Model Selection for Panel Data Models with Fixed Effects: A Simulation Study*

Minchul Yum

July 2021

Abstract

This study considers model selection criteria, such as the Akaike's Information Criterion (AIC), the corrected Akaike's Information Criterion (AIC_C), and the Bayesian Information Criterion (BIC), for panel data models with fixed effects. Applying these information criteria to fixed effects panel models is not a trivial matter due to the incidental parameters problem that might adversely affect their practical performance, especially when it comes to short panel data. Monte Carlo experiments suggest that the information criteria are quite successful in selecting the true model. In particular, the AIC_C and the AIC operate successfully unless a time dimension is extremely small.

Keywords: Panel data, fixed effects, model selection criteria

JEL codes: C23, C53

*Address: L7, 3-5, Department of Economics, University of Mannheim, 68161, Mannheim, Germany. E-mail: minchul.yum@uni-mannheim.de. This note is based on my master's thesis submitted to Sogang University. I am indebted to the thesis supervisor, In Choi, for his guidance and support. I am also grateful to the anonymous referee for the insightful and constructive comments and to Seung C. Ahn, Hanbat Jeong, and participants at 2010 Korean Economic Association Joint Conference at Seoul National University for their helpful comments.

1 Introduction

Panel data have become increasingly available and popular in economics. Consider a popular linear regression model:

$$y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + u_{it}, \qquad (i = 1, ..., N,\ t = 1, ..., T), \tag{1}$$

where $\mathbf{X}_{it}$ is a matrix of observable regressors, and $\boldsymbol{\beta}$ is an l-vector of parameters we are interested in. We can decompose u_{it} into three statistically independent parts (Balestra and Nerlove 1966):

$$u_{it} = \alpha_i + \lambda_t + \varepsilon_{it}, \tag{2}$$

where α_i represents time-invariant individual fixed effects, capturing unobserved individual specific effects (e.g., innate ability), and λ_t captures time-specific effects which do not vary across individuals.

This paper focuses on estimating (1) with (2) while treating α_i and λ_t as fixed—the fixed effect estimation. Considering them as parameters to be estimated, the OLS estimator of $\boldsymbol{\beta}$ in (1) is also the Least-Squares Dummy Variable (LSDV) estimator, which is consistent even when unobserved heterogeneity, α_i and λ_t, is correlated with $\mathbf{X}_{it}$. Two issues arise. First, the LSDV estimator may involve a substantially large number of parameters to be estimated. For example, if the true data generating process (DGP) contains only λ_t, the efficiency of the estimator $\hat{\boldsymbol{\beta}}$ in (1) can be adversely affected since there are many additional parameters for non-meaningful dummies (overspecification). Second, assuming the same DGP as above, if a researcher mistakenly chooses a model which only includes individual fixed effects, it would lead to a misspecification problem, resulting in an inconsistent estimator of $\boldsymbol{\beta}$.

It would be reasonable to allow data to select which structure of error-components gives the best fit in panel data fixed-effects models, unless there is a firm reason (e.g., based on economic theory) for choosing a particular structure of the error term. In this regard, various procedures to compare models such as the F-test and various asymptotic testing methods in the form of LM tests or LR tests have been mainly used. Since these hypothesis tests require a certain set of null and alternative hypotheses, their results may vary depending on the different set of hypotheses or significance levels. Furthermore, these tests can be applied only when the two competing models are nested with each other.

An alternative approach, which is the focus of this study, is to apply information criteria to the model selection problem among candidate fixed-effects models.[1] The core philosophy

[1] The model selection approach has led to significant advances in statistics, and has been applied to economics as well (see Chao and Phillips 1999, Bai and Ng 2002, Baltagi and Wang 2007, Choi and Kurozumi 2012, Choi and Jeong 2019 for some recent examples). The exercise presented herein is another attempt to apply model selection to an important empirical model in economics.

of model selection is to balance the quality of fit against complexity. In this sense, the model selection can be thought of as a good alternative to solve the previously suggested problems involving a large number of parameters. However, this application needs careful analysis when T is small and constant in fixed-effects estimation because the number of parameters for individual dummies increases with N. In other words, the incidental parameters problem (Neyman and Scott 1948) could adversely affect the operation of the information criteria by preventing many dummies from being consistently estimated.

Through Monte Carlo simulations, this study examines the practical performance of model selection criteria in choosing the true fixed effects model in the presence of the incidental parameters problem. Overall, the experiment results suggest that the three widely used information criteria, including the Akaike's Information Criterion (AIC), the corrected Akaike's Information Criterion (AIC_C), and the Bayesian Information Criterion (BIC), are quite successful in selecting the true model. In particular, AIC_C and AIC are found to operate more successfully given that a time dimension is not extremely small.

2 Model Selection Criteria

Utilizing a simple transformation of the sum of squared residuals (SSR), AIC (Akaike 1973) for the fixed-effects estimation is given by

$$AIC = NT \log(\hat{\sigma}_\varepsilon^2) + 2K, \tag{3}$$

where $\hat{\sigma}_\varepsilon^2 = \frac{SSR}{NT}$, and K is the number of parameters. The second is often called a penalty term.

Since AIC is derived asymptotically, a short length of T in actual panel data might affect their practical performance. Moreover, since the derivation depends on the maximum likelihood, $\hat{\sigma}_\varepsilon^2$ should be the MLE of σ_ε^2 and be included in K. Finally, K includes the number of dummies, as we treat them to be estimated.

Hurvich and Tsai (1989) develop the corrected AIC (AIC_C) that improves the finite-sample properties of AIC. Adopting the criterion into the current setting, we have

$$AIC_C = AIC + \frac{2K(K+1)}{NT - K - 1}. \tag{4}$$

One might prefer AIC_C instead of AIC with short panels. However, this is not clear because the second correction term does not vanish as N goes to infinity holding T fixed, since K includes increasing N as well. Therefore we can infer that although the second correction term of AIC_C can mitigate the tendency of AIC to overfit the true model, it might also undesirably lead to a rather underspecified model.

Table 1: Error structures in DGPs

	Structure of the error term
DGP I	$u_{it} = \varepsilon_{it}$
DGP II	$u_{it} = \alpha_i + \varepsilon_{it}$
DGP III	$u_{it} = \lambda_t + \varepsilon_{it}$
DGP IV	$u_{it} = \alpha_i + \lambda_t + \varepsilon_{it}$

where $\varepsilon_{it} \sim i.i.d.N(0,1)$, $\alpha_i \sim i.i.d.N(0,\sigma_\alpha^2)$, and $\lambda_t \sim i.i.d.N(0,\sigma_\lambda^2)$

Unlike AIC and AIC$_C$ that seek to find the best model that minimizes the estimated Kullback-Leibler discrepancy, the Bayesian Information Criterion (BIC) (Schwarz 1978, Akaike 1978) finds the best model that maximizes the Bayesian posterior probability. BIC in the current setting is:

$$BIC = NT\log(\hat{\sigma}_\varepsilon^2) + K\log(NT). \tag{5}$$

Comparing the penalty terms in (3) and (5), we can easily infer that BIC would choose a parsimonious model due to its harsher penalty term, especially when there are many parameters.

3 Monte Carlo Simulation

The true model or the data generating process (DGP) is set to be

$$y_{it} = \beta x_{it} + u_{it}, \qquad (i = 1, ..., N; \quad t = 1, ..., T) \tag{6}$$

where $x_{it} \sim i.i.d.N(0,1)$ and β is set to 1. As summarized in Table 1, u_{it} in (6) is assumed to have four different structures. This experiment design follows the error-components model assumptions by Wallace and Hussain (1969): α_i, λ_t, and ε_{it} have zero means with variances $\sigma_\alpha^2, \sigma_\lambda^2$, and σ_ε^2, respectively, and are independent of each other.

The assumption that individual-specific effects are not correlated with the regressor might be strong. Thus, we also consider (6) with the second error structure in Table 1 while replacing the independent regressor with the one having $corr(x_{it}, \alpha_i) = 0.3$ (denoted as *DGP V*).

For the magnitudes of σ_α^2 and σ_λ^2, we consider three different values of $0.5, 1$, and 4, chosen around the normalized variance of idiosyncratic errors (ε_{it}). The results with the three different values are expected to give the distinct benchmarks for practical use. We are interested in a practical setting with a large N and a small T. The results presented here include 12 combinations of $N = 50, 100, 200, 500$ and $T = 2, 4, 6$.

Finally, the four different structures of u_{it} in Table 1 are also considered for the structure

Table 2: Four candidate models

Candidate	Specification
Model 1: *Pooled OLS*	$y_{it} = \mu + \beta x_{it} + \varepsilon_{it}$
Model 2: *1-way FE (w.r.t. i)*	$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$
Model 3: *1-way FE (w.r.t. t)*	$y_{it} = \beta x_{it} + \lambda_t + \varepsilon_{it}$
Model 4: *2-way FE*	$y_{it} = \beta x_{it} + \alpha_i + \lambda_t + \varepsilon_{it}$

of candidate models. Table 2 summarizes the candidate models.

Through 1,000 replications for each combination of N, T, and σ_α^2 (or σ_λ^2) according to each DGP, the relative frequency that each candidate model is selected by the model selection criteria is calculated.[2] The OLS residuals after the corresponding within transformations are used to calculate information criteria for easier computation. Fixed seeds are used for generating random numbers.

The first case with *DGP I* is shown in Table 3. While BIC is found to be very accurate, AIC and AIC$_C$ tend to slightly overfit the true model. This overspecification is not a critical issue because (i) the number of additional dummies in *Model* 3 is only $T - 1$, which is very small in our case; and, (ii) they correct their selection behavior in favor of the true model as T increases. Thus we can conclude that using any information criteria is safe and accurate under *DGP I*.

Table 4 summarizes the results with *DGP II*. AIC performs fairly successfully when $T \geq 4$, and becomes more accurate as either N or T increases. Although it tends to overfit the true model, it rarely selects misspecified models (i.e., *Model* 1 and *Model* 3). The overspecification problem here is not a serious concern either. More importantly, the relative frequency of overspecification decreases as T becomes larger. Unlike AIC, AIC$_C$ does not tend to overfit the true model regardless of the length of T. Instead, AIC$_C$ is subject to the serious misspecification problem when T is extremely small. As T increases, however, this problem quickly vanishes, and AIC$_C$ performs slightly better than AIC. The threshold T that entirely alters the performance of AIC$_C$ depends heavily on the size of σ_α^2. Next, BIC mostly selects either a underspecified model (when the length of T is not enough) or the true model (when T reaches the sufficient levels that depend on σ_α^2). In addition, the performance of BIC worsens as N increases holding T fixed; that is, BIC appears to be more exposed to the

[2] Table 8 reports the relative likelihood of model i (Burnham and Anderson, 2002):

$$\exp\left(\frac{AIC_{\min} - AIC_i}{2}\right)$$

for each different true model under consideration, with largest sample number ($N = 500$ and $T = 6$). This statistic could be helpful to evaluate prediction accuracy losses from model selection. The reported numbers are the averages across replications for $\sigma_\alpha^2 = \sigma_\lambda^2 = 1$. The results with other values of σ_α^2 and σ_λ^2 are essentially the same.

incidental parameters problem.

We now turn to the case where the true model does not contain the individual specific effects in the error term. This setting is non-problematic in the sense that even if there exist time effects in the error term, we have a fairly large number of N which allows the time dummies to be accurately estimated. Table 5 presents the results under $DGP\ III$. All of the three information criteria demonstrate similar practical performances; given $T \geq 4$, all of them select the true model very accurately. It is interesting to note that the performance of BIC now improves as N increases even with fixed T, like AIC and AIC_C.

The case of $DGP\ IV$ is presented in Table 6. The performance of AIC given a sufficient T is even more accurate, as compared to the case under $DGP\ II$. Even when T is extremely small (e.g., $T = 2$), an increase in N allows AIC to operate significantly better. The performance of AIC_C is as great as under $DGP\ II$; given the variance of specific effects which is larger than that of idiosyncratic errors and $T \geq 4$, its capability of choosing the true model is almost perfect. However, unlike the case under $DGP\ II$ where AIC_C outperforms AIC, they seem to be equally effective under $DGP\ IV$. By contrast, BIC seems vulnerable to the incidental parameters problem again; an increase in N is not beneficial for the operation of BIC as long as T is fixed.

Finally, $DGP\ V$ that contains a correlated regressor is investigated in Table 7. The results are mostly similar to those in Table 4. Given $T \geq 4$, both AIC and AIC_C operate fairly correctly with a negligible concern of the overspecification for AIC. Even with $T = 2$, a high σ_α^2 and a high N would warrant a great practical performance of AIC_C. By contrast, BIC tends to underestimate the true model. This misspecification is more serious since using the pooled OLS model would lead the estimator of β to be biased in the presence of the correlation between regressors and unobservable individual specific effects.

4 Concluding remarks

This study finds that the model selection criteria, especially AIC and AIC_C, are generally applicable to the problem of determining the structure of fixed-effects error term. BIC also can be useful only when T is relatively large with a substantial variance of individual-specific effects. Therefore, when T is short, even if BIC often outperforms depending on the true model, we should generally put more weight on AIC and AIC_C since true models are unknown in practice.

Based on the simulation results that exhibit successful practical performances, future work in this direction seems promising. It is worth noting that the conclusion obtained from simulations in a basic environment is by no means complete for the applications to a more general environment. It would be important to consider a simulation study with such generalizations as the presence of multivariate independent variables and the widely adopted environment of a dynamic panel model (Hahn and Kuersteiner 2002). Moreover, although the current study examines the issue only through simulations using select popular information criteria, further theoretical research such as devising an alternative information criterion could be valuable.

References

Akaike, H. (1973): "Information Theory and an Extension of the Maximum Likelihood Principle," *Second International Symposium on Information Theory*, in B.N. Petrov and F. Caski, eds., Akademiai Kiado, Budapest, 267-281.

Akaike, H. (1978): "A Bayesian Analysis of the Minimum AIC procedure," *Annals of the Institute of Statistical Mathematics*, 30(1), 9-14.

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191-221.

Balestra P., and M. Nerlove (1966): "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: the Demand for Natural Gas," *Econometrica*, 34(3), 585-612.

Baltagi, B.H. and Z. Wang. (2007): "Testing for Cointegrating Rank Via Model Selection: Evidence from 165 Data Sets." *Empirical Economics* 33 (1): 41-49.

Burnham, K.P., and D.R. Anderson (2002): *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach,* 2nd ed., Springer-Verlag, New York.

Chao, J.C., and P.C.B. Phillips (1999): "Model Selection in Partially Nonstationary Vector Autoregressive Processes with Reduced Rank Structure," *Journal of Econometrics*, 91(2), 227-271.

Choi, I. and H. Jeong. 2019. "Model Selection for Factor Analysis: Some New Criteria and Performance Comparisons." *Econometric Reviews* 38 (6): 577-596.

Choi, I. and E. Kurozumi. 2012. "Model Selection Criteria for the Leads-and-Lags Cointegrating Regression." *Journal of Econometrics* 169 (2): 224-238.

Hahn, Jinyong and Guido Kuersteiner. 2002. "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when both n and T are Large." *Econometrica* 70 (4): 1639-1657.

Table 3: Model selection relative frequency under DGP I

		AIC				AIC$_C$				BIC			
N	T	M1*	M2	M3	M4	M1*	M2	M3	M4	M1*	M2	M3	M4
50	2	.796	.028	.162	.014	.848	.000	.152	.000	.967	.000	.033	.000
	4	.879	.000	.121	.000	.893	.000	.107	.000	.998	.000	.002	.000
	6	.922	.001	.077	.000	.932	.000	.068	.000	1.000	.000	.000	.000
100	2	.832	.001	.166	.001	.844	.000	.156	.000	.981	.000	.019	.000
	4	.900	.000	.100	.000	.908	.000	.092	.000	1.000	.000	.000	.000
	6	.926	.000	.074	.000	.931	.000	.069	.000	1.000	.000	.000	.000
200	2	.827	.000	.173	.000	.833	.000	.167	.000	.984	.000	.016	.000
	4	.890	.000	.110	.000	.894	.000	.106	.000	1.000	.000	.000	.000
	6	.924	.000	.076	.000	.927	.000	.073	.000	1.000	.000	.000	.000
500	2	.835	.000	.165	.000	.838	.000	.162	.000	.992	.000	.008	.000
	4	.884	.000	.116	.000	.884	.000	.116	.000	1.000	.000	.000	.000
	6	.916	.000	.084	.000	.917	.000	.083	.000	1.000	.000	.000	.000

Note: *$Model$ 1 is the true model.

Hurvich, C.M., and C. Tsai (1989): "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76(2), 297-307.

Neyman, J., and E.L. Scott (1948): "Consistent Estimation from partially consistent observations," *Econometrica*, 16, 1-32.

Schwarz, G. (1978): "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461-464.

Wallace, T.D., and A. Hussain (1969): "The Use of Error Components Models in Combining Cross Section with Time Series Data," *Econometrica*, 37(1), 55-72.

Table 4: Model selection relative frequency under DGP II

			AIC				AIC$_C$				BIC			
σ^2_α	N	T	M1	M2*	M3	M4	M1	M2*	M3	M4	M1	M2*	M3	M4
.5	50	2	.231	.499	.022	.248	.929	.000	.071	.000	.989	.000	.011	.000
		4	.032	.752	.004	.212	.462	.496	.020	.022	1.000	.000	.000	.000
		6	.001	.860	.000	.139	.019	.942	.002	.037	1.000	.000	.000	.000
	100	2	.172	.543	.024	.261	.923	.000	.077	.000	.993	.000	.007	.000
		4	.004	.800	.000	.196	.407	.558	.018	.017	1.000	.000	.000	.000
		6	.000	.865	.000	.135	.004	.964	.000	.032	1.000	.000	.000	.000
	200	2	.122	.601	.018	.259	.918	.000	.082	.000	.997	.000	.003	.000
		4	.000	.799	.000	.201	.327	.627	.019	.027	1.000	.000	.000	.000
		6	.000	.871	.000	.129	.000	.961	.000	.039	1.000	.000	.000	.000
	500	2	.041	.662	.001	.296	.928	.000	.072	.000	.998	.000	.002	.000
		4	.000	.795	.000	.205	.277	.679	.008	.036	1.000	.000	.000	.000
		6	.000	.853	.000	.147	.000	.959	.000	.041	1.000	.000	.000	.000
1	50	2	.015	.667	.001	.317	.960	.001	.039	.000	.997	.001	.002	.000
		4	.000	.779	.000	.221	.015	.945	.000	.040	.989	.011	.000	.000
		6	.000	.861	.000	.139	.000	.961	.000	.039	.721	.279	.000	.000
	100	2	.000	.668	.000	.332	.959	.000	.041	.000	.996	.000	.004	.000
		4	.000	.803	.000	.197	.000	.963	.000	.037	1.000	.000	.000	.000
		6	.000	.865	.000	.135	.000	.968	.000	.032	.980	.020	.000	.000
	200	2	.000	.688	.000	.312	.948	.000	.052	.000	1.000	.000	.000	.000
		4	.000	.799	.000	.201	.000	.951	.000	.049	1.000	.000	.000	.000
		6	.000	.871	.000	.129	.000	.961	.000	.039	1.000	.000	.000	.000
	500	2	.000	.693	.000	.307	.963	.000	.037	.000	.999	.000	.001	.000
		4	.000	.795	.000	.205	.000	.955	.000	.045	1.000	.000	.000	.000
		6	.000	.853	.000	.147	.000	.959	.000	.041	1.000	.000	.000	.000
4	50	2	.000	.682	.000	.318	.267	.706	.001	.026	.421	.497	.000	.082
		4	.000	.779	.000	.221	.000	.959	.000	.041	.002	.987	.000	.011
		6	.000	.861	.000	.139	.000	.961	.000	.039	.000	1.000	.000	.000
	100	2	.000	.668	.000	.332	.105	.859	.002	.034	.961	.035	.000	.004
		4	.000	.803	.000	.197	.000	.963	.000	.037	.000	.997	.000	.003
		6	.000	.865	.000	.135	.000	.968	.000	.032	.000	1.000	.000	.000
	200	2	.000	.688	.000	.312	.022	.932	.000	.046	1.000	.000	.000	.000
		4	.000	.799	.000	.201	.000	.951	.000	.049	.011	.984	.000	.005
		6	.000	.871	.000	.129	.000	.961	.000	.039	.000	1.000	.000	.000
	500	2	.000	.693	.000	.307	.000	.961	.000	.039	1.000	.000	.000	.000
		4	.000	.795	.000	.205	.000	.955	.000	.045	.536	.463	.000	.001
		6	.000	.853	.000	.147	.000	.959	.000	.041	.000	1.000	.000	.000

Note: **Model* 2 is the true model.

Table 5: Model selection relative frequency under DGP III

			AIC				AIC$_C$				BIC			
σ_λ^2	N	T	M1	M2	M3*	M4	M1	M2	M3*	M4	M1	M2	M3*	M4
.5	50	2	.231	.010	.723	.036	.252	.000	.748	.000	.328	.000	.672	.000
		4	.038	.000	.961	.001	.041	.000	.959	.000	.125	.000	.875	.000
		6	.004	.000	.995	.001	.005	.000	.995	.000	.050	.000	.950	.000
	100	2	.170	.001	.827	.002	.173	.000	.827	.000	.252	.000	.748	.000
		4	.010	.000	.990	.000	.010	.000	.990	.000	.061	.000	.939	.000
		6	.001	.000	.999	.000	.001	.000	.999	.000	.020	.000	.980	.000
	200	2	.122	.000	.878	.000	.122	.000	.878	.000	.203	.000	.797	.000
		4	.004	.000	.996	.000	.004	.000	.996	.000	.024	.000	.976	.000
		6	.000	.000	1.000	.000	.000	.000	1.000	.000	.001	.000	.999	.000
	500	2	.080	.000	.920	.000	.080	.000	.920	.000	.138	.000	.862	.000
		4	.001	.000	.999	.000	.001	.000	.999	.000	.007	.000	.993	.000
		6	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	.999	.000
1	50	2	.161	.008	.793	.038	.185	.000	.815	.000	.250	.000	.750	.000
		4	.016	.000	.983	.001	.017	.000	.983	.000	.053	.000	.947	.000
		6	.000	.000	.999	.001	.000	.000	1.000	.000	.014	.000	.986	.000
	100	2	.116	.001	.881	.002	.120	.000	.880	.000	.190	.000	.810	.000
		4	.002	.000	.998	.000	.002	.000	.998	.000	.024	.000	.976	.000
		6	.000	.000	1.000	.000	.000	.000	1.000	.000	.003	.000	.997	.000
	200	2	.089	.000	.911	.000	.091	.000	.909	.000	.149	.000	.851	.000
		4	.001	.000	.999	.000	.001	.000	.999	.000	.007	.000	.993	.000
		6	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000	.000
	500	2	.061	.000	.939	.000	.061	.000	.939	.000	.104	.000	.896	.000
		4	.000	.000	1.000	.000	.000	.000	1.000	.000	.002	.000	.998	.000
		6	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000	.000
4	50	2	.089	.003	.865	.043	.099	.000	.901	.000	.139	.000	.861	.000
		4	.002	.000	.997	.001	.002	.000	.998	.000	.008	.000	.992	.000
		6	.000	.000	.999	.001	.000	.000	1.000	.000	.000	.000	1.000	.000
	100	2	.057	.000	.940	.003	.058	.000	.942	.000	.100	.000	.900	.000
		4	.001	.000	.999	.000	.001	.000	.999	.000	.002	.000	.998	.000
		6	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000	.000
	200	2	.033	.000	.967	.000	.034	.000	.966	.000	.065	.000	.935	.000
		4	.000	.000	1.000	.000	.000	.000	1.000	.000	.001	.000	.999	.000
		6	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000	.000
	500	2	.033	.000	.967	.000	.033	.000	.967	.000	.059	.000	.941	.000
		4	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000	.000
		6	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000	.000

Note: **Model* 3 is the true model.

Table 6: Model selection relative frequency under DGP IV

			AIC				AIC_C				BIC			
$\sigma_\alpha^2, \sigma_\lambda^2$	N	T	M1	M2	M3	M4*	M1	M2	M3	M4*	M1	M2	M3	M4*
.5	50	2	.078	.118	.170	.634	.292	.000	.708	.000	.395	.000	.605	.000
		4	.001	.021	.029	.949	.031	.029	.493	.447	.200	.000	.800	.000
		6	.000	.002	.000	.998	.001	.008	.032	.959	.104	.000	.896	.000
	100	2	.044	.078	.149	.729	.198	.000	.802	.000	.309	.000	.691	.000
		4	.001	.006	.002	.991	.010	.009	.440	.541	.113	.000	.887	.000
		6	.000	.000	.000	1.000	.000	.001	.004	.995	.043	.000	.957	.000
	200	2	.015	.079	.124	.782	.149	.000	.851	.000	.243	.000	.757	.000
		4	.000	.003	.000	.997	.002	.004	.373	.621	.040	.000	.960	.000
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.004	.000	.996	.000
	500	2	.005	.053	.036	.906	.091	.000	.909	.000	.173	.000	.827	.000
		4	.000	.000	.000	1.000	.001	.000	.303	.696	.012	.000	.988	.000
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.001	.000	.999	.000
1	50	2	.005	.131	.012	.852	.246	.000	.753	.001	.339	.000	.660	.001
		4	.000	.013	.000	.987	.000	.022	.021	.957	.131	.000	.856	.013
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.036	.003	.648	.313
	100	2	.000	.080	.000	.920	.171	.000	.829	.000	.251	.000	.749	.000
		4	.000	.002	.000	.998	.000	.006	.000	.994	.063	.000	.937	.000
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.019	.000	.956	.025
	200	2	.000	.061	.000	.939	.130	.000	.870	.000	.203	.000	.797	.000
		4	.000	.001	.000	.999	.000	.002	.000	.998	.023	.000	.997	.000
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.002	.000	.998	.000
	500	2	.000	.040	.000	.960	.082	.000	.918	.000	.138	.000	.862	.000
		4	.000	.000	.000	1.000	.000	.000	.000	1.000	.007	.000	.993	.000
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.001	.000	.999	.000
4	50	2	.000	.071	.000	.929	.057	.095	.275	.573	.123	.051	.286	.540
		4	.000	.000	.000	1.000	.000	.003	.000	.997	.000	.006	.001	.993
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000
	100	2	.000	.049	.000	.951	.024	.067	.106	.803	.204	.002	.756	.038
		4	.000	.001	.000	.999	.000	.001	.000	.999	.000	.002	.000	.998
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000
	200	2	.000	.023	.000	.977	.003	.050	.022	.925	.171	.000	.829	.000
		4	.000	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.010	.990
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000
	500	2	.000	.024	.000	.976	.000	.045	.000	.955	.117	.000	.883	.000
		4	.000	.000	.000	1.000	.000	.000	.000	1.000	.001	.000	.518	.481
		6	.000	.000	.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000

Note: *$Model$ 4 is the true model.

Table 7: Model selection relative frequency under DGP V

$corr(X,\alpha)=.3$			AIC				AIC_C				BIC			
σ_α^2	N	T	M1	M2*	M3	M4	M1	M2*	M3	M4	M1	M2*	M3	M4
.5	50	2	.294	.456	.038	.212	.910	.000	.090	.000	.987	.000	.013	.000
		4	.048	.740	.002	.210	.544	.409	.028	.019	1.000	.000	.000	.000
		6	.001	.860	.000	.139	.043	.919	.001	.037	1.000	.000	.000	.000
	100	2	.242	.486	.031	.241	.908	.000	.092	.000	.993	.000	.007	.000
		4	.011	.792	.002	.195	.544	.416	.028	.012	1.000	.000	.000	.000
		6	.000	.865	.000	.135	.006	.962	.000	.032	1.000	.000	.000	.000
	200	2	.199	.534	.028	.239	.900	.000	.100	.000	.997	.000	.003	.000
		4	.002	.797	.000	.201	.534	.418	.029	.019	1.000	.000	.000	.000
		6	.000	.871	.000	.129	.000	.961	.000	.039	1.000	.000	.000	.000
	500	2	.100	.609	.015	.276	.910	.000	.090	.000	.997	.000	.003	.000
		4	.000	.795	.000	.205	.578	.378	.027	.017	1.000	.000	.000	.000
		6	.000	.853	.000	.147	.000	.959	.000	.041	1.000	.000	.000	.000
1	50	2	.024	.663	.001	.312	.939	.000	.061	.000	.992	.000	.008	.000
		4	.000	.779	.000	.221	.022	.937	.001	.040	.997	.003	.000	.000
		6	.000	.861	.000	.139	.000	.961	.000	.039	.840	.160	.000	.000
	100	2	.002	.666	.000	.332	.942	.000	.058	.000	.996	.000	.004	.000
		4	.000	.803	.000	.197	.001	.962	.000	.037	1.000	.000	.000	.000
		6	.000	.865	.000	.135	.000	.968	.000	.032	.999	.001	.000	.000
	200	2	.000	.688	.000	.312	.939	.000	.061	.000	.998	.000	.002	.000
		4	.000	.799	.000	.201	.000	.951	.000	.049	1.000	.000	.000	.000
		6	.000	.871	.000	.129	.000	.961	.000	.039	1.000	.000	.000	.000
	500	2	.000	.693	.000	.307	.947	.000	.053	.000	1.000	.000	.000	.000
		4	.000	.795	.000	.205	.000	.955	.000	.045	1.000	.000	.000	.000
		6	.000	.853	.000	.147	.000	.959	.000	.041	1.000	.000	.000	.000
4	50	2	.000	.682	.000	.318	.386	.590	.007	.017	.541	.394	.000	.065
		4	.000	.779	.000	.221	.000	.959	.000	.041	.001	.988	.000	.011
		6	.000	.861	.000	.139	.000	.961	.000	.039	.000	1.000	.000	.000
	100	2	.000	.668	.000	.332	.195	.769	.005	.031	.987	.010	.000	.003
		4	.000	.803	.000	.197	.000	.963	.000	.037	.004	.993	.000	.003
		6	.000	.865	.000	.135	.000	.968	.000	.032	.000	1.000	.000	.000
	200	2	.000	.688	.000	.312	.058	.895	.002	.045	1.000	.000	.000	.000
		4	.000	.799	.000	.201	.000	.951	.000	.049	.049	.946	.000	.005
		6	.000	.871	.000	.129	.000	.961	.000	.039	.000	1.000	.000	.000
	500	2	.000	.693	.000	.307	.003	.958	.000	.039	1.000	.000	.000	.000
		4	.000	.795	.000	.205	.000	.955	.000	.045	.907	.093	.000	.000
		6	.000	.853	.000	.147	.000	.959	.000	.041	.000	1.000	.000	.000

Note: **Model* 2 is the true model.

Table 8: Akaike relative likelihood

$N = 500$, $T = 6$	Candidate model			
True model	*Model* 1	*Model* 2	*Model* 3	*Model* 4
DGP I	.968	.000	.177	.000
DGP II	.000	.925	.000	.257
DGP III	.000	.000	1.000	.000
DGP IV	.000	.000	.000	1.000
DGP V	.000	.925	.000	.258