Model selection for explosive models

Yubo TAO
Singapore Management University, yubo.tao.2014@phdecons.smu.edu.sg

Jun YU
Singapore Management University, yujun@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/soe_research
Part of the Econometrics Commons, and the Economic Theory Commons

Citation
TAO, Yubo and Jun YU. Model selection for explosive models. (2016). 1-31. Research Collection School Of Economics.
Available at: https://ink.library.smu.edu.sg/soe_research/1801
Model Selection for Explosive Models

Yubo Tao & Jun Yu

March 2016

Paper No. 06-2016
Model Selection for Explosive Models*

Yubo Tao & Jun Yu

Singapone Management University

March 29, 2016

Abstract

This paper examines the limit properties of information criteria for distinguishing between the unit root model and the various kinds of explosive models. The information criteria include AIC, BIC, HQIC. The explosive models include the local-to-unit-root model, the mildly explosive model and the regular explosive model. Initial conditions with different order of magnitude are considered. Both the OLS estimator and the indirect inference estimator are studied. It is found that BIC and HQIC, but not AIC, consistently select the unit root model when data come from the unit root model. When data come from the local-to-unit-root model, both BIC and HQIC select the wrong model with probability approaching 1 while AIC has a positive probability of selecting the right model in the limit. When data come from the regular explosive model or from the mildly explosive model in the form of \(1 + n^\alpha/n\) with \(\alpha \in (0, 1)\), all three information criteria consistently select the true model. Indirect inference estimation can increase or decrease the probability for information criteria to select the right model asymptotically relative to OLS, depending on the information criteria and the true model. Simulation results confirm our asymptotic results in finite sample.

Keywords: Model Selection; Information Criteria; Local-to-unit-root Model; Mildly Explosive Model; Unit Root Model; Indirect Inference.

1 Introduction

Information criteria have found a wide range of practical applications in empirical work. Examples include choosing explanatory variables in regression models and selecting lag lengths in time series models. Frequently used information criteria are AIC of Akaike (1969, 1973),

---

*Yubo Tao, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: yubo.tao.2014@phdecons.smu.edu.sg. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: yu-jun@smu.edu.sg.
BIC of Schwarz (1978), HQIC of Hannan and Quinn (1979). A major nice feature in these information criteria is that the penalty term is trivial to compute and hence the implementation of them is straightforward and can be made automatic.

With a growing interest in nonstationarity in time series analysis, researchers have examined the properties of information criteria in the context of nonstationary models with the unit root behavior. Contributions include Tsay (1984), Pötscher (1989), Leeb and Pötscher (2005), and Phillips (2008). Another form of nonstationarity in time series involves explosive roots. Recent global financial crisis has motivated researchers to study explosive behavior in economic and financial time series; see, for example, Phillips and Yu (2011), Phillips, Wu and Yu (2011) and Phillips, Shi and Yu (2015a, b).

In this paper, we study the limit properties of information criteria for distinguishing between the unit root model and the explosive models. The information criteria considered in this paper have a general form and include AIC, BIC and HQIC as the special cases. The impact of the initial condition on the limit properties is examined by allowing for an initial condition of three different orders of magnitude. Moreover, both the OLS estimator and the indirect inference estimator are studied when investigating the limit properties of information criteria. The motivation for the use of indirect inference estimator comes from the existence of finite sample bias in the OLS estimator and the ability that the indirect inference method can reduce the bias.

It is found that information criteria consistently choose the unit root model against the explosive alternatives when data comes from the unit root model. Second, we prove that the probability for information criteria to correctly select the explosive model models against the unit root model depends crucially on both the degree of explosiveness and the size of the penalty term in information criteria. Finally, we show that indirect inference estimation can increase or decrease the probability for information criteria to select the right model asymptotically relative to OLS, depending on the information criteria and the true model.

The rest of this paper is organized as follows. Section 2 introduces the models and information criteria, and briefly reviews the literature. Section 3 gives the limit properties of information criteria for distinguishing models with an explosive root from the unit root model when the OLS estimator is used. Section 4 gives the limit properties of information criteria when the indirect inference estimator is used. Section 5 provides Monte Carlo evidence to support the theoretical results. Section 6 concludes. All the detailed proofs are provided in the appendix. To compress notation, we denote \( \int_0^1 B dB \) and \( \int_0^1 B^2 \) in short for \( \int_0^1 B(r) dB(r) \) and \( \int_0^1 B(r)^2 dr \) respectively throughout the paper, and \( \Rightarrow \) denotes weak convergence.
2 Models, Information Criteria and A Literature Review

The model considered in the present paper is of the form:

\[ X_t = \rho_n X_{t-1} + u_t, \quad t = 1, \ldots, n, \]  

(2.1)

where \( u_t \) is independent and identically distributed as \( (0, \sigma^2) \) and the model is initialized at \( t = 0 \) with some \( X_0 \). The autoregressive (AR) coefficient \( \rho_n \) is the crucial parameter that determines the dynamic behavior of \( X_t \). When \( \rho_n = \rho \) and \( |\rho| < 1 \), \( X_t \) is stationary. When \( \rho_n = 1 \), \( X_t \) has a unit root (UR hereafter). When \( \rho_n = 1 \) and \( c_n/n = 1 - c/n \) for \( c > 0 \), \( X_t \) is near-stationary and has a root that is local-to-unity (LTUS hereafter) (Phillips, 1987b; Chan and Wei, 1987). When \( \rho_n = \rho \) and \( |\rho| > 1 \), \( X_t \) has an explosive root (EX hereafter). When \( \rho_n = 1 + c_n/n = 1 + c/n \) for \( c > 0 \), \( X_t \) is near-explosive and also has a root that is local-to-unity (LTUE hereafter). When \( \rho_n = \rho \) and \( c_n/n \to \infty \) but \( c_n/n \to 0 \), the root represents moderate deviations from unity and \( X_t \) is near-stationary (Phillips and Magdalinos, 2007). When \( \rho_n = 1 + c_n/n \) for \( c_n \to \infty \) but \( c_n/n \to 0 \), \( X_t \) is mildly explosive (hereafter ME).

The asymptotic properties of the OLS estimator of the AR coefficient in the stationary AR(1) model is well known. The rate of convergence is \( \sqrt{n} \) and the limiting distribution is Gaussian. Phillips (1987a) provided the limiting theory for the OLS estimator in the UR model and the rate of convergence is \( n \). Phillips (1987b) and Chan and Wei (1987) established the asymptotic theory for the LTUS and LTUE models. The asymptotic theory is similar to that in the UR model and the rate of convergence is also \( n \). In the cases of UR and LTU, \( u_t \) can be weakly dependent stationary. Anderson (1959) studied the limiting distribution of the OLS estimator in the EX model under the condition that \( u_t \) is independent and identically distributed as \( N(0, \sigma^2) \) and \( X_0 = 0 \). The limiting distribution is Cauchy and the rate of convergence is \( n \). However, no invariance principle applies. Assuming \( X_0 = o_p(\sqrt{n/c_n}) \), Phillips and Magdalinos (2007) developed the asymptotic theory for the model with \( \rho_n = 1 - c_n/n \) for \( c_n \to \infty \) but \( c_n/n \to 0 \) and showed that the asymptotic distribution is invariant to the error distribution. The rate of convergence is \( n/\sqrt{c_n} \). If \( c_n = n^a \) with \( \alpha \in (0, 1) \), this rate of convergence bridges that of UR/LTU models and that of the stationary process. Phillips and Magdalinos (2007) also developed the asymptotic theory for the ME model. The rate of convergence is \( n\rho_n^2/c_n \). The limiting distribution is Cauchy which is the same as in the EX model. Interestingly, in the ME case, the asymptotic theory is independent of the initial condition as long as \( X_0 = o_p(\sqrt{n/c_n}) \).

It is known that the OLS estimator of \( \rho_n \) is biased downward when \( \rho_n = 1 \) or when \( \rho_n \) is in the vicinity of unity. In this case, the indirect inference estimation is effective in reducing the bias. Phillips (2012) derives the asymptotic theory of the indirect inference estimator when

\[ 1 \text{ There is a separate literature where the asymptotic distribution is developed in continuous time models; see Perron (1991), Tang and Chen (2009) and Wang and Yu (2016).} \]
the model is UR or LTU and $u_t \sim \mathcal{N}(0, \sigma^2)$. The rate of convergence remains unchanged while the limiting distribution is different from that of the OLS estimator.

Information criteria for model selection have been proposed by Akaike (1969, 1973), Schwarz (1978), Hannan and Quinn (1979), among many others. The general form of these criteria is

$$IC_k = \log \hat{\sigma}_k^2 + \frac{kp_n}{n},$$

where $k$ is the number of parameters to be estimated, $\hat{\sigma}_k^2$ is the estimated $\sigma^2$ when $k$ parameters are estimated. In general, $IC_k$ trades off the term that measures the goodness-of-fit (i.e. $\log \hat{\sigma}_k^2$) and the penalty term that measures the complexity of the model (i.e. $kp_n/n$).

Coefficient $p_n = 2, \log n, 2 \log \log n$ corresponds to AIC of Akaike (1973), BIC of Schwarz (1978) and HQIC of Hannan and Quinn (1979). Other forms of $p_n$ are possible.

In the time series literature, information criteria have been widely used to select the lag length both in the family of stationary models and in the family of nonstationary models. The important contributions in the context of nonstationary models include Tsay (1984), Pötscher (1989), Ng and Perron (1995), Ploberger and Phillips (2003), Leeb and Pötscher (2005). The information criteria can also be used to evaluate whether $\rho_n = 1$ (i.e. $k = 0$) or $\rho_n \neq 1$ (i.e. $k = 1$) in Model (2.1). For example, Phillips (2008) obtained limit properties of $IC_k$ for distinguishing between the unit root model and the stationary model. Phillips and Lee (2015) show that BIC can successfully distinguish the UR model from the ME model. This is a surprising result as it is well known that BIC cannot consistently distinguish between the UR model and the LTU model; see Ploberger and Phillips (2003) and Leeb and Pötscher (2005).

In this paper we focus our attention to distinguishability between the unit root model and the three explosive models (i.e., LTUE, ME and EX). Our paper is related to Phillips (2008) in the sense that a general form of information criteria is considered. However, unlike Phillips (2008) who compares the UR model with the stationary model, we compare the UR model with the explosive models. Our paper is also related to Phillips and Lee (2015) who compare the UR model and the ME model based on BIC. However, we consider a general form of information criteria and allow for different initial conditions. Our study is further different from Phillips (2008) and Phillips and Lee (2015) in that the indirect inference method is also used to estimate the autoregressive parameter.

To visually understand the difference between the UR model, the LTU model and the ME model, we simulate a sample path of different length ($n = 100, 200, 500, 1000$) with $y_0 = 0$, based on the same realizations of the error process, iid $\mathcal{N}(0, 1)$, from the following four models, $\rho_n = 1$ (UR), $\rho_n = 1 + 1/n$ (LTUE), $\rho_n = 1 + n^{0.1}/n$ (ME1), and $\rho_n = 1 + n^{0.5}/n$ (ME2). Figures 1-3 give the time series plot of UR against LTU, UR against ME1, UR against ME2, respectively. It can be seen from Figure 1 that it is very difficult to distinguish between the UR process and the LTU process, even when the sample size is as large as 1,000. When the sample size increases, the gap between the UR process and the two ME processes
becomes larger and larger, as apparent in Figure 2 and more so in Figure 3.

3 Limit Properties Based on the OLS Estimator

Following Phillips (2008), when the data generating process (DGP) is the UR model, since $\rho_n = 1$, we set the parameter count to $k = 0$. For the LTU model, the ME model and the EX model, we need to estimate the AR coefficient and hence set the parameter count to $k = 1$. Throughout the paper we denote $\hat{\rho}$ the OLS estimator of $\rho$. $\hat{k}_{IC} = 0$ or 1 means the information criterion of the UR model is smaller or larger than that of the competing model when $\rho$ is estimated by OLS. We aim to find the limit of the following probabilities:

$$
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 | k = 0 \right\} ; \\
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 | k = 0 \right\} ; \\
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} ; \\
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} .
$$

As shown in Phillips and Magdalinos (2009), the unit root asymptotic distribution is sensitive to initial conditions in the distant past. To understand how the initial condition
Figure 2: A realization of the UR model and the ME process with $\rho_n = 1 + n^{0.1}/n$ (ME1).

Figure 3: A realization of the UR model and the ME model with $\rho_n = 1 + n^{0.5}/n$ (ME2).
affects the property of $\hat{k}_{IC}$, we follow Phillips and Magdalinos (2009) by assuming alternative initial conditions.

**Assumption 1 (IN)** The initial condition has the form

$$X_0(n) = \sum_{j=0}^{\kappa_n} u_j,$$

where $\kappa_n$ is a sequence of integers satisfying $\kappa_n \to \infty$ and

$$\frac{\kappa_n}{n} \to \tau \in [0, \infty], \text{ as } n \to \infty.$$  

The following cases are distinguished:

(i) If $\tau = 0$, $X_0(n)$ is said to be a recent past initialization.

(ii) If $\tau \in (0, \infty)$, $X_0(n)$ is said to be a distant past initialization.

(iii) If $\tau = \infty$, $X_0(n)$ is said to be an infinite past initialization.

**Theorem 3.1** Under Assumption 1 (i) or (ii) or (iii), we have

(1) when $p_n \to \infty$ and $p_n/n \to 0$ as $n \to \infty$,

$$\lim_{n \to \infty} \mathbb{P} \left\{ \hat{k}_{IC} = 0 \mid k = 0 \right\} = \lim_{n \to \infty} \mathbb{P} \{ IC_0 - IC_1 \leq 0 \} = 1,$n \to \infty, \quad \lim_{n \to \infty} \mathbb{P} \{ IC_0 - IC_1 > 0 \} = 0.$$

(2) when $p_n = 2$, the asymptotic distribution under the AIC criterion is

$$\lim_{n \to \infty} \mathbb{P} \left\{ \hat{k}_{AIC} = 0 \mid k = 0 \right\} = \lim_{n \to \infty} \mathbb{P} \{ AIC_0 - AIC_1 \leq 0 \} = P \left( \xi^2 < 2 \right),$$

$$\lim_{n \to \infty} \mathbb{P} \left\{ \hat{k}_{AIC} = 1 \mid k = 0 \right\} = \lim_{n \to \infty} \mathbb{P} \{ AIC_0 - AIC_1 > 0 \} = 1 - P \left( \xi^2 < 2 \right).$$

where

$$\xi^2 = \begin{cases} \left( \int_0^1 B dB \right)^2, & \text{if } \tau = 0 \\ \int_0^1 B^2, & \text{if } \tau \in (0, \infty) \\ B(1)^2, & \text{if } \tau = \infty \end{cases},$$

with $B(s)$ being a Brownian motion, and

$$B_\tau(s) = B(s) + \sqrt{\tau}B_0(1),$$

with $B_0(s)$ being an independent Brownian motion.
Remark 3.2 Theorem 3.1 is the same as Theorem 1 in Phillips (2008) for distinguishing between the UR model and the stationary model. The condition that $p_n \to \infty$ and $p_n/n \to 0$ covers BIC and HQIC and hence, both BIC and HQIC can consistently select the UR model. The AIC criterion is inconsistent and its asymptotic distribution depends on $\xi^2$, the squared unit root $t$-statistic for the OLS estimator. Hence, we extend the asymptotic theory of Phillips (2008) to cover the LTUE, ME and EX alternatives.

Remark 3.3 The validity of Theorem 3.1 does not require the iid assumption for the error term $u_t$. If we follow Phillips (2008) by denoting $F(L) = \sum_{j=0}^{\infty} F_j L^j$, with $F_0 = 1$ and $F(1) \neq 0$, and letting $u_s$ have Wold representation

$$u_s = F(L) \varepsilon_s = \sum_{j=0}^{\infty} F_j \varepsilon_{s-j}, \text{ with } \sum_{j=0}^{\infty} j^{1/2} |F_j| < \infty,$$

(3.7)

where $\varepsilon_t \simiid (0, \sigma^2)$, the results in Theorem 3.1 continue to hold. However, both $B_0$ and $\xi^2$ need to be modified to accommodate the dependence in $u_t$ as in Phillips (2008).

Theorem 3.4 Case 1: Let Assumption 1 (i) or (ii) holds. Assume the true DGP is the LTUE model.

(1) When $p_n \to \infty$ and $p_n/n \to 0$ as $n \to \infty$,

$$\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) > 0 \right\} = 1,$$

$$\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) \leq 0 \right\} = 0.$$

(2) When $p_n = 2$, the asymptotic distribution of the AIC criterion is

$$\lim_{n \to \infty} P \left\{ \hat{k}_{AIC} = 0 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ n (AIC_1 - AIC_0) > 0 \right\} = 1 - P (\xi^2 > 2),$$

$$\lim_{n \to \infty} P \left\{ \hat{k}_{AIC} = 1 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ n (AIC_1 - AIC_0) \leq 0 \right\} = P (\xi^2 > 2),$$

where

$$\xi^2 = \left( \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} \right)^2 + 2c \int_0^1 J_c dB + c^2 \int_0^1 J_c^2 ,$$

with

$$J_c(r) = \int_0^r \exp \{c(r - s)\} dB(s).$$

Case 2: Let Assumption 1 (i) or (ii) holds. Assume the true DGP is the ME model.
(1) When \( \lim_{n \to \infty} \frac{p_n}{\rho_{2n}^2} = 0 \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_{2n}^2} (IC_1 - IC_0) > 0 \right\} = 0,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_{2n}^2} (IC_1 - IC_0) \leq 0 \right\} = 1.
\]

(2) When \( \lim_{n \to \infty} \frac{p_n}{\rho_{2n}^2} = \pi \in (0, +\infty) \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) > 0 \right\} = P \left( \chi^2(1) < 4\pi \right),
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) \leq 0 \right\} = 1 - P \left( \chi^2(1) < 4\pi \right).
\]

(3) When \( \lim_{n \to \infty} \frac{p_n}{\rho_{2n}^2} \to +\infty \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) > 0 \right\} = 1,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) \leq 0 \right\} = 0.
\]

**Case 3:** Let Assumption 1 (i) holds. Assume the true DGP is the EX model.

(1) When \( \lim_{n \to \infty} \frac{p_n}{\rho_{2n}^2} = 0 \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_{2n}^2} (IC_1 - IC_0) > 0 \right\} = 0,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_{2n}^2} (IC_1 - IC_0) \leq 0 \right\} = 1.
\]

(2) When \( \lim_{n \to \infty} \frac{p_n}{\rho_{2n}^2} = \pi \in (0, +\infty) \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_{2n}^2} (IC_1 - IC_0) > 0 \right\} = P \left( \chi^2(1) < (1 + \rho)^2\pi \right),
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_{2n}^2} (IC_1 - IC_0) \leq 0 \right\} = 1 - P \left( \chi^2(1) < (1 + \rho)^2\pi \right).
\]

(3) When \( \lim_{n \to \infty} \frac{p_n}{\rho_{2n}^2} \to +\infty \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) > 0 \right\} = 1,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) \leq 0 \right\} = 0.
\]
Remark 3.5 Case 1 in Theorem 3.4 shows that all the information criteria are inconsistent in distinguishing between the LTUE model and the UR models when data comes from the LTUE model. AIC selects the wrong model with probability going to $1 - P(\zeta^2 > 2)$, which depends on the localization constant $c$. This problem worsens for BIC and HQIC as the probability of selecting the wrong model goes to one. Note that BIC is well known to be blind to local alternatives; see, for example, Ploberger and Phillips (2003) and Leeb and Pötscher (2005).

Remark 3.6 Case 2 in Theorem 3.4 shows that the limit probability of selecting the correct model by information criteria under the ME model depends critically on two parameters, $c_n$, $p_n$. As expected, the larger $c_n$, the further the model away from the UR model and the higher probability for the information criteria to select the correct model. Interestingly, the smaller $p_n$, the higher probability for the information criteria to select the correct model. From Phillips and Magdalinos (2009), we know $\rho_n^{-n} = o(c_n^{-1})$ and hence $\rho_n^p/c_n \rightarrow +\infty$. In the special case where $c_n = n^\alpha$, for $\alpha \in (0, 1)$, $\lim_{n \rightarrow \infty} p_n/\rho^{2n} = 0$ no matter whether $p_n = 2$ or $\log n$ or $2 \log \log n$. In this case, all the well-known information criteria can consistently select the true model.

Remark 3.7 Case 3 in Theorem 3.4 shows that the limit probability of selecting the correct model by information criteria under the EX model depends also critically on two parameters, $\rho$, $p_n$. As expected, the larger $\rho$, the higher probability for the information criteria to select the correct model. Interestingly, the smaller $p_n$, the higher probability for the information criteria to select the correct model. If $p_n = 2$ or $\log n$ or $2 \log \log n$, $\lim_{n \rightarrow \infty} p_n/\rho^{2n} = 0$ and hence Case 3 (1) applies, suggesting that all the well-known information criteria can consistently select the true model.

Results in Theorem 3.4 can be extended to cover the LTUE model and the ME model with weakly dependent errors. The following proposition establishes the results for the ME model.

Proposition 3.8 Let Assumption 1 (i) or (ii) and the assumption specified in Equation (3.7) hold. Assume the true DGP is the ME model.

(1) When $\lim_{n \rightarrow \infty} \frac{p_n}{\rho_n^{2n}} = 0$,

$$
\lim_{n \rightarrow \infty} P\{\hat{k}_{IC} = 0 | k = 1\} = \lim_{n \rightarrow \infty} P\left\{\frac{n}{\rho_n^{2n}} (IC_1 - IC_0) > 0\right\} = 0,
$$

$$
\lim_{n \rightarrow \infty} P\{\hat{k}_{IC} = 1 | k = 1\} = \lim_{n \rightarrow \infty} P\left\{\frac{n}{\rho_n^{2n}} (IC_1 - IC_0) \leq 0\right\} = 1.
$$
(2) When \( \lim_{n \to \infty} \frac{p_n}{\rho_n^{2n}} = \pi \in (0, +\infty) \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 \middle| k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) > 0 \right\} = P \left( \chi^2(1) < \frac{4\pi}{\omega^2} \right),
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 \middle| k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) \leq 0 \right\} = 1 - P \left( \chi^2(1) < \frac{4\pi}{\omega^2} \right).
\]

where \( \omega^2 = \left( \sum_{j=0}^{\infty} F_j \right)^2 \).

(3) When \( \frac{p_n}{\rho_n^{2n}} \to +\infty \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 \middle| k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) > 0 \right\} = 1,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 \middle| k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) \leq 0 \right\} = 0.
\]

4 Limit Properties Based on the Indirect Inference Estimator

The OLS estimator of \( \rho_n \) in Model (2.1) is known to be biased and the bias is acute when \( \rho_n \) is close to unity. To reduce the bias, the indirect inference method of Smith (1993) and Gouréroux et al (1993) can be used if Model (2.1) is fully specified. Phillips (2012) derives the asymptotic theory of the indirect inference estimator when the model is UR or LTU and \( u_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \). Throughout the paper we denote \( \hat{\rho} \) the indirect inference estimator of \( \rho \). Let \( h(c) = c + g(c) \) and \( g(c) = g^-(c)1_{c \leq 0} + g^+(c)1_{c > 0} \) with

\[
g^-(c) = -\frac{3}{4} \int_0^\infty e^{-\frac{v}{c}} k^-(v; c)^{1/2} dv + \frac{1}{4} \int_0^\infty e^{-\frac{v}{c}} k^-(v; c)^{3/2} dv
\]

\[
- \frac{e^{2c}}{8} \int_0^\infty e^{-\frac{v}{c}} k^-(v; c)^{3/2} dv,
\]

\[
g^+(c) = \frac{3}{4} \int_0^\infty e^{\frac{v}{c}} k^+(w; c)^{1/2} dw - \frac{1}{4} \int_0^\infty e^{\frac{v}{c}} k^+(w; c)^{3/2} dw
\]

\[
- \frac{e^{2c}}{8} \int_0^\infty e^{\frac{v}{c}} k^+(w; c)^{3/2} dw,
\]

\[
k^-(v; c) = \frac{2v - 4c}{v + e^{2c} e^{-v} - 4c},
\]

\[
k^+(w; c) = \frac{2w + 4c}{w + e^{2c} w e^{w} + 4c}.
\]

Phillips (2012) shows that under the UR model,
\[ n(\hat{p} - 1) \Rightarrow h^{-1} \left( \int_0^1 B dB / \int_0^1 B^2 \right) \quad \text{as } n \to +\infty, \]

and under the LTUE model,
\[ n(\hat{p} - \rho_n) \Rightarrow h^{-1} \left( \int_0^1 J_c dB / \int_0^1 J_c^2 + c \right) - c \quad \text{as } n \to +\infty. \]

Let \( \check{k}_{IC} = 0 \) or \( 1 \) mean the information criterion of the UR model is smaller or larger than that of the competing model when the model is estimated by the indirect inference method. We aim to find is the limit of the following probabilities:
\[
\begin{align*}
\lim_{n \to \infty} P \left\{ \check{k}_{IC} = 0 \mid k = 1 \right\} &; \quad (4.1) \\
\lim_{n \to \infty} P \left\{ \check{k}_{IC} = 1 \mid k = 1 \right\} &; \quad (4.2) \\
\lim_{n \to \infty} P \left\{ \check{k}_{IC} = 0 \mid k = 0 \right\} &; \quad (4.3) \\
\lim_{n \to \infty} P \left\{ \check{k}_{IC} = 1 \mid k = 0 \right\} &. \quad (4.4)
\end{align*}
\]

**Theorem 4.1** Under Assumption 1(i) or (ii) or (iii), we have

(1) when \( p_n \to \infty \) and \( p_n/n \to 0 \) as \( n \to \infty \),
\[
\begin{align*}
\lim_{n \to \infty} P \left\{ \check{k}_{IC} = 0 \mid k = 0 \right\} & = \lim_{n \to \infty} P \left\{ IC_0 - IC_1 \leq 0 \right\} = 1, \\
\lim_{n \to \infty} P \left\{ \check{k}_{IC} = 1 \mid k = 0 \right\} & = \lim_{n \to \infty} P \left\{ IC_0 - IC_1 > 0 \right\} = 0;
\end{align*}
\]

(2) when \( p_n = 2 \), the asymptotic distribution under the AIC criterion is
\[
\begin{align*}
\lim_{n \to \infty} P \left\{ \check{k}_{AIC} = 0 \mid k = 0 \right\} & = P \left( \zeta^2 < 2 \right), \\
\lim_{n \to \infty} P \left\{ \check{k}_{AIC} = 1 \mid k = 0 \right\} & = 1 - P \left( \zeta^2 < 2 \right),
\end{align*}
\]

where
\[
\zeta^2 = \begin{cases} 
\int_0^1 B^2 \cdot h^{-1} \left( \left( \int_0^1 B dB / \int_0^1 B^2 \right)^2 \right) - 2 \int_0^1 B dB \cdot h^{-1} \left( \int_0^1 B dB / \int_0^1 B^2 \right) , & \text{if } \tau = 0 \\
\int_0^1 B^2 \cdot h^{-1} \left( \left( \int_0^1 B \tau dB / \int_0^1 B^2 \right)^2 \right) - 2 \int_0^1 B \tau dB \cdot h^{-1} \left( \int_0^1 B \tau dB / \int_0^1 B^2 \right) , & \text{if } \tau \in (0, \infty) \}
\end{cases}
\]

\[
\int_0^1 B \tau dB \cdot h^{-1} \left( \left( \int_0^1 B \tau dB / \int_0^1 B^2 \right)^2 \right) - 2 \int_0^1 B \tau dB \cdot h^{-1} \left( \int_0^1 B \tau dB / \int_0^1 B^2 \right) , & \text{if } \tau \in (0, \infty) \}
\end{cases}
\]

\[
\int_0^1 B^2 \cdot h^{-1} \left( \left( \int_0^1 B dB / \int_0^1 B^2 \right)^2 \right) - 2 \int_0^1 B dB \cdot h^{-1} \left( \int_0^1 B dB / \int_0^1 B^2 \right) , & \text{if } \tau = \infty
\end{cases}
\]

with \( \mathcal{C} \) being a standard Cauchy variate.
Remark 4.2 According to Theorem 4.1, as long as $p_n \rightarrow \infty$ and $p_n/n \rightarrow 0$, information criteria based on the indirect inference estimator is consistent in selecting the UR model. Hence, BIC and HQIC based on the indirect inference estimator can consistently select the UR model. Like the AIC criterion that is based on the OLS estimator, the AIC criterion based on the indirect inference estimator continues to be inconsistent. However, its asymptotic distribution depends on $\xi^2$, the squared unit root t-statistic for the indirect inference estimator.

Remark 4.3 As shown in Phillips (2012), the squared unit root t-statistic for the indirect inference estimator has a smaller variance than that of the squared unit root t-statistic for the OLS estimator. Consequently, $P(\xi^2 < 2) > P(\xi^2 < 2)$, suggesting that AIC based on the indirect inference estimator can select the true model (i.e. the UR model) with a larger probability than that based on the OLS estimator.

Theorem 4.4 Case 1: Let Assumption 1 (i) or (ii) holds. Assume the true DGP is the LTUE model.

(1) When $p_n \rightarrow \infty$ and $p_n/n \rightarrow 0$ as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} = \lim_{n \rightarrow \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) > 0 \right\} = 1,$$

$$\lim_{n \rightarrow \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} = \lim_{n \rightarrow \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) \leq 0 \right\} = 0.$$

(2) When $p_n = 2$, the asymptotic distribution under the AIC criterion is

$$\lim_{n \rightarrow \infty} P \left\{ \hat{k}_{AIC} = 0 | k = 1 \right\} = \lim_{n \rightarrow \infty} P \left\{ n (AIC_1 - AIC_0) > 0 \right\} = 1 - P(\vartheta^2 > 2),$$

$$\lim_{n \rightarrow \infty} P \left\{ \hat{k}_{AIC} = 1 | k = 1 \right\} = \lim_{n \rightarrow \infty} P \left\{ n (AIC_1 - AIC_0) \leq 0 \right\} = P(\vartheta^2 > 2),$$

where

$$\vartheta^2 \equiv 2h^{-1} \left( \int_0^1 J_c dB + c \right) \left( \int_0^1 J_c dB + c \int_0^1 J_c^2 \right) - h^{-1} \left( \int_0^1 J_c dB + c \right) \int_0^1 J_c^2.$$

Case 2: Let Assumption 1 (i) or (ii) holds. Assume the true DGP is the ME model.

(1) When $\lim_{n \rightarrow \infty} \frac{p_n}{\rho_n^{2n}} = 0$,

$$\lim_{n \rightarrow \infty} P \left\{ \hat{k}_{IC} = 0 | k = 1 \right\} = \lim_{n \rightarrow \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) > 0 \right\} = 0,$$

$$\lim_{n \rightarrow \infty} P \left\{ \hat{k}_{IC} = 1 | k = 1 \right\} = \lim_{n \rightarrow \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) \leq 0 \right\} = 1.$$

13
(2) When \( \lim_{n \to \infty} \frac{p_n}{\rho_n^{2n}} = \pi \in (0, +\infty) \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) > 0 \right\} = P \left( \chi^2(1) < 4\pi \right) ,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) \leq 0 \right\} = 1 - P \left( \chi^2(1) < 4\pi \right) .
\]

(3) When \( \frac{p_n}{\rho_n^{2n}} \to +\infty \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) > 0 \right\} = 1 ,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) \leq 0 \right\} = 0 .
\]

**Case 3:** Let Assumption 1 (i) holds. Assume the true DGP is the EX model.

1. When \( \lim_{n \to \infty} \frac{p_n}{\rho_n^{2n}} = 0 \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) > 0 \right\} = 0 ,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) \leq 0 \right\} = 1 .
\]

2. When \( \lim_{n \to \infty} \frac{p_n}{\rho_n^{2n}} = \pi \in (0, +\infty) \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) > 0 \right\} = P \left( \chi^2(1) < (1 + \rho)^2\pi \right) ,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{\rho_n^{2n}} (IC_1 - IC_0) \leq 0 \right\} = 1 - P \left( \chi^2(1) < (1 + \rho)^2\pi \right) .
\]

3. When \( \lim_{n \to \infty} \frac{p_n}{\rho_n^{2n}} \to +\infty \),

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 0 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) > 0 \right\} = 1 ,
\]

\[
\lim_{n \to \infty} P \left\{ \hat{k}_{IC} = 1 \mid k = 1 \right\} = \lim_{n \to \infty} P \left\{ \frac{n}{p_n} (IC_1 - IC_0) \leq 0 \right\} = 0 .
\]

**Remark 4.5** Case 1 in Theorem 4.4 shows that all the information criteria continue to be inconsistent in distinguishing between the LTUE model and the UR models when data come
from the LTUE model even when the indirect inference estimation is employed. AIC selects the wrong model with probability going to $1 - P(\hat{\sigma}^2 > 2)$. Since the variance of $\zeta^2$ is bigger than that of $\hat{\sigma}^2$, the tail probability of $\zeta^2$ is larger than that of $\hat{\sigma}^2$, suggesting that AIC based on OLS selects the true model (i.e. LTUE model) with a greater probability than AIC based on the indirect inference estimator.

**Remark 4.6** The results under Case 2 in Theorem 4.4 are the same as those under Case 2 in Theorem 3.4, suggesting all the well-known information criteria can consistently select the true model (i.e. ME model) when \(c_n = n^\alpha\), for \(\alpha \in (0, 1)\).

**Remark 4.7** The results under Case 3 in Theorem 4.4 are the same as those under Case 3 in Theorem 3.4, suggesting that all the well-known information criteria can consistently select the true model (i.e. EX model).

## 5 Monte Carlo Study

In this section, we examine the performance of alternative information criteria, namely, AIC, BIC and HQIC, in finite sample via simulated data and check the reliability of the asymptotic results developed in Section 3 and Section 4. In the simulation study, we use both OLS and the indirect inference method to estimate $\rho_n$ from sample paths that are simulated from different DGPs. In total we design four experiments. In the first experiment we simulate data from the UR model. In the second experiment we simulate data from the LTUE model with $c = 1$ (i.e. $\rho_n = 1 + 1/n$). In the third experiment we simulate data from two ME models with $c_n = n^{0.1}, n^{0.3}$, respectively. In the last experiment we simulate data from the EX model with $\rho = 1.01, 1.05$, respectively. In all experiments, we simulate 10,000 sample paths with initial value $X_0 = 0$ and four sample sizes are considered, $n = 100, 200, 500, 1000$. In each experiment, we report the fraction of the number of times in which the correct model is selected out of 10,000 replications.

Table 1 reports the results when the true DGP is UR. Several results can be found here. First, the probability for BIC and HQIC to select the true model grows as $n$ grows. However, the probability for AIC to select the true model does not seem to increase or decrease as $n$ grows. This observation is consistent with the asymptotic results reported in Theorem 3.1. Second, the probability for BIC to select the true model is larger than that in HQIC which is in turn larger than AIC in these four sample sizes. So we can conclude that the probability grows as $p_n$ increases since $2 < 2\log n < \log n$ when $100 \leq n \leq 1000$. Third, the probability implied by AIC based on the indirect inference estimator is larger than that based on OLS. This finding is consistent with Theorem 4.1 and Remark 4.3.

Table 2 report the results when the true DGP is the LTUE model with $c_n = 1$. Also reported is the value of $p_n/\rho_n^{2n}$. Several results can be found here. First, the probability for BIC and HQIC to select the true model becomes smaller as $n$ grows. However, the probability for AIC to select the true model does not seem to increase or decrease as $n$ grows.
Table 1: Probability of Selecting the Correct Model when Data Come from the UR Model

| \( n \) | 100          | 200          |
|--------|--------------|--------------|
| IC     | AIC BIC HQIC | AIC BIC HQIC |
| OLS    | 0.8160 0.9604 0.9020 | 0.8155 0.9751 0.9249 |
| IIE    | 0.8731 0.9702 0.9292 | 0.8742 0.9810 0.9445 |
| \( n \) | 500          | 1000         |
| IC     | AIC BIC HQIC | AIC BIC HQIC |
| OLS    | 0.8127 0.9849 0.9335 | 0.8195 0.9895 0.9402 |
| IIE    | 0.8704 0.9881 0.9508 | 0.8759 0.9918 0.9566 |

This observation is consistent with the asymptotic results reported in Case 1 in Theorem 3.4. Second, the probability implied by AIC based on the indirect inference estimator is smaller than that based on OLS. This finding is consistent with Case 1 (2) in Theorem 4.4 and Remark 4.5. Finally, it seems that AIC performs better than BIC and HQIC in all cases.

Table 2: Probability of Selecting the Correct Model when Data Come from the LTUE Model with \( c_n = 1 \)

| \( n \) | 100          | 200          |
|--------|--------------|--------------|
| IC     | AIC BIC HQIC | AIC BIC HQIC |
| \( p_n/\rho_n^{2n} \) | 0.2734 0.6295 0.4175 | 0.2720 0.7206 0.4536 |
| OLS    | 0.3516 0.1475 0.2420 | 0.3406 0.1305 0.2156 |
| IIE    | 0.1485 0.0445 0.0922 | 0.1235 0.0269 0.0663 |
| \( n \) | 500          | 1000         |
| IC     | AIC BIC HQIC | AIC BIC HQIC |
| \( p_n/\rho_n^{2n} \) | 0.2712 0.8427 0.4955 | 0.2709 0.9358 0.5236 |
| OLS    | 0.3474 0.1019 0.1933 | 0.3416 0.0871 0.1823 |
| IIE    | 0.1169 0.0134 0.0517 | 0.1089 0.0090 0.0394 |

Table 3 report the results when the true DGP is the ME model with \( c_n = n^{0.1}, n^{0.3} \). Also reported is the value of \( p_n/\rho_n^{2n} \). Several results can be found here. First, the probability for all three information criteria to select the true model grows as \( n \) increases. This observation is consistent with the asymptotic results reported in Case 2(1) in Theorem 3.4 and Remark 4.6. Second, comparing the results for \( c_n = n^{0.1} \) and those for \( c_n = n^{0.3} \), the probability for all three information criteria to select the true model increases when \( c_n \) is bigger. Third, the probability based on the indirect inference estimator is smaller than that based on OLS. Finally, it seems that AIC performs better than BIC and HQIC in all cases.
Table 3: Probability of Selecting the Correct Model when Data Come from the ME Model with $c_n = n^{0.1}$ and $c_n = n^{0.3}$

| ME Model with $c_n = n^{0.1}$ | 100       | 200       |
|-------------------------------|-----------|-----------|
| $n$                           | IC  | AIC   | BIC   | HQIC | IC  | AIC   | BIC   | HQIC |
| $p_n/\rho_n^{2n}$             |     |       |       |      |     |       |       |      |
| OLS                           | 0.0861 | 0.1983 | 0.1316 |      | 0.0679 | 0.1799 | 0.1132 |      |
| IIE                           | 0.3071 | 0.1741 | 0.2406 |      | 0.3211 | 0.1624 | 0.2250 |      |
| $n$                           |     |       |       |      |     |       |       |      |
| IC  | AIC   | BIC   | HQIC  | AIC  | BIC   | HQIC  |
| $p_n/\rho_n^{2n}$             | 0.0486 | 0.1512 | 0.0889 | 0.0371 | 0.1282 | 0.0718 |
| OLS                           | 0.6151 | 0.4083 | 0.5048 | 0.6469 | 0.4374 | 0.5494 |
| IIE                           | 0.3544 | 0.2008 | 0.2815 | 0.3925 | 0.2351 | 0.3129 |

| ME Model with $c_n = n^{0.3}$ |
|-------------------------------|-----------|-----------|
| $n$                           | IC  | AIC   | BIC   | HQIC |
| $p_n/\rho_n^{2n}$             | 0.0008 | 0.0019 | 0.0012 |      |
| OLS                           | 0.9374 | 0.9066 | 0.9235 |      |
| IIE                           | 0.9274 | 0.8979 | 0.9163 |      |
| $n$                           |     |       |       |      |
| IC  | AIC   | BIC   | HQIC  |
| $p_n/\rho_n^{2n}$             | 1.0e-06 | 1.0e-05 | 1.0e-06 | 1.0e-07 | 1.0e-07 | 1.0e-07 |
| OLS                           | 0.9948 | 0.9907 | 0.9938 | 0.9988 | 0.9985 | 0.9986 |
| IIE                           | 0.9938 | 0.9901 | 0.9933 | 0.9986 | 0.9985 | 0.9985 |

17
Table 4 report the results when the true DGP is the EX model with \( \rho = 1.01, 1.05 \). Also reported is the value of \( p_n/\rho^{2n} \). Several results can be found here. First, when \( \rho = 1.01 \), which is larger than the unity by 1%, the probability for information criteria to select the correct model is small in all cases when the sample size is small. However, it grows very quickly with the sample size. When \( \rho = 1.05 \), the probability for information criteria to select the correct model is almost 1 in all cases even when the sample size is small and increases with the sample size. Finally, it seems that AIC performs better than BIC and HQIC in all cases.

Table 4: Probability of Selecting the Correct Model when Data Come from the Regular Explosive Model with \( \rho = 1.01, 1.05 \).

| \( n \) | \( p_n/\rho^{2n} \) | IC | AIC | BIC | HQIC |
|---|---|---|---|---|---|
| 100 | 0.2734 | 0.6295 | 0.4175 | 0.0374 | 0.0990 | 0.0623 |
| 200 | 0.0485 | 0.0445 | 0.0922 | 0.4740 | 0.3059 | 0.3845 |
| 500 | 1.0e-4 | 1.0e-4 | 1.0e-4 | 1.0e-9 | 1.0e-8 | 1.0e-9 |
| 1000 | 0.9775 | 0.9599 | 0.9704 | 0.9998 | 0.9997 | 0.9998 |
|  | 0.9733 | 0.9563 | 0.9681 | 0.9998 | 0.9997 | 0.9998 |

| \( n \) | \( p_n/\rho^{2n} \) | IC | AIC | BIC | HQIC |
|---|---|---|---|---|---|
| 100 | 0.0001 | 0.0003 | 0.0002 | 1.0e-07 | 1.0e-07 | 1.0e-07 |
| 200 | 0.9741 | 0.9643 | 0.9681 | 0.9999 | 0.9998 | 0.9998 |
| 500 | 0.9703 | 0.9626 | 0.9655 | 0.9999 | 0.9998 | 0.9998 |
| 1000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 18 |
6 Conclusion

This paper studies the limit properties of information criteria for distinguishing between unit root model and three types of explosive models. Both the OLS estimator and the indirect inference estimator are employed to estimate the AR coefficient in the candidate model. This paper contributes to the literature in three aspects. First, our results extends results in Phillips (2008) to the other side of the unit root, and we find that information criteria consistently choose the unit root model when the unit root model is the true model. Second, we show that the limiting probabilities for information criteria to select the explosive model depends on both the distance of autoregressive coefficient from unity and the size of penalty term in the information criteria. When the penalty term is not too large and the root is not too close to unit root, all the information criteria consistently select the true model. It is known that the indirect inference method is effective in reducing the bias in OLS estimation in all cases as well as reducing the variance in OLS estimation in the UR model and in the LTU model. However, when information criteria are used in connection with the indirect inference estimation, the limiting probabilities for information criteria to select the correct model can go up or down relative to that with the OLS estimation, depending on the true DGP. When the true DGP is the UR model, the indirect inference estimation increases the probability. When the true DGP is the LTUE model or the ME model or the EX model, the indirect inference estimation decreases the probability.

Appendix

A Proof of Theorem 3.1

The proof is same as the proof for Theorem 1 in Phillips (2008), and hence omitted.

B Proof of Theorem 3.4

Case 1: when the true DGP is the LTUE model, we have $0 < c < \infty$ and

$$IC_0 = \log \hat{\sigma}_0^2 = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - X_{t-1})^2 \right\}$$

$$= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - 1)X_{t-1} + u_t]^2 \right\}$$

$$= \log \left\{ \frac{1}{n} (\rho_n - 1)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - 1) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}.$$
By Lemma 1 in Phillips (1987b), when the process is initialized at \( X_0 \), we know
\[
\frac{1}{n^2} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 J_c^2, \tag{6.1}
\]
and
\[
\frac{1}{n} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \sigma^2 \int_0^1 J_c dB, \tag{6.2}
\]
where
\[
J_c(r) = \int_0^r e^{c(r-s)} dB(s).
\]

Therefore, by Equation (6.1) and (6.2) we have
\[
IC_0 = \log \left\{ \frac{\sigma^2 e^2}{n} \int_0^1 J_c^2 + \frac{2c \sigma^2}{n} \int_0^1 J_c dB + \sigma^2 + o_p(n^{-1}) \right\}
= \log \sigma^2 + \log \left\{ 1 + \frac{2c}{n} \int_0^1 J_c(r) dB + \frac{c^2}{n} \int_0^1 J_c^2 + o_p(n^{-1}) \right\}. \tag{6.3}
\]

We also know from Phillips (1987b) that
\[
n(\hat{\rho}_n - \rho_n) \Rightarrow \int_0^1 \frac{J_c dB}{\int_0^1 J_c^2}, \tag{6.4}
\]
Hence,
\[
IC_1 = \log \sigma_1^2 + \frac{p_n}{n}
= \log \left\{ n^{-1} \sum_{t=1}^{n} (X_t - \hat{\rho}_n X_{t-1})^2 \right\} + \frac{p_n}{n}
= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - \hat{\rho}_n) X_{t-1} + u_t]^2 \right\} + \frac{p_n}{n}
= \log \left\{ \frac{1}{n} (\rho_n - \hat{\rho}_n)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - \hat{\rho}_n) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}
\Rightarrow \log \left\{ \frac{-\sigma^2 \left( \int_0^1 J_c dB \right)^2}{n \int_0^1 J_c^2} + \sigma^2 \right\} + \frac{p_n}{n}
= \log \sigma^2 + \log \left\{ 1 - \frac{1}{n} \left( \int_0^1 J_c dB \right)^2 \right\} + \frac{p_n}{n}. \tag{6.5}
\]
Therefore, by Equation (6.3) and (6.5), we have

\[ IC_1 - IC_0 \Rightarrow \log \left\{ 1 - \frac{1}{n} \left( \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} \right)^2 \right\} - \log \left\{ 1 + \frac{2c}{n} \int_0^1 J_c dB + \frac{c^2}{n} \int_0^1 J_c^2 \right\} + \frac{p_n}{n}. \]

Hence, if \( p_n = 2 \) (as in AIC), as \( n \to \infty \), we have

\[ n (IC_1 - IC_0) \Rightarrow 2 - \frac{\left( \int_0^1 J_c dB \right)^2}{\int_0^1 J_c^2} - 2c \int_0^1 J_c dB - c^2 \int_0^1 J_c^2. \quad (6.6) \]

If \( p_n \to \infty \) and \( \frac{p_n}{n} \to 0 \), we have

\[ \frac{n}{p_n} (IC_1 - IC_0) \Rightarrow 1. \]

**Case 2:** When the true DGP is the ME model, we have

\[
IC_0 = \log \hat{\sigma}_0^2 = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - X_{t-1})^2 \right\} \\
= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - 1) X_{t-1} + u_t]^2 \right\} \\
= \log \left\{ \frac{1}{n} (\rho_n - 1)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - 1) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}. \quad (6.7)
\]

According to Phillips and Magdalinos (2007), when the process is initialized at \( X_0 = \alpha_p(\sqrt{n}/c_n) \), we have

\[
\frac{cn\rho_n^{-2n}}{n^2} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \frac{\sigma^2}{4} Y^2, \quad (6.8)
\]

\[
\frac{cn\rho_n^{-n}}{n} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \frac{\sigma^2}{2} XY, \quad (6.9)
\]

and

\[
\frac{n\rho_n^n}{2cn} \left( \hat{\rho}_n - \rho_n \right) \Rightarrow \frac{X}{Y} \sim C, \quad (6.10)
\]

where \( X, Y \sim N(0, 1) \) and \( C \) is a standard Cauchy variate.

Therefore, by (6.8) and (6.9) we have

\[
IC_0 \Rightarrow \log \left\{ \frac{\sigma^2}{4n\rho_n^{-2n}} Y^2 + \frac{\sigma^2}{n\rho_n^{-n}} XY + \sigma^2 \right\} \\
= \log \sigma^2 + \log \left\{ \frac{1}{4n\rho_n^{-2n}} Y^2 + \frac{1}{n\rho_n^{-n}} XY + 1 \right\}. \quad (6.11)
\]
On the other hand,
\[ IC_1 = \log \hat{\sigma}^2 + \frac{p_n}{n} \]
\[ = \log \left\{ n^{-1} \sum_{t=1}^{n} (X_t - \hat{\rho}_n X_{t-1})^2 \right\} + \frac{p_n}{n} \]
\[ = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - \hat{\rho}_n) X_{t-1} + u_t]^2 \right\} + \frac{p_n}{n} \]
\[ = \log \left\{ \frac{1}{n} (\rho_n - \hat{\rho}_n)^2 \sum_{t=1}^{n} X^2_{t-1} + \frac{2}{n} (\rho_n - \hat{\rho}_n) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}. \]

By equation (6.8) to (6.10), we obtain
\[
IC_1 \Rightarrow \log \left\{ \frac{1}{n} \frac{4 c_n^2}{n^2 \rho_n^2} \sigma^2 - \frac{2 c_n}{n \rho_n^2} Y^2 X \sigma^2 - \frac{2 c_n}{n \rho_n^2} \right\} + \frac{p_n}{n} \]
\[ = \log \sigma^2 + \log \left\{ -\frac{1}{n} X^2 + 1 \right\} + \frac{p_n}{n}. \]

Therefore, by equation (6.11) and (6.12), we have
\[
IC_1 - IC_0 \Rightarrow \log \left\{ 1 - \frac{1}{n} X^2 \right\} - \log \left\{ 1 + \frac{1}{4 n \rho_n^{-2 n}} Y^2 - \frac{1}{n \rho_n^{-2 n}} X Y + \sigma^2 \right\} + \frac{p_n}{n}.
\]

Note \( X^2, Y^2 \sim \chi^2(1) \) and \( \rho_n^{-n} = o(c_n^{-1}) \). If \( \lim_{n \to \infty} \frac{p_n}{\rho_n^{2 n}} = \pi \),
\[
\frac{n}{k_n} (IC_1 - IC_0) \Rightarrow \begin{cases} 
\pi - \frac{1}{4} \chi^2(1), & \text{if } \pi \in [0, \infty) \\
1, & \text{if } \pi = \infty
\end{cases},
\]
where
\[
k_n = \begin{cases} 
{\rho_n^{2 n}}, & \text{if } \pi \in [0, \infty) \\
p_n, & \text{if } \pi = \infty
\end{cases}.
\]

**Case 3:** When the true DGP is EX model, we have
\[
IC_0 = \log \hat{\sigma}^2 = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - X_{t-1})^2 \right\} \]
\[ = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho - 1) X_{t-1} + u_t]^2 \right\} \]
\[ = \log \left\{ \frac{1}{n} (\rho - 1)^2 \sum_{t=1}^{n} X^2_{t-1} + \frac{2}{n} (\rho - 1) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}.
\]
By results established in Anderson (1959), we know

\[ \frac{1}{\rho^{2n}} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \frac{\sigma^2 Y^2}{(\rho^2 - 1)^2}, \quad (6.13) \]

\[ \frac{1}{\rho^n} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \frac{\sigma^2 XY}{\rho^2 - 1}, \quad (6.14) \]

\[ \frac{\rho^n}{\rho^2 - 1} (\rho - \rho) \Rightarrow C, \quad (6.15) \]

where \( X, Y \sim iid \mathcal{N}(0, 1) \) and \( C \) is a standard Cauchy variate. Then we have

\[ IC_0 = \log \left\{ \frac{\sigma^2 \rho^{2n}}{n(\rho + 1)^2} X^2 + \frac{2\sigma^2 \rho^n}{n(\rho + 1)} XY + \sigma^2 \right\} \]

\[ = \log \sigma^2 + \log \left\{ \frac{\rho^{2n}}{n(\rho + 1)^2} X^2 + \frac{2\rho^n}{n(\rho + 1)} XY + 1 \right\}. \quad (6.16) \]

For the OLS estimator for the general explosive series, we have

\[ IC_1 = \log \sigma_1^2 + \frac{p_n}{n} \]

\[ = \log \left\{ n^{-1} \sum_{t=1}^{n} (X_t - \hat{\rho} X_{t-1})^2 \right\} + \frac{p_n}{n} \]

\[ = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho - \hat{\rho}) X_{t-1} + u_t]^2 \right\} + \frac{p_n}{n} \]

\[ = \log \left\{ \frac{1}{n} (\rho - \hat{\rho})^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho - \hat{\rho}) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}. \]

By equation (6.13) to (6.15), we have

\[ IC_1 = \log \sigma^2 + \log \left\{ 1 - \frac{1}{n} X^2 \right\} + \frac{p_n}{n}. \quad (6.17) \]

Now, by equation (6.16) and (6.17), we obtain

\[ IC_1 - IC_0 = \log \left\{ 1 - \frac{1}{n} X^2 \right\} - \log \left\{ 1 + \frac{2\rho^n}{n(\rho + 1)} XY + \frac{\rho^{2n}}{n(\rho + 1)^2} X^2 \right\} + \frac{p_n}{n}. \]

Since \( \lim_{n \to \infty} \rho^{2n} = \pi \), we have

\[ \frac{n}{k_n} (IC_1 - IC_0) \Rightarrow \begin{cases} \pi - \frac{1}{(1 + \rho)^2} X^2(1), & \text{if } \pi \in [0, \infty) \\ 1, & \text{if } \pi = \infty \end{cases}, \]

where

\[ k_n = \begin{cases} \rho^{2n}, & \text{if } \pi \in [0, \infty) \\ p_n, & \text{if } \pi = \infty \end{cases}. \]

23
C Proof of Proposition 3.8

When the true DGP is ME model, we have $0 < c < \infty$, and

$$IC_0 = \log \hat{\sigma}_0^2 = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - X_{t-1})^2 \right\}$$

$$= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - 1) X_{t-1} + u_t]^2 \right\}$$

$$= \log \left\{ \frac{1}{n} (\rho_n - 1)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - 1) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}.$$  

When the process is initialized at $X_0$, by Lemma 5 in Magdalinos (2012), we know

$$\frac{c_n}{\omega^2 n^2} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \frac{\sigma^2}{4} Z^2,$$  \hspace{1cm} (6.18)

and

$$\frac{c_n}{\omega^2 n^2} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \frac{\sigma^2}{2} Y Z,$$  \hspace{1cm} (6.19)

where by Lemma 2 in Magdalinos (2012), we know $Y$ and $Z$ are independent $\mathcal{N}(0,1)$ variates with $\omega^2 = \left( \sum_{j=0}^{\infty} F_j \right)^2$.

Therefore, by Equation (6.18) and (6.19) we have

$$IC_0 \Rightarrow \log \left\{ \frac{\omega^2 \sigma^2}{4n\rho_n^{-2n}} Z^2 + \frac{\omega^2 \sigma^2}{n\rho_n^{-n}} Y Z + \sigma^2 + o_p(n^{-1}) \right\}$$

$$= \log \sigma^2 + \log \left\{ 1 + \frac{\omega^2}{n\rho_n^{-n}} Y Z + \frac{\omega^2}{4n\rho_n^{-2n}} Z^2 + o_p(n^{-1}) \right\}.$$  \hspace{1cm} (6.20)

We also know from Magdalinos (2012) that

$$\frac{n\rho_n^n}{2c_n} (\hat{\rho}_n - \rho_n) \Rightarrow C.$$  \hspace{1cm} (6.21)
Hence,

\[
\text{IC}_1 = \log \sigma_1^2 + \frac{p_n}{n}
\]

\[
= \log \left\{ n^{-1} \sum_{t=1}^{n} (X_t - \hat{\rho}_n X_{t-1})^2 \right\} + \frac{p_n}{n}
\]

\[
= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - \hat{\rho}_n) X_{t-1} + u_t]^2 \right\} + \frac{p_n}{n}
\]

\[
= \log \left\{ \frac{1}{n} (\rho_n - \hat{\rho}_n)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - \hat{\rho}_n) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}
\]

\[
\Rightarrow \log \left\{ -\frac{\omega^2 \sigma^2 Y^2}{n} + \sigma^2 \right\} + \frac{p_n}{n}
\]

\[
= \log \sigma^2 + \log \left\{ 1 - \frac{\omega^2 Y^2}{n} \right\} + \frac{p_n}{n}.
\]  

(6.22)

Therefore, by Equation (6.20) and (6.22), we have

\[
\text{IC}_1 - \text{IC}_0 \Rightarrow \log \left\{ 1 - \frac{\omega^2 Y^2}{n} \right\} - \log \left\{ 1 + \frac{\omega^2}{n \rho_n^{-n}} Y Z + \frac{\omega^2}{4n \rho_n^{-2n}} Z^2 \right\} + \frac{p_n}{n}.
\]

Note \(Y^2, Z^2 \sim \chi^2(1)\) and \(\rho_n^{-n} = o(\rho_n^{-1})\). If \(\lim_{n \to \infty} \frac{p_n}{n \rho_n^{-2n}} = \pi\),

\[
\frac{n}{k_n} (\text{IC}_1 - \text{IC}_0) \Rightarrow \begin{cases} 
\pi - \frac{\omega^2}{4} \chi^2(1), & \text{if } \pi \in [0, \infty), \\
1, & \text{if } \pi = \infty
\end{cases}
\]

where

\[
k_n = \begin{cases} 
\rho_n^{-2n}, & \text{if } \pi \in [0, \infty) \\
p_n, & \text{if } \pi = \infty
\end{cases}
\]

**D Proof of Theorem 4.1**

When the true DGP is the UR model, we have

\[
\text{IC}_0 = \log \hat{\sigma}_0^2 = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} = \log \sigma^2.
\]
Also, we have

\[ IC_1 = \log \sigma_1^2 + \frac{p_n}{n} = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - \tilde{\rho} X_{t-1})^2 \right\} + \frac{p_n}{n} \]

\[ = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(1 - \tilde{\rho}) X_{t-1} + u_t]^2 \right\} + \frac{p_n}{n} \]

\[ = \log \left\{ \frac{1}{n} (1 - \tilde{\rho})^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (1 - \tilde{\rho}) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}. \]

According to Phillips (2012), we have

\[
\tilde{\rho} - 1 \Rightarrow \begin{cases} 
\frac{1}{n} h^{-1} \left( \frac{1}{0} BdB \right), & \text{if } \tau = 0 \\
\frac{1}{n} h^{-1} \left( \frac{1}{0} B^2 \right), & \text{if } \tau \in (0, \infty) \\
\frac{1}{\sqrt{n^2/c_n}} h^{-1} (C), & \text{if } \tau = \infty 
\end{cases}
\]

where \( h(c) \) was defined in Section 4.

According to Phillips and Magdalinos (2009), we have

\[ \frac{1}{n^2} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \begin{cases} 
\sigma^2 \int_0^1 B^2, & \text{if } \tau = 0 \\
\sigma^2 \int_0^1 B_{\tau}^2, & \text{if } \tau \in (0, \infty) \\
\sigma^2 B_0(1)^2/c_n, & \text{if } \tau = \infty 
\end{cases} 
\]

\[ \frac{1}{n} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \begin{cases} 
\sigma^2 \int_0^1 B dB, & \text{if } \tau = 0 \\
\sigma^2 \int_0^1 B_{\tau} dB, & \text{if } \tau \in (0, \infty) \\
\sqrt{1/c_n \sigma^2 B(1) B_0(1)}, & \text{if } \tau = \infty 
\end{cases} 
\]

Therefore, we have

\[ IC_0 - IC_1 \Rightarrow \begin{cases} 
- \log \left\{ \int_0^1 B^2 n h^{-1} \left( \frac{1}{0} BdB \right)^2 - \frac{2}{n} \left( \int_0^1 B dB \right) h^{-1} \left( \frac{1}{0} BdB \right) + 1 \right\} - \frac{p_n}{n} \\
- \log \left\{ \int_0^1 B_{\tau}^2 n h^{-1} \left( \frac{1}{0} B_{\tau} dB \right)^2 - \frac{2}{n} \left( \int_0^1 B_{\tau} dB \right) h^{-1} \left( \frac{1}{0} B_{\tau} dB \right) + 1 \right\} - \frac{p_n}{n} \\
- \log \left\{ \frac{1}{n} h^{-1} (C)^2 B_0(1)^2 - \frac{2}{n} h^{-1} (C) B(1) B_0(1) + 1 \right\} - \frac{p_n}{n} 
\end{cases} \]
E Proof of Theorem 4.4

Case 1: when the true DGP is the LTUE model, we have $0 < c < \infty$. There is no difference between $IC_0$ based on the OLS estimator and that based on the indirect inference estimator. For $IC_1$, we have

$$ IC_1 = \log \hat{c}^2 + \frac{p_n}{n} $$

$$ = \log \left\{ n^{-1} \sum_{t=1}^{n} (X_t - \hat{\rho}_n X_{t-1})^2 \right\} + \frac{p_n}{n} $$

$$ = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (\rho_n - \hat{\rho}_n) X_{t-1} + \frac{2}{n} (\rho_n - \hat{\rho}_n) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}. $$

By the limit theory for the indirect inference estimator developed in Phillips (2012), we have

$$ n (\hat{\rho}_n - \rho_n) \Rightarrow h^{-1} \left( \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c \right) - c. \quad (6.23) $$

By equation (6.1), (6.2) and (6.23), we have

$$ IC_1 \Rightarrow \log \left\{ 1 - \frac{2}{n} \left[ h^{-1} \left( \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c \right) - c \right] \int_0^1 J_c dB + \frac{1}{n} \left[ h^{-1} \left( \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c \right) - c \right]^2 \int_0^1 J_c^2 \right\} $$

$$ + \log \sigma^2 + \frac{p_n}{n}. \quad (6.24) $$

Therefore, by equation (6.11) and (6.23), we have

$$ IC_1 - IC_0 \Rightarrow \log \left\{ 1 - \frac{2}{n} \int_0^1 J_c dB \left[ h^{-1} \left( \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c \right) - c \right] + \frac{\int_0^1 J_c^2}{n} \left[ h^{-1} \left( \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c \right) - c \right]^2 \right\} $$

$$ - \log \left\{ 1 + \frac{2c}{n} \int_0^1 J_c dB + \frac{c^2}{n} \int_0^1 J_c^2 \right\} + \frac{p_n}{n}. $$

When $p_n = 2$, as $n \to \infty$ we have

$$ n (IC_1 - IC_0) \Rightarrow 2 - \vartheta^2. $$

where

$$ \vartheta^2 \equiv 2h^{-1} \left( \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c \right) \left( \int_0^1 J_c dB + c \int_0^1 J_c^2 \right) - h^{-1} \left( \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c \right)^2 \int_0^1 J_c^2. $$

27
When $p_n \to \infty$ and $\frac{p_n}{n} \to 0$, we have
\[
\frac{n}{p_n} (IC_1 - IC_0) \Rightarrow 1.
\]

**Case 2:** When the true DGP is the ME model, we have
\[
IC_0 = \log \hat{\sigma}_0^2 = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - X_{t-1})^2 \right\}
= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - 1) X_{t-1} + u_t]^2 \right\}
= \log \left\{ \frac{1}{n} (\rho_n - 1)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - 1) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}.
\]

By equation (6.8) and (6.9) we have
\[
IC_0 \Rightarrow \log \left\{ \frac{\sigma^2}{4n p_n^{-2n}} Y^2 + \frac{\sigma^2}{n \rho_n^{-n}} X Y + \sigma^2 \right\}
= \log \sigma^2 + \log \left\{ \frac{1}{4n p_n^{-2n}} Y^2 + \frac{1}{n \rho_n^{-n}} X Y + 1 \right\}. \tag{6.25}
\]

Similarly, for $IC_1$ based on the indirect inference estimator, we have
\[
IC_1 = \log \hat{\sigma}_1^2 + \frac{p_n}{n}
= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - \hat{\rho}_n X_{t-1})^2 \right\} + \frac{p_n}{n}
= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - \hat{\rho}_n) X_{t-1} + u_t]^2 \right\} + \frac{p_n}{n}
= \log \left\{ \frac{1}{n} (\rho_n - \hat{\rho}_n)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - \hat{\rho}_n) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}
\]

Using the results in Phillips (2012), equation (6.8) and (6.9), we obtain
\[
IC_1 \Rightarrow \log \left\{ \frac{1}{n} \frac{4c_n^2}{n^2 \rho_n^{2n}} \left( \mathcal{C} + O \left( \frac{1}{2c_n} \right) \right)^2 \frac{n^2 \sigma^2}{4c_n^2 \rho_n^{-2n}} Y^2 - \frac{2}{n} \frac{2c_n}{n \rho_n^{n}} \left( \mathcal{C} + O \left( \frac{1}{2c_n} \right) \right) \frac{n \sigma^2}{2c_n \rho_n^{-n}} X Y + \sigma^2 \right\} + \frac{p_n}{n}
\Rightarrow \log \sigma^2 + \log \left\{ 1 - \frac{1}{n} X^2 + O \left( \frac{1}{c_n n} \right) \right\} + \frac{p_n}{n}. \tag{6.26}
\]

Therefore, the similar results to those in Theorem 3.4 are obtained.
Case 3: When the true DGP is the EX model, for the indirect inference estimator, we know that for $IC_0$, it is the same as OLS estimator. Therefore, we only need to derive the $IC_1$. Note that for $IC_1$, we have

$$IC_1 = \log \left\{ \frac{1}{n} (\rho - \hat{\rho})^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho - \hat{\rho}) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}.$$ 

According to the results in Phillips (2012), for $|\rho| > 1$, we know the binding function for $\rho$ is

$$b_n(\rho) = \rho + O(\rho^{-n}).$$

Therefore, we obtain

$$IC_1 = \log \sigma^2 + \log \left\{ \frac{1}{n} \left( C + O \left( \frac{1}{n \rho^2 - 1} \right) \right)^2 Y^2 - \frac{2}{n} \left( C + O \left( \frac{1}{n \rho^2 - 1} \right) \right) XY + 1 \right\} + \frac{p_n}{n}$$

$$= \log \sigma^2 + \log \left\{ 1 - \frac{1}{n} X^2 + O \left( \frac{1}{n \rho^2 - 1} \right) \right\} + \frac{p_n}{n}. \quad (6.27)$$

Now, by equation (6.16) and (6.27), we obtain

$$IC_1 - IC_0 = \log \left\{ 1 - \frac{1}{n} X^2 + O \left( \frac{1}{n \rho^2 - 1} \right) \right\} - \log \left\{ 1 + \frac{2 \rho^n}{n \rho + 1} XY + \frac{\rho^{2n}}{n \rho + 1} X^2 \right\} + \frac{p_n}{n}.$$ 

Since $\lim_{n \to \infty} \frac{p_n}{n \rho^{2n}} = \pi$, we have

$$\frac{n}{k_n} (IC_1 - IC_0) \Rightarrow \begin{cases} \frac{\pi - (1 + \rho) X^2(1)}{(1 + \rho^2)^2}, & \text{if } \pi \in [0, \infty) \\ 1, & \text{if } \pi = \infty \end{cases},$$

where

$$k_n = \begin{cases} \rho^{2n}, & \text{if } \pi \in [0, \infty) \\ p_n, & \text{if } \pi = \infty \end{cases}.$$ 

References

[1] Akaike, Hirotugu. 1969. “Fitting autoregressive models for prediction.” Annals of the Institute of Statistical Mathematics, 21(1): 243–247.

[2] Akaike, Hirotugu. 1973. “Information theory and an extension of the maximum likelihood principle.” Second International Symposium on Information Theory, Springer Verlag, 1, 267-281.

[3] Anderson, Theodore W. 1959. “On asymptotic distributions of estimates of parameters of stochastic difference equations.” The Annals of Mathematical Statistics, 30, 676–687.
[4] Andrews, Donald WK. 1993. “Exactly median-unbiased estimation of first order autoregressive/unit root models.” Econometrica, 61, 139–165.

[5] Chan, Ngai H, and Ching-Zong Wei. 1987. “Asymptotic inference for nearly nonstationary AR (1) processes.” The Annals of Statistics, 15, 1050–1063.

[6] Gouriéroux, Christian, Alain Monfort, and Eric Renault. 1993. “Indirect inference.” Journal of Applied Econometrics, 8: S85–S85.

[7] Hannan, Edward J, and Barry G Quinn. 1979. “The determination of the order of an autoregression.” Journal of the Royal Statistical Society, Series B, 41, 190–195.

[8] Leeb, Hannes, and Benedikt M Pötscher. 2005. “Model selection and inference: Facts and fiction.” Econometric Theory, 21(01): 21–59.

[9] Magdalinos, T., 2012. “Mildly explosive autoregression under weak and strong dependence.” Journal of Econometrics, 169(2), pp.179-187.

[10] Ng, Serena, and Pierre Perron. 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(429): 268–281.

[11] Perron, P., 1991. “A continuous time approximation to the unstable first order autoregressive processes: The case without an intercept.” Econometrica, 59, 211–236.

[12] Phillips, Peter C.B. 1987a. “Time series regression with a unit root.” Econometrica, 55, 277–301.

[13] Phillips, Peter C.B. 1987b. “Towards a unified asymptotic theory for autoregression.” Biometrika, 74(3): 535–547.

[14] Phillips, Peter C.B. 2008. “Unit Root Model Selection.” Journal of Japan Statistical Society, 38(1): 65–74.

[15] Phillips, Peter C.B. 2012. “Folklore theorems, implicit maps, and indirect inference.” Econometrica, 80(1): 425–454.

[16] Phillips, Peter C.B., and Ji Hyung Lee. 2015. “Limit theory for VARs with mixed roots near unity.” Econometric Reviews, 34(6-10): 1035–1056.

[17] Phillips, Peter C.B., and Jun Yu. 2011. “Dating the timeline of financial bubbles during the subprime crisis.” Quantitative Economics, 2(3): 455–491.

[18] Phillips, Peter C.B., and Tassos Magdalinos. 2007. “Limit theory for moderate deviations from a unit root.” Journal of Econometrics, 136(1): 115–130.

30
[19] Phillips, Peter C.B., and Tassos Magdalinos. 2009. “Unit root and cointegrating limit theory when initialization is in the infinite past.” Econometric Theory, 25(06): 1682–1715.

[20] Phillips, Peter C.B., Shu-Ping Shi, and Jun Yu. 2015a. “Testing For Multiple Bubbles: Limit Theory Of Real-Time Detectors.” International Economic Review, 56(4): 1079-1134.

[21] Phillips, Peter C.B., Shu-Ping Shi, and Jun Yu. 2015b. “Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500.” International Economic Review, 56(4): 1043-1078..

[22] Phillips, Peter C.B., Yangru Wu, and Jun Yu. 2011. “Explosive Behavior in the 1990s NASDAQ: When did Exuberance Escalate Asset Values?” International Economic Review, 52(1): 201–226.

[23] Ploberger, Werner, and Peter C.B. Phillips. 2003. “Empirical limits for time series econometric models.” Econometrica, 71(2): 627–673.

[24] Pötscher, BM. 1989. “Model selection under nonstationarity: Autoregressive models and stochastic linear regression models.” The Annals of Statistics, 17, 1257–1274.

[25] Schwarz, Gideon. 1978. “Estimating the dimension of a model.” The Annals of Statistics, 6(2): 461–464.

[26] Smith, Anthony A. 1993. “Estimating nonlinear time series models using simulated vector autoregressions.” Journal of Applied Econometrics, 8(S1): S63–S84.

[27] Tang, C.Y., Chen, S.X., 2009. “Parameter estimation and bias correction for diffusion processes.” Journal of Econometrics, 149, 65–81.

[28] Tsay, Ruey S. 1984. “Order selection in nonstationary autoregressive models.” The Annals of Statistics, 12, 1425–1433.

[29] Wang, Xiaohu, Jun Yu, 2016, “Double asymptotics for explosive continuous time models”, Journal of Econometrics, 193, 35–53.