On least-squares bias in the AR(p) models: Bias correction using the bootstrap methods

Tanizaki, Hisashi
Hamori, Shigeyuki
Matsubayashi, Yoichi

(Citation)
Statistical Papers,47(1):109-124

(Issue Date)
2006-01

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000128

On Least-Squares Bias in the AR(p) Models: Bias Correction Using the Bootstrap Methods*

Hisashi Tanizaki, Shigeyuki Hamori and Yoichi Matsubayashi

Graduate School of Economics, Kobe University, Kobe 657-8501, Japan.

Abstract In the case where the lagged dependent variables are included in the regression model, it is known that the ordinary least squares estimates (OLSE) are biased in small sample and that the bias increases as the number of the irrelevant variables increases. In this paper, based on the bootstrap methods, an attempt is made to obtain the unbiased estimates in autoregressive and non-Gaussian cases. We propose the residual-based bootstrap method in this paper. Some simulation studies are performed to examine whether the proposed estimation procedure works well or not. We obtain the results that it is possible to recover the true parameter values and that the proposed procedure gives us the less biased estimators than OLSE.

Key words AR(p) Model, OLSE, Unbiased Estimator, Exogenous Variables, Nonnormal Error, Bootstrap Method.

1 Introduction

In the case where the lagged dependent variables are included in the regression model, it is known that the OLSEs of autoregressive (AR) models are biased in small sample. Hurwicz (1950), Marriott and Pope (1954), Kendall (1954) and

* This paper is a substantial revision of Tanizaki (2000). The normality assumption is adopted in Tanizaki (2000), but it is not required in this paper. The authors are grateful to an anonymous referee for valuable suggestions and comments. This research was partially supported by Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C)(2) #14530033, 2002–2005, for H. Tanizaki and Grants-in-Aid for the 21st Century COE Program.

Correspondence to: Hisashi Tanizaki, Graduate School of Economics, Kobe University, Rokkodai 2-1, Nadaku, Kobe 657-8501, Japan; e-mail: tanizaki@kobe-u.ac.jp

White (1961) discussed the mean-bias of the OLSE. Quenouille (1956) introduced the jackknife estimator of the AR parameter which is median-unbiased to order $1/T$ as T goes to infinity, where the trend term is not taken into account. Orcutt and Winokur (1969) constructed approximately mean-unbiased estimates of the AR parameter in stationary models. Sawa (1978), Tanaka (1983) and Tsui and Ali (1994) also examined the AR(1) models, where the exact moments of OLSE are discussed. Shaman and Stine (1988) established the mean-bias of the OLSE to order $1/T$ in stationary AR(p) (also see Maekawa (1987) for the AR(p) models). Grubb and Symons (1987) gave an expression to order $1/T$ for bias to the estimated coefficient on a lagged dependent variable when all other regressors are exogenous (also see Tse (1982) and Maekawa (1983) for the AR models including the exogenous variables). Peters (1989) studied the finite sample sensitivity of OLSE of the AR(1) term with nonnormal errors. In Abadir (1993), an analytical formula was derived to approximate the finite sample bias of OLSE of the AR(1) term when the underlying process has a unit root.

Moreover, in the case where the true model is the first-order AR model, Andrews (1993) examined the cases where the estimated models are the AR(1), the AR(1) with a constant term and the AR(1) with constant ane trend terms, where the exact median-unbiased estimator of the first-order autoregressive model is derived by utilizing the Imhof (1961) algorithm. Andrews and Chen (1994) obtained the approximately median-unbiased estimator of autoregressive models, where Andrews (1993) is applied by transforming AR(p) models into AR(1) and taking the iterative procedure.

Thus, the AR models have been studied with respect to various aspects, i.e., (i) a stationary model or a unit root model, (ii) the first-order autoregressive model or the higher-order autoregressive models, (iii) an autoregressive model with or without exogenous variables, and (iv) a normal error or a nonnormal error. Tanizaki (2000) proposed the median- and mean-unbiased estimators using simulation techniques, where the underlying assumption is that the error term is normal. In this paper, in more general formulation which can be applied to all the cases (i) – (iv), using the bootstrap methods we derive the unbiased estimates of the regression coefficients in small sample.

2 Bias Correction Method

We take the autoregressive model which may include the exogenous variables, say x_t. That is, consider the following simple regression model:

$$y_t = x_t\beta + \sum_{j=1}^{p} \alpha_j y_{t-j} + u_t = z_t\theta + u_t, \tag{1}$$

for $t = p+1, p+2, \cdots, T$, where x_t and β are a $1 \times k$ vector and a $k \times 1$ vector, respectively. θ and z_t are given by $\theta = (\beta', \alpha')'$ and $z_t = (x_t, y_{t-1}, y_{t-2}, \cdots, y_{t-p})$, where $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_p)'$. u_t is assumed to be distributed with mean zero and variance σ^2. We will discuss later for the distribution function of the error term

u_t. In this paper, the initial values $y_p, y_{p-1}, \cdots, y_1$ are assumed to be constant for simplicity.

Since it is well known that OLSE of the autoregressive coefficient vector in the AR(p) model is biased in small sample (see, for example, Andrews (1993), Andrews and Chen (1994), Diebold and Rudebusch (1991), Hurwicz (1959), Kendall (1954), Marriott and Pope (1954), Quenouille (1956) and so on), OLSE of θ, $\hat{\theta} = (\hat{\beta}', \hat{\alpha}')'$, is clearly biased.

To obtain the unbiased estimator of θ, the underlying idea in this paper is described as follows. Let θ be an unknown parameter and $\hat{\theta}$ be the biased estimate of θ. Suppose that the distribution function of $\hat{\theta}$ is given by $f_{\hat{\theta}}(\cdot)$, which is not obtained analytically in the case where the lagged dependent variables are included in the explanatory variables. Since $\hat{\theta}$ is biased, we have $\theta \neq \mathrm{E}(\hat{\theta})$, where the expectation $\mathrm{E}(\hat{\theta})$ is defined as follows:

$$\mathrm{E}(\hat{\theta}) \equiv \int_{-\infty}^{+\infty} x f_{\hat{\theta}}(x) \mathrm{d}x. \tag{2}$$

To obtain the relationship between $\hat{\theta}$ and θ, let $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \cdots, \hat{\theta}_n^*\}$ be a sequence of the biased estimates of θ, which are taken as the random draws generated from $f_{\hat{\theta}}(\cdot)$. Note that $\hat{\theta}$ implies the OLSE obtained from the actual data while $\hat{\theta}_i^*$ denotes the ith OLSE based on the simulated data given the true parameter value θ. Therefore, $\hat{\theta}_i^*$ depends on θ, i.e., $\hat{\theta}_i^* = \hat{\theta}_i^*(\theta)$ for all $i = 1, 2, \cdots, n$. Suppose that given θ we can generate the n random draws $\hat{\theta}_1^*, \hat{\theta}_2^*, \cdots, \hat{\theta}_n^*$ from $f_{\hat{\theta}}(\cdot)$. Using the n random draws, the integration in equation (2) can be represented as follows:

$$\mathrm{E}(\hat{\theta}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i^*(\theta). \tag{3}$$

Let us define the unbiased estimator of θ as $\bar{\theta}$. Equation (3) implies that $\bar{\theta}$ is given by the θ which satisfies the following equation:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i^*(\theta) \equiv g(\theta), \tag{4}$$

where $\hat{\theta}$ in the left-hand side represents the OLSE of θ based on the original data y_t and x_t. $g(\cdot)$ denotes a $(k + p) \times 1$ vector function, which is defined as equation (4). The solution of θ obtained from equation (4) is denoted by $\bar{\theta}$, which corresponds to the unbiased estimator of θ. Conventionally, it is impossible to obtain an explicit functional form of $g(\cdot)$ in equation (4). Therefore, equation (4) is numerically solved by an iterative procedure or a simple grid search. In this paper, the computational implementation is shown as the following iterative procedure.

1. Given the actual time series data (i.e., x_t and y_t), estimate θ and σ^2 in equation (1) by OLS, Let us denote the OLS estimates of θ and σ^2 by $\hat{\theta} = (\sum_{t=p+1}^{T} z_t' z_t)^{-1}$ $(\sum_{t=p+1}^{T} z_t' y_t)$ and $\hat{\sigma}^2 = \sum_{t=p+1}^{T} (y_t - z_t\hat{\theta})^2/((T - p) - (k + p))$.
2. Let u_t^* be the random draw with mean zero and variance $\hat{\sigma}^2$. Suppose for now that the random draws $u_{p+1}^*, u_{p+2}^*, \cdots, u_T^*$ are available. We will discuss later for the random number generation method of u_t^*.

3. Let $\theta^{(j)}$ be the jth iteration of θ and y_t^* be the random draw of y_t. Given the initial values $\{y_p^*, y_{p-1}^*, \cdots, y_1^*\}$, the exogenous variable x_t for $t = p+1, p+2, \cdots, T$ and $\theta^{(j)}$, we obtain y_t^* substituting u_t^* into u_t in equation (1), where the initial values $\{y_p^*, y_{p-1}^*, \cdots, y_1^*\}$ may be taken as the actual data $\{y_p, y_{p-1}, \cdots, y_1\}$. That is, y_t^* is generated from $z_t^*\theta^{(j)} + u_t^*$ given $\theta^{(j)}$ and the random draw u_t^*, where $z_t^* = (x_t, y_{t-1}^*, y_{t-2}^*, \cdots, y_{t-p}^*)$. For $j = 1$ we may set $\theta^{(1)} = \hat{\theta}$.
4. Given $\{y_1^*, y_2^*, \cdots, y_T^*\}$ and $\{x_{p+1}, x_{p+2}, \cdots, x_T\}$, compute the OLSE of $\theta^{(j)}$, which is denoted by $\hat{\theta}^*$.
5. Repeat Steps 2 – 4 n times, where $n = 10^4$ is taken in this paper. Then, n OLSEs of $\theta^{(j)}$ are obtained, which correspond to $\hat{\theta}_i^*(\theta^{(j)})$, $i = 1, 2, \cdots, n$, in equation (4). Based on the n OLSEs of $\theta^{(j)}$, compute the arithmetic mean for each element of θ, i.e., the function $g(\theta^{(j)})$.
6. As in equation (4), $\hat{\theta}$ should be equal to the arithmetic average $g(\theta^{(j)}) \equiv (1/n)\sum_{i=1}^{n}\hat{\theta}_i^*(\theta^{(j)})$. For each element of θ, therefore, $\theta^{(j+1)}$ should be smaller than $\theta^{(j)}$ if $\hat{\theta}$ is less than the arithmetic average, and it should be larger than $\theta^{(j)}$ otherwise. Here, we consider that each element of $g(\theta)$ is a monotone increasing function of the corresponding element of θ. Thus, $\theta^{(j)}$ is updated to $\theta^{(j+1)}$. The optimization procedure taken in this paper is described in Appendix 2.
7. Repeat Steps 2 – 6 until $\theta^{(j+1)}$ is stable, where the limit of $\theta^{(j+1)}$ is taken as $\bar{\theta}$. Note that the random draws of u_t, generated in Step 2, should be same for all j, i.e., $n \times (T-p)$ random draws are required.

Now, in Step 2 we need to consider generating the random draw of u_t, i.e., u_t^*, which is assumed to be distributed with mean zero and variance $\hat{\sigma}^2$. In the regression model (1), the underlying distribution of u_t is conventionally unknown. To examine whether the suggested procedure is robust or not, using the bootstrap methods we take the following four types of random draws for u_t^*, i.e., the normal error (N), the chi-squared error (X), the uniform error (U) and the residual-based error (R).

(N) $u_t^* = \hat{\sigma}\epsilon_t$, where $\epsilon_t \sim N(0,1)$ and $\hat{\sigma}$ denotes the standard error of regression by OLS.
(X) $u_t^* = \hat{\sigma}\epsilon_t$, where $\epsilon_t = (v_t - 1)/\sqrt{2}$ and $v_t \sim \chi^2(1)$.
(U) $u_t^* = \hat{\sigma}\epsilon_t$, where $\epsilon_t = 2\sqrt{3}(v_t - .5)$ and $v_t \sim U(0,1)$.
(R) u_t^* is resampled from $\{c\hat{u}_{p+1}, c\hat{u}_{p+2}, \cdots, c\hat{u}_T\}$ with equal probability $1/(T-p)$, where $\hat{u}_t$ denotes the OLS residual at time t, i.e., $\hat{u}_t = y_t - z_t\hat{\theta}$, and $c = \sqrt{(T-p)/((T-p)-(k+p))}$ is taken (see Wu (1986) for c). Note as follows. If the constant term is not included in the regression model, $\sum_{t=p+1}^{n}\hat{u}_t = 0$ is not guaranteed, which contradicts the assumption of mean zero error. Accordingly, in this case, $\hat{u}_t$ should be replaced by $\hat{\sigma}(\hat{u}_t - \bar{u}) \Big/ \sqrt{\sum_{t=p+1}^{T}(\hat{u}_t - \bar{u})^2/(T-p)}$, which represents mean zero and variance $\hat{\sigma}^2$, where $\bar{u} = \sum_{t=p+1}^{T}\hat{u}_t/(T-p)$.

Thus, for $t = p+1, p+2, \cdots, T$, it is necessary to generate the random draws of u_t in Step 2. In practice we often have the case where the underlying distribution of the true data series is different from that of the simulated one, because the distribution of u_t is not known. (R) does not assume any distribution for u_t. Therefore, it might be expected that (R) is more robust, compared with (N), (X) and (U).

In the next section, for the true distribution of u_t we consider the three types of the error term, i.e., the standard normal error, the chi-squared error and the uniform error. For each of the three errors, (N), (X), (U) and (R) are examined through Monte Carlo experiments.

3 Monte Carlo Experiments

3.1 AR(1) Models

Let Model A be the case of $k = 0$, Model B be the case of $k = 1$ and $x_t = 1$ and Model C be the case of $k = 2$ and $x_t = (1,\ x_{1t})$, i.e,

$$\text{Model A: } y_t = \sum_{j=1}^{p} \alpha_j y_{t-j} + u_t,$$

$$\text{Model B: } y_t = \beta_1 + \sum_{j=1}^{p} \alpha_j y_{t-j} + u_t,$$

$$\text{Model C: } y_t = \beta_1 + \beta_2 x_{1t} + \sum_{j=1}^{p} \alpha_j y_{t-j} + u_t,$$

for $t = p+1, p+2, \ldots, T$, given the initial condition $y_1 = y_2 = \cdots = y_p = 0$. In Model C, we take x_{1t} as the trend term, i.e., $x_{1t} = t$.

The true distribution of the error term u_t is assumed to be normal in Table 1, chi-squared in Table 2 and uniform in Table 3, respectively. For all the tables, mean and variance of the error are normalized to be zero and one. Since the true distribution of the error term is not known in practice, we examine (N), (X), (U) and (R) for Models A – C.

The case of $p = 1$ is examined in Tables 1 – 3, although the suggested procedure would be applied to any p (see Table 4 for $p = 2$ and Table 5 for $p = 3$). The sample size is $T = 20, 40$ in Table 1 and $T = 20$ in Tables 2 and 3. For the parameter values, $\alpha_1 = 0.6, 0.9, 1.0$, $\beta_1 = 0.0, 1.0, 2.0$ and $\beta_2 = 0.0$ are taken. We perform 10^3 simulation runs. The arithmetic averages from the 10^3 estimates of α_1 are shown in Tables 1 – 3. The values in the parentheses are the root mean square errors (RMSE) from the 10^3 estimates. In Tables 1 – 3, (O) represents OLSE, and (O)′ denotes the bias-corrected estimate based on Grubb and Symons (1987), which is discussed in Appendix 3. (R)′ indicates the bias-corrected estimate suggested by Efron and Tibshirani (1993), which is discribed in Remark of Appendix 2. (N), (X), (U) and (R) are discussed in Section 2. We compare (O), (O)′, (N), (X), (U), (R) and (R)′ in Table 1, and (O), (N) and (R) in Tables 2 and 3. (1) – (5) in each table denote as follows:

(1) The true model is $\alpha_1 = 0.6, 0.9, 1.0$ and $(\beta_1, \beta_2) = (0, 0)$, i.e., Model A, while the estimated model is Model A.
(2) The true model is $\alpha_1 = 0.6, 0.9, 1.0$ and $(\beta_1, \beta_2) = (0, 0)$, i.e., Model A, while the estimated model is Model B.
(3) The true model is $\alpha_1 = 0.6, 0.9, 1.0$ and $(\beta_1, \beta_2) = (1, 0)$, i.e., Model B, while the estimated model is Model B.

Table 1 Estimates of α_1: Case of $\beta_2 = 0$ and $N(0,1)$ Error

T	Est.	Model	(1) A	(2) B	(3) B	(4) B	(5) C
		$\alpha_1 \setminus \beta_1$	0.0	0.0	1.0	2.0	0.0
20	(O)	0.6	.537 (.209)	.435 (.274)	.471 (.233)	.524 (.162)	.336 (.350)
	(O)′		.600 (.223)	.579 (.260)	.622 (.231)	.684 (.190)	.558 (.294)
	(N)		.592 (.219)	.594 (.269)	.588 (.226)	.591 (.156)	.593 (.321)
	(U)		.564 (.225)	.573 (.264)	.572 (.227)	.581 (.155)	.567 (.314)
	(X)		.591 (.221)	.588 (.271)	.585 (.228)	.589 (.157)	.585 (.323)
	(R)		.591 (.219)	.591 (.269)	.587 (.226)	.591 (.156)	.590 (.322)
	(R)′		.587 (.218)	.561 (.260)	.570 (.221)	.585 (.155)	.516 (.295)
	(O)	0.9	.818 (.181)	.662 (.312)	.837 (.117)	.882 (.048)	.518 (.445)
	(O)′		.914 (.181)	.848 (.244)	1.056 (.195)	1.110 (.217)	.790 (.311)
	(N)		.892 (.170)	.838 (.236)	.898 (.103)	.899 (.045)	.800 (.311)
	(U)		.876 (.179)	.820 (.244)	.891 (.103)	.896 (.044)	.772 (.319)
	(X)		.893 (.171)	.837 (.239)	.899 (.104)	.899 (.045)	.796 (.316)
	(R)		.892 (.170)	.837 (.237)	.899 (.103)	.899 (.045)	.798 (.313)
	(R)′		.891 (.174)	.815 (.251)	.894 (.103)	.899 (.045)	.724 (.331)
	(O)	1.0	.919 (.167)	.753 (.315)	.982 (.048)	.996 (.021)	.544 (.512)
	(O)′		1.027 (.165)	.957 (.235)	1.229 (.234)	1.245 (.246)	.823 (.344)
	(N)		.990 (.146)	.901 (.224)	1.005 (.050)	.999 (.020)	.818 (.341)
	(U)		.980 (.155)	.887 (.234)	.999 (.047)	.998 (.020)	.791 (.355)
	(X)		.991 (.146)	.900 (.227)	1.006 (.050)	1.000 (.020)	.815 (.345)
	(R)		.991 (.146)	.900 (.225)	1.006 (.050)	.999 (.020)	.817 (.342)
	(R)′		.995 (.152)	.890 (.241)	1.003 (.049)	.999 (.020)	.749 (.376)
40	(O)	0.6	.574 (.135)	.527 (.162)	.539 (.146)	.558 (.118)	.475 (.196)
	(O)′		.606 (.140)	.598 (.156)	.611 (.144)	.632 (.124)	.586 (.169)
	(N)		.606 (.140)	.604 (.160)	.605 (.144)	.604 (.117)	.599 (.178)
	(U)		.592 (.140)	.598 (.158)	.598 (.143)	.600 (.116)	.592 (.176)
	(X)		.603 (.140)	.602 (.161)	.603 (.145)	.603 (.117)	.597 (.179)
	(R)		.604 (.140)	.603 (.160)	.603 (.144)	.603 (.117)	.598 (.178)
	(R)′		.602 (.139)	.595 (.157)	.598 (.142)	.601 (.116)	.580 (.171)
	(O)	0.9	.861 (.103)	.789 (.162)	.866 (.070)	.890 (.032)	.713 (.231)
	(O)′		.907 (.101)	.883 (.129)	.966 (.093)	.992 (.097)	.852 (.159)
	(N)		.904 (.098)	.893 (.130)	.903 (.064)	.901 (.031)	.874 (.165)
	(U)		.895 (.101)	.886 (.132)	.902 (.064)	.901 (.031)	.865 (.167)
	(X)		.903 (.099)	.892 (.131)	.904 (.064)	.902 (.031)	.874 (.166)
	(R)		.902 (.099)	.892 (.130)	.903 (.064)	.901 (.031)	.874 (.165)
	(R)′		.902 (.100)	.882 (.133)	.901 (.063)	.901 (.031)	.843 (.166)
	(O)	1.0	.958 (.089)	.874 (.162)	.996 (.015)	.999 (.008)	.763 (.272)
	(O)′		1.010 (.083)	.975 (.114)	1.107 (.108)	1.110 (.111)	.907 (.175)
	(N)		.997 (.076)	.954 (.108)	1.001 (.016)	1.000 (.007)	.916 (.170)
	(U)		.992 (.079)	.950 (.112)	1.000 (.016)	1.000 (.007)	.907 (.175)
	(X)		.996 (.076)	.954 (.109)	1.001 (.016)	1.000 (.007)	.915 (.170)
	(R)		.996 (.076)	.954 (.108)	1.001 (.016)	1.000 (.007)	.915 (.170)
	(R)′		.999 (.079)	.953 (.116)	1.001 (.016)	1.000 (.007)	.891 (.185)

Table 2 Estimates of α_1: Case of $\beta_2 = 0$, $(\chi^2(1) - 1)/\sqrt{2}$ Error and $T = 20$

Est. Model		(1) A	(2) B	(3) B	(4) B	(5) C
	$\alpha_1 \setminus \beta_1$	0.0	0.0	1.0	2.0	0.0
(O)		.575 (.197)	.454 (.237)	.493 (.202)	.537 (.162)	.352 (.319)
(O)′		.642 (.222)	.602 (.222)	.648 (.209)	.700 (.203)	.580 (.257)
(N)		.632 (.215)	.613 (.229)	.606 (.197)	.597 (.161)	.608 (.280)
(U)	0.6	.605 (.217)	.590 (.224)	.590 (.197)	.588 (.160)	.579 (.272)
(X)		.632 (.216)	.607 (.231)	.603 (.199)	.595 (.162)	.601 (.281)
(R)		.612 (.218)	.598 (.224)	.597 (.198)	.592 (.160)	.593 (.274)
(R)′		.609 (.216)	.572 (.217)	.582 (.192)	.588 (.159)	.528 (.254)
(O)		.841 (.163)	.691 (.274)	.848 (.124)	.886 (.052)	.540 (.416)
(O)′		.940 (.174)	.883 (.211)	1.070 (.217)	1.114 (.222)	.817 (.277)
(N)		.915 (.158)	.864 (.197)	.899 (.121)	.900 (.050)	.825 (.276)
(U)	0.9	.900 (.165)	.845 (.205)	.892 (.119)	.898 (.049)	.794 (.285)
(X)		.916 (.159)	.863 (.200)	.901 (.122)	.900 (.050)	.822 (.280)
(R)		.905 (.161)	.855 (.200)	.897 (.120)	.899 (.049)	.815 (.278)
(R)′		.903 (.165)	.833 (.211)	.892 (.118)	.899 (.049)	.740 (.295)
(O)		.936 (.149)	.780 (.284)	.988 (.045)	.997 (.021)	.565 (.484)
(O)′		1.046 (.157)	.989 (.214)	1.236 (.242)	1.247 (.248)	.850 (.309)
(N)		1.007 (.132)	.922 (.192)	1.006 (.049)	1.000 (.021)	.850 (.308)
(U)	1.0	.997 (.139)	.910 (.202)	1.000 (.046)	1.000 (.020)	.822 (.323)
(X)		1.008 (.132)	.922 (.194)	1.007 (.049)	1.001 (.021)	.848 (.312)
(R)		1.001 (.136)	.919 (.195)	1.003 (.047)	1.000 (.020)	.843 (.312)
(R)′		1.004 (.142)	.906 (.210)	1.001 (.045)	1.000 (.020)	.771 (.344)

Table 3 Estimates of α_1: Case of $\beta_2 = 0$, $U(-\sqrt{3}, \sqrt{3})$ Error and $T = 20$

Est. Model		(1) A	(2) B	(3) B	(4) B	(5) C
	$\alpha_1 \setminus \beta_1$	0.0	0.0	1.0	2.0	0.0
(O)		.541 (.207)	.441 (.271)	.474 (.233)	.525 (.164)	.343 (.348)
(O)′		.605 (.222)	.586 (.261)	.626 (.234)	.687 (.193)	.568 (.298)
(N)		.596 (.218)	.600 (.269)	.595 (.229)	.595 (.159)	.602 (.324)
(U)	0.6	.568 (.224)	.580 (.265)	.579 (.230)	.585 (.158)	.575 (.316)
(X)		.596 (.220)	.595 (.271)	.592 (.231)	.593 (.160)	.594 (.326)
(R)		.598 (.219)	.600 (.270)	.596 (.230)	.596 (.159)	.600 (.325)
(R)′		.594 (.218)	.569 (.260)	.578 (.224)	.590 (.158)	.527 (.299)
(O)		.819 (.177)	.668 (.304)	.837 (.120)	.883 (.049)	.526 (.438)
(O)′		.915 (.177)	.855 (.237)	1.056 (.197)	1.111 (.218)	.799 (.305)
(N)		.893 (.166)	.845 (.227)	.901 (.107)	.900 (.046)	.809 (.303)
(U)	0.9	.877 (.175)	.826 (.236)	.893 (.106)	.897 (.046)	.781 (.312)
(X)		.894 (.166)	.843 (.230)	.902 (.108)	.900 (.046)	.806 (.307)
(R)		.896 (.166)	.845 (.228)	.903 (.108)	.900 (.046)	.810 (.304)
(R)′		.894 (.169)	.824 (.243)	.898 (.107)	.900 (.046)	.734 (.322)
(O)		.918 (.166)	.760 (.300)	.983 (.049)	.996 (.021)	.550 (.505)
(O)′		1.026 (.163)	.964 (.217)	1.229 (.235)	1.245 (.247)	.830 (.337)
(N)		.990 (.145)	.911 (.203)	1.006 (.051)	1.000 (.021)	.826 (.332)
(U)	1.0	.979 (.154)	.896 (.214)	.999 (.047)	.999 (.020)	.799 (.347)
(X)		.991 (.145)	.910 (.205)	1.008 (.051)	1.000 (.021)	.824 (.337)
(R)		.992 (.145)	.912 (.204)	1.007 (.051)	1.000 (.021)	.827 (.333)
(R)′		.996 (.151)	.900 (.221)	1.004 (.049)	1.000 (.020)	.758 (.368)

(4) The true model is $\alpha_1 = 0.6, 0.9, 1.0$ and $(\beta_1, \beta_2) = (2, 0)$, i.e., Model B, while the estimated model is Model B.
(5) The true model is $\alpha_1 = 0.6, 0.9, 1.0$ and $(\beta_1, \beta_2) = (0, 0)$, i.e., Model A, while the estimated model is Model C.

Suppose that the true model is represented by Model A with $p = 1$ and that $x_{1t} = t$ is taken in Model C. Then, it is well known that the OLSE of α_1 estimated by Model C gives us the largest bias and the OLSE of α_1 estimated by Model A yields the smallest one (see, for example, Andrews (1993)). That is, the OLSE bias of the AR(1) coefficient increases as the number of exogenous variables increases. In order to check this fact, first we compare (1), (2) and (5) with respect to (O). Note that (O) represents the arithmetic average and the root mean square error from the 10^3 OLSEs. For (O) in Table 1, in the case of $T = 20$ and $\alpha_1 = 0.6$, (1) is 0.537, (2) is 0.435 and (5) is 0.336. For all the cases of T and α_1 in Tables 1 – 3, (O) is biased as the number of exogenous variables increases.

Next, we compare (2) – (4), taking the case $T = 20$ in Table 1, where the true model is Model A or B while the estimated model is Model B. We examine whether the intercept influences precision of OLSE. The results are as follows. When the intercept increases, the OLSE approaches the true parameter value, i.e., 0.435 for (O) in (2), 0.471 for (O) in (3) and 0.524 for (O) in (4), and in addition, the root mean square error (RMSE) of the OLSE is small, i.e., 0.274 for (O) in (2), 0.233 for (O) in (3) and 0.162 for (O) in (4). Thus, as the intercept increases the better OLSE is obtained. As for (O), the similar results are obtained in both Tables 2 and 3.

The error term is assumed to be normal in Table 1, chi-squared in Table 2 and uniform in Table 3. It is observed from Tables 1 – 3 that the bias of OLSE depends on the underlying distribution of the error term. That is, the OLSE with the uniform error is very close to the OLSE with the normal error (see (O) in Tables 1 and 3), while the OLSE with the chi-squared error yields the smallest bias and RMSE of the three (see (O) in each table).

In Table 1, under the condition that for the true distribution of the error term u_t is normal, we compute the unbiased estimate of the AR(1) coefficient assuming that the error term follows the normal distribution (N), the chi-square distribution (X), the uniform distribution (U) and the residual-based distribution (R). Accordingly, it might be expected that (N) shows the best performance, because the estimated model is consistent with the underlying true one. Similarly, (X) in Table 2 and (U) in Table 3 should be better than any other procedures. That is, the best estimator should be (N) in Table 1, (X) in Table 2 and (U) in Table 3. In Table 1, as expected, (N) shows the best estimator because (N) is close to the true parameter values in almost all the cases. However, (X), (U) and (R) are not too different from (N). We cannot see that (X) is the best in Table 2 and that (U) is the best in Table 3. It is clear that (N),(X), (U) and (R) are much better than (O). (R) is distribution-free, while (N), (X) and (U) are based on the distribution of the error term. The underlying distribution is conventionally unknown. In the three tables, (R) is not necessarily the best estimator, but it is relatively good. Therefore, it might be better to choose (R), rather tan the distribution-based error terms such as (N), (X) and (U).

Table 4 AR(2) Model: $N(0,1)$ Error and $T = 20$

(λ_1, λ_2)	(α_1, α_2)		Est. of α_1	Est. of α_2
(0.0, 0.0)	(0.0, 0.00)	(O)	−.008 (.248)	−.062 (.250)
		(N)	−.009 (.262)	−.008 (.282)
		(R)	−.010 (.262)	−.009 (.282)
(0.5, 0.0)	(0.5, 0.00)	(O)	.465 (.249)	−.051 (.242)
		(N)	.490 (.260)	.003 (.274)
		(R)	.489 (.261)	.002 (.274)
(1.0, 0.0)	(1.0, 0.00)	(O)	.933 (.254)	−.020 (.248)
		(N)	.988 (.266)	.002 (.280)
		(R)	.989 (.266)	.002 (.280)
(0.5, 0.5)	(1.0,−0.25)	(O)	.937 (.245)	−.254 (.228)
		(N)	.989 (.251)	−.239 (.262)
		(R)	.989 (.251)	−.239 (.262)
(1.0, 0.5)	(1.5,−0.50)	(O)	1.393 (.252)	−.438 (.243)
		(N)	1.487 (.247)	−.492 (.263)
		(R)	1.488 (.248)	−.493 (.263)
(1.0, 1.0)	(2.0,−1.00)	(O)	1.863 (.246)	−.865 (.260)
		(N)	1.985 (.208)	−.986 (.230)
		(R)	1.986 (.207)	−.988 (.229)

(O)′ denotes the bias-corrected estimate based on OLSE, which is discussed in Appendix 3. Also, see Grubb and Symons (1987). (O)′ gives us the asymptotic result and accordingly it is expected that (O)′ still has bias and large RMSE. As a result, in Tables 1 – 3 we have a lot of cases where (O)′ is overestimated. Moreover, RMSE is extremely large for (3) and (4) in Tables 1 – 3, because of large bias. Thus, we can conclude that in small sample (O)′ is not useful in practice. Furthermore, as discussed in Appendix 3, (O)′ cannot be applied to higher-order AR models, which implies that (O)′ is quite restrictive.

(R)′ corresponds to the bias-corrected estimate based on Efron and Tibshirani (1993, Chap.10), which is discussed in Remark of Appendix 2. Using the OLS residuals $\{c\hat{u}_1, c\hat{u}_2, \cdots, c\hat{u}_T\}$ with $c = \sqrt{(T-p)/((T-p)-(k+p))}$, (R) is obtained by the iterative procedure shown in Steps 1 – 7 of Section 2. (R)′ is equivalent to the case of $j = 1$, i.e., $\theta^{(2)}$, in Steps 1 –7, where the iterative procedure is not taken. In this paper, we suggest taking the iterative procedure to obtain less biased estimate. In Tables 1 – 3, (R)′ is much less biased than (O), but (R) is more unbiased and more efficient in the sense of the smaller RMSE, compared with (R)′.

3.2 AR(p) Models

In this section, we consider the AR(p) models, where $p = 2, 3$ is taken. Assume that the true model is represented by Model A, i.e.,

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + u_t,$$

for $t = p+1, p+2, \cdots, T$, where u_t is assumed to be distributed as a standard normal random variable and the initial values are given by $y_1 = y_2 = \cdots = y_p = 0$.

Table 5 AR(3) Models: $N(0,1)$ Error and $T = 20$

$(\lambda_1, \lambda_2, \lambda_3)$	$(\alpha_1, \alpha_2, \alpha_3)$		Est. of α_1	Est. of α_2	Est. of α_3
(0.0, 0.0, 0.0)	(0.0, 0.00, 0.000)	(O)	−.006 (.261)	−.073 (.265)	.005 (.261)
		(N)	−.005 (.277)	−.016 (.297)	.006 (.317)
		(R)	−.010 (.277)	−.021 (.297)	.003 (.316)
(0.5, 0.0, 0.0)	(0.5, 0.00, 0.000)	(O)	.466 (.266)	−.069 (.292)	.021 (.268)
		(N)	.496 (.281)	−.013 (.332)	.021 (.324)
		(R)	.491 (.281)	−.014 (.331)	.019 (.323)
(1.0, 0.0, 0.0)	(1.0, 0.00, 0.000)	(O)	.931 (.280)	−.062 (.355)	.057 (.290)
		(N)	.993 (.289)	−.014 (.405)	.019 (.338)
		(R)	.988 (.289)	−.013 (.404)	.022 (.337)
(0.5, 0.5, 0.0)	(1.0,−0.25, 0.000)	(O)	.938 (.274)	−.278 (.344)	.027 (.275)
		(N)	.997 (.286)	−.262 (.402)	.022 (.328)
		(R)	.991 (.286)	−.259 (.401)	.022 (.328)
(1.0, 0.5, 0.0)	(1.5,−0.50, 0.000)	(O)	1.399 (.288)	−.483 (.419)	.049 (.294)
		(N)	1.491 (.290)	−.505 (.486)	.016 (.341)
		(R)	1.485 (.290)	−.501 (.485)	.017 (.340)
(1.0, 1.0, 0.0)	(2.0,−1.00, 0.000)	(O)	1.853 (.306)	−.864 (.511)	.016 (.317)
		(N)	1.979 (.297)	−.980 (.579)	.005 (.366)
		(R)	1.973 (.297)	−.973 (.578)	.006 (.365)
(0.5, 0.5, 0.5)	(1.5,−0.75, 0.125)	(O)	1.403 (.283)	−.693 (.407)	.120 (.281)
		(N)	1.494 (.288)	−.752 (.470)	.138 (.331)
		(R)	1.488 (.288)	−.747 (.469)	.137 (.331)
(1.0, 0.5, 0.5)	(2.0,−1.25, 0.250)	(O)	1.857 (.300)	−1.098 (.495)	.223 (.300)
		(N)	1.984 (.288)	−1.234 (.544)	.251 (.344)
		(R)	1.979 (.289)	−1.228 (.544)	.251 (.344)
(1.0, 1.0, 0.5)	(2.5,−2.00, 0.500)	(O)	2.296 (.332)	−1.666 (.621)	.369 (.344)
		(N)	2.468 (.295)	−1.944 (.614)	.479 (.368)
		(R)	2.461 (.295)	−1.934 (.613)	.476 (.367)
(1.0, 1.0, 1.0)	(3.0,−3.00, 1.000)	(O)	2.740 (.371)	−2.455 (.790)	.710 (.449)
		(N)	2.957 (.292)	−2.905 (.656)	.948 (.393)
		(R)	2.949 (.292)	−2.889 (.656)	.940 (.392)

In Section 3.1 we have examined the three kinds of distributions for u_t but in this section we consider only the normal error. The above AR(p) model is rewritten as:

$$(1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L) y_t = u_t,$$

where L denotes the lag operator. $\lambda_1, \lambda_2, \cdots, \lambda_p$ are assumed to be real numbers. Taking the cases of $p = 2, 3$, we estimate the true model by Model A. However, the true model is not necessarily equivalent to the estimated model. The case of $\lambda_1 \neq 0$ and $\lambda_2 = \cdots = \lambda_p = 0$ implies that the true model is AR(1) but the estimated one is AR(p). The results are in Table 4 for AR(2) and Table 5 for AR(3), where the arithmetic averages and the root mean squares errors (RMSE) from the 10^3 coefficient estimates of α_1, α_2 and α_3 are shown. In both Tables 4 and 5, $T = 20$ is taken as the sample size. As shown in Tables 1, the cases of $T = 40$ are similar to those of $T = 20$ except that the former cases are less biased than the latter ones. We examine the cases where λ_i takes 0.0, 0.5 or 1.0 for $i = 1, 2, 3$ and $\lambda_1 \geq \lambda_2 \geq \lambda_3$ holds.

Under the assumption that the error term u_t is normally distributed in the true model, we obtain the unbiased estimates using (N), (X), (U) and (R) in Section 2, (R)′ in Remark of Appendix 2, and (O)′ in Appendix 3. In Tables 4 and 5 we focus on (O), (N) and (R), i.e., (X), (U) and (R)′ are omitted, because we have computed all the cases but the obtained results were very similar to those in Table 1. Since (O)′ can be applicable to the AR(1) cases (see Appendix 3), it is also excluded from Tables 4 and 5.

In Table 4, RMSEs of (N) and (R) are smaller than those of OLSE only in the case of $\alpha_1 = 2$ and $\alpha_2 = -1$, i.e., $\lambda_1 = \lambda_2 = 1$. Therefore, (O) is better than (N) and (R) in the RMSE criterion. For all the estimates of α_1 and α_2, however, the arithmetic averages of (N) and (R) are closer to the true parameter values than those of (O). Therefore, it might be concluded that the OLSE bias is corrected by the suggested estimators. (R) is very close to (N) for both bias and RMSE criteria. In Table 4, the true distribution of the error is normal and (N) requires the normal random draws in the estimation procedure. Thus, (N) should be the best in Table 4, because the estimation procedure is consistent with the true distribution. As a result, since (R) is very close to (N), (R) might be useful in empirical studies.

Next, we examine the AR(3) models and the results are in Table 5. For estimation of zero coefficient, all the three estimators are close to the true parameter value. For example, in the case of $(\alpha_1, \alpha_2, \alpha_3) = (0, 0, 0)$, the estimates of α_1 are between $-.010$ and $-.005$, those of α_2 are from $-.073$ to $-.016$ (the estimate of α_2 in (O) is $-.073$, which is slightly biased), and those of α_3 are between .003 and .006. However, for estimation of non-zero coefficient, the suggested estimators are superior to OLSE, which implies that (N) and (R) are less biased than (O). For example, see the estimate of α_1 in the case of $(\alpha_1, \alpha_2, \alpha_3) = (0.5, 0, 0)$, which is .466 for (O), .496 for (N) and .491 for (R), where (N) is very close to (R) and both (N) and (R) are less biased than (O).

Thus, for all the cases of AR(p) for $p = 1, 2, 3$, it is shown from Tables 1 – 5 that OLSE bias is corrected using the proposed estimator (R), even if the data generating process is not known. Finally, note as follows. In Table 4, the case of $\alpha_1 \neq 0$ and $\alpha_2 = 0$ (i.e., $\lambda_1 \neq 0$ and $\lambda_2 = 0$) implies that the data generating process is AR(1). In Table 5, the case of $\alpha_1 \neq 0$ and $\alpha_2 = \alpha_3 = 0$ (i.e., $\lambda_1 \neq 0$ and $\lambda_2 = \lambda_3 = 0$) implies that the data generating process is AR(1) and the case of $\alpha_1 \neq 0$, $\alpha_2 \neq 0$ and $\alpha_3 = 0$ (i.e., $\lambda_1 \neq 0$, $\lambda_2 \neq 0$ and $\lambda_3 = 0$) implies that the data generating process is AR(2). For example, in the case of $(\alpha_1, \alpha_2) = (0.5, 0.0)$, $(1.0, 0.0)$ in Table 4, both (N) and (R) of α_1 are close to 0.5 or 1.0 and those of α_2 are also close to 0.0. Thus, even if the true model is AR(1) but it is estimated as AR(2), i.e., even if the true model is different from the estimated model, we can obtain the bias-corrected coefficient estimate based on the suggested estimators such as (N) and (R).

4 Summary

It is well known that the OLS estimates are biased when the autoregressive terms are included in the explanatory variables. In this paper, we have proposed the bias

correction method using the simulation techniques, where the bootstrap methods are applied. We obtain the unbiased estimate of θ, i.e., $\bar{\theta}$, which is the θ such that the OLSE computed from the original data is equal to the arithmetic average of the OLSEs obtained from the simulated data given θ. When we simulate the data series, we need to assume a distribution of the error term. Since the underlying true distribution of the error term is not known in practice, the four types of random draws are examined for the error term, i.e., the normal error (N), the chi-squared error (X), the uniform error (U) and the residual-based error (R). Because the residual-based approach is distribution-free, it is easily expected that (R) shows a good performance for all the simulation studies even in the case where the distribution of the error term is misspecified. As a result, we can conclude that (R) is practically very useful, because it is relatively unbiased efficient and robust in the sense that (R) can be applicable to any underlying distribution of the error term.

Appendices

Appendix 1: OLSE Bias

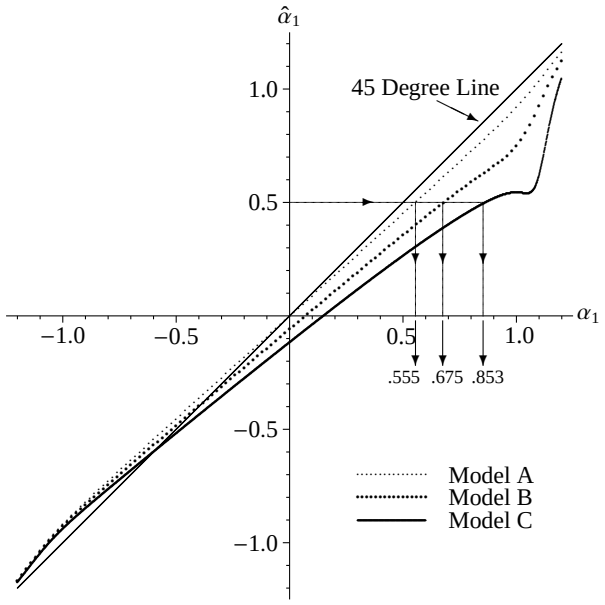
In this appendix, we examine by Monte Carlo simulations how large the OLSE bias is. We focus only on the case of $p = 1$, i.e., the AR(1) model. Suppose that the true model is represented by Model A with $p = 1$. When $x_{1t} = t$ (time trend) is taken in Model C, it is known that OLSE of α_1 from Model C gives us the largest bias and OLSE of α_1 from Model A yields the smallest one (see, for example, Andrews (1993)). Figure 1 shows the relationship between the true autoregressive coefficient (i.e., α_1), the arithmetic mean of OLSEs from 10^4 simulation runs. In order to draw the figure, we take the following simulation procedure.

(i) Generate $y_2, y_3, \cdots, y_T$ by Model A given α_1, $u_t \sim N(0,1)$ and $y_1 = 0$, where $T = 20$.
(ii) Compute OLSE of α_1 by estimating Model A, that of (β_1, α_1) by Model B, and that of $(\beta_1, \beta_2, \alpha_1)$ by Model C.
(iii) Repeat (i) and (ii) 10^4 times.
(iv) Obtain the arithmetic mean from the 10^4 OLSEs of α_1.
(v) Repeat (i) – (iv) given the exactly same random draws for u_t (i.e., $10^4 \times (T - p)$ random draws for $T = 20$ and $p = 1$ have to be stored) and the different parameter value for α_1 (i.e., $\alpha_1 = -1.20, -1.19. - 1.18, \cdots, 1.20$).

Thus, we have the arithmetic mean from the 10^4 OLSEs corresponding to the true parameter value for each model. The true model is given by Model A and it is estimated by Models A – C. The horizontal line implies the true parameter value of the AR(1) coefficient and the vertical line indicates the OLSE corresponding to the true parameter value. Unless the OLSE is biased, the 45 degree line represents the relationship between the true parameter value and the OLSE.

The obtained results are as follows. There is the largest bias around $\alpha_1 = 1$ for all the Models A – C. The bias drastically increases as number of the exogenous variables increases. That is, in the case where α_1 is positive, OLSE of Model C has the largest downward-bias and OLSE of Model A represents the smallest

Fig. 1 The Arithmetic Average from the 10^4 OLSEs of AR(1) Coeff.
— $N(0, 1)$ Error and $n = 20$ —



downward-bias, which implies that inclusion of more extra variables results in larger bias of OLSE.

Now we consider correcting the OLSE bias. In Figure 1, we see the arithmetic mean from the 10^4 OLSEs given the true coefficient. It is also possible to read the figure reversely. For example, when OLSE is obtained as $\hat{\alpha}_1 = 0.5$ from actually observed data, the true parameter value α_1 can be estimated as 0.555 for Model A, 0.675 for Model B and 0.853 for Model C.

In practice, no one knows the true model. What we can do is to estimate the model assumed by a researcher. When we estimate Model C and OLSE of the AR(1) coefficient is obtained as 0.5, the unbiased estimate of the AR(1) coefficient is taken as 0.853. Thus, Figure 1 indicates that inclusion of more extra variables possibly yields serious biased OLSE and furthermore that the true parameter values can be recovered from the estimated model even if we do not know the true model. In Section 2, based on this idea, we obtain the unbiased estimator, which can be applied to any case of the higher-order autoregressive models, the nonnormal error term and inclusion of the exogenous variables other than the constant and trend terms. Here, we take the constant term and the time trend as x_t, although any exogenous variables can be included in the model.

Appendix 2: Optimization Procedure

As for a solution of equation (4), in this appendix we take an example of an iterative procedure, where the numerical optimization procedure is utilized.

Since $\hat{\theta}_i^*$ is computed based on the artificially simulated data, $\bar{\theta}$ is numerically derived by the simulation technique (see, for example, Tanizaki (1995) for the numerical optimization procedure). Using equation (4), we update the parameter θ as follows:

$$\theta^{(j+1)} = \theta^{(j)} + \gamma^{(j)} \left(\hat{\theta} - g(\theta^{(j)}) \right), \tag{5}$$

where j denotes the jth iteration. In the first iteration, OLSE of θ is taken for $\theta^{(1)}$, i.e., $\theta^{(1)} = \hat{\theta}$. $\gamma^{(j)}$ is a scalar, which may depend on number of iteration, i.e., j.

When $\theta^{(j+1)}$ is stable, we take it as the unbiased estimate of θ, i.e., $\bar{\theta}$. As for convergence criterion, in this paper, when each element of $\theta^{(j+1)} - \theta^{(j)}$ is less than 0.001 in absolute value, we consider that $\theta^{(j+1)}$ is stable.

For an interpretation of $\gamma^{(j)}$ in equation (5), it might be appropriate to consider that the Newton-Raphson optimization procedure is taken. which is described as follows. Approximating $\hat{\theta} - g(\theta)$ about $\theta = \theta^*$, we have:

$$0 = \hat{\theta} - g(\theta) \approx \hat{\theta} - g(\theta^*) - \frac{\partial g(\theta^*)}{\partial \theta'}(\theta - \theta^*).$$

Then, we can rewrite as:

$$\theta = \theta^* + \left(\frac{\partial g(\theta^*)}{\partial \theta'} \right)^{-1} \left(\hat{\theta} - g(\theta^*) \right).$$

Regarding θ as $\theta^{(j+1)}$ and θ^* as $\theta^{(j)}$, the following equation is derived:

$$\theta^{(j+1)} = \theta^{(j)} + \left(\frac{\partial g(\theta^{(j)})}{\partial \theta'} \right)^{-1} \left(\hat{\theta} - g(\theta^{(j)}) \right),$$

which is equivalent to equation (5) with the following condition:

$$\left(\frac{\partial g(\theta^{(j)})}{\partial \theta'} \right)^{-1} = \gamma^{(j)} I_{k+p},$$

where I_{k+p} denotes a $(k+p)\times(k+p)$ identity matrix. Since $g(\theta)$ cannot be explicitly specified, we take the first derivative of $g(\theta)$ as the diagonal matrix. Moreover, taking into account the convergence speed, $\gamma^{(j)} = c^{j-1}$ is used in this paper, where $c = 0.9$.

Remark: In Efron and Tibshirani (1993, Chap.10), the bias correction method using the bootstrap is discussed. Let $\hat{\theta}$ be an estimate of θ and $\bar{\theta}$ be the bias-corrected estimate. Then, $\bar{\theta}$ is obtained as follows:

$$\bar{\theta} = \hat{\theta} - \widehat{\text{bias}}, \qquad \widehat{\text{bias}} = \hat{\theta}^* - \hat{\theta}, \qquad \hat{\theta}^* = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i^*(\hat{\theta}),$$

where $\hat{\theta}_i^*(\hat{\theta})$ denotes the ith bootstrap estimate given $\hat{\theta}$. The $\bar{\theta}$ shown above corresponds to $\theta^{(2)}$ with $\theta^{(1)} = \hat{\theta}$, $j = 1$ and (R) in Steps 1 – 7 of Section 2. This procedure is denoted by (R)′ in Tables 1 – 3.

Appendix 3: Bias Correction by Grubb and Symons (1987)

Grubb and Symons (1987) derived an expression to order $1/T$ for bias to the estimated coefficient on a lagged dependent variable when all other regressors are exogenous. Their procedure is applied to only the case of $p = 1$ in equation (1). They have shown that the maximum bias for $|\alpha_1| < 1$ is given by:

$$\mathrm{E}(\hat{\alpha}_1 - \alpha_1) = -\frac{k(1+\alpha_1) + 2\alpha_1}{T-1} + O(T^{-1}).$$

Ignoring the second term in the right-hand side of the above equation, α_1 is approximately written as:

$$\alpha_1 \approx \mathrm{E}\left(\frac{(T-1)\hat{\alpha}_1 + k}{T-k-3}\right).$$

Therefore, $((T-1)\hat{\alpha}_1 + k)/(T-k-3)$ is taken as the bias-corrected estimate, which is given by (O)′ in Tables 1 – 3. Note that the sample size is given by $T-1$ when $p = 1$, i.e., $T-1$ in the above bias-corrected estimator is replaced by T in Grubb and Symons (1987).

The above bias formula has been derived under the assumption $|\alpha_1| < 1$ by Grubb and Symons (1987). At $\alpha_1 = 1$, the bias with no exogenous variables (i.e., the case of $k = 0$) is $-1.785/(T-1)$, the bias with a constant term (i.e., the case of $k = 1$) is $-5.6/(T-1)$, and the bias with a constant term and trend term (i.e., the case of $k = 2$) is $-9.4/(T-1)$. However, practically we do not know the true value of α_1. Therefore, for (O)′ in Tables 1 – 3, $((T-1)\hat{\alpha}_1 + k)/(T-k-3)$ is taken as the bias-corrected estimate irrespective of α_1.

References

Abadir, K.M. (1993) "OLS Bias in A Nonstationary Autoregression," *Econometric Theory*, **9** (1), 81 – 93.

Andrews, D.W.K. (1993) "Exactly Median-Unbiased Estimation of First Order Autoregressive / Unit Root Models," *Econometrica*, **61** (1), 139 – 165.

Andrews, D.W.K. and Chen, H.Y. (1994) "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics*, **12** (2), 187 – 204.

Diebold, F.X. and Rudebusch, G.D. (1991) "Unbiased Estimation of Autoregressive Models," unpublished manuscript, University of Pennsylvania.

Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*.

Enders, W. and Falk, B. (1998) "Threshold-Autoregressive, Median-Unbiased, and Cointegration Tests of Purchasing Power Parity," *International Journal of Forecasting*, **14** (2), 171 – 186.

Greene, W.H. (1993) *Econometric Analysis* (Second Edition), Prentice Hall.

Grubb, D. and Symons, J. (1987) "Bias in Regressions with a Lagged Dependent Variable," *Econometric Theory*, **3**, 371 – 386.

Hurwicz, L. (1950) "Least-Squares Bias in Time Series," in *Statistical Inference in Dynamic Economic Models*, ed. T.C. Koopmans, New York. John Wiley, 365 – 383.

Imhof, J.P. (1961) "Computing the Distribution of Quadratic Forms in Normal Variates," *Biometrika*, **48**, 419 – 426.

Kendall, M.G. (1954) "Note on Bias in the Estimation of Autocorrelations," *Biometrika*, **41**, 403 – 404.

MacKinnon, J.G., and Smith, A.A., Jr. (1998) "Approximate Bias Correction in Econometrics," *Journal of Econometrics*, **85** (2), 205 – 230.

Maekawa, K. (1983) "An Approximation to the Distribution of the Least Squares Estimator in an Autoregressive Model with Exogenous Variables," *Econometrica*, **51** (1), 229 – 238.

Maekawa, K. (1987) "Finite Sample Properties of Several Predictors from an Autoregressive model," *Econometric Theory*, **3**, 359 – 370.

Marriott, F.H.C. and Pope, J.A. (1954) "Bias in the Estimation of Autocorrelations," *Biometrika*, **41**, 390 – 402.

Orcutt, G.H. and Winokur, H.S. (1969) "First Order Autoregression: Inference, Estimation, and Prediction," *Econometrica*, **37** (1), 1 – 14.

Peters, T.A. (1989) "The Exact Moments of OLS in Dynamic Regression Models with Non-Normal Errors," *Journal of Econometrics*, **40**, 279 – 305.

Quenouille, M.H. (1956) "Notes on Bias in Estimation," *Biometrika*, **43**, 353 – 360.

Sawa, T. (1978) "The Exact Moments of the Least Squares Estimator for the Autoregressive Model," *Journal of Econometrics*, **8**, 159 – 172.

Shaman, P. and Stine, R.A. (1988) "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association*, **83**, 842 – 848.

Tanaka, K. (1993) "Asymptotic Expansions Associated with AR(1) Model with Unknown Mean," *Econometrica*, **51** (4), 1221 – 1231.

Tanizaki, H. (1995) "Asymptotically Exact Confidence Intervals of CUSUM and CUSUMSQ Tests: A Numerical Derivation Using Simulation Technique," *Communications in Statistics, Simulation and Computation*, **24** (4), 1019 – 1036.

Tanizaki, H. (2000) "Bias Correction of OLSE in the Regression Model with Lagged Dependent Variables," *Computational Statistics and Data Analysis*, **34** (4), 495 – 511.

Tse, Y.K. (1982) "Edgeworth Approximations in First-Order Stochastic Difference Equations with Exogenous Variables," *Journal of Econometrics*, **20**, 175 – 195.

Tsui, A.K. and Ali, M.M. (1994) "Exact Distributions, Density Functions and Moments of the Least Squares Estimator in a First-Order Autoregressive Model," *Computational Statistics and Data Analysis*, **17** (4), 433 – 454.

White, J.S. (1961) "Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient," *Biometrika*, **48**, 85 – 94.

Wu, C.F.J. (1986) "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *Annals of Statistics*, **14**, 1261 – 1350 (with discussion).