Research Collection School Of Economics   School of Economics

1-2020

Model selection for explosive models

Yubo TAO
Singapore Management University, ybtao@smu.edu.sg

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

Citation

1

MODEL SELECTION FOR EXPLOSIVE MODELS

Yubo Tao[a] and Jun Yu[b]

[a]School of Economics, Singapore Management University, Singapore
*[b]School of Economics and Lee Kong Chian School of Business, Singapore
Management University, Singapore*

ABSTRACT

*This chapter examines the limit properties of information criteria (such as
AIC, BIC, and HQIC) for distinguishing between the unit-root (UR) model
and the various kinds of explosive models. The explosive models include the
local-to-unit-root model from the explosive side the mildly explosive (ME)
model, and the regular explosive model. Initial conditions with different
orders 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 UR model when data come from the UR model. When
data come from the local-to-unit-root model from the explosive side, 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 ME 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 asymptoti-
cally 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 Akaike information criterion (AIC) of Akaike (1969, 1973), Bayesian information criterion (BIC) of Schwarz (1978), and Hannan-Quinn information criterion (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 (UR) behavior. An important form of nonstationarity in time series involves explosive roots (EXR). 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), Phillips, Shi and Yu (2015a, 2015b), Long, Li, and Li (2015).

In this chapter, we study the limit properties of information criteria for distinguishing between the UR model and the explosive models. The information criteria considered in this chapter 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 ordinary least squares (OLS) estimator and the indirect inference estimator (IIE) 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 UR model against the explosive alternatives when data comes from the UR model. Second, we prove that the probability for information criteria to correctly select the explosive model models against the UR model depends crucially on both the degree of explosiveness and the size of the penalty term in information criteria. Finally and surprisingly, 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, although the indirect inference method can reduce the bias.

The rest of this chapter 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 UR 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 chapter, and $\Rightarrow$ denotes weak convergence.

2. MODELS, INFORMATION CRITERIA, AND A LITERATURE REVIEW

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

$$X_t = \rho_n X_{t-1} + u_t, \quad t = 1, \ldots, n, \tag{2.1}$$

where $u_t \overset{iid}{\sim} (0, \sigma^2)$ and the model is initialized at $t = 0$ with some X_0. The autoregressive (AR) coefficient ρ_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 UR. When $\rho_n = 1 - c_n/n = 1 - c/n$ for $c > 0$, X_t is near-stationary the model has a local-to-unit-root (LTU hereafter) from the stationary side (LTUS hereafter) or is near-stationary (Chan & Wei, 1987; Phillips, 1987b). When $\rho_n = \rho$ and $|\rho| > 1$, X_t has an EXR. When $\rho_n = 1 + c_n/n = 1 + c/n$ for $c > 0$, X_t is near-explosive and the model is a local-to-unit-root model from the explosive side (LTUE hereafter). When $\rho_n = 1 - c_n/n$ for $c_n \to \infty$ but $c_n/n \searrow 0$, the root represents moderate deviations from unity and X_t is near-stationary (Phillips & Magdalinos, 2007). When $\rho_n = 1 + c_n/n$ for $c_n \to \infty$ but $c_n/n \searrow 0$, X_t is mildly explosive (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 explosive model under the condition that $u_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ and $X_0 = 0$. The limiting distribution is Cauchy and the rate of convergence is ρ^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 $c_n/n \searrow 0$ and showed that the asymptotic distribution is invariant to the error distribution. The rate of convergence is $n/\sqrt{c_n}$. If $c_n = n^\alpha$ with $\alpha \in (0, 1)$, this rate of convergence bridges that of UR/LTUR 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 $n\rho_n^n/c_n$. The limiting distribution is Cauchy which is the same as in the explosive 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 ρ_n is biased downward when $\rho_n = 1$ or when ρ_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 the model has a UR or LTU and $u_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. 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

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

where k is the number of parameters to be estimated, $\hat{\sigma}_k^2$ is the estimated σ^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). Clearly, for any n, $\log n > 2 \log \log n$; if $n \geq 8$, $\log n > 2$; if $n \geq 16$, $2 \log \log n > 2$. 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 non-stationary models; see, for example, Ng and Perron (1995) and Ploberger and Phillips (2003). 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 UR 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).

In this chapter, we focus our attention to distinguishability between the UR model and the three explosive models (i.e., LTUE, ME, and EX) after the candidate models are estimated by OLS or by the indirect inference method. As a result, we make contributions in two strands of literature: explosive time series and indirect inference. Although we only use IC to do pairwise comparison in this chapter, IC is more widely used to compare multiple models.

To visually understand the difference between the UR model, the LTU model, and the ME model, we simulate a sample path of different length ($n = 100, 200, 500, 1,000$) 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). Figs. 1–3 illustrate the time series plot of UR against LTU, UR against ME1, and UR against ME2, respectively. It can be seen from Fig. 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 Fig. 2 and more so in Fig. 3.

3. LIMIT PROPERTIES BASED ON THE OLS ESTIMATOR

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

$$\lim_{n \to \infty} P\left\{\hat{k}_{IC} = 0 \middle| k = 0\right\}; \tag{3.1}$$

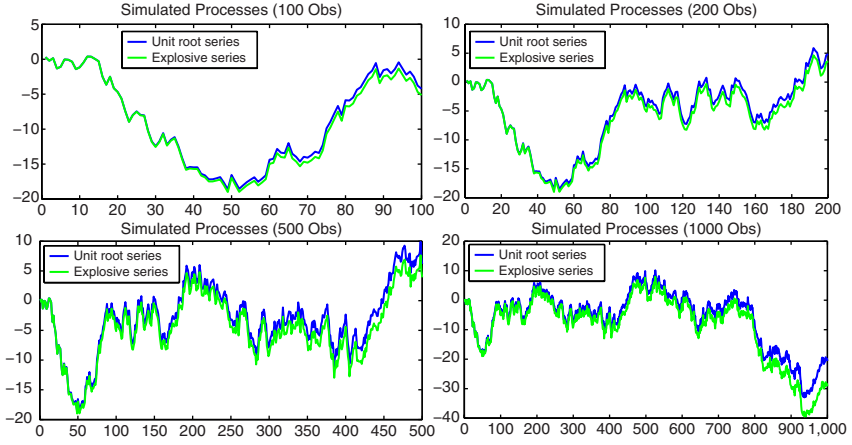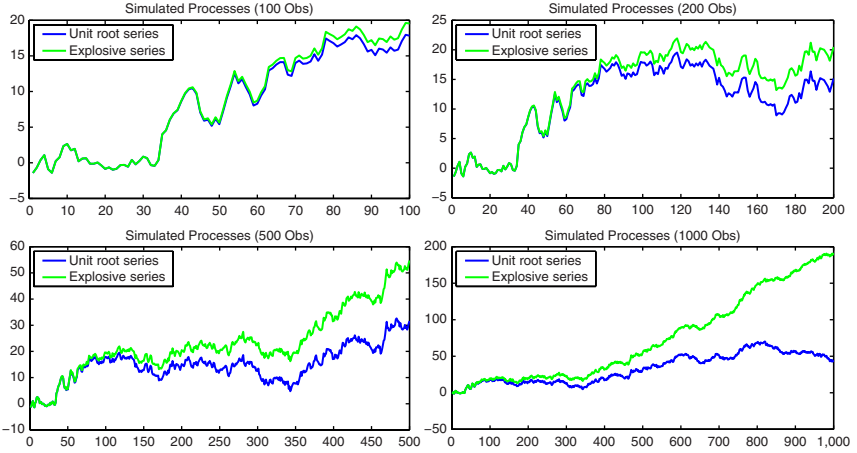*Fig. 1.* A Realization of the UR Model and the LTU Model with $1 + 1/n$.

Fig. 2. A Realization of the UR Model and the ME Process with
$$\rho_n = 1 + n^{0.1}/n \text{ (ME1)}.$$

$$\lim_{n\to\infty} P\left\{\hat{k}_{IC} = 1 \middle| k = 0\right\}; \tag{3.2}$$

$$\lim_{n\to\infty} P\left\{\hat{k}_{IC} = 0 \middle| k = 1\right\}; \tag{3.3}$$

$$\lim_{n\to\infty} P\left\{\hat{k}_{IC} = 1 \middle| k = 1\right\}; \tag{3.4}$$
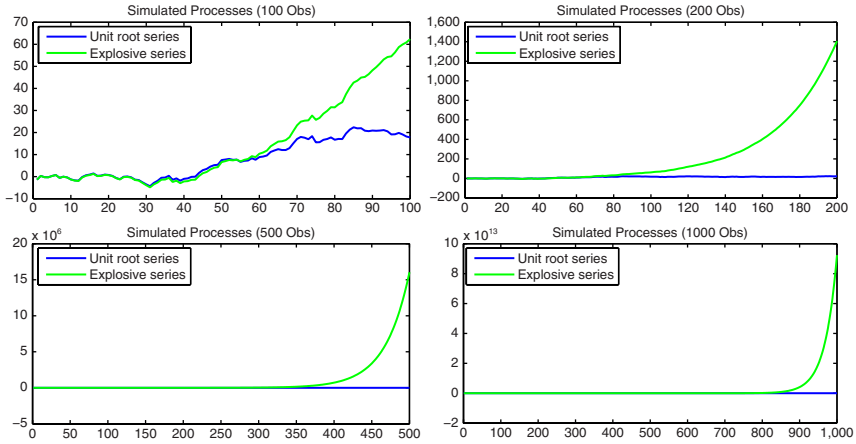
Fig. 3. A Realization of the UR Model and the ME Model with $\rho_n = 1 + n^{0.5}/n$ (ME2).

Given that $\log(n) > 2\log(\log(n)) > 2$ for $n \geq 16$, the probability for BIC to select the UR model cannot be smaller than for HQIC and then for AIC.

As shown in Phillips and Magdalinos (2009), the UR asymptotic distribution is sensitive to initial conditions in the distant past. To understand how the initial condition 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

$$X_0(n) = \sum_{j=0}^{\kappa_n} u_{-j}, \tag{3.5}$$

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

$$\frac{\kappa_n}{n} \rightarrow \tau \in [0, \infty], \text{ as } n \rightarrow \infty. \tag{3.6}$$

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 $p_n \rightarrow \infty$ and $p_n/n \rightarrow 0$ as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} P\left\{\hat{k}_{IC} = 0 \middle| k = 0\right\} = \lim_{n \rightarrow \infty} P\left\{IC_0 - IC_1 \leq 0\right\} = 1,$$

$$\lim_{n \rightarrow \infty} P\left\{\hat{k}_{IC} = 1 \middle| k = 0\right\} = \lim_{n \rightarrow \infty} P\left\{IC_0 - IC_1 > 0\right\} = 0.$$

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

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

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

where

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

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 Phillips (2008, Theorem 1) 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 ξ^2, the squared UR t-statistic for the OLS estimator.

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 $F(1) \ne 0$, and letting u_s have Wold representation

$$u_s = F(L)\varepsilon_s = \sum_{j=0}^{\infty} F_j \varepsilon_{s-j}, \text{ with } \sum_{j=0}^{\infty} j^{1/2} |F_j| < \infty, \tag{3.7}$$

where $\varepsilon_t \overset{iid}{\sim} (0, \sigma_\varepsilon^2)$, the results in *Theorem 3.1* continue to hold. However, both B_0 and ξ^2 need to be modified to accommodate the dependence in u_t as in Phillips (2008).

Theorem 3.4. 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 $n \to \infty$,

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

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

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

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

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

where

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

with

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

Remark 3.5. *Theorem 3.4* shows that all the information criteria are inconsistent in distinguishing between the LTUE model and the UR models when data come 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).

Theorem 3.6. Let *Assumption 1* (i) or (ii) holds. Assume the true DGP is the ME model.

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

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

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

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

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

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

(3) When $\lim\limits_{n\to\infty}\dfrac{p_n}{\rho_n^{2n}}\to+\infty$,

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

$$\lim_{n\to\infty}P\left\{\hat{k}_{IC}=1\big|k=1\right\}=\lim_{n\to\infty}P\left\{\frac{n}{p_n}(IC_1-IC_0)\le0\right\}=0.$$

Remark 3.7. *Theorem 3.6* 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 $\rho_n^n/c_n\to+\infty$. In the special case where $c_n=n^\alpha$, for $\alpha\in(0,1)$, $\lim\limits_{n\to\infty}p_n/\rho_n^{2n}=0$ 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.

Theorem 3.8. Let *Assumption 1* (i) holds. Assume the true DGP is the explosive model.

(1) When $\lim\limits_{n\to\infty}\dfrac{p_n}{\rho^{2n}}=0$,

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

$$\lim_{n\to\infty}P\left\{\hat{k}_{IC}=1\big|k=1\right\}=\lim_{n\to\infty}P\left\{\frac{n}{\rho^{2n}}(IC_1-IC_0)\le0\right\}=1.$$

(2) When $\lim\limits_{n\to\infty}\dfrac{p_n}{\rho^{2n}}=\pi\in(0,+\infty)$,

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

$$\lim_{n\to\infty}P\left\{\hat{k}_{IC}=1\big|k=1\right\}=\lim_{n\to\infty}P\left\{\frac{n}{\rho_n^{2n}}(IC_1-IC_0)\le0\right\}=1-P\left(\chi^2(1)<(1+\rho)^2\pi\right).$$

(3) When $\lim\limits_{n\to\infty}\dfrac{p_n}{\rho^{2n}}\to+\infty$,

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

$$\lim_{n\to\infty}P\left\{\hat{k}_{IC}=1\big|k=1\right\}=\lim_{n\to\infty}P\left\{\frac{n}{p_n}(IC_1-IC_0)\le0\right\}=0.$$

Remark 3.9. *Theorem 3.8* shows that the limit probability of selecting the correct model by information criteria under the explosive model depends also critically on two parameters, ρ, p_n. As expected, the larger ρ, 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 $2 \log \log n$, $\lim_{n \to \infty} p_n / \rho^{2n} = 0$ and hence case (1) applies, suggesting that all the well-known information criteria can consistently select the true model.

Results in *Theorem 3.6* 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 3.10. 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 $\lim_{n \to \infty} \dfrac{p_n}{\rho_n^{2n}} = 0$,

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

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

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

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

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

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

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

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

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

4. LIMIT PROPERTIES BASED ON THE INDIRECT INFERENCE ESTIMATOR

The OLS estimator of ρ_n in Model (2.1) is known to be biased and the bias is acute when ρ_n is close to unity. To reduce the bias, the indirect inference method

of Smith (1993) and Gouriéroux 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 chapter, we denote $\breve{\rho}$ as the indirect inference estimator of ρ. Let $h(c) = c + g(c)$ and $g(c) = g^-(c) 1_{\{c \le 0\}} + g^+(c) 1_{\{c > 0\}}$ with

$$g^-(c) = -\frac{3}{4} \int_0^\infty e^{-\frac{v}{4}} k^-(v;c)^{1/2}\, dv + \frac{1}{4} \int_0^\infty e^{-\frac{v}{4}} k^-(v;c)^{3/2}\, dv$$
$$- \frac{e^{2c}}{8} \int_0^\infty e^{-\frac{5v}{4}} k^-(v;c)^{3/2} v\, dv,$$

$$g^+(c) = \frac{3}{4} \int_0^\infty e^{\frac{w}{4}} k^+(w;c)^{1/2}\, dw - \frac{1}{4} \int_0^\infty e^{\frac{w}{4}} k^+(w;c)^{3/2}\, dw$$
$$- \frac{e^{2c}}{8} \int_0^\infty e^{\frac{5w}{4}} k^+(w;c)^{3/2} w\, dw,$$

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

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

Phillips (2012) shows that under the UR model,

$$n(\breve{\rho} - 1) \Rightarrow h^{-1}\left(\int_0^1 B\, dB \bigg/ \int_0^1 B^2 \right) \text{ as } n \to +\infty,$$

and under the LTUE model,

$$n(\breve{\rho} - \rho_n) \Rightarrow h^{-1}\left(\int_0^1 J_c\, dB \bigg/ \int_0^1 J_c^2 + c \right) - c \text{ as } n \to +\infty.$$

Let $\breve{k}_{IC} = 0$ 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:

$$\lim_{n \to \infty} P\left\{ \breve{k}_{IC} = 0 \big| k = 1 \right\}; \tag{4.1}$$

$$\lim_{n \to \infty} P\left\{ \breve{k}_{IC} = 1 \big| k = 1 \right\}; \tag{4.2}$$

$$\lim_{n \to \infty} P\left\{ \breve{k}_{IC} = 0 \big| k = 0 \right\}; \tag{4.3}$$

$$\lim_{n \to \infty} P\left\{ \breve{k}_{IC} = 1 \big| k = 0 \right\}. \tag{4.4}$$

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 $n \to \infty$,

$$\lim_{n \to \infty} P\left\{ \breve{k}_{IC} = 0 \big| k = 0 \right\} = \lim_{n \to \infty} P\left\{ (IC_0 - IC_1) \le 0 \right\} = 1,$$

$$\lim_{n \to \infty} P\left\{ \breve{k}_{IC} = 1 \big| k = 0 \right\} = \lim_{n \to \infty} P\left\{ (IC_0 - IC_1) > 0 \right\} = 0;$$

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

$$\lim_{n\to\infty} P\left\{\breve{k}_{AIC} = 0 \middle| k = 0\right\} = P\left(\varsigma^2 < 2\right),$$

$$\lim_{n\to\infty} P\left\{\breve{k}_{AIC} = 1 \middle| k = 0\right\} = 1 - P\left(\varsigma^2 < 2\right),$$

where

$$\varsigma^2 = \begin{cases} \int_0^1 B^2 \cdot h^{-1}\left(\left(\dfrac{\int_0^1 B\,dB}{\int_0^1 B^2}\right)^2\right) - 2\int_0^1 B\,dB \cdot h^{-1}\left(\dfrac{\int_0^1 B\,dB}{\int_0^1 B^2}\right), & \text{if } \tau = 0 \\[3em] \int_0^1 B_\tau^2 \cdot h^{-1}\left(\left(\dfrac{\int_0^1 B_\tau\,dB}{\int_0^1 B_\tau^2}\right)^2\right) - 2\int_0^1 B_\tau\,dB \cdot h^{-1}\left(\dfrac{\int_0^1 B_\tau\,dB}{\int_0^1 B_\tau^2}\right), & \text{if } \tau \in (0,\infty), \\[3em] h^{-1}(\mathcal{C})^2\,B_0^2(1) - 2h^{-1}(\mathcal{C})\,B(1)\,B_0(1), & \text{if } \tau = \infty \end{cases}$$

with $\mathcal{C}$ being a standard Cauchy variate.

Remark 4.2. According to *Theorem 4.1*, as long as $p_n \to \infty$ and $p_n/n \to 0$, 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 ς^2, the squared UR t-statistic for the indirect inference estimator.

Remark 4.3. As shown in Phillips (2012), the squared UR t-statistic for the indirect inference estimator has a smaller variance than that of the squared UR t-statistic for the OLS estimator. Consequently, $P(\varsigma^2 < 2) > P(\xi^2 < 2)$, 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. 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 $n \to \infty$,

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

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

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

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

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

where

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

Remark 4.5. *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 $1 - P(\vartheta^2 > 2)$. Since the variance of ζ^2 is bigger than that of υ^2, the tail probability of ζ^2 is larger than that of υ^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. This is a rather surprising result and suggests that the superiority in estimation does not necessarily translate to the superiority in model selection.

Theorem 4.6. Let *Assumption 1* (i) or (ii) holds. Assume the true DGP is the ME model.

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

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

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

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

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

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

(3) When $\lim\limits_{n\to\infty}\dfrac{p_n}{\rho_n^{2n}}\to+\infty$,

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

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

Remark 4.7. The results in *Theorem 4.6* are the same as those in *Theorem 3.6*, suggesting all the well-known information criteria can consistently select the true model (i.e., ME model) when $c_n=n^\alpha$, for $\alpha\in(0,1)$.

Theorem 4.8. Let *Assumption 1* (i) holds. Assume the true DGP is the explosive model.

(1) When $\lim\limits_{n\to\infty}\dfrac{p_n}{\rho^{2n}}=0$,

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC}=0\big|k=1\right\}=\lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}(IC_1-IC_0)>0\right\}=0,$$

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC}=1\big|k=1\right\}=\lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}(IC_1-IC_0)\leq 0\right\}=1.$$

(2) When $\lim\limits_{n\to\infty}\dfrac{p_n}{\rho^{2n}}=\pi\in(0,+\infty)$,

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

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

(3) When $\lim\limits_{n\to\infty}\dfrac{p_n}{\rho^{2n}}\to+\infty$,

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

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

Remark 4.9. The results in *Theorem 4.8* are the same as those in *Theorem 3.8*, suggesting that all the well-known information criteria can consistently select the true model (i.e., the explosive 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 Sections 3 and 4. In the simulation study, we use both OLS and the indirect inference method to estimate ρ_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 explosive 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, $n = 100, 200, 500, 1,000$. 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 $2 < 2 \log \log n < \log n$ 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/ρ_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. This observation is consistent with the asymptotic results 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 *Theorem 4.4* and *Remark 4.5*. Finally, it seems that AIC performs better than BIC and HQIC in all cases.

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/ρ_n^{2n}. Several results can be found here. First,

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			1,000	
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

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/ρ_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			1,000		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_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. 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 $c_n = n^{0.1}$						
n	100			200		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	0.0861	0.1983	0.1316	0.0679	0.1799	0.1132
OLS	0.5183	0.3403	0.4349	0.5554	0.3638	0.4629
IIE	0.3071	0.1741	0.2406	0.3211	0.1624	0.2250
n	500			1,000		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_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	100			200		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	0.0008	0.0019	0.0012	0.0001	0.0003	0.0002
OLS	0.9374	0.9066	0.9235	0.9749	0.9608	0.9683
IIE	0.9274	0.8979	0.9163	0.9716	0.9578	0.9648
n	500			1,000		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_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

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 *Theorem 3.6* and *Remark 4.7*. 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 4 report the results when the true DGP is the explosive model with $\rho = 1.01$, 1.05. Also reported is the value of p_n/ρ_n^2. 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.

	Explosive Model with $\rho = 1.01$					
n		100			200	
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	0.2734	0.6295	0.4175	0.0374	0.0990	0.0623
OLS	0.3516	0.1475	0.2420	0.6449	0.4820	0.5555
IIE	0.1485	0.0445	0.0922	0.4740	0.3059	0.3845
n		500			1,000	
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	1.0e-4	1.0e-4	1.0e-4	1.0e-9	1.0e-8	1.0e-9
OLS	0.9775	0.9599	0.9704	0.9998	0.9997	0.9998
IIE	0.9733	0.9563	0.9681	0.9998	0.9997	0.9998
	Explosive Model with $\rho = 1.05$					
n		100			200	
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	0.0001	0.0003	0.0002	1.0e-07	1.0e-07	1.0e-07
OLS	0.9741	0.9643	0.9681	0.9999	0.9998	0.9998
IIE	0.9703	0.9626	0.9655	0.9999	0.9998	0.9998
n		500			1,000	
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	1.0e-20	1.0e-20	1.0e-20	1.0e-41	1.0e-41	1.0e-41
OLS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
IIE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

6. CONCLUSION

This chapter studies the limit properties of information criteria for distinguishing between the UR 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 chapter contributes to the literature in three aspects. First, our results extends results in the literature to the explosive side of the UR, and we find that information criteria consistently choose the UR model when the UR 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 AR 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 UR, 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 explosive model, the indirect inference estimation decreases the probability. This rather surprising result suggests that the superiority in estimation does not necessarily translate to the superiority in model selection.

ACKNOWLEDGMENTS

REFERENCES

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, *21*(1), 243–247.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory*, Tsahkadsor, Armenia, USSR. (pp. 267–281).

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics*, *30*, 676–687.

Chan, N. H, & Wei, C.-Z. (1987). Asymptotic inference for nearly nonstationary AR (1) processes. *The Annals of Statistics*, *15*, 1050–1063.

Gouriéroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, *8*, S85–S85.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B, 41*, 190–195.

Long, W., Li, D., & Li, Q. (2015). Testing explosive behavior in the gold market. *Empirical Economics*, *51*(3), 1151–1164.

Magdalinos, T. (2012). Mildly explosive autoregression under weak and strong dependence. *Journal of Econometrics*, *169*(2), 179–187.

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*(429), 268–281.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, *55*, 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, *74*(3), 535–547.

Phillips, P. C. B. (2008). Unit Root Model Selection. *Journal of Japan Statistical Society*, *38*(1): 65–74.

Phillips, P. C. B. (2012). Folklore theorems, implicit maps, and indirect inference. *Econometrica*, *80*(1), 425–454.

Phillips, P. C. B., & Lee, J. H. (2015). Limit theory for VARs with mixed roots near unity. *Econometric Reviews*, *34*(6–10), 1035–1056.

Phillips, P. C. B., & Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, *136*(1), 115–130.

Phillips, P. C. B., & Magdalinos, T. (2009). Unit root and cointegrating limit theory when initialization is in the infinite past. *Econometric Theory*, *25*(06), 1682–1715.

Phillips, P. C. B., Shi, S.-P., & Yu, J. (2015a). Testing For multiple bubbles: Limit theory of real-time detectors. *International Economic Review*, *56*(4), 1079–1134.

Phillips, P. C. B., Shi, S.-P., & Yu, J. (2015b). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, *56*(4), 1043–1078.

Phillips, P. C. B., Wu, Y., & Yu, J. (2011). Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review*, *52*(1), 201–226.

Phillips, P. C. B., & Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics*, *2*(3), 455–491.

Ploberger, W., & Phillips, P. C. B. (2003). Empirical limits for time series econometric models. *Econometrica*, *71*(2), 627–673.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Smith, A. A. (1993). Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics*, *8*(S1), S63–S84.

APPENDIX

A. Proof of Theorem 3.1

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

B. Proof of Theorem 3.4

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

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

By *Lemma 1* in Phillips (1987b), when the process is initialized at X_0, we know

$$
\frac{1}{n^2} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 J_c^2, \tag{A.1}
$$

and

$$
\frac{1}{n} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \sigma^2 \int_0^1 J_c \, dB, \tag{A.2}
$$

where

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

Therefore, by Equations (A.1) and (A.2), we have

$$
\begin{aligned}
IC_0 &= \log \left\{ \frac{\sigma^2 c^2}{n} \int_0^1 J_c^2 + \frac{2c\sigma^2}{n} \int_0^1 J_c \, dB + \sigma^2 + o_p\left(n^{-1}\right) \right\} \\
&= \log \sigma^2 + \log \left\{ 1 + \frac{2c}{n} \int_0^1 J_c(r) \, dB + \frac{c^2}{n} \int_0^1 J_c^2 + o_p\left(n^{-1}\right) \right\}.
\end{aligned} \tag{A.3}
$$

We also know from Phillips (1987b) that

$$
n\left(\hat{\rho}_n - \rho_n\right) \Rightarrow \frac{\int_0^1 J_c \, dB}{\int_0^1 J_c^2}, \tag{A.4}
$$

Hence,

$$
IC_1 = \log \hat{\sigma}_1^2 + \frac{p_n}{n}
$$

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

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

$$
= \log \left\{ \frac{1}{n} \left(\rho_n - \hat{\rho}_n \right)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} \left(\rho_n - \hat{\rho}_n \right) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n} \qquad \text{(A.5)}
$$

$$
\Rightarrow \log \left\{ -\frac{\sigma^2}{n} \frac{\left(\int_0^1 J_c \, dB \right)^2}{\int_0^1 J_c^2} + \sigma^2 \right\} + \frac{p_n}{n}
$$

$$
= \log \sigma^2 + \log \left\{ 1 - \frac{1}{n} \frac{\left(\int_0^1 J_c \, dB \right)^2}{\int_0^1 J_c^2} \right\} + \frac{p_n}{n}.
$$

Therefore, by Equations (A.3) and (A.5), we have

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

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

$$
n \left(IC_1 - IC_0 \right) \Rightarrow 2 - \frac{\left(\int_0^1 J_c \, dB \right)^2}{\int_0^1 J_c^2} - 2c \int_0^1 J_c \, dB - c^2 \int_0^1 J_c^2. \qquad \text{(A.6)}
$$

If $p_n \to \infty$ and $\dfrac{p_n}{n} \to 0$, we have

$$
\frac{n}{p_n} \left(IC_1 - IC_0 \right) \Rightarrow 1.
$$

C. Proof of Theorem 3.6

When the true DGP is the ME model, we have

$$
\begin{aligned}
IC_0 = \log \hat{\sigma}_0^2 &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - X_{t-1})^2 \right\} \\
&= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - 1) X_{t-1} + u_t]^2 \right\} \\
&= \log \left\{ \frac{1}{n} (\rho_n - 1)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - 1) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}.
\end{aligned}
\tag{A.7}
$$

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

$$
\frac{c_n^2 \rho_n^{-2n}}{n^2} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \frac{\sigma^2}{4} Y^2,
\tag{A.8}
$$

$$
\frac{c_n \rho_n^{-n}}{n^2} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \frac{\sigma^2}{2} XY,
\tag{A.9}
$$

and

$$
\frac{n \rho_n^n}{2 c_n} (\hat{\rho}_n - \rho_n) \Rightarrow \frac{X}{Y} \sim \mathcal{C},
\tag{A.10}
$$

where $X, Y \sim \mathcal{N}(0, 1)$ and ζ is a standard Cauchy variate.

Therefore, by Equations (A.8) and (A.9), we have

$$
\begin{aligned}
IC_0 &\Rightarrow \log \left\{ \frac{\sigma^2}{4n \rho_n^{-2n}} Y^2 + \frac{\sigma^2}{n \rho_n^{-n}} XY + \sigma^2 \right\} \\
&= \log \sigma^2 + \log \left\{ \frac{1}{4n \rho_n^{-2n}} Y^2 + \frac{1}{n \rho_n^{-n}} XY + 1 \right\}.
\end{aligned}
\tag{A.11}
$$

On the other hand,

$$
\begin{aligned}
IC_1 &= \log \hat{\sigma}_1^2 + \frac{p_n}{n} \\
&= \log \left\{ n^{-1} \sum_{t=1}^{n} (X_t - \hat{\rho}_n X_{t-1})^2 \right\} + \frac{p_n}{n} \\
&= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho_n - \hat{\rho}_n) X_{t-1} + u_t]^2 \right\} + \frac{p_n}{n} \\
&= \log \left\{ \frac{1}{n} (\rho_n - \hat{\rho}_n)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - \hat{\rho}_n) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}.
\end{aligned}
$$

By Equations (A.8)–(A.10), we obtain

$$IC_1 \Rightarrow \log\left\{ \frac{1}{n} \frac{4c_n^2}{n^2 \rho_n^{2n}} \mathcal{C}^2 \frac{n^2 \sigma^2}{4c_n^2 \rho_n^{-2n}} Y^2 - \frac{2}{n} \frac{2c_n}{n\rho_n^n} \mathcal{C} \frac{n\sigma^2}{2c_n \rho_n^{-n}} XY + \sigma^2 \right\} + \frac{p_n}{n}$$

$$= \log \sigma^2 + \log\left\{ -\frac{1}{n} X^2 + 1 \right\} + \frac{p_n}{n}. \tag{A.12}$$

Therefore, by Equations (A.11) and (A.12), we have

$$IC_1 - IC_0 \Rightarrow \log\left\{ 1 - \frac{1}{n} X^2 \right\} - \log\left\{ 1 + \frac{1}{4n\rho_n^{-2n}} Y^2 + \frac{1}{n\rho_n^{-n}} XY \right\} + \frac{p_n}{n}.$$

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

$$\frac{n}{k_n}(IC_1 - IC_0) \Rightarrow \begin{cases} \pi - \frac{1}{4}\chi^2(1), & \text{if } \pi \in [0, \infty) \\ 1, & \text{if } \pi = \infty \end{cases},$$

where

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

D. Proof of Theorem 3.8

When the true DGP is the explosive model, we have

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

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

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

By results established in Anderson (1959), we know

$$\frac{1}{\rho^{2n}} \sum_{t=1}^n X_{t-1}^2 \Rightarrow \frac{\sigma^2 Y^2}{(\rho^2 - 1)^2}, \tag{A.13}$$

$$\frac{1}{\rho^n} \sum_{t=1}^n X_{t-1}u_t \Rightarrow \frac{\sigma^2 XY}{\rho^2 - 1}, \tag{A.14}$$

$$\frac{\rho^n}{\rho^2 - 1}(\hat{\rho} - \rho) \Rightarrow \mathcal{C}, \tag{A.15}$$

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

$$
\begin{aligned}
IC_0 &\Rightarrow \log \left\{ \frac{\sigma^2 \rho^{2n}}{n(\rho+1)^2} X^2 + \frac{2\sigma^2 \rho^n}{n(\rho+1)} XY + \sigma^2 \right\} \\
&= \log \sigma^2 + \log \left\{ \frac{\rho^{2n}}{n(\rho+1)^2} X^2 + \frac{2\rho^n}{n(\rho+1)} XY + 1 \right\}.
\end{aligned} \tag{A.16}
$$

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

$$
\begin{aligned}
IC_1 &= \log \hat{\sigma}_1^2 + \frac{p_n}{n} \\
&= \log \left\{ n^{-1} \sum_{t=1}^{n} (X_t - \hat{\rho} X_{t-1})^2 \right\} + \frac{p_n}{n} \\
&= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} [(\rho - \hat{\rho}) X_{t-1} + u_t]^2 \right\} + \frac{p_n}{n} \\
&= \log \left\{ \frac{1}{n} (\rho - \hat{\rho})^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho - \hat{\rho}) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}.
\end{aligned}
$$

By Equation (A.13)–(A.15), we have

$$IC_1 = \log \sigma^2 + \log \left\{ 1 - \frac{1}{n} X^2 \right\} + \frac{p_n}{n}. \tag{A.17}$$

Now, by Equations (A.16) and (A.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 $\lim_{n \to \infty} \frac{p_n}{\rho^{2n}} = \pi$, we have

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

where

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

E. Proof of Proposition 3.10

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

$$IC_0 = \log \hat\sigma_0^2 = \log\left\{\frac{1}{n}\sum_{t=1}^n (X_t - X_{t-1})^2\right\}$$

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

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

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

$$\frac{c_n^2 \rho_n^{-2n}}{\omega^2 n^2}\sum_{t=1}^n X_{t-1}^2 \Rightarrow \frac{\sigma^2}{4}Z^2, \tag{A.18}$$

and

$$\frac{c_n \rho_n^{-n}}{\omega^2 n}\sum_{t=1}^n X_{t-1}u_t \Rightarrow \frac{\sigma^2}{2}YZ, \tag{A.19}$$

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 Equations (A.18) and (A.19), we have

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

$$= \log \sigma^2 + \log\left\{1 + \frac{\omega^2}{n\rho_n^{-n}}YZ + \frac{\omega^2}{4n\rho_n^{-2n}}Z^2 + o_p\left(n^{-1}\right)\right\}. \tag{A.20}$$

We also know from Magdalinos (2012) that

$$\frac{n\rho_n^n}{2c_n}\left(\hat\rho_n - \rho_n\right) \Rightarrow \mathcal{C}. \tag{A.21}$$

Hence,

$$IC_1 = \log \hat\sigma_1^2 + \frac{p_n}{n}$$

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

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

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

$$\Rightarrow \log\left\{-\frac{\omega^2\sigma^2}{n}Y^2 + \sigma^2\right\} + \frac{p_n}{n} \tag{A.22}$$

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

Therefore, by Equations (A.20) and (A.22), we have

$$IC_1 - IC_0 \Rightarrow \log\left\{1 - \frac{\omega^2}{n}Y^2\right\} - \log\left\{1 + \frac{\omega^2}{n\rho_n^{-n}}YZ + \frac{\omega^2}{4n\rho_n^{-2n}}Z^2\right\} + \frac{p_n}{n}.$$

Note $Y^2, Z^2 \sim \chi^2(1)$ and $\rho_n^{-n} = o\left(c_n^{-1}\right)$. If $\lim\limits_{n\to\infty}\dfrac{p_n}{\rho_n^{2n}} = \pi$,

$$\frac{n}{k_n}(IC_1 - IC_0) \Rightarrow \begin{cases} \pi - \dfrac{\omega^2}{4}\chi^2(1), & \text{if } \pi \in [0, \infty) \\ 1, & \text{if } \pi = \infty \end{cases},$$

where

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

F. Proof of Theorem 4.1

When the true DGP is the UR model, we have

$$IC_0 = \log\breve{\sigma}_0^2 = \log\left\{\frac{1}{n}\sum_{t=1}^{n}u_t^2\right\} = \log\sigma^2.$$

Also, we have

$$IC_1 = \log\breve{\sigma}_1^2 + \frac{p_n}{n} = \log\left\{\frac{1}{n}\sum_{t=1}^{n}(X_t - \breve{\rho}X_{t-1})^2\right\} + \frac{p_n}{n}$$

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

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

According to Phillips ([2012](#)), we have

$$\breve{\rho}-1 \Rightarrow \begin{cases} \dfrac{1}{n}h^{-1}\left(\dfrac{\int_0^1 B\,dB}{\int_0^1 B^2}\right), & \text{if } \tau = 0 \\[4mm] \dfrac{1}{n}h^{-1}\left(\dfrac{\int_0^1 B_\tau\,dB}{\int_0^1 B_\tau^2}\right), & \text{if } \tau \in (0,\infty), \\[4mm] \dfrac{1}{\sqrt{n^2/c_n}}h^{-1}(\mathcal{C}), & \text{if } \tau = \infty \end{cases}$$

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

According to Phillips and Magdalinos ([2009](#)), we have

$$\frac{1}{n^2}\sum_{t=1}^n X_{t-1}^2 \Rightarrow \begin{cases} \sigma^2\int_0^1 B^2, & \text{if } \tau = 0 \\[3mm] \sigma^2\int_0^1 B_\tau^2, & \text{if } \tau \in (0,\infty), \\[3mm] \sigma^2 B_0(1)^2/c_n, & \text{if } \tau = \infty \end{cases}$$

$$\frac{1}{n}\sum_{t=1}^n X_{t-1}u_t \Rightarrow \begin{cases} \sigma^2\int_0^1 B\,dB, & \text{if } \tau = 0 \\[3mm] \sigma^2\int_0^1 B_\tau\,dB, & \text{if } \tau \in (0,\infty). \\[3mm] \sqrt{1/c_n}\,\sigma^2 B(1)\,B_0(1), & \text{if } \tau = \infty \end{cases}$$

Therefore, we have

$$IC_0 - IC_1 \Rightarrow \begin{cases} -\log\left\{\dfrac{\int_0^1 B^2}{n}h^{-1}\left(\dfrac{\int_0^1 B\,dB}{\int_0^1 B^2}\right)^2 - \dfrac{2\left(\int_0^1 B\,dB\right)}{n}h^{-1}\left(\dfrac{\int_0^1 B\,dB}{\int_0^1 B^2}\right)+1\right\} - \dfrac{p_n}{n} \\[6mm] -\log\left\{\dfrac{\int_0^1 B_\tau^2}{n}h^{-1}\left(\dfrac{\int_0^1 B_\tau\,dB}{\int_0^1 B_\tau^2}\right)^2 - \dfrac{2\left(\int_0^1 B_\tau\,dB\right)}{n}h^{-1}\left(\dfrac{\int_0^1 B_\tau\,dB}{\int_0^1 B_\tau^2}\right)+1\right\} - \dfrac{p_n}{n}. \\[6mm] -\log\left\{\dfrac{1}{n}h^{-1}(\mathcal{C})^2 B_0(1)^2 - \dfrac{2}{n}h^{-1}(\mathcal{C})B(1)B_0(1)+1\right\} - \dfrac{p_n}{n} \end{cases}$$

G. Proof of Theorem 4.4

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

$$
IC_1 = \log \breve{\sigma}_1^2 + \frac{p_n}{n}
$$

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

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

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

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

$$
n(\breve{\rho}_n - \rho_n) \Rightarrow h^{-1} \left(\frac{\int_0^1 J_c \, dB}{\int_0^1 J_c^2} + c \right) - c. \tag{A.23}
$$

By Equations (A.1), (A.2), and (A.23), we have

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

$$
\left. + \frac{1}{n} \left[h^{-1} \left(\frac{\int_0^1 J_c \, dB}{\int_0^1 J_c^2} + c \right) - c \right]^2 \int_0^1 J_c^2 \right\} + \log \sigma^2 + \frac{p_n}{n}. \tag{A.24}
$$

Therefore, by Equations (A.11) and (A.23), we have

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

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

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

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

where

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

When $p_n \to \infty$ and $\frac{p_n}{n} \to 0$, we have n

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

H. Proof of Theorem 4.6

When the true DGP is the ME model, we have

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

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

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

By Equations (A.8) and (A.9), we have

$$IC_0 \Rightarrow \log\left\{\frac{\sigma^2}{4n\rho_n^{-2n}}Y^2 + \frac{\sigma^2}{n\rho_n^{-n}}XY + \sigma^2\right\}$$

$$= \log\sigma^2 + \log\left\{\frac{1}{4n\rho_n^{-2n}}Y^2 + \frac{1}{n\rho_n^{-n}}XY + 1\right\}. \tag{A.25}$$

Similarly, for IC_1 based on the indirect inference estimator, we have

$$IC_1 = \log\breve{\sigma}_1^2 + \frac{p_n}{n}$$

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

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

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

Using the results in Phillips (2012), Equations (A.8) and (A.9), we obtain

$$IC_1 \Rightarrow \log\left\{\frac{1}{n}\frac{4c_n^2}{n^2\rho_n^{2n}}\left(\mathcal{C} + O\left(\frac{1}{2c_n}\right)\right)^2\frac{n^2\sigma^2}{4c_n^2\rho_n^{-2n}}Y^2\right.$$

$$\left. -\frac{2}{n}\frac{2c_n}{n\rho_n^n}\left(\mathcal{C} + O\left(\frac{1}{2c_n}\right)\right)\frac{n\sigma^2}{2c_n\rho_n^{-n}}XY + \sigma^2\right\} + \frac{p_n}{n} \tag{A.26}$$

$$\Rightarrow \log\sigma^2 + \log\left\{1 - \frac{1}{n}X^2 + O\left(\frac{1}{c_n n}\right)\right\} + \frac{p_n}{n}.$$

Therefore, the similar results to those in *Theorem 3.6* are obtained.

I. Proof of Theorem 4.8

When the true DGP is the explosive model, for the indirect inference estimator, we know that for IC_0, it is the same as the OLS estimator. Therefore, we only need to derive the IC_1. Note that for IC_1, we have

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

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

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

Therefore, we obtain

$$IC_1 = \log\sigma^2 + \log\left\{\frac{1}{n}\left(\mathcal{C} + O\left(\frac{1}{\rho^2 - 1}\right)\right)^2Y^2 - \frac{2}{n}\left(\mathcal{C} + O\left(\frac{1}{\rho^2 - 1}\right)\right)XY + 1\right\} + \frac{p_n}{n} \tag{A.27}$$

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

Now, by Equations (A.16) and (A.27), we obtain

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

$$+ \frac{p_n}{n}.$$

Since $\lim\limits_{n\to\infty} \dfrac{p_n}{\rho^{2n}} = \pi$, we have

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

where

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