

Edgeworth Approximation of a Finite Sample Distribution for an AR(1) Model with Measurement Error

Shuichi Nagata

Department of Mathematical Sciences, Kwansei Gakuin University, Sanda, Japan
Email: nagatas@kwansei.ac.jp

Received July 25, 2012; revised August 27, 2012; accepted September 9, 2012

ABSTRACT

In this paper, we consider the finite sample property of the ordinary least squares (OLS) estimator for an AR(1) model with measurement error. We present the Edgeworth approximation for a finite distribution of OLS up to $O(T^{1/2})$. We introduce an instrumental variable estimator that is consistent in the presence of measurement error. Finally, a simulation study is conducted to assess the theoretical results and to compare the finite sample performances of these estimators.

Keywords: Edgeworth Expansion; OLS; Measurement Error; Instrumental Variables Estimator

1. Introduction

The Ordinary Least Squares (OLS) estimator for the AR(1) model is one of the most general estimators in econometrics, and a number of studies considering the properties of the OLS estimator under certain conditions have been conducted by several authors. For example, it is well known that the OLS estimator for the AR(1) model has a non-negligible bias when the sample size T is not large. This problem is known as the small sample problem ([1, 2]).

Another problem of the OLS estimator is that the observation data are sometimes contaminated by noise, which also affects the estimation result. In this case, the OLS estimator in the AR(1) model is not consistent. This problem is commonly known as the measurement error problem. Following [3] that summarized the early results on this topic, numerous articles have been published on this topic. For example, with respect to time series analysis, some estimators in an AR model with measurement errors in [4] and statistical a test for the existence of noise is proposed in [5].

In this paper, we deal with these two important problems simultaneously. In particular, we consider the OLS estimation when the sample size T is not large, and when an measurement error is present but ignored. To evaluate the effect of a small sample size and ignoring measurement error, we derive finite sample properties of the OLS estimator with noise using the Edgeworth expansion, which is a traditional technique in econometrics to approximate a finite sample distribution. For example, the OLS estimator was studied in [6] for pure AR(1), in [7]

for AR(1) with an unknown mean, and in [8] for AR(1) with exogenous variables. Following these studies, we apply the algorithm proposed in [9] to calculate the Edgeworth coefficients.

In our setting, some problems are the result of noise, which make calculation difficult. First, if data are affected by noise, it is difficult to obtain variables that are related to the autocovariance function of the observation process. To obtain these variables, we use the result in [10], which shows that an AR(1) process together with noise can be represented by an ARMA(1, 1) process. Second, the OLS estimator is inconsistent with noise, and it is impossible to apply the formula in [9] in this case; hence we use a corrected error function that follows [8] and [11] to avoid this problem.

In the simulation section, we also consider a instrumental variable (IV) estimation, which is the consistent estimation in our setting. We compare the finite sample performances of the OLS estimator with those of our proposed IV estimator using simulations.

This paper is organized as follows. In the next section, we provide our setting and the main result for the Edgeworth approximation of the OLS estimator up to $O(T^{1/2})$. In Section 3, we examine a Monte Carlo simulation and graphical comparison. Finally, Section 4 concludes this paper.

2. Setup and Main Results

We consider a following measurement error model given by

$$\begin{aligned}
 y_t &= x_t + u_t, \quad u_t \sim i.i.d.N(0, \sigma_u^2), \\
 x_t &= \beta x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2),
 \end{aligned}
 \tag{1}$$

where only y_t is observable, x_t is a stationary AR(1) process with $|\beta| < 1$ and u_t is the measurement error or noise. For a given sample period $t = 0, 1, \dots, T$, the OLS estimate can be written as follows:

$$\hat{\beta} = \frac{y' C_1 y}{y' C_2 y}, \tag{2}$$

where $y = (y_0, \dots, y_T)$,

$$C_1 = \begin{pmatrix} 0 & \frac{1}{2} & \dots & 0 & 0 \\ \frac{1}{2} & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \frac{1}{2} \\ 0 & 0 & \dots & \frac{1}{2} & 0 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

The result of this paper relies on the following well known result given in [10]. If x_t is an AR(1) process with AR parameter β , and u_t is white noise with constant variance σ_u^2 , then y_t follows an ARMA(1, 1) process, which is given by the following equations:

$$(1 - L\beta)y_t = (1 - L\gamma)\zeta_t, \quad \zeta_t \sim i.i.d.N(0, \sigma_\zeta^2), \tag{3}$$

where L is the lag operator. The parameters σ_ζ^2 and γ (the MA parameter) can be related to β , σ_u^2 , and σ_ε^2 as follows:

$$\sigma_\zeta^2 = \frac{\beta}{\gamma} \sigma_u^2, \quad \gamma = \frac{1 - \sqrt{1 - 4q^2}}{2q},$$

where

$$q = \frac{\beta \sigma_u^2}{(1 + \beta^2) \sigma_u^2 + \sigma_\varepsilon^2}.$$

Then, we obtain the following theorem.

Theorem 1. The finite sample distribution of the OLS estimator up to $O(T^{1/2})$ is given by

$$\begin{aligned}
 &P\left(\frac{\sqrt{T}(\hat{\beta} - \beta + d)}{\omega} < w\right) \\
 &= I(w) + \frac{i(w)}{\sqrt{T}} \frac{P_1}{3P_2\beta(P_6P_1)^{3/2}} \{(6P_4P_6P_1 - Q) + Qw^2\},
 \end{aligned}$$

where $i(w) = (1/2\pi)^{1/2} \exp(-w^2/2)$, $I(w) = \int_{-\infty}^w i(t) dt$. ω , $P_i (i = 1, \dots, 6)$ and Q are as follows:

$$\omega = (P_6P_1P_2^4\beta^2)^{1/2}, \quad P_1 = (\beta - r)(1 - r\beta),$$

$$P_2 = 1 - 2r\beta + r^2,$$

$$P_3 = 2\beta^5r^2 - \beta^4r - \beta^4r^3 + \beta^3 + r^4\beta^3 + 2r^2\beta^3 - 4r\beta^2 - 4r^3\beta^2 + \beta + r^4\beta + r + r^3,$$

$$P_4 = 3\beta^4r^2 - 2r\beta^3 - 2r^3\beta^3 - 2r^2\beta^2 + \beta^2 + r^4\beta^2 + r^2,$$

$$\begin{aligned}
 P_5 &= 3\beta^3 + 62r^4\beta^3 + 6r^3 + 3r^2\beta^3 - 8r^3\beta^2 - 32r^4\beta \\
 &+ 6r^7 + 16r^5 + 24\beta^5r^2 - 12\beta^4r - 26\beta^4r^3 - 16\beta^6r^5 \\
 &- 20\beta^6r^3 + 32\beta^5r^4 - 12r^9\beta^4 + 3\beta^3r^{10} - 84r^5\beta^4 \\
 &+ 8r^5\beta^2 + 62r^6\beta^3 + 4\beta^7r^4 + 4\beta^8r^5 + 4\beta^7r^6 \\
 &+ 32r^6\beta^5 - 20\beta^6r^7 - 26r^7\beta^4 + 24r^8\beta^5 - 32r^6\beta \\
 &- 8r^7\beta^2 + 3r^8\beta^3,
 \end{aligned}$$

$$\begin{aligned}
 P_6 &= r - 10r^3\beta^2 - 3r^4\beta + \beta + r^5 - 6\beta^5r^2 \\
 &- 3r^2\beta - 4r\beta^2 + 3r\beta^4 + 9r^2\beta^3 - \beta^3 \\
 &+ r^6\beta - 4r^5\beta^2 + 4\beta^4r^3 + 9r^4\beta^3 \\
 &+ 2\beta^6r^3 - 6\beta^5r^4 + 4r^3 - r^6\beta^3 + 3r^5\beta^4
 \end{aligned}$$

$$Q = -P_5P_2\beta + P_5P_2\beta^3 + 6P_4P_1P_3P_2 - 12P_4^2P_1^2.$$

Proof. The proof is given in Appendix.

Here, we also examine the IV estimator, which is defined as follows:

$$\tilde{\beta} = \frac{\sum_{i=2}^T y_i y_{i-2}}{\sum_{i=2}^T y_i y_{i-1}}. \tag{4}$$

The IV estimator is consistent in the presence of the noise. It is easy to show that the asymptotic variance of the IV estimator is $(1 - \beta^2)/\beta^2$. When there is no noise, the asymptotic variance of the OLS estimator is $1 - \beta^2$. Therefore, the OLS estimator is more efficient than the IV estimator in the absence of noise.

3. Simulation and Graphical Comparison

In this section, we examine the finite sample property of the OLS estimator, and evaluate the approximate distribution generated in Section 2 by Monte Carlo simulation.

Data were simulated from Equation (1) with $\sigma_\varepsilon^2 = 1$. Therefore, the noise-to-signal ratio $\rho = \sigma_u^2/\sigma_\varepsilon^2 = \sigma_u^2$ throughout this section. In addition to the OLS estimator, we also compute the IV estimator defined in the previous section. The number of replications was 20,000. We computed the mean (Mean) and the root mean squared error (RMSE) for each estimator. The simulation results are summarized in **Tables 1-3**.

From **Tables 1-3**, we confirm that the OLS estimator is biased. As expected, the smaller the sample size and

Table 1. Simulation results ($\rho = 0.2$).

T	$\beta = 0.4$				$\beta = 0.8$			
	Mean		RMSE		Mean		RMSE	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
20	0.31	0.91	0.24	68.73	0.68	0.71	0.23	1.68
40	0.33	0.58	0.17	45.50	0.71	0.75	0.16	0.22
100	0.34	0.20	0.12	23.74	0.73	0.78	0.11	0.10
500	0.34	0.40	0.07	0.13	0.74	0.80	0.07	0.04
800	0.34	0.40	0.07	0.10	0.74	0.80	0.06	0.03

Table 2. Simulation results ($\rho = 0.4$).

T	$\beta = 0.4$				$\beta = 0.8$			
	Mean		RMSE		Mean		RMSE	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
20	0.27	-0.06	0.25	48.96	0.63	0.69	0.27	3.20
40	0.29	4.23	0.19	540.2	0.66	0.75	0.20	0.29
100	0.29	0.36	0.15	2.53	0.68	0.78	0.15	0.10
500	0.30	0.40	0.11	0.15	0.70	0.80	0.11	0.04
800	0.30	0.40	0.11	0.11	0.70	0.80	0.11	0.03

Table 3. Simulation results ($\rho = 0.7$).

T	$\beta = 0.4$				$\beta = 0.8$			
	Mean		RMSE		Mean		RMSE	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
20	0.23	0.62	0.28	109.62	0.56	0.59	0.32	12.6
40	0.24	0.99	0.22	72.55	0.60	0.77	0.26	2.89
100	0.25	0.37	0.18	4.39	0.62	0.78	0.20	0.12
500	0.25	0.40	0.16	0.18	0.64	0.80	0.17	0.05
800	0.25	0.40	0.15	0.14	0.64	0.80	0.17	0.04

the larger the noise variance, the larger will be the bias. On the other hand, as the IV estimator is a consistent estimator, IV may converge to the true value. The simulation results are consistent with this hypothesis. However, in small samples such as for $T = 20$ and 40 , the RMSE of the IV estimator is rather large, as seen in all tables.

From the simulation results for the RMSE in **Tables 2** and **3**, we find that the IV estimator is more efficient than the OLS estimator when $T \geq 800$ ($\beta = 0.4$) and $T \geq 100$ ($\beta = 0.8$). Therefore, we can conclude that, if the sample size is not large ($T = 20, 40$), or both β and ρ are small as in **Table 1** ($\beta = 0.4$ and $\rho = 0.2$), then the OLS estimator is better than IV in terms of the RMSE.

Next, we compare the exact cdf with the asymptotic normal distribution. **Figure 1** depicts the approximate distributions of the OLS and the IV with $\beta = 0.4$, $T = 20$, and various values of ρ . **Figure 1** indicates that the OLS

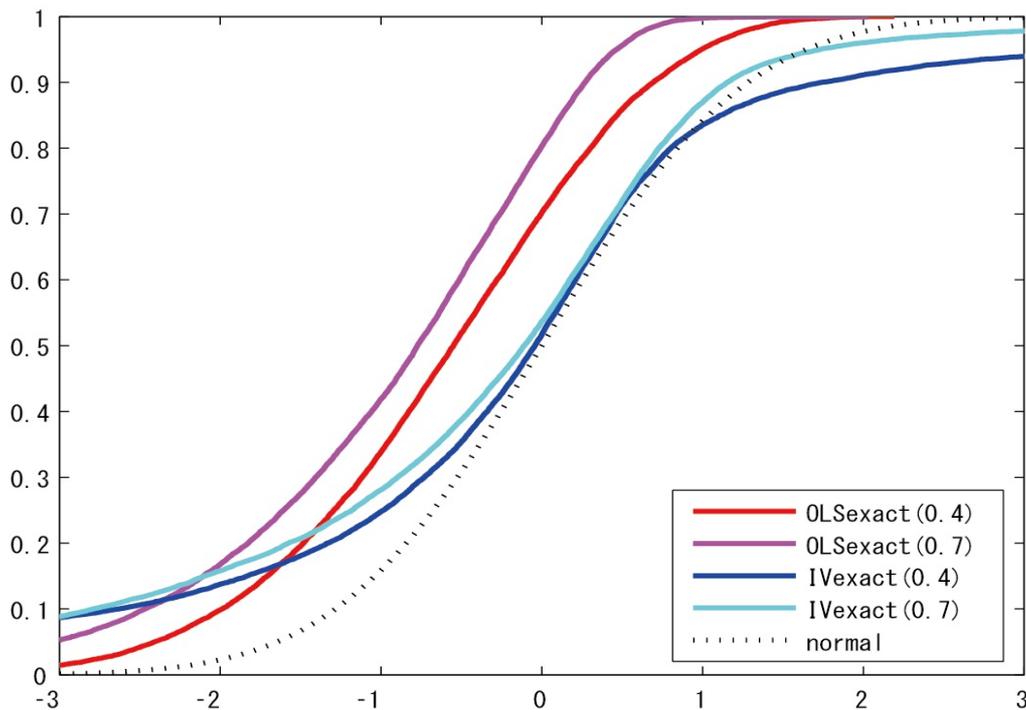


Figure 1. Exact distributions of OLS and IV.

values have a downward bias. The IV exhibits good behavior in the central region of the distribution; however, its distribution is fatter in the tails compared to the normal distribution.

Finally, we evaluate the approximate distributions obtained in Section 2, and compare them with the exact cdf and asymptotic normal distributions. To enable a com-

parison of the shapes of the distributions, the asymptotic bias of the OLS estimator is corrected hereinafter. **Figure 2** shows the approximate distributions of the OLS estimator with $T = 20$, $\rho = 0.2$, and three different values of β . From this figure, we can observe the same result as those obtained in [6]. **Figure 3** depicts the approximate distributions of the OLS with $\rho = 0.7$, where the other pa-

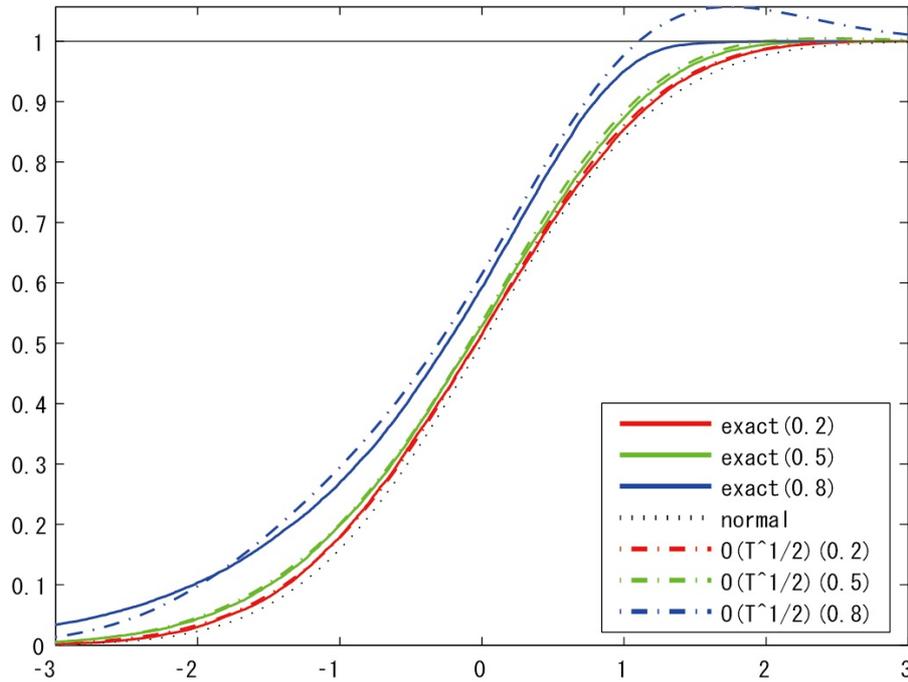


Figure 2. Exact and approximate distributions of OLS.

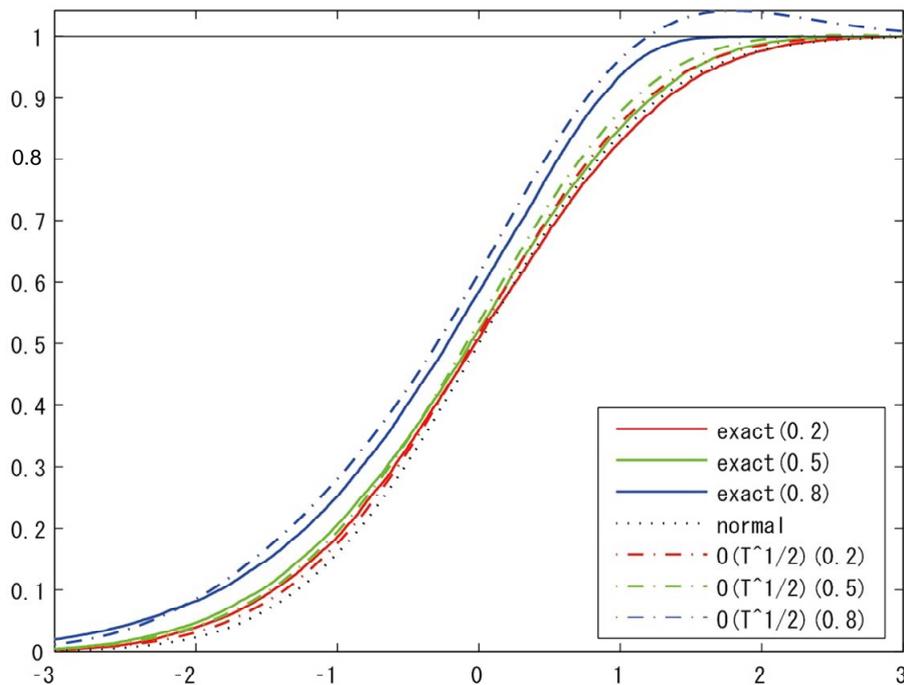


Figure 3. Exact and approximate distributions of OLS.

parameter values are the same as those for **Figure 2**. We note that the shapes of the distributions are almost the same, even if the noise ratio ρ is changed. From these figures, the noise variance has only a small effect on the shape of the OLS distribution.

4. Discussion

In this paper, we considered finite sample properties of the OLS estimator for the AR(1) model with measurement error. Using the formula in [9], we obtained the Edgeworth expansion for finite sample distributions of the OLS estimator up to $O(T^{1/2})$.

In the simulation work, we have compared naive OLS estimator with the IV estimator which is a consistent estimator in the presence of noise. We can confirm that, even if the measurement errors exist, the OLS estimator is more efficient than the IV estimator when the sample size is small such as $T = 20$ and 40 . If the noise-to-signal ratio is not so small ($\rho \geq 0.4$), the IV estimator is more efficient than the OLS estimator when $T \geq 800$ ($\beta = 0.4$) or $T \geq 100$ ($\beta = 0.8$). From the graph of the normalized OLS distributions, we find similar properties to those of the distributions, which correspond to the no noise situation examined by [6]. This result implies that measurement error mainly distorts the OLS distributions for mean and variance; hence we can separately deal with the two problems of small sample size and observation errors.

Recently, the differenced-AR(1) estimator was discussed in [12,13]. Even if the sample size T is relatively small, this estimator has a small bias. To obtain the finite sample distribution and to examine the robustness of such estimators with respect to observation errors, we can apply the technique of this paper, and this will be dealt with in a future study.

5. Acknowledgements

I am grateful to Professor Koichi Maekawa for his guidance on this topic and his valuable comments on this paper. I am also grateful to Professor Yasuyoshi Tokutsu for his valuable comments.

REFERENCES

- [1] F. H. C. Marriott and J. A. Pope, "Bias in the Estimation of Autocorrelations," *Biometrika*, Vol. 41, No. 3-4, 1954, pp. 390-402. [doi:10.2307/2332719](https://doi.org/10.2307/2332719)
- [2] M. G. Kendall, "Note on the Bias in the Estimation of Autocorrelation," *Biometrika*, Vol. 41, No. 3-4, 1954, pp. 403-404. [doi:10.2307/2332720](https://doi.org/10.2307/2332720)
- [3] W. A. Fuller, "Measurement Error Models," John Wiley, New York, 1987. [doi:10.1002/9780470316665](https://doi.org/10.1002/9780470316665)
- [4] J. Staudenmayer and P. Buonaccorsi, "Measurement Error in Linear Autoregressive Model," *Journal of the American Statistical Association*, Vol. 100, No. 471. 2005, pp. 841-851. [doi:10.1198/016214504000001871](https://doi.org/10.1198/016214504000001871)
- [5] K. Tanaka, "A Unified Approach to the Measurement Error Problem in Time Series Models," *Econometric Theory*, Vol. 18, No. 2, 2002, pp. 278-296. [doi: 10.1017/S026646660218203X](https://doi.org/10.1017/S026646660218203X)
- [6] P. C. B. Phillips, "Approximations to Some Sample Distributions Associated with a First Order Stochastic Difference Equation," *Econometrica*, Vol. 45, No. 2, 1977, pp. 463-485. doi.org/10.2307/1911222
- [7] K. Tanaka, "Asymptotic Expansions Associated with the AR(1) Model with Unknown Mean," *Econometrica*, Vol. 51, No. 4, 1983, pp. 1221-1231. [doi:10.2307/1912060](https://doi.org/10.2307/1912060)
- [8] K. Maekawa "An Approximation to the Distribution of the Least Squares Estimator in an Autoregressive Model with Exogenous Variables," *Econometrica*, Vol. 51, No.1, 1983, pp. 229-238. [doi:10.2307/1912258](https://doi.org/10.2307/1912258)
- [9] J. D. Sargan, "Econometric Estimators and the Edgeworth Expansions," *Econometrica*, Vol. 44, No. 3, 1976, pp. 421-448. [doi:10.2307/1913972](https://doi.org/10.2307/1913972)
- [10] C. W. J. Granger and M. J. Morris, "Time Series Modeling and Interpretation," *Journal of the Royal Statistical Society A*, Vol. 139, No. 2, 1976, pp. 246-257. [doi:10.2307/2345178](https://doi.org/10.2307/2345178)
- [11] K. Maekawa, "Edgeworth Expansion for the OLS Estimator in a Time Series Regression Model," *Econometric Theory*, Vol. 1, No. 2, 1985, pp. 223-239. [doi:10.1017/S0266466600011154](https://doi.org/10.1017/S0266466600011154)
- [12] K. Hayakawa, "A Note on Bias in First-Differenced AR(1) Models," *Economics Bulletin*, Vol. 3 No. 27, 2006, pp. 1-10.
- [13] P. C. B. Phillips and C. Han, "Gaussian Inference in AR(1) Times Series with or without Unit Root," *Econometric Theory*, Vol. 24, No. 3, 2008, pp. 631-650. [doi:10.1017/S0266466608080262](https://doi.org/10.1017/S0266466608080262)

Appendix

Proof of Theorem 1

The OLS estimator for β is given by Equation (2). Introducing $\mu_i = E[y'C_i y]$, we can write the derivation of the estimation as follows:

$$\hat{\beta} - \beta = \frac{(y'C_1 y - \mu_1) - \beta(y'C_2 y - \mu_2) - \sigma_u^2 \beta T}{(y'C_2 y - \mu_2) + \mu_2}.$$

In addition, we introduce $q_i = (y'C_i y - \mu_i)/T$ and $q = (q_1, q_2)'$. Then, we have the error function as:

$$\hat{\beta} - \beta = \frac{q_1 - \beta q_2 - \sigma_u^2 \beta}{q_2 + \mu_2/T}.$$

In order to develop the Edgeworth expansion, we define a modified function $e(q)$:

$$e(q) = \hat{\beta} - \beta + \frac{\sigma_u^2 \beta T}{\mu_2} = \frac{q_1 - \beta q_2 - \sigma_u^2 \beta}{q_2 + \mu_2/T} + \frac{\sigma_u^2 \beta T}{\mu_2}.$$

It is easy to obtain the cumulant generating function of $T^{1/2} q$ is

$$\psi(\theta) = -\frac{1}{2} \log \det \left\{ I - \frac{2}{\sqrt{T}} (\theta_1 C_1 + \theta_2 C_2) \Sigma \right\} - \frac{\mu_1 \theta_1}{\sqrt{T}} - \frac{\mu_2 \theta_2}{\sqrt{T}},$$

where $\theta_j = it_j$ and $\theta = (\theta_1, \theta_2)'$. The matrices I and Σ are the identity matrix and the covariance matrix of y , respectively.

Edgeworth expansion requires the partial derivatives of $e(q)$ and $\varphi(\theta)$ up to the third order. In the current paper these derivatives are denoted as e_j, e_{jk}, ψ_{jk} and ψ_{jkl} , which are all evaluated at the origin. Using the tensor summation convention, Edgeworth coefficients of Sargan's formula are obtained by these derivatives as follows:

$$\begin{aligned} \omega^2 &= -\psi_{jk} e_j e_k, & \alpha_1 &= \psi_{jkl} e_j e_k e_l, \\ \alpha_3 &= \gamma_j e_{jk} \gamma_k, & \alpha_4 &= \psi_{jk} e_{jk}, \\ \gamma_a &= \psi_{aj} e_j. \end{aligned}$$

Although we only show Edgeworth coefficients related to approximations of up to $O(T^{1/2})$, the original formula of Sargan can approximate up to $O(T)$, see [6].

After many calculations, we finally obtain the Edgeworth coefficients. To save space, we show only the results as follows:

$$\begin{aligned} \alpha_1 &= \frac{-2iP_1 P_5 (1 - \beta^2)}{P_2^6 \beta^3 \sqrt{T}}, & \alpha_3 &= \frac{4(-P_3 P_2 + 2P_4 P_1) P_1^2 P_4}{P_2^7 \beta^4}, \\ \alpha_4 &= \frac{4P_4 P_1}{P_2^3 \beta^2}, & \omega^2 &= \frac{P_6 P_1}{P_2^4 \beta^2}, \end{aligned}$$

where P_i and Q are polynomials defined in Section 2. The approximation of the OLS estimator up to $O(T^{1/2})$ is derived in [6] as follows.

$$p(\sqrt{T}e(q) < x) = I\left(\frac{x}{\omega}\right) + i\left(\frac{x}{\omega}\right) \left\{ c_0 + c_2 \left(\frac{x}{\omega}\right)^2 \right\}, \tag{5}$$

where c_0 and c_2 are

$$c_0 = \frac{\alpha_4}{2\omega\sqrt{T}} + \frac{i\alpha_1}{6\omega^3} + \frac{\alpha_3}{2\omega^3\sqrt{T}} \quad c_2 = -\frac{1}{\omega^3} \left(\frac{i\alpha_1}{6} + \frac{\alpha_3}{2\sqrt{T}} \right).$$

Using these results, we obtain the following equations.

$$c_0 = \frac{P_1(6P_4 P_6 P_1 - Q)}{3P_2 \beta (P_6 P_1)^{3/2} \sqrt{T}} \quad c_2 = \frac{P_1 Q}{3P_2 \beta (P_6 P_1)^{3/2} \sqrt{T}}$$

Substituting these results into Equation (5), we obtain Theorem 1.