}}(\boldsymbol{X_i})} - E\frac{Y_{\boldsymbol{i}}\Psi_{j\boldsymbol{k}}^{(l)}(\boldsymbol{X_i})}{f_{\boldsymbol{X}}(\boldsymbol{X_i})}$

Proof. The proof is constructed by the following 5 steps.

Step 1 $E\int_{[0,1]^2}\Big(\hat{g}(\boldsymbol{x}) - g(\boldsymbol{x})\Big)^2 d\boldsymbol{x} = J_1(j_1) + J_2(\alpha_{j_0\boldsymbol{k}}) + J_3(j_0, j_s) + J_4(j_s, j_1)$.

Step 2 (Lemma 4.7) $J_1(j_1) = o(\boldsymbol{n}_\pi^{-2s/(2s+2)})$.
This is shown by Lemma 4.1 and the definition of Besov space and j_1.

Step 3 (Lemma 4.8) $J_2(\alpha_{j_0\boldsymbol{k}}) = o(\boldsymbol{n}_\pi^{-2s/(2s+2)})$.
This is shown by Lemma 4.3 and the definition of j_0.

Step 4 (Lemma 4.9) $J_3(j_0, j_s) \le C\ln \boldsymbol{n}_\pi \cdot \boldsymbol{n}_\pi^{-2s/(2s+2)}$.
Note that $J_3(j_0, j_s) = J_3^{(1)} + J_3^{(2)} + J_3^{(3)}$. Also divide $J_3^{(1)}$ into

$$J_3^{(1)} = J_{311} + J_{312} + J_{32}.$$

The bound of J_{311} is shown by the definition of δ and j_s.
The bound of J_{312} is shown by Lemma 4.6 and the definition of j_0.
The bound of J_{32} is shown by the definition of j_s.
The arguments for both $J_3^{(2)}$, $J_3^{(3)}$ are the same.

Step 5 (Lemma 4.10) $J_4(j_s, j_1) \le C\ln \boldsymbol{n}_\pi \cdot \boldsymbol{n}_\pi^{-2s/(2s+2)}$.
Note that $J_4(j_s, j_1) = J_4^{(1)} + J_4^{(2)} + J_4^{(3)}$. Also divide $J_4^{(1)}$ into

$$J_4^{(1)} = J_{411} + J_{412} + J_{421} + J_{422}.$$

The bound of J_{411} is shown by the definition of j_s for the case of $p \ge 2$ and by the definition of δ and j_s for the case of $1 \le p < 2$.
The bound of J_{412} is shown by Lemma 4.6.
The bound of J_{421} is shown by Lemma 4.3 for the case of $1 \le p < 2$ and by Lemma 4.3 and the definition of δ.
The bound of J_{422} is shown by the result of Lemma 4.4, Lemma 4.5.
The arguments for both $J_4^{(2)}$, $J_4^{(3)}$ are the same.

□

2.2. **Complement.**

(i) Lemma 4.1 is given in Cai (1999).
(ii) Lemma 4.2 is given in Tran (1990).
(iii) Lemma 4.4 is given in Carbon et al. (1997).
(iv) Lemma 4.5 is given in Härdle et al. (1998, p.243).
(v) Lemma 4.3 is shown under Assumption (2.1) and (A1)-(A3).

(vi) Lemma 4.6 is shown by the method of large blocks and small blocks as Tran (1990) and Hallin et al. (2004). The result depends on Lemma 4.4 of independent approximation and Lemma 4.5 of Bernstein's inequality.

3. Three Gaussian models (GWN, NR, GS)

3.1. notations.

1. φ_j	an orthonormal basis in $L_2[0,1]$
2.	
3. $I_n = \{\boldsymbol{i} : \boldsymbol{i} \in \mathbb{Z}^3, 1 \le i_k \le n_k, k = 1,2,3\}$	a rectangular region on $\mathbb{Z}^3$
4. $\boldsymbol{i} = (i_1, i_2, i_3)$	a site
5. $(Y_{\boldsymbol{j}}, X_{\boldsymbol{j}}) \in \mathbb{R}^3$, where $\boldsymbol{j} \in I_n$	a random field indexed by $\mathbb{Z}^3$
6. $X_{\boldsymbol{i}} = (X_{1\boldsymbol{i}}, X_{2\boldsymbol{i}})$	observation
7. $g : \boldsymbol{x} \mapsto g(\boldsymbol{x}) := E[\,Y_{\boldsymbol{i}} \vert \boldsymbol{X}_{\boldsymbol{i}} = \boldsymbol{x}\,]$	the spatial regression function $g : \mathbb{R}^2 \to \mathbb{R}$
8. $\phi(x)$	father wavelets (scaling functions)
9. $\psi(x)$	mother wavelets
10. $\boldsymbol{n}_\pi = n_1 n_2 n_3$	
11. $f_{\boldsymbol{X}}(x)$	the spatial marginal density function
12. $2^{j_0} \simeq \ln \boldsymbol{n}_\pi$	
13. $2^{j_1} \simeq \boldsymbol{n}_\pi^{1/2} / (\ln \boldsymbol{n}_\pi)^7$	

3.2. classification.

$$dY(t) = f(t)dt + \epsilon dW(t), \qquad t \in [0,1], \qquad 0 < \epsilon < 1, f : [0,1] \to \mathbb{R}. \tag{1}$$

$$y_j = \theta_j + \epsilon \xi_j, \qquad j = 1, 2, \ldots. \tag{2}$$

$$Y_i = f(i/n) + \xi_i, \qquad i = 1, \ldots, n, \qquad \text{where } \xi \sim \text{i.i.d. } \mathcal{N}(0,1). \tag{3}$$

The three models above are respectively,

(1) Gaussian White Noise (GWN)
(2) Gaussian Sequence (GS)
(3) Nonparametric Regression (NR).

3.3. (1) → (2).

$$\begin{aligned} y_j &= \int_0^1 \varphi_j(t) dY(t), \\ \theta_j &= \int_0^1 f(t)\varphi_j(t) dt, \\ \xi_j &= \int_0^1 \varphi_j(t) dW(t). \end{aligned}$$

The other two relationships are also valid under some conditions which mainly need a good approximation.

3.4. **Projection Estimator.**

(i) $n \geq 1$,
(ii) the statistic S is

$$\hat{f}_n(x) = \sum_{j=1}^{n} \hat{\theta}_j \varphi_j(x).$$

S is called a *projection estimator* of the regression function f at the point x.
note. In time series analysis,

$$\begin{aligned} x &\rightarrow \lambda, \\ \hat{\theta}_j &\rightarrow \hat{\gamma}(j), \\ \varphi_j(x) &\rightarrow e^{ij\lambda}. \end{aligned}$$

Remark 3.1. The main difference between the trigonometric basis and wavelet bases consists in the fact that the trigonometric bases "localizes" the function f in the frequency domain only, while the wavelet bases "localize" it both in the frequency domain and time domain if we interpret x as a time variable and the index j corresponds to frequency and k characterizes position in time.

4. WORDS

1. synthetic	総合的な
2. exposition	説明
3. extraordinary	ずばぬけた
4. preparatory	予備的な
5. an extensive literature	多方面にわたる文献
6. the references therein	その中にある文献
7. for the sake of the clarity of exposition	説明のわかりやすさの為に
8. denote constants whose values	are unimportant and may vary from line to line
9. circumvent	出し抜く

SUMMARY

YAN LIU

References

[1] Fujisawa, H. and S. Eguchi (2008) "Robust parameter estimation with a small bias against heavy contamination," *Journal of Multivariate Analysis*, Vol. 99, No. 9, pp. 2053–2081.

1. preliminaries

1.1. notations.

1. γ — a positive constant
2. $f(x)$ — the underlying probability density function
3. $\delta(x)$ — the contamination probability density function related to outliers
4. $g(x)$ — the contaminated probability density function
5. $f_\theta(x)$ — a parametric probability density function
6. $x_1, \dots, x_n$ — the observations
7. $\hat{\theta}$ — the estimator of the parameter θ
8. x^* — an outlier

1.2. preliminaries.

(i) The observations which are draw from

$$g(x) = (1 - \epsilon)f(x) + \epsilon\delta(x).$$

1.3. assumptions.

$$\nu_f = \left\{ \delta(x) f(x)^{\gamma_0} dx \right\}^{1/\gamma_0}$$

is sufficiently small for an appropriately large $\gamma_0 > 0$.

Remark 1.1. If $\delta(x)$ is the Dirac function at x^*, then the assumption is

$$\nu_f = f(x^*).$$

Date: October 18, 2014.

1.4. **cross entropy.**

(i) cross entropy

$$d_\gamma(g,f) = -\log\Big[\big\{\int g(x)f(x)^\gamma dx\big\}^{1/\gamma}/\big\{f(x)^{1+\gamma}\big\}^{1/(1+\gamma)}\Big].$$

(ii) divergence

$$D_\gamma(g,f) = d_\gamma(g,f) - d_\gamma(g,g).$$

(iii) the robust estimator

$$\hat{\theta}_\gamma = \arg\min_\theta d_\gamma(\hat{g}, f_\theta).$$

(iv) the minimizer between the observation and the parametric model

$$\theta^*_\gamma = \arg\min_\theta d_\gamma(g, f_\theta).$$

(v) the minimizer between the true model and the parametric model

$$\theta^* = \arg\min_\theta d_\gamma(f, f_\theta).$$

(vi) Restricted parameter space

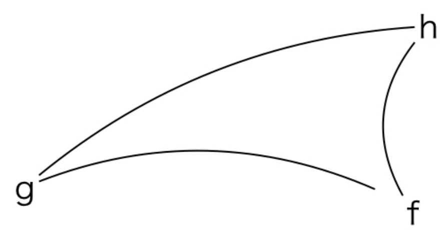
$$\Omega_{\nu_\omega} = \{\theta; \nu_{f_\theta} \le \nu_\omega\}$$

(vii) Another cross entropy considered by Basu et al.

$$m_\beta(g,f) = -\frac{1}{\beta}\int g f^\beta dx + \frac{1}{1+\beta}\int f^{1+\beta} dx.$$

(viii) the estimator

$$\hat{\theta}^{(m)}_\beta = \arg\max m_\beta(g, f_\theta).$$



1.5. **idea.** Hope for robust parameter estimation that the bias $\theta^*_\gamma - \theta^*$ is sufficiently small.

2. Main Results

Theorem 2.1 (**?**, Theorem 3.2)**.** *Suppose*

$$\nu = \max\{\nu_f, \nu_h\}.$$

Then

$$D_\gamma(g,h) - D_\gamma(g,f) - D_\gamma(f,h) = O(\epsilon\nu^\gamma).$$

Proof. Note that

$$d_\gamma(f,h) - d_\gamma((1-\epsilon)f,h) = \frac{1}{\gamma}\log(1-\epsilon) + O(\epsilon\nu_h^\gamma).$$

□

Remark 2.2. (i) If γ is too small, then ν^γ not small.

Theorem 2.3 (**?**, Theorem 3.3)**.** *Suppose*

$$\theta \in \Omega_{\nu_\omega}.$$

Then

$$\theta_\gamma^* = \theta^* + O(\epsilon\nu^\gamma).$$

Theorem 2.4 (**?**, Theorem 5.1)**.**

$$\sqrt{n}\big(\hat{\theta}_\gamma - \theta_\gamma^*\big) \to \mathcal{N}(0, \Sigma_g(\theta_\gamma^*)),$$

where

$$\Sigma_g(\theta) = J_g(\theta)^{-1} I_g(\theta) J_g(\theta)'^{-1}.$$

Here,

$$\begin{aligned} J_g(\theta_\gamma^*) &= (1-\epsilon)J_f(\theta^*) + O(\epsilon\nu^\gamma), \\ I_g(\theta_\gamma^*) &= (1-\epsilon)I_f(\theta^*) + O(\epsilon\nu^\gamma). \end{aligned}$$

Theorem 2.5 (**?**, Theorem 6.1)**.** *Under the cross entropy*

$$d(g,f) = \psi\Big(\int g\chi(f)dx, \int \rho(f)dx\Big),$$

where $\psi(u,v)$, $\chi(s)$ and $\rho(s)$ are twice differentiable real-valued functions. Further assume:

(i) $d(\lambda g, f)$ is uniquely minimized at $f = g$ for any $\lambda > 0$, the Hessian is positive-definite.
(ii) $\chi(0) = 0$
(iii) $\{s\chi'(s)\}' > 0$?

Then there exists a monotone increasing real-valued function ϕ such that

$$d(g,f) = \phi(d_\gamma(g,f)).$$

3. Comment in the paper

We expect the robust estimate to have a small bias when the influence function is redescending, but this is not always clear in the case of heavy contamination. This paper clearly shows that the robust estimate $\hat{\theta}_\gamma$ has a small bias even in the case of heavy contamination.

We often suppose that $\epsilon < 1/2$. Some results obtained in this paper seem to hold even for $\epsilon \geq 1/2$. This is not unreasonable because the underlying density f is always the object in interest in this paper.

One of the remaining problem is how to set a tuning parameter γ. Basu et al. said that there could be no universal way of selecting an appropriate tuning parameter when we used the cross entropy. They persisted in given priority to either robustness or efficiency.

$\Rightarrow$ Future Work.

4. WORDS

1. synthetic	総合的な
2. exposition	説明
3. extraordinary	ずばぬけた
4. preparatory	予備的な
5. an extensive literature	多方面にわたる文献
6. the references therein	その中にある文献
7. for the sake of the clarity of exposition	説明のわかりやすさの為に
8. denote constants whose values	are unimportant and may vary from line to line
9. circumvent	出し抜く

HALLIN AND WERKER(2003)

GEN RYU

1. DEFINITIONS

1.1. **somewhere parametrically efficiency.** A method is called somewhere parametrically efficient if it is efficient in the parametric model induced by some f_0. If the method is not parametrically efficient at some f_0, but at any f, then it is called parametrically efficient.

1.2. **somewhere semi-parametric efficiency.** Methods that are semi-parametrically efficient at some f_0 are called somewhere semi-parametrically efficient, and a method that is semi-parametrically efficient at any f is simply called semi-parametrically efficient.

1.3. **adaptive.** The parametrically efficient method is also called adaptive. When the parametric and semi-parametric lower bounds coincide at some f_0, the model is called somewhere adaptive. If the two bounds coincide for all f, the model is called adaptive.

2. ASSUMPTIONS

Assumption 2.1 (Assumption A). *For any bounded sequence* $(\boldsymbol{\tau}_n)$ *in* $\mathbb{R}^k$, *we have*

$$\frac{dP^{(n)}_{\boldsymbol{\theta}_0+\boldsymbol{\tau}/\sqrt{n},\phi_0}}{dP^{(n)}_0}$$

Date: May 15, 2012.

MATERIALS FOR 5/12

GEN RYU

Part 1. Definitions

1. UNIMODALITY

1.1. the definition.

Definition 1.1. *If $-\log f(x)$ is a convex function within some open interval (a, b) such that $-\infty \le a < b \le \infty$ and $\int_a^b f(x) = 1$, the density $f(x)$ is called strongly unimodal.*

Such densities are absolutely continuous within (a, b) and

$$[-\log f(x)]' = -\frac{f'(x)}{f(x)}$$

is a non-decreasing function. Another interpretation of unimodality is that a unimodal probability distribution is a probability distribution which has a single mode. A mode of a discrete probability distribution is a value at which the probability mass function takes its maximum value. On the other hand, a mode of a continuous probability distribution is a value at which the probability density function attains its maximum value.

The unimodality is also defined in the way that the cdf of the distribution is convex for $x < m$ and concave for $x > m$.

1.2. Properties of unimodality.

- A first important result is Gaussian's inequality; Gaussian's inequality gives an upper bound on the probability that a value lies more than any given distance from its mode.
- Another result is Vysochanskii-Petunin inequality. The inequality is a extension of Chebychev's inequality, and it is more accurate than than the latter.

Theorem 1.2. *A convolution of two strongly unimodal densities is again a strongly unimodal densities.*

Lemma 1.3. *If $f(x)$ is strongly unimodal, then $\varphi(u, f)(\equiv -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))})$ is non-decreasing.*

Corollary 1.4. *The density of Stable is unimodal.*

Date: May 15, 2012.

2. Gaussian log-likelihood and Green's function p.10

The Green's function associated with $\mathbf{\Delta}^{(n)}_{\mathcal{G};\boldsymbol{\theta}}$ is characterized by

$$\left(1-\sum_{i=1}^{p}\theta_i z^i\right)^{-1} \equiv \sum_{u=0}^{\infty} g_u(\boldsymbol{\theta}) z^u \quad |z|<1.$$

Using the notation, the asymptotic covariance matrix $\mathbf{\Gamma}_{\mathcal{G}(\boldsymbol{\theta})}$ is the autocovariance matrix of order p of the stationary AR(p) solution of $\mathbf{\Delta}^{(n)}_{\mathcal{G};\boldsymbol{\theta}}$ under standard normal innovations, and is denoted by

$$\mathbf{\Gamma}_{\mathcal{G}(\boldsymbol{\theta})} = \left(\sum_{u=0}^{\infty} g_u(\boldsymbol{\theta}) g_{u+|i-j|}(\boldsymbol{\theta})\right),$$

which the term in the parentheses is the (i,j) element.

Part 2. Optimal testing for semi-parametric AR models

Let $(X_{-p+1},\dots,X_0,X_1,\dots,X_t,\dots,X_n)'$ be an observed series of length $n+p$. Throughout the paper, we assumer that $\boldsymbol{X}^{(n)}=(X_1,\dots,X_n)'$ satisfies the stochastc difference equation (AR(p) model)

$$X_t-\sum_{i=1}^{p}\theta_i X_{t-i}=\epsilon_t, \quad t=0,\pm1,\pm2,\dots,$$

where $\{\epsilon_t\}, t=0,\pm1,\pm2,\dots$ is an i.i.d. sequence with mean zero antdprobability density f.

It is also assumed that the parameter $\boldsymbol{\theta}=(\theta_1,\dots,\theta_p)'\in\mathbb{R}^p$ is such that all the roots of the characteristic polynomial

$$\theta(z)=1-\sum_{i=1}^{p}\theta_i z^i, \quad z\in\mathbb{C}$$

lie out side the unit disk.

3. linear hypotheses

The null hypotheses we are interested in are the linear hypotheses, under which $\boldsymbol{\theta}$ belongs to some linear restriction of Θ or, equivalently, satisfies some given set of linear constraints. Such hypotheses are characterized by a $p\times r$ matrix Ω, of full rank $r\le p$, and by an element $\boldsymbol{\theta}_0$ of $\mathbb{R}^p$: denoting by $\mathcal{M}(\Omega)$ the r-dimensional linear subspace of $\mathbb{R}^p$ spanned by the columns of Ω, we consider the hypothesis under which $\boldsymbol{\theta}-\boldsymbol{\theta}_0$ belongs to $\Theta\cap\mathcal{M}(\Omega)$, and thus satisfies a set of $p-r$ linearly independent linear constraints on $\boldsymbol{\theta}$. we tacitly assume that either $\boldsymbol{\theta}_0=\mathbf{0}$, or $\boldsymbol{\theta}_0\in\mathbb{R}^p\setminus\mathcal{M}(\Omega)$.

3.1. **Explanation.** Here, Θ and $\mathcal{M}(\Omega)$ is p-dimensional subspace of $\mathbb{R}^p$. You can see that $\mathcal{M}(\Omega)$ is a space of linear map, and the easy way to understand is to think

$$\Omega'(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \in \mathbb{R}^r.$$

This means that the strain is given by Ω and $\Omega'\boldsymbol{\theta}_0$. The virtue of this representation is that $\Omega'(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ is ??.

3.2. **Hypothesis.** we denote by $H_f^{(n)}(\boldsymbol{\theta}_0; \Omega)$ the linear hypothesis characterized by $\boldsymbol{\theta}_0$ and Ω. This is a representation depending on the density f in a semi-parametric model. We will overcome this obstacle by using notation as

$$H^{(n)}(\boldsymbol{\theta}_0; \Omega) = \{P^{(n)}_{f;\boldsymbol{\theta}} | \, f \in \mathcal{F}; \boldsymbol{\theta} - \boldsymbol{\theta}_0 \in \Theta \cap \mathcal{M}(\Omega)\} = \bigcup_{f \in \mathcal{F}} H_f^{(n)}(\boldsymbol{\theta}_0; \Omega)$$

4. SAMPLE SPLITTING DEVISE

Let the number of the observations $Z_t(\boldsymbol{\theta})$ be $2n$. Split the sample into the two groups half and half. Then estimate f from the sample in each group and use the estimated $\hat{f}$ to compute the central sequence for the other group. The statement can be shown as the splits respectively,

$$n^{-1/2} \sum_{t=n+1}^{2n} \phi_{\hat{f}_1^{(n)}}(Z_t(\boldsymbol{\theta}))\boldsymbol{W}_{t-1};$$

$$n^{-1/2} \sum_{t=1}^{n} \phi_{\hat{f}_2^{(n)}}(Z_t(\boldsymbol{\theta}))\boldsymbol{W}_{t-1}.$$

5. ASSUMPTIONS AND THEOREMS

5.1. **Sets of assumptions.**

(A1) $f(x) > 0$, $x \in \mathbb{R}$; $\int_{-\infty}^{\infty} x f(x)\, dx = 0$; $\int_{-\infty}^{\infty} x^2 f(x)\, dx = \sigma^2 < \infty$;

(A2) f is absolutely continuous on finite intervals, i.e., there exists $\dot{f}$ such that for all $-\infty < a < b < \infty$, $f(b) - f(a) = \int_a^b \dot{f}(x)\, dx$;

(A3) letting $\phi_f \equiv -\frac{\dot{f}}{f}$, the generalized Fisher information $\int_{-\infty}^{\infty} \phi_f^2(x) f(x)\, dx \equiv \mathcal{I}_f = \sigma^{-2}\mathcal{I}_{f_1}$ is finite.

(A4) the score function ϕ_f is piecewise Lipschitz, i.e., there exist a finite partition of $\mathbb{R}$ into nonoverlapping intervals $J_1, \dots, J_k$ and a constant A_f such that

$$|\phi_f(x) - \phi_f(y)| \le A_f |x - y| \qquad \forall x, y \in J_i, \forall i = 1, \dots, k.$$

(A5) f is strongly unimodal, i.e., ϕ_f is monotone increasing.

note. All stable densities are unimodal. *note2.* $(A5) \Rightarrow (A2)$.

5.2. **Interpretation of the assumptions.**

(A1)—(A3) The LAN result holds under these assumptions;

(A4) This assumption induces that the influence of starting values on residual autocorrelations is to be asymptotically negligible.

(A5) This assumption is given meaning to by the Proposition below.

The more general assumption to have the initial joint distribution to be negligible is given in Kreiss (1987).

Proposition 5.1. *Let the densities f and g both satisfy assumptions (A1)-(A3); assume moreover that f satisfies (A5). Then, under $H_g^{(n)}(\boldsymbol{\theta})$, as $n \to \infty$,*

$$\overline{r}_{f;u}^{(n)}(\boldsymbol{\theta}) = \left\{ (n-u)^{-1} \sum_{t=u+1}^{n} \phi_{f_1} \circ F_1^{-1}(G(\epsilon_t)) F_1^{-1}(G(\epsilon_{t-u})) \right\} / \mathcal{I}_{f_1}^{1/2} + o_P(n^{-1/2})$$

with $G(x) \equiv \int_{-\infty}^{x} g(z)\, dz$.

Corollary 5.2. *Let the densities f and g both satisfy assumptions (A1)-(A3); assume moreover that f satisfies (A5). Then, under $H_g^{(n)}(\boldsymbol{\theta})$, as $n \to \infty$, any k-tuple*

$$\left((n-i_1)^{1/2} \overline{r}_{f;i_1}^{(n)}(\boldsymbol{\theta}), \ldots, (n-i_k)^{1/2} \overline{r}_{f;i_k}^{(n)}(\boldsymbol{\theta}) \right)'$$

is asymptotically $\mathcal{N}(\mathbf{0}, I_{p\times p})$.

5.3. **theorems for adaptive rank tests.**

Proposition 5.3. *Assume that f and g both satisfy (A1)-(A4); assume that f moreover satisfies (A5). For all $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p) \in \mathbb{R}^p$, consider the sequence $a_u(\boldsymbol{\tau};\boldsymbol{\theta})$ characterized by*

$$\sum_{u=1}^{\infty} a_u(\boldsymbol{\tau};\boldsymbol{\theta}) z^u \equiv \frac{\sum_{i=1}^{p} \tau_i z^i}{1 - \sum_{i=1}^{p} \theta_i z^i} = (\sum_{i=1}^{p} \tau_i z^i)(\sum_{u=0}^{\infty} g_u(\boldsymbol{\theta}) z^u), \quad |z| < 1.$$

Then,

$$n^{1/2}[\, \overline{r}_{f;u}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}) - \overline{r}_{f;u}^{(n)}(\boldsymbol{\theta})\,] = \overline{\sigma}(f;g)\overline{\mathcal{I}}(f;g)(\mathcal{I}_{f_1})^{-1/2} a_u(n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta});\boldsymbol{\theta})) + o_P(1)$$

and

$$\overline{\boldsymbol{\Delta}}_{f;\hat{\boldsymbol{\theta}}^{(n)}}^{(n)} - \overline{\boldsymbol{\Delta}}_{f;\boldsymbol{\theta}}^{(n)} = \overline{\sigma}(f;g)\overline{\mathcal{I}}(f;g)(\mathcal{I}_{f_1})\Gamma_{\mathcal{G}}(\boldsymbol{\theta}) n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}) + o_P(1)$$

under $H_g^{(n)}(\boldsymbol{\theta})$, as $n \to \infty$, where

$$\overline{\sigma}(f;g) \equiv \int_0^1 F_1^{-1}(u) G_1^{-1}(u)\, du$$

and

$$\overline{I}(f;g) \equiv \int_0^1 \phi_{f_1}(F_1^{-1}(u)) \phi_{g_1}(G_1^{-1}(u))\, du;$$

$\Gamma_{\mathcal{G}}(\boldsymbol{\theta})$ and $\mathcal{I}_f$ are defined in (5) and (1), respectively.

note. From this proposition, neither $\bar{r}^{(n)}_{f;u}(\hat{\boldsymbol{\theta}}^{(n)})$ nor $\bar{\Delta}^{(n)}_{f;\hat{\boldsymbol{\theta}}^{(n)}}$ are asymptotically invariant, since the right-hand sides of equations both depend on the unspecified under lying density g.

After the proposition above, we can have a more general statement as follows.

Proposition 5.4. *Let the assumption of Proposition 5.2 be satisfied. Suppose that the estimator $\hat{f}^{(n)}$ of f_1 is based on the order statistics of the residuals $Z_t(\boldsymbol{\theta})$ and is consistent in the sense that*

$$E_{f_1}\left\{\left[\phi_{f_1}\circ F_1^{-1}\left(\frac{R_t^{(n)}}{n+1}\right)F_1^{-1}\left(\frac{R_{t-u}^{(n)}}{n+1}\right)-\phi_{\hat{f}^{(n)}}\circ(\hat{F}^{(n)})^{-1}\left(\frac{R_t^{(n)}}{n+1}\right)(\hat{F}^{(n)})^{-1}\left(\frac{R_{t-u}^{(n)}}{n+1}\right)\right]^2\middle|\hat{f}^{(n)}\right\} = o_P(1),$$

where $\hat{F}^{(n)}$ denotes the distribution function associated with $\hat{f}^{(n)}$. Then, under $H_f^{(n)}(\boldsymbol{\theta})$,

$$\bar{r}^{(n)}_{\hat{f}^{(n)};u}(\boldsymbol{\theta}) = r^{(n)}_{f;u}(\boldsymbol{\theta}) + o_P(n^{-1/2}).$$

Part 3. Reference

- Optimal testing for semi-parametric AR models from Gaussian Lagrange Multipliers to Autoregression Rank Scores and Adaptive Tests (2006)
- Unimodality wikipedia
- Theory of Rank tests

MATERIALS FOR 5/19 (THE FIRST PART TO HALLIN&WERKER)

GEN RYU

REFERENCE

(1) [article] Kreiss(1987)
On Adaptive Estimation in Stationary ARMA Process
(2) [article] Hallin and Werker(2003)
Semi-parametric efficiency, distribution-freeness and invariance
(3) [article] Hallin and Werker(1999)
Optimal testing for semiparametric AR models: from Lagrange multipliers to autoregression rank scores and adaptive tests
(4) [article] Fabian and Hannan(1982)
On estimation and adaptive estimation for locally asymptotically normal families
(5) [book] Bickel et al.(1998)
Efficient and Adaptive Estimation for Semiparametric Models

1. ARMA Model

The process $\{X_t; t \in \mathbb{Z}\}$ which satisfies

$$X_t - \sum_{i=1}^{p} a_i X_{t-i} = e_t + \sum_{i=1}^{q} b_i e_{t-i} \quad \text{for all } t \in \mathbb{R},$$

are concerned in the paper. Denote the density function of e_t by $f(x)$, and the common distribution of $(e_{1-q}, \ldots, e_0; X_{1-p}, \ldots, X_n)$ by $g_n(\cdot;\theta)$, where $\theta \in \Theta$ is the underlying parameter.

note. one can see that

$$g_n(\cdot;\theta) = g_0(e_{1-q}, \ldots, e_0; X_0; \theta) \prod_{t=1}^{n} f(e_t\{e_{1-q}, \ldots, X_t\}),$$

where

$$e_t\{e_{1-q}, \ldots, X_t\} = \sum_{k=1}^{t} \beta_{k-1} \left(- \sum_{i=0}^{p} a_i X_{t+1-k-i} \right) + \sum_{s=0}^{q-1} e_{-s} \left(\sum_{k=0}^{s} \beta_{t+s-k} b_k \right).$$

Date: May 19, 2012.

note2. Using the representation of the common distribution above and Lemma 2.2 in Appendix, we have

$$\frac{dP_{n,\theta}}{dP_{n,\theta_0}} = \frac{g_0(e_{1-q},\ldots,X_0;\theta)}{g_0(e_{1-q},\ldots,X_0;\theta_0)} \prod_{j=1}^{n} \frac{f(e_j^0 - (\theta-\theta_0)' Z(j-1;\theta,\theta_0))}{f(e_j^0)}.$$

With the following additional abbreviation,

$$\phi_j^2(\theta_0,\theta) = \frac{f(e_j(\theta_0) - (\theta-\theta_0)' Z(j-1;\theta,\theta_0))}{f(e_j(\theta_0))},$$

we have

$$\log \frac{dP_{n,\theta}}{dP_{n,\theta_0}} = \log \frac{g_0(e_{1-q},\ldots,X_0;\theta)}{g_0(e_{1-q,\ldots,X_0};\theta_0)} + 2\sum_{j=1}^{n} \log \phi_j(\theta_0,\theta).$$

2. Definitions

2.1. **LAN.** The LAN property here is defined the same as Fabian and Hannan(1982). LAN $\langle \theta, M_n, \gamma_n \rangle$ holds if, for each $n \in \mathbb{Z}$, $\theta \in \Theta_n$, M_n is a $k \times k$ positive definite matrix, γ_n a k-dimensional random vector on $\boldsymbol{X}_n$ such that

$$E_{n,\theta}^{\gamma_n} \xrightarrow{d} \mathcal{N}.$$

and if, for each bounded sequence $\langle t_n \rangle \subset \mathbb{R}^k$, $\delta_n = \theta + M_n^{-1/2} t_n$ is eventually in Θ_n, and

$$g_n \in dE_{n,\delta_n}/dE_{n,\theta}$$

implies

$$g_n/\rho_{t_n}(\gamma_n) \to 1 \quad \text{in } \langle E_{n,\theta} \rangle\text{-prob.}$$

Here, $\mathcal{N}$ denotes the integral with respect to the normal $(0, I)$, and $\rho_t(x) = e^{t'x - ||t||^2/2}$.

note. The convergence in distribution at first shows the normality and the statement under the convergence in distribution shows the two density is contiguity.

2.2. **LAM.** [Fabian and Hannan(1982) page 463] If Condition LAN $\langle \theta, M_n, \gamma_n \rangle$ holds then $\langle Z_n \rangle$ is LAM(θ) (locally asymptotically minimax at θ) if $\langle Z_n \rangle$ is a sequence of estimates for which

$$\lim_{k\to\infty} \lim_{n\to\infty} \sup_{||M_n^{1/2}(\delta-\theta)|| \le K} E_{n,\delta} l(Q_n M_n^{1/2}(Z_n - \delta)) = \mathcal{N} l$$

holds for every sequence $\langle Q_n \rangle$ in the collection of all orthogonal $k \times k$ matrices and for every bounded loss function l on $\mathbb{R}^k$.

Definition 2.1. *A sequence $\langle Z_n \rangle$ of estimates is called regular(θ) if*

$$M_n^{1/2}(Z_n - \theta) - \gamma_n \to 0 \quad \textit{in } \langle E_{n,\theta} \rangle\textit{-prob.}$$

note. γ_n is the part which converges to the standard normal distribution in the LAN.

Theorem 2.2. *Let $\langle Z_n \rangle$ be a sequence of estimates. Then the regularity(θ) of $\langle Z_n \rangle$ implies*

$$E_{n,\delta_n}^{M_n^{1/2}(Z_n-\delta_n)} \Rightarrow \mathcal{N}$$

for every sequence $\delta_n = \theta + M_n^{-1/2} t_n$ such that $\langle t_n \rangle$ is bounded; the latter property, in turn, implies that $\langle Z_n \rangle$ is LAM (θ).

2.3. **Discrete sequences of estimators.** [Bickel(1982)] The discrete sequences of estimators $\{\bar{\theta}_n\}$ satisfies that $\bar{\theta}_n$ is given by one of the vertices of $\{\theta : \theta = n^{-1/2}(i_1, \ldots, i_{p+q}), i_j \in Z\}$ nearest to θ_n, which is a sequence with

$$\sqrt{n}(\theta_n - \theta_0) \quad \text{is bounded by a constant } c > 0.$$

This idea is due to Le Cam(1960), (1969), (1970) for construction of an efficient estimator.

Theorem 2.3. *If $\boldsymbol{P} = \{P_\theta; \theta \in \Theta\}$ is a regular parametric model on a Euclidean space $\boldsymbol{X}$ and θ is identifiable, then there exist uniformly $\sqrt{n}$-consistent estimates of θ.*

The steps are as follows:

(1) Construct $\tilde{\theta}_n$ uniformly $\sqrt{n}$-consistent as in theorem 2.3 below.
(2) Form a grid of cubes with sides of length $cn^{-1/2}$ over $\mathbb{R}^k$, given $\tilde{\theta}_n$, define θ_n^* to be the midpoint of the cube into which $\tilde{\theta}_n$ fallen. (This means that θ_n^* is also uniformly $\sqrt{n}$ consistent.)
(3) Define

$$\hat{\theta}_n = \theta_n^* + n^{-1} \sum_{i=1}^{n} I^{-1}(\theta) \dot{\boldsymbol{l}}(X_i, \theta_n^*).$$

Theorem 2.4. *If $\boldsymbol{P}$ is a regular parametric model and if there exists a uniformly $\sqrt{n}$-consistent estimator $\tilde{\theta}_n$ of θ, then the estimator $\hat{\theta}_n$ given above is a uniformly efficient estimator of θ.*

note. It is important for the result above that the sample space is Euclidean.
note2. The result is also important since even if the maximum likelihood estimate $\hat{\theta}_n$ does not exist, we can define a one-step Newton-Raphson approximate 'solution' by

$$\hat{\theta}_n^{\text{approx}} = \tilde{\theta}_n + \left[-\frac{1}{n} \sum_{i=1}^{n} \ddot{\boldsymbol{l}}(X_i, \tilde{\theta}_n) \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} \dot{\boldsymbol{l}}(X_i, \tilde{\theta}_n).$$

Part 1. Local asymptotic normality for ARMA process

3. Assumptions

3.1. **Assumption for stationary and invertibility and etc.**

(S1) The polynomials $A(z) = 1 + \sum_{i=1}^{p} -a_i z^i$ and $B(z) = 1 + \sum_{i=1}^{q} b_i z^q$ have no zeros with magnitude less or equal to one.
(S2) The two polynomials have no zeros in common and $a_p \neq 0$ or $b_q \neq 0$.

3.2. Assumptions for LAN.

(A1) $f(x) > 0$, $x \in \mathbb{R}$; $\int_{-\infty}^{\infty} x f(x)\,dx = 0$; $\int_{-\infty}^{\infty} x^2 f(x)\,dx = \sigma^2 < \infty$;

(A2) f is absolutely continuous on finite intervals, i.e., there exists $\dot{f}$ such that for all $-\infty < a < b < \infty$, $f(b) - f(a) = \int_a^b \dot{f}(x)\,dx$;

(A3) letting $\phi_f \equiv -\frac{\dot{f}}{f}$, the generalized Fisher information $\int_{-\infty}^{\infty} \phi_f^2(x) f(x)\,dx \equiv \mathcal{I}_f = \sigma^{-2}\mathcal{I}_{f_1}$ is finite.

(A4) the score function ϕ_f is piecewise Lipschitz, i.e., there exist a finite partition of $\mathbb{R}$ into nonoverlapping intervals $J_1, \dots, J_k$ and a constant A_f such that

$$|\phi_f(x) - \phi_f(y)| \le A_f |x - y| \qquad \forall x, y \in J_i, \forall i = 1, \dots, k.$$

(A5) f is strongly unimodal, i.e., ϕ_f is monotone increasing.

(A6) $g_0(\boldsymbol{e}_0, \boldsymbol{X}_0, \theta_n) \to g_0(\boldsymbol{e}_0, \boldsymbol{X}_0, \theta_0)$, in P_{θ_0}-probability if $\theta_n \to \theta$.

(A7) There exists a sequence $\{\bar{\theta}_n\}$ of estimators which satisfies

$$\sqrt{n}(\bar{\theta}_n - \theta_0) = O_{P_{\theta_0}}(1).$$

(A8) $\dot{\varphi}$ is assumed to satisfy

$$\lim_{h \to 0} \int \{\dot{\varphi}(x+h) - \dot{\varphi}(x)\}^2 f(x)\,dx = 0,$$

$$\lim_{h \to 0} \int \frac{\dot{\varphi}(x+h) - \dot{\varphi}(x)}{h} f(x)\,dx = -\frac{1}{2} I(f).$$

(A9) In order to construct the adaptive estimator, the following conditions on the densities f are required:

$$\int x^4 f(x)\,dx < \infty,$$

and f is symmetric about the origin.

note. All stable densities are unimodal. *note2.* $(A5) \Rightarrow (A2)$.

3.3. Interpretation of the assumptions.

(A1)—(A3) The LAN result holds under these assumptions;

(A4) This assumption induces that the influence of starting values on residual autocorrelations is to be asymptotically negligible. In other words, it can be shown that

$$E_\theta |\, \Delta_n(\theta) - \hat{\Delta}_n(\theta) \,| = o(1)$$

holds true, where

$$\hat{\Delta}_n(\theta) := \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(\hat{e}_j(\theta)) \sum_{k=1}^{j} \beta_{k-1} \begin{pmatrix} Y(j-k) \\ \hat{E}(j-k;\theta) \end{pmatrix},$$

and

$$\hat{e}_t := \sum_{k=1}^{t} \beta_{k-1}(X_{t+1-k} - a_1 X_{t-k} - \cdots - a_p X_{t+1-k-p}).$$

(A5) This assumption is given meaning to by the Proposition below(in Hallin&Werker part).
(A6) It is necessary to assume the convergency of the initial observation.
(A7) The existence of $\sqrt{n}$-consistent initial estimators $\{\bar{\theta}_n\}$ is assumed to construct regular estimates. In fact, (A7) holds for estimators for which the usual CLT is valid, i.e., for all the standard estimators.

As an example in Anderson(1971), consider AR model as follows:

$$\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{y}_{t-1} = \boldsymbol{u}_t. \tag{3.1}$$

or

$$y_t + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} = u_t. \tag{3.2}$$

Under the assumption (A.1) below, we can write the process as

$$y_t = \sum_{s=0}^{\infty} (-\boldsymbol{B})^s \boldsymbol{u}_{t-s}, \quad t = \ldots, -1, 0, 1, \ldots.$$

Let $\boldsymbol{F}$ be the sample variance, then it can be written in the form like

$$\boldsymbol{F} = E\boldsymbol{y}_t\boldsymbol{y}_t' = \sum_{s=0}^{\infty} (-B)^s \Sigma (-B')^s.$$

Also define $\tilde{\boldsymbol{F}}$ as

$$\tilde{\boldsymbol{F}} = \sum_{s=0}^{\infty} \tilde{\boldsymbol{B}}^s \tilde{\Sigma} \tilde{\boldsymbol{B}}'^s,$$

where

$$\tilde{\Sigma} = \begin{pmatrix} \sigma^2 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{O} \end{pmatrix}, \quad \tilde{\boldsymbol{B}} = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_p \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Assuming conditions in the set 1 and set 2 respectively below,

(A.1) $\{\boldsymbol{y}_t\}$ is a sequence of random vectors satisfying (3.1) with $\{\boldsymbol{u}_t\}$ independently and identically distributed with $E\boldsymbol{u}_t = \boldsymbol{0}$ and $E\boldsymbol{u}_t\boldsymbol{u}_t' = \boldsymbol{\Sigma}$;
(A.2) $-\boldsymbol{B}$ has all characteristic roots less than 1 in absolute value;
(A.3) $\boldsymbol{F}$ is positive definite;
(A.1)' $\{y_t\}$ is a sequence of random vectors satisfying (3.2) with $\{u_t\}$ independently and identically distributed with $Eu_t = 0$ and $Eu_t^2 = \sigma^2$;
(A.2)' The roots of the associated polynomial equation are less than 1 in absolutely value,

then we have theorems which give the estimators satisfing the assumption(A7).

Theorem 3.1. *Under the set 1 of assumptions, $\sqrt{T}(\hat{\boldsymbol{B}}' - \boldsymbol{B}')$ has a limiting normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{F}^{-1} \otimes \boldsymbol{\Sigma}$.*

Theorem 3.2. *Under the set 2 of assumptions,* $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ *has a limiting normal distribution with mean* $\mathbf{0}$ *and covariance matrix* $\sigma^2 \tilde{\boldsymbol{F}}^{-1}$*.*

Instead of (A.1) or (A.1)', the result holds true even if we assume
(A.1)" $E|\, u_{it}\, |^{2+\epsilon} < m, \quad i = 1, 2, \ldots, p, t = 1, 2, \ldots,$ for some $\epsilon > 0$ and some m.
In conclusion, such estimators satisfying assumption (A7) exist under moment conditions.

(A8) This is the definition of regularity on the model $\boldsymbol{P}$, which guarantees the L^2-continuity of $\dot{\varphi}$ and the existence of it.

(A9) This is a condition for Theorem 5.7 to be true.

4. Theorems

The LAN property is established for ARMA model by using the assumptions of Roussas(1979). Similar conditions sufficient for the LAN property are given in Swensen(1985).

Theorem 4.1 ((K-Theorem 3.1)LAN property for ARMA models)**.** *Let* $\{h_n\} \subset \mathbb{R}^{p+q}$ *be a bounded sequence and* $\theta_n = \theta_0 + n^{-1/2} h_n$*. Under our assumptions (A1)–(A4) and (A6), we have for*

$$\Delta_n(\theta) = \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(e_j(\theta)) Z(j-1; \theta, \theta), \quad \dot{\varphi} = -f'/2f,$$

the following two results:

$$\log[\, dP_{n,\theta_n}/dP_{n,\theta_0}\,] - h_n^T \Delta_n(\theta_0) + \frac{1}{2} h_n^T I(f) \Gamma(\theta_0) h_n \to 0,$$

in P_{n,θ_0}*-probability, where* $\Gamma(\theta_0)$ *is defined in Theorem 3.5 below (approximation of the log-likelihood ratio).*

$$\mathcal{L}(\Delta_n(\theta_0) | P_{n,\theta_0}) \Rightarrow \mathcal{N}(0, I(f)\Gamma(\theta_0)),$$

where "$\Rightarrow$*" denotes weak convergence (asymptotic normality of the approximating statistic).*

Corollary 4.2. *Under the same assumption as above* $\{P_{n,\theta_0}\}$ *and* $\{P_{n,\theta_n}\}$ *are contiguous in the sense of Definition 2.1, Roussas (1972), page 7, and*

$$\mathcal{L}(\Delta_n(\theta_0) - I(f)\Gamma(\theta_0) h_n | P_{n,\theta_n}) \Rightarrow \mathcal{N}(0, I(f)\Gamma(\theta_0)).$$

5. The sufficient conditions for local asymptotic normality

The 4 theorems below guarantee that the sufficient conditions for the LAN in Roussas (1979) are fulfilled.

Theorem 5.1. *For each* $\theta_0 \in \Theta$*, the random functions* $\phi_j(\theta_0, \cdot)$ *are differentiable in q.m.* $[\, P_{\theta_0}\,]$ *uniformly in* $j \geq 1$*. That is, there are* $(p+q)$*-dimensional r.v.'s* $\dot{\phi}_j(\theta_0) = \dot{\varphi}(e_j^0) Z(j - 1; \theta_0, \theta_0) = \dot{\varphi}(e_j^0) Z^0(j-1)$ *[the q.m. derivative of* $\phi_j(\theta_0, \theta)$ *with respect to* θ *at* θ_0*] such that*

$$\frac{\phi_j(\theta_0, \theta_0 + \lambda h) - 1}{\lambda} - h^T \dot{\phi}_j(\theta_0) \to 0, \quad \text{in q.m. } [P_{\theta_0}] \text{ as } \lambda \to 0$$

uniformly on bounded sets of $h \in \mathbb{R}^{p+q}$ *and uniformly in* $j \in \mathbb{N}$*. Finally,* $\dot{\phi}_j(\theta_0)$ *is measurable with respect to* $\mathcal{A}_j$*.*

Theorem 5.2. *For each* $\theta_0 \in \Theta$ *and each* $h \in \mathbb{R}^{p+q}$*, the sequence* $\{(h^T \dot{\phi}_j(\theta_0))^2\}, j \in \mathbb{N}$*, is uniformly integrable with respect to* P_{θ_0}*.*

Theorem 5.3. *For each* $\theta_0 \in \Theta$ *and* $j \geq 1$ *let the* $(p+q) \times (p+q)$*-dimensional covariance matrix* $\Gamma_j(\theta_0)$ *be defined by*

$$\Gamma_j(\theta_0) = 4E_{\theta_0}[\,\dot{\phi}_j(\theta_0)\dot{\phi}_j^T(\theta_0)\,] = I(f)E_{\theta_0}[\,Z(j-1;\theta_0,\theta_0)Z^T(j-1;\theta_0,\theta_0)\,].$$

Then $\Gamma_j(\theta_0) \to \Gamma(\theta_0)I(f)$*, as* $j \to \infty$*, in any one of the standard norms in* $\mathbb{R}^{p+q}$*, and* $\Gamma(\theta_0)$ *is positive definite.*

Theorem 5.4. *(i) For each* $\theta_0 \in \Theta$*, each* $h \in \mathbb{R}^{p+q}$ *and for the probability measure* P_{θ_0}*, the WLLN holds for the sequence* $\{[\,h^T \dot{\varphi}_j(\theta_0)\,]^2, j \in \mathbb{N}\}$*. Also*
(ii)

$$\frac{1}{n}\sum_{j=1}^{n}\{E_{\theta_0}[\,(h^T\dot{\phi}_j(\theta_0))^2|\mathcal{A}_{j-1}\,] - [\,h^T\dot{\phi}_j(\theta_0)\,]^2\} \to 0, \quad \text{as } n \to \infty,$$

in P_{θ_0}*-probability.*

Part 2. Existence and construction of LAM estimates

Lemma 5.5 (K-Lemma 4.1)**.** *Under assumptions (A1)–(A3) and (A6), we have for any sequence* $\{Z_n\}$ *of estimates the following implication:*

$$\sqrt{n}(Z_n - \theta_0) - \frac{\Gamma(\theta_0)^{-1}}{I(f)}\Delta_n(\theta_0) = o_{P_{\theta_0}}(1) \quad (\{Z_n\} \text{ is called } \theta_0\text{-regular})$$

implies that $\{Z_n\}$ *is LAM.*

Theorem 5.6 (Existence of LAM estimators)**.** *Assume* $\{\bar{\theta}_n\} \subset \Theta$ *is discrete and* $\sqrt{n}$*-consistent for* $\theta_0 \in \Theta$*. Then* $\hat{\theta}_n$ *defined below is regular:*

$$\hat{\theta}_n = \bar{\theta}_n + \frac{1}{\sqrt{n}}\frac{\hat{\Gamma}_n(\bar{\theta}_n)^{-1}}{I(f)}\Delta_n(\bar{\theta}_n),$$

$$\hat{\Gamma}_n(\theta) = \frac{1}{n}\sum_{j=1}^{n} Z(j-1;\theta,\theta)Z^T(j-1;\theta,\theta).$$

Part 3. Construction of adaptive estimates

Theorem 5.7. *Let* $\{\bar{\theta}_n\} \subset \Theta$ *be a discrete and* $\sqrt{n}$*-consistent sequence of estimators of* θ_0*. Under our assumptions (A1)–(A3),(A6)–(A9) and*

$$\tilde{\Delta}_n(\bar{\theta}_n) - \Delta_n(\bar{\theta}_n) = o_{P_{\theta_0}}(1)$$

holds, if $c_n \to \infty$*,* $g_n \to \infty$*,* $\sigma(n) \to 0$*,* $d_n \to 0$*,* $\sigma(n)c_n \to 0$*,* $g_n\sigma(n)^{-4}/n \to 0$ *and* $n\sigma(n)$ *stays bounded.*

6. Summary

The parameter θ_0(the coefficients in ARMA model) are considered in ARMA model with independent and identically, but not necessary normally distributed errors. LAN is proven in this general model where the sample is not i.i.d. Under the adequate conditions, the LAM(Locally asymptotically minimax) estimators are proposed, and strongly adaptive estimators are obtained. The approach is exactly the starting point of semi-parametric method in ARMA models, which Hallin&Werker(2003) follows later on.

7. Appendix

7.1. Formulas.

$$\beta_s + b_1\beta_{s-1} + \cdots + b_q\beta_{s-q} = 0 \quad \text{for } \forall s \geq 1.$$

$$(1 + b_1 L + \cdots + b_q L^q)^{-1} = \sum_{k=0}^{\infty} \beta_k L^k.$$

$$e_j = \sum_{k=1}^{j} \beta_{k-1} \left(-\sum_{i=0}^{p} a_i X_{j+1-k-i} \right) + \sum_{s=0}^{q-1} e_{-s} \left(\sum_{k=0}^{s} \beta_{j+s-k} b_k \right).$$

$$\sum_{k=j+1}^{\infty} \beta_{k-1} \left(\sum_{i=0}^{q} b_i e_{j+1-k-i} \right) = \sum_{s=0}^{q-1} e_{-s} \left(\sum_{k=0}^{s} \beta_{j+s-k} b_k \right).$$

$$\sum_{k=1}^{j} \beta_{k-1} \left(\sum_{i=0}^{q} b_i e_{j+1-k-i} \right) = e_{j+1-k} - \sum_{s=0}^{q-1} e_{-s} \left(\sum_{k=0}^{s} \beta_{j+s-k} b_k \right).$$

$$\sum_{i=0}^{p} a_i^0 X_{t-i} = \sum_{i=0}^{q} b_i^0 e_{t-i}(\theta_0).$$

7.2. Lemmas.

Lemma 7.1 (K-Lemma 2.2)**.** *With*

$$\begin{aligned} Z(j-1;\theta,\theta_0) &= \sum_{k=1}^{j} \beta_{k-1}(\theta)(X_{j-k},\ldots,X_{j-k+1-p}; e_{j-k}^0,\ldots,e_{j-k+1-q}^0)' \\ &= \sum_{k=1}^{j} \beta_{k-1}(\theta)(Y'(j-k); E(j-k;\theta_0)')', \end{aligned}$$

$$e_j(\theta_0) - e_j(\theta) = (\theta - \theta_0)' Z(j-1;\theta,\theta_0)$$

holds true.

SUMMARY OF HOSOYA-TANIGUCHI(1982)

GEN RYU

1. MODELS AND NOTATIONS

1.1. Scalar linear processes.

$$x(n) = \sum_{j=0}^{\infty} a_j(\theta) e(n-j), \quad n \in \mathbb{Z} \tag{1.1}$$

We define $I_X(\omega)$ and $f(\omega)$ for (1.1).

$$I_X(\omega) = (2\pi n)^{-1} \left| \sum_{t=1}^{n} x(t) e^{it\omega} \right|^2 \tag{1.2}$$

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma(h) e^{ih\omega} \tag{1.3}$$

1.2. Vector linear processes.

$$z(n) = \sum_{j=0}^{\infty} G_j(\theta) e(n-j), \quad n \in \mathbb{Z} \tag{1.4}$$

$$(s * 1) = (s * p) * (p * 1) \tag{1.5}$$

We define $I_X(\omega)$ and $f(\omega)$ similarly for (1.2).

2. HISTORY

Discoverer and Year	Model	$\hat{\theta}$
Whittle(1962)	scalar;$e(n) \sim i.i.d(0, \sigma^2)$	$\int_{-\pi}^{\pi} \frac{I_X(\omega)}{f_\theta(\omega)} \, d\omega$
Walker(1964) Hannan(1973)	σ^2 depends on θ	
Hosoya(1974)	same to the above	$\int_{-\pi}^{\pi} \log f_\theta(\omega) + \frac{I_X(\omega)}{f_\theta(\omega)} \, d\omega$
Hannan(1976) Dunsmuir(1976,1979)	vector;$e(n) \sim i.i.d(0, K(\theta))$	$2\pi \log \det K(\theta) + \int_{-\pi}^{\pi} \mathrm{tr}\{f_\theta(\omega)^{-1} I_X(\omega)\} d\omega$

TABLE 1. Transition of the estimators

Date: September 2, 2011.

In Hosoya-Taniguchi(1982), a criterion is proposed as follows.

$$D(f_t, f) = \int_{-\pi}^{\pi} \log\det f_t(\omega) + \mathrm{tr}\{f_t(\omega)^{-1} f(\omega)\} d\omega \tag{2.1}$$

3. THEOREMS

Here, the model is vector-valued and is represented by

$$\left\{ \begin{array}{rr} z(n) = & \sum_{j=0}^{\infty} G(j) e(n-j), \quad n \in \mathbb{Z} \\ (s*1) & (s*p) \quad (p*1) \\ & \\ E\{e(n)\} = & 0 \\ E\{e(m)e(n)'\} = & \delta(m,n)K \end{array} \right. \tag{3.1}$$

In the paper, an assumption is assumed throughout.

Assumption 3.1.

$$\sum_{j=0}^{\infty} \mathrm{tr} G(j) K G(j)' < \infty. \tag{3.2}$$

Under this assumption, the process $\{z(n)\}$ is a second-order stationary process. The spectral density matrix of the process is shown as

$$f(\omega) = \frac{1}{2\pi} k(\omega) K k(\omega)^*, \quad -\pi \le \omega \le \pi. \tag{3.3}$$

Theorem 3.2. $\{x(t)\}$*: zero-mean second-order stationary process.* $\mathcal{F}_t \equiv \mathcal{F}_t^x$.
Assumptions:

(1) $\forall \epsilon > 0, \mathrm{Var}\{E(x_\alpha(t+\tau)|\mathcal{F}_t)\} = O(\tau^{-2-\epsilon})$ *uniformly in t, for* $\alpha = 1, \dots, p$.
(2) $\forall l, m > t$, $\forall \eta > 0$,
$E|E\{x_\alpha(l)x_\beta(m)|\mathcal{F}_t)\} - E\{x_\alpha(l)x_\beta(m)\}| = O[\{\min(|l-t|, |m-t|)\}^{-1-\eta}]$ *uniformly in t, for* $\alpha = 1, \dots, p$.
(3) *Any element of* $f(\omega) = \{f_{\alpha\beta}(\omega); \alpha, \beta = 1, \dots, p\}$ *is continuous at the origin; f(0) is non-degenerate.*

Result: $\xi_N = N^{-\frac{1}{2}} \sum_{n=1}^{N} x(n) \to N(0, 2\pi f(0))$.

Theorem 3.3. $\mathcal{B} \equiv \mathcal{B}_e(t)$.
Assumptions:

(1) $\forall \beta_1, \beta_2, m, \forall \epsilon > 0$,
$\mathrm{Var}[E\{e_{\beta_1}(n) e_{\beta_2}(n+m)|\mathcal{B}(n-\tau)\} - \delta(m,0)K_{\beta_1\beta_2}] = O(\tau^{-2-\epsilon})$ *uniformly in n.*
(2) $\forall \eta > 0$,
$E|E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)|\mathcal{B}(n_1-\tau)\} - E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)\}| = O(\tau^{-1-\eta})$*, uniformly in* n_1*, where* $n_1 \le n_2 \le n_3 \le n_4$.

(3) $f_{\beta\beta}$ *are square-integrable.*
(4) $\sum_{j_1,j_2,j_3=infty}^{\infty} |Q^e_{\mathcal{B}_1\ldots\mathcal{B}_4}(j_1,j_2,j_3)| < \infty$.

Results:

(1) $\sqrt{N}\{C^Z_{\alpha_1\alpha_2}(m) - \gamma^Z_{\alpha_1\alpha_2}(m)\} \to N(0,...)$

(2)

$$\mathrm{Cov}(\sqrt{N}\{C^Z_{\alpha_1\alpha_2}(m_1) - \gamma^Z_{\alpha_1\alpha_2}(m_1)\}, \sqrt{N}\{C^Z_{\alpha_3\alpha_4}(m_2) - \gamma^Z_{\alpha_3\alpha_4}(m_2)\})$$
$$\to 2\pi \int_{-\pi}^{\pi} [f_{\alpha_1\alpha_3}(\omega)\bar{f}_{\alpha_2\alpha_4}(\omega)\exp\{-i(m_2-m_1)\omega\} + f_{\alpha_1\alpha_4}(\omega)\bar{f}_{\alpha_2\alpha_3}(\omega)\exp\{i(m_1+m_2)\omega\}]\,d\omega$$
$$+ 2\pi \sum_{\beta_1,\ldots,\beta_4=1}^{p} \int\int_{-\pi}^{\pi} \exp\{im_1\omega_1 + im_2\omega_2)\}k_{\alpha_1\beta_1}(\omega_1)$$
$$k_{\alpha_2\beta_2}(-\omega_1)k_{\alpha_3\beta_3}(\omega_2)k_{\alpha_4\beta_4}(-\omega_2)\tilde{Q}^e_{\beta_1\ldots\beta_4}(\omega_1,-\omega_2,\omega_2)\,d\omega_1 d\omega_2$$

Lemma 3.4. $D(f_{T(f)}, f) = \min_{t\in\Theta} D(f_t, f)$.

Assumptions:

(1) $\Theta : (\subset \mathbb{R}^q)$ *compact.*
(2) $\theta_1 \neq \theta_2 \Rightarrow f_{\theta_1} \neq f_{theta_2}$.
(3) $f_\theta(\omega)$*: positive definite.*
(4) $f_\theta(\omega)$ *is continuous w.r.t* θ, ω.

Results:

(1) $\forall f \in \mathcal{P}, \exists T(f) \in \Theta$ *s.t.* $D(f_{T(f)}, f) = \min_{t\in\Theta} D(f_t, f)$.
(2) *T(f): unique,* $T(f_N) \to_\omega f \Longrightarrow$ *as* $N \to \infty, T(f_N) \to_\omega f$.
(3) $\forall\theta \in \Theta, T(f_\theta) = \theta$.

Theorem 3.5. $\exists 1\, T(f) \in \Theta^\circ$ *;*

$$M_f = \int_{-\pi}^{\pi} \left[\frac{\partial^2}{\partial\theta\partial\theta'}\mathrm{tr}\{f_t(\omega)^{-1}f(\omega)\} + \frac{\partial^2}{\partial\theta\partial\theta'}\log\det f_\theta(\omega)\right]_{\theta=T(f)} d\omega,$$

where M_f *is nonsingular matrix.*

Assumptions:

(1) $\forall\beta_1, \beta_2, m, \forall\epsilon > 0$,
$\mathrm{Var}[E\{e_{\beta_1}(n)e_{\beta_2}(n+m)|\mathcal{B}(n-\tau)\} - \delta(m,0)K_{\beta_1\beta_2}] = O(\tau^{-2-\epsilon})$ *uniformly in n.*
(2) $\forall\eta > 0$,
$E|E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)|\mathcal{B}(n_1-\tau)\} - E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)\}| = O(\tau^{-1-\eta})$, *uniformly in* n_1, *where* $n_1 \leq n_2 \leq n_3 \leq n_4$.
(3) $f_{\beta\beta}$ *are square-integrable.*
(4) $\sum_{j_1,j_2,j_3=infty}^{\infty} |Q^e_{\mathcal{B}_1\ldots\mathcal{B}_4}(j_1,j_2,j_3)| < \infty$.
(5) $f(\omega) \in \mathrm{Lip}(\alpha)$ *where* $\alpha > \frac{1}{2}$.

Results:

(1) *p-lim*$_{N\to\infty} T(I_z) = T(f)$.

(2) *as* $N \to \infty$, $\sqrt{N}\{T(I_z) - T(f)\} \to N(0, M_f^{-1}\tilde{V}M_f^{-1})$. *where*

$$\tilde{V}_{jl} = 4\pi \int_{-\pi}^{\pi} \mathrm{tr}\left[f(\omega)\frac{\partial}{\partial\theta_j}\{f_t(\omega)\}^{-1} f(\omega)\frac{\partial}{\partial\theta_l}\{f_t(\omega)\}^{-1}\right]_{\theta=T(f)} d\omega$$
$$+ 2\pi \sum_{r,t,u,v=1}^{s} \int\int_{-\pi}^{\pi} \left\{\frac{\partial}{\partial\theta_j} f_\theta^{(r,t)}(\omega_1)\frac{\partial}{\partial\theta_l} f_\theta^{(u,v)}(\omega_2)\right\}_{\theta=T(f)} \tilde{Q}^z_{rtuv}(-\omega_1, \omega_2, -\omega_2)\, d\omega_1 d\omega_2,$$

where $f_\theta^{(r,t)}(\omega)$ *is the (r,t) element of* $\{f_\theta(\omega)\}^{-1}$.

Corollary 3.6.

$$\tilde{V}_{jl} = 4\pi \int_{-\pi}^{\pi} \mathrm{tr}\left[f(\omega)\frac{\partial}{\partial\theta_j}\{f_t(\omega)\}^{-1} f(\omega)\frac{\partial}{\partial\theta_l}\{f_t(\omega)\}^{-1}\right]_{\theta=T(f)} d\omega$$
$$+ 2\pi \sum_{a,b,c,d=1}^{p} \sum_{r,t,u,v=1}^{s} \int\int_{-\pi}^{\pi} \left\{\frac{\partial}{\partial\theta_j} f_\theta^{(r,t)}(\omega_1)\frac{\partial}{\partial\theta_l} f_\theta^{(u,v)}(\omega_2)\right\}_{\theta=T(f)}$$
$$k_{ra}(-\omega_1)k_{tb}(\omega_1)k_{uc}(-\omega_2)k_{vd}(\omega_2)\tilde{Q}^e_{abcd}(-\omega_1, \omega_2, -\omega_2)\, d\omega_1 d\omega_2$$

Proposition 3.7. *Assumption:*

$$\mathrm{cum}\{e_a(n_1), e_b(n_2), e_c(n_3), e_d(n_4)\} = \begin{cases} \kappa_{abcd} & \textit{if } n_1 = n_2 = n_3 = n_4 \\ 0 & \textit{otherwise.} \end{cases} \tag{3.4}$$

Result:

$$\tilde{V}_{jl} = 4\pi \int_{-\pi}^{\pi} \mathrm{tr}\left[f(\omega)\frac{\partial}{\partial\theta_j}\{f_t(\omega)\}^{-1} f(\omega)\frac{\partial}{\partial\theta_l}\{f_t(\omega)\}^{-1}\right]_{\theta=T(f)} d\omega$$
$$+ \sum_{a,b,c,d=1}^{s} \kappa_{abcd}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} k^*(\omega)\frac{\partial}{\partial\theta_j}\{f_\theta(\omega)\}^{-1}k(\omega)d\omega\right]_{ab}$$
$$\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} k^*(\omega)\frac{\partial}{\partial\theta_l}\{f_\theta(\omega)\}^{-1}k(\omega)d\omega\right]_{cd} \Big|_{\theta=T(f)}.$$

In the case where $f(\omega) = f_\theta(\omega)$ and θ is the innovation-free parameter, the second term in the right-hand side will be 0. On the other hand, in the case of $f(\omega) \neq f_\theta(\omega)$, even if (3.4) is satisfied, the quasi-Gaussian maximum likelihood estimates for the innovation-free parameters are generally not robust against the fourth cumulant. In the case $s = 1$, that is in the scalar case, the quasi-Gaussian maximum likelihood estimates for the innovation-free parameters are robust against fourth cumulant even if $f(\omega) \neq f_\theta(\omega)$.

4. APPLICATION

4.1. **An autoregressive signal with white noise.**

$$\begin{cases} \sum_{j=1}^{q} \theta_j s(t-j) = \eta(t), \quad n \in \mathbb{Z} \\ E\{\eta(t)\} = 0 \\ E\{\eta(t)\eta(s)\} = \theta_{q+1}\delta(t,s), \end{cases} \tag{4.1}$$

where all zeros of $\sum \theta_j z^j$ are assumed to be outside the unit circle.

$$\begin{cases} X(t) = s(t) + e(t) \\ E\{e(t)\} = 0 \\ E\{e(t)e(s)\} = \theta_{q+2}\delta(t,s) \\ E\{e(t)\eta(s)\} = 0 \qquad \textit{for all } t \textit{ and } s. \end{cases} \tag{4.2}$$

Proposition 4.1. *For the model above, we give the assumptions:*

(1) $\{e(t)\}$ *and* $\{\eta(t)\}$ *are fourth-order stationary processes.*
(2) *the vector-valued process* $\{e(t), \eta(t)\}$ *satisyies*
   (a) $\forall \beta_1, \beta_2, m, \forall \epsilon > 0$,
   $\mathrm{Var}[E\{e_{\beta_1}(n)e_{\beta_2}(n+m)|\mathcal{B}(n-\tau)\} - \delta(m,0)K_{\beta_1\beta_2}] = O(\tau^{-2-\epsilon})$ *uniformly in* n.
   (b) $\forall \eta > 0$,
   $E|E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)|\mathcal{B}(n_1-\tau)\} - E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)\}| = O(\tau^{-1-\eta})$, *uniformly in* n_1, *where* $n_1 \le n_2 \le n_3 \le n_4$.
   (c) $f_{\beta\beta}$ *are square-integrable.*
   (d) $\sum_{j_1,j_2,j_3=infty}^{\infty} |Q^e_{\mathcal{B}_1\ldots\mathcal{B}_4}(j_1,j_2,j_3)| < \infty$.
(3) θ^0 *(true value of* θ*)* $\in B \times K_1 \times K_2$, *where* B, K_1, K_2 *is compact subset respectively.*

Results:

$$\sqrt{N}\{T(I_X) - \theta^0\} \to N(0, M_f^{-1} V M_f^{-1}) \tag{4.3}$$

where

$$V_{jl} = 4\pi \int_{-\pi}^{\pi} \{f_{\theta^0}(\omega)\}^2 \frac{\partial}{\partial\theta_j}\{f_{\theta^0}(\omega)\}^{-1} \frac{\partial}{\partial\theta_l}\{f_{\theta^0}(\omega)\}^{-1}\, d\omega$$
$$+ 2\pi \sum_{\alpha_1,\alpha_2,\alpha_3,\alpha_4=1}^{2} \int\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_j}\{f_{\theta^0}(\omega_1)\}^{-1} \frac{\partial}{\partial\theta_l}\{f_{\theta^0}(\omega_2)\}^{-1}$$
$$k_{\alpha_1}(-\omega_1)k_{\alpha_2}(\omega_1)k_{\alpha_3}(-\omega_2)k_{\alpha_4}(\omega_2)\tilde{Q}^{e,\eta}_{abcd}(-\omega_1,\omega_2,-\omega_2)\, d\omega_1 d\omega_2.$$

ISOMETRIC APPROXIMATION

YAN LIU

1. Essential points

Definition 1.1 (ϵ-nearisometry and $(1+\epsilon)$-bilipschitz)**.** f is an ϵ-nearisometry if

$$|x-y|-\epsilon \leq |fx-fy| \leq |x-y|+\epsilon$$

for all $x, y \in A$, and that f is $(1+\epsilon)$-bilipschitz if

$$|x-y|/(1+\epsilon) \leq |fx-fy| \leq (1+\epsilon)|x-y|$$

for all $x, y \in A$.

Lemma 1.2. *Suppose that $A \subset \mathbb{R}^n$ is a bounded set and that $f : A \to l_2$ is a map satisfying the nearisometry condition. Then there is an isometry $S : A \to l_2$ such that*

$$|Sx - fx| \leq c_n\sqrt{\epsilon}$$

for all $x \in A$. Further, if A satisfies a thickness condition, then

$$|Sx - fx| \leq c_n\epsilon/q,$$

where q is a thickness parameter.

(The thickness condition is not so clear in the paper.)

Date: August 3, 2014.

SUMMARY–ASYMPTOTICS OF TESTS FOR A UNIT ROOT IN AUTOREGRESSION

GEN RYU

Suppose that $\{Y_t : t = 1, \ldots, n\}$ is generated by the first-order autoregressive process

$$Y_t = \theta Y_{t-1} + e_t, \quad Y_0 = 0,\ t = 1, \ldots, \tag{0.1}$$

where e_t's are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables and

$$\theta = \exp(\frac{c}{n}).$$

As a generation of the LSE $\hat{\theta}$ and $\hat{\hat{\theta}}_{c_1,c_2}$,

$$\hat{\theta}_{c_1,c_2} = \frac{\sum_{t=2}^{n} Y_t Y_{t-1}}{\sum_{t=2}^{n-1} Y_t^2 + c_1 Y_1^2 + c_2 Y_n^2}, \quad c_1, c_2 \geq 0.$$

$$\hat{\hat{\theta}}_{c_1,c_2} = \frac{\sum_{t=2}^{n} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=2}^{n-1} (Y_t - \bar{Y})^2 + c_1 (Y_1 - \bar{Y})^2 + c_2 (Y_n - \bar{Y})^2}, \quad c_1, c_2 \geq 0,\ \bar{Y} = \sum_{t=1}^{n} Y_t / n,$$

are supposed.

The hypothesis is supposed as

$$H : \theta = 1 \quad vs \quad A : \theta \in (0, 1).$$

For the testing problem, the following tests are introduced:

$$K_{1n} = \frac{\sqrt{2}}{n\hat{\sigma}^2} \sum_{t=2}^{n} (\hat{\theta}_{c_1,c_2} - 1); \tag{0.2}$$

$$K_{2n} = \frac{n}{\sqrt{2}} (\hat{\theta}_{c_1,c_2} - 1); \tag{0.3}$$

$$K_{3n} = \left(\sum_{t=2}^{n} \frac{Y_{t-1}^2}{\hat{\sigma}^2} \right)^{1/2} (\hat{\theta}_{c_1,c_2} - 1), \tag{0.4}$$

where $\hat{\sigma}^2 = n^{-1} \sum_{t=2}^{n} (Y_t - \hat{\theta}_{c_1,c_2} Y_{t-1})^2$.

Date: April 10, 2012.

1. DEFINITIONS

1.1. **Ornstein-Uhlenbeck process.** Let $J_c(t)$ be an Ornstein-Uhlenbeck process

$$J_c(t) = \int_0^t \exp\{(t-s)c\}\, dW(s),$$

which is generated by

$$dJ_c(t) = cJ_c(t)dt + dW(t),$$

with initial condition $J_c(0) = 0$.

1.2. **integrated process.**
For the process above,
when $c \neq 0$, it is called a near-integrated process;
when $c = 0$, it is called an integrated process.

2. ASSUMPTIONS

(1) $E(e_t) = 0$ for all t,
(2) $\sup_t E|e_t|^{\beta+\epsilon} < \infty$ for some $\beta > 2$ and $\epsilon > 0$,
(3) $\sigma^2 = \lim_{n\to\infty} E(n^{-1}S_n^2)$ exists and $\sigma^2 > 0$ where $S_t = \sum_{s=1}^t e_s$,
(4) $\{e_t\}$ is strong mixing with mixing coefficients α_m that satisfy $\sum_{m=1}^\infty \alpha_m^{1-2/\beta} < \infty$.

3. THEOREMS AND LEMMAS

Lemma 3.1 (Phillips(1987b))**.** *If $\{Y_t\}$ is a near-integrated time series generated by* (0.1), *then, as $n \to \infty$,*

(1) $n^{1/2}Y_{[nt]} \xrightarrow{d} \sigma J_c(t)$;
(2) $n^{3/2}\sum_{t=1}^n Y_t \xrightarrow{d} \sigma \int_0^1 J_c(t)\, dt$;
(3) $n^{-2}\sum_{t=1}^n Y_t^2 \xrightarrow{d} \sigma^2 \int_0^1 J_c(t)^2\, dt$;
(4) $n^{-1}\sum_{t=1}^n Y_{t-1}e_t \xrightarrow{d} \sigma^2 \int_0^1 J_c(t)\, dW(t) + \frac{1}{2}(\sigma^2 - \sigma_e^2)$ *with* $\sigma_e^2 = \lim_{n\to\infty} n^{-1}\sum_{t=1}^n E(e_t^2)$.

Theorem 3.2. *If $\{Y_t\}$ is a near-integrated time series generated by the model above, then, as $n \to \infty$,*

$$n(\hat{\theta}_{c_1,c_2} - \theta) \xrightarrow{d} \frac{(1-2c_2)J_c(1)^2 - 2c\int_0^1 J_c(t)^2\, dt - \sigma_e^2/\sigma^2}{2\int_0^1 J_c(t)^2\, dt}.$$

Corollary 3.3. *If $\theta = 1$ (i.e., $c = 0$), then*

$$n(\hat{\theta}_{c_1,c_2} - 1) \xrightarrow{d} \frac{(1-2c_2)W(1)^2 - \sigma_e^2/\sigma^2}{2\int_0^1 W(t)^2\, dt}.$$

Theorem 3.4. *If $\{Y_t\}$ is a near-integrated time series generated by the model above, then, as $n \to \infty$,*

$$n(\hat{\hat{\theta}} - \theta) \xrightarrow{d} \frac{-2cG + (1-2c_2)T^2 + 4c_2TH - 2(c_1+c_2-1)H^2 - 2HW(1) - \sigma_e^2/\sigma^2}{2(G - H^2)},$$

where $G = \int_0^1 J_c(t)^2\,dt$, $T = J_c(1)$ *and* $H = \int_0^1 J_c(t)\,dt$.

Corollary 3.5. *If* $\theta = 1$, *then, as* $n \to \infty$,

$$n(\hat{\hat{\theta}}_{c_1,c_2} - 1) \xrightarrow{d} \frac{(1-2c_2)T_w^2 + 2(2c_2-1)T_w H_w - 2(c_1+c_2-1)H_w^2 - \sigma_e^2/\sigma^2}{2(G_w - H_w^2)},$$

where $G_w = \int_0^1 W(t)^2\,dt$, $T_w = W(1)$ *and* $H_w = \int_0^1 W(t)\,dt$.

Theorem 3.6. *Under* H, *as* $n \to \infty$, *we have*

$$K_{1n} \xrightarrow{d} \frac{(1-2c_2)W(1)^2 - 1}{\sqrt{2}}; \tag{3.1}$$

$$K_{2n} \xrightarrow{d} \frac{(1-2c_2)W(1)^2 - 1}{2\sqrt{2}\int_0^1 W(t)^2\,dt}; \tag{3.2}$$

$$K_{3n} \xrightarrow{d} \frac{(1-2c_2)W(1)^2 - 1}{2(\int_0^1 W(t)^2\,dt)^{1/2}}. \tag{3.3}$$

Theorem 3.7. *Under* A_n, *as* $n \to \infty$, *we have*

$$K_{1n} \xrightarrow{d} \frac{(1-2c_2)J_c(1)^2 - 2c\int_0^1 J_c(t)^2\,dt - 1}{\sqrt{2}}; \tag{3.4}$$

$$K_{2n} \xrightarrow{d} \frac{(1-2c_2)J_c(1)^2 - 2c\int_0^1 J_c(t)^2\,dt - 1}{2\sqrt{2}\int_0^1 J_c(t)^2\,dt}; \tag{3.5}$$

$$K_{3n} \xrightarrow{d} \frac{(1-2c_2)J_c(1)^2 - 2c\int_0^1 J_c(t)^2\,dt - 1}{2(\int_0^1 J_c(t)^2\,dt)^{1/2}}. \tag{3.6}$$

PAIRWISE SCORE FUNCTION

YAN LIU

Something has to be specified

Why do we use non-negative weights there in the bivariate marginal densities? In **?**, they mentioned that if the weights are all equal then they can be ignored: selection of unequal weights to improve efficiency is discussed in the context of particular application in their paper. A. Since there are many scores in the estimating function. $\omega_{i,j}$ are weights which can be used for example to reduce the number of pairs included in the estimation. The parameter estimates are obtained by maximizing expression.

Or

If the weights are all constant, then they can be ignored. Selection of unequal weights may improve efficiency, as explained in **?**.

Date: May 29, 2014.

1. Background

- Y q dimensional random vector;
- $\theta \in \Theta \subset \mathbb{R}^d$, $d \geq 1$;
- $(Y_1, \ldots, Y_n)$ are independent

Suppose that it is difficult to evaluate $f(y;\theta)$ and the corresponding likelihood $L(\theta)$. On the other hand, we can compute the likelihoods for pairs of observations (y_h^i, y_k^i) for $i = 1, \ldots, n$, and $h, k = 1, \ldots, q$. The pairwise likelihood defined from the bivariate marginal densities $f_{hk}(\cdot, \cdot; \theta)$ is given by

$$pL(\theta) = \prod_{i=1}^{n} \prod_{h=1}^{q-1} \prod_{k=h+1}^{q} f_{hk}(y_h^i, y_k^i; \theta)^{w_{hk}^i}, \tag{1.1}$$

where w_{hk}^i are non-negative weights which do not depend on the parameter θ or on y.

Then the maximum pairwise likelihood estimator $\hat{\theta}$ is the solution of

$$pU(\theta) = \frac{\partial}{\partial\theta} \log pL(\theta) = \sum_{i=1}^{n} \sum_{h=1}^{q-1} \sum_{k=h+1}^{q} w_{hk}^i \, pU_{hk}^i(\theta), \tag{1.2}$$

where

$$pU_{hk}^i(\theta) = \frac{\partial}{\partial\theta} \log f_{hk}(y_h^i, y_k^i; \theta). \tag{1.3}$$

Remark 1.1. *Here, the joint distribution of random vector is not assumed. The assumptions are correspond to the marginal distribution for bivariate part.*

Note that (**??**) is a likelihood and the empirical likelihood can be defined. For the profile empirical likelihood ratio function

$$R_E(\theta) = \sup\{\prod_{i=1}^{n} np_i | p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g(x_i, \theta) = 0\}, \tag{1.4}$$

we only have to assume

$$g(x_i, \theta) = \sum_{h=1}^{q-1} \sum_{k=h+1}^{q} w_{hk}^i \, pU_{hk}^i(\theta). \tag{1.5}$$

Or.... Change the order of the summation! There are two variations of the empirical likelihood.

(i) Usual one:

$$pw_n(\theta) = 2 \sum_{i=1}^{n} \log\{1 + \xi' g(x_i, \theta)\} \tag{1.6}$$

(ii) Or dimensional one:

$$pw_m(\theta) = 2 \sum_{r=1}^{m} \log\{1 + \xi' \, pU_r(\theta)\}. \tag{1.7}$$

Proposition 1.2. *Under H,*

$$pw_n(\theta) \xrightarrow{\mathcal{L}} \chi^2_d. \tag{1.8}$$

Proposition 1.3. *Under H,*

$$pw_m(\theta) \xrightarrow{\mathcal{L}} \sum_{j=1}^{d} \omega_j \chi^2_d, \tag{1.9}$$

where ω_j are the eigenvalues of the matrix $H(\theta)^{-1}J(\theta)$.

2. SEMIVARIOGRAM

If

$$\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) = \gamma(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|), \tag{2.1}$$

that is, the semivariogram depends only on the distance between the locations, then they are called isotropic semivariograms.

One of the commonly used isotropic semivariogram models is the exponential model

$$\gamma(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|; \phi) = c_0 + \sigma(1 - \rho^{\|\boldsymbol{s}_i - \boldsymbol{s}_j\|}). \tag{2.2}$$

3. WORDS

1.	isotropic	等長的な
2.	anisotropic	非等長的な
3.	nugget	金塊
4.	sill	土台

BRIEF ARTICLE

GEN RYU

1. Levy process

we denote its characteristic exponent by Ψ,

$$E(\exp\{i\langle\lambda, X_1\rangle\}) = \exp\{-\Psi(\lambda)\}.$$

Then the poisson process and Brownian motion can be shown as

$$\Psi(\lambda) = c(1 - e^{i\lambda}), \quad \Psi(\lambda) = \frac{1}{2}|\lambda|^2.$$

For every $\alpha \in (0, 2]$, a Lévy process with characteristic exponent Ψ is called a stable process with index α, if $\Psi(k\lambda) = k^\alpha\Psi(\lambda)$. for every $k > 0$ and $\lambda \in \mathbb{R}^d$. For $\alpha \neq 2$, the Lévy measure of a stable process of index α can be expresses in polar coordinates $(r, \theta) \in [0, \infty) \times S_{d-1}$ in the form

$$\Pi(dr, d\theta) = r^{-\alpha-d} dr v(d\theta),$$

where v is some finite measure on S_{d-1}.

Date: February 7, 2012.

SUMMARY OF LINEAR ALGEBRA AND THE WAY TO SPECTRAL MEASURE

GEN RYU

Referecen

(1) William Arveson [2002]
A short course on spectral theory
(2) Sam Raskin [2006]
Spectral Measures and the Spectral Theorem

Part 1. Two Ideas

The fundamental problem of linear algebra over the complex numbers is the solution of systems of linear equations. One can write the problem like:

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n = b_n \end{cases}$$

where a_{11}, $\cdots$, a_{nn}, b_1, b_n are given, and one attempts to solve for x_1, $\cdots$, x_n. It is also usual to write the system in the matrix way, that is

$$A\boldsymbol{x} = \boldsymbol{b}.$$

1. Linear Operator

The existence of solutions for any choice of $\boldsymbol{b}$ is equivalent to surjectivity of A; uniqueness of solutions is equivalent to injectivity of A. Thus the system of equations is uniquely solvable for all choices of $\boldsymbol{b}$ if and only if the linear operator A is invertible.

However, in infinite dimensions the difficulty lies deeper than the things above, because for most operators on an infinite-dimensional Banach space there is no meaningful concept of determinant. In other words, there is no numerical invariant for operators that determines invertibility in infinite dimensions as the determinant does in finite dimensions.

Date: June 9, 2012.

2. Eigenvalues

In finite dimensions, eigenvalues and eigenvectors for an operator A occur in pairs $(\lambda, \boldsymbol{x})$, where $Af = \lambda\boldsymbol{x}$, and $\boldsymbol{x}$ is nonzero vector in $\mathbb{C}^n$ and λ is a complex number.

Let think the case more general. $V_\lambda = \{\boldsymbol{x} \in \mathbb{C}^n; A\boldsymbol{x} = \lambda\boldsymbol{x}\} \subset \mathbb{C}^n$ is always a linear subspace of $\mathbb{C}^n$. However, V_λ is nontrivial if and only if the operator $A - \lambda I$ has nontrivial kernel: that is to say, if and only if the operator $A - \lambda I$ is not invertible.

Assuming that A is invertible, one can find $\boldsymbol{x}$ by the decomposition of $\boldsymbol{b}$. Let $\boldsymbol{b}$ be decomposed by

$$\boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{b}_2 + \cdots + \boldsymbol{b}_n,$$

where $\boldsymbol{b}_j$ is in V_{λ_j}, $\lambda_1, \ldots, \lambda_k$ being eigenvalues of A, then $\boldsymbol{x}$ is shown as

$$\boldsymbol{x} = \lambda_1^{-1}\boldsymbol{b}_1 + \lambda_2^{-1}\boldsymbol{b}_2 + \cdots + \lambda_k^{-1}\boldsymbol{b}_k.$$

Part 2. Spectral Measure

3. notations

We denote any Hilbert spaces by $\mathcal{H}$.

4. Riesz Lemma

Lemma 4.1. *If ψ is a linear functionsl on $\mathcal{H}$, then $\psi(v) = (v, w)$ for suitable choice of $w \in \mathcal{H}$.*

note. one can decompose $v = v_1 + v_2$, where $v_1 = \frac{\psi(v)}{|w|^2}w$. Then $\psi(v_2) = 0$ leads the lemma.

Lemma 4.2 (Riesz). *If ψ is a bounded bilinear functional on $\mathcal{H}$, then there exists a unique operator A such that $\psi(v, w) = (Av, w)$ for all $v, w \in \mathcal{H}$.*

5. adjoints

Theorem 5.1. *For A an operator, there exists a unique operator A^*, the adjoint of A, satisfying the identity $(Av, w) = (v, Aw)$ for all $v, w \in \mathcal{H}$.*

Definition 5.2. *An operator A is Hermitian if $A = A^*$. An operator A is normal if $||Av|| = ||A^*v||$ for all $v \in \mathcal{H}$.*

Proposition 5.3. *An operator A is nomal iff $AA^* = A^*A$.*

note. Note that in the Riesz lemma, if φ is symmetric, that is, $\varphi(v, w) = \overline{\varphi(w, v)}$, then the resulting operator will be Hermitian.

6. Projections

Definition 6.1. *If $\mathcal{M}$ is a closed subspace of $\mathcal{H}$, then elementary Hilebert space theory tells us that every vector $v \in \mathcal{H}$ has a unique decomposition $v = v_1 + v_2$, where $v_1 \in \mathcal{M}$ and $v_2 \in \mathcal{M}^\perp$. We define the projection onto $\mathcal{M}$ to be the map $P : v \mapsto v_1$. Note that P is necessarily an operator. If $\mathcal{M} = \mathbb{C}$, we denote P by 1, and if $\mathcal{M} = \{0\}$, we denote P by 0.*

Definition 6.2. *Let $\{P_i\}_{i\in I}$ be projections onto $\mathcal{M}_i$. We partially order these by $P_i \leq P_j$ if $\mathcal{M}_i \subset \mathcal{M}_j$. We further define $\sum_{i\in I} P_i$ to be the projection onto $\cup_{i\in I}\mathcal{M}_i$.*

Theorem 6.3. *An operator P is a projection if and only if it is Helmitian and idempotent ($P^2 = P$).*

note. It suffices to prove that for all $v \in \mathcal{H}$, $(Pv, v - Pv) = 0$.
note2. In finite case, the projection can be written in the form of elements in the operator, that is,

$$P = A_{12}A_{22}^{-1}$$

where A_{ij} is (i, j)-element of A.

Corollary 6.4. *If P is a projection, then for all $v \in \mathcal{H}$, $||Pv||^2 = (Pv, v)$.*

7. Spectral Measures

Let $\mathcal{B}(\mathbb{C})$ be the set of Borel sets in $\mathbb{C}$ and $P(\mathcal{H})$ the set of projections on $\mathcal{H}$.

Definition 7.1. *A spectral measure is a function $E : \mathcal{B}(\mathbb{C}) \to P(\mathcal{H})$ satisfying the following properties:*

(1) *$E(\emptyset) = 0$ and $E(\mathbb{C}) = 1$;*
(2) *If $\{B_n\}_{n\in\mathbb{N}}$ is a family of disjoint Borel sets, then $E(\bigcup B_n) = \sum E(B_n)$.*

8. Properties of Spectral measure

8.1. **Equalities and Inequalities.**

$$E(B_0 \cup B_1) + E(B_0 \cap B_1) = E(B_0) + E(B_1);$$

$$E(B_0)E(B_0 \cup B_1) = E(B_0);$$

$$E(B_0 \cap B_1) = E(B_0)E(B_1).$$

Furthermore, if $B_0 \subset B_1$, then

$$E(B_0) \leq E(B_1),$$

where the definition of "$\leq$" is given in (6.2).

8.2. **The sufficient condition for spectral measure.**

Proposition 8.1. *Suppose $E : \mathcal{B}(\mathbb{C}) \to P(\mathcal{H})$ is any function such that $\forall v, w \in \mathcal{H}$, the function $E^*(B) = (E(B)v, w)$ satisfies $E^*(\bigcup B_n) = \sum E^*(B_n)$ and that $E(\mathbb{C}) = 1$. Then E is a spectral measure.*

9. Spectral Integrals and their Associated Operators

Definition 9.1. *Given a spectral measure E,* spectral integral *is defined as the Lebesgue-Stieltjes integral*

$$\int f(\lambda)d(E(\lambda)v,w), \quad \forall v,w \in \mathcal{H}.$$

Definition 9.2. *The spectrum of a spectral measure E is $\Lambda(E) = \mathbb{C}\backslash \cup U_i$, where the union is taken over all open sets U_i for which $E(U_i) = 0$. We say that E is compact if $\Lambda(E)$ is compact.*

Theorem 9.3. *For E a compact spectral measure, there is a unique normal operator A such that $\forall v,w \in \mathcal{H}$, $\int \lambda d(E(\lambda)v,w) = (Av,w)$.*

note. First, show the boundedness of $\varphi(v,w)$ for the existence of operator. Next, show the uniqueness of the adjoint. Last, show the normality of the operator.

10. The spectrum of an Operator

Definition 10.1. *The spectrum of an operator A is the set $\Lambda(A) = \{\lambda \in \mathbb{C} | A - \lambda I \text{ is not invertible}\}$*

Theorem 10.2. *If A is an operator, then $\Lambda(A)$ is compact. In particular, if $\lambda \in \Lambda(A)$, then $||\lambda|| \leq ||A||$.*

To prove this theorem, we need a proposition here.

Proposition 10.3. *If A is any operator such that $||A - I|| < 1$, then A is invertible.*

Finally, what we assumed on the spectrum can be justified by the theorem below.

Theorem 10.4. *If E is a compact spectral measure and $A = \int \lambda dE(\lambda)$, then $\Lambda(E) = \Lambda(A)$.*

LONG RANGE DEPENDENCE

YAN LIU

1. Reference

McElroy and Holan (2014), AS.

2. notations

2.1. Notations.

1.	Z_j	$e^{-i\lambda_j}$
2.	$\langle F Z_1^{-h_1} Z_2^{-h_2} \rangle$	$\gamma_{h_1,h_2}(F)$

3. Fundamental Setting

3.1. Basics.

(i) autocovariance function (acf)

$$\mathrm{Cov}(\mathbb{Y}_{s_1,s_2}, \mathbb{Y}_{r_1,r_2} = E[\mathbb{W}_{s_1,s_2}\mathbb{W}_{r_1,r_2}] = \gamma_{s_1-r_1,s_2-r_2}.$$

for all s_1, s_2, r_1, $r_2 \in \mathbb{Z}$.

(ii) the commutativity of the field $\mathbb{Y}$

$$\gamma_{h_1,h_2}(F) = \gamma_{-h_1,-h_2}(F)$$

(iii) the spectral density

$$F(\lambda_1,\lambda_2) = \sum_{h_1,h_2\in\mathbb{Z}} \gamma_{h_1,h_2}(F) Z_1^{h_1} Z_2^{h_2}.$$

We have the following by the Fourier inversion

$$\gamma_{h_1,h_2}(F) = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} F(\lambda_1,\lambda_2) Z_1^{-h_1} Z_2^{-h_2} d\lambda_1 d\lambda_2.$$

Date: November 30, 2014.

3.2. **Assumptions.** There exists $0 < \alpha(\theta) < 1$ such that for each $\delta > 0$,

(A.1) $g(\theta) = \int_{-\pi}^{\pi} \log f(x,\theta) dx$ can be differentiated twice under the integral sign.

(A.2) $f(x,\theta)$ is continuous at all (x,θ), $x \neq 0$, $f^{-1}(x,\theta)$ is continuous at all (x,θ), and

$$f(x,\theta) = O(|x|^{-\alpha(\theta)-\delta}), \quad \text{as } x \to 0.$$

(A.3) $\partial/\partial\theta_j f^{-1}(x,\theta)$ and $\partial^2/\partial\theta_j\partial\theta_k f^{-1}(x,\theta)$ are continuous at all (x,θ),

$$\frac{\theta}{\partial\theta_j} f^{-1}(x,\theta) = O(|x|^{-\alpha(\theta)-\delta}) \quad \text{as } x \to 0, \quad 1 \leq j \leq p,$$

and

$$\frac{\partial^2}{\partial\theta_j\partial\theta_k} f^{-1}(x,\theta) = O(|x|^{\alpha(\theta)-\delta}) \quad \text{as } x \to 0, \quad 1 \leq j,k \leq p.$$

(A.4) $\partial/\partial x f(x,\theta)$ is continuous at all (x,θ), $x \neq 0$, and

$$\frac{\partial}{\partial x} f(x,\theta) = O(|x|^{-\alpha(\theta)-1-\delta}), \quad \text{as } x \to 0.$$

(A.5) $\partial^2/\partial x\partial\theta_j f^{-1}(x,\theta)$ is continuous at all (x,θ), $x \neq 0$, and

$$\frac{\partial^2}{\partial x\partial\theta_j} f^{-1}(x,\theta) = O(|x|^{\alpha(\theta)-1-\delta}) \quad \text{as } x \to 0, \quad 1 \leq j \leq p.$$

(A.6) $\partial^3/\partial x^2\partial\theta_j f^{-1}(x,\theta)$ is continuous at all (x,θ), $x \neq 0$, and

$$\frac{\partial^3}{\partial^2 x\partial\theta_j} f^{-1}(x,\theta) = O(|x|^{\alpha(\theta)-2-\delta}) \quad \text{as } x \to 0, \quad 1 \leq j \leq p.$$

4. Main Results

Proposition 4.1. *The act of the cepstral model is given by*

$$\gamma_{h_1,h_2}(F) = e^{\Theta_{0,0}} \sum_{j_1,j_2\in\mathbb{Z}} \gamma_{j_1,j_2}(\Phi)\Big[\sum_{k_1,k_2\in\mathbb{Z}} \gamma_{h_1+j_1-k_1,h_2-j_2-k_2}(\Psi)\gamma_{k_1}(\Xi)\gamma_{k_2}(\Omega)\Big],$$

where $\gamma(\Phi)$, $\gamma(\Psi)$, $\gamma(\Xi)$ and $\gamma(\Omega)$ can be calculated in terms of their coefficients, which are recursively given by

$$\begin{aligned}
\psi_{j_1,j_2} &= \frac{1}{j_1}\sum_{k_1=1}^{p_1} k_1\Big(\sum_{k_2=1}^{j_2}\psi_{j_1-k_1,j_2-k_2}\Theta_{k_1,k_2}\Big),\\
\phi_{j_1,j_2} &= \frac{1}{j_1}\sum_{k_1=1}^{p_1} k_1\Big(\sum_{k_2=1}^{j_2}\phi_{j_1-k_1,j_2-k_2}\Theta_{-k_1,k_2}\Big),\\
\xi_{j_1} &= \frac{1}{j_1}\sum_{k_1=1}^{p_1} k_1\Theta_{k_1,0}\xi_{j_1-k_1},\\
\omega_{j_2} &= \frac{1}{j_2}\sum_{k_2=1}^{p_1} k_2\Theta_{0,k_2}\omega_{j_2-k_2}
\end{aligned}$$

for $j_1 \geq 1$ and $j_2 \geq 1$.

5. Further Reading

- See Sinai (1976, TPA) for the derivation of the spectral density of long-range dependent process
- See Granger and Joyeux (1980) and Hosking (1981) for the modeling of strongly dependent phenomena.
- Fox and Taqqu (1983), technical report 590, Cornell Univ.
- Fox and Taqqu (1985), AP

6. notations

1. X_t, $t \in \mathbb{Z}$	a strongly dependent time series
2. $f(x)$	the spectral density of the time series
3. L	a slowly varying function at infinity
4. α	the exponent
5. G	a polynomial
6. Y_t	$= G(X_t)$
7. $s_\theta(x) = \sigma^2 g_\theta(x)$	the spectral density of the process Y_t
8. $L_{G,\theta}$	a slowly varying function
9. (θ, σ^2)	the parameters
10. $A_{N,\theta}$	$= \{a_\theta(t-s)\}_{t,s=1,\ldots,N}$
11. ρ_1	$= 2\sum_{t\in\mathbb{Z}}[E\dot{G}(X_t)G(X_0)]\nabla a_{\theta_0}(t)$

7. Fundamental Setting

7.1. Basics.

(i) Estimators

$$\hat{\theta}_N = \arg\min_{\theta} N^{-1} Y' A_{N,\theta} Y,$$

where $Y = (Y_1, \ldots, Y_N)$.

(ii) $a_\theta(t)$ is defined by

$$a_\theta(t) = \int_{-\pi}^{\pi} e^{its} g_\theta^{-1}(x) dx.$$

(iii) $v_{m,n}(t-s)$

$$v_{m,n}(t-s) = \frac{1}{m!n!}[EG^m(X_t)G^{(n)}(X_s)]\nabla a_{\theta_0}(t-s).$$

(iv) ρ_k

$$\rho_k = \sum_{m,n\geq 0;\, m+n=k} \sum_{t\in\mathbb{Z}} v_{m,n}(t)$$

⟨ss:7.2⟩ 7.2. Assumptions.

(i) The spectral density $f(x)$ satisfies

$$f(x) = |x|^{-\alpha} L(1/|x|), \quad x \in [-\pi, \pi], \quad (0 < \alpha < 1).$$

Remark 7.1. Note that $\alpha = 1 - 2H$ $(1/2 < H < 1)$.

(ii) g_θ satisfies

$$g_\theta(x) = |x|^{-\alpha_G(\theta)} L_{G,\theta}(1/|x|), \quad |x| \leq \pi,$$

where $0 \leq \alpha_G(\theta) < 1$.

(iii) Suppose

$$\int_{-\pi}^{\pi} \log g_\theta(x) dx = 0, \quad \theta \in \Theta. \tag{7.1}$$ eq2.3:gt1999

(iv) $(\partial^2/\partial\theta_i\partial\theta_j) g_\theta^{-1}(x)$ is a continuous function in (x, θ).

(v) For any small fixed number $\epsilon > 0$,

$$\left|\frac{\partial}{\partial\theta_j} g_\theta^{-1}(x)\right| \leq C|x|^{\alpha_G(\theta)-\epsilon}, \quad |x| \leq \pi \quad \text{for } \theta = \theta_0,$$

$$\left|\frac{\partial^2}{\partial x\partial\theta_j} g_\theta^{-1}(x)\right| \leq C|x|^{\alpha_G(\theta)-1-\epsilon}, \quad |x| \leq \pi \quad \text{for } \theta = \theta_0.$$

(vi) the spectral density f of the Gaussian sequence (X_t) satisfies

$$\left|\frac{d}{dx}f(x)\right| \le C|x|^{-\alpha-1\epsilon}, \quad |x| \le \pi,$$

where $\epsilon = \epsilon(\theta) > 0$ is any fixed number.

8. Main Results

Theorem 8.1. *Assume that* (7.1) *holds and that* $g_\theta^{-1}(x)$ *is a continuous function. Then almost surely,*

$$\lim_{N\to\infty} \hat{\theta}_N = \theta_0.$$
$$\lim_{N\to\infty} \hat{\sigma}_N^2 = \sigma_0^2.$$

⟨thm2.2:gt1999⟩ **Theorem 8.2.** *Suppose that Assumptions 7.2 hold, that* $W_{\theta_0}^{-1}$ *exists and* $\rho_1 \neq 0$. *Then*

$$\hat{\theta}_N - \theta_0 = -(2\pi\sigma_0^2)^{-1} W_{\theta_0}^{-1} \rho_1 \Big(N^{-1}\sum_{j=1}^{N} X_j\Big)(1 + o_P(1)).$$

Corollary 8.3. Theorem 8.2 implies that

$$[N^{1-\alpha}L^{-1}(N)]^{1/2}(\hat{\theta}_N - \theta_0) \xrightarrow{\mathcal{L}} (2\pi\sigma_0^2)^{-1} W_{\theta_0}^{-1}\rho_1\xi,$$

where ξ is a Gaussian random variable with zero mean and variance $E\xi^2 = 2/(\alpha(\alpha+1))$.

Example 1. In the case of $G(X_t) = X_t$, $\dot{G}(X_t) = 1$ and $E\dot{G}(X_t)G(X_t) = EX_t = 0$ and therefore $\rho_1 = 0$.

Theorem 8.4. *Let* $\rho_1 = 0$, $\rho_2 \neq 0$.

(i) If $1/2 < \alpha < 1$, *then*

$$N^{(1-\alpha)}L^{-1}(N)(\hat{\theta}_N - \theta_0) \xrightarrow{\mathcal{L}} (2\pi\sigma_0^2)^{-1} W_{\theta_0}^{-1}\rho_2 I_2,$$

where I_2 *has the Rosenblatt distribution, i.e.,*

$$I_2 = \int_{\mathbb{R}^2} \frac{\exp(it(x_1+x_2)) - 1}{i(x_1+x_2)} |x_1|^{-\alpha}|x_2|^{-\alpha} Z(dx_1)Z(dx_2), \quad \alpha > 1/2.$$

(ii) If $0 < \alpha < 1/2$, *then*

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (2\pi\sigma_0^2)^{-2} W_{\theta_0}^{-1} D W_{\theta_0}^{-1}),$$

where D *is a* $p \times p$ *matrix with entries*

$$d(i,j) = \sum_{t\in\mathbb{Z}} \left[\sum_{s_1,s_2\in\mathbb{Z}} \dot{a}_{\theta_0}^{(i)}(s_1)\dot{a}_{\theta_0}^{(j)}(s_2)\mathrm{Cov}(G(X_t)G(X_{t+s_1}), G(X_0)G(X_{s_2}))\right].$$

9. words

1. rather the exception than the rule　どちらかといえば例外的で

10. New knowledge

- The compensation effect in the Whittle estimator appears when the observations X_t are pure Gaussian or linear is rather the exception than the rule!!

MINIMAX PROBLEM

YAN LIU

1. Reference

Hosoya (1978), AP

2. Notations

1.	$\mathcal{D}_0$	the class of all probability distribution functions
2.	$\mathcal{D}_1 \subset \mathcal{D}_0$	the class of ... absolutely continuous w.r.t. the Lebesgue measure
3.	$\mathcal{F}_0(F) \subset \mathcal{D}_0$	$\{H \in \mathcal{D}_0; H = (1-\epsilon)F + \epsilon G, G \in \mathcal{D}_0\}$
4.	$\mathcal{F}_1(F) \subset \mathcal{D}_0$	$\{H \in \mathcal{D}_0; H = (1-\epsilon)F + \epsilon G, G \in \mathcal{D}_1\}$
5.	$E_m(f)$	$= \{\omega \in (-\pi, \pi]; m \geq (1-\epsilon)f(\omega)\}$
6.	$F_m(f)$	$= \{\omega \in (-\pi, \pi]; m < (1-\epsilon)f(\omega)\}$
7.	$\mathcal{L}(H)$	the class of a linear predictor
8.	$\mathcal{L}_0(f)$	$= \cap_{H \in \mathcal{F}_0(F)} \mathcal{L}(H)$
9.	$\mathcal{L}_1(f)$	$= \cap_{H \in \mathcal{F}_1(F)} \mathcal{L}(H)$
10.	λ_i	jumps at countable points with the corresponding saltuses $\Delta F(\lambda_i)$

3. Concepts and definitions

3.1. Contrast function.

$$f_m(\omega) = \begin{cases} (1-\epsilon)f(\omega) & \omega \in F_{m(f)} \\ m(f) & \omega \in E_{m(f)} \end{cases}$$

4. Results

Theorem 4.1. *There exists an optimal predictor $\phi_m \in \mathcal{L}_1$ for the spectral density f_m such that*

$$\max_{H \in \mathcal{F}_1(F)} V(\phi_m, H) = 2\pi \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_m(\omega) d\omega \right\}.$$

Proposition 4.2.

$$\max_{H \in \mathcal{F}_1(F)} V(\phi_m, H) = \min_{\phi \in \mathcal{L}_1(F)} \max_{H \in \mathcal{F}_1(F)} V(\phi, H) = 2\pi \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_m(\omega) d\omega \right\}.$$

Date: December 16, 2014.

FOR NON-STANDARD RANK TESTS

GEN RYU

Reference

- Janssen and Mason[1990]
  Non-Standard Rank Tests

1. Preface

In the models with the standard regularity assumptions, the rank tests is efficient and powerful, since the rank statistics with the exact score of the approximate score is distributed as normal distribution, and if the logarithm of ratio of the densities of null hypotheses and alternative hypotheses is distributed as normal distribution, by Cramer device, we can see that the rank statisticss are efficient and the distribution of the statistics is normal under both hypotheses from Le Cam's third lemma.

It has been well-known since Hájek and Sidak that rank tests work well under standard regularity conditions. These are assumptions concerning the differentiability of the underlying parametric model. L^1-differentiability of the densities is needed to derive locally most powerful rank tests at a finite sample size, whereas, the now famous L^2-differentiability of the square root of the densities is required to prove the asymptotic efficiency of rank tests under certain parametric alternatives.

note. Weibull location models with shape parameter $a \leq 1$ are excluded from the class.

In this book, the methodology to construct rank tests for models where the standard regularity assumptions do not hold.

note2. The treatable non-standard models where the L^2-differentiability assumption is violated can be divided into two classes:

(1) the Fisher information "just" fails to be finite, almost regular models
(2) non-regular or irregular models.

1.1. **Simple Linear Rank Statistics.** Here the simple linear rank statistic

$$T_N(R) = \sum_{i=m+1}^{N} a_N(R_i)$$

Date: June 15, 2012.

with wights given by

$$a_N(i) = E[\,(-\dot{f}(F^{-1}(U_i)))/f(F^{-1}(U_i))\,], \quad i = 1, 2, \ldots, N.$$

1.2. **Efficiency.** Efficiency means that the test reaches asymptotically the same power under the local alternatives as the corresponding Neyman-Pearson tests.

1.3. **Quick consistency.** Quick consistency means that the test has the optimal rate of convergence as measured in terms of local alternatives. More precisely, consider simple alternatives

$$\theta_N = \theta F^{-1}(1/n), \quad \theta > 0,$$

and let Ψ_N denote the Neyman-Pearson test for $\{0\}$ against $\{\theta_N\}$ at sample size N.

For level α tests and sufficiently small $\theta > 0$

$$1 > \lim_{N\to\infty} E_{\theta_N} \Psi_N > \alpha$$

and

$$1 > \lim_{N\to\infty} E_{\theta_N} \Phi_N > \alpha.$$

Theorem 1.1.

- *The score rank test is asymptotically efficient for* $a = 1$.
- *The score rank test is quickly consistent for* $0 < a < 1$.

OGATA(2010)

GEN RYU

REFERENCE

(1) [article] H. Ogata(2010)
Empirical Likelihood Estimation for a class of Stable Processes
(2) [book] M. Taniguchi and Y. Kakizawa(2000)
Asymptotic Theory of Statistical Inference for Time Series

1. MA Model with stable innovations

$$X_t = \sum_{j=0}^{\infty} \varphi_{j,\boldsymbol{\xi}} Z_{t-j}, \quad t \in \mathbb{Z}, \quad \boldsymbol{\xi} \in \Xi \subset \mathbb{R}^q \tag{1.1}$$

Here, the parameters is denoted by $\boldsymbol{\theta} = (\alpha, \gamma, \xi')' \in \Theta \subset \mathbb{R}^{2+q}$.

2. Definitions

2.1. A summary for mixing condition. [Taniguchi & Kakizawa]

2.1.1. *Uniform Mixing Condition.*

Definition 2.1. *The process $\{\boldsymbol{X}_t; t \in \mathbb{Z}\}$ is said to satisfy a uniform mixing condition if*

$$\sup_{A \in \mathcal{F}^t_{-\infty},\, B \in \mathcal{F}^{\infty}_{t+\tau}} \frac{|P(A \cap B) - P(A)P(B)|}{P(A)} \equiv \phi(\tau) \to 0 \quad \textit{as } \tau \to \infty.$$

2.1.2. *Strong Mixing Condition.*

Definition 2.2. $\{\boldsymbol{X}_t; t \in \mathbb{Z}\}$ *is said to satisfy a strong mixing condition if there exist a positive function g satisfying $g(n) \to 0$ as $n \to \infty$ so that*

$$|\, P(A \cap B) - P(A)P(B)\,| < g(r - q), \quad A \in \mathcal{F}^q_{-\infty}, \quad B \in \mathcal{F}^{\infty}_r,$$

where notation follows $\mathcal{F}^q_{-\infty} = \sigma\{\boldsymbol{X}_q, \boldsymbol{X}_{q-1} \dots\}$ and $\mathcal{F}^{\infty}_r = \sigma\{\boldsymbol{X}_r, \boldsymbol{X}_{r+1} \dots\}$.

2.1.3. *Mixing.*

Definition 2.3. $\{\boldsymbol{X}_t; t \in \mathbb{Z}\}$ *is said to be mixing if*

$$\lim_{n \to \infty} P(A \cap T^{-n}B) = P(A)P(B) \quad A, B \in \mathcal{F}.$$

Date: May 21, 2012.

2.1.4. *Ergodicity.*

Definition 2.4. *The process $\{\boldsymbol{X}_t; t \in \mathbb{Z}\}$ is said to be ergodic if for all $A \in \mathcal{A}$, either $P(A) = 0$ or $P(A) = 1$.*

note. The relation of logic above is (2.1) ⇒ (2.2) ⇒ (2.3) ⇒ (2.4).

Part 1. Empirical likelihood estimation

3. Assumptions and their interpretation

3.1. Assumptions. [Ogata(2010)]

Assumption 3.1. *The coefficient ϕ_j's satisfies the following conditions.*

- (A1) $\varphi_0 = 1$.
- (A2) $\sum_{j=0}^{\infty} | \varphi_j |^{\alpha} < \infty$.
- (A3) $\{X_t\}_{t \in \mathbb{Z}}$ *satisfies the uniform mixing condition and that the mixing coefficient $\phi(\tau)$ satisfies $\sum_{\tau} \{\phi(\tau)\}^{1/2} < \infty$.*

3.2. Interpretation of assumptions.

- (A1) It means that the model is standardized by this assumption.
- (A2) The assumption guarantees that a.s. convergence of series (**??**). See Embrechts et al.(1997, Sec 7.2) for details.
- (A3) This assumption is used for the CLT in the proof of Lemma 1 in section 6.

4. The important equation and its transformation

4.1. The theoretical characteristic function.

$$\phi_{\boldsymbol{\theta}}$$

SUMMARY ON THE SPECTRAL ANALYSIS OF TIME SERIES MODELS

GEN RYU

Consider the linear process

$$X(t) = \sum_{j=0}^{\infty} A(j)U(t-j), \quad t \in \mathbb{Z},$$

where the innovations $U(j)$ are i.i.d. s-vector random variables. The process $\{X(t)\}$ has spectral density matrix which is expressed as

$$g(\omega) = (2\pi)^{-1}k(\omega)k(\omega)^*, \quad -\pi \leq \omega \leq \pi,$$

where $k(\omega) = \sum_{j=0}^{\infty} A(j)e^{i\omega j}$. The periodogram of the process is defined as

$$I_T(\omega) = (2\pi T)^{-1}d_T(\omega)d_T(\omega)^*, \quad -\pi \leq \omega \leq \pi,$$

where $d_T(\omega) = \sum_{t=1}^{T} X(t)e^{-i\omega t}$.

1. Whittle likelihood

The multivariate Whittle likelihood is given by

$$W(\boldsymbol{\theta}) = \int_{-\pi}^{\pi} [\log\det f_\theta(\omega) + \mathrm{tr}\{f_\theta(\omega)^{-1}I_T(\omega)\}]d\omega.$$

The spectral form of a general linear process is given by

$$f_\theta(\omega) = \left(\sum_{j=0}^{\infty} B_j(\theta)e^{ij\omega}\right) \Sigma \left(\sum_{j=0}^{\infty} B_j(\theta)e^{ij\omega}\right)^*,$$

where the $B_j(\theta)$ are $s \times s$ matrices, $B_0(\theta)$ is an $s \times s$ identity matrix and Σ is an $s \times s$ symmetric matrix. Assuming that the parameter θ does not depend on Σ, which corresponds to the covariance matrix of the innovation, while the B_j depend on θ, we call θ "innovation-free".

Let θ_0 be the value defined by

$$\frac{\partial}{\partial\theta}\int_{-\pi}^{\pi} [\log\det f_\theta(\omega) + \mathrm{tr}\{f_\theta(\omega)^{-1}g(\omega)\}]d\omega\bigg|_{\theta=\theta_0} = 0, \tag{1.1}$$

which is called the pseudo-true value of θ. Here, θ_0 means the point minimizing the $D(f_\theta, g)$ under natural conditions.

Date: June 11, 2012.

SOME USEFUL RESULTS

GEN RYU

Reference

(1) [book] L.Breiman(1968)
Probability
(2) [book] P.Billingsley(1968)
Convergence of Probability Measures

Part 1. The relation between characteristic function and the distribution

1. Definitions

1.1. Definition for all probability measures.

Definition 1.1. *Let $\mathcal{D}$ denote the set of all distribution functions. A subset $\mathcal{L} \subset \mathcal{D}$ will be said to be mass-preserving if for any $\epsilon > 0$, there is a finite interval I such that*

$$F(I^c) < \epsilon \quad \text{for all } F \in \mathcal{D}.$$

In general, we have an extension of the definition as follows:

Definition 1.2. *A family Π of probability measures on the general metric space S is said to be tight if for every positive ϵ there exists a compact set K such that $P(K) > 1 - \epsilon$ for all P in Π.*

This definition introduces another definition, which is a little weaker than *tight*.

Definition 1.3. *A family Π on $(S, \mathcal{S})$ is relatively compact if every sequence of elements of Π contains a weakly convergent subsequence; that is, if for every sequence $\{P_n\}$ in Π there exist a subsequence $\{P_{n'}\}$ and a probability measure Q which defined on $(S, \mathcal{S})$ such that $P_{n'} \xrightarrow{d} Q$.*

1.2. infinitely divisible distribution.

Definition 1.4. *X will be said to have an infinitely divisible distribution if for every n, there are independent and identically distributed random variables $X_1^{(n)}$, $X_2^{(n)}$, $X_n^{(n)}$ such that $\mathcal{L}(X) = \mathcal{L}(X_1^{(n)} + X_2^{(n)} + \cdots + X_n^{(n)})$.*

Date: May 31, 2012.

2. Theorems

2.1. Theorems concerning the family of probability measures.

Theorem 2.1 (Billingsley(1968) Thm 6.1). *If Π is tight, then it is relatively compact.*

Theorem 2.2 (Billigsley(1968) Thm 6.2). *Suppose S is separable and complete. If Π is relatively compact, then it is tight.*

Corollary 2.3 (Breiman(1968) Cor 8.11). *If $F_n \xrightarrow{d} F$, $F \in \mathcal{D}$, then $\{F_n\}$ is mass-preserving.*

3. Definition 2

Definition 3.1. *A set $\mathcal{E}$ of bounded continuous functions on $\mathbb{R}$ will be called $\mathcal{D}$-separating if for any $F, G \in \mathcal{D}$,*

$$\int f dF = \int f dG, \quad \forall f \in \mathcal{E}$$

implies $F = G$.

4. The result from definition

Proposition 4.1 (Breiman Prop 8.15). *Let $\mathcal{E}$ be $\mathcal{D}$-separating, and $\{F_n\}$ mass-preserving. Then there exists an $F \in \mathcal{D}$ such that $F_n \xrightarrow{d} F$ if and only if*

$$\lim_n \int f dF_n \text{exists, for all } f \in \mathcal{E}.$$

If this holds, then $\lim_n \int f dF_n = \int f dF$, for all $f \in \mathcal{E}$.

Corollary 4.2 (Breiman Cor 8.16). *Let $\mathcal{E}$ be $\mathcal{D}$-separating and $\{F_n\}$ mass-preserving. If $F \in \mathcal{D}$ is such that $\int f dF_n \to \int f dF$, for all $f \in \mathcal{E}$, then $F_n \xrightarrow{d} F$.*

Proposition 4.3 (Breiman Prop 8.17). *$\mathcal{E}_0$, which is the family of trapezoid functions is $\mathcal{D}$-separating.*

Proposition 4.4 (Breiman Prop 8.18). *Let $\mathcal{E}$ be a class of continuous bounded functions on $\mathbb{R}$ with the property that for any $f_0 \in \mathcal{E}_0$, there exist $f_n \in \mathcal{E}$ such that $\sup |f_n(x)| < M$, for all n, and $\lim_n f_n(x) = f_0(x)$ for every $x \in \mathbb{R}$. Then $\mathcal{E}$ is $\mathcal{D}$-separating.*

5. The justification of characteristic functions

Theorem 5.1 (Breiman thm 8.28–The continuity theorem). *If F_n are distribution functions with characteristic functions $f_n(u)$ such that*

1. $\lim_n f_n(u)$ *exists for every u,*
2. $\lim_n f_n(u) = h(u)$ *is continuous at $u = 0$,*

then there is a distribution function F such that $F_n \xrightarrow{d} F$ and $h(u)$ is the characteristic function of F.

The theorem holds because the class of characteristic functions makes up the property of $\mathcal{D}$-separating, and the continuity at $u = 0$ implies mass-preserving.

Theorem 5.2 (Breiman 8.24)**.** *The set of all complex exponentials* $\{e^{iux}\}$, $u \in \mathbb{R}$ *is* $\mathcal{D}$-*separating.*

Proposition 5.3 (Breiman 8.29)**.** *There exists a constant* α, $0 < \alpha < \infty$, *such that for any distribution* F *with characteristic function* f, *and any* $u > 0$,

$$F\left(\left[-\frac{1}{u}, \frac{1}{u}\right]^c\right) \leq \frac{\alpha}{u}\int_0^u (1 - \mathrm{Ref(v)})\mathrm{dv}.$$

Part 2. The Infinitely divisible laws and Stable distributions

6. Definitions

Definition 6.1. X *will be said to have an infinitely divisible distribution if for every* n, *there are independent and identically distributed random variables* $X_1^{(n)}, \ldots, X_n^{(n)}$ *such that*

$$\mathcal{L}(X) = \mathcal{L}(X_1^{(n)} + \cdots + X_n^{(n)}).$$

Let

$$S_n = X_1^{(n)} + \cdots + X_n^{(n)},$$

then we have following proposition.

Proposition 6.2 (Breiman 9.10)**.** *If* $S_n \xrightarrow{d} X$, *then* $X_1^{(n)} \xrightarrow{d} 0$.

6.1. **Some useful equations.** Let $f(u)$ be the characteristic function of X. Therefore, since $\mathcal{L}(X) = \mathcal{L}(X_1^{(n)} + \cdots X_n^{(n)})$, there is a characteristic function $f_n(u)$ such that $f(u) = [\, f_n(u)\,]^n$ and from the proposition $f_n(u) \to 1$ uniformly in u. Then,

$$\log f(u) = n\log[\,1 - (1 - f_n(u))\,] = n(f_n(u) - 1)(1 + \epsilon_n(u)),$$

where $\epsilon \to 0$ uniformly in u. Denote by F_n the distribution function of $X_1^{(n)}$, then

$$\log f(u) = (1 + \epsilon_n(u))\int (e^{iux} - 1)n dF_n.$$

Theorem 6.3 (Breiman thm 9.17)**.** X *has infinitely divisible distribution if and only if its characteristic function* $f(u)$ *is given by*

$$\log f(u) = i\beta u - \frac{\sigma^2 u^2}{2} + \int \left(e^{iux} - 1 - \frac{iux}{1 + x^2}\frac{1 + x^2}{x^2}\nu(dx),\right)$$

where ν *is a finite measure that assigns zero mass to the origin.*

note. The time honored custom is to take $\frac{x^2}{1+x^2}$ for the change of measure.

note2. The order of the change of measures is

$$F_n \to \mu_n \to \nu_n \quad \text{(the change has using the term above)} \to \nu.$$

7. Stable laws

Definition 7.1. *A random variable X is said to have stable law if for every integer $k > 0$, and $X_1 \dots, X_k$ independent with the same distribution as X, there are constants $a_k > 0$, b_k such that*

$$\mathcal{L}(X_1 + \cdots + X_k) = \mathcal{L}(a_k X + b_k).$$

Theorem 7.2 (Breiman 9.27)**.** *Let X have a stable law. Then either X has a normal distribution or there is a number α, $0 < \alpha < 2$, called the exponent of the law and constants $m_1 \geq 0$, $m_2 \geq 0$, β such that*

$$\log f_X(u) = iu\beta + m_1 \int_0^\infty \left(e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{dx}{x^{1+\alpha}} + m_2 \int_{-\infty}^0 \left(e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{dx}{|x|^{1+\alpha}}$$

note. An equation used in Hosoya(1978) is very difficult to understand, so we give an explanation here. Define a measure μ:

$$\mu(B) = \int_B \frac{1+x^2}{x^2} \gamma(dx).$$

Then μ is σ-finite, $\mu[-a,a]^c < \infty$, for any $a > 0$, $\int_{[-a,a]} x^2 d\mu < \infty$, and

$$\begin{aligned} \varphi(a_k u) &= i a_k \beta u + \int \left(e^{iua_k x} - 1 - \frac{iua_k x}{1+x^2} \right) \mu(dx) \\ &= i d_k u + \int \left(e^{iua_k x} - 1 - \frac{iua_k x}{1+a_k^2 x^2} \right) \mu(dx), \end{aligned}$$

where

$$d_k = a_k \beta + a_k \int \left[\frac{x}{1+a_k^2 x^2} - \frac{x}{1+x^2} \right] \mu(dx).$$

Then define a change of variable measure μ_k by

$$\mu_k(B) = \mu(z; a_k z \in B),$$

to get

$$\varphi(a_k u) = i d_k u + \int \left(e^{iux} - 1 - \frac{iux}{1+x^2} \right) \mu_k(dx).$$

Theorem 7.3 (Breiman 9.32)**.** *$f(u) = e^{\varphi(u)}$ is the characteristic function of a stable law of exponent α, $0 < \alpha < 1$, and $1 < \alpha < 2$ if and only if it has the form*

$$\varphi(u) = iuc - d|u|^\alpha \left(1 + i\theta \frac{u}{|u|} \tan \frac{\pi}{2}\alpha \right),$$

where c is real, d is real and positive, and θ real such that $|\theta| \leq 1$. For $\alpha = 1$, the form of the characteristic function is given by

$$\varphi(u) = iuc - d|u| \left(1 + i\theta \frac{u}{|u|} \frac{2}{\pi} \log|u| \right),$$

with c, d, θ as above.

PROFILE LIKELIHOOD

1. Reference

Mupphy and van der Vaart (2000), JASA.

2. notations

2.1. **Notations.**

1.	θ	a low-dimensional parameter of interest
2.	η	a higher-dimensional nuisance parameter
3.	(θ, η)	the parameter
4.	$\mathrm{l}_n(\theta, \eta)$	the full likelihood
5.	$\mathrm{pl}_n(\theta)$	the profile likelihood for θ
6.	$\tilde{l}_0$	the efficient score function for θ
7.	$\tilde{I}_0$	the efficient Fisher information matrix

3. Fundamental Setting

(i) the profile likelihood for θ

$$\mathrm{pl}_n(\theta) = \sup_{\eta} l_n(\theta, \eta).$$

(ii) For any random sequence $\tilde{\theta}_n \to_p \theta_0$,

$$\log \mathrm{pl}_n(\tilde{\theta}_n) = \log \mathrm{pl}_n(\theta_0) + (\tilde{\theta}_n - \theta_0)^T \sum_{i=1}^{n} \tilde{l}_0(X_i) - \frac{1}{2} n (\tilde{\theta}_n - \theta_0)^T \tilde{I}_0 (\tilde{\theta}_n - \theta_0) + o_{P_{\theta_0,\eta_0}}(\sqrt{n}\|\tilde{\theta}_n - \theta_0\| + 1)^2. \quad (3.1)$$

eq:3.1

4. Main Results

Corollary 4.1. If (3.1) holds, $\tilde{I}_0$ is invertible, and $\hat{\theta}_n$ is consistent, then they hold that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{I}_0^{-1} \tilde{l}_0(X_i) + o_{P_{\theta_0,\eta_0}}(1). \quad (4.1)$$

{?}

$$\log \mathrm{pl}_n(\tilde{\theta}_n) = \log \mathrm{pl}_n(\hat{\theta}_n) - \frac{1}{2} n (\tilde{\theta}_n - \hat{\theta}_n)^T \tilde{I}_0 (\tilde{\theta}_n - \hat{\theta}_n) + o_{P_{\theta_0,\eta_0}}(\sqrt{n}\|\tilde{\theta}_n - \theta_0\| + 1)^2. \quad (4.2)$$

{?}

In particular, the MLE is asymptotically normal with mean 0 and covariance matrix the inverse of $\tilde{I}_0$.

Corollary 4.2. If (3.1) holds, $\tilde{I}_0$ is invertible, and $\hat{\theta}_n$ is consistent, then under the null hypothesis $H_0 : \theta = \theta_0$, the sequence $2\log(\mathrm{pl}_n(\hat{\theta}_n)/\mathrm{pl}_n(\theta_0))$ is asymptotically chi-squared distributed with d degrees of freedom.

5. Question

(1) What is d?

6. Further Reading

7. Structure

(1) Introduction
(2) Least favorable submodels
(3) Main result
(4) Examples
(5) Disscussion

ROBUSTNESS

YAN LIU

1. Reference

Li (2008), JASA.

2. notations

2.1. **Notations.**

1.	μ_t	deterministic sequence
2.	ϵ_t	random process
3.	$F_t(x)$	marginal distribution
4.	$f_t(x)$	density function
5.	$F_{ts}(u,v)$	bivariate distribution
6.	$r_{ts}(u,v)$	below
7.	w_{jt}	below
8.	d, c, N_0	positive numbers
9.	W_0, Q_0	positive definite matrixes

2.2. **Fundamental Setting.**

(i) model

$$y_t = \mu_t + \epsilon_t$$

(ii) regression coefficient

$$\hat{\beta}_{jn} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{n} |y_t - x_{jt}^T \beta|$$

(iii) difference between dependence and independence

$$r_{ts}(u,v) = F_{ts}(u,v) - F_t(u)F_s(v)$$

(iv) mean difference

$$w_{jt} = x_{jt}^T \beta_0 - \mu_t$$

Date: September 16, 2014.

2.3. **Assumptions.**

(A1) The derivative of $G_t(u)$ with respect to u, denoted as $g_t(u) := \dot{G}_t(u)$, exists for all u and satisfies $g_t(w_{jt}) = O(1)$ uniformly.

(A2) $G_t(u + w_{jt}) - G_t(w_{jt}) = g_t(w_{jt})u + O(u^{d+1})$ uniformly for $|u| \le u_0$, where $d > 0$ and $u_0 > 0$ are some constants.

(A3) $\Lambda_{jn} := n^{-1}\sum_{t=1}^{n} g_t(w_{jt})\boldsymbol{x}_{jt}\boldsymbol{x}_{jt}^T \ge \Lambda_0$ for all j and for large n, where Λ_0 is a positive definite matrix.

(A4) $C_{jkn} := n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} r_{ts}(w_{jt}, w_{ks})\boldsymbol{x}_{jt}\boldsymbol{x}_{ks}^T \ge C_0$ for all j, k, and for large n, where C_0 is a positive definite matrix and

$$r_{ts}(u, v) := G_{ts}(u, v) - G_t(u)G_s(v).$$

(A5) $\{\epsilon_t\}$ is an m-dependence process or a linear process of the form $\epsilon_t := \sum_{l=-\infty}^{\infty} \phi_l e_{t-l}$, where $\{e_t\}$ is an i.i.d. random sequence with $E|e_t| < \infty$ and $\{\phi_l\}$ is an absolutely summable deterministic sequence such that $\sum_{|l|>n^r} \phi_l = o(n^{-1})$ as $n \to \infty$ for some constant $r \in [0, 1/4)$.

3. Main results

Lemma 3.1 (Quantile Regression Lemma)**.** *Under assumptions above,*

$$\sqrt{n}\mathrm{vec}[\hat{\boldsymbol{\beta}}_{\mathrm{jn}} - \boldsymbol{\beta}_{\mathrm{j}}]_{\mathrm{j=1}}^{\mathrm{q}} \xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{\mu}_{\mathrm{n}}, \Sigma_{\mathrm{n}}), \tag{3.1}$$

as $n \to \infty$. Furthermore,

$$\sum_{t=1}^{n} \rho_\alpha(Y_t - \boldsymbol{x}_{jt}^T\hat{\boldsymbol{\beta}}_{jn}) = \sum_{t=1}^{n} \rho_\alpha(Y_t - \boldsymbol{x}_{jt}^T\boldsymbol{\beta}_j) - \frac{1}{2}\boldsymbol{\zeta}_{jn}^T\Lambda_{jn}^{-1}\boldsymbol{\zeta}_{jn} + o_P(1), \tag{3.2}$$

and the $\boldsymbol{\zeta}_{jn}$ are asymptotically jointly normal such that

$$\mathrm{vec}[\boldsymbol{\zeta}_{\mathrm{jn}}]_{\mathrm{j=1}}^{\mathrm{q}} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathrm{vec}[\mathrm{h_{jn}}]_{\mathrm{j=1}}^{\mathrm{q}}, [\mathrm{C_{jkn}}]_{\mathrm{j,k=1}}^{\mathrm{q}}). \tag{3.3}$$

4. Further setting

4.1. **Notations.**

(1) model

$$y_t = \epsilon_t$$

(2) a new regression coefficient

$$\hat{\beta}_n(\omega) := \arg\min_{\beta\in\mathbb{R}^2} \sum_{t=1}^{n} |y_t - x_t^T(\omega)\beta|$$

(3) a new periodogram

$$L_n(\omega) = \frac{1}{4}n\|\hat{\beta}_n(\omega)\|^2$$

(4) distribution $F(x)$

  (a) median

$$F(0) = 1/2$$

(b) density $f(x)$

$$f(0) > 0$$

(5)

$$\gamma_\tau = P(\epsilon_t \epsilon_{t+\tau} < 0)$$

(6)

$$S(\omega) = \sum_{\tau=-\infty}^{\infty} (1 - 2\gamma_\tau) \cos(\omega\tau)$$

(7) coef. of Gaussian

$$\eta^2 = \frac{1}{4f^2(0)}$$

(8) coef. of chi squared

$$S = \text{diag}\{S(\omega_1), S(\omega_1), \ldots, S(\omega_q), S(\omega_q)\}$$

(9) chi squared

$$Z_j \sim i.i.d.\ \chi^2(2).$$

4.2. **Assumptions.**

(B1) (A5)
(B2) $\sum_{\tau=0}^{\infty} |1 - 2\gamma_\tau| < \infty$
(B3) $f(x)$ is continuously differentiable in a neighborhood of $x = 0$
(B4) matrix condition

$$D_{jkn} = n^{-1} \sum_{t=1}^{n} x_t(\omega_j) x_t(\omega_k) = \frac{1}{2}\delta_{j-k} I + O(1).$$

5. Main results

$$n^{1/2}\text{vec}\hat{\beta}_n(\omega_j) \sim N(0, 2\eta^2 S),$$
$$L_n(\omega_j) \sim \frac{1}{2}\eta^2 S(\omega_j) Z_j$$

6. words

1.	commence	始める
2.	dwell	いる
3.	minutiae	ささいな事柄
4.	predator-prey	捕食者-犠牲者
5.	sonar	ソナー

7. Further Reading

7.1. **robustness.** Robust statistics, Maronna, Martin and Yohai

毎回のまとめの時、余った定理など

GEN RYU

REFERENCE

- cargo[1966]
  some extension of the integral test

Part 1. Cargo[1966]

1. THEOREMS

Theorem 1.1. *If f is a real-valued function defined on $[0, \infty)$ such that* $\sup\{ V_0^n f :=$ $1, 2, \dots \} < \infty$*, then* $\sum_{k=1}^{\infty} f(k)$ *and* $\int_0^{\infty} f(t)\,dt$ *converge or diverge together.*

note. Of course, the spectral density and periodogram are real-valued functions defined on $[0, \infty)$. It means that $\boldsymbol{m}$ we defined in the paper is also real-valued function. What does V mean? Let me check!

Theorem 1.2. *Let f be a nonnegative function defined on $[0, \infty)$. Then* $\sum_{k=1}^{\infty} f(k)$ *and* $\int_0^{\infty} f(t)\,dt$ *converge or diverge together provided*

$$\sup\{ V_0^n f : n = 1, 2, \dots \} < \infty,$$

where $V_0^n f$ denotes the total variation of f on $[0, n]$

note. What is the sufficient condition for finite total variation? Or spectral density already satisfies?

2. IMPORTANT EQUATION

Corollary 2.1 (Pólya p.37)**.** *If a function g ha finite total variation V on $[0, 1]$, then*

$$\left| \int_0^1 g(x)\,dx - \frac{1}{n} \sum_{k=1}^{n} g(\frac{k}{n}) \right| \leq \frac{V}{n}.$$

Calculation. If f is a function satisfying condition (5),

$$\sup\{V_0^n f : n = 1, 2, \dots \} < \infty,$$

Date: July 11, 2012.

then we have

$$\begin{aligned}\int_0^n f(t)\,dt - \sum_{k=1}^n f(k) &= n\int_0^1 f(nx)\,dx - \sum_{k=1}^n f(k)\\ &= n\left[\int_0^1 g_n(x)\,dx - \frac{1}{n}\sum_{k=1}^n g_n\left(\frac{l}{n}\right)\right],\end{aligned}$$

where $g_n(x)$ is, by definition, equal to $f(nx)$ for all x in $[\,0,1\,]$.

Theorem 2.2 (Hardy(1910))**.** *Let f be a nonnegative function which is defined and has a continuous derivative on $[\,0,\infty]$. Then $\sum_{k=1}^\infty f(k)$ and $\int_0^\infty f(t)\,dt$ converge or diverge together provided*

$$\int_0^\infty |f'(t)|\,dt < \infty.$$

2.1. others.

(3)

$$\sum_{k=1}^\infty \sup\{|f(k)-f(t)| : k-1 \le t \le k\} < \infty.$$

(6)

$$|\sum_{k=1}^n f(k) - \int_0^n f(t)\,dt| \le V_0^n f \quad n=1,2,\ldots.$$

(7)

$$\int_0^\infty f(t)\,dt - \sum_{k=1}^n f(k).$$

Part 2. something concerning stable

3. For stable

Corollary 3.1 (Can et al.(2010) Cor 3.3)**.** *Let $\alpha \in (0,2)$ and*

$$\mathcal{F}\{f \in L^2[0,\pi] : \boldsymbol{a}(f) = (a_1(f), a_2(f), \ldots,) \in l^\alpha \log l\},$$

where $\boldsymbol{a}(f)$ is defined by

$$a_h(f) = \int_0^\pi \cos(\lambda h) f(\lambda) d\lambda, \quad h \in \mathbb{Z}.$$

Then we have

$$\begin{aligned} n(\log n)^{-1/\alpha}[\,J_{n,\epsilon}(f) - a_0(f)\gamma_{n,\epsilon}(0)\,]_{f\in\mathcal{F}} &\xrightarrow{fidi} 2[\,Y(\boldsymbol{a}(f))\,]_{f\in\mathcal{F}},\\ (n/\log n)^{1/\alpha}[\,\tilde{J}_{n,\epsilon}(f) - a_0(f)\,]_{f\in\mathcal{F}} &\xrightarrow{fidi} 2[\,\tilde{Y}(\boldsymbol{a}(f))\,]_{f\in\mathcal{F}}.\end{aligned}$$

Proposition 3.2 (K&M(1996) Prop 3.5)**.** *Let* $(X_t)_{t\in\mathbb{Z}}$ *be a linear process with coefficients* $(\psi_j)_{j\in\mathbb{Z}}$ *satisfying* $\sum_{j=-\infty}^{\infty}|j||\psi_j|^\delta < \infty$ *and suppse that* $\alpha \in (0,2)$*. Furthermore, assume that* f *is defined on* $[-\pi,\pi]$ *such that* $g(\cdot) = f(\cdot)|\phi(\cdot)|^2$ *is continuous and*

$$\sum_{t=1}^{\infty}\left|\int_{-\pi}^{x} g(\lambda)\cos(t\lambda)\,d\lambda\right|^{\mu} < \infty$$

for some $x \in [-\pi,\pi]$ *and some* $0 < \mu < \alpha$*. Then*

$$(\gamma^2_{n,X}, T_n, x_n\int_{-\pi}^{x}(I_{n,X}(\lambda) - |\psi(\lambda)|^2 T_n)f(\lambda)d\lambda) \Rightarrow \left(Y_0\psi^2, Y_0, 2Z_1\left(\sum_{t=1}^{\infty}\left|\int_{\pi}^{x} g(\lambda)\cos(t\lambda)\,d\lambda\right|^{\alpha}|\right)^{1/\alpha}\right),$$

with Y_0 *independent of* $(Z_t)_{t\in\mathbb{Z}}$*.*

Lemma 3.3 (Can&Mikosch(2010) Lem 3.1)**.** *For every* $m \geq 1$*,*

$$\left(\frac{n\gamma_{n,Z}(0)}{n^{2/\alpha}}, \frac{n\gamma_{n,Z}(h)}{(n\log n)^{1/\alpha}}, h = 1,\dots,m\right) \Rightarrow (Y_0, Y_1, \dots, Y_m).$$

what is important here is that the self-normalized periodogram is employed in the functional form, that is

$$J_{n,X}(f) = \int_0^{\pi} I_{n,X}(\lambda)f(\lambda)\,d\lambda,$$

where f is any function for appropriate classes of real-valued functions $f \in \mathcal{F}$ on $[0,\pi]$.

Let $Y(\boldsymbol{a})$ and $\tilde{Y}(\boldsymbol{a})$ be defined as

$$Y(\boldsymbol{a}) = \sum_{k=1}^{\infty} a_k Y_k,$$
$$\tilde{Y}(\boldsymbol{a}) = Y(\boldsymbol{a})/Y_0.$$

Before defining the class $\mathcal{F}$, we will have a proper class for $\boldsymbol{a}$,

$$\boldsymbol{a} \in l^\alpha \log l = \left\{\boldsymbol{a} = (a_1, a_2, \dots) \in l^\alpha : \sum_{k=1}^{\infty}|a_k|^\alpha \log^+\frac{1}{|a_k|} < \infty\right\}.$$

ROBUSTNESS

YAN LIU

1. Reference

Künsch (1984), AS.

2. notations

2.1. Notations.

1. $h(x)$	a known probability density on $\mathbb{R}$
2. σ^2	variance of U_i
3. $\rho(x,n)^m$	m-dimensional marginal distribution of stationary processes
4. $\mathcal{M}^m$	the set of $\rho(x,n)^m$
5. $\theta \in \Theta \subset \mathbb{R}^q$ $(q \leq p+2)$	unknown parameter
6. T	a functional $\mathcal{M}^m \to \Theta$ (or restrict T to a certain subset of $\mathcal{M}^m$)
7. $\gamma^* = \sup_x \lvert \mathrm{IC}_T(x,\theta) \rvert$	gross error sensitivity
8. $\theta = (\theta_1, \theta_2)$	
9. $\theta_1 = \sigma$ and $\theta_2 = (\eta, \beta_1, \ldots, \beta_p)$	
10. $\kappa = (\kappa_1, \kappa_2)$	
11. $\psi = (\psi_1, \psi_2)$	

2.2. Fundamental Setting.

(i) AR(p) process

$$(X_i - \eta) = \sum_{k=1}^{p} \beta_k (X_{i-k} - \eta) + U_i, \quad \text{i.i.d. } U_i$$

Using $x_i^* = x_i - \eta$, κ is defined by

$$\kappa(x_1, \ldots, x_{p+1}; \theta) = \frac{\partial}{\partial \theta} \log \frac{1}{\sigma} h\left(\frac{x_{p+1}^* - \sum \beta_k x_{p+1-k}^*}{\sigma} \right)$$

Furthermore, let u denote

$$u = x_{p+1}^* - \sum \beta_k x_{p+1-k}^*$$

(ii) m-dimensional marginal distributions

Date: August 24, 2014.

(a) Define $x_i = x_{i-kn}$ if $i > kn$ for $k \in \mathbb{N}$;
(b) m-dimensional marginal distributions

$$\rho(x,n)^m = n^{-1}\sum_{i=1}^{n}\delta(x_i,\ldots,x_{i+m-1}),$$

where $\delta(x_i,\ldots,x_{i+m-1})$ is the point mass at $x \in \mathbb{R}^m$.

(iii) M-estimator defined by

$$\sum_{j=1}^{n-m+1}\psi(x_j,\cdots,x_{j+m-1};\hat{\theta}_n) = 0.$$

(iv) Choice of the functional T : for $T(\mu_\theta^m) = \theta$,

$$\hat{\theta}_n(x_1,\ldots,x_n) = T(\rho(x,n)^m).$$

(v) Any version of the influence function $\mathrm{IC}_T(x,\theta)$

$$\int \mathrm{IC}_T(x,\theta)\mu_\theta(dx_m|x_{m-1},\cdots,x_{m-p}) = 0.$$

(vi) Asymptotical variance-covariance matrix

$$C(T,\theta) = \int \mathrm{IC}_T(x,\theta)\mathrm{IC}_T(x,\theta)^T\mu_\theta^m(dx).$$

Lemma 2.1. *If $\{X_i\}_{i\in\mathbb{Z}}$ is stationary ergodic process, then*

$$\rho(x,n)^m \to \mu^m \quad as\ n \to \infty.$$

2.3. **Hampel's optimality problem.** Minimize the trace of the asymptotic covariance matrix $C(T,\theta)$ among all estimators of (iv) which have an influence function and for which

$$\gamma^* = \sup_x|\mathrm{IC}_T(x,\theta)| \le c(\theta).$$

2.4. **Huber function.**

$$H_c(x) = x\min(1,\frac{c}{|x|})$$

3. Fundamental Theorems

Theorem 3.1 (Künsch (1984), Theorem 1.1)**.** *A functional $L : \mathcal{M}^m \to \mathbb{R}$ is of the form $L(\nu^m) = \int t(x)\nu^m(dx)$ with t bounded and continuous iff L is affine and weakly continuous.*

Theorem 3.2 (Künsch (1984), Theorem 1.2)**.** *$\int t(x)\nu^m(dx) = 0$ for all $\nu^m \in \mathcal{M}^m$ iff $t(x_1,\ldots,x_m) = g(x_1,\ldots,x_{m-1}) - g(x_2,\cdots,x_m)$ with an arbitrary g.*

Theorem 3.3 (Künsch (1984), Theorem 1.3)**.** *Let μ denote the distribution of an AR(p)-process. If $\mathbb{R}^m \to \mathbb{R}$, $m > p$, is continuous, $\sup|f(x)|/(1+|x|) < \infty$ and $\int f(x)\mu^m(dx) = 0$, then there exists a continuous function $g : \mathbb{R}^{m-1} \to \mathbb{R}$ with $\sup|g(x)|/(1+|x|) < \infty$ and*

$$\int f(x_1,\ldots,x_m) + g(x_1,\ldots,x_{m-1}) - g(x_2,\cdots,x_m)\mu(dx_m|x_{m-1},\cdots,x_{m-p}) = 0$$

for all $x_1, \ldots, x_{m-1}$. g *is unique up to an additive constant.*

4. OPTIMAL ROBUST ESTIMATORS

Theorem 4.1. *Suppose* σ *is known and* $h(x) = h(-x)$. *If the bound* $c(\theta)$ *is such that*

$$\int H_{c(\theta)}(A(\theta)\kappa(x,\theta))\kappa(x,\theta)^T \mu_\theta^{p+1}(dx) = \mathrm{Id}$$

has a solution $A(\theta)$ *for all* θ, *then the solution of Hampel's optimal problem is given by*

$$\begin{cases} m = p+1, \\ \psi(x,\theta) = H_{c(\theta)}(A(\theta)\kappa(x,\theta)). \end{cases}$$

Theorem 4.2. *If* $\psi_1(x_1, \ldots, x_{p+1}; \theta) = \chi\left(\frac{u}{\sigma}\right)$ *with* $\chi(\cdot)$ *even and* ψ_2 *is one of the optimal solutions of Hampel's problem, then* $\hat{\sigma}$ *is asymptotically independent of* $\hat{\theta}_2$ *and the asymptotic covariance for* $\hat{\theta}_2$ *is the same as for known* σ.

5. WORDS

6. Further Reading

6.1. **robustness.** Robust statistics, Maronna, Martin and Yohai

SUMMARY=LINEAR SERIAL RANK TESTS FOR RANDOMNESS AGAINST ARMA ALTERNATIVES

GEN RYU

Let $a_1, \dots, a_{p_1}, b_1, \dots, b_{p_2}$ be an arbitrary $(p_1 + p_2)-$tuple of real numbers, and consider the sequence

$$X_t - n^{-1/2} \sum_{i=1}^{p_1} a_i X_{t-i} = \epsilon_t + n^{-1/2} \sum_{i=1}^{p_2} b_i \epsilon_{t-i}, \quad t \in \mathbb{Z}, \quad n = 1, 2, \dots$$

of stochastic difference equations. For n sufficiently large, all the roots of the characteristic equation

$$z^{p_1} - n^{-1/2} \sum_{i=1}^{p_1} a_i z^{p_1 - i} = 0, \quad z \in \mathbb{C}$$

lie inside the unit-circle.

The hypothesis is denoted by H_0:

$$l_n^0(\boldsymbol{x}) = \prod_{t=1}^{n} f(x_t),$$

The alternative is denoted by H_1:

$$l_n^1(\boldsymbol{x}) = \int \prod_{t=1}^{n} f(x_t - n^{-1/2} \sum_{i=1}^{p_1} a_i x_{t-i} + \sum_{u=1}^{t-1} g_u (x_{t-u} - n^{-1/2} \sum_{i=1}^{p_1} a_i x_{t-u-i}) + \sum_{u=t}^{t+p_2-1} g_u e_{t-u})$$
$$dG_{ab}(x_{-p_1+1}, \dots, x_0) f(e_{-p_2+1}) \dots f(e_0)\, de_{-p_2+1} \dots de_o.$$

Furthermore, consider the likelihood ratio

$$L_n(\boldsymbol{x}) = \begin{cases} \dfrac{l_n^1(\boldsymbol{x})}{l_n^0(\boldsymbol{x})} & \text{if } l_n^0(\boldsymbol{x}) > 0 \\ 1 & \text{if } l_n^1(\boldsymbol{x}) = l_n^0(\boldsymbol{x}) = 0 \\ \infty & \text{if } l_n^1(\boldsymbol{x}) > l_n^0(\boldsymbol{x}) = 0. \end{cases}$$

Date: February 4, 2013.

1. DEFINITIONS

1.1. **linear serial rank statistics in the paper.**

$$S_n = \frac{1}{n-p}\sum_{t=p+1}^{n} a_n(R_t^{(n)}, R_{t-1}^{(n)}, \ldots, R_{t-p}^{(n)})$$

where $a_n(\cdots)$ is some given score function and $R_t^{(n)}$ is the rank of the observation made at time t in an observed series of length n.

Lemma 1.1. *The variance of* $(n-p)S_n = \sum_{t=p+1}^{n} a_n(R_t^{(n)}, R_{t-1}^{(n)}, \ldots, R_{t-p}^{(n)})$ *is*

$$D^2((n-p)S_n) = \tag{1.1}$$

$$(n-p)\mathrm{Var}(a(R_{p+1},\ldots,R_1)) + 2\sum_{i=1}^{p}(n-p-i)\mathrm{Cov}(a(R_{p+1+i},\ldots,R_{i+i}), a(R_{p+1},\ldots,R_1))$$
$$+ [\,(n-3p)(n-3p-1) + p(2n-5p-1)\,]\mathrm{Cov}(a(R_{2p+2},\ldots,R_{p+2}), a(R_{p+1},\ldots,R_1)).$$

Corollary 1.2.

$$D^2((n-p)S_n) \le (n-p)(2p+1)\mathrm{Var}(a(R_{p+1},\ldots,R_1)) + n\mathrm{Cov}(a(R_{2p+2},\ldots,R_{p+2}), a(R_{p+1},\ldots,R_1)). \tag{1.2}$$

Lemma 1.3.

$$E[\,a(R_{p+1},\ldots,R_1)\,] \tag{1.3}$$
$$= \frac{(n-p-1)(n-p-2)\cdots(n-2p-1)}{n(n-1)\cdots(n-p)} E[\,a(R_{p+1},\ldots,R_1)|R_{2p+2},\ldots,R_{p+2}\,]$$
$$+ \sum_{l=1}^{p+1}\frac{(n-p-1)(n-p-2)\cdots(n-2p-1+l)}{n(n-1)\cdots(n-p)} \sum_{p+2\le j_1\ne\cdots j_l\le 2p+2}\ \sum_{1\le k_1<k_2<\cdots<k_l\le p+1}$$
$$E[\,a(R_{p+1},\ldots,R_{k_l+1}R_{j_l}R_{k_l-1},\ldots,R_{k_1+1}R_{j_1}R_{k_1-1},\ldots,R_1)|R_{2p+2},\ldots,R_{p+2}\,]$$

Lemma 1.4.

$$n|\mathrm{Cov}(a(R_{p+1},\ldots,R_1), a(R_{2p+2},\ldots,R_{p+2}))| \le KE[\,a^2(R_{p+1},\ldots,R_1)\,]$$

1.2. **score functions and score-generating function.** The authors assume that the score functions $a_n(\cdots)$ are such that there exists a function $J = J(v_{p+1}, v_p, \ldots, v_1)$, defined over $[0,1]^{p+1}$, such that

$$0 < \int_{[0,1]^{p+1}} J^2(v_{p+1},\cdots,v_1)\,dv_{p+1}\cdots dv_1 < \infty$$

and

$$\lim_{n\to\infty} E[\,(J(U_{p+1},\cdots,U_1) - a_n(R_{p+1}^{(n)},\cdots,R_1^{(n)}))^2|H_0^{(n)}] = 0$$

This assumption is satisfied most of the time when a_n is of the form

$$a_n(i_1, i_2, \cdots, i_{p+1}) = J(\frac{i_1}{n+1}, \frac{i_2}{n+1}, \cdots, \frac{i_{p+1}}{n+1}).$$

1.3. **a discrete-time stationary white noise.** a sequence $\{\epsilon_t; t \in \mathbb{Z}\}$ of independent and identically distributed random variables with means $E[\epsilon_t] = 0$, $t \in \mathbb{Z}$.

1.4. **finite Fisher's information related to the location parameter.** $f(x)$ is absolutely continuous on finite intervals, and

$$0 < I(f) = \int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\,dx < \infty.$$

1.5. **distribution function.**

$$F^{-1}(u) = \inf\{x | F(x) \geq u\}, \quad 0 < u < 1.$$

Put

$$\phi(F^{-1}(u)) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}, \quad 0 < u < 1.$$

This function can also be written a.e. as

$$\phi(x) = -\frac{f'(x)}{f(x)}, \quad x \in \mathbb{R}.$$

There are many properties like:

(1)
$$\int_{-\infty}^{\infty} \phi(x) f(x)\,dx = 0;$$

(2)
$$\int_{-\infty}^{\infty} \phi^2(x) f(x)\,dx = I(f);$$

(3)
$$\int_{-\infty}^{\infty} x\phi(x) f(x)\,dx = 1;$$

(4)
$$\int_{-\infty}^{\infty} \phi'(x) f(x)\,dx = I(f).$$

(5) $\sigma^2 I(f)$ is independent of the scale transformation.

1.6. **the Green's function.** g_u associated with the operator $C(L) = 1 + \sum_{i=1}^{p} c_i L^i$ (the c_i's are real and $c_p \neq 0$) is the value in $t = u$ of ψ_t of the homogeneous difference equation

$$\psi_t + \sum_{i=1}^{p} c_i \psi_{t-i} = 0, \quad t \in \mathbb{Z},$$

taking on initial values $\psi_0 = 1, \psi_{-1} = \cdots = \psi_{-p+1} = 0$.

Lemma 1.5.

$$\sum_{u=v}^{w} |g_u| \quad \leq \sum_{u=[\,(v-1)/p]+1}^{w} (pb_M n^{-1/2})^u \quad \forall w \geq v \geq 0.$$

Corollary 1.6. *For* $n > 4(pb_M)^2$,

$$\sum_{u=pv+1}^{\infty} \leq 2(pb_M n^{-1/2})^{v+1} = o(n^{-v/2}) \quad \forall v \geq 0.$$

1.7. **the distribution function of p_1 successive values of $\{X_t\}$.** $G_{ab}(x_{t+1}, \ldots, x_{t+p_1})$.

1.8. **ARE–asymptotic relative efficiency.** (TRT pp.317) Assume that the asymptotically most powerful test for H against q may be based on a statistic S_0 which is asymptotically normal $(0, \sigma^2)$ under H_0 and which is asymptotically normal (μ_0, σ_o^2) under q. Further, consider another test based on a statistic S, which is asymptotically normal $(0, \sigma^2)$ under H_0, and which is asymptotically normal $(\mu.\sigma^2)$ under q. Then the number

$$e = \left(\frac{\mu \sigma_o}{\mu_0 \sigma} \right)^2$$

will be called the asymptotic efficiency of the S-test. It is also called the Pitman asymptotic efficiency.

For the definition in the paper, the efficiency is extended to any two linear serial rank statistics. A test statistic $\bar{S}_n$ such that $e(\bar{S}_n, S_n) \geq 1$ for any linear serial rank statistic S_n will be asymptotically the most efficient statistic (in Pitman's sense) within the class of linear serial rank statistics.

1.9. **run statistic.**

$$a_n(i_1, i_2) = \begin{cases} 1 & \text{if} \quad (2i_1 - n - 1)(2i_2 - n - 1) < 0 \\ 0 & \text{if} \quad (2i_1 - n - 1)(2i_2 - n - 1) \geq 0. \end{cases}$$

1.10. **turning point statistic.**

$$a_n(i_1, i_2, i_3) = \begin{cases} 1 & \text{if} \quad i_1 > i_2 < i_3 \\ 1 & \text{if} \quad i_1 < i_2 < i_3 \\ 0 & \text{elsewhere.} \end{cases}$$

2. THE MAIN POINTS OF THE PAPER

(1) A class of linear serial rank statistics is introduced to test white noise against alternatives of ARMA serial dependence.
(2) Using LeCam's notion of contiguity, the asymptotic normality of the proposed statistics is established.
(3) An explicit formulation of the asymptotically most efficient score generating functions is provided.
(4) The asymptotic relative efficiency of the proposed procedures is studied with respect to their normal theory counterparts based on sample autocorrelations.

この論文のすごいところは、n samples $\sim (0, ?)$ を手にした時、この n samples は確かに中心極限定理により、漸近的に正規分布に従いそうだが、実際正規分布に従ったところで、これらの標本が独立か、ARMA の形を持つ従属標本かについて rank statistics による最適検定が可能であることを示した点である。

3. THE PROCESS OF THE PAPER

(1) Introduction
(2) Notation and basic assumption
(3) Asymptotic distribution of likelihood ratios (pp.1160)
(4) Asymptotic distribution of linear serial rank statistics (pp.1162)
(5) Asymptotic efficiency of linear serial rank statistics (pp.1166)
(6) Examples and the table (pp.1172)
(7) Appendix 1-6 (pp.1173/pp.1173/pp.1176/pp.1178/pp.1179/pp.1180)
(8) References (pp.1181)

4. BASIC ASSUMPTIONS AND THEOREMS

4.1. **assumptions.**

Let $\{\epsilon_t; t \in \mathbb{Z}\}$ be a discrete-time stationary white noise. Assume that it has a density $f(x)$, and that the following conditions are satisfied:

(1) ϵ_t has finite moments up to the third order; denote its variance by σ^2.
(2) $f(x)$ is a.e. derivative, and its derivative $f'(x)$ satisfies

$$\int_{-\infty}^{\infty} |f'(x)| dx < \infty$$

(3) $f(x)$ has finite Fisher's information $I(f)$.
(4) Assume $\phi(x)$ is a.e. derivative, and its derivative $\phi'(x)$ satisfies a Lipschitz condition

$$|\phi'(x) - \phi'(y)| < A|x - y|, \quad a.e.$$

4.2. **asymptotic distribution of likelihood ratios.** First, the authors established the contiguity between the hypotheses:

Proposition 4.1. *Under* H_0,

$$\log L_n(X_1,\dots,X_n) = \mathcal{L}_n^0(X_1,\dots,X_n) - \frac{d^2}{2} + 0_p,$$

where

$$\mathcal{L}_n^0(\boldsymbol{X}) = n^{-1/2}\sum_{t=p+1}^{n}\phi(X_t)\sum_{i=1}^{p} d_i X_{t-i}$$

$$d_i = \begin{cases} a_i + b_i & 1 \le i \le \min(p_1,p_2) \\ a_i & p_2 < i \le p_1 \qquad \text{if } p_2 < p_1 \\ b_i & p_1 < i \le p_2 \qquad \text{if } p_1 < p_2 \end{cases}$$
$$p = \max(p_1,p_2) \quad \text{and} \quad d^2 = \textstyle\sum_{i=1}^{p} d_i^2\sigma^2 I(f).$$

Moreover, $\mathcal{L}_n^0 \xrightarrow{d} \mathcal{N}(0,d^2)$.

The form of this asymptotic distribution shows that, for n sufficiently large, there will be little difference, from a statistical point of view, between AR, MA and ARMA models!!

4.3. **asymptotic distribution of linear serial rank statistics.** In the paper, the authors proposed the linear serial rank statistics for the models as follows:

$$S_n = \frac{1}{n-p}\sum_{t=p+1}^{n} a_n(R_t^{(n)}, R_{t-1}^{(n)},\dots,R_{t-p}^{(n)}),$$

$$m_n = E[S_n|H_0^{(n)}] = \frac{1}{n(n-1)\cdots(n-p)}\sum_{1\le i_1\ne\cdots\ne i_{p+1}\le n} a_n(i_1,\cdots,i_{p+1}).$$

The authors established the asymptotic equivalence of $(n-p)^{1/2}(S_n - m_n)$ with $\mathcal{S}_n - \mathcal{E}_n$, where

$$\mathcal{S}_n(\boldsymbol{X}) = (n-p)^{-1/2}\sum_{t=p+1}^{n} J(F(X_t), F(X_{t-1}),\dots,F(X_{t-p}))$$

$$\mathcal{E}_n(\boldsymbol{X}) = \frac{(n-p)^{1/2}}{n(n-1)\cdots(n-p)}\sum_{1\le t_1\ne\cdots\ne t_{p+1}\le n} J(F(X_{t_1}),\dots,F(X_{t_{p+1}})).$$

It is also established that $n^{-1/2}(\mathcal{S}_n - \mathcal{E}_n)$ and $n^{-1/2}\mathcal{L}_n^0$ are asymptotically equivalent to U-statistics.

Proposition 4.2. *Under* H_0,

$$\begin{pmatrix}\sqrt{n}(S_n - m_n)\\ \log L_n\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix}0\\ -\frac{1}{2}\sum_{i=1}^{p} d_i^2\sigma^2 I(f)\end{pmatrix}, \begin{pmatrix}V^2 & \sum_{i=1}^{p} d_i C_i\\ \sum_{i=1}^{p} d_i C_i & \sum_{i=1}^{p} d_i\sigma^2 I(f)\end{pmatrix}\right),$$

where

$$V^2 = \int_{[0,1]^{p+1}} [\, J^*(v_{p+1}, \cdots, v_1)\,]^2 dv_1 \cdots dv_{p+1} + 2\sum_{j=1}^{p} \int_{[0,1]^{p+1+j}} J^*(v_{p+1}, \cdots, v_1) J^*(v_{p+1+j}, \cdots, v_{1+j}) dv_1 \cdots dv_{p+1+j} \tag{4.1}$$

and

$$C_i = \int_{[0,1]^{p+1}} J^*(v_{p+1}, \cdots, v_1) \sum_{j=0}^{p-i} \phi(F^{-1}(v_{p+1-j}))F^{-1}(v_{p+1-j-i}) dv_1 \cdots dv_{p+1}$$

Proposition 4.3. *Under H_1,*

$$\sqrt{n}(S_n - m_n) \xrightarrow{d} \mathcal{N}(\sum_{i=1}^{p} d_i C_i, V^2).$$

4.4. **asymptotic efficiency of linear serial rank statistics.**

Proposition 4.4. *An asymptotically optimal linear serial rank test for H_0 against H_d is provided by any statistic S_n^d with score-generating function (up to additive and multiplicative constants) given by*

$$J^d(v_{p+1}, \cdots, v_1) = \sum_{i=1}^{p} \frac{d_i}{p+1-i} \sum_{j=0}^{p-i} \phi(F^{-1}(v_{p+1-j}))F^{-1}(v_{p+1-j-i}).$$

Under $H_h (h \in \mathbb{R}^p)$,

$$n^{1/2}(S_n^d - m_n^d) \xrightarrow{d} \mathcal{N}(\sum_{i=1}^{p} h_i d_i \sigma^2 I(f), V_d^2),$$

where $V_d^2 = \sum_{i=1}^{p} d_i^2 \sigma^2 I(f)$.

This optimality result relies on the following lemma.

Lemma 4.5. *Let S_n be a linear rank statistic with score-generating function $J^*(v_{p+1}, \cdots, v_1)$, and let*

$$J_0^*(v_{p+1}, \cdots, v_1) = (\sigma^2 I(f))^{-1} \sum_{i=1}^{p} \frac{C_i}{p+1-i} \sum_{j=0}^{p-i} \phi(F^{-1}(v_{p+1-j}))F^{-1}(v_{p+1-j-i}).$$

Denote by S_n^0 a linear serial rank statistic associated with J_0^*. Then $e(S_n, S_n^0) \leq 1$ for any alternative H_n^d.

Proposition 4.6. *Under H_0,*

$$\begin{pmatrix} \sqrt{n} r_1 \\ \vdots \\ \sqrt{n} r_p \\ \log L_n \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\frac{1}{2}\sum_{i=1}^{p} d_i^2 \sigma^2 I(f) \end{pmatrix}, \begin{pmatrix} & & & d_1 \\ & I & & \vdots \\ & & & d_p \\ d_1 & \dots & d_p & \sum d_i^2 \sigma^2 I(f) \end{pmatrix}\right).$$

Corollary 4.7. *Under H_d,*

$$n^{1/2}\sum_{k=1}^{p}\alpha_k r_k \xrightarrow{d} \mathcal{N}(\sum_{i=1}^{p}\alpha_i d_i, \sum_{i=1}^{p}\alpha_i^2)$$

Corollary 4.8. *The asymptotically most efficient (in Pitman's sense) linear combination of the r_k's against H_d is $\sum_{k=1}^{p} d_k r_k$.*

Proposition 4.9. *Denote by e the ARE of the asymptotically optimal serial rank statistic with respect to the asymptotically optimal combination of autocorrelations $\sum_{k=1}^{p} d_k r_k$, then we obtain*

$$e = \sigma^2 I(f).$$

5. THE HISTORY OF NONPARAMETRIC METHODS

(1) Dufour et al. (1982) and
(2) Hotelling and Pabst (1936) → run test / turning point test
(3) Wald and Wolfowitz (1943) → Spearman's autocorrelation coefficient
(4) Jogdeo (1968) → not adapted to time-seiries situations
(5) Knoke (1977) → the power of serial rank procedures / ARE / Spearman's first-order autocorrelation coefficient / turning point statistic
(6) Gupta and Govindarajulu (1980) → locally most powerful rank statistic (particular case)
(7) Aiyar (1981) → van der Waerden statistic (particular case)
(8) Bartels (1982) → von Neumann's test / more efficient than the run test / parametric von Neumann test
(9) Bell et al. (1970) → a highly systematic and theoretically-based approach
(10) Dufour (1982) → sign / Wilcoxon / signed rank / van der Waerden tests
(11) Govindarajulu and Dwass (1983) the same to the above
(12) Govindarajulu (1983) An overall review of some of these procedures

6. WORDS

(1) tractable 扱いやすい
(2) viz. = that is, namely すなわち
(3) scattered 散在している
(4) piecemeal すこしずつの、ばらばらの

MATERIALS FOR 5/12

GEN RYU

1. Definitions

1.1. **discrete sequences of estimators.** [Bickel(1982)] The discrete sequences of estimators $\{\bar{\theta}_n\}$ satisfies that $\bar{\theta}_n$ is given by one of the vertices of $\{\theta : \theta = n^{-1/2}(i_1, \dots, i_{p+q}), i_j \in Z\}$ nearest to θ_n.

Part 1. Local asymptotic normality for ARMA process

2. Theorems

Theorem 2.1 (LAN property for ARMA models)**.** *Let $\{h_n\} \subset \mathbb{R}^{p+q}$ be a bounded sequence and $\theta_n = \theta_0 + n^{-1/2}h_n$. Under our assumptions (A1), (A2) and (A3), we have for*

$$\Delta_n(\theta) = \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(e_j(\theta)) Z(j-1; \theta_0, \theta), \quad \dot{\varphi} = -f'/2f,$$

the following two results:

$$\log[\, dP_{n,\theta_n}/dP_{n,\theta_0}\,] - h_n^T \Delta_n(\theta_0) + \frac{1}{2} h_n^T I(f)\Gamma(\theta_0) h_n \to 0,$$

in P_{n,θ_0}-probability, where $\Gamma(\theta_0)$ is defined in Theorem 3.5 below (approximation of the log-likelihood ratio).

$$\mathcal{L}(\Delta_n(\theta_0)|P_{n,\theta_0}) \Rightarrow \mathcal{N}(0, I(f)\Gamma(\theta_0)),$$

where "$\Rightarrow$" denotes weak convergence (asymptotic normality of the approximating statistic).

Corollary 2.2. *Under the same assumption as above $\{P_{n,\theta_0}\}$ and $\{P_{n,\theta_n}\}$ are contiguous in the sense of Definition 2.1, Roussas (1972), page 7, and*

$$\mathcal{L}(\Delta_n(\theta_0) - I(f)\Gamma(\theta_0)h_n | P_{n,\theta_n}) \Rightarrow \mathcal{N}(0, I(f)\Gamma(\theta_0)).$$

Date: May 15, 2012.

3. The sufficient conditions for local asymptotic normality

The 4 theorems below guarantee that the sufficient conditions are fulfilled.

Theorem 3.1. *For each $\theta_0 \in \Theta$, the random functions $\phi_j(\theta_0, \cdot)$ are differentiable in q.m. $[P_{\theta_0}]$ uniformly in $j \geq 1$. That is, there are $(p+q)$-dimensional r.v.'s $\dot{phi}_j(\theta_0) = \dot{\varphi}(e_j^0)Z(j-1; \theta_0, \theta_0) = \dot{\varphi}(e_j^0)Z^0(j-1)$ [the q.m. derivative of $\phi_j(\theta_0, \theta)$ with respect to θ at θ_0] such that*

$$\frac{\phi_j(\theta_0, \theta_0 + \lambda h) - 1}{\lambda} - h^T \dot{\phi}_j(\theta_0) \to 0, \quad \textit{in q.m. } [P_{\theta_0}] \textit{ as } \lambda \to 0$$

uniformly on bounded sets of $h \in \mathbb{R}^{p+q}$ and uniformly in $j \in \mathbb{N}$. Finally, $\dot{\phi}_j(\theta_0)$ is measurable with respect to $\mathcal{A}_j$.

Theorem 3.2. *For each $\theta_0 \in \Theta$ and each $h \in \mathbb{R}^{p+q}$, the sequence $\{(h^T \dot{\phi}_j(\theta_0))\}, j \in \mathbb{N}$, is uniformly integrable with respect to P_{θ_0}.*

Theorem 3.3. *For each $\theta_0 \in \Theta$ and $j \geq 1$ let the $(p+q) \times (p+q)$-dimensional covariance matrix $\Gamma_j(\theta_0)$ be defined by*

$$\Gamma_j(\theta_0) = 4E_{\theta_0}[\, \dot{\phi}_j(\theta_0) \dot{(\phi)}_j^T(\theta_0)\,] = I(f)E_{\theta_0}[\, Z(j-1; \theta_0, \theta_0)Z^T(j-1; \theta_0, \theta_0)\,].$$

Then $\Gamma_j(\theta_0) \to \Gamma(\theta_0)I(f)$, as $j \to \infty$, in any one of the standard norms in $\mathbb{R}^{p+q}$, and $\Gamma(\theta_0)$ is positive definite.

Theorem 3.4. *(i) For each $\theta_0 \in \Theta$, each $h \in \mathbb{R}^{p+q}$ and for the probability measure P_{θ_0}, the WLLN holds for the sequence $\{[\, h^T \dot{\varphi}_j(\theta_0)\,]^2, j \in \mathbb{N}\}$. Also*
(ii)

$$\frac{1}{n}\sum_{j=1}^{n}\{E_{\theta_0}[\,(h^T \dot{\varphi}_j(\theta_0))^2 | \mathcal{A}_{j-1}\,] - [\, h^T \dot{\varphi}_j(\theta_0)\,]^2\} \to 0, \quad \textit{as } n \to \infty,$$

in P_{θ_0}-probability.

Part 2. Existence and construction of LAM estimates

Lemma 3.5. *Under assumptions (A.1), (A.2), (A.3) we have for any sequence $\{Z_n\}$ of estimates the following implication:*

$$\sqrt{n}(Z_n - \theta_0) - \frac{\Gamma(\theta_0)^{-1}}{I(f)}\Delta_n(\theta_0) = o_{P_{\theta_0}}(1) \quad (\{Z_n\} \textit{ is called } \theta_0\textit{-regular})$$

implies that $\{Z_n\}$ is LAM.

4. Assumptions and the interpretation of it

Assumption 4.1. *There exists a sequence $\{\bar{\theta}_n\}$ of estimators which satisfies*

$$\sqrt{n}(\bar{\theta}_n - \theta_0) = O_{P_{\theta_0}}(1).$$

This assumption holds for estimators for which the usual CLT is valid, i.e., for all the standard estimators. (See Anderson(1971) Thm 5.5.7). Fuller(1976) sec 8.4.

Theorem 4.2 (Existence of LAM estimators)**.** *Assume* $\{\bar{\theta}_n\} \subset \Theta$ *is discrete and* $\sqrt{n}$*-consistent for* $\theta_0 \in \Theta$*. Then* $\hat{\theta}_n$ *defined by (4.2) and (4.3) below is regular:*

$$\hat{\theta}_n = \bar{\theta}_n + \frac{1}{\sqrt{n}} \frac{\hat{\Gamma}_n(\bar{\theta}_n)^{-1}}{I(f)} \Delta_n(\bar{\theta}_n),$$

$$\hat{\Gamma}_n(\theta) = \frac{1}{n} \sum_{j=1}^{n} Z(j-1;\theta,\theta) Z^T(j-1;\theta,\theta).$$

Part 3. Construction of adaptive estimates

Theorem 4.3. *Let* $\{\bar{\theta}_n\} \subset \Theta$ *be a discrete and* $\sqrt{n}$*-consistent sequence of estimators of* θ_0*. Under our assumptions (A.1)-(A.6)*

$$\tilde{\Delta}_n(\bar{\theta}_n) - \Delta_n(\bar{\theta}_n) = o_{P_{\theta_0}}(1)$$

holds, if $c_n \to \infty$*,* $g_n \to \infty$*,* $\sigma(n) \to 0$*,* $d_n \to 0$*,* $\sigma(n)c_n \to 0$*,* $g_n \sigma(n)^{-4}/n \to 0$ *and* $n\sigma(n)$ *stays bounded.*

LONG RANGE DEPENDENCE

YAN LIU

1. Reference

Shibata (1980), AS.

2. notations

2.1. Notations.

1.	$\{x_t\}$	Gaussian stationary process
2.	r_l	$= E(x_t x_{t+l})$, autocovariance

3. Fundamental Setting

3.1. Basics.

(i) Model

$$x_t + \sum_{j=1}^{\infty} a_j x_{t-j} = e_t, \quad t = \dots, -1, 0, 1, \dots,$$

where a_1, a_2, $\dots$ are real numbers, $e_t \sim \mathcal{N}(0, \sigma^2)$.

(ii) the $k \times k$ covariance matrix

$$R(k) = (r_{ij}, 1 \leq i, j \leq k),$$

where $r_{ij} = r_{|i-j|}$.

(iii) the associated power series

$$A(z) = 1 + \sum_{j=1}^{\infty} a_j z^j.$$

3.2. Assumptions.

(A.1) $\sum_{1 \leq j < \infty} |a_j| < \infty$.
(A.2) $A(z)$ is nonzero for $|z| \leq 1$.
(A.3)

Date: November 12, 2014.

4. Main Results

Lemma 4.1. *For any* $1 \le k \le K_n$,

$$NE\|\sum_{K_n \le t \le n-1} X_t(k)(e_{t+1,k} - e_{t+1})/N\|^2 \le k\|a - a(k)\|^2\|R\|(\sum_{-\infty<j<\infty}|r_j| + \|R\|),$$

where $r_{-j} = r_j$ $(j = 1,\ 2,\ \ldots)$.

5. Further Reading

•

6. words

1.

7. New knowledge

•

(S)ON ESTIMATING THE INTENSITY OF LONG-RANGE DEPENDENCE IN FINITE AND INFINITE VARIANCE TIME SERIES

GEN RYU

1. Summary

The study includes both distributions with finite variance and infinite variance innovations. The model is also assumed not only with long-range dependence, the short dependence is also used in the paper.

When generating series with infinite variance, we will use independent symmetric α-stable variables as innovations in FARIMA(p, d, q) series, and skewed stable and Pareto distributions as the innovations in a FARIMA$(0, d, 0)$ series. The parameter d is restricted to interval $[\,0, 1 - 1/\alpha)$.

note. The relation between the index of the similarity and "d" in FARIMA is shown as

$$H = d + 1/\alpha.$$

note2. The parameter d plays the role of a differencing parameter in the FARIMA model.

note3. The $X^{(m)}$ defined below has the relation with S,

$$X^{(m)} \sim_d m^{H-1} S,$$

where S is a process which depends on the distribution of X but does not depend on m.

2. Methods

2.1. **Whittle Method.** The *Whittle estimator* gave the best performance for the series used in the study. If the parametric form of a time series is known, then the Whittle estimator is to be recommended. Even if the exact form is not known, but the maximum order (p, q) is known, this estimator can give good results.

$$Q(\eta) = \int_{-\pi}^{\pi} \frac{I(\eta)}{f(\nu;\eta)} d\nu + \int_{-\pi}^{\pi} \log f(\nu;\eta) d\nu.$$

note. It is d and not H which is estimated even in the infinite variance case.

note2. If the model is under specified, the Whittle estimator becomes more biased than any of the others used here.

Date: May 28, 2012.

2.2. **Local Whittle Method.** The second recommended method is Local Whittle estimator, which is proposed by Robinson(1995).

One estimates d by minimizing

$$R(d) = \log\left(\frac{1}{M}\sum_{j=1}^{M}\frac{I(\nu_j)}{\nu_j^{-2d}}\right) - 2d\frac{1}{M}\sum_{j=1}^{M}\log\nu_j.$$

note. There are as yet no corresponding theoretical results.

2.3. **Periodogram Method.** The periodogram is defined as

$$I(\nu) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N}X(j)e^{ij\nu}\right|^2,$$

where ν is the frequency, N is the length of the series, and X is the time series. $I(\nu)$ is an estimator of the spectral density of X, and a series with long-range dependence will have a spectral density proportional to $|\nu|^{-2d}$ close to the origin. A log-log regression thus provides an estimate of d. In the infinite variance case the problem is significantly more complicated.

note. There are no theoretical results for the periodogram regression method. The proportionality to $|\nu|^{-2d}$ as $\nu \to 0$, however, seems to hold empirically in the infinite variance case as well.

2.4. **The estimator for H in robust order.**

2.4.1. *Variance of Residuals Method(VR).* The Variance of Residuals method was introduced by Peng et al.(1994). First the series is divided into blocks of size m. Then, within each block, the partial sums of the series are calculated,

$$Y(t) = \sum_{i=1}^{t}X_i.$$

A least-squares line, $a + bt$, is fitted to the partial sums within each block, and the sample variance of the residuals is computed,

$$\frac{1}{m}\sum_{t=1}^{m}(Y(t) - a - bt)^2.$$

As the calculation in the paper, the slope of the log-log plot shows $2H$. In practice, this is not recommended because the scatter is too large for the infinite variance series.

2.4.2. *Absolute Value Method.* Let $X^{(m)}(k)$ be

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad k = 1, 2, \ldots, [N/m].$$

Then we take the first absolute moment of this series,

$$AM^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k) - \bar{X}|,$$

where $\bar{X}$ is the overall series mean. The same thing can be concluded like above, that is, the slope of log-log lime is the function of H, $H-1$.

note. The method loses efficiency, especially for values for α close to 1.

3. Interesting facts

3.1. **About S in the first section.** In the finite variance case S is Fractional Gaussian Noise (FGN) and in the infinite variance case it is Linear Fractional Stable Noise (LFSN).

The FGN series $\{X_i, i \geq 1\}$ is a zero mean, stationary, Gaussian time series whose autocovariance function at lag h is:

$$\gamma(h) = \frac{1}{2}\{(h+1)^{2d+1} - 2h^{2d+1} + |h-1|^{2d+1}\}, h \geq 0, -1/2 < d < 1/2.$$

For $d \neq 0$, the autocovariance satisfies

$$\gamma(h) \sim d(2d+1)h^{2d-1} \quad \text{as } h \to \infty.$$

In addition, the spectral density is given by:

$$f(\nu) = C_d(2\sin\frac{\nu}{2})^2 \sum_{k=-\infty}^{\infty} \frac{1}{|\nu + 2\pi k|^{2d+2}} \sim C_d|\nu|^{-2d} \quad \text{as } \nu \to 0.$$

3.2. **About stable distribution.** If either Pareto or stable with parameter α, then $P(\epsilon > x) \sim Cx^{-\alpha}$ as $x \to \infty$, that is, the probability tails decrease slowly, like a power function. Moreover, $\mathrm{Var}(\epsilon) = \infty$ if $\alpha < 2$, and $E|\epsilon| = \infty$ if $0 < \alpha \leq 1$.

3.3. **About long dependence.** The FARIMA(p, d, q) family of models is widely used in the modeling of time series with long-range dependence. These are moving averages

$$X_n = \sum_{i=-\infty}^{n} c_{n-i}\epsilon_i,$$

where c_k behaves like k^{d-1} for large k and the ϵ_i's are independent, identically distributed random variables.

SUMMARY-HALLIN1987

GEN RYU

1. THE MAIN POINTS OF THE PAPER

(1) Quadratic serial rank statistics are introduced to be some optimal tests which is sensitive against a whole subclass of the alternative in the case of unspecified alternative.
(2) The asymptotically maximin most powerful quadratic serial rank tests (rank portmanteau tests) are obtained after deriving their asymptotic distribution tunder the null hypothesis and contiguous ARMA alternatives.
(3) The asymptotic relative efficiencies of the rank portmanteau tests are derived.

2. NOTATIONS

2.1. $H_{\boldsymbol{d},f}$. the contiguous ARMA alternatives $H_{\boldsymbol{d},f}$ are completely specified by a vector $\boldsymbol{d} = (d_1, \ldots, d_p)'$ of real coefficients and a density type f.

2.2. f. an unspecified member of the family of the densities is denoted by $f(x)$, while the specified is denoted by $f_\sigma(x)$.

3. DEFINITIONS

3.1. **strongly unimodal density.** The density $f(x)$ is called strongly unimodal if

(1) $-\log f(x)$ is a convex function within some open interval (a, b) such that $-\infty \leq a < b \leq \infty$;
(2) $\int_a^b f(x)\, dx = 1$.

note. Such densities are absolutely continuous within (a, b) and

$$[\,-\log f(x)\,]' = -\frac{f'(x)}{f(x)}$$

is a non-decreasing function.

note2. Normal, Double exponential, Exponential, Logistic, Uniform, Triangular and etc are strongly unimodal density.

Date: February 4, 2013.

3.2. **asymptotically maximin most powerful tests.** a sequence ϕ^n of tests is said to be asymptotically maximin most powerful for the sequence of null hypotheses H^n against the sequence of alternatives K^n if its power asymptotically reached the envelope power function $\beta(\alpha, H^n, K^n)$, i.e. if

$$\limsup_{n\to\infty}[\, E_{l^{(n)}}\phi^n - \alpha\,] \le 0, \quad \text{for all } l^{(n)} \in H^{(n)}.$$

and

$$\liminf_{n\to\infty}[\, E_{l^{(n)}}\phi^{(n)} - \beta(\alpha, H^{(n)}, K^{(n)})\,] \ge 0 \quad \text{for all } l^{(n)} \in K^{(n)}.$$

4. ASSUMPTIONS AND PROPOSITIONS

4.1. **assumptions(1985).**
Let $\{\epsilon_t; t \in \mathbb{Z}\}$ be a discrete-time stationary white noise. Assume that it has a density $f(x)$, and that the following conditions are satisfied:

(1) ϵ_t has finite moments up to the third order; denote its variance by σ^2.
(2) $f(x)$ is a.e. derivative, and its derivative $f'(x)$ satisfies

$$\int_{-\infty}^{\infty} |f'(x)| dx < \infty$$

(3) $f(x)$ has finite Fisher's information $I(f)$.
(4) Assume $\phi(x)$ is a.e. derivative, and its derivative $\phi'(x)$ satisfies a Lipschitz condition

$$|\phi'(x) - \phi'(y)| < A|x-y|, \quad a.e.$$

4.2. **assumptions(1987).** Assume (2) and (3) in the assumption(1985) hold.

(1) a density type f means the family of densities $\{f_\sigma(x) = \frac{1}{\sigma} f_1(\frac{x}{\sigma}); \sigma > 0\}$ indexed by scale parameter.
(2) The assumptions below is the same to the definition of white noise.

$$\int x f(x)\, dx = 0,$$
$$\int x^2 f(x)\, dx = \sigma^2.$$

note. The existence of the third moment is not assumed.

(3) Letting

$$\phi(x) = \frac{f_1'(x)}{f_1(x)} \quad \text{a.e.}$$

note. From this assumption, it is seen that

$$I(f) = \sigma^{-2} \int \phi^2(x) f_1(x)\, dx,$$

that is, $\sigma^2 I(f) = I(f_1)$.

(4) Assume $d_p \neq 0$.

4.3. **quadratic serial rank statistics.**

Proposition 4.1. *Denote by* $J_\mu(v_{p_\mu+1},\dots,v_1)$ *the score-generating functions associated with* S_μ $(\mu = 1,\dots,q)$*, and assume that they satisfy*

$$\int_{[0,1]^{p+1}} J(v_{p+1},\dots,v_1)\prod_{j\neq i} dv_j = 0, \quad (i = 1,\dots,p+1).$$

Then, under H_0*,*

$$n^{1/2}(\boldsymbol{S}-\boldsymbol{m}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\boldsymbol{V^2})$$

where $\boldsymbol{V}^2 = (V_{\mu\nu})$ *i.e. (writing* $J_\mu(v_{p+1},\dots,v_1)$ *for* $J_\mu(v_{p\,u+1},\dots,v_1)$*),*

$$V_{\mu\nu} = \int_{[0,1]^{p+1}} J_\mu(v_{p+1},\dots,v_1)J_\nu(v_{p+1},\dots,v_1)\,dv_{p+1}\dots dv_1 + \sum_{k=1}^{p}\int_{[0,1]p+1+k}\{J_\mu(v_{p+1},\dots,v_1)J_\nu(v_{p+1+k},\dots,v_{k+1}) + J_\mu(v_{p+1+k},\dots,v_{1+k})J_\nu(v_{p+1},\dots,v_1)\,dv_{p+1+k}\dots dv_1 \quad (\mu,\nu = 1,\dots,q). \tag{4.1}$$

Under $H_{\boldsymbol{d},f}$*,*

$$n^{1/2}(\boldsymbol{S}-\boldsymbol{m}) \xrightarrow{d} (\boldsymbol{C}'\boldsymbol{d},\boldsymbol{V}^2)$$

where $\boldsymbol{C}$ *is a* $p\times q$ *matrix with entries*

$$C_{i\mu} = \sum_{j=0}^{p-i}\int_{[0,1]^{p+1}} J_\mu(v_{p+1},\dots,v_1)\phi(F_1^{-1}(v_{p+1-j}))F_1^{-1}(v_{p+1-j-i})\,dv_{p+1}\dots dv_1, \quad (i = 1,\dots,p;\mu = 1,\dots,q). \tag{4.2}$$

Proposition 4.2. *Using the same notation as in Prop 2.1, and assuming that* $\boldsymbol{V}^2$ *is of rank q, the rank statistic*

$$\boldsymbol{Q} = n(\boldsymbol{S}-\boldsymbol{m})'\boldsymbol{V}^{-2}(\boldsymbol{S}-\boldsymbol{m}).$$

is, under H_0*,*

$$\boldsymbol{Q} \xrightarrow{d} \chi^2(q),$$

and under $H_{\boldsymbol{d},f}$

$$\boldsymbol{Q} \xrightarrow{d} non\chi^2 q$$

and with non-centrality parameter

$$\lambda_f(\boldsymbol{d}) = \frac{1}{2}\boldsymbol{d}'\boldsymbol{C}\boldsymbol{V^{-2}}\boldsymbol{C}'\boldsymbol{d}.$$

4.4. **testing H_0 against $H_{\boldsymbol{d},f}$ and asymptotically most powerful test against $H_{\boldsymbol{d},f}$.** Consider the following linear serial rank statistic of order i:

$$r_{(i)f} = \left\{ (n-1)^{-1} \sum_{t=i+1}^{n} \phi\left(F_1^{-1}\left(\frac{R_t}{n+1}\right)\right) F_1^{-1}\left(\frac{R_{t-i}}{n+1}\right) - m \right\} / s$$

where

$$m = \frac{1}{n(n-1)} \sum_{1 \leq i_1 \neq i_2 \leq n} \phi_1\left(F_1^{-1}\left(\frac{i_1}{n+1}\right)\right) F_1^{-1}\left(\frac{i_2}{n+1}\right)$$

and

$$(s)^2 = \{n(n-1)\}^{-1} \sum_{1 \leq i_1 \neq i_2 \leq n} \left\{ \phi\left(F_1^{-1}\left(\frac{i_1}{n+1}\right)\right) F_1^{-1}\left(\frac{i_2}{n+1}\right) \right\}^2$$
$$+ \frac{1}{3}(n-2i)(n-i)^{-1}$$
$$\binom{n}{3}^{-1} \sum_{1 \leq i_1 \neq i_2 \neq i_3 \leq n} \left\{ \phi\left(F_1^{-1}\left(\frac{i_1}{n+1}\right)\right) \phi\left(F_1^{-1}\left(\frac{i_2}{n+1}\right)\right) \left(F_1^{-1}\frac{i_2}{n+1}\right) F_1^{-1}\left(\frac{i_3}{n+1}\right) \right\}$$
$$+ \frac{1}{24}\{n^2 - n(2i+3) + i^2 + 5i\}(n-i)^{-1}$$
$$\binom{n}{4}^{-1} \sum_{1 \leq i_1 \neq i_2 \neq i_3 \neq i_4 \leq n} \left\{ \phi\left(F_1^{-1}\left(\frac{i_1}{n+1}\right)\right) \phi\left(F_1^{-1}\left(\frac{i_2}{n+1}\right)\right) \left(F_1^{-1}\frac{i_3}{n+1}\right) F_1^{-1}\left(\frac{i_4}{n+1}\right) \right\}$$
$$- (n-i)(m)^2.$$

Assumption 4.3. *Assume f to be a strongly unimodal density.*

Proposition 4.4.

(1) *$r_{(i)f}$ admits the score-generating function*

$$J_{(i)f}(v_{i+1}, \ldots, v_1) = \phi(F_1^{-1}(v_{i+1}))F_1^{-1}(v_1)\{\sigma^2 I(f)\}^{-1/2}.$$

(2) *Denote by r_f the vector of f-rank autocorrelations of orders 1 through q (q arbitrary-assume $q > p$). Then, under H_0*

$$n^{1/2} r_f \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_{q \times q});$$

under $H_{\boldsymbol{d},f}$

$$n^{1/2} r_f \xrightarrow{d} \mathcal{N}\left(\{\sigma^2 I(f)\}^{1/2} \begin{pmatrix} d_1 \\ \vdots \\ d_p \\ \mathbf{0} \end{pmatrix}, I_{q \times q} \right).$$

Proposition 4.5. *Consider the linear serial rank statistic*

$$S^*_{\boldsymbol{d},f} = \sum_{i=1}^{p} \frac{d_i}{||\boldsymbol{d}||} r_{(i)f} \qquad \left(\text{with } ||\boldsymbol{d}|| = \left(\sum_{i=1}^{p} d_i^2 \right)^{1/2} \right).$$

Then

(1) *under* H_0,

$$n^{1/2} S^*_{\boldsymbol{d},f} \xrightarrow{d} \mathcal{N}(0,1);$$

under $H_{\boldsymbol{h},f}$, $\boldsymbol{h} = (h_1, \dots, h_p) \in \mathbb{R}^p$

$$n^{1/2} S^*_{\boldsymbol{d},f} \xrightarrow{d} \mathcal{N}\left(\{\sigma^2 I(f)\}^{1/2} \sum_{i=1}^{p} \frac{h_i d_i}{||\boldsymbol{d}||}, 1 \right).$$

note. the mean is a form of inner product between $\boldsymbol{h}$ *and the mean in prop 4.4.*

(2) *The test based on* $S^*_{\boldsymbol{d},f}$ *(at level* α*) is asymptotically most powerful, within the class of all tests of level* α*, for testing* H_0 *against* $H_{\boldsymbol{d},f}$*.*

4.5. Testing H_0 against $H_{\boldsymbol{d},f}$ ($\boldsymbol{d}$ unspecified, f specified).

Assumption 4.6. *Assume* f *and* g *to be a strongly unimodal density.*

Proposition 4.7. *The quadratic serial rank statistic*

$$Q^*_f = \sum_{i=1}^{p} (n-i)(r_{(i)f})^2 = n \sum_{i=1}^{p} (r_{(i)f})^2 + o_p(1)$$

is

(1) *under* H_0

$$Q^*_f \xrightarrow{d} \chi^2(p);$$

under $H_{\boldsymbol{d},g}$,

$$Q^*_f \xrightarrow{d} non\chi^2(p);$$

with non-centrality parameter

$$\text{(a)} \qquad \lambda^*_{f,g}(\boldsymbol{d}) = \frac{1}{2} ||\boldsymbol{d}||^2 \left\{ \int \phi(F_1^{-1}(u)) \phi_g(G^{-1}(u))\, du \int F_1^{-1}(v) G^{-1}(v)\, dv \right\}^2 \{\sigma^2 I(f)\}^{-1}$$

Under $H_{\boldsymbol{d},f}$*, (a) reaches, for given* $||\boldsymbol{d}|| = d$*, its maximal value* $\lambda^*_f(\boldsymbol{d}) = \frac{1}{2} ||\boldsymbol{d}||^2 \sigma^2 I(f)$.

(2) Q^*_f *provides, at any level* $\alpha \in (0,1)$ *and for any value of* $d > 0$*, an asymptotically maximin most powerful test for* H_0 *against* $K(d)$*.*

Proposition 4.8. *The ARE of an* $f-$ *rank portmanteau statistic* $Q_g^{(n)*}$ *with respect to another one,* $Q_h^{(n)*}$*, when testing against* $H_{\boldsymbol{d},f}$*, is given by*

$$e_f(Q_g^{(n)*}, Q_h^{(n)*}) = \frac{\sigma^2 I(h)}{\sigma^2 I(g)} \left(\frac{\int \phi_g(G^{-1}(u)) \phi_f(F^{-1}(u))\, du \int G^{-1}(v) F^{-1}(v)\, dv}{\int \phi_h(H^{-1}(u)) \phi_f(F^{-1}(u))\, du \int H^{-1}(v) F^{-1}(v)\, dv} \right)^2$$

5. WORDS

(1) insensitive 感受性の鈍い
(2) vast 膨大な
(3) diffuse 広がった
(4) intuitively appealing 直感に訴えるような
(5) remedy 手段
(6) allied 同類の
(7) overall 全体にわたる
(8) revisited 再度確認される
(9) in the sequel 後ろで
(10) This provides the incentive for our considering quadratic statistics. これによって我々の 2 次統計量に対する目的意識を感じた。
(11) in the subsequent sections 次の節で
(12) inferential 推定の
(13) eventual 究極的な
(14) superiority 優越
(15) allow for 考慮する
(16) contradictions 矛盾

(S)ON ESTIMATING THE INTENSITY OF LONG-RANGE DEPENDENCE IN FINITE AND INFINITE VARIANCE TIME SERIES

GEN RYU

1. Summary

The study includes both distributions with finite variance and infinite variance innovations. The model is also assumed not only with long-range dependence, the short dependence is also used in the paper.

When generating series with infinite variance, we will use independent symmetric α-stable variables as innovations in FARIMA(p, d, q) series, and skewed stable and Pareto distributions as the innovations in a FARIMA$(0, d, 0)$ series. The parameter d is restricted to interval $[0, 1 - 1/\alpha)$.

note. The relation between the index of the similarity and "d" in FARIMA is shown as

$$H = d + 1/\alpha.$$

note2. The parameter d plays the role of a differencing parameter in the FARIMA model.

note3. The $X^{(m)}$ defined below has the relation with S,

$$X^{(m)} \sim_d m^{H-1} S,$$

where S is a process which depends on the distribution of X but does not depend on m.

2. Methods

2.1. **Whittle Method.** The *Whittle estimator* gave the best performance for the series used in the study. If the parametric form of a time series is known, then the Whittle estimator is to be recommended. Even if the exact form is not known, but the maximum order (p, q) is known, this estimator can give good results.

$$Q(\eta) = \int_{-\pi}^{\pi} \frac{I(\eta)}{f(\nu;\eta)} d\nu + \int_{-\pi}^{\pi} \log f(\nu;\eta) d\nu.$$

note. It is d and not H which is estimated even in the infinite variance case.

note2. If the model is under specified, the Whittle estimator becomes more biased than any of the others used here.

Date: May 28, 2012.

2.2. **Local Whittle Method.** The second recommended method is Local Whittle estimator, which is proposed by Robinson(1995).

One estimates d by minimizing

$$R(d) = \log\left(\frac{1}{M}\sum_{j=1}^{M}\frac{I(\nu_j)}{\nu_j^{-2d}}\right) - 2d\frac{1}{M}\sum_{j=1}^{M}\log\nu_j.$$

note. There are as yet no corresponding theoretical results.

2.3. **Periodogram Method.** The periodogram is defined as

$$I(\nu) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N}X(j)e^{ij\nu}\right|^2,$$

where ν is the frequency, N is the length of the series, and X is the time series. $I(\nu)$ is an estimator of the spectral density of X, and a series with long-range dependence will have a spectral density proportional to $|\nu|^{-2d}$ close to the origin. A log-log regression thus provides an estimate of d. In the infinite variance case the problem is significantly more complicated.

note. There are no theoretical results for the periodogram regression method. The proportionality to $|\nu|^{-2d}$ as $\nu \to 0$, however, seems to hold empirically in the infinite variance case as well.

2.4. **The estimator for H in robust order.**

2.4.1. *Variance of Residuals Method(VR).* The Variance of Residuals method was introduced by Peng et al.(1994). First the series is divided into blocks of size m. Then, within each block, the partial sums of the series are calculated,

$$Y(t) = \sum_{i=1}^{t}X_i.$$

A least-squares line, $a + bt$, is fitted to the partial sums within each block, and the sample variance of the residuals is computed,

$$\frac{1}{m}\sum_{t=1}^{m}(Y(t) - a - bt)^2.$$

As the calculation in the paper, the slope of the log-log plot shows $2H$. In practice, this is not recommended because the scatter is too large for the infinite variance series.

2.4.2. *Absolute Value Method.* Let $X^{(m)}(k)$ be

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad k = 1, 2, \ldots, [\,N/m\,].$$

Then we take the first absolute moment of this series,

$$AM^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k) - \bar{X}|,$$

where $\bar{X}$ is the overall series mean. The same thing can be concluded like above, that is, the slope of log-log lime is the function of H, $H-1$.

note. The method loses efficiency, especially for values for α close to 1.

3. Interesting facts

3.1. **About S in the first section.** In the finite variance case S is Fractional Gaussian Noise (FGN) and in the infinite variance case it is Linear Fractional Stable Noise (LFSN).

The FGN series $\{X_i, i \geq 1\}$ is a zero mean, stationary, Gaussian time series whose autocovariance function at lag h is:

$$\gamma(h) = \frac{1}{2}\{(h+1)^{2d+1} - 2h^{2d+1} + |h-1|^{2d+1}\}, h \geq 0, -1/2 < d < 1/2.$$

For $d \neq 0$, the autocovariance satisfies

$$\gamma(h) \sim d(2d+1)h^{2d-1} \quad \text{as } h \to \infty.$$

In addition, the spectral density is given by:

$$f(\nu) = C_d(2\sin\frac{\nu}{2})^2 \sum_{k=-\infty}^{\infty} \frac{1}{|\nu + 2\pi k|^{2d+2}} \sim C_d|\nu|^{-2d} \quad \text{as } \nu \to 0.$$

3.2. **About stable distribution.** If either Pareto or stable with parameter α, then $P(\epsilon > x) \sim Cx^{-\alpha}$ as $x \to \infty$, that is, the probability tails decrease slowly, like a power function. Moreover, $\mathrm{Var}(\epsilon) = \infty$ if $\alpha < 2$, and $E|\epsilon| = \infty$ if $0 < \alpha \leq 1$.

3.3. **About long dependence.** The FARIMA(p, d, q) family of models is widely used in the modeling of time series with long-range dependence. These are moving averages

$$X_n = \sum_{i=-\infty}^{n} c_{n-i}\epsilon_i,$$

where c_k behaves like k^{d-1} for large k and the ϵ_i's are independent, identically distributed random variables.

SUMMARY-TIME SERIES(APPLICATIONS TO FINANCE)

GEN RYU

1. ARIMA and SARIMA

1.1. **ARIMA models (autoregressive integrated moving average model).** ARIMA model is a generalization of ARMA model. For ARMA(p,q) model W_t, the ARIMA(p,d,q) model Y_t is defined as

$$(1-B)^d Y_t = W_t.$$

If we write ARMA(p,q) in an explicit way,

$$\phi(B)W_t = \theta(B)Z_t,$$

then the ARIMA(p,d,q) can be shown as

$$\phi(B)(1-B)^d Y_t = \theta(B)Z_t.$$

note. The definition varies from books to books. So you must be careful to put the definition in mind.

Examples:

(1) $Y_t = \alpha + \beta t + N_t$ can be written as ARIMA(0,1,1).
(2) ARIMA(0,1,0) is a Random Walk model.
(3) (conti of (2)) the price of a stock at the end of day t can be written as ARIMA(0,1,0).

1.2. **Another definition for ARIMA model.**

Definition 1.1 (The ARIMA(p,d,q) Process). *If d is a non-negative integer, then $\{X_t\}$ is said to be an ARIMA(p,d,q) process if $Y_t := (1-B)^d X_t$ is a causal ARMA(p,q) process.*

note. This definition is from "Time seris: Theory and Methods".

1.3. **the extension of the ARIMA model-SARIMA model.**

Definition 1.2 (The SARIMA(p,d,q) $\times$ $(P,D,Q)_s$ Process). *If d and D are non-negative integers, then $\{X_t\}$ is said to be a seasonal ARIMA(p,d,q) $\times$ $(P,D,Q)_s$ process with period s if the diffenenced process $Y_t := (1-B)^d(1-B^s)^D X_t$ is a causal ARMA process,*

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \quad \{Z_t\} \sim WN(0,\sigma^2)$$

where $\phi(z) = 1-\phi_1 z - \cdots - \phi_p z^p$, $\Phi(z) = 1-\Phi_1 z - \cdots - \Phi_P z^P$, $\theta(z) = 1+\theta_1 z + \cdots + \theta_q z^q$, $\Theta(z) = 1+\Theta_1 z - \cdots - \Theta_Q z^Q$.

Date: January 29, 2012.

2. INVERTIBLE AND NONCAUSAL

2.1. invertible.

Theorem 2.1. *An MA(q) model* $\{Y_t\}$ *is invertible if the roots of the equation* $\theta(B) = 0$ *all lie outside the unit circle.*

2.2. causal.

Definition 2.2. *A process* $\{Y_t\}$ *is said to be causal if there exists a sequence of constants* $\{\psi_j\}$*'s such that* $Y_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ *with* $\sum_{j=0}^{\infty} |\psi| < \infty$*.*

Theorem 2.3. *An AR(p) process is causal if the roots of the characteristic polynomial* $\phi(z) = 1 - \phi z - \cdots - \phi_p z^p$ *all lie outside the unit circle.*

3. ACF AND PACF

3.1. ACF.

Definition 3.1. *Let* $\{X_t\}$ *be a stationary process. Then*

(1) $\gamma(\tau) = \mathrm{Cov}(X_t, X_{t+\tau})$ *is called the autocovariance function.*
(2) $\rho(\tau) = \gamma(\tau)/\gamma(0)$ *is called the autocorrelation function.*

3.2. PACF.

Definition 3.2. *The PACF of a stationary time series is defined as*

$$\phi_{11} = \rho(1), \tag{3.1}$$

$$\phi_{kk} = \mathrm{corr}(Y_{k+1} - P_{\bar{sp}\{Y_2,\cdots,Y_k\}} Y_{k+1}, Y_1 - P_{\bar{sp}\{Y_2,\cdots,Y_k\}} Y_1) \tag{3.2}$$

where $P_{\bar{sp}\{Y_2,\cdots,Y_k\}} Y$ *denotes the projection of the random variable* Y *onto the closed linear subspace spanned by the random variables* $\{Y_2, \ldots, Y_k\}$*.*

Examples:

(1) $\phi_{11} = \mathrm{corr}(Y_{k+1} - P_{\bar{sp}\{Y_2,\cdots,Y_k\}} Y_{k+1}, Y_1 - P_{\bar{sp}\{Y_2,\cdots,Y_k\}} Y_1)$
(2) ARIMA(0,1,0) is a Random Walk model.
(3) (conti of (2)) the price of a stock at the end of day t can be written as ARIMA(0,1,0).

SUMMARY–DISCRIMINANT ANALYSIS FOR DYNAMICS OF STABLE PROCESSES

GEN RYU

1. Definition

On the linear process

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}, \tag{1.1}$$

where the innovation $\{Z_t\}_{t\in\mathbb{Z}}$ is a seq of iid symmetric α-stable r.v. for $\alpha \in (0, 2)$, many key words are defined as follows.

1.1. **characteristic exponent.**

α is called characteristic exponent.

1.2. **the power transfer function of the linear filter.**

The power transfer function of the linear filter is defined as

$$|\psi(\lambda)|^2 = |\sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}|^2, \quad \lambda \in [-\pi, \pi].$$

1.3. **the normalized power transfer function of $\{X_t\}$.**

The normalized power transfer function of $\{X_t\}$ is written as

$$\tilde{f}(\lambda) \equiv \frac{|\psi(\lambda)|^2}{\psi^2} = \frac{|\sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}|^2}{\sum_{j=-\infty}^{\infty} \psi_j^2} = \sum_{k=-\infty}^{\infty} \rho(k) e^{-ik\lambda}, \quad \lambda \in [-\pi, \pi],$$

where

$$\rho(k) = \frac{\sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k}}{\sum_{j=-\infty}^{\infty} \psi_j^2}, \quad k \in \mathbb{Z}.$$

Date: February 19, 2012.

note. Since

$$
\begin{aligned}
|\sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}|^2 &= (\sum_{j=-\infty}^{\infty} \psi_j e^{ij\lambda})(\sum_{l=-\infty}^{\infty} \psi_l e^{-il\lambda}) \\
&= \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \psi_j \psi_l e^{-ij\lambda} e^{-il\lambda} \\
&= \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \psi_j \psi_l e^{-i(l-j)\lambda} \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_{j+k} e^{-ik\lambda}, \quad \text{where } k = l - j.
\end{aligned}
$$

1.4. **smoothed self-normalized periodogram.** It is called the smoothed self-normalized periodogram if it takes the form

$$\sum_{|k| \le m} W_n \tilde{I}_{n,X}(\lambda_k).$$

2. THE MAIN POINT OF THIS PAPER

2.1. **consistency.** The misclassification probabilities converge to 0 as the sample size tends to infinity.

2.2. **the goodness of fit.** The evaluation of $I_n(\tilde{f}, \tilde{g})$ is done in terms of misclassification probabilities when one density is contiguous to another.

3. ASSUMPTIONS AND THEOREMS

The hypotheses is defined as

$$\pi_1 : \tilde{f}(\lambda) \qquad \pi_2 : \tilde{g}(\lambda),$$

and the classification statistic is defined as

$$I_n(\tilde{f}, \tilde{g}) = \int_{-\pi}^{\pi} \left\{ \log\left(\frac{\tilde{g}(\lambda)}{\tilde{f}(\lambda)}\right) + \left(\frac{1}{\tilde{g}(\lambda)} - \frac{1}{\tilde{f}(\lambda)}\right) \tilde{I}_{n,X}(\lambda)\} d\lambda \right\}.$$

3.1. **consistency.**

Assumption 3.1. *The linear filter* $\{\psi_j\}_{j\in\mathbb{Z}}$ *satisfies*

$$\sum_{j=-\infty}^{\infty} |j||\psi_j|^{\delta} < \infty$$

for some $\delta < \min(1, \alpha)$.

note1. Under this assumption, we also have

$$\sum_{k=-\infty}^{\infty} |k||\rho(k)|^{\delta} < \infty.$$

note2. This assumption also implies the normalized power transfer function can be defined as (1.3) by using Hölder's inequality.

Assumption 3.2. *$\tilde{f}(\lambda)$ and $\tilde{g}(\lambda)$ are positive on $[\pi, \pi]$, and $\tilde{f}(\lambda) \not\equiv \tilde{g}(\lambda)$ on a set of positive Lebesgue measure.*

Theorem 3.3. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be defined by (1.1) and suppose that Asp1-2 hold. Then, for the hypothesis, the misclassification probabilities tend to 0 as $n \to \infty$. That is,*

$$P(2|1) \equiv P(I_n(\tilde{f}, \tilde{g}) \le 0|\pi_1) \to 0, \quad P(1|2) \equiv P(I_n(\tilde{f}, \tilde{g}) > 0|\pi_2) \to 0.$$

3.2. **the goodness of fit.** Now we consider the goodness of fit in the case that $\tilde{g}(\lambda)$ is contiguous to $\tilde{f}(\lambda)$. That is

$$\pi_1 : \tilde{f}(\lambda|\boldsymbol{\theta}) \qquad \pi_2 : \tilde{g}(\lambda) = \tilde{f}(\lambda|\boldsymbol{\theta} + a_n\boldsymbol{h}),$$

where

$$a_n = \left(\frac{\log n}{n}\right)^{1/\alpha}.$$

Assumption 3.4. *$\tilde{f}(\lambda|\theta)$ is continuously three times differentiable with respect to $\boldsymbol{\theta} \in \Theta$, and*

$$\sum_{j,k,l=1}^{q} |\frac{\partial^3}{\partial\theta_j\partial_k\partial_l}\tilde{f}(\lambda|\boldsymbol{\theta})| < \infty.$$

Assumption 3.5. *For some $\delta < \min(1, \alpha)$, and for $k, l = 1, \ldots, q$.*

$$\sum_{t=1}^{\infty} |\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_k}\tilde{f}(\lambda|\boldsymbol{\theta})\, \tilde{f}^{-1}(\lambda|\boldsymbol{\theta}) \cos(t\lambda)\, d\lambda|^{\delta} < \infty,$$

$$\sum_{t=1}^{\infty} |\int_{-\pi}^{\pi} \frac{\partial^2}{\partial\theta_k\partial\theta_l}\tilde{f}(\lambda|\boldsymbol{\theta})\, \tilde{f}^{-1}(\lambda|\boldsymbol{\theta}) \cos(t\lambda)\, d\lambda|^{\delta} < \infty,$$

Theorem 3.6. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be defined by (1.1) and suppose that Asp1-4 hold. Then under the contiguous condition,*

$$\lim_{n\to\infty} P(2|1) = P(\frac{Z_1}{Y_0} \ge \frac{\boldsymbol{h}^T\mathcal{F}(\boldsymbol{\theta})\boldsymbol{h}}{4(C_\alpha \sum_{t=1}^{\infty} |\mathcal{E}_t(\boldsymbol{\theta})^T\boldsymbol{h}|^{\alpha})^{\frac{1}{\alpha}}}),$$

where

$$\begin{aligned}
\mathcal{F}(\boldsymbol{\theta}) &= \left[\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_k}\tilde{f}(\lambda|\boldsymbol{\theta})\,\frac{\partial}{\partial\theta_l}\tilde{f}(\lambda|\boldsymbol{\theta})\tilde{f}^{-2}(\lambda|\boldsymbol{\theta})\,d\lambda\right]_{k,l} \quad (q\times q)-\textit{matrix};\\
\mathcal{E}_t(\boldsymbol{\theta}) &= \left[\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_k}\tilde{f}(\lambda|\boldsymbol{\theta})\,\tilde{f}^{-1}(\lambda|\boldsymbol{\theta})(\cos(t\lambda)-\rho(t))\,d\lambda\right]_{k,1} \quad (q\times 1)-\textit{vector};\\
C_\alpha &= \begin{cases} \dfrac{(1-\alpha)\sigma}{2\Gamma(2-\alpha)\cos\pi\alpha/2}, & \textit{if } \alpha\neq 1,\\ \sigma/\pi, & \textit{if } \alpha = 1.\end{cases}
\end{aligned}$$

and Y_0 *is an* $\frac{\alpha}{2}$*-stable positive r.v. which is independent of* $\{Z_t\}_{t\in\mathbb{Z}}$.

3.3. **the estimation of** $\tilde{f}(\lambda)$**.** We can estimate the npt function by the smoothed self-normalized periodogram

$$\sum_{|k|\leq m} W_n(k)\tilde{I}_{n,X}(\lambda_k)$$

where

$$\lambda_k = \lambda + \frac{k}{n}, \quad |k|\leq m,$$

$m = m_n$ is a seq in $\mathbb{N}$ such that

$$m_n \to \infty \quad \text{and} \quad \frac{m_n}{n}\to 0,\ n\to\infty.$$

$\{W_n\}_{n\in\mathbb{N}}$ is a seq of weight functions satisfying the conditions:

$$\begin{aligned}
\sum_{|k|\leq m} W_n(k) &= 1;\\
\sum_{|k|\leq m} W_n(k)^2 &= o(1), \quad n\to\infty;\\
W_n(k) = W_n(-k) \quad &, \quad W_n(k)\geq 0.
\end{aligned}$$

SUMMARY=LOCAL ASYMPTOTIC NORMALITY FOR REGRESSION MODELS WITH LONG-MEMORY DISTURBANCE

GEN RYU

1. THE MAIN POINT OF THE PAPER

1.1. **The local asymptotic normality property.** In the paper, the local asymptotic normality property is established for a regression model with fractional ARIMA(p,d,q) errors.

1.2. **applications for inference problems in the long-memory context.**

(1) testing linear constraints on the parameters;
(2) the discriminant analysis problem;
(3) the construction of locally asymptotically minimax adaptive estimators.

1.3. **other applications in the long-memory context.**

(1) hypothesis testing;
(2) discriminant analysis;
(3) rank-based testing;
(4) locally asymptotically minimax;
(5) adaptive estimation.

2. QUESTION

(1) What is Durbin-Watson test?

Date: January 31, 2012.

SUMMARY=LIMIT THEORY FOR THE SAMPLE COVARIANCE AND CORRELATION FUNCTIONS OF MOVING AVERAGES

GEN RYU

1. DEFINITION

We consider the discrete time moving average process

$$X_t = \sum_{j=-\infty}^{\infty} c_j Z_{t-j}, \quad t \in \mathbb{Z}, \tag{1.1}$$

with

$$\sum_{j=-\infty}^{\infty} |c_j| < \infty.$$

in this summary.

1.1. regularly varying tail probabilities.

$\{Z_t, -\infty < t < \infty\}$ is an independent and identically distributed(iid) sequence of random variables with regularly varying tail probabilities, that is,

$$P(|Z_k| > x) = x^{-\alpha} L(x)$$

with $\alpha > 0$ and $L(x)$ a slowly varying function at ∞ and,

$$\frac{P(Z_k > x)}{P(|Z_k| > x)} \to p \quad \text{and} \quad \frac{P(Z_k < -x)}{P(|Z_k| > x)} \to q$$

as $x \to \infty$, $0 \leq p \leq 1$ and $q = 1 - p$.

1.2. the sample correlation function.

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{n-h} X_t X_{t+h}}{\sum_{t=1}^{n} X_t^2}, \quad h > 0.$$

1.3. the correlation function.

$$\rho(h) = \frac{\sum_{j=-\infty}^{\infty} c_j c_{j+h}}{\sum_{j=-\infty}^{\infty} c_j^2}.$$

1.4. the sample covariance function.

$$\hat{\gamma(h)} = \frac{1}{n} \sum_{t=1}^{n} X_t X_{t+h}, \quad h > 0.$$

Date: January 31, 2012.

2. The main points of this paper

2.1. **the limit distribution of the sample covariance function.** The limit distribution is derived in the case that the process has a finite variance but an infinite fourth moment.

2.2. **convergence($0 < \alpha < 2$).** The sample correlation function converges in distribution to the ratio of two independent stable random variables with indices α and $\alpha/2$, respectively.

2.3. **the limit distribution for the least squares estimates.** The limit distribution for the least squares estimates of the parameters in an AR model.

3. Propositions

Proposition 3.1. *If $2 \leq \alpha < 4$ and $EZ_t = 0$, then for every positive integer h,*

$$a_n^{-2}(n\hat{\gamma}(h) - \sum_{t=1}^{n} \sum_{i=-\infty}^{\infty} c_i c_{i+h} Z_{t-i}^2) \to_p 0,$$

where

$$a_n = \inf\{x; P(|Z_1| > x) \leq n^{-1}\}.$$

SUMMARY-MARTINGALE CENTRAL LIMIT THEOREMS

GEN RYU

1. NOTATIONS

Let $\{S_n, \mathcal{F}_n, n = 1, 2, \dots\}$ be a martingale on the probability space $\{\Omega, \mathcal{F}, P\}$, with

$$\begin{aligned} S_0 &= 0, \\ X_n &= S_n - S_{n-1}, \qquad n = 1, 2, \dots. \end{aligned}$$

note. $\mathcal{F}_0$ need not be the trivial σ-field $\{\emptyset, \Omega\}$.

$$\phi_j(t) = E(e^{itX_j} | \mathcal{F}_{j-1}) = E_{j-1}(e^{itX_j}), \tag{1.1}$$

$$\sigma_n^2 = E_{n-1}(X_n^2), \tag{1.2}$$

$$V_n^2 = \sum_{j=1}^{n} \sigma_j^2; \tag{1.3}$$

$$s_n^2 = EV_n^2 = ES_n^2, \tag{1.4}$$

$$b_n = s_n^{-2} \max_{j \le n} \sigma_j^2. \tag{1.5}$$

2. DEFINITION

2.1. **the Lindeberg condition.** the Lindeberg condition is said to hold if the martingale satisfies

$$V_n^2 s_n^{-2} \to_p 1$$

and

$$s_n^{-2} \sum_{j=1}^{n} EX_j^2 I(|X_j| \ge \epsilon s_n) \to_p 0 \quad \text{as } n \to \infty.$$

for all $\epsilon > 0$.

Date: January 28, 2012.

3. ASSUMPTIONS

Assumption 3.1.

$$V_n^2 s_n^{-2} \to_p 1$$

For this class of martingales, the Lindeberg condition is said to hold if

Assumption 3.2.

$$s_n^{-2} \sum_{j=1}^{n} EX_j^2 I(|X_j| \geq \epsilon s_n) \to_p 0 \quad \textit{as } n \to \infty.$$

for all $\epsilon > 0$.

ROBUSTNESS

YAN LIU

1. Reference

Künsch (1984), AS.

2. notations

2.1. Notations.

1. $h(x)$	a known probability density on $\mathbb{R}$
2. σ^2	variance of U_i
3. $\rho(x,n)^m$	m-dimensional marginal distribution of stationary processes
4. $\mathcal{M}^m$	the set of $\rho(x,n)^m$
5. $\theta \in \Theta \subset \mathbb{R}^q$ $(q \leq p+2)$	unknown parameter
6. T	a functional $\mathcal{M}^m \to \Theta$ (or restrict T to a certain subset of $\mathcal{M}^m$)
7. $\gamma^* = \sup_x \lvert \mathrm{IC}_T(x,\theta) \rvert$	gross error sensitivity
8. $\theta = (\theta_1, \theta_2)$	
9. $\theta_1 = \sigma$ and $\theta_2 = (\eta, \beta_1, \ldots, \beta_p)$	
10. $\kappa = (\kappa_1, \kappa_2)$	
11. $\psi = (\psi_1, \psi_2)$	

2.2. Fundamental Setting.

(i) AR(p) process

$$(X_i - \eta) = \sum_{k=1}^{p} \beta_k (X_{i-k} - \eta) + U_i, \quad \text{i.i.d. } U_i$$

Using $x_i^* = x_i - \eta$, κ is defined by

$$\kappa(x_1, \ldots, x_{p+1}; \theta) = \frac{\partial}{\partial \theta} \log \frac{1}{\sigma} h \left(\frac{x_{p+1}^* - \sum \beta_k x_{p+1-k}^*}{\sigma} \right)$$

Furthermore, let u denote

$$u = x_{p+1}^* - \sum \beta_k x_{p+1-k}^*$$

(ii) m-dimensional marginal distributions

Date: September 1, 2014.

(a) Define $x_i = x_{i-kn}$ if $i > kn$ for $k \in \mathbb{N}$;
(b) m-dimensional marginal distributions

$$\rho(x,n)^m = n^{-1} \sum_{i=1}^{n} \delta(x_i, \dots, x_{i+m-1}),$$

where $\delta(x_i, \dots, x_{i+m-1})$ is the point mass at $x \in \mathbb{R}^m$.

(iii) M-estimator defined by

$$\sum_{j=1}^{n-m+1} \psi(x_j, \cdots, x_{j+m-1}; \hat{\theta}_n) = 0.$$

(iv) Choice of the functional T : for $T(\mu_\theta^m) = \theta$,

$$\hat{\theta}_n(x_1, \dots, x_n) = T(\rho(x,n)^m).$$

(v) Any version of the influence function $\mathrm{IC}_T(x,\theta)$

$$\int \mathrm{IC}_T(x,\theta)\mu_\theta(dx_m|x_{m-1}, \cdots, x_{m-p}) = 0.$$

(vi) Asymptotical variance-covariance matrix

$$C(T,\theta) = \int \mathrm{IC}_T(x,\theta)\mathrm{IC}_T(x,\theta)^T \mu_\theta^m(dx).$$

Lemma 2.1. *If $\{X_i\}_{i\in\mathbb{Z}}$ is stationary ergodic process, then*

$$\rho(x,n)^m \to \mu^m \quad \textit{as } n \to \infty.$$

2.3. **Hampel's optimality problem.** Minimize the trace of the asymptotic covariance matrix $C(T,\theta)$ among all estimators of (iv) which have an influence function and for which

$$\gamma^* = \sup_x |\mathrm{IC}_T(x,\theta)| \leq c(\theta).$$

2.4. **Huber function.**

$$H_c(x) = x \min(1, \frac{c}{|x|})$$

3. Fundamental Theorems

Theorem 3.1 (Künsch (1984), Theorem 1.1)**.** *A functional $L : \mathcal{M}^m \to \mathbb{R}$ is of the form $L(\nu^m) = \int t(x)\nu^m(dx)$ with t bounded and continuous iff L is affine and weakly continuous.*

Theorem 3.2 (Künsch (1984), Theorem 1.2)**.** *$\int t(x)\nu^m(dx) = 0$ for all $\nu^m \in \mathcal{M}^m$ iff $t(x_1, \dots, x_m) = g(x_1, \dots, x_{m-1}) - g(x_2, \cdots, x_m)$ with an arbitrary g.*

Theorem 3.3 (Künsch (1984), Theorem 1.3)**.** *Let μ denote the distribution of an AR(p)-process. If $\mathbb{R}^m \to \mathbb{R}$, $m > p$, is continuous, $\sup|f(x)|/(1+|x|) < \infty$ and $\int f(x)\mu^m(dx) = 0$, then there exists a continuous function $g : \mathbb{R}^{m-1} \to \mathbb{R}$ with $\sup|g(x)|/(1+|x|) < \infty$ and*

$$\int f(x_1, \dots, x_m) + g(x_1, \dots, x_{m-1}) - g(x_2, \cdots, x_m)\mu(dx_m|x_{m-1}, \cdots, x_{m-p}) = 0$$

for all $x_1, \ldots, x_{m-1}$. g is unique up to an additive constant.

4. OPTIMAL ROBUST ESTIMATORS

Theorem 4.1. *Suppose σ is known and $h(x) = h(-x)$. If the bound $c(\theta)$ is such that*

$$\int H_{c(\theta)}(A(\theta)\kappa(x,\theta))\kappa(x,\theta)^T \mu_\theta^{p+1}(dx) = \mathrm{Id}$$

has a solution $A(\theta)$ for all θ, then the solution of Hampel's optimal problem is given by

$$\begin{cases} m = p+1, \\ \psi(x,\theta) = H_{c(\theta)}(A(\theta)\kappa(x,\theta)). \end{cases}$$

Theorem 4.2. *If $\psi_1(x_1, \ldots, x_{p+1}; \theta) = \chi\big(\frac{u}{\sigma}\big)$ with $\chi(\cdot)$ even and ψ_2 is one of the optimal solutions of Hampel's problem, then $\hat{\sigma}$ is asymptotically independent of $\hat{\theta}_2$ and the asymptotic covariance for $\hat{\theta}_2$ is the same as for known σ.*

5. WORDS

1. commence 始める
2. dwell いる
3. minutiae ささいな事柄
4. predator-prey 捕食者-犠牲者
5. sonar ソナー

6. Further Reading

6.1. robustness.
Robust statistics, Maronna, Martin and Yohai

7. NOTATIONS

1. $\{y_t, x_t, r_t\}_{t=1}^T$ a stationary and weakly depedent process
2. $x_t \in \mathcal{X} \subset \mathbb{R}^{d_x}$
3. $y_t \in \mathcal{Y} \subset \mathbb{R}$
4. H_0 the null hypothesis
5. $\nu(y, g)$ a mean zero Gaussian process
6. $\mathcal{B}$ $\{(y,g) \in (\mathcal{Y} \times \mathcal{G}); E\,1(y_t \le y)g(x_t)(\pi_0^+(x_t) - 1(r_{t-1} \ge 0)) = 0\}$.

7.1. Hypotheses.

- the null hypothesis H_0

$$F^+(y|x) \ge F^-(y|x) \quad \text{a.s. for all } (y,x) \in \mathcal{Y} \times \mathcal{X}.$$

  Rewrite the notations by π_0^+ where

$$\pi_0^+(x) = P(r_{t-1} \ge 0 | x_t = x).$$

Then the null hypothesis H_0 is

$$E\Big[1(y_t \le y)\Big(\frac{1(r_{t-1} < 0)}{\pi_0^-(x_t)} - \frac{1(r_{t-1} \ge 0)}{\pi_0^+(x_t)}\Big)\Big|x_t = x\Big] \le 0$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$. Moreover using instrument $g \in \mathcal{G}$ where

$$\mathcal{G} = \{g_{a,b}; g_{a,b}(x) = \prod_{i=1}^{d_x} 1(a_i < x_i \le b_i) \quad \text{for some } a, b \in \mathcal{X}\},$$

the hypothesis can be simplified more.

7.2. **Test statistic.**

$$S_T = \sup_{(y,g)\in\mathcal{Y}\times\mathcal{G}} \sqrt{T}\bar{m}_T(y, g, \hat{\pi}^+),$$

where

$$\bar{m}_T(y, g, \pi) = \frac{1}{T}\sum_{t=1}^{T} 1(y_t \le y)g(x_t)\{\pi(x_t) - 1(r_{t-1} \ge 0)\},$$

and Nadaraya-Watson's kernel

$$\hat{\pi}^+(x) = \frac{\sum_{t=2}^{T} 1(r_{t-1} > 0K_h(x - x_t))}{\sum_{t=2}^{T} K_h(x - x_t)},$$

where $K : \mathbb{R}^{d_x} \to \mathbb{R}$.

7.3. **Empirical process.**

$$\nu_T(y, g) = \sqrt{T}\{\xi_T(y, g) - E\xi_T(y, g)\},$$

where

$$\xi_T(y, g) = \frac{1}{T}\sum_{t=1}^{T}\{1(y_t \le y) - F(y|x_t)\}g(x_t)\{\pi_0^+(x_t) - 1(r_{t-1} \ge 0)\}.$$

7.4. **Covariance of $\nu(y, g)$.**

$$C((y_1, g_1), (y_2, g_2)) = \lim_{T\to\infty} \mathrm{Cov}(\nu_T(y_1, g_1), \nu_T(y_2, g_2)).$$

8. Results

Theorem 8.1. *Under Assumptions and the null hypothesis H_0,*

$$S_T \Rightarrow \begin{cases} \sup_{(y,g)\in\mathcal{B}} \nu(y, g) & \text{if } \mathcal{B} \ne \emptyset \\ -\infty & \text{if } \mathcal{B} = \emptyset \end{cases}.$$

THEORY OF RANK TESTS

GEN RYU

1. WORDS

(1) treatise 専門書
(2) substantially 実質上
(3) sadly 不幸にも
(4) untimely 時を誤った
(5) flourish 栄える
(6) former colleagues 前の同僚
(7) press ... to 人に... をせがむ、懇願する
(8) come to life 生まれる
(9) tribute 捧げもの
(10) substantial かなりの
(11) incorporated 取り入れられる
(12) designated つけられる
(13) painstaking 骨をおる
(14) patience 辛抱
(15) long-lasting 長く続く
(16) inspire 奮い立たせる
(17) be acquainted with 知っている
(18) striving するよう努力する
(19) lucidity 明晰、明快
(20) in this respect この点において
(21) formulae = formula
(22) supplied with 与えられる
(23) complement 補完
(24) omitting を省略する
(25) coverage 範囲、概観
(26) comprise を含む
(27) sector 部門
(28) due 当然与えられるべきな
(29) dual 二重の
(30) genesis 起源
(31) composite 複合的な
(32) prescribe 規定する
(33) stringent きびしい
(34) go far beyond 優に超える
(35) blended 混ぜ合わせられた
(36) harmoniously 調和的に
(37) ally と同類である
(38) amenable 従いやすい
(39) tenable 持続できる
(40) thrust 目標
(41) appraisal 評価
(42) constitute を構成する
(43) pertain 付属する
(44) sequential 順次的な
(45) depict 表現する
(46) intricate 入り組んだ、錯綜した
(47) encompass を含む
(48) sibling きょうだい
(49) duality 二重性
(50) alignment 整列
(51) invincible 無敵の
(52) annex 併合する
(53) intervention 介入
(54) at the cost of sacrificing を犠牲にして
(55) in favor of 賛成して
(56) solid 確固たる
(57) crop up 現れる
(58) intricate 入り組んだ
(59) tie up with つなぐ
(60) stood 立てられた
(61) for a while しばらくの間
(62) breakthrough 前進
(63) appraise 評価する
(64) nuisance 邪魔者
(65) emerge 浮かび上がる
(66) viable 生きていける
(67) focal 焦点の
(68) pertinent 適切な
(69) depiction 描写
(70) flavor 風味
(71) bread and butter 必要最低限な

Date: February 14, 2012.

(72) contrast 対照する
(73) gear 合うように調整する
(74) genuine 本物の
(75) facilitate を容易にする
(76) contemplated 熟慮した
(77) vigorous たくましい
(78) intrude 侵入する
(79) synopsis 概要
(80) auxiliary 補助
(81) incur 招く
(82) realistically 現実的に
(83) universally いつも
(84) attach くっつける
(85) conjecture 推測、憶測
(86) affix 添加物
(87) induce 気にさせる
(88) reserved position 控えめな
(89) abundance 多数
(90) conceivable 想像できる
(91) tendency 傾向
(92) adhere 執着
(93) commit 犯す
(94) reluctant 気の進まない
(95) there is no need to する必要がない
(96) there is no reasonable rule for する理由がない
(97) irreversible 不可逆の
(98) utilize 役立たせる
(99) clear-cut 明らかな
(100) embrace (考え) を抱く
(101) reverse 逆の
(102) envelope 包絡線
(103) in no way = not at all 決して
(104) definitive 決定的な
(105) quote 持ち出す
(106) discus 円盤
(107) indeterminate 不確定な
(108) advantages and shortcomings 善し悪し
(109) in ascending order 昇順
(110) on account of = because of の為に
(111) a fortiori 結果的に
(112) the question of how... question の使い方
(113) amount to と言っていい
(114) expedient 便利だ
(115)
(116)
(117) validly 妥当に
(118) appealing 魅力的な
(119) prominent 卓越した
(120) tacitly 暗黙の

WAVELET TRANSFORMATION

YAN LIU

1. Introduction

Both the Fourier transformation and the wavelet transformation transform the function from time domain to frequency domain. The main idea is based on the theory of the basis for functions. The difference between two methods is that the Fourier transformation is based on the basis

$$\mathfrak{B} = \{\mathfrak{b}(t - nq_0)e^{imp_0t}; m, n \in \mathbb{Z}\},$$

while the wavelet transformation is based on the basis

$$\mathfrak{B} = \{|p_0|^{-m/2}\psi(p_0^{-m}t - nq_0); m, n \in \mathbb{Z}\}.$$

Here, $\psi(\cdot)$ is called *mother wavelet.*

2. Theory for basis

2.1. the basis for wavelets.

Definition 2.1 (MRA)**.** The closed subspace $\{V_j; j \in \mathbb{Z}\} \subset L^2(\mathbb{R})$ is called *multiresolution analysis (MRA)* if

(i) $V_j \subset V_{j+1}, j \in \mathbb{Z}$,
(ii) $\cap_{j\in\mathbb{Z}} V_j = \{0\}$, $(\cup_{j\in\mathbb{Z}} V_j)^c = L^2(\mathbb{R})$,
(iii) $f(x) \in V_j$ if and only if $f(2x) \in V_{j+1}$,
(iv) there exists a function $\varphi(x) \in V_0$ such that $\{\varphi(x - k); k \in \mathbb{Z}\}$ is the orthornormal basis for V_0.

Here, $\varphi(\cdot)$ is called *scaling function.*

Note that $L^2(\mathbb{R})$ can be always represented by

$$L^2(\mathbb{R}) = V_J \oplus \sum_{s=J}^{\infty} \oplus W_s.$$

Define

$$h_k = \sqrt{2}\int_{\mathbb{R}} \varphi(x)\overline{\varphi(2x - k)}dx.$$

Date: July 28, 2014.

The scaling function $\varphi(x)$ satisfies

$$\varphi(x) = \sqrt{2}\sum_k h_k \varphi(2x-k).$$

Mother wavelet $\psi(x)$ is defined by

$$\psi(x) = \sum_k (-1)^k \overline{h}_{1-k}\varphi(2x-k).$$

As a result, any function $f(x) \in L^2(\mathbb{R})$ has a representation

$$\begin{aligned} f(x) &= \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} b_{j,k}\psi_{j,k}(x) \\ &= \sum_{k=-\infty}^{\infty} \alpha_{J,k}\varphi_{J,k}(x) + \sum_{j=J}^{\infty}\sum_{k=-\infty}^{\infty} \beta_{j,k}\psi_{j,k}(x). \end{aligned}$$

Obviously,

$$\begin{aligned} b_{j,k} &= \langle\, f, \psi_{j,k}\,\rangle, \quad j,k \in \mathbb{Z}, \\ \alpha_{J,k} &= \langle\, f, \varphi_{J,k}\,\rangle, \quad k \in \mathbb{Z}, \\ \beta_{j,k} &= \langle\, f, \psi_{j,k}\,\rangle, \quad j \geq J, k \in \mathbb{Z}. \end{aligned}$$

2.2. **norm.** Define $\tau_h f(x) = f(x-h)$. The Besov space $B^s_{p,q}$ is defined as follows: for $f \in L^p(\mathbb{R})$, $1 \leq p \leq \infty$,

(1) for $s \in (0,1)$,

$$\begin{aligned} \gamma_{s,p,q}(f) &= \Big\{\int_{\mathbb{R}} \Big(\frac{\|\tau_h f - f\|_{L^p}}{|h|^s}\Big)^q \frac{dh}{|h|}\Big\}^{1/q}, \\ \gamma_{s,p,\infty}(f) &= \sup_{h\in\mathbb{R}} \frac{\|\tau_h f - f\|_{L^p}}{|h|^s}, \end{aligned}$$

(2) for $s = 1$,

$$\begin{aligned} \gamma_{1,p,q}(f) &= \Big\{\int_{\mathbb{R}} \Big(\frac{\|\tau_h f + \tau_{-h} f - 2f\|_{L^p}}{|h|}\Big)^q \frac{dh}{|h|}\Big\}^{1/q}, \\ \gamma_{1,p,\infty}(f) &= \sup_{h\in\mathbb{R}} \frac{\|\tau_h f + \tau_{-h} f - 2f\|_{L^p}}{|h|^s}, \end{aligned}$$

then we say

$$f \in B^s_{p,q} \iff \gamma_{s,p,q}(f) < \infty, f \in L^p(\mathbb{R}).$$

The Besov norm for $f(x) \in L^2(\mathbb{R}) \cap B^s_{2,q}$ is well defined and

$$\|f\|^s_{2,q} = \Big(\sum_k |\alpha_{0,k}|^2\Big)^{1/2} + \Big[\sum_{l\geq 0}\Big(2^{ls}\Big(\sum_k |\beta_{l,k}|^2\Big)^{1/2}\Big)^q\Big]^{1/q}$$

3. Further reading

Kato and Masry (1999) for wavelet transform of fractional Brownian motion, Donoho and Johnstone (1994, 1995) and Donoho et al. (1995, 1997) for wavelet for statistics. Also see Japanese work like Kawasaki and Shibata (1995) and Shibata and Takagiwa (1997).

4. Idea

The main purpose is to transmit functions using some finite device. Suppose $f(x) \in L^2$. It is known that f can be represented by the basis in L^2.

$$f = \sum_n a_n f_n.$$

Precisely, the wavelet is defined as follows.

Definition 4.1. A wavelet is a function $\Psi(t) \in L^2(\mathbb{R})$ such that the family of functions

$$\Psi_{j,k} = 2^{j/2}\Psi(2^j t - k)$$

where j and k are arbitrary integers, is an orthonormal basis in the Hillbert space $L^2(\mathbb{R})$.

5. words

1. archetypal	原型の、典型的な
2. recipient	受取人
3. acoustics	音響学
4. seismology	地震学
5. depict	表現する
6. for the time being	差し当たって
7. spring to mind	頭に浮かぶ
8. intrinsic	固有の