SHRINKAGE ESTIMATION FOR THE AUTOCOVARIANCE MATRIX OF VECTOR-VALUED GAUSSIAN STATIONARY PROCESSES

Yoshihiro Suto

Received July 3, 2014; revised September 12, 2014

ABSTRACT. We discuss the problem of shrinkage estimation for the autocovariance matrix of a Gaussian stationary vector-valued process to improve on the usual sample autocovariance matrix with respect to the mean squares error. We propose a kind of empirical Bayes estimators when the mean of the stochastic process is zero and non-zero. We show that the shrinkage estimators dominate the usual estimators, and the asymptotic risk differences are similar to that of scalar-valued Gaussian stationary processes. This result seems to be useful for the autocovariance estimation with vector-valued dependent observations.

1 Introduction There have been many discussions on shrinkage estimation to improve on the sample mean and the sample covariance of independent observations. Stein [6] showed the inadmissibility of the sample mean for k-dimensional independent normal observations when $k \ge 3$. James and Stein [5] suggested a shrinkage estimator which dominates the sample mean with respect to the mean squares error when $k \ge 3$. Furthermore, in the univariate case, Stein [7] proposed a truncated estimator and showed the estimator improves on the usual sample variance. Also in the multivariate case, Haff [2] proposed an empirical Bayes estimator for the normal covariance matrix and showed the estimator improves on the sample covariance matrix.

All mentioned above are the discussions for independent normal observations. However, it is natural that the actual data are dependent. Therefore, it is important to consider the shrinkage estimators which dominate the usual sample mean and the autocovariance when the observations are dependent. For a vector-valued Gaussian process, Taniguchi and Hirukawa [8] gave a sufficient condition for James-Stein type estimator to dominate the sample mean. Furthermore, for the scalar-valued Gaussian stationary process, Taniguchi *et al.* [9] suggested an empirical Bayes estimator motivated by Haff [2] and discussed on the improvement by the estimator.

Since it is useful to represent the actual time series data by dependent and multivariate statistical models, in this paper, we consider improved autocovariance estimation for vector-valued Gaussian stationary processes motivated by Taniguchi *et al.* [9]. We propose shrinked autocovariance estimators, and show that the estimators dominate the usual autocovariance estimators in case of vector-valued Gaussian stationary processes.

This paper is organized as follows. In Section 2, we introduce empirical Bayes estimators in view of Taniguchi *et al.* [9] when the mean of the stochastic process is zero and non-zero. Then we evaluate the asymptotic risk differences by the mean squares error between the shrinkage estimator and the usual sample autocovariance matrix. The improvements by the shrinkage estimators are expressed in terms of the spectral density of the process. Section 3 provides the proofs of theorems in Section 2.

²⁰⁰⁰ Mathematics Subject Classification. 60G10, 60G15, 62H12.

Key words and phrases. shrinkage estimation, Gaussian stationary process, autocovariance, spectral density, mean squares error.

Y. SUTO

Throughout this paper, Z denotes the set of all integers, and \otimes denotes the Kronecker product of matrices, and \circ denotes the Hadamard product (entrywise product) of matrices.

2 Shrinkage estimators for autocovariance matrix Let $\{X(t), t \in Z\}$ be an *m*dimensional Gaussian stationary process with mean $E(X(t)) = \mu$ and autocovariance matrix $\gamma(s) = E[(X(t) - \mu)(X(t+s) - \mu)']$ for $s \in \mathbb{Z}$ and all $t \in \mathbb{Z}$. We assume that $\gamma(s)$'s satisfy

Assumption 1.

$$\sum_{s=-\infty}^{\infty} |s| \cdot \|\boldsymbol{\gamma}(s)\| < \infty,$$

where $\|\cdot\|$ is the Euclidean norm. Then the spectral density matrix of the process is given by

(1)
$$\boldsymbol{f}(\lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \boldsymbol{\gamma}(s) e^{-\boldsymbol{i}s\lambda}$$

Here we consider to estimate the autocovariance matrix

(2)
$$\Gamma = \begin{pmatrix} \gamma(0) & \gamma(-1) & \dots & \gamma(1-p) \\ \gamma(1) & \gamma(0) & \dots & \gamma(2-p) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{pmatrix}$$

for positive integer p. Since $\gamma(-s) = \gamma(s)'$, Γ is symmetric. Suppose that an observed stretch $\{X(1), \ldots, X(n)\}$ of the process $\{X(t)\}$ is available. When $\mu = 0$, the usual estimator for Γ is

(3)
$$\hat{\Gamma}_0 = \frac{1}{n-k} S_n,$$

where

$$S_n = \sum_{t=p}^n \boldsymbol{Y}(t)\boldsymbol{Y}(t)', \quad \boldsymbol{Y}(t) = (\boldsymbol{X}(t)', \dots, \boldsymbol{X}(t-p+1)')',$$

and k = 0 or p - 1. When $\mu \neq 0$, the usual estimator for Γ is

(4)
$$\tilde{\Gamma}_0 = \frac{1}{n-k}\tilde{S}_n,$$

where

$$\tilde{S}_n = \sum_{t=p}^n \tilde{\mathbf{Y}}(t)\tilde{\mathbf{Y}}(t)', \quad \tilde{\mathbf{Y}}(t) = ((\mathbf{X}(t) - \bar{\mathbf{X}}_n)', \dots, (\mathbf{X}(t-p+1) - \bar{\mathbf{X}}_n)')',$$

with $\bar{\mathbf{X}}_n = n^{-1} \sum_{t=1}^n \mathbf{X}(t)$ and k = 0 or p - 1. We measure the goodness of $\hat{\Gamma}_0$ by the following mean squares error loss function

(5)
$$L(\hat{\Gamma}_0, \Gamma) = \operatorname{tr}\{(\hat{\Gamma}_0 \Gamma^{-1} - \boldsymbol{I}_{mp})^2\}$$
 (\boldsymbol{I}_{mp} is the $mp \times mp$ identity matrix)

and the risk $R(\hat{\Gamma}_0, \Gamma) = E\{L(\hat{\Gamma}_0, \Gamma)\}$. Similarly, for $\tilde{\Gamma}_0$ we also define $L(\tilde{\Gamma}_0, \Gamma)$ and $R(\tilde{\Gamma}_0, \Gamma)$. Next, we consider to improve the estimators $\hat{\Gamma}_0$ and $\tilde{\Gamma}_0$ with respect to the risk $R(\cdot, \cdot)$. When $\{X(t)\}$ is a scalar-valued process, Taniguchi *et al.* [9] introduced the following empirical Bayes estimators

(6)
$$\hat{\Gamma} = \frac{1}{n-k} \left(S_n + \frac{b}{n \operatorname{tr}(S_n^{-1}C)} C \right)$$

and

(7)
$$\tilde{\Gamma} = \frac{1}{n-k} \left(\tilde{S}_n + \frac{b}{n \operatorname{tr}(\tilde{S}_n^{-1}C)} C \right)$$

to improve $\hat{\Gamma}_0$ and $\tilde{\Gamma}_0$, respectively, where *b* is a constant and *C* is a positive definite matrix of the same size as Γ , and showed that $\hat{\Gamma}$ and $\tilde{\Gamma}$ dominate $\hat{\Gamma}_0$ and $\tilde{\Gamma}_0$, respectively, with respect to the risk. Similarly, when $\{\boldsymbol{X}(t)\}$ is a vector-valued process, we use the estimators in the form of (6) and (7), and show that $\hat{\Gamma}$ and $\tilde{\Gamma}$ dominate $\hat{\Gamma}_0$ and $\tilde{\Gamma}_0$, respectively. To evaluate the improvement of the estimator, we need the following assumption.

Assumption 2. C is symmetric.

The assumption seems to be natural because S_n in (6) and S_n in (7) are symmetric and $\hat{\Gamma}$ in (6) and $\tilde{\Gamma}$ in (7) should be symmetric. Then, the following theorem holds.

Theorem 1. When $\mu = 0$, suppose that Assumptions 1 and 2 hold. Then the asymptotic risk difference for the estimator $\hat{\Gamma}_0$ and $\hat{\Gamma}$ is

(8)
$$\lim_{n \to \infty} n^2 [R(\hat{\Gamma}_0, \Gamma) - R(\hat{\Gamma}, \Gamma)] = -b \frac{\operatorname{tr}\{(C\Gamma^{-1})^2\}}{\{\operatorname{tr}(\Gamma^{-1}C)\}^2} [b+B],$$

where (0)

$$B = \begin{cases} 2(-p+1)\frac{\{\operatorname{tr}(\Gamma^{-1}C)\}^{2}}{\operatorname{tr}\{(C\Gamma^{-1})^{2}\}} \\ +\frac{8\pi}{\operatorname{tr}\{(C\Gamma^{-1})^{2}\}} \int_{-\pi}^{\pi} \operatorname{tr}\{[\{(G(\lambda) \otimes \mathbf{I}_{m})\Gamma^{-1}C\Gamma^{-1}\} \circ (\mathbf{I}_{p} \otimes U_{m})\}(U_{p} \otimes \mathbf{f}(\lambda))]^{2}\}d\lambda, \\ (\text{if } k = 0), \\ \frac{8\pi}{\operatorname{tr}\{(C\Gamma^{-1})^{2}\}} \int_{-\pi}^{\pi} \operatorname{tr}\{[\{(G(\lambda) \otimes \mathbf{I}_{m})\Gamma^{-1}C\Gamma^{-1}\} \circ (\mathbf{I}_{p} \otimes U_{m})\}(U_{p} \otimes \mathbf{f}(\lambda))]^{2}\}d\lambda, \\ (\text{if } k = p - 1). \end{cases}$$

with $G(\lambda) = (e^{-i(h-l)\lambda})_{h,l=1,\ldots,p}$ ($p \times p$ matrix), $U_m = \mathbf{1}_m \mathbf{1}'_m$ and $\mathbf{1}_m = (1,\ldots,1)'$ ($m \times 1$ vector).

We can see that this result includes Theorem 1 of Taniguchi et al. [9] as special case.

When $\mu \neq 0$, we can show the following theorem for $\tilde{\Gamma}_0$ and $\tilde{\Gamma}$.

Theorem 2. When $\mu \neq 0$, suppose that Assumptions 1 and 2 hold. Then the asymptotic risk difference for the estimator $\tilde{\Gamma}_0$ and $\tilde{\Gamma}$ is

(10)
$$\lim_{n \to \infty} n^2 [R(\tilde{\Gamma}_0, \Gamma) - R(\tilde{\Gamma}, \Gamma)] = -b \frac{\operatorname{tr}\{(C\Gamma^{-1})^2\}}{\{\operatorname{tr}(\Gamma^{-1}C)\}^2} [b + \tilde{B}],$$

where

(11)
$$\tilde{B} = B - 4\pi \frac{\operatorname{tr}\{(U_p \otimes f(0))\Gamma^{-1}C\Gamma^{-1}\} \cdot \operatorname{tr}\{\Gamma^{-1}C\}}{\operatorname{tr}\{(C\Gamma^{-1})^2\}}.$$

We can see that this result includes Theorem 2 of Taniguchi et al. [9] as special case.

Y. SUTO

3 Proofs This section provides the proofs of the theorems. We need the following lemma to prove Theorem 1 (for the proofs, see Lemma A2.3 of Hosoya and Taniguchi [4] and Theorem 4.5.1 of Brillinger [1]).

Lemma 1. Suppose that Assumption 1 holds.

(a) Denote the α -th component of $\mathbf{X}(t)$ by $X_{\alpha}(t)$, and denote the (α, β) -th component of $f(\lambda)$ by $f_{\alpha\beta}(\lambda)$. If $\{\mathbf{X}(t)\}$ is Gaussian, Then

$$\lim_{n \to \infty} n \operatorname{Cov} \left\{ \frac{1}{n} \sum_{t=p}^{n} X_{\alpha_1}(t-j_1) X_{\alpha_2}(t-j_2), \ \frac{1}{n} \sum_{t=p}^{n} X_{\alpha_3}(t-j_3) X_{\alpha_4}(t-j_4) \right\}$$
$$= 2\pi \int_{-\pi}^{\pi} \{ f_{\alpha_1 \alpha_3}(\lambda) \overline{f_{\alpha_2 \alpha_4}(\lambda)} e^{-i(j_1-j_2+j_4-j_3)\lambda} + f_{\alpha_1 \alpha_4}(\lambda) \overline{f_{\alpha_2 \alpha_3}(\lambda)} e^{i(j_2-j_1+j_4-j_3)\lambda} \} d\lambda$$
$$= W_{j_1,\dots,j_4}^{\alpha_1,\dots,\alpha_4} \ (say) \ (0 \le j_1,\dots,j_4 \le p-1).$$

(b) Denote the (α, β) -th component of $\gamma(s)$ by $\gamma_{\alpha\beta}(s)$. Then

$$\frac{1}{\sqrt{n}} \sum_{t=p}^{n} \{ X_{\alpha}(t-j_1) X_{\beta}(t-j_2) - \gamma_{\alpha\beta}(j_1-j_2) \} = O(\sqrt{\log n}), \quad a.s.$$

Proof of Theorem 1 We can calculate the asymptotic risk difference in the vector-valued case as same as (19) of Taniguchi *et al.* [9]. In the proof of Theorem 1 of [9], we can use the form of (23) of [9]. Therefore we only evaluate the numerator in the expectation of (23) of [9]. The numerator is given by

(12)
$$E\left[\left(\operatorname{tr}\left\{\sqrt{n}\left(\frac{1}{n-k}S_n - E\left(\frac{1}{n-k}S_n\right)\right)\Gamma^{-1}C\Gamma^{-1}\right\}\right)^2\right]$$

Here we set $Z = \frac{1}{n-k}S_n - E\left(\frac{1}{n-k}S_n\right)$ and $V = \Gamma^{-1}C\Gamma^{-1}$. Then (12) is equal to

(13)
$$nE\left[\left\{\sum_{h=1}^{p}\sum_{l=1}^{p}\operatorname{tr}\left(Z^{hl}V^{lh}\right)\right\}^{2}\right],$$

where Z^{hl} and V^{hl} are the (h, l)-th $m \times m$ block matrices of Z and V, respectively. Denote the (i, j)-th component of Z^{hl} and V^{lh} by Z^{hl}_{ij} and V^{lh}_{ij} , respectively. Then (13) is equal to

(14)
$$\sum_{h,l,h',l'=1}^{p} \sum_{i,j,i',j'=1}^{m} nE[Z_{ij}^{hl} Z_{i'j'}^{h'l'}] V_{ji}^{lh} V_{j'i'}^{l'h'}.$$

Let S_n^{hl} be the (h, l)-th $m \times m$ block matrix of S_n . Since $Z^{hl} = \frac{1}{n-k} \{S_n^{hl} - E[S_n^{hl}]\}$ and $S_n^{hl} = \sum_{t=p}^n \mathbf{X}(t-h+1)\mathbf{X}(t-l+1)'$, (14) is equal to

(15)
$$\sum_{\substack{h,l,h',l'=1\\ i,j,i',j'=1}}^{p} \sum_{\substack{i,j,i',j'=1\\ (n-k)^2}}^{m} \frac{n^2}{V_{ji}^{lh} V_{j'i'}^{l'h'}} \times n \operatorname{Cov}\left(\frac{1}{n} \sum_{\substack{t=p\\ t=p}}^{n} X_i(t-h+1) X_j(t-l+1), \frac{1}{n} \sum_{\substack{t=p\\ t=p}}^{n} X_{i'}(t-h'+1) X_{j'}(t-l'+1)\right).$$

Using Lemma 1(a), as $n \to \infty$, (15) converges to

(16)
$$\sum_{\substack{h,l,h',l'=1\\ \lambda \geq \pi}}^{p} \sum_{\substack{i,j,i',j'=1\\ j'i'}}^{m} V_{ji}^{lh} V_{j'i'}^{l'h'} \times 2\pi \int_{-\pi}^{\pi} \{f_{ii'}(\lambda) f_{j'j}(\lambda) e^{-i(h-l+l'-h')\lambda} + f_{ij'}(\lambda) f_{i'j}(\lambda) e^{i(l-h+l'-h')\lambda} \} d\lambda.$$

Here, by Assumption 2, V is symmetric and then $(V^{lh})' = V^{hl}$. Therefore (16) is equal to

(17)
$$4\pi \int_{-\pi}^{\pi} \operatorname{tr} \left\{ \sum_{h,l=1}^{p} e^{-(h-l)\lambda} V^{lh} \boldsymbol{f}(\lambda) \sum_{h',l'=1}^{p} e^{-(h'-l')\lambda} V^{l'h'} \boldsymbol{f}(\lambda) \right\} d\lambda.$$

Therefore (17) can be expressed as

(18)
$$4\pi \int_{-\pi}^{\pi} \operatorname{tr}\{[\{(G(\lambda) \otimes \boldsymbol{I}_m)V\} \circ (\boldsymbol{I}_p \otimes U_m)\}(U_p \otimes \boldsymbol{f}(\lambda))]^2\} d\lambda,$$

which completes the proof of Theorem 1.

Next, we prove Theorem 2. To prove the theorem we need the following lemma.

Lemma 2. Suppose that Assumption 1 holds. Then, (a)

$$nE[(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu})(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu})'] = 2\pi \boldsymbol{f}(0) + o(1).$$

(b)

$$E\left(\frac{1}{n-k}\tilde{S}_n\right) = \left(1 + \frac{k-p+1}{n-k}\right)\Gamma - \frac{2\pi}{n-k}(U_p \otimes \boldsymbol{f}(0)) + o(n^{-1}).$$

(c) Denote the α -th component of $\bar{\mathbf{X}}_n$ by \bar{X}_n^{α} . Then

$$\lim_{n \to \infty} n \operatorname{Cov} \left\{ \frac{1}{n} \sum_{t=p}^{n} (X_{\alpha_1}(t-j_1) - \bar{X}_n^{\alpha_1}) (X_{\alpha_2}(t-j_2) - \bar{X}_n^{\alpha_2}), \\ \frac{1}{n} \sum_{t=p}^{n} (X_{\alpha_3}(t-j_3) - \bar{X}_n^{\alpha_3}) (X_{\alpha_4}(t-j_4) - \bar{X}_n^{\alpha_4}) \right\}$$
$$= W_{j_1,\dots,j_4}^{\alpha_1,\dots,\alpha_4}.$$

(d)

$$\frac{1}{\sqrt{n}}(\tilde{S}_n - n\Gamma) = O(\sqrt{\log n}), \ a.s.$$

Proof of Lemma 2 (a) is due to [3] (p.208, Corollary 4).

Y. SUTO

(b) \tilde{S}_n^{hl} denotes the (h, l)-th $m \times m$ block matrix of \tilde{S}_n . Then

$$(19) \quad \frac{1}{n-k}\tilde{S}_{n}^{hl} \\ = \frac{1}{n-k}\sum_{t=p}^{n}(\boldsymbol{X}(t-h+1)-\bar{\boldsymbol{X}}_{n})(\boldsymbol{X}(t-l+1)-\bar{\boldsymbol{X}}_{n})' \\ = \frac{1}{n-k}\sum_{t=p}^{n}(\boldsymbol{X}(t-h+1)-\boldsymbol{\mu}+\boldsymbol{\mu}-\bar{\boldsymbol{X}}_{n})(\boldsymbol{X}(t-l+1)-\boldsymbol{\mu}+\boldsymbol{\mu}-\bar{\boldsymbol{X}}_{n})' \\ = \frac{1}{n-k}\sum_{t=p}^{n}(\boldsymbol{X}(t-h+1)-\boldsymbol{\mu})(\boldsymbol{X}(t-l+1)-\boldsymbol{\mu})'+\frac{n-p+1}{n-k}(\bar{\boldsymbol{X}}_{n}-\boldsymbol{\mu})(\bar{\boldsymbol{X}}_{n}-\boldsymbol{\mu})' \\ + \frac{1}{n-k}(\boldsymbol{\mu}-\bar{\boldsymbol{X}}_{n})\sum_{t=p}^{n}(\boldsymbol{X}(t-l+1)-\boldsymbol{\mu})'+\frac{1}{n-k}\sum_{t=p}^{n}(\boldsymbol{X}(t-l+1)-\boldsymbol{\mu})(\boldsymbol{\mu}-\bar{\boldsymbol{X}}_{n})' \\ = \frac{1}{n-k}\sum_{t=p}^{n}(\boldsymbol{X}(t-h+1)-\boldsymbol{\mu})(\boldsymbol{X}(t-l+1)-\boldsymbol{\mu})'-\frac{n}{n-k}(\bar{\boldsymbol{X}}_{n}-\boldsymbol{\mu})(\bar{\boldsymbol{X}}_{n}-\boldsymbol{\mu})' \\ + \frac{1}{n-k}o_{p}(1).$$

From (a), we obtain

$$E\left(\frac{1}{n-k}\tilde{S}_{n}^{hl}\right) = \frac{n-p+1}{n-k}\gamma(h-l) - \frac{1}{n-k}(2\pi f(0) + o(1)) + \frac{1}{n-k}o(1)$$
$$= \left(1 + \frac{k-p+1}{n-k}\right)\gamma(h-l) - \frac{2\pi}{n-k}f(0) + o(n^{-1}).$$

Then we get the result.

(c) From (19), Gaussianity of $\{X_t\}$, and the properties of cumulant, we can show this lemma.

(d) Noting that Theorem 4.5.1 of Brillinger [1], we obtain

$$\sqrt{n}(\bar{X}_n - \mu) = O(\sqrt{\log n}) \quad a.s.$$

From Lemma 1 (b), we can see that (d) holds.

 $Proof \ of \ Theorem \ 2 \ \ We \ can prove the theorem similarly to Theorem 1, except for the evaluation of$

(20)
$$-\frac{2n^2b}{n-k}\operatorname{tr}\left[\left\{E\left(\frac{1}{n-k}\tilde{S}_n\right)-\Gamma\right\}\Gamma^{-1}C\Gamma^{-1}\right]$$

in (21) of [9]. From Lemma 2 (b) it is seen that

(21)
$$\lim_{n \to \infty} (20) = -2b[(k-p+1)\operatorname{tr}\{C\Gamma^{-1}\} - 2\pi \operatorname{tr}\{(U_p \otimes \boldsymbol{f}(0))\Gamma^{-1}C\Gamma^{-1}\}].$$

Therefore we obtain the Theorem 2.

Acknowledgments The author thanks to Professor Masanobu Taniguchi for his comments and encouragements. The author also thanks to the referees for their understandings and constructive comments to improve this paper.

288

| - | | |
|-------|--|---|
| С | | т |
| | | L |

References

- Brillinger, D. R., (1981), Time Series: Data Analysis and Theory, Expanded Edition, San Francisco: Holden-Day.
- [2] Haff, L. R., (1980), Empirical Bayes estimation of the multivariate normal covariance matrix, *The Annals of Statistics*, 586-597.
- [3] Hannan, E. J., (1970), Multiple time series, John Wiley & Sons.
- [4] Hosoya, Y., and Taniguchi, M., (1982), A central limit theorem for stationary processes and the parameter estimation of linear processes, *The Annals of Statistics*, 10, 132-153. Correction: (1993), 21, 1115-1117.
- [5] James, W., and Stein, C., (1961), Estimation with quadratic loss. In Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, Vol. 1, No. 1961, 361-379.
- [6] Stein, C., (1956), Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In Proceedings of the Third Berkeley symposium on mathematical statistics and probability, Vol. 1, No. 399, 197-206.
- [7] Stein, C., (1964), Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean, Annals of the Institute of Statistical Mathematics, **16**(1), 155-160.
- [8] Taniguchi, M., and Hirukawa, J., (2005), The Stein-James estimator for short- and long- memory Gaussian processes, *Biometrika*, 92(3), 737-746.
- [9] Taniguchi, M., Shiraishi, H., and Ogata, H., (2007), Improved estimation for the autocovariances of a Gaussian stationary process, *Statistics*, 41(4), 269-277.

Communicated by Masanobu Taniguchi

Yoshihiro Suto

Department of Pure and Applied Mathematics, Graduate School of Fundamental Science and Engineering, Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan E-mail: y-sutoh-honjo@akane.waseda.jp