Practical Unit-Root Analysis Using Information Criteria: Simulation Evidence

Kosei Fukuda
Nihon University, Japan

Recommended Citation

Practical Unit-Root Analysis Using Information Criteria: Simulation Evidence

Kosei Fukuda
Nihon University

The information-criterion-based model selection method for detecting a unit root is proposed. The simulation results suggest that the performances of the proposed method are usually comparable to and sometimes better than those of the conventional unit-root tests. The advantages of the proposed method in practical applications are also discussed.

Key words: Information criteria, model selection, Monte Carlo simulation, pre-testing problem, unit root

Introduction

Since the seminal work of Dickey and Fuller (1979), numerous alternative methods have been developed to improve the size and power properties of unit-root tests. However, little attention has been paid to two practical problems encountered in unit-root tests. Consider the following augmented DF (ADF) regression for an observed time series y_t $(t = 1,...,T)$:

$$\Delta y_t = \mu + \beta t + \rho y_{t-1} + \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + e_t,$$
(1)

where $e_t \sim \text{NID}(0, \sigma^2)$. Three model classes and the corresponding t-statistics (denoted by $\hat{\tau}$) for a unit root are obtained by considering parameter restrictions on (1).

Kosei Fukuda is Associate Professor, College of Economics, at Nihon University, Tokyo, Japan. He had served as an economist in the Economic Planning Agency of the Japanese Government (1986-2000). Email him at fukuda@eco.nihon-u.ac.jp or at kosefuku@crocus.ocn.ne.jp.

Model 1 (Statistics, Null, and Alternative)
$\hat{\tau}$ $\mu = \beta = \rho = 0$ $\mu = \beta = 0$ and $\rho < 0$

Model 2 (Statistics, Null, and Alternative)
$\hat{\tau}_\mu$ $\mu = \beta = \rho = 0$ $\beta = 0$ and $\rho < 0$

Model 3 (Statistics, Null, and Alternative)
$\hat{\tau}_\tau$ $\beta = \rho = 0$ $\rho < 0$

Although there is no discussion on a statistical method for selecting a suitable model class from among these three alternatives, different statistics can lead to different conclusions. For example, in the seminal work of Nelson and Plosser (1982), the result obtained applying Model 3 suggested that the annual time series of U.S. unemployment rate is generated from the trend-stationary process. Forecasting the unemployment rate with the trend-stationary model in the very long horizon would provide a value less than zero or more than 1. If Model 1 is applied in place of Model 3, the null hypothesis of a unit root cannot be rejected. Which conclusion should be embraced? As discussed by Phillips (2005), there is little guidance from economic theory about the source and nature of the trending behavior. Thus, model selection criteria are expected to be applied in selecting a suitable model from among three alternatives.

Furthermore, the pretesting problem also arises. In the conventional ADF regression, the lag length k is selected by applying the Ng and

Perron (1995) (NP) general-to-specific procedure. In this procedure, given the maximum lag length k_{max}, by working backward from $k = k_{max}$, the first value of k is selected such that the t-statistic on $\hat{\phi}_k$ is significant. Thus, in total, hypothesis testing has to be implemented at two stages. Every test has a nonzero frequency of rejecting the null hypothesis, thus causing Type I errors to accumulate. As discussed by Krolzig and Hendry (2001), it is important to distinguish between the individual test sizes and the overall test size.

These two practical problems have remained even in the recent literature. For example, although Elliott et al. (1996) have proposed more powerful unit root tests, there is no criterion in selecting deterministic components and there is the pretesting problem in selecting the lag length. The ADF test still has the most popularity even now in empirical analyses, while it is the oldest unit-root test and has low power. Thus, the ADF test remains a benchmark method in the present study.

The purpose of this article is to propose an information criterion (IC)-based model selection method for detecting a unit root in order to provide a solution to the above two problems. In this method, the following three steps are taken. First, several alternative models are considered by changing the model class (Models 1, 2, and 3 with and without a unit root) and the lag length, and each model is estimated with the corresponding IC. Second, the best model is selected from among the alternative models by using the minimum IC procedure. Finally, on the basis of the selected model, it is determined whether the observed data contain a unit root. In this article, the Akaike information criterion (AIC) proposed by Akaike (1974) and the Bayesian information criterion (BIC) proposed by Schwartz (1978) are applied. The AIC and BIC for Model (1) are obtained as follows:

$$\text{AIC} = (T - k - 1)\ln\hat{\sigma}^2 + 2p,$$
$$\text{BIC} = (T - k - 1)\ln\hat{\sigma}^2 + p\ln(T - k - 1),$$

where p denotes the number of parameters. In the full model (Model 3 without a unit root), $p = 3 + k$, and the other values of p are obtained in correspondence with the number of parameter restrictions.

Simulation Studies: The Case Where the Model Specification is Known

Unlike the ADF tests described earlier, the DF likelihood ratio (DFLR, Dickey & Fuller, 1981) tests are now considered. This is because the AIC and BIC are both penalized likelihoods. However, DF argued that the limiting distribution of the LR test statistics is too complex to provide an analytical solution. DF presented empirical distributions using Monte Carlo simulations. In their simulation study, the following three cases are considered:

Null   Alternative

Case1: $y_t = y_{t-1} + e_t \qquad y_t = \mu + \rho\, y_{t-1} + e_t$

Case 2: $y_t = y_{t-1} + e_t \qquad y_t = \mu + \beta t + \rho y_{t-1} + e_t$

Case 3: $y_t = \alpha + y_{t-1} + e_t \quad y_t = \mu + \beta t + \rho y_{t-1} + e_t.$

In the subsequent study presented in this section, it is assumed that the model specification is known in each case. In the Monte Carlo simulation, the assumed data generating process (DGP) is

$$y_t = \theta\, y_{t-1} + e_t \text{ and } e_t \sim \text{NID}(0,1).$$

Each experiment is performed as follows. First, artificial time series are generated from the assumed DGP. Second, in each case, the DFLR test and the IC-based model selection are performed. Finally, the presence or absence of a unit root is determined in each method. In the DFLR test, three significance levels—10%, 5%, and 1%—are applied. In each experiment, three values of θ (0.9, 0.95, 1) and two values of T (100, 250) are considered. The number of replications in each experiment is 5,000.

Table 1 shows the frequency count of selecting stationary models. In Case 1, the frequency count of incorrectly selecting stationary models, which corresponds to test size in the terminology of hypothesis testing, is high in the AIC-based method. In the case of $T = 100$, this count is 0.41 and it is 0.43 in the case of $T = 250$. On the other hand, the performances of the BIC-based method are comparable to those of the DFLR tests. Interestingly, in the case of $T = 100$, the performances of the BIC-based method are identical to those of the DFLR tests at the 5% significance level. This is because the penalty on the likelihood of the stationary model in the BIC-based method is accidentally identical to that in the DFLR test. In Case 2, similar results are obtained. However, in Case 3, the frequency count of incorrectly selecting stationary models is slightly high in the BIC-based method in the case of $T = 100$. Thus, it can be concluded that

the performances of the BIC-based method are roughly comparable to those of the DFLR tests.

The Case Where Only the Lag Length is Unknown

The IC-based method is compared with the ADF tests using the NP lag length selection. The DGP considered here is partially similar to that considered by NP. Artificial time series are generated using the following process:

$$y_t = \rho\, y_{t-1} + \sum_{i=1}^{3} \phi_i \Delta y_{t-i} + e_t,$$

where $e_t \sim \mathrm{NID}(0,1)$. Three values of ρ— $\rho = 0.85,\ 0.95,\ 1$—are considered, and four vectors of (ϕ_1, ϕ_2, ϕ_3)—$(\phi_1, \phi_2, \phi_3) = (0.6,\ 0,\ 0)$, $(-0.6,\ 0,\ 0)$, $(0.4,\ 0.2,\ 0)$, $(0.2,\ 0.2,\ 0.2)$—are considered. The maximum lag length k_{max} is assumed to be $k_{max} = 8$. Given the observed

Table 1. Frequency count of selecting stationary models

θ	T	Methods				
		10%	5%	1%	AIC	BIC
Null: $y_t = y_{t-1} + e_t$, Alternative: $y_t = \mu + \rho\, y_{t-1} + e_t$.						
0.9	100	0.39	0.24	0.06	0.86	0.24
0.9	250	0.99	0.93	0.64	1.00	0.80
0.95	100	0.15	0.08	0.01	0.56	0.08
0.95	250	0.54	0.36	0.11	0.95	0.20
1	100	0.10	0.05	0.01	0.41	0.05
1	250	0.10	0.05	0.01	0.43	0.02
Null: $y_t = y_{t-1} + e_t$, Alternative: $y_t = \mu + \beta t + \rho\, y_{t-1} + e_t$.						
0.9	100	0.18	0.09	0.02	0.78	0.10
0.9	250	0.82	0.64	0.28	1.00	0.41
0.95	100	0.09	0.04	0.01	0.57	0.04
0.95	250	0.25	0.13	0.03	0.88	0.05
1	100	0.10	0.05	0.01	0.58	0.05
1	250	0.10	0.05	0.01	0.58	0.02
Null: $y_t = \alpha + y_{t-1} + e_t$, Alternative: $y_t = \mu + \beta t + \rho\, y_{t-1} + e_t$.						
0.9	100	0.28	0.15	0.04	0.96	0.39
0.9	250	0.90	0.77	0.41	1.00	0.88
0.95	100	0.13	0.07	0.01	0.83	0.21
0.95	250	0.36	0.22	0.06	0.98	0.33
1	100	0.11	0.06	0.01	0.75	0.18
1	250	0.09	0.05	0.01	0.74	0.08

Note: DGP: $y_t = \theta\, y_{t-1} + e_t$.

time series, the lag length selection is performed as follows. In the ADF tests, by working backward from $k = k_{max}$, the first value of k is selected such that the t-statistic on $\hat{\phi}_k$ is significant. Three significance levels—10%, 5%, and 1%—are applied in the case of the ADF unit-root tests and the t-tests for the lag length selection.

In the IC-based method, alternative models are considered by changing the lag length as $k = 0, 1, ..., k_{max}$, and the best model is selected from $(1 + k_{max})$ alternative models. Two values of T (100, 250) are considered, and each experiment is replicated 5,000 times.

Table 2 shows the frequency count of selecting stationary models. First, the size distortion does not occur, while it is well known that the incorrect lag length can cause the size distortion. As shown by NP, the general-to-specific procedure in selecting the lag length can prevent the size distortion. Second, the pretesting problem is not shown, similar to the results obtained by NP. Finally, the performances of the BIC-based method are roughly comparable to those of the ADF 5% tests in the case of $T = 100$ and to those of the ADF 1% tests in the case of $T = 250$.

Table 2. Frequency count of selecting stationary models

Parameters				Methods				
ρ	ϕ_1	ϕ_2	ϕ_3	10%	5%	1%	AIC	BIC
				T = 100				
0.85	0.6	0	0	0.99	0.98	0.97	1.00	0.97
0.85	-0.6	0	0	0.78	0.63	0.26	0.89	0.49
0.85	0.4	0.2	0	0.98	0.97	0.90	0.99	0.95
0.85	0.2	0.2	0.2	0.98	0.94	0.64	0.99	0.78
0.95	0.6	0	0	0.86	0.76	0.43	0.94	0.66
0.95	-0.6	0	0	0.35	0.19	0.05	0.44	0.11
0.95	0.4	0.2	0	0.84	0.68	0.27	0.92	0.51
0.95	0.2	0.2	0.2	0.80	0.63	0.22	0.87	0.37
1	0.6	0	0	0.12	0.05	0.01	0.20	0.03
1	-0.6	0	0	0.11	0.05	0.01	0.19	0.04
1	0.4	0.2	0	0.11	0.06	0.01	0.19	0.04
1	0.2	0.2	0.2	0.11	0.05	0.01	0.20	0.06
				T = 250				
0.85	0.6	0	0	1.00	1.00	1.00	1.00	1.00
0.85	-0.6	0	0	1.00	0.99	0.95	1.00	0.95
0.85	0.4	0.2	0	1.00	1.00	1.00	1.00	1.00
0.85	0.2	0.2	0.2	1.00	1.00	1.00	1.00	0.99
0.95	0.6	0	0	1.00	1.00	0.98	1.00	0.96
0.95	-0.6	0	0	0.76	0.55	0.20	0.87	0.28
0.95	0.4	0.2	0	1.00	1.00	0.96	1.00	0.96
0.95	0.2	0.2	0.2	1.00	0.99	0.92	1.00	0.93
1	0.6	0	0	0.11	0.05	0.01	0.20	0.02
1	-0.6	0	0	0.10	0.05	0.01	0.19	0.02
1	0.4	0.2	0	0.11	0.05	0.01	0.19	0.02
1	0.2	0.2	0.2	0.11	0.05	0.01	0.19	0.02

Note: DGP: $y_t = \rho\, y_t + \sum_{i=1}^{3} \phi_i \Delta y_{t-i} + e_t$.

The Case Where the Model Class as Well as the Lag Length is Unknown

In this phase, the model class (Models 1, 2, and 3 in Section 1) as well as the lag length is unknown. Thus, the ADF tests should determine which model class is applied. No study has been devoted to this problem. Furthermore, there is little guidance from economic theory about the source and nature of the trending behavior. In the present study, therefore, the selection of the model class is performed based on statistical tests. Motivated by NP, general-to-specific modeling is performed in this subsection as follows. First, using Model 3, the ADF regression is performed with the NP lag length selection. If the t-statistic on $\hat{\beta}$ of the selected model is significant, Model 3 is obtained and the unit-root test is implemented. Otherwise, the ADF regression is performed using Model 2. If the t-statistic on $\hat{\mu}$ of the selected model is significant, Model 2 is obtained and the unit-root test is implemented. Otherwise, the ADF unit-root test is performed using Model 1. Three significance levels—10%, 5%, and 1%—are applied in the case of the ADF unit-root tests, the t-tests for the lag length selection, and the t-tests for $\hat{\beta}$ and $\hat{\mu}$. If each test is independent at three stages and is evaluated at the 10% significance level, the overall rejection probability under the null is $1-(1-0.1)^3 = 0.271$, which is substantial.

In this simulation, artificial time series are generated using the following process:

$$y_t = \mu + \beta t + \rho\, y_{t-1} + \sum_{i=1}^{3}\phi_i \Delta y_{t-i} + e_t,$$

where $e_t \sim \text{NID}(0,1)$. In order to obtain the stationary Models 1–3, the parameter vectors considered are $(\mu = 0, \beta = 0)$, $(\mu = 1, \beta = 0)$, and $(\mu = 0, \beta = 0.1)$, respectively. With regard to the nonstationary models, the parameter vectors considered are $(\mu = 0, \beta = 0)$ and $(\mu = 0.1, \beta = 0)$. The other parameter setting is implemented as follows. Two values of ρ — $\rho = 0.95, 1$ —are considered, and four vectors of (ϕ_1, ϕ_2, ϕ_3) —

$(\phi_1, \phi_2, \phi_3) = (0.6, 0, 0)$, $(-0.6, 0, 0)$, $(0.4, 0.2, 0)$, $(0.2, 0.2, 0.2)$—are considered. The maximum lag length k_{max} is assumed to be $k_{max} = 8$. Two values of T (100, 250) are considered, and each experiment is replicated 5,000 times.

Table 3 shows the frequency count of selecting stationary models. Unlike in the preceding subsection, in this case, the pretesting problem is clearly shown. Consider the results of applying the 10% significance level for eight unit-root processes with $T = 100$. The frequency count of incorrectly selecting stationary models is from 26% to 31%. In the preceding subsection, it was shown that the lag length selection has little effect on the results of the unit-root tests under the assumption of the known model class. The size distortion is caused by the assumed method for selecting the model class. The selection of the model class with statistically significant deterministic components such as μ and β has a bias toward selecting a stationary model. The same results are obtained in the case of $T = 250$; however, in this case, the extent of size distortion is smaller. The frequency count of incorrectly selecting stationary models by the BIC-based method is from 12% to 19% in the case of $T = 100$ and from 5% to 9% in the case of $T = 250$. The performances of the BIC-based method are roughly comparable to those of the ADF 5% tests in the case of $T = 100$ and to those of the ADF 2.5% tests (not shown here) in the case of $T = 250$.

In particular, in the case of the trending process, it can be concluded that the BIC-based method outperforms the hypothesis-testing method. For example, compare the two cases of $(\rho, \mu, \beta, \phi_1, \phi_2, \phi_3) = (1, 0.1, 0, 0.6, 0, 0)$ and $(0.95, 0, 0.1, 0.6, 0, 0)$ with $T = 100$. In the case of the 10% significance level, the frequency counts of incorrectly and correctly selecting stationary models are 0.27 and 0.45, respectively. On the other hand, in the case of the BIC-based method, the frequency counts of incorrectly and correctly selecting stationary models are 0.15 and 0.49, respectively. In the terminology of hypothesis testing, the BIC-based method shows lower size and higher

Table 3. Frequency count of selecting stationary models

Parameters						Methods				
ρ	μ	β	ϕ_1	ϕ_2	ϕ_3	10%	5%	1%	AIC	BIC
T = 100										
0.95	0	0	0.6	0	0	0.84	0.78	0.46	0.98	0.72
0.95	0	0	-0.6	0	0	0.45	0.29	0.07	0.76	0.18
0.95	0	0	0.4	0.2	0	0.83	0.71	0.29	0.96	0.55
0.95	0	0	0.2	0.2	0.2	0.80	0.66	0.25	0.94	0.40
0.95	1	0	0.6	0	0	0.64	0.48	0.18	0.91	0.34
0.95	1	0	-0.6	0	0	0.29	0.17	0.04	0.65	0.08
0.95	1	0	0.4	0.2	0	0.58	0.40	0.11	0.88	0.22
0.95	1	0	0.2	0.2	0.2	0.55	0.36	0.10	0.84	0.15
0.95	0	0.1	0.6	0	0	0.45	0.32	0.10	0.96	0.49
0.95	0	0.1	-0.6	0	0	0.36	0.24	0.09	1.00	1.00
0.95	0	0.1	0.4	0.2	0	0.41	0.26	0.07	0.93	0.35
0.95	0	0.1	0.2	0.2	0.2	0.39	0.24	0.06	0.91	0.27
1	0	0	0.6	0	0	0.31	0.19	0.04	0.75	0.13
1	0	0	-0.6	0	0	0.30	0.17	0.04	0.72	0.12
1	0	0	0.4	0.2	0	0.30	0.17	0.03	0.74	0.13
1	0	0	0.2	0.2	0.2	0.28	0.16	0.04	0.76	0.16
1	0.1	0	0.6	0	0	0.27	0.16	0.04	0.77	0.15
1	0.1	0	-0.6	0	0	0.26	0.15	0.03	0.74	0.14
1	0.1	0	0.4	0.2	0	0.27	0.15	0.04	0.76	0.15
1	0.1	0	0.2	0.2	0.2	0.26	0.14	0.04	0.78	0.19
T = 250										
0.95	0	0	0.6	0	0	1.00	1.00	0.99	1.00	1.00
0.95	0	0	-0.6	0	0	0.73	0.57	0.21	0.94	0.31
0.95	0	0	0.4	0.2	0	1.00	1.00	0.96	1.00	0.99
0.95	0	0	0.2	0.2	0.2	0.99	0.99	0.92	1.00	0.95
0.95	1	0	0.6	0	0	0.97	0.96	0.85	1.00	0.89
0.95	1	0	-0.6	0	0	0.41	0.25	0.07	0.81	0.07
0.95	1	0	0.4	0.2	0	0.97	0.94	0.73	1.00	0.77
0.95	1	0	0.2	0.2	0.2	0.97	0.92	0.67	1.00	0.70
0.95	0	0.1	0.6	0	0	0.92	0.88	0.66	1.00	0.94
0.95	0	0.1	-0.6	0	0	1.00	1.00	0.99	1.00	1.00
0.95	0	0.1	0.4	0.2	0	0.90	0.83	0.55	1.00	0.84
0.95	0	0.1	0.2	0.2	0.2	0.89	0.80	0.50	1.00	0.79
1	0	0	0.6	0	0	0.24	0.14	0.03	0.71	0.06
1	0	0	-0.6	0	0	0.24	0.14	0.03	0.70	0.05
1	0	0	0.4	0.2	0	0.25	0.13	0.03	0.71	0.05
1	0	0	0.2	0.2	0.2	0.25	0.15	0.03	0.73	0.06
1	0.1	0	0.6	0	0	0.19	0.10	0.02	0.76	0.08
1	0.1	0	-0.6	0	0	0.18	0.10	0.02	0.74	0.07
1	0.1	0	0.4	0.2	0	0.20	0.11	0.02	0.77	0.08
1	0.1	0	0.2	0.2	0.2	0.19	0.11	0.02	0.76	0.09

Note: DGP: $y_t = \mu + \beta t + \rho y_t + \sum_{i=1}^{3} \phi_i \Delta y_{t-i} + e_t$.

power than the hypothesis-testing method. Similar results are obtained in other trending process of $(\rho = 1, \mu = 0.1)$ and $(\rho = 0.95, \beta = 0.1)$.

Conclusion

This article focused on the two problems encountered in the conventional unit-root tests: the absence of a criterion for selecting a suitable model class and the presence of the pretesting problem. In order to provide a solution to these problems, the IC-based model selection method was proposed. In this method, alternative models with and without a unit root are considered by changing the model class and the lag length. All the possible models are estimated and the corresponding IC values are stored. Finally, the best model is selected from among the alternatives. Thus, on the basis of the selected model, it is determined whether the observed time series contain a unit root. The simulation results suggested that the performances of the BIC-based method are usually comparable to and sometimes better than those of the DFLR and ADF unit-root tests.

In comparison with the conventional hypothesis testing methods, this BIC-based model selection method has two advantages. First, by the introduction of the minimum BIC procedure, the subjective judgment required in the hypothesis testing procedure for determining the levels of significance is completely eliminated, thus enabling a semiautomatic execution. The well-known criticism of the IC-based method is that it cannot control the test size. However, as shown in Table 3, the conventional hypothesis testing method causes the pretesting problem and cannot control the overall test size.

Second, the selection of the model class can be performed automatically and consistently using the IC-based method. Furthermore, flexible time-series modeling, such as the introduction of measurement error (Fukuda, 2005a) and/or regime switching (Fukuda, 2005b), is applicable in the proposed method, and the efficacy of a model change can be consistently evaluated via the minimum BIC procedure. In the case of hypothesis testing, different models require different statistics; this makes time-series analyses very complex.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19, 716-723.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*, 427-431.

Dickey, D. A. & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49*, 1057-1072.

Elliot, G., Rothnberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica, 64*, 813-836.

Fukuda, K. (2005a). Unit-root detection allowing for measurement error. *Statistics and Probability Letters, 74*, 373-377.

Fukuda, K. (2005b). Detection of regime switches between stationary and nonstationary processes and economic forecasting. *Journal of Forecasting, 24*, 255-267.

Krolzig, H.-M. & Hendry, D.F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control, 25*, 831-866.

Nelson, C. R. & Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics, 10*, 139-162.

Ng, S. & Perron, P. (1995). Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association, 90*, 268-281.

Phillips, P. C. B. (2005). Challenges of trending time series econometrics. *Mathematics and Computers in Simulation, 68*, 401-416.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*, 461-464.