CCES Discussion Paper Series
Center for Research on Contemporary Economic Systems

Graduate School of Economics
Hitotsubashi University

CCES Discussion Paper Series, No.6
December 2008

Model Selection Criteria for the Leads-and-Lags Cointegrating Regression

In Choi
(Sogang University)
Eiji Kurozumi
(Hitotsubashi University)

Naka 2-1, Kunitachi, Tokyo 186-8601, Japan
Phone: +81-42-580-9076   Fax: +81-42-580-9102
URL: http://www.econ.hit-u.ac.jp/~cces/index.htm
E-mail: cces@econ.hit-u.ac.jp

Model Selection Criteria for the Leads-and-Lags Cointegrating Regression*

In Choi† and Eiji Kurozumi‡

First Draft: April, 2008; Revised: August, 2008

Abstract

In this paper, Mallows' (1973) C_p criterion, Akaike's (1973) AIC, Hurvich and Tsai's (1989) corrected AIC and the BIC of Akaike (1978) and Schwarz (1978) are derived for the leads-and-lags cointegrating regression. Deriving model selection criteria for the leads-and-lags regression is a nontrivial task since the true model is of infinite dimension. This paper justifies using the conventional formulas of those model selection criteria for the leads-and-lags cointegrating regression. The numbers of leads and lags can be selected in scientific ways using the model selection criteria. Simulation results regarding the bias and mean squared error of the long-run coefficient estimates are reported. It is found that the model selection criteria are successful in reducing bias and mean squared error relative to the conventional, fixed selection rules. Among the model selection criteria, the BIC appears to be most successful in reducing MSE, and C_p in reducing bias. We also observe that, in most cases, the selection rules without the restriction that the numbers of the leads and lags be the same have an advantage over those with it.

Keywords: Cointegration, Leads-and-lags regression, AIC, Corrected AIC, BIC, C_p

1 Introduction

Several methods have been proposed for efficient estimation of cointegrating relations. Phillips and Hansen (1990) and Park (1992) use semiparametric approaches to derive

*This paper was presented in a conference honoring the 60th birthday of Professor Peter C. B. Phillips that was organized by Robert Mariano, Zhijie Xiao and Jun Yu and held in Singapore in July, 2008. The authors are thankful to the seminar participants for their valuable comments. Kurozumi's research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology under Grants-in-Aid No. 18730142.

†Corresponding author. Department of Economics, Sogang University, #1 Shinsu-dong, Mapo-gu, Seoul, 121-742 Korea. E-mail: inchoi@gmail.com

‡Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo, Japan.

efficient estimators that have a mixture normal distribution in the limit. Saikkonen (1991), Phillips and Loretan (1991), and Stock and Watson (1993) use the regression augmented with the leads and lags of the regressors' first differences, yielding estimators as efficient as those based on the semiparametric approach. This is called leads-and-lags regression or dynamic OLS regression. This method was also used for cointegrating smooth-transition regression by Saikkonen and Choi (2004). Johansen (1988) uses vector autoregression to derive the maximum likelihood estimator of cointegrating spaces under the assumption of a normal distribution. In addition, Pesaran and Shin (1999) use an autoregressive distributed modelling approach for the inference on cointegrating vectors. Finite-sample properties of the aforementioned methods are studied by Hargreaves (1994) and Panopoulou and Pittis (2004), among others.

This paper focuses on the leads-and-lags regression among the methods mentioned above. Though it has been used intensively in empirical applications[1] due to its optimal property and simplicity, the practical question of how to select the numbers of the leads and lags has not been resolved yet. Most empirical studies use arbitrary numbers of leads and lags and empirical results can differ depending on this choice. Furthermore, a restriction that the numbers of leads and lags be the same is often imposed out of convenience. This situation is certainly undesirable from empirical viewpoints and indicates a need for methods that select the numbers of leads and lags in nonarbitrary ways.

The main purpose of this paper is to propose methods for the selection of the numbers of leads and lags in cointegrating regressions. More specifically, we will derive model selection criteria for the leads-and-lags regression so that the numbers of leads and lags can be chosen scientifically. These will make the leads-and-lags

[1] Examples are Ball (2001) for the money-demand equation; Bentzen (2004) for the rebound effect in energy consumption; Caballero (1994) for the elasticity of the U.S. capital-output ratio to the cost of capital; Hai, Mark, and Wu (1997) for spot and forward exchange rate regressions; Hussein (1998) for the Feldstein–Horioka puzzle; Masih and Masih (1996) for elasticity estimates of coal demand for China; Weber (1995) for estimates of Okun's coefficient; Wu, Fountas, and Chen (1996) for current account deficits; and Zivot (2000) for the forward rate unbaisedness hypothesis.

regression more useful for empirical applications. Deriving model selection criteria for the leads-and-lags regression is a nontrivial task since the true model is of infinite dimension. Most model selection criteria in time series analysis are derived assuming that the true model is contained in a set of candidate models. The only exception that we are aware of is Hurvich and Tsai (1991), which considers a bias-corrected Akaike information criterion (AIC) for the infinite-order autoregressive model using a frequency-domain approach for the approximation of the variance–covariance matrices for the true and approximating models. Notably, the resulting formula from this study is different from that for the finite-order autoregressive model (cf. Hurvich and Tsai, 1989). By contrast, we will show that all the model selection criteria that we derive for the leads-and-lags regression are the same as those for the case of a finite-dimensional true model.

In this paper, we will consider four model selection criteria: Mallows' (1973) C_p criterion, Akaike's (1973) AIC, Hurvich and Tsai's (1989) corrected AIC, and the Bayesian information criterion (BIC) of Akaike (1978) and Schwarz (1978). These methods are the most common in practice,[2] though there are many other methods available as documented by Rao and Yu (2001).[3]

We will also report extensive simulation results that compare the model selection criteria using bias and mean squared error (MSE) of the long-run coefficient estimate as benchmarks. The simulation results show that the model selection criteria are successful in reducing bias and MSE relative the fixed selection rules. the BIC appears to be most successful in reducing MSE, and C_p in reducing bias. We also observe that, in most cases, the selection rules without the restriction that the numbers of the leads and lags be the same have an advantage over those with it.

[2] The Google citation numbers for these articles are 1,060, 4,546, 741, 205 and 5,720, respectively, as of August 23, 2008.

[3] Unfortunately, this review article focuses on the statistics literature only and neglects contributions to the subject of model selection that appeared in the econometrics literature. Some of the important works neglected there are Phillips and Ploberger (1994, 1996). However, since the Phillips–Ploberger criterion assumes that the true model is contained in a set of candidate models (see Section 3 of Phillips and Ploberger, 1994), it is inapplicable to our problem.

A recent paper related to the current one is Kejriwal and Perron (2008). Since such selection rules as the AIC and BIC yield logarithmic rates of increase of the chosen numbers of leads and lags, they do not satisfy the upper bound condition for leads-and-lags regression (condition (9) in Section 2). However, Kejriwal and Perron (2008) show that the condition can be weakened without bringing changes to the asymptotic mixture normality of the long-run coefficient estimates. The weakened condition is satisfied by the AIC and BIC so that their use is justified in practice. Kejriwal and Perron (2008) use the AIC and BIC without considering their appropriateness for the leads-and-lags regression. This paper establishes rigorously that using them and others is proper from the viewpoint of model selection.

This paper is organized as follows. Section 2 briefly explains cointegrating leads-and-lags regression. Section 3 derives model selection criteria for leads-and-lags regression. Section 4 reports simulation results that compare the performance of the model selection criteria in finite samples. Section 5 summarizes and concludes. All the proofs are contained in appendices.

A few words on our notation. Weak convergence is denoted by $\Rightarrow$ and all limits are taken as $T \to \infty$. The largest integer not exceeding x is denoted by $[x]$. For an arbitrary matrix A, $\|A\| = [tr(A'A)]^{1/2}$ and $\|A\|_1 = \sup\{\|Ax\| : \|x\| \leq 1\}$. When applied to matrices, the inequality signs $>$ and $\geq$ mean the usual ordering of positive definite and semidefinite matrices, respectively. Last, for a matrix A, $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$.

2 Leads-and-lags regression

This section briefly introduces the leads-and-lags regression of Saikkonen (1991)[4] and some required assumptions. Consider the cointegrating regression model

$$y_t = \mu + \beta' x_t + u_t, \quad (t = 1, 2, \ldots, T), \tag{1}$$

where x_t ($p \times 1$) is an $I(1)$ regressor vector and u_t a zero-mean stationary error term. The main purpose of the leads-and-lags regression is to estimate the cointegrating

[4] See also Phillips and Loretan (1991) and Stock and Watson (1993).

vector β efficiently such that it has a mixture normal distribution in the limit. Leads and lags will augment the regression model (1) for this purpose.

For the regressors and error terms, we assume that $w_t = (\Delta x_t' \ u_t)' = (v_t' \ u_t)'$ satisfy conditions for the multivariate invariance principle such that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} w_t \Rightarrow B(r), \ r \in (0, 1], \tag{2}$$

where $B(r)$ is a vector Brownian motion with a positive-definite variance–covariance matrix $\Omega = \begin{array}{cc} \begin{bmatrix} \Omega_{vv} & \omega_{vu} \\ \omega_{uv} & \omega_{uu} \end{bmatrix} & \begin{array}{c} p \\ 1 \end{array} \end{array}$. More primitive conditions for this are available in the literature (cf., e.g., Phillips and Durlauf, 1986).

Furthermore, the summability condition

$$\sum_{j=-\infty}^{\infty} \left\| E\left(w_t w_{t+j}'\right)\right\| < \infty \tag{3}$$

needs to be satisfied. This implies that the process w_t has a continuous spectral density matrix $f_{ww}(\lambda)$, which we assume to satisfy

$$f_{ww}(\lambda) \geq \varepsilon I_{p+1}, \quad \varepsilon > 0. \tag{4}$$

This assumption means that the spectral density matrix $f_{ww}(\lambda)$ is bounded away from zero. Last, denoting the fourth-order cumulants of w_t as κ_{ijkl}, we also require a technical assumption:

$$\sum \sum_{m_1, m_2, m_3 = -\infty}^{\infty} \sum |\kappa_{ijkl}(m_1, m_2, m_3)| < \infty.$$

Note that all of these assumptions are taken from Saikkonen (1991).

Under conditions (3) and (4), the error term u_t can be expressed as

$$u_t = \sum_{j=-\infty}^{\infty} \pi_j' v_{t-j} + e_t, \tag{5}$$

where e_t is a zero-mean stationary process such that $E e_t v_{t-j}' = 0$ for all $j = 0, \pm 1, \ldots$, and

$$\sum_{j=-\infty}^{\infty} \|\pi_j\| < \infty.$$

As is well known, the long-run variance of the process e_t can be expressed as $\omega_e^2 = \omega_{uu} - \omega_{uv}\Omega_{vv}^{-1}\omega_{vu}$.

Using equation (5), we can write model (1) as

$$y_t = \mu + \beta' x_t + \sum_{j=-\infty}^{\infty} \pi_j' \Delta x_{t-j} + e_t, \tag{6}$$

where Δ signifies the difference operator. Truncating the infinite sum in model (6) at K_U and K_L, we obtain

$$\begin{aligned} y_t &= \mu + \beta' x_t + \sum_{j=-K_L}^{K_U} \pi_j' \Delta x_{t-j} + e_{Kt}, \quad (t = K_L + 2, K_L + 3, \ldots, T - K_U), \\ &= \theta_K' z_{Kt} + e_{Kt}, \end{aligned} \tag{7}$$

where $\theta_K = [\mu \ \beta' \ \pi'_{-K_L}, \ldots, \pi'_{K_U}]'$, $z_{Kt} = [1, x_t', \Delta x'_{t+K_L}, \ldots, \Delta x'_{t-K_U}]'$ and

$$e_{Kt} = e_t + \sum_{j>K_U,\ j<-K_L} \pi_j' v_{t-j} = e_t + V_{t,K}.$$

In regression model (7), leads and lags are used as additional regressors. Saikkonen (1991) uses a common value for K_U and K_L for simplicity, but the results he obtained apply to the current case with some minor changes in notation.

The numbers of leads and lags should be large enough to make the effect of truncation negligible, but should not be too large because this will bring inefficiency in estimating the coefficient vector β. Conditions on K_U and K_L that provide asymptotic mixture normality of the OLS estimator of β and asymptotic normality of the OLS estimator of $[\pi'_{-K_L}, \ldots, \pi'_{K_U}]'$ are

$$K_U^3/T,\ K_L^3/T \to 0 \tag{8}$$

and

$$\sqrt{T} \sum_{j>K_U,\ j<-K_L} \|\pi_j\| \to 0. \tag{9}$$

In fact, Saikkonen (1991) did not derive asymptotic normality of the OLS estimator of $[\pi'_{-K_L}, \ldots, \pi'_{K_U}]'$, but this can be done using the same methods as for Theorem 4 of Lewis and Reinsel (1985) (see also Berk, 1974). Conditions (8) and (9) are sufficient to derive the asymptotic distributions of the OLS estimators for model (7), but do not provide practical guidance in selecting K_U and K_L in finite samples. The next section will consider various methods for their optimal selection.

3 Methods for selecting K_U and K_L

This section considers various procedures for selecting K_U and K_L. These are basically model selection procedures that have often been used for regressions and time-series analysis. For the derivations of these procedures, we assume that the conditions of the leads-and-lags regression in Section 2 hold.

For later use, write model (7) in obvious matrix notation as $y = Z_K\theta_K + e_K$. The OLS estimator of the parameter vector θ_K using model (7) is denoted by $\hat{\theta}_K$. In addition, model (6) is written in vector notation as $y = \tau + e$, where $e = [e_{K_L+2}, \ldots, e_{T-K_U}]'$ is the vector of errors.

3.1 C_p criterion

The C_p criterion of Mallows (1973) is an estimator of the expected squared sum of forecast errors. Assume that $E(e_t \mid Z_K) = 0$ for any K_U and K_L. Then the forecast error for the C_p criterion is defined by $f_t = \hat{y}_t - E(y_t \mid Z_K) = \hat{\theta}_K' z_{Kt} - \theta_K' z_{Kt} - E(V_{t,K} \mid Z_K)$. This measures the distance between the fitted value $\hat{y}_t$ and the conditional expectation of y_t. The expected squared sum of the forecast errors standardized by σ_e^2 $(= E(e_t^2))$ is

$$
\begin{aligned}
\Delta_K &= \frac{1}{\sigma_e^2}\sum_{t=K_L+2}^{T-K_U} E\left(f_t^2 \mid Z_K\right) \\
&= \frac{1}{\sigma_e^2} E\left[(\hat{\theta}_K - \theta)' Z_K' Z_K (\hat{\theta}_K - \theta) \mid Z_K\right] + \frac{1}{\sigma_e^2}\sum_{t=K_L+2}^{T-K_U} V_{t,K}^2 \\
&\quad - 2\frac{1}{\sigma_e^2}\sum_{t=K_L+2}^{T-K_U} E\left[\left(\hat{\theta}_K - \theta_K\right)' z_{Kt} E\left[V_{t,K} \mid Z_K\right] \mid Z_K\right] \\
&= A + B - C, \text{ say.}
\end{aligned}
$$

Since $\frac{1}{\sigma_e^2}(\hat{\theta}_K - \theta)' Z_K' Z_K (\hat{\theta}_K - \theta)_{|Z_K} \Rightarrow \chi^2(p(K_L + K_U + 2) + 1)$ (cf. Theorem 4.1 of Saikkonen, 1991, and Theorem 4 of Lewis and Reinsel, 1985), A is approximated by $p(K_L + K_U + 2) + 1$. Using the relation $E(e_{Kt}^2) = \sigma_e^2 + E(V_{t,K}^2)$, we approximate B by $\frac{1}{\sigma_e^2}\sum_{t=K_L+2}^{T-K_U} \hat{e}_{t,K}^2 - (T - K_U - K_L - 1)$. The third term C is approximated by zero since $\hat{\theta}_K - \theta_K$ converges to zero in probability. Let $K_{U,\max}$ and $K_{L,\max}$ be the maximum values used for selecting K_U and K_L, respectively. We estimate σ_e^2 by $\hat{\sigma}_e^2 = \frac{1}{T-K_{U,\max}-K_{L,\max}-1}\sum_{t=K_{L,\max}+2}^{T-K_{U,\max}} \hat{e}_{t,K_{\max}}^2$, where $\hat{e}_{t,K\max}$ denotes the regression residual using $K_{U,\max}$ and $K_{L,\max}$ for K_U and K_L, respectively. Then, using the aforementioned approximations, the C_p criterion that approximates Δ_K is defined by

$$C_p = \frac{1}{\hat{\sigma}_e^2}\sum_{t=K_L+2}^{T-K_U} \hat{e}_{t,K}^2 + (p+1)(K_L + K_U + 2) - T.$$

In practice, we choose K_U and K_L so that the C_p criterion is minimized. Note that this requires preselecting $K_{U,\max}$ and $K_{L,\max}$.

3.2 Akaike information criterion

The AIC of Akaike (1973) is an estimator of the expected Kullback–Leibler information measure and is often used for regression and time-series models. In deriving the AIC and its variants, it is usually assumed[5] that candidate models include the true model (cf. Akaike, 1973; Hurvich and Tsai, 1989), though it does not have to be so because the Kullback–Leibler information measure simply indicates how far apart the true and any candidate models are.

Since the true model of the current study (equation (6)) involves an infinite number of parameters, candidate models cannot include the true model. However, it is still possible to derive the AIC using a general formula.[6] Assume that the error

[5] A notable exception is Hurvich and Tsai (1991), which considers a bias-corrected AIC for the autoregressive model of infinite order.

[6] The general formula assumes $\sqrt{T}$ asymptotics, but it can straightforwardly be extended to the current case because the leads-and-lags coefficient estimators for the nonstationary regressors have a mixture-normal distribution in the limit. The general formula is related to Takeuchi's (1976) information criterion. See Chapter 7, Section 2 of Burnham and Anderson (2002) for further details on this.

terms of the candidate model (7) follow an iid normal distribution with mean 0 and variance σ^2_{eK} conditional on Z_K, and denote the conditional log-likelihood function of the candidate model by $l(\theta_K, \sigma^2_{eK}) = l(\delta_K)$. Then, letting $n = T - K_U - K_L - 1$, the general formula can be written as

$$n \ln \left(\frac{\sum_{t=K_L+2}^{T-K_U} \hat{e}^2_{t,K}}{n} \right) + 2tr\left[J(\delta_o) I(\delta_o)^{-1} \right],$$

where $\delta_o = [\theta'_o \ \sigma^2_o]'$ minimizes the Kullback–Leibler information measure, $J(\delta_o) = E\left[\frac{\partial l(\delta_K)}{\partial \delta_K} \frac{\partial l(\delta_K)}{\partial \delta_K}_{|\delta_K=\delta_o} \right]$ and $I(\delta_o) = E\left[-\frac{\partial^2 l(\delta_K)}{\partial \delta_K \partial \delta'_K}_{|\delta_K=\delta_o} \right]$. Since $tr\left[J(\delta_o) I(\delta_o)^{-1} \right] = \frac{\sigma^2_e}{\sigma^2_o} \left(p(K_U + K_L + 2) + 1 + \frac{2\sigma^2_o - \sigma^2_e}{\sigma^2_o} \right)$ as shown in Appendix I and $\sigma^2_o \xrightarrow{p} \sigma^2_e$ as shown in Lemma A.2 in Appendix II, the AIC is defined by

$$\text{AIC} = n \ln \left(\frac{\sum_{t=K_L+2}^{T-K_U} \hat{e}^2_{t,K}}{n} \right) + 2\left(p(K_U + K_L + 2) + 2 \right).$$

Notably, this is the same as the usual AIC that assumes that candidate models include the true model. However, notice that the sample size n depends on the chosen model unlike in conventional regression models.

3.3 Corrected Akaike information criterion

Hurvich and Tsai's (1989) corrected AIC is also an estimator of the Kullback–Leibler information measure. First, it calculates the Kullback–Leibler information measure using unknown parameter values. Next, the unknown parameter values are replaced with the maximum likelihood estimators. These steps provide the corrected AIC.

In our application, letting $E_F(\cdot)$ be an expectation operator using the true model (6) and assuming $E_F(e^2_t \mid Z_\infty) = \sigma^2_e$ where Z_∞ denotes $\{z_{Kt}\}_{t=-\infty}^{\infty}$, the Kullback–Leibler information measure using unknown parameter values is

$$\begin{aligned} -2E_F(l(\delta_K)) &= E_F\left[n \ln(\sigma^2_{eK}) + \frac{(y - Z_K\theta_K)'(y - Z_K\theta_K)}{\sigma^2_{eK}} \mid Z_\infty \right] \\ &= E_F\left(n \ln(\sigma^2_{eK}) \mid Z_\infty \right) + \frac{(\tau - Z_K\theta_K)'(\tau - Z_K\theta_K)}{\sigma^2_{eK}} + \frac{n\sigma^2_e}{\sigma^2_{eK}} \\ &= f(\theta_K, \sigma^2_{eK}), \text{ say.} \end{aligned}$$

Replacing θ_K and σ^2_{eK} with corresponding maximum likelihood estimators, we obtain

$$f(\hat{\theta}_K, \hat{\sigma}^2_{eK}) = E_F\left(n\ln(\hat{\sigma}^2_{eK})\right) + \frac{(\tau - Z_K\hat{\theta}_K)'(\tau - Z_K\hat{\theta}_K)}{\hat{\sigma}^2_{eK}} + \frac{n\sigma^2_e}{\hat{\sigma}^2_{eK}}, \tag{10}$$

where $\hat{\sigma}^2_{eK} = \frac{\sum_{t=K_L+2}^{T-K_U}\hat{e}^2_{t,K}}{n}$. The corrected AIC is an approximation to $f(\hat{\theta}_K, \hat{\sigma}^2_{eK})$.

As shown in Lemma A.3 in Appendix II, the second and third terms in relation (10) follow $\frac{nm}{n-m}F(m, n-m)$ and $\frac{n^2}{\chi^2(n-m)}$, respectively, under a normality assumption when T is large. Thus, using the mean values of the second and third terms in relation (10), we approximate $f(\hat{\theta}_K, \hat{\sigma}^2_{eK})$ by

$$\mathrm{AIC}_C = n\ln\left(\frac{\sum_{t=K_L+2}^{T-K_U}\hat{e}^2_{t,K}}{n}\right) + \frac{nm}{n-m-2} + \frac{n^2}{n-m-2}.$$

This is the corrected AIC. Notice that this is exactly the same as the corrected AIC of Hurvich and Tsai (1989), which assumes that candidate models include the true model. By contrast, Hurvich and Tsai's (1991) corrected AIC derived for the AR(∞) model is different from that of the current work.

Comparison of the AIC and the corrected AIC reveals that the major difference between them lies at which stage the maximum likelihood estimators are plugged into the Kullback–Leibler information measure. The AIC calculates the Kullback–Leibler information measure using the maximum likelihood estimators from the beginning, but the corrected AIC calculates the Kullback–Leibler information measure using unknown parameter values and then uses the maximum likelihood estimators for the computed Kullback–Leibler information measure.

3.4 Bayesian information criterion

The BIC of Akaike (1978) and Schwarz (1978) is an approximation to a transformation of the Bayesian posterior probability of a candidate model. Unlike the AIC, it does not require the probability density of the true model. Therefore, the true model of infinite dimension as in this paper does not require any separate treatment and the usual formula,

$$\mathrm{BIC} = n\ln\left(\frac{\sum_{t=K_L+2}^{T-K_U}\hat{e}^2_{t,K}}{n}\right) + (p(K_U + K_L + 2) + 2)\ln(n),$$

can be used. Because we never use the true model in the leads-and-lags regression, consistency of the BIC cannot be an issue.

4 Simulation results

The ultimate purpose of the leads-and-lags regression is to estimate the long-run coefficient β precisely. This section investigates how model selection criteria of the previous section perform in relation to the estimation of the long-run coefficient β. To this end, we consider the data generating process

$$y_t = \mu + \beta x_t + u_t, \quad x_t = x_{t-1} + v_t,$$

where x_t is a scalar unit root process with $x_0 = 0$. We set $\mu = 1$ and $\beta = 1$ throughout the simulations and set the sample size at 100 or 300. The error term $w_t = (v_t, u_t)'$ is generated from a VARMA(1,1) process:

$$w_t = Aw_{t-1} + \varepsilon_t - \Theta\varepsilon_{t-1},$$

where

$$\begin{aligned}
A &= \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}, \qquad \Theta = \begin{bmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{bmatrix}, \\
\varepsilon_t &= \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim i.i.d.(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix}, \\
w_0 &= \varepsilon_0 = 0.
\end{aligned}$$

The parameters a_{ii} and θ_{ii} $(i = 1, 2)$ are related to the strength of the serial correlation in w_t, whereas σ_{12} signifies contemporaneous correlation. In the simulations, the parameters a_{ii} and θ_{ii} take values from $\{0, 0.4, 0.8\}$ while σ_{12} is equal to either 0.4 or 0.8. We consider two distributions for ε_t: a standard normal distribution and a log-normal distribution. Since AIC$_C$ is based on the assumption that the error terms are normally distributed, the nonnormal distribution of ε_t implies that the correction for the AIC$_C$ does not make much sense. In order to check the robustness of AIC$_C$ to nonnormal error terms, we try a log-normal distribution for the error terms. We also note that all the selection criteria depend on the maximum numbers of the leads and lags. We will use $K_4 = [4 \times (T/100)^{1/4}]$ and $K_{12} = [12 \times (T/100)^{1/4}]$ commonly for $K_{U,\max}$ and $K_{L,\max}$ in the simulations. All computations are carried out using the GAUSS matrix language with 10,000 replications.

Before reporting our simulation results, we note that the leads of $\Delta x_t = v_t$ do not have to be included in the augmented regression model (7) when $a_{11} = \theta_{11} = 0$. In this case, v_t becomes equal to ε_{1t} and

$$\begin{aligned} u_t &= (1 - a_{22}L)^{-1}(1 - \theta_{11}L)\varepsilon_{2t} \\ &= \phi(L)\sigma_{12}v_t + \phi(L)\varepsilon_{2\cdot 1,t}, \end{aligned} \tag{11}$$

where $\phi(L) = (1 - a_{22}L)^{-1}(1 - \theta_{11}L)$ with L being the lag operator and $\varepsilon_{2\cdot 1,t} = \varepsilon_{2t} - \sigma_{12}\varepsilon_{1t}$. Since the first term of (11) includes $v_{t-j} = \Delta x_{t-j}$ only for $j \geq 0$, while $\varepsilon_{2\cdot 1,t}$ is independent of $v_s = \varepsilon_{1s}$ for all s and t, we can see that u_t can be expressed as in equation (5) without the leads of Δx_t.

Tables 1–4 present empirical bias of the estimate of β while Tables 5–8 report empirical MSE. Part I of each table deals with the case where the leads of Δx_t are not required (i.e., $a_{11} = \theta_{11} = 0$). Part II handles the case where either $a_{11} \neq 0$ or $\theta_{11} \neq 0$ or both. Note that a different scale is used in each table depending on the sample size and the distribution of ε_t. For the purpose of comparison, we also select the leads and lags by the fixed rules $K_U = K_L = 1$, 2, 3, K_4 and K_{12}. Although it is often the case that K_U is conventionally set equal to K_L in practice, we do not have to use the same numbers for the leads and lags. Especially, when $a_{11} = \theta_{11} = 0$, we do not have to include the leads and then the selection rules without the restriction of $K_U = K_L$ are expected to have an advantage over those with $K_U = K_L$.

We first summarize the simulation results regarding the bias.

(b-i) C_p without the restriction of $K_U = K_L$ tends to be most successful in reducing bias. Especially when there are high serial correlations in the data (see the last two columns in each table), C_p shows much better performance than the other selection criteria and the fixed selection rules.

(b-ii) When $a_{11} = \theta_{11} = 0$, the absolute value of the bias without the restriction of $K_U = K_L$ is smaller than that with $K_U = K_L$ in most cases. This is well expected because the leads of Δx_t are not required in this case for the augmented regression (7).

(b-iii) When $a_{11} = \theta_{11} = 0$, the bias resulting from the use of C_p is smallest whereas the BIC leads to the most biased estimates among the four selection criteria.

(b-iv) The AIC tends to perform slightly better than the AIC$_c$, especially when $a_{11} = \theta_{11} = 0$. But the differences are marginal.

(b-v) There are a few cases where the model selection criteria with the restriction $K_U = K_L$ result in the smaller bias, but the differences between the restricted and unrestricted cases are relatively small. In most cases, the model selection criteria without the restriction $K_U = K_L$ perform better than those with it.

(b-vi) The fixed selection rules sometimes result in large biases. In most cases, they are dominated by one of the model selection criteria.

(b-vii) Overall, the bias with $K_{\max} = K_{12}$ tends to be smaller than that with $K_{\max} = K_4$.

(b-viii) The log-normal distribution does not bring any noticeable changes in evaluating the selection rules. In particular, performance of the AIC$_c$ does not change much with the log-normal distribution.

(b-ix) As sample size grows, bias decreases as expected, but qualitative differences are not observed with the increasing sample sizes in evaluating the selection rules.

Regarding to the MSE, the simulation results are summarized as follows.

(m-i) The BIC tends to perform best in almost all the cases, and the AIC$_c$ tends to follow. The MSE with C_p tends to become largest in most cases. Overall, however, the differences of the MSEs are relatively small among the four selection criteria.

(m-ii) In most cases, the MSE without the restriction of $K_U = K_L$ is smaller than that with $K_U = K_L$ though there are exceptions, especially in Part II of the tables, but the differences between the restricted and unrestricted cases are quite small.

(m-iii) The AIC$_c$ tends perform slightly better than the AIC.

(m-iv) In most cases, the fixed selection rules are dominated by one of the model selection criteria.

(m-v) Overall, the MSE with $K_{\max} = K_{12}$ tends to be larger than that with $K_{\max} = K_4$.

(m-vi) The log-normal distribution does not bring any noticeable changes in evaluating the selection rules. In particular, performance of the AIC$_c$ does not change much with the log-normal distribution.

(m-vii) As sample size grows, MSE decreases as expected, but qualitative differences are not observed with the increasing sample sizes in evaluating the selection rules.

We infer from the simulation results that the model selection criteria are successful in reducing bias and MSE relative the fixed selection rules. The BIC appears to be most successful in reducing MSE, and C_p in reducing bias. We also observe that the selection rules without the restriction $K_L = K_U$ have an advantage over those with $K_U = K_L$ in most cases. For practitioners, therefore, we recommend using the selection rules of this paper without the restriction of $K_U = K_L$.

5 Conclusion and further remarks

We have derived Mallows' (1973) C_p criterion, Akaike's (1973) AIC, Hurvich and Tsai's (1989) corrected AIC, and the BIC of Akaike (1978) and Schwarz (1978) for the leads-and-lags cointegrating regression. Our results justify using conventional formulas of those model selection criteria for the leads-and-lags cointegrating regression. These model selection criteria allow us to choose the numbers of leads and lags in scientific ways. Simulation results regarding the bias and mean squared error of the long-run coefficient estimates are also reported. The model selection criteria are shown to be successful in reducing bias and mean squared error relative to the conventional, fixed selection rules. Among them, the BIC appears to be most successful in reducing MSE, and C_p in reducing bias. We also observe that the selection rules without the restriction that the numbers of the leads and lags be the same have an advantage over those with it in most cases. The model selection criteria in this paper were derived for linear regression, but we note that they can also be used for the nonlinear leads-and-lags regression of Saikkonen and Choi (2004).

The ultimate purpose of the model selection criteria for the leads-and-lags cointegrating regression is to estimate the long-run slope coefficient efficiently. Though it was shown through simulations that they improve on the fixed selection rules in terms of bias and mean squared error, a better rule that directly minimizes the mean squared error (or other efficiency measures) of the long-run slope coefficient estimate may exist. How this rule, if it exists, and the model selection criteria of this paper are related is a question one may be interested in investigating.

References

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle. In: B. N. Petrov and F. Csáki (Eds.), 2nd International Symposium on Information Theory, Akadémiai Kiadó, Budapest, pp. 267–281.

Akaike, H., 1978, A Bayesian analysis of the minimum AIC procedure. Annals of the Institute of Statistical Mathematics, 30, 9–14.

Ball, L., 2001, Another look at long-run money demand. Journal of Monetary Economics, 47, 31–44.

Bentzen, J., 2004, Estimating the rebound effect in U.S. manufacturing energy consumption. Energy Economics, 26, 123–134.

Berk, K. N., 1974, Consistent autoregressive spectral estimates. The Annals of Statistics, 2, 489–502.

Burnham, K. P., D. R. Anderson, 2002, Model selection and multi-model inference: A practical information-theoretic approach, 2nd ed. Springer-Verlag.

Caballero, R. J., 1994, Small sample bias and adjustment costs. The Review of Economics and Statistics, 76, 52–58.

Hai, W., Nelson C. Mark, Wu, Y., 1997, Understanding spot and forward exchange rate regressions. Journal of Applied Econometrics, 12, 715–734.

Hargreaves, C. P., 1994, A review of methods of estimating cointegrating relationships. In: C. P. Hargreaves (Ed.), Nonstationary time series analysis and cointegration, Oxford University Press, pp. 87–132.

Hurvich, C. M., Tsai, C., 1989, Regression and time series model selection in small samples. Biometrika, 76, 297–307.

Hurvich, C. M., Tsai, C., 1991, Bias of the corrected AIC criterion for underfitted regression and time series models. Biometrika, 78, 499–509.

Hussein, K. A., 1998, International capital mobility in OECD countries: The Feldstein–Horioka 'puzzle' revisited. Economics Letters, 59, 237–242.

Johansen, S., 1988, Statistical anaysis of cointgrating vectors. Journal of Economic Dynamics and Control, 12, 231-254.

Kejriwal, M., Perron, P., 2008, Data dependent rule for selection of the number of leads and lags in the dynamic OLS cointegrating regression. Econometric Theory, 24, 1425–1441.

Lewis, R., Reinsel G. C., 1985, Prediction of multivariate time series by autoregressive model fitting. Journal of Multivariate Analysis, 16, 393–411.

Mallows, C. L., 1973, Some comments on C_p, Technometrics, 15, 661-675.

Masih, R., Masih, A. M. M., 1996, Stock–Watson dynamic OLS (DOLS) and error-correction modelling approaches to estimating long- and short-run elasticities in a demand function: New evidence and methodological implications from an application to the demand for coal in mainland china. Energy Economics, 18, 315–334.

Panopoulou, E., Pittis, N., 2004, A comparison of autoregressive distributed lag and dynamic OLS cointegration estimators in the case of a serially correlated cointegration error. Econometrics Journal, 7, 545–617.

Park, J. Y. 1992, Canonical cointegrating regressions. Econometrica, 60, 119–143.

Pesaran, H. M., Shin, Y., 1999, An autoregressive distributed lag modelling approach to cointegration analysis. In: S. Strom (Ed.), Econometrics and economic theory in the twentieth century: The Ragnar Frisch centennial symposium, Cambridge University Press, Cambridge, UK, pp.371-413.

Phillips, P. C. B., Durlauf, S. N., 1986, Multiple time series regression with integrated processes. The Review of Economic Studies, 53, 473–495.

Phillips, P. C. B., Hansen, B. E., 1990, Statistical inference in instrumental variables regression with I(1) processes. The Review of Economic Studies, 57, 99–125.

Phillips, P. C. B., Loretan, M., 1991, Estimating long-run economic equilibria. The Review of Economic Studies, 58, 407–436.

Phillips, P. C. B., Ploberger, W., 1994, Posterior odds testing for a unit root with data-based model selection. Econometric Theory, 10, 774–808.

Phillips, P. C. B., Ploberger, W., 1996, An asymptotic theory of Bayesian inference for time series. Econometrica, 64, 381–412.

Rao, C. R., Wu, Y., 2001, On model selection. In: P. Lahiri (Ed.), Model selection, Institute of Mathematical Statistics, Beachwood, Ohio, pp. 1-57.

Saikkonen, P., 1991, Asymptotically efficient estimation of cointegration regressions. Econometric Theory, 7, 1–21.

Saikkonen, P., Choi, I., 2004, Cointegrating smooth transition regressions. Econometric Theory, 20, 301–340.

Schwarz, G., 1978, Estimating the dimension of a model. The Annals of Statistics, 6, 461–464.

Stock, J. H., Watson, M. W., 1993, A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica, 61, 783–820.

Takeuchi, K., 1976, Distribution of informational statistics and a criterion of model fitting. Suri-Kagaku (Mathematic Sciences), 153, 12–18.

Weber, C. E., 1995, Cyclical output, cyclical unemployment, and Okun's coefficient: A new approach. Journal of Applied Econometrics, 10, 433–445.

Wu, J., Fountas, S., Chen, S., 1996, Testing for the sustainability of the current account deficit in two industrial countries. Economics Letters, 52, 193-198.

Zivot, E., 2000, Cointegration and forward and spot exchange rate regressions. Journal of International Money and Finance, 19, 785–812.

Appendix I: Calculation of $tr\left\{J(\delta_o)I(\delta_o)^{-1}\right\}$

Ignoring a constant, the Kullback–Leibler information measure is written as

$$\begin{aligned}-E_F(l(\delta_K)) &= E_F\left[\frac{n}{2}\ln(\sigma_{eK}^2)+\frac{(y-Z_K\theta_K)'(y-Z_K\theta_K)}{2\sigma_{eK}^2}\mid Z_\infty\right]\\ &= \frac{n}{2}\ln(\sigma_{eK}^2)+\frac{(\tau-Z_K\theta_K)'(\tau-Z_K\theta_K)}{2\sigma_{eK}^2}+\frac{n\sigma_e^2}{2\sigma_{eK}^2}.\end{aligned}$$

This is minimized by $\theta_o=(Z_K'Z_K)^{-1}Z_K'\tau$ and $\sigma_o^2=\frac{(\tau-Z_K\theta_o)'(\tau-Z_K\theta_o)}{n}+\sigma_e^2$.

Since $\frac{\partial l(\delta_K)}{\partial\delta_K}=\begin{bmatrix}\frac{1}{\sigma_{eK}^2}(Z_K'y-Z_K'Z_K\theta_K)\\ -\frac{n}{2\sigma_{eK}^2}+\frac{1}{2\sigma_{eK}^4}(y-Z_K\theta_K)'(y-Z_K\theta_K)\end{bmatrix}$, $Z_K'y-Z_K'Z_K\theta_o=Z_K'e$, and $y-Z_K\theta_o=M_{Z_K}\tau+e$, we may write

$$\frac{\partial l(\delta_K)}{\partial\delta_K}\frac{\partial l(\delta_K)}{\partial\delta_K'}\mid_{\theta=\theta_o}=\begin{bmatrix}A_{11} & A_{12}\\ A_{12}' & A_{22}\end{bmatrix},$$

where $A_{11}=\frac{1}{\sigma_o^4}Z_K'ee'Z_K,\ A_{12}=\frac{1}{\sigma_o^2}Z_K'e\times\left(-\frac{n}{2\sigma_o^2}+\frac{1}{2\sigma_o^4}(\tau'M_{Z_K}\tau+e'e+2e'M_{Z_K}\tau)\right)$ and $A_{22}=\left(-\frac{n}{2\sigma_o^2}+\frac{1}{2\sigma_o^4}\left(\tau'M_{Z_K}\tau+e'e+2e'M_{Z_K}\tau\right)\right)^2$. Thus,

$$J(\delta_o)=\begin{bmatrix}\frac{\sigma_e^2}{\sigma_o^4}Z_K'Z_K & 0\\ 0 & \frac{n\sigma_e^4}{2\sigma_o^8}+\frac{\sigma_e^2}{\sigma_o^8}\tau'M_{Z_K}\tau\end{bmatrix}. \tag{A.1}$$

Moreover, using

$$\frac{\partial^2 l(\delta_K)}{\partial\theta\partial\theta'}=\begin{bmatrix}-\frac{1}{\sigma_{eK}^2}Z_K'Z_K & -\frac{1}{\sigma_{eK}^4}(Z_K'y-Z_K'Z_K\theta_K)\\ -\frac{1}{\sigma_{eK}^4}(y'Z_K-\theta_K'Z_K'Z_K) & \frac{n}{2\sigma_{eK}^4}-\frac{1}{\sigma_{eK}^6}(y-Z_K\theta_K)'(y-Z_K\theta_K)\end{bmatrix},$$

we find

$$I(\delta_o)=\begin{bmatrix}\frac{1}{\sigma_o^2}Z_K'Z_K & 0\\ 0 & \frac{n\sigma_e^2}{\sigma_o^6}-\frac{n}{2\sigma_o^4}+\frac{\tau'M_{Z_K}\tau}{\sigma_o^6}\end{bmatrix}. \tag{A.2}$$

Using (A.1), (A.2) and the relation $\tau'M_{Z_K}\tau=n(\sigma_o^2-\sigma_e^2)$, we obtain

$$\begin{aligned}tr\left\{J(\delta_o)I(\delta_o)^{-1}\right\} &= \frac{\sigma_e^2}{\sigma_o^2}tr\left\{\begin{bmatrix}I_{p(K_U+K_L+2)+1} & 0\\ 0 & \frac{2\sigma_o^2-\sigma_e^2}{\sigma_o^2}\end{bmatrix}\right\}\\ &= \frac{\sigma_e^2}{\sigma_o^2}\left(p(K_U+K_L+2)+1+\frac{2\sigma_o^2-\sigma_e^2}{\sigma_o^2}\right).\end{aligned}$$

Appendix II: Proofs

The following lemma is required for the derivation of the corrected AIC in Subsection 3.

Lemma A.1 *(i)* $\frac{V'M_{Z_K}V}{n-m} = o_p(1)$. *(ii)* $\frac{V'M_{Z_K}e}{n-m} = o_p(1)$.

Proof: (i) Write $V'M_{Z_K}V = V'V - V'Z_K(Z_K'Z_K)^{-1}Z_KV$. Since

$$\begin{aligned}
E\left(V_{t,K}^2\right) &= E\left(\sum_{j>K_U,\ j<-K_L} \pi_j' v_{t-j}\right)^2 \\
&= \sum_{j>K_U,\ j<-K_L} E\left(\pi_j' v_{t-j} v_{t-j}' \pi_j\right) + \sum_{j>K_U,\ j<-K_L} \sum_{l>K_U,\ l<-K_L} E\left(\pi_j' v_{t-j} v_{t-l}' \pi_l\right) \\
&\leq \sum_{j>K_U,\ j<-K_L} \pi_j' \Omega_{vv} \pi_j + \sum_{j>K_U,\ j<-K_L} \sum_{l>K_U,\ l<-K_L} \sqrt{\pi_j' \Omega_{vv} \pi_j} \sqrt{\pi_l' \Omega_{vv} \pi_l} \\
&\leq l_{vv} \left(\sum_{j>K_U,\ j<-K_L} \|\pi_j\|^2 + \sum_{j>K_U,\ j<-K_L} \sum_{l>K_U,\ l<-K_L} \|\pi_j\| \|\pi_l\| \right) \\
&= l_{vv} \left(\sum_{j>K_U,\ j<-K_L} \|\pi_j\| \right)^2,
\end{aligned}$$

where l_{vv} is the maximum eigenvalue of Ω_{vv} and the Cauchy-Schwarz inequality is used for the first inequality. Assumption (9) implies $E\left(V_{t,K}^2\right) = o(T^{-1})$. Thus,

$$V'V = o_p(1). \tag{A.3}$$

Next, let $D_T = diag[n^{-1/2}, n^{-1}I_p, n^{-1/2}I_p, \ldots, n^{-1/2}I_p]$ and $R = diag[1, n^{-2}\sum_{t=K_L+2}^{T-K_U} x_t x_t', \Gamma]$ where $\Gamma = E\left(v_{t+K_L}', \ldots, v_{t-K_U}'\right)'\left(v_{t+K_L}', \ldots, v_{t-K_U}'\right)$. Then, we have the following equality for the second term of $V'M_{Z_K}V$.

$$\begin{aligned}
&\left\|V'Z_K D_T (D_T Z_K' Z_K D_T)^{-1} D_T Z_K V\right\| \\
&\qquad \leq \left\|V'Z_K D_T\right\| \left\|R^{-1}\right\|_1 \left\|D_T Z_K V\right\| \\
&\qquad\quad + \left\|V'Z_K D_T\right\| \left\|D_T (Z_K' Z_K)^{-1} D_T - R^{-1}\right\|_1 \left\|D_T Z_K V\right\|,
\end{aligned}$$

where Lemma A1 of Saikkonen is used for the inequality. As shown in Saikkonen (1991),[7] $\|V'Z_K D_T\| = o_p(K^{1/2})$, $\left\|R^{-1}\right\|_1 = O_p(1)$, and $\left\|D_T(Z_K'Z_K)^{-1}D_T - R^{-1}\right\|_1 =$

[7] Saikkonen (1991) does not consider an intercept term in his linear regression model, but extending his results to the model with an intercept term is straightforward.

$O_p(K/T^{1/2})$, where K/K_L, $K/K_U = O(1)$. Therefore, $V'Z_K(Z_K'Z_K)^{-1}Z_KV = o_p(K)$, which gives the stated result along with Assumption (8).

(ii) Write $V'M_{Z_K}e = V'e - V'Z_K(Z_K'Z_K)^{-1}Z_Ke$. Then, $V'e = o_p(\sqrt{n})$ because

$$E\left\|V'e\right\| \le E\left\|V\right\|\left\|e\right\| \le \sqrt{E\left(V'V\right)}\sqrt{E(e'e)} = o(1)\times O(\sqrt{n}) = o(\sqrt{n}).$$

Since $\|D_TZ_Ke\| = O_p(K^{1/2})$, as shown in Saikkonen (1991), $V'Z_K(Z_K'Z_K)^{-1}Z_Ke = o_p(K)$. Thus, the stated result follows.

Lemma A.2 $\sigma_o^2 \xrightarrow{p} \sigma_e^2$.

Proof: Since $\sigma_o^2 = \frac{(\tau - Z_K\theta_o)'(\tau - Z_K\theta_o)}{n} + \sigma_e^2$ and $\frac{(\tau - Z_K\theta_o)'(\tau - Z_K\theta_o)}{n} = \frac{V'M_{Z_K}V}{n} \xrightarrow{p} 0$ by Lemma A.1 (i), we obtain the result.

Lemma A.3 *Assume* $e_t \sim iid\ N(0,\sigma_e^2)$.
(i) $\frac{(\tau - Z_K\hat{\theta}_K)'(\tau - Z_K\hat{\theta}_K)}{\hat{\sigma}_{eK}^2}$ *follows* $\frac{nm}{n-m}F(m, n-m)$ *when* T *is large.*
(ii) $\frac{n\sigma_e^2}{\hat{\sigma}_{eK}^2}$ *follows* $\frac{n^2}{\chi^2(n-m)}$ *when* T *is large.*

Proof: (i) Since $\tau - Z_K\hat{\theta}_K = \tau - P_{Z_K}y = M_{Z_K}\tau - P_{Z_K}e$, the second term in relation (10) is written as

$$\begin{aligned} \frac{(\tau - Z_K\hat{\theta}_K)'(\tau - Z_K\hat{\theta}_K)}{\hat{\sigma}_{eK}^2} &= \frac{e'P_{Z_K}e}{\hat{\sigma}_{eK}^2} + \frac{\tau'M_{Z_K}\tau}{\hat{\sigma}_{eK}^2} \\ &= \frac{nm}{n-m}\frac{\left(\frac{e'P_{Z_K}e}{\sigma_e^2}\right)/m}{\left(\frac{n\hat{\sigma}_{eK}^2}{\sigma_e^2}\right)/(n-m)} + \frac{\tau'M_{Z_K}\tau}{\hat{\sigma}_{eK}^2}, \qquad \text{(A.4)} \end{aligned}$$

where $m = p(K_U + K_L + 2) + 1$. For the first term in relation (A.4), note first that $\frac{e'P_{Z_K}e}{\sigma_e^2} \overset{d}{=} \chi^2(m)$. Next, letting $V = [V_{K_L+2,K}, \ldots, V_{T-K_U,K}]'$, we have

$$\begin{aligned} \left(\frac{n\hat{\sigma}_{eK}^2}{\sigma_e^2}\right)/(n-m) &= \frac{e'M_{Z_K}e}{\sigma_e^2(n-m)} + \frac{V'M_{Z_K}V}{\sigma_e^2(n-m)} + \frac{2V'M_{Z_K}e}{\sigma_e^2(n-m)} \\ &= \frac{e'M_{Z_K}e}{\sigma_e^2(n-m)} + o_p(1), \qquad \text{(A.5)} \end{aligned}$$

and $\frac{e'M_{Z_K}e}{\sigma_e^2(n-m)} \overset{d}{=} \chi^2(n-m)/(n-m)$. Note that the second equality of relation (A.5) holds due to Lemma A.1. Since $\frac{e'P_{Z_K}e}{\sigma_e^2}$ and $\frac{e'M_{Z_K}e}{\sigma_e^2}$ are independent, the first term in (A.4) is approximately distributed as $\frac{nm}{n-m}F(m, n-m)$. For the second term in (A.4), note that relation (A.3) yields $\tau' M_{Z_K} \tau = V' M_{Z_K} V \leq V'V = o_p(1)$ and that $\hat{\sigma}_{eK}^2 = O_p(1)$. Thus, the second term is $o_p(1)$. This completes the proof.

(ii) Relation (A.5) implies that $\frac{n\sigma_e^2}{\hat{\sigma}_{eK}^2}$ is distributed as $\frac{n^2}{\chi^2(n-m)}$.

Table 1: Bias of the Estimates

Part I: $a_{11} = \theta_{11} = 0$ (bias$\times$10, $T = 100$, ε_t: normal)

	$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$	$\sigma_{12} = .8$
(a_{11}, a_{22})	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .8)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, .8)$	$(0, .8)$
$(\theta_{11}, \theta_{22})$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .4)$
$(K_{\max} = K_4)$										
C_p	.0082	.3433	.0135	.6369	.0084	−.0051	−.0019	−.0018	.1745	.3243
AIC	.0304	.3835	.0258	.6389	−.0197	−.0115	−.0026	−.0021	.2055	.3347
AIC_c	.0364	.4023	.0312	.6392	−.0235	−.0149	−.0026	−.0020	.2165	.3391
BIC	.0777	.5010	.0594	.6665	−.0467	−.0402	−.0029	−.0012	.2768	.3872
C_p ($K_L = K_U$)	.0106	.3548	.0155	.6474	−.0116	−.0069	−.0017	−.0015	.1790	.3323
AIC ($K_L = K_U$)	.0482	.4357	.0393	.6722	−.0306	−.0250	−.0029	−.0017	.2328	.3644
AIC_c ($K_L = K_U$)	.0591	.4664	.0491	.6916	−.0366	−.0310	−.0027	−.0011	.2524	.3857
BIC ($K_L = K_U$)	.1077	.5957	.0935	.8024	−.0684	−.0876	−.0087	−.0018	.3142	.4946
$K_L = K_U = K_{\max}$	.1400	.7226	.2896	1.4632	−.0960	−.1868	−.1847	−.3670	.3587	.7304
$(K_{\max} = K_{12})$										
C_p	−.0155	.0564	−.0028	.1664	−.0017	−.0016	−.0049	−.0018	.0270	.0825
AIC	.0220	.1515	.0063	.2386	−.0153	−.0092	−.0021	−.0014	.1066	.1473
AIC_c	.0283	.2610	.0261	.3068	−.0237	−.0163	−.0018	−.0013	.1591	.1984
BIC	.0763	.4242	.0577	.4229	−.0463	−.0398	−.0029	−.0012	.2624	.3107
C_p ($K_L = K_U$)	−.0206	.0658	−.0073	.1637	−.0066	−.0043	−.0053	−.0009	.0268	.0823
AIC ($K_L = K_U$)	.0409	.2016	.0233	.2858	−.0275	−.0231	−.0017	−.0011	.1401	.1868
AIC_c ($K_L = K_U$)	.0556	.3810	.0462	.4619	−.0355	−.0301	−.0017	−.0004	.2344	.3099
BIC ($K_L = K_U$)	.1073	.5693	.0933	.6541	−.0684	−.0876	−.0087	−.0017	.3111	.4694
$K_L = K_U = K_{\max}$	.1629	.8344	.3390	1.6892	−.1168	−.2248	−.2199	−.4357	.4128	.8425
$K_L = K_U = 1$	.0480	.5513	.1086	1.1267	−.0043	−.0027	−.0022	−.0017	.2727	.5620
$K_L = K_U = 2$	.0148	.4524	.0425	.9314	−.0026	−.0001	−.0013	−.0001	.2237	.4644
$K_L = K_U = 3$	.0009	.3703	.0143	.7690	−.0026	−.0001	−.0017	−.0002	.1826	.3829

Part II: $a_{11} \neq 0$ and/or $\theta_{11} \neq 0$ (bias$\times$10, $T = 100$, ε_t: normal)

	$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$	$\sigma_{12} = 0.8$
(a_{11}, a_{22})	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.8, .8)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(.8, .8)$	$(.8, .8)$
$(\theta_{11}, \theta_{22})$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.4, .4)$
$(K_{\max} = K_4)$										
C_p	−.0059	−.0065	−.0024	−.0005	−.0048	−.0094	−.0031	−.0027	.1745	−.0012
AIC	−.0056	−.0071	−.0022	−.0015	−.0048	−.0045	−.0044	−.0040	.2055	−.0013
AIC_c	−.0059	−.0062	−.0019	−.0010	−.0046	−.0059	−.0039	−.0037	.2165	−.0013
BIC	−.0049	−.0056	−.0016	−.0008	−.0046	−.0041	−.0023	−.0041	.2768	−.0013
C_p ($K_L = K_U$)	−.0060	−.0066	−.0031	−.0004	−.0065	−.0114	−.0031	−.0052	.1790	−.0010
AIC ($K_L = K_U$)	−.0077	−.0056	−.0024	−.0014	−.0050	−.0018	−.0048	−.0003	.2328	−.0017
AIC_c ($K_L = K_U$)	−.0061	−.0061	−.0024	−.0015	−.0063	−.0033	−.0036	−.0010	.2524	−.0016
BIC ($K_L = K_U$)	−.0049	−.0051	−.0021	−.0012	−.0048	−.0071	−.0027	−.0057	.3142	−.0015
$K_L = K_U = K_{\max}$	−.0049	−.0050	−.0018	−.0011	−.0058	−.0104	−.0031	−.0066	.3587	−.0013
$(K_{\max} = K_{12})$										
C_p	−.0137	−.0121	−.0054	−.0033	.0026	.0095	−.0072	−.0095	.0270	−.0047
AIC	−.0058	−.0094	−.0075	−.0036	.0000	.0051	−.0054	−.0037	.1066	−.0067
AIC_c	−.0069	−.0084	−.0022	−.0029	−.0067	−.0048	−.0035	−.0042	.1591	−.0026
BIC	−.0054	−.0077	−.0012	−.0031	−.0040	−.0027	−.0023	−.0032	.2624	−.0007
C_p ($K_L = K_U$)	−.0149	−.0119	−.0069	−.0032	−.0094	−.0054	−.0086	−.0096	.0268	−.0035
AIC ($K_L = K_U$)	−.0074	−.0132	−.0040	−.0058	−.0017	.0070	−.0037	−.0006	.1401	−.0065
AIC_c ($K_L = K_U$)	−.0073	−.0068	−.0023	−.0002	−.0060	−.0013	−.0031	.0006	.2344	−.0016
BIC ($K_L = K_U$)	−.0048	−.0047	−.0021	−.0013	−.0048	−.0073	−.0027	−.0057	.3111	−.0011
$K_L = K_U = K_{\max}$	−.0086	−.0067	−.0038	−.0015	−.0086	−.0087	−.0046	−.0069	.4128	−.0019
$K_L = K_U = 1$	−.0054	−.0050	−.0023	−.0014	−.0071	−.0111	−.0035	−.0058	.2727	−.0016
$K_L = K_U = 2$	−.0053	−.0052	−.0024	−.0014	−.0043	.0021	−.0022	.0011	.2237	−.0016
$K_L = K_U = 3$	−.0054	−.0054	−.0027	−.0012	−.0046	−.0008	−.0034	−.0030	.1826	−.0016

Table 2: Bias of the Estimates

Part I: $a_{11} = \theta_{11} = 0$ (bias$\times 10^2$, $T = 100$, ε_t: log-normal)

	$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$	$\sigma_{12} = .8$
(a_{11}, a_{22})	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .8)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, .8)$	$(0, .8)$
$(\theta_{11}, \theta_{22})$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .4)$
$(K_{\max} = K_4)$										
C_p	.0078	.4234	.0093	.4088	−.0064	−.0047	.0007	.0006	.2250	.2103
AIC	.0172	.4444	.0132	.4078	−.0085	−.0078	.0009	.0007	.2527	.2147
AIC_c	.0205	.4544	.0142	.4079	−.0096	−.0084	.0005	.0006	.2633	.2169
BIC	.0345	.5515	.0202	.4169	−.0138	−.0158	−.0006	.0007	.3372	.2368
C_p $(K_L = K_U)$	.0127	.4528	.0089	.4189	−.0044	−.0032	.0007	.0001	.2436	.2162
AIC $(K_L = K_U)$	.0257	.5120	.0151	.4290	−.0101	−.0091	.0000	.0005	.2978	.2317
AIC_c $(K_L = K_U)$	.0293	.5443	.0173	.4366	−.0108	−.0110	−.0001	.0005	.3200	.2409
BIC $(K_L = K_U)$	.0440	.6975	.0275	.4800	−.0185	−.0243	−.0032	−.0001	.3921	.2883
$K_L = K_U = K_{\max}$	.0667	.9647	.0727	.9509	−.0276	−.0551	−.0315	−.0634	.4823	.4757
$(K_{\max} = K_{12})$										
C_p	−.0083	.1096	.0001	.0994	−.0013	−.0008	−.0008	.0020	.0593	.0575
AIC	.0045	.1780	.0113	.1416	−.0060	−.0097	.0010	.0023	.1198	.1022
AIC_c	.0123	.2424	.0125	.1819	−.0083	−.0113	−.0003	.0015	.1722	.1309
BIC	.0316	.4053	.0180	.2384	−.0131	−.0178	−.0008	.0016	.2894	.1771
C_p $(K_L = K_U)$	−.0085	.1071	−.0044	.0938	−.0006	.0013	−.0012	.0001	.0591	.0494
AIC $(K_L = K_U)$	.0135	.2280	.0108	.1517	−.0060	−.0089	−.0001	.0018	.1612	.1097
AIC_c $(K_L = K_U)$	.0236	.3859	.0161	.2560	−.0090	−.0134	.0001	.0019	.2542	.1808
BIC $(K_L = K_U)$	.0418	.6072	.0275	.3448	−.0173	−.0250	−.0031	.0002	.3610	.2553
$K_L = K_U = K_{\max}$	.0762	1.1273	.0776	1.0753	−.0325	−.0639	−.0355	−.0696	.5631	.5369
$K_L = K_U = 1$	.0244	.7429	.0285	.7383	−.0012	−.0009	−.0004	−.0006	.3707	.3690
$K_L = K_U = 2$	.0093	.6158	.0128	.6101	−.0012	−.0010	.0003	−.0001	.3073	.3054
$K_L = K_U = 3$	.0030	.5114	.0064	.5048	−.0005	−.0001	.0009	.0005	.2553	.2530

Part II: $a_{11} \neq 0$ and/or $\theta_{11} \neq 0$ (bias$\times 10^2$, $T = 100$, ε_t: log-normal)

	$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$	$\sigma_{12} = 0.8$
(a_{11}, a_{22})	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.8, .8)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(.8, .8)$	$(.8, .8)$
$(\theta_{11}, \theta_{22})$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.4, .4)$
$(K_{\max} = K_4)$										
C_p	.0003	.0287	−.0004	−.0071	−.0059	−.0300	.0012	.0039	.2250	−.0064
AIC	.0026	.0321	−.0010	−.0079	−.0082	−.0437	.0019	.0088	.2527	−.0074
AIC_c	.0034	.0334	−.0008	−.0085	−.0085	−.0473	.0019	.0084	.2633	−.0076
BIC	.0055	.0420	−.0016	−.0105	−.0093	−.0669	.0016	.0115	.3372	−.0096
C_p $(K_L = K_U)$	.0014	.0301	−.0001	−.0067	−.0041	−.0250	.0014	.0045	.2436	−.0062
AIC $(K_L = K_U)$	.0040	.0360	−.0004	−.0080	−.0068	−.0458	.0022	.0076	.2978	−.0077
AIC_c $(K_L = K_U)$	.0041	.0383	−.0005	−.0088	−.0078	−.0502	.0018	.0087	.3200	−.0082
BIC $(K_L = K_U)$	.0067	.0480	−.0018	−.0111	−.0097	−.0697	.0013	.0131	.3921	−.0103
$K_L = K_U = K_{\max}$	.0123	.0600	−.0024	−.0136	−.0122	−.0734	.0034	.0184	.4823	−.0122
$(K_{\max} = K_{12})$										
C_p	−.0062	.0006	−.0022	−.0049	−.0018	−.0115	−.0009	.0030	.0593	−.0033
AIC	−.0024	.0073	.0001	−.0038	−.0058	−.0487	.0011	.0108	.1198	−.0025
AIC_c	−.0004	.0124	−.0005	−.0050	−.0085	−.0606	.0003	.0137	.1722	−.0049
BIC	.0048	.0297	−.0012	−.0087	−.0092	−.0760	.0017	.0160	.2894	−.0075
C_p $(K_L = K_U)$	−.0069	.0015	−.0034	−.0056	.0005	−.0031	−.0017	−.0001	.0591	−.0042
AIC $(K_L = K_U)$	.0002	.0117	.0005	−.0040	−.0036	−.0446	.0013	.0106	.1612	−.0045
AIC_c $(K_L = K_U)$	.0018	.0228	−.0003	−.0077	−.0058	−.0602	.0029	.0134	.2542	−.0062
BIC $(K_L = K_U)$	.0062	.0405	−.0014	−.0104	−.0084	−.0729	.0011	.0134	.3610	−.0091
$K_L = K_U = K_{\max}$	.0139	.0704	−.0047	−.0169	−.0160	−.0896	.0021	.0183	.5631	−.0154
$K_L = K_U = 1$	.0039	.0461	−.0009	−.0105	−.0022	−.0054	−.0009	−.0027	.3707	−.0095
$K_L = K_U = 2$	.0012	.0384	.0004	−.0084	−.0023	−.0081	.0003	−.0016	.3073	−.0076
$K_L = K_U = 3$	−.0003	.0320	.0007	−.0071	−.0003	−.0007	.0018	.0043	.2553	−.0063

Table 3: Bias of the Estimates

Part I: $a_{11} = \theta_{11} = 0$ (bias×10, $T = 300$, ε_t: normal)

	$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$	$\sigma_{12} = .8$
(a_{11}, a_{22})	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .8)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, .8)$	$(0, .8)$
$(\theta_{11}, \theta_{22})$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .4)$
$(K_{\max} = K_4)$										
C_p	.0049	.1046	.0025	.1847	.0002	.0001	.0000	.0000	.0560	.0926
AIC	.0092	.1139	.0054	.1839	−.0008	.0000	.0001	.0001	.0658	.0928
AIC_c	.0096	.1155	.0059	.1837	−.0008	−.0001	.0001	.0000	.0667	.0927
BIC	.0241	.1597	.0142	.1827	−.0068	−.0008	.0000	.0000	.1010	.0983
C_p $(K_L = K_U)$	.0059	.1100	.0033	.1857	.0000	.0003	.0000	.0001	.0604	.0935
AIC $(K_L = K_U)$	.0131	.1311	.0080	.1861	−.0021	.0002	.0000	−.0001	.0788	.0957
AIC_c $(K_L = K_U)$	.0141	.1346	.0088	.1863	−.0023	.0000	−.0001	−.0001	.0811	.0963
BIC $(K_L = K_U)$	.0366	.2177	.0229	.1965	−.0161	−.0045	−.0002	−.0001	.1260	.1182
$K_L = K_U = K_{\max}$	.0510	.2826	.0975	.5509	−.0294	−.0598	−.0595	−.1187	.1418	.2754
$(K_{\max} = K_{12})$										
C_p	.0005	.0262	.0003	.0319	.0000	.0001	.0000	.0000	.0153	.0172
AIC	.0069	.0522	.0035	.0448	−.0005	.0002	.0000	−.0001	.0370	.0300
AIC_c	.0080	.0586	.0041	.0489	−.0006	.0001	.0001	−.0001	.0425	.0333
BIC	.0241	.1457	.0142	.0944	−.0068	−.0008	.0000	.0000	.0983	.0708
C_p $(K_L = K_U)$	.0012	.0304	.0001	.0349	.0002	.0004	−.0002	−.0003	.0171	.0191
AIC $(K_L = K_U)$	.0113	.0786	.0068	.0587	−.0019	.0002	−.0001	−.0001	.0573	.0417
AIC_c $(K_L = K_U)$	.0128	.0925	.0078	.0698	−.0022	.0001	.0000	−.0001	.0657	.0493
BIC $(K_L = K_U)$	.0366	.2148	.0229	.1501	−.0161	−.0045	−.0002	−.0001	.1257	.1092
$K_L = K_U = K_{\max}$	.0539	.3014	.1047	.5904	−.0320	−.0644	−.0636	−.1272	.1510	.2952
$K_L = K_U = 1$	.0211	.2232	.0382	.4321	.0003	−.0002	−.0002	−.0001	.1120	.2160
$K_L = K_U = 2$	.0096	.1817	.0155	.3503	.0008	.0004	.0001	.0001	.0914	.1752
$K_L = K_U = 3$	.0045	.1474	.0061	.2836	.0004	.0001	−.0001	.0000	.0741	.1418

Part II: $a_{11} \neq 0$ and/or $\theta_{11} \neq 0$ (bias×10, $T = 300$, ε_t: normal)

	$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$	$\sigma_{12} = 0.8$
(a_{11}, a_{22})	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.8, .8)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(.8, .8)$	$(.8, .8)$
$(\theta_{11}, \theta_{22})$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.4, .4)$
$(K_{\max} = K_4)$										
C_p	.0008	.0005	−.0004	−.0003	.0008	−.0003	.0002	.0004	.0560	−.0003
AIC	.0007	.0006	−.0001	−.0002	.0010	.0007	.0001	.0000	.0658	−.0002
AIC_c	.0009	.0007	.0000	−.0002	.0011	.0005	.0001	.0001	.0667	−.0002
BIC	.0010	.0007	−.0001	−.0001	.0009	.0009	.0001	.0010	.1010	−.0001
C_p $(K_L = K_U)$	.0006	.0006	−.0002	−.0002	.0008	.0006	.0003	.0010	.0604	−.0001
AIC $(K_L = K_U)$	.0007	.0006	.0000	−.0001	.0008	.0006	.0002	.0008	.0788	−.0002
AIC_c $(K_L = K_U)$	.0008	.0006	−.0001	−.0002	.0009	.0010	.0001	.0007	.0811	−.0001
BIC $(K_L = K_U)$	.0010	.0007	−.0001	−.0002	.0010	.0022	.0000	.0012	.1260	−.0002
$K_L = K_U = K_{\max}$	.0008	.0005	−.0002	−.0003	.0012	.0034	−.0001	.0002	.1418	−.0003
$(K_{\max} = K_{12})$										
C_p	−.0006	−.0005	−.0003	−.0005	.0005	.0006	−.0003	−.0008	.0153	−.0004
AIC	.0005	−.0006	−.0004	−.0004	.0009	.0003	.0001	.0000	.0370	−.0001
AIC_c	.0006	−.0004	−.0002	−.0003	.0010	−.0001	.0000	.0001	.0425	−.0003
BIC	.0010	.0007	−.0001	.0000	.0009	.0009	.0001	.0010	.0983	−.0001
C_p $(K_L = K_U)$	−.0007	−.0005	−.0005	−.0003	.0005	.0008	−.0003	−.0009	.0171	−.0005
AIC $(K_L = K_U)$	.0004	.0001	−.0001	−.0001	.0009	.0004	.0002	.0005	.0573	−.0001
AIC_c $(K_L = K_U)$	.0006	.0001	−.0002	−.0001	.0010	.0005	.0001	.0007	.0657	−.0002
BIC $(K_L = K_U)$	.0010	.0007	−.0001	−.0001	.0010	.0022	.0000	.0012	.1257	−.0002
$K_L = K_U = K_{\max}$	.0004	.0004	.0000	.0001	.0006	.0012	−.0001	−.0005	.1510	.0001
$K_L = K_U = 1$	.0009	.0007	−.0001	−.0002	.0007	−.0014	−.0002	−.0014	.1120	−.0002
$K_L = K_U = 2$	.0010	.0007	.0000	−.0002	.0013	.0027	.0001	.0008	.0914	−.0002
$K_L = K_U = 3$	.0007	.0006	−.0002	−.0003	.0008	.0012	−.0001	−.0001	.0741	−.0003

Table 4: Bias of the Estimates

Part I: $a_{11} = \theta_{11} = 0$ (bias$\times 10^3$, $T = 300$, ε_t: log-normal)

	$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$	$\sigma_{12} = .8$
(a_{11}, a_{22})	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .8)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, .8)$	$(0, .8)$
$(\theta_{11}, \theta_{22})$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .8)$	$(0, .4)$	$(0, .4)$
$(K_{\max} = K_4)$										
C_p	−.0013	.3427	.0025	.3568	−.0059	−.0035	−.0008	−.0001	.1836	.1807
AIC	.0049	.3594	.0047	.3569	−.0062	−.0056	−.0008	.0004	.2137	.1830
AIC_c	.0057	.3640	.0046	.3574	−.0064	−.0058	−.0009	.0003	.2169	.1833
BIC	.0277	.5095	.0098	.3599	−.0099	−.0064	−.0015	.0001	.3470	.1950
C_p $(K_L = K_U)$	.0077	.3856	.0006	.3587	−.0024	−.0013	−.0013	.0003	.2171	.1802
AIC $(K_L = K_U)$	.0159	.4423	.0040	.3606	−.0063	−.0036	−.0013	.0008	.2766	.1853
AIC_c $(K_L = K_U)$	.0163	.4522	.0046	.3611	−.0061	−.0035	−.0016	.0007	.2849	.1868
BIC $(K_L = K_U)$	.0418	.7135	.0163	.3821	−.0168	−.0105	−.0018	.0007	.4385	.2268
$K_L = K_U = K_{\max}$	.0652	1.0822	.0758	1.0932	−.0335	−.0620	−.0391	−.0742	.5386	.5446
$(K_{\max} = K_{12})$										
C_p	−.0088	.0729	.0048	.0886	−.0046	−.0048	−.0006	−.0001	.0511	.0580
AIC	.0009	.1451	.0053	.1143	−.0048	−.0075	−.0001	.0014	.1196	.0800
AIC_c	.0018	.1673	.0053	.1216	−.0044	−.0083	−.0003	.0014	.1330	.0842
BIC	.0264	.4492	.0090	.1941	−.0104	−.0087	−.0014	.0007	.3231	.1394
C_p $(K_L = K_U)$	.0010	.1245	.0036	.0774	−.0024	−.0029	−.0021	−.0003	.0801	.0487
AIC $(K_L = K_U)$	.0129	.2692	.0037	.1274	−.0054	−.0052	−.0010	.0009	.1926	.0895
AIC_c $(K_L = K_U)$	.0116	.3046	.0033	.1459	−.0055	−.0057	−.0012	.0007	.2124	.1054
BIC $(K_L = K_U)$	.0383	.6787	.0167	.2783	−.0173	−.0113	−.0017	.0011	.4186	.2053
$K_L = K_U = K_{\max}$	.0748	1.1625	.0805	1.1551	−.0302	−.0605	−.0370	−.0725	.5813	.5768
$K_L = K_U = 1$	.0250	.8506	.0288	.8606	.0003	.0019	−.0006	.0006	.4246	.4294
$K_L = K_U = 2$	.0074	.6851	.0098	.6935	−.0020	−.0014	−.0014	−.0009	.3413	.3458
$K_L = K_U = 3$	.0003	.5508	.0019	.5574	−.0007	.0006	−.0004	.0007	.2744	.2780

Part II: $a_{11} \neq 0$ and/or $\theta_{11} \neq 0$ (bias$\times 10^3$, $T = 300$, ε_t: log-normal)

	$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$	$\sigma_{12} = 0.8$
(a_{11}, a_{22})	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.8, .8)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(.8, .8)$	$(.8, .8)$
$(\theta_{11}, \theta_{22})$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.8, .8)$	$(.4, .4)$	$(.4, .4)$
$(K_{\max} = K_4)$										
C_p	.0007	.0364	−.0039	−.0121	−.0079	−.0353	−.0013	.0058	.1836	−.0113
AIC	.0030	.0469	−.0048	−.0145	−.0116	−.0593	−.0010	.0082	.2137	−.0136
AIC_c	.0032	.0483	−.0048	−.0147	−.0117	−.0601	−.0013	.0090	.2169	−.0138
BIC	.0068	.0675	−.0059	−.0194	−.0160	−.0917	−.0020	.0136	.3470	−.0174
C_p $(K_L = K_U)$	.0011	.0403	−.0044	−.0130	−.0081	−.0336	−.0022	.0042	.2171	−.0121
AIC $(K_L = K_U)$	.0030	.0534	−.0049	−.0161	−.0137	−.0688	−.0035	.0052	.2766	−.0148
AIC_c $(K_L = K_U)$	.0034	.0549	−.0051	−.0166	−.0140	−.0690	−.0032	.0047	.2849	−.0152
BIC $(K_L = K_U)$	.0071	.0718	−.0059	−.0202	−.0174	−.0951	−.0018	.0131	.4385	−.0181
$K_L = K_U = K_{\max}$	.0100	.0793	−.0072	−.0224	−.0204	−.0951	−.0008	.0162	.5386	−.0196
$(K_{\max} = K_{12})$										
C_p	−.0049	.0102	−.0004	−.0031	−.0071	−.0348	−.0006	.0071	.0511	−.0020
AIC	−.0008	.0278	−.0029	−.0078	−.0082	−.0680	.0000	.0101	.1196	−.0078
AIC_c	.0010	.0306	−.0023	−.0087	−.0086	−.0725	−.0004	.0127	.1330	−.0082
BIC	.0052	.0634	−.0053	−.0179	−.0171	−.0998	−.0022	.0158	.3231	−.0163
C_p $(K_L = K_U)$	−.0023	.0149	−.0001	−.0034	−.0050	−.0309	−.0030	.0008	.0801	−.0026
AIC $(K_L = K_U)$	.0023	.0372	−.0045	−.0108	−.0114	−.0677	−.0029	.0077	.1926	−.0109
AIC_c $(K_L = K_U)$	.0017	.0413	−.0044	−.0121	−.0115	−.0695	−.0031	.0064	.2124	−.0116
BIC $(K_L = K_U)$	.0051	.0671	−.0055	−.0194	−.0183	−.0980	−.0018	.0134	.4186	−.0171
$K_L = K_U = K_{\max}$	.0158	.0888	−.0051	−.0222	−.0158	−.0933	.0024	.0225	.5813	−.0192
$K_L = K_U = 1$	.0037	.0623	−.0033	−.0174	−.0002	.0052	−.0014	.0008	.4246	−.0151
$K_L = K_U = 2$	.0000	.0501	−.0024	−.0146	−.0034	−.0073	−.0025	−.0047	.3413	−.0127
$K_L = K_U = 3$	−.0015	.0401	−.0021	−.0124	−.0016	.0014	−.0008	.0033	.2744	−.0107

Table 5: MSE of the Estimates

Part I: $a_{11} = \theta_{11} = 0$ (MSE×10, $T = 100$, ε_t: normal)

	$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$	$\sigma_{12} = .8$
(a_{11}, a_{22})	(0, .4)	(0, .8)	(0, .4)	(0, .8)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, .8)	(0, .8)
$(\theta_{11}, \theta_{22})$	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, .4)	(0, .8)	(0, .4)	(0, .8)	(0, .4)	(0, .4)
($K_{\max} = K_4$)										
C_p	.0398	.3031	.0172	.1863	.0060	.0023	.0026	.0009	.1081	.0619
AIC	.0354	.2853	.0157	.1828	.0054	.0022	.0023	.0008	.0985	.0611
AIC_c	.0339	.2807	.0152	.1808	.0053	.0022	.0022	.0008	.0959	.0607
BIC	.0302	.2581	.0141	.1811	.0051	.0028	.0020	.0007	.0854	.0637
C_p ($K_L = K_U$)	.0398	.3074	.0175	.1948	.0061	.0024	.0027	.0010	.1079	.0643
AIC ($K_L = K_U$)	.0340	.2854	.0163	.2004	.0054	.0025	.0022	.0008	.0951	.0670
AIC_c ($K_L = K_U$)	.0326	.2765	.0158	.2045	.0052	.0026	.0022	.0008	.0916	.0686
BIC ($K_L = K_U$)	.0289	.2448	.0155	.2300	.0052	.0040	.0022	.0007	.0800	.0785
$K_L = K_U = K_{\max}$	.0373	.3005	.0285	.4206	.0069	.0079	.0083	.0249	.1021	.1161
($K_{\max} = K_{12}$)										
C_p	.1459	.9807	.0680	.4590	.0207	.0092	.0097	.0035	.3620	.1699
AIC	.0699	.7069	.0345	.3506	.0080	.0033	.0035	.0014	.2183	.1136
AIC_c	.0449	.4926	.0203	.2572	.0058	.0025	.0025	.0009	.1472	.0819
BIC	.0309	.3311	.0146	.2198	.0051	.0028	.0020	.0007	.0939	.0697
C_p ($K_L = K_U$)	.1563	1.0208	.0717	.4752	.0235	.0108	.0110	.0043	.3713	.1755
AIC ($K_L = K_U$)	.0637	.6961	.0334	.3700	.0072	.0034	.0031	.0014	.2044	.1205
AIC_c ($K_L = K_U$)	.0360	.3979	.0172	.2604	.0053	.0027	.0022	.0008	.1114	.0789
BIC ($K_L = K_U$)	.0289	.2736	.0155	.2607	.0052	.0040	.0022	.0007	.0810	.0811
$K_L = K_U = K_{\max}$	.0642	.4999	.0464	.6442	.0123	.0137	.0138	.0401	.1715	.1802
$K_L = K_U = 1$	.0295	.2376	.0142	.2709	.0046	.0016	.0020	.0007	.0825	.0773
$K_L = K_U = 2$	.0340	.2583	.0144	.2277	.0054	.0020	.0023	.0008	.0918	.0682
$K_L = K_U = 3$	.0394	.2868	.0163	.2027	.0064	.0023	.0026	.0010	.1035	.0636

Part II: $a_{11} \neq 0$ and/or $\theta_{11} \neq 0$ (MSE×10, $T = 100$, ε_t: normal)

	$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$	$\sigma_{12} = 0.8$
(a_{11}, a_{22})	(.4, .4)	(.8, .8)	(.4, .4)	(.8, .8)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(.8, .8)	(.8, .8)
$(\theta_{11}, \theta_{22})$	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(.4, .4)	(.8, .8)	(.4, .4)	(.8, .8)	(.4, .4)	(.4, .4)
($K_{\max} = K_4$)										
C_p	.0148	.0158	.0063	.0065	.0149	.0416	.0065	.0184	.1081	.0062
AIC	.0128	.0140	.0054	.0058	.0130	.0419	.0055	.0184	.0985	.0055
AIC_c	.0123	.0135	.0052	.0056	.0126	.0430	.0054	.0186	.0959	.0053
BIC	.0105	.0113	.0044	.0048	.0112	.0484	.0047	.0204	.0854	.0045
C_p ($K_L = K_U$)	.0147	.0158	.0062	.0066	.0154	.0464	.0065	.0194	.1079	.0062
AIC ($K_L = K_U$)	.0120	.0134	.0051	.0056	.0127	.0487	.0052	.0199	.0951	.0053
AIC_c ($K_L = K_U$)	.0114	.0125	.0048	.0053	.0120	.0486	.0050	.0204	.0916	.0050
BIC ($K_L = K_U$)	.0099	.0105	.0042	.0044	.0109	.0526	.0045	.0222	.0800	.0042
$K_L = K_U = K_{\max}$	.0129	.0129	.0053	.0053	.0138	.0374	.0057	.0163	.1021	.0052
($K_{\max} = K_{12}$)										
C_p	.0589	.0591	.0256	.0259	.0516	.1355	.0257	.0575	.3620	.0248
AIC	.0267	.0414	.0129	.0188	.0201	.0591	.0085	.0255	.2183	.0155
AIC_c	.0160	.0260	.0069	.0112	.0144	.0465	.0060	.0195	.1472	.0086
BIC	.0107	.0149	.0045	.0066	.0113	.0485	.0047	.0205	.0939	.0049
C_p ($K_L = K_U$)	.0616	.0616	.0264	.0269	.0590	.1557	.0282	.0702	.3713	.0256
AIC ($K_L = K_U$)	.0235	.0398	.0114	.0181	.0172	.0596	.0075	.0261	.2044	.0145
AIC_c ($K_L = K_U$)	.0124	.0200	.0052	.0082	.0123	.0500	.0051	.0206	.1114	.0059
BIC ($K_L = K_U$)	.0099	.0117	.0042	.0050	.0109	.0526	.0045	.0222	.0810	.0043
$K_L = K_U = K_{\max}$	.0225	.0222	.0093	.0094	.0251	.0599	.0101	.0243	.1715	.0091
$K_L = K_U = 1$	.0112	.0114	.0047	.0048	.0119	.0399	.0051	.0173	.0825	.0047
$K_L = K_U = 2$	.0130	.0132	.0054	.0055	.0142	.0423	.0059	.0176	.0918	.0053
$K_L = K_U = 3$	.0151	.0153	.0062	.0063	.0167	.0444	.0068	.0181	.1035	.0061

Table 6: MSE of the Estimates

Part I: $a_{11} = \theta_{11} = 0$ (MSE$\times 10^3$, $T = 100$, ε_t: log-normal)

	$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$	$\sigma_{12} = .8$
(a_{11}, a_{22})	(0, .4)	(0, .8)	(0, .4)	(0, .8)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, .8)	(0, .8)
$(\theta_{11}, \theta_{22})$	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, .4)	(0, .8)	(0, .4)	(0, .8)	(0, .4)	(0, .4)
$(K_{\max} = K_4)$										
C_p	.0287	.2775	.0105	.1315	.0039	.0012	.0014	.0004	.0975	.0430
AIC	.0262	.2735	.0098	.1276	.0036	.0011	.0013	.0004	.0950	.0419
AIC$_c$	.0256	.2720	.0096	.1261	.0036	.0011	.0013	.0004	.0944	.0416
BIC	.0240	.2717	.0092	.1228	.0035	.0011	.0012	.0003	.0921	.0417
C_p $(K_L = K_U)$	.0287	.2869	.0109	.1383	.0041	.0014	.0015	.0005	.0993	.0455
AIC $(K_L = K_U)$	.0262	.2878	.0103	.1396	.0038	.0013	.0014	.0004	.0974	.0461
AIC$_c$ $(K_L = K_U)$	.0255	.2874	.0101	.1403	.0037	.0013	.0013	.0004	.0960	.0465
BIC $(K_L = K_U)$	.0243	.2881	.0097	.1455	.0036	.0014	.0013	.0004	.0931	.0491
$K_L = K_U = K_{\max}$	.0296	.3450	.0147	.2809	.0048	.0039	.0031	.0071	.1108	.0791
$(K_{\max} = K_{12})$										
C_p	.0695	.5939	.0252	.2186	.0087	.0037	.0033	.0015	.2075	.0761
AIC	.0367	.4696	.0136	.1773	.0044	.0016	.0017	.0006	.1452	.0565
AIC$_c$	.0269	.3687	.0105	.1417	.0037	.0012	.0014	.0004	.1128	.0447
BIC	.0229	.3087	.0092	.1255	.0033	.0011	.0012	.0003	.0943	.0413
C_p $(K_L = K_U)$	.0750	.6210	.0277	.2341	.0106	.0048	.0040	.0020	.2168	.0814
AIC $(K_L = K_U)$	.0363	.4773	.0148	.1980	.0049	.0020	.0019	.0008	.1426	.0625
AIC$_c$ $(K_L = K_U)$	.0268	.3473	.0109	.1588	.0038	.0012	.0014	.0004	.1050	.0499
BIC $(K_L = K_U)$	.0240	.3132	.0099	.1576	.0037	.0013	.0013	.0004	.0958	.0498
$K_L = K_U = K_{\max}$	.0474	.5321	.0255	.4677	.0086	.0094	.0063	.0164	.1693	.1301
$K_L = K_U = 1$	.0257	.2739	.0098	.1869	.0038	.0010	.0014	.0004	.0912	.0549
$K_L = K_U = 2$	.0280	.2690	.0101	.1614	.0042	.0012	.0015	.0004	.0920	.0493
$K_L = K_U = 3$	.0309	.2738	.0110	.1464	.0048	.0015	.0017	.0005	.0956	.0464

Part II: $a_{11} \neq 0$ and/or $\theta_{11} \neq 0$ (MSE$\times 10^3$, $T = 100$, ε_t: log-normal)

	$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$	$\sigma_{12} = 0.8$
(a_{11}, a_{22})	(.4, .4)	(.8, .8)	(.4, .4)	(.8, .8)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(.8, .8)	(.8, .8)
$(\theta_{11}, \theta_{22})$	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(.4, .4)	(.8, .8)	(.4, .4)	(.8, .8)	(.4, .4)	(.4, .4)
$(K_{\max} = K_4)$										
C_p	.0103	.0113	.0036	.0040	.0103	.0239	.0037	.0087	.0975	.0037
AIC	.0092	.0105	.0033	.0037	.0093	.0206	.0033	.0075	.0950	.0034
AIC$_c$	.0090	.0103	.0032	.0036	.0091	.0200	.0033	.0072	.0944	.0033
BIC	.0083	.0092	.0030	.0032	.0087	.0176	.0032	.0063	.0921	.0030
C_p $(K_L = K_U)$	.0102	.0112	.0037	.0040	.0105	.0256	.0038	.0093	.0993	.0037
AIC $(K_L = K_U)$	.0090	.0103	.0032	.0036	.0094	.0218	.0033	.0076	.0974	.0033
AIC$_c$ $(K_L = K_U)$	.0088	.0099	.0032	.0035	.0092	.0201	.0033	.0071	.0960	.0032
BIC $(K_L = K_U)$	.0084	.0088	.0030	.0031	.0089	.0173	.0032	.0062	.0931	.0029
$K_L = K_U = K_{\max}$	.0102	.0102	.0036	.0036	.0111	.0215	.0040	.0078	.1108	.0035
$(K_{\max} = K_{12})$										
C_p	.0259	.0315	.0093	.0116	.0241	.0694	.0086	.0254	.2075	.0103
AIC	.0128	.0222	.0047	.0080	.0113	.0282	.0043	.0113	.1452	.0059
AIC$_c$	.0097	.0159	.0036	.0054	.0095	.0210	.0035	.0075	.1128	.0042
BIC	.0081	.0110	.0030	.0038	.0083	.0162	.0031	.0058	.0943	.0032
C_p $(K_L = K_U)$	.0277	.0336	.0102	.0122	.0281	.0847	.0104	.0327	.2168	.0109
AIC $(K_L = K_U)$	.0127	.0216	.0048	.0078	.0127	.0342	.0046	.0122	.1426	.0059
AIC$_c$ $(K_L = K_U)$	.0094	.0131	.0035	.0047	.0095	.0190	.0035	.0072	.1050	.0038
BIC $(K_L = K_U)$	.0084	.0097	.0031	.0035	.0089	.0166	.0032	.0062	.0958	.0031
$K_L = K_U = K_{\max}$	.0160	.0156	.0057	.0054	.0181	.0410	.0064	.0153	.1693	.0053
$K_L = K_U = 1$	.0094	.0092	.0033	.0033	.0103	.0240	.0037	.0088	.0912	.0032
$K_L = K_U = 2$	.0103	.0101	.0037	.0036	.0114	.0262	.0041	.0095	.0920	.0035
$K_L = K_U = 3$	.0114	.0111	.0041	.0040	.0129	.0310	.0046	.0109	.0956	.0039

Table 7: MSE of the Estimates

Part I: $a_{11} = \theta_{11} = 0$ (MSE$\times 10^2$, $T = 300$, ε_t: normal)

	$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$	$\sigma_{12} = .8$
(a_{11}, a_{22})	(0, .4)	(0, .8)	(0, .4)	(0, .8)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, .8)	(0, .8)
$(\theta_{11}, \theta_{22})$	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, .4)	(0, .8)	(0, .4)	(0, .8)	(0, .4)	(0, .4)
$(K_{\max} = K_4)$										
C_p	.0345	.3112	.0145	.1769	.0046	.0009	.0019	.0004	.1104	.0579
AIC	.0329	.3070	.0140	.1727	.0044	.0009	.0018	.0004	.1096	.0566
AIC$_c$	.0328	.3069	.0139	.1719	.0044	.0009	.0018	.0004	.1093	.0565
BIC	.0312	.3137	.0135	.1660	.0044	.0009	.0017	.0003	.1092	.0572
C_p $(K_L = K_U)$	.0346	.3192	.0147	.1814	.0046	.0010	.0020	.0004	.1130	.0596
AIC $(K_L = K_U)$	.0330	.3251	.0143	.1819	.0044	.0009	.0018	.0004	.1135	.0607
AIC$_c$ $(K_L = K_U)$	.0329	.3260	.0142	.1820	.0044	.0009	.0018	.0004	.1133	.0609
BIC $(K_L = K_U)$	.0322	.3409	.0143	.1910	.0048	.0013	.0018	.0004	.1128	.0695
$K_L = K_U = K_{\max}$	.0369	.3941	.0293	.5945	.0060	.0071	.0080	.0246	.1288	.1608
$(K_{\max} = K_{12})$										
C_p	.0490	.4356	.0211	.1905	.0064	.0013	.0027	.0005	.1564	.0682
AIC	.0377	.3852	.0163	.1722	.0047	.0010	.0020	.0004	.1322	.0607
AIC$_c$	.0361	.3720	.0157	.1675	.0046	.0009	.0019	.0004	.1275	.0591
BIC	.0312	.3236	.0135	.1511	.0044	.0009	.0017	.0003	.1106	.0556
C_p $(K_L = K_U)$	.0500	.4404	.0214	.1987	.0065	.0013	.0027	.0005	.1587	.0703
AIC $(K_L = K_U)$	.0355	.3827	.0158	.1877	.0046	.0009	.0019	.0004	.1292	.0652
AIC$_c$ $(K_L = K_U)$	.0343	.3691	.0151	.1841	.0045	.0009	.0019	.0004	.1240	.0641
BIC $(K_L = K_U)$	.0322	.3437	.0143	.1888	.0048	.0013	.0018	.0004	.1130	.0701
$K_L = K_U = K_{\max}$	.0459	.4862	.0365	.7217	.0077	.0088	.0097	.0294	.1596	.1965
$K_L = K_U = 1$	.0305	.3224	.0149	.3955	.0041	.0008	.0018	.0004	.1081	.1103
$K_L = K_U = 2$	.0313	.3106	.0135	.3017	.0043	.0009	.0018	.0004	.1066	.0874
$K_L = K_U = 3$	.0327	.3072	.0138	.2414	.0045	.0009	.0019	.0004	.1072	.0730

Part II: $a_{11} \neq 0$ and/or $\theta_{11} \neq 0$ (MSE$\times 10^2$, $T = 300$, ε_t: normal)

	$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$	$\sigma_{12} = 0.8$
(a_{11}, a_{22})	(.4, .4)	(.8, .8)	(.4, .4)	(.8, .8)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(.8, .8)	(.8, .8)
$(\theta_{11}, \theta_{22})$	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(.4, .4)	(.8, .8)	(.4, .4)	(.8, .8)	(.4, .4)	(.4, .4)
$(K_{\max} = K_4)$										
C_p	.0125	.0128	.0052	.0055	.0124	.0261	.0052	.0111	.1104	.0054
AIC	.0118	.0121	.0049	.0052	.0115	.0324	.0049	.0138	.1096	.0050
AIC$_c$	.0116	.0121	.0049	.0051	.0115	.0330	.0049	.0140	.1093	.0050
BIC	.0106	.0110	.0045	.0046	.0107	.0474	.0047	.0207	.1092	.0046
C_p $(K_L = K_U)$	.0124	.0127	.0052	.0054	.0124	.0306	.0053	.0131	.1130	.0053
AIC $(K_L = K_U)$	.0113	.0118	.0048	.0050	.0113	.0403	.0049	.0168	.1135	.0049
AIC$_c$ $(K_L = K_U)$	.0112	.0117	.0048	.0050	.0112	.0410	.0048	.0171	.1133	.0048
BIC $(K_L = K_U)$	.0105	.0107	.0044	.0045	.0107	.0526	.0047	.0226	.1128	.0045
$K_L = K_U = K_{\max}$	.0119	.0121	.0050	.0051	.0122	.0228	.0051	.0095	.1288	.0050
$(K_{\max} = K_{12})$										
C_p	.0181	.0193	.0076	.0081	.0169	.0294	.0071	.0118	.1564	.0080
AIC	.0135	.0157	.0056	.0067	.0123	.0306	.0052	.0132	.1322	.0062
AIC$_c$	.0128	.0151	.0054	.0064	.0120	.0309	.0051	.0133	.1275	.0059
BIC	.0106	.0112	.0045	.0047	.0107	.0474	.0047	.0207	.1106	.0046
C_p $(K_L = K_U)$	.0182	.0193	.0076	.0080	.0173	.0323	.0072	.0138	.1587	.0079
AIC $(K_L = K_U)$	.0120	.0144	.0052	.0061	.0117	.0395	.0050	.0166	.1292	.0056
AIC$_c$ $(K_L = K_U)$	.0115	.0135	.0050	.0057	.0113	.0404	.0049	.0169	.1240	.0053
BIC $(K_L = K_U)$	.0105	.0108	.0044	.0045	.0107	.0526	.0047	.0226	.1130	.0045
$K_L = K_U = K_{\max}$	.0151	.0152	.0064	.0064	.0155	.0245	.0066	.0102	.1596	.0064
$K_L = K_U = 1$	.0109	.0112	.0046	.0047	.0113	.0309	.0048	.0132	.1081	.0047
$K_L = K_U = 2$	.0114	.0118	.0048	.0050	.0118	.0257	.0049	.0109	.1066	.0049
$K_L = K_U = 3$	.0120	.0124	.0051	.0052	.0122	.0230	.0051	.0097	.1072	.0051

Table 8: MSE of the Estimates

Part I: $a_{11} = \theta_{11} = 0$ (MSE×10^5, $T = 300$, ε_t: log-normal)

	$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$		$\sigma_{12} = .8$		$\sigma_{12} = .4$	$\sigma_{12} = .8$
(a_{11}, a_{22})	(0, .4)	(0, .8)	(0, .4)	(0, .8)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, .8)	(0, .8)
$(\theta_{11}, \theta_{22})$	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, .4)	(0, .8)	(0, .4)	(0, .8)	(0, .4)	(0, .4)
$(K_{\max} = K_4)$										
C_p	.1041	.9804	.0379	.4089	.0135	.0022	.0049	.0008	.3496	.1400
AIC	.1010	.9678	.0369	.4027	.0131	.0021	.0047	.0008	.3453	.1377
AIC_c	.1008	.9670	.0368	.4018	.0131	.0022	.0047	.0008	.3446	.1376
BIC	.0984	.9680	.0362	.3930	.0131	.0021	.0047	.0008	.3453	.1362
C_p $(K_L = K_U)$	.1051	1.0073	.0388	.4229	.0137	.0023	.0049	.0008	.3576	.1450
AIC $(K_L = K_U)$	.1017	1.0089	.0379	.4233	.0133	.0022	.0048	.0008	.3559	.1454
AIC_c $(K_L = K_U)$	.1013	1.0082	.0378	.4233	.0133	.0022	.0048	.0008	.3554	.1455
BIC $(K_L = K_U)$	.0998	1.0206	.0374	.4279	.0134	.0024	.0048	.0008	.3548	.1495
$K_L = K_U = K_{\max}$	.1113	1.1701	.0529	.8979	.0159	.0093	.0098	.0205	.3940	.2617
$(K_{\max} = K_{12})$										
C_p	.1175	1.1857	.0435	.4308	.0147	.0026	.0054	.0009	.4128	.1512
AIC	.1014	1.0732	.0379	.3979	.0130	.0022	.0047	.0008	.3660	.1384
AIC_c	.1000	1.0503	.0373	.3910	.0129	.0021	.0047	.0008	.3590	.1359
BIC	.0956	.9593	.0355	.3657	.0128	.0021	.0046	.0007	.3369	.1302
C_p $(K_L = K_U)$	.1199	1.2028	.0447	.4510	.0153	.0028	.0056	.0010	.4162	.1586
AIC $(K_L = K_U)$	.1026	1.0976	.0388	.4302	.0133	.0023	.0049	.0008	.3739	.1490
AIC_c $(K_L = K_U)$	.1015	1.0704	.0379	.4206	.0132	.0022	.0048	.0008	.3652	.1465
BIC $(K_L = K_U)$	.0983	1.0130	.0371	.4136	.0132	.0023	.0047	.0008	.3512	.1468
$K_L = K_U = K_{\max}$	.1297	1.3728	.0635	1.1105	.0189	.0121	.0122	.0274	.4601	.3203
$K_L = K_U = 1$	.1010	1.0260	.0383	.6532	.0135	.0023	.0049	.0008	.3530	.1985
$K_L = K_U = 2$	.1035	.9990	.0375	.5523	.0140	.0024	.0050	.0008	.3489	.1740
$K_L = K_U = 3$	.1068	.9906	.0384	.4876	.0144	.0024	.0052	.0009	.3498	.1589

Part II: $a_{11} \neq 0$ and/or $\theta_{11} \neq 0$ (MSE×10^5, $T = 300$, ε_t: log-normal)

	$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$		$\sigma_{12} = 0.8$		$\sigma_{12} = 0.4$	$\sigma_{12} = 0.8$
(a_{11}, a_{22})	(.4, .4)	(.8, .8)	(.4, .4)	(.8, .8)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(.8, .8)	(.8, .8)
$(\theta_{11}, \theta_{22})$	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(.4, .4)	(.8, .8)	(.4, .4)	(.8, .8)	(.4, .4)	(.4, .4)
$(K_{\max} = K_4)$										
C_p	.0369	.0389	.0133	.0137	.0368	.0506	.0133	.0184	.3496	.0134
AIC	.0356	.0375	.0128	.0132	.0356	.0479	.0129	.0174	.3453	.0129
AIC_c	.0355	.0374	.0128	.0132	.0356	.0477	.0129	.0173	.3446	.0129
BIC	.0346	.0353	.0124	.0125	.0351	.0450	.0126	.0160	.3453	.0123
C_p $(K_L = K_U)$	.0368	.0385	.0133	.0136	.0369	.0512	.0133	.0185	.3576	.0134
AIC $(K_L = K_U)$	.0353	.0368	.0128	.0130	.0357	.0478	.0129	.0172	.3559	.0128
AIC_c $(K_L = K_U)$	.0353	.0367	.0127	.0129	.0357	.0477	.0129	.0170	.3554	.0127
BIC $(K_L = K_U)$	.0347	.0350	.0125	.0124	.0354	.0448	.0127	.0159	.3548	.0123
$K_L = K_U = K_{\max}$	.0382	.0378	.0137	.0133	.0394	.0507	.0141	.0181	.3940	.0133
$(K_{\max} = K_{12})$										
C_p	.0420	.0484	.0154	.0170	.0396	.0588	.0147	.0215	.4128	.0163
AIC	.0358	.0423	.0131	.0148	.0353	.0485	.0128	.0176	.3660	.0139
AIC_c	.0353	.0413	.0129	.0144	.0351	.0478	.0127	.0174	.3590	.0137
BIC	.0337	.0349	.0122	.0125	.0343	.0430	.0124	.0155	.3369	.0122
C_p $(K_L = K_U)$	.0424	.0476	.0156	.0170	.0415	.0642	.0153	.0232	.4162	.0164
AIC $(K_L = K_U)$	.0359	.0403	.0130	.0142	.0359	.0508	.0130	.0172	.3739	.0134
AIC_c $(K_L = K_U)$	.0354	.0390	.0128	.0137	.0355	.0482	.0129	.0170	.3652	.0132
BIC $(K_L = K_U)$	.0343	.0348	.0124	.0124	.0350	.0434	.0127	.0158	.3512	.0123
$K_L = K_U = K_{\max}$	.0444	.0439	.0157	.0153	.0461	.0647	.0164	.0237	.4601	.0153
$K_L = K_U = 1$	.0363	.0360	.0130	.0128	.0374	.0537	.0134	.0193	.3530	.0127
$K_L = K_U = 2$	.0375	.0371	.0134	.0132	.0386	.0547	.0137	.0196	.3489	.0131
$K_L = K_U = 3$	.0387	.0382	.0139	.0137	.0399	.0573	.0143	.0205	.3498	.0136