Nearly Unstable Processes: A Prediction Perspective

Ngai Hang Chan
The Chinese University of Hong Kong

Ching-Kang Ing
National Tsing Hua University

Rongmao Zhang
Zhejiang University

Abstract

Prediction has long been a vibrant topic in modern probability and statistics. In addition to finding optimal forecast and model selection, it is argued in this paper that the prediction principle can also be used to analyze critical phenomena, in particular, stationary and unstable time series. Although the notion of nearly unstable models has become one of the important concepts in time series econometrics, its role from a prediction perspective is less developed. Based on moment bounds for the extreme-value (EV) and least squares (LS) estimates, asymptotic expressions for the mean squared prediction errors (MSPE) of the EV and LS predictors are obtained for a nearly unstable first-order autoregressive (AR(1)) model with positive error. These asymptotic expressions are further extended to a general class of nearly unstable models, thereby allowing one to understand to what degree such general models can be used to establish a link between stationary and unstable models from a prediction perspective. As applications, we illustrate the usefulness of these results in conducting finite sample approximations of the MSPE for near unit-root time series.

Keywords: Extreme-value predictor, least squares predictor, mean squared prediction error, nearly unstable process, positive error.

1 Introduction

Prediction has long been a vibrant topic in modern probability and statistics. The seminal monograph of Whittle (1963) illustrates the importance of linear prediction. There are several objectives in prediction studies. The first one is computing an optimal forecast, based on either finite or infinite samples. The second one is to use prediction methods for model selection. The third one, which is less well known but of no less importance, is to use the prediction principle to understand critical phenomena, in particular, stationary and unstable processes, see for example Wei (1992). This goal also constitutes the main focus of the present paper.

To achieve this goal, moment bounds become indispensable tools. For example, based on maximal moment inequalities for martingales, Wei (1987, 1992) provided an asymptotic expression for the accumulated prediction error (APE) of a linear stochastic regression model, which in turn leads to the Fisher information criterion (FIC) for model selection. Findley and Wei (2002) and Chan and Ing (2011) established inverse moment bounds for the Fisher information matrices of time series models. These useful bounds enable one not only to calculate the mean squared prediction errors (MSPE) of the least squares predictors, but also to derive the Akaike information criterion (AIC; Akaike (1974)) and the final prediction error criterion (FPE; Akaike (1969)) in a rigorous manner.

Studies of moment bounds and MSPE have mostly been focused around the least squares procedures, much less have been conducted on the so-called extreme-value estimates (EVE), which is mainly used for heavy-tailed dependent data. Due to the emergence of big data, dependent heavy-tailed phenomena have been reported in various disciplines, see for example, the exemplary monograph by Finkenstädt and Rootzén (2004) and the examples therein. To appreciate the significance of such types of estimates, suppose that the data are generated from the first-order autoregressive (AR(1)) model:

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, n, \tag{1.1}$$

where $0 \le \rho < 1$ and ε_t's are i.i.d. positive noise with regularly varying density $f_\varepsilon(x)$ at zero, i.e.,

$$\lim_{x \to 0} \frac{f_\varepsilon(x)}{cx^{\alpha-1}} = 1, \text{ for some unknown } \alpha > 0 \text{ and } c > 0. \tag{1.2}$$

One of the most popular methods for estimating ρ in (1.1) is the least squares estimator (LSE),

$$\tilde{\rho}_n = \sum_{i=2}^{n} (y_{i-1} - \bar{y})(y_i - \bar{y}) / \sum_{i=2}^{n} (y_{i-1} - \bar{y})^2, \tag{1.3}$$

where $\bar{y} = \bar{y}_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n-1} y_i$. However, when the noise has a density like (1.2), LSE may not be efficient and other estimation procedures are required. When the parametric form of the distribution of ε_t is known, a natural alternative to $\tilde{\rho}_n$ is the maximum likelihood estimator (MLE), yet as argued in Davis and McCormick (1989) and Ing and Yang (2014), the MLE is in general analytically difficult to work with. A remedy for this difficulty is to use the EVE, $\hat{\rho}_n$, instead, where

$$\hat{\rho}_n = \min_{1 \le i \le n-1} y_{i+1}/y_i. \tag{1.4}$$

Note that $\hat{\rho}_n$ is also the MLE when ε_t has an exponential distribution or is uniformly distributed over $[0, a]$ for some $a > 0$; see Bell and Smith (1986). Under an assumption more general than (1.2), Bell and Smith (1986) showed that $\hat{\rho}_n$ is consistent. When (1.2) holds, it is shown in Corollaries 2.4 and 2.5 of Davis and McCormick (1989) that the limit distributions of $\hat{\rho}_n$ satisfies

$$\lim_{n \to \infty} P\{(cM_\alpha(\rho)/\alpha)^{1/\alpha} n^{1/\alpha} (\hat{\rho}_n - \rho) > t\} = \exp\{-t^\alpha\}, \tag{1.5}$$

where $M_\alpha(\rho) = \mathrm{E}(\sum_{j=0}^{\infty} \rho^j \varepsilon_{1-j})^\alpha$. Equation (1.5) reveals that when $\alpha < 2$ ($\alpha > 2$), the convergence rate of $\hat{\rho}_n$ ($\tilde{\rho}_n$) is faster than that of $\tilde{\rho}_n$ ($\hat{\rho}_n$); see Section 2 of Ing and Yang (2014) for a more comprehensive comparison of $\hat{\rho}_n$ and $\tilde{\rho}_n$.

Model (1.1) with ε_t satisfying (1.2) has found broad applications in hydrology, economics, finance, epidemiology and quality control; see, among others, Gaver and Lewis (1980), Bell and Smith (1986), Lawrance and Lewis (1985), Davis and McCormick (1989), Smith (1994), Barndorff-Nielsen and Shephard (2001), Nielsen and Shephard (2003), Sarlak (2008) and Ing and Yang (2014). In particular, Bell and

Smith (1986) analyzed two sets of pollution data from the Willamette River, Oregon, using model (1.1) with ε_t following the uniform distribution or exponential distribution; both are special cases of (1.2). In addition, Sarlak (2008) adopted model (1.1) with a Weibull error to analyze the annual streamflow data from the Kizilirmak River in Turkey. On the other hand, model (1.1), focusing exclusively on the stationary case $0 \leq \rho < 1$, fails to accommodate data that may fluctuate around an upward trend with variance increasing over time. Ing and Yang (2014) therefore generalized (1.1) to $\rho = 1$, which is referred to as the unit-root model, and established the limit distribution of (1.4) in this case. In addition, they derived asymptotic expressions for the mean squared prediction errors (MSPE) of the EV predictor ($\hat{y}_{n+1}$) and the LS predictor ($\tilde{y}_{n+1}$), i.e. $\text{MSPE}_A = \text{E}(y_{n+1} - \hat{y}_{n+1})^2$ and $\text{MSPE}_B = \text{E}(y_{n+1} - \tilde{y}_{n+1})^2$, as

$$\lim_{n\to\infty} n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2) = R_A^o(\alpha, \rho), \ \lim_{n\to\infty} n(\text{MSPE}_B - \sigma^2) = R_B^o(\rho), \tag{1.6}$$

where $0 \leq \rho \leq 1$, $R_A^o(\alpha, \rho)$ is a positive constant depending on α, ρ and $f_{\varepsilon_1}(\cdot)$ and $R_B^o(\rho)$ is a positive constant depending on ρ and $\sigma^2 = \text{Var}(\varepsilon_1) > 0$; see (37), (42), (47) and (49) of Ing and Yang (2014).

While (1.6) suggests that $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ and $n(\text{MSPE}_B - \sigma^2)$ can be approximated by $R_A^o(\alpha, \rho)$ and $R_B^o(\rho)$, such an approximation may become unsatisfactory when ρ is near one; see Tables 1–6 of Section 3.2. This phenomenon is a reminiscence of the nearly unstable autoregressive model that was discussed in Chan and Wei (1987). By virtue of the order of the observed Fisher's information number, they argued that neither the stationary normal limit nor the unit-root limit distributions would be a good approximation to the finite sample behavior of the LSE for the situation where ρ is close to 1. Putting it differently, a main difficulty in using (1.6) when ρ is close to 1 may be due to the critical behaviors of the limit distributions associated with the EVE and LSE. Such critical behaviors perpetuate in the performance of the corresponding predictors.

To circumvent this difficulty, consider the following family of nearly unstable models:

$$y_t = \rho_n y_{t-1} + \varepsilon_t, \tag{1.7}$$

in which $\rho_n = 1 - b/n$, b is a positive constant, and ε_t is defined as in (1.2). The notion of nearly unstable models has become one of the most important concepts in time series econometrics since the papers of Chan and Wei (1987) and Phillips (1987). It has found widespread applications in the analysis of time series data; for more background information, see the surveying articles of Chan (2006) and Chan (2009). By means of the moment bounds of the EVE and LSE for this class of models, asymptotic expressions for the MSPEs of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ under (1.7),

$$\lim_{n\to\infty} n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2) = R_A(\alpha, b), \ \lim_{n\to\infty} n(\text{MSPE}_B - \sigma^2) = R_B(b), \tag{1.8}$$

are established, where $R_A(\alpha, b)$ is a positive constant depending on α, b and σ^2, and $R_B(b)$ is a positive constant depending on b and σ^2. When data are generated from model (1.1) with ρ fixed but close to 1, $R_A(\alpha, b)$ and $R_B(b)$, with $b = n(1 - \rho)$, can be used in place of $R_A^o(\alpha, \rho)$ and $R_B^o(\rho)$ to approximate $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ and $n(\text{MSPE}_B - \sigma^2)$. Since $R_A(\alpha, n(1 - \rho))$ and $R_B(n(1 - \rho))$ vary with the sample size n, they are referred to as the finite sample approximations. It is shown in Section 3.2 that

$R_A(\alpha, n(1-\rho))$ and $R_B(n(1-\rho))$ substantially outperform $R^o_A(\alpha, \rho)$ and $R^o_B(\rho)$ in situation where $n(1-\rho)$ is small to moderate.

One of the most intriguing features of $R^o_B(\rho)$ is that it exhibits a jump behavior at the point $\rho = 1$. Specifically, according to (47) and (48) of Ing and Yang (2014),

$$R^o_B(\rho) = \begin{cases} 2\sigma^2, \text{ for } 0 \leq \rho < 1, \\ \\ 4\sigma^2, \text{ for } \rho = 1. \end{cases} \tag{1.9}$$

This phenomenon is analogous to the "quantum jump" behavior observed in physics, where the state of a system remains unchanged until a critical amount of energy is accumulated. With the MSPE jump in (1.9), it will be interesting to explore if a connection between the stationary and the unstable regimes can be established via a smooth transition mechanism such as (1.7). In other words, would the following relationship

$$R_B(b) \to \begin{cases} 2\sigma^2, \text{ as } b \to \infty\,, \\ \\ 4\sigma^2, \text{ as } b \to 0\,, \end{cases} \tag{1.10}$$

remain valid? Again, such an issue is a reminiscence of Corollary 1 and Theorem 2 of Chan and Wei (1987), who showed that under (1.7), the limit distribution of the suitably normalized LSE converges to the unit-root case as $b \to 0$; and converges to the stationary case as $b \to \infty$. In the current scenario, although the lower half of (1.10) remains valid for $b \to 0$, the upper half fails to hold and $R_B(b)$ converges to σ^2 as $b \to \infty$; see (2.15) and (2.20) of Section 2.

This discrepancy in the upper half of (1.10) suggests that $1 - (b/n)$ may be converging to unity too rapidly and as a result, $R_B(b)$ does not attain the limiting value $R^o_B(\rho)$ of the stationary case. An immediate remedy would be to determine if there exists a constant $\beta \in (0, 1)$ such that the corresponding AR coefficient $1 - (b/n^\beta)$ (which approaches 1 at a slower rate) would give rise to a limiting MSPE behaving like (1.10). Unfortunately, the answer to this question is still negative. In Section 3.1, we derive the limiting value, $\Lambda_1(\beta, b)$, of $n(\text{MSPE}_B - \sigma^2)$ for the general near unit-root model, namely, (1.7) with $\rho_n = 1 - b/n^\beta, 0 < \beta \leq 1$ and $b > 0$. We show that $\Lambda_1(\beta, b)$ remains at the stationary state, $2\sigma^2$, for $0 < \beta < 1/2$; transits to the intermediate state, σ^2, at $\beta = 1/2$; remains at the intermediate state for $1/2 < \beta < 1$; and transits to the unit-root state, $4\sigma^2$, at $\beta = 1$. Although no single β directly connects $\Lambda_1(\beta, b)$ from $2\sigma^2$ to $4\sigma^2$, our result reveals that a connection can be established through two critical values of β, namely, $\beta = 1/2$ and $\beta = 1$, which connect $\Lambda_1(\beta, b)$ for the stationary and intermediate states, and for the intermediate and unit-root states, respectively. More precisely, we have $\lim_{b\to\infty} \Lambda_1(1/2, b) = 2\sigma^2$, $\lim_{b\to 0} \Lambda_1(1/2, b) = \sigma^2$, $\lim_{b\to\infty} \Lambda_1(1, b) = \lim_{b\to\infty} R_B(b) = \sigma^2$, and $\lim_{b\to 0} \Lambda_1(1, b) = \lim_{b\to 0} R_B(b) = 4\sigma^2$. This feature is not only of theoretical interest, but it also provides an alternative finite sample approximation, $\Lambda_1(1/2, n^{1/2}(1-\rho))$, for $n(\text{MSPE}_B - \sigma^2)$ that surpasses $R_B(n(1-\rho))$ when $n(1-\rho)$ stays far away from 0. This notion is further elaborated in Section 3.2.

Although the EV predictor also encounters the MSPE jump at $\rho = 1$ in the sense that

$$\lim_{\rho\to 1} R_A^o(\alpha,\rho) \neq R_A^o(\alpha,1), \tag{1.11}$$

this discrepancy can be eliminated by $R_A(\alpha, b)$, which satisfies for any $0 < \alpha < \infty$,

$$\lim_{\rho\to 1} R_A^o(\alpha,\rho) = \lim_{b\to\infty} R_A(\alpha,b), \quad R_A^o(\alpha,1) = \lim_{b\to 0} R_A(\alpha,b). \tag{1.12}$$

For more details, see Remark 1 of Section 2. To deepen our understanding of the EV predictor in the near unit-root region, we also obtain an asymptotic expression for MSPE_A in the general near unit-root model. This result leads to an alternative finite sample approximation of $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$, which notably improves upon $R_A(\alpha, n(1-\rho))$ when $n(1-\rho)$ is relatively large. For more details, see Section 3.2.

In short, the focuses of this paper can be succinctly summarized as follows:

(1) Analysis of near unit-root processes from a prediction perspective.

(2) Analysis of general near unit-root processes, thereby allowing one to understand to what degree such general models can be used to establish a link between stationary and unstable models from a prediction perspective.

(3) An illustration of the importance of finite sample approximations of the MSPE derived from near unit-root and general near unit-root processes.

The rest of the paper is organized as follows. In Section 2, asymptotic properties of the EVE and LSE for the near unit-root case with $\rho_n = 1-(b/n)$ are developed, which include: (a) the limit distributions and the moment bounds; (b) the asymptotic expressions for MSPE_A and MSPE_B. In Section 3.1, extensions of the results in Section 2 to the case of $\rho_n = 1-(b/n^\beta), 0 < \beta < 1$, are given. In Section 3.2, we offer finite sample approximations of $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ and $n(\text{MSPE}_B - \sigma^2)$ based on the asymptotic expressions obtained in Sections 2 and 3.1. The resultant performance is illustrated using AR(1) models with Beta$(\alpha, 1)$ errors, in which the AR coefficient lies between 0.86 and 0.99 and $1.5 \leq \alpha \leq 4$ (see also Section 3.2). We conclude in Section 4. With the help of Section 3.2, we further provide a simple rule for choosing a finite sample approximation from those derived in the near unit-root and the general near unit-root models. This result, together with the proofs of the theorems in Sections 2 and 3.1, is deferred to the supplementary document.

2 Near Unit-Root Models

In this section, we provide asymptotic expressions for the MSPEs of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ under (1.7), where

$$\hat{y}_{n+1} = \hat{\mu}_n + \hat{\rho}_n y_n, \tag{2.1}$$

with $\hat{\mu}_n = \frac{1}{n-1}\sum_{t=1}^{n-1}(y_{t+1} - \hat{\rho}_n y_t)$, and

$$\tilde{y}_{n+1} = \tilde{\mu}_n + \tilde{\rho}_n y_n, \tag{2.2}$$

noting that with $\boldsymbol{x}_j = (1, y_j)^T$,

$$(\tilde{\mu}_n, \tilde{\rho}_n) = (\sum_{j=1}^{n-1} \boldsymbol{x}_j \boldsymbol{x}_j^T)^{-1} \sum_{j=1}^{n-1} \boldsymbol{x}_j y_{j+1}. \tag{2.3}$$

Let $z_n = (y_n - y_1)/(n-1)$. Then the MSPE of $\hat{y}_{n+1}$ can be expressed as

$$\text{MSPE}_A = \text{E}(y_{n+1} - \hat{y}_{n+1})^2 = \sigma^2 + \text{E}\big\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y}) + [(1-\rho_n)\bar{y} - \mu + z_n]\big\}^2, \tag{2.4}$$

where $\mu = \text{E}(\varepsilon_1)$. In addition, it is straightforward to see that the MSPE of $\tilde{y}_{n+1}$ satisfies

$$\text{MSPE}_B = \text{E}(y_{n+1} - \tilde{y}_{n+1})^2 = \sigma^2 + \text{E}\big\{n^{-1} \sum_{i=1}^{n} \eta_i + (y_n - \bar{y})(\tilde{\rho}_n - \rho_n)\big\}^2, \tag{2.5}$$

where $\eta_i = \varepsilon_i - \mu$. To simplify the exposition, we shall assume that $y_0 = 0$ for the rest of this paper. We begin by deriving the limit distributions of $\hat{\rho}_n$ and $\tilde{\rho}_n$ and providing asymptotic expressions for their mean squared errors (MSEs); see Theorems 2.1 and 2.2. As is clear from (2.4) and (2.5), these results play important roles in analyzing MSPE_A and MSPE_B.

Theorem 2.1. *Assume (1.7) with $0 < b < \infty$. Suppose ε_1 obeys $\text{E}\varepsilon_1^\kappa < \infty$ for some $\kappa > 0$. Then,*

$$\text{E}\{n^{1+1/\alpha}(\hat{\rho}_n - \rho_n)\}^q < \infty, \text{ for any } q > 0. \tag{2.6}$$

Further, if $\text{E}(\varepsilon_1^{1+\tau}) < \infty$ for some $\tau > 0$, then

$$\lim_{n\to\infty} P\{(cM_{\alpha,b}/\alpha)^{1/\alpha} n^{1+1/\alpha}(\hat{\rho}_n - \rho_n) > t\} = \exp\{-t^\alpha\},\ t > 0 \tag{2.7}$$

and

$$\lim_{n\to\infty} \text{E}[(cM_{\alpha,b}/\alpha)^{1/\alpha} n^{1+1/\alpha}(\hat{\rho}_n - \rho_n)]^2 = \Gamma((\alpha+2)/\alpha), \tag{2.8}$$

where

$$M_{\alpha,b} \;=\; \mu^\alpha \int_0^1 [(1 - \exp(-bx))/b]^\alpha \, dx. \tag{2.9}$$

We offer a brief explanation of why $n^{1+1/\alpha}$ is the valid normalization factor in Theorem 2.1. It is not difficult to see that $\hat{\rho}_n - \rho_n = \min_{2\le i\le n} \varepsilon_i / y_{i-1}$, and $\exp(-b) S_t \le y_t = \sum_{j=0}^{t-1} \rho_n^j \varepsilon_{t-j} \le S_t$, where $S_t = \sum_{j=1}^{t} \varepsilon_j$. Moreover, it is shown in the proof of Theorem 2 of Ing and Yang (2014) that there exists $1/2 < \theta < 1$ for which $n^{1+1/\alpha} \min_{2\le i\le n} \varepsilon_i / S_{i-1} = n^{1+1/\alpha} \min_{n^\theta \le i\le n} \varepsilon_i / [(i-1)\mu] + o_p(1)$. Since the limit distribution of $n^{1+1/\alpha} \min_{n^\theta\le i\le n} \varepsilon_i/[(i-1)\mu]$ is non-degenerate, this explains why $n^{1+1/\alpha}$ is required by $\hat{\rho}_n - \rho_n$ to yield the desired results.

Theorem 2.2. *Assume (1.7) with $0 < b < \infty$. Suppose ε_1 satisfies $\text{E}\varepsilon_1^{2+\tau} < \infty$ for some $\tau > 0$. Then,*

$$\frac{n^{3/2}}{\sigma_{1,b}} (\tilde{\rho}_n - \rho_n) \xrightarrow{d} N(0,1), \tag{2.10}$$

where $\sigma_{1,b}^2 = \sigma^2/[\mu^2(I_1(b) - I_2(b))]$, *with* $I_1(b) = 2^{-1}b^{-3}[2b + 4\exp(-b) - \exp(-2b) - 3]$ *and* $I_2(b) = b^{-4}[b - 1 + \exp(-b)]^2$. *Further, assume that* $\mathrm{E}\varepsilon_1^s < \infty$ *for some* $s > 10$ *and there exist positive constants* K, a *and* δ *such that for all* $|x - y| \le \delta$ *and all large* m,

$$|F_m(x) - F_m(y)| \le K|x - y|^a, \tag{2.11}$$

where F_m *is the distribution function of* $m^{-1/2}\sum_{t=1}^m(\varepsilon_t - \mathrm{E}\varepsilon_1)$. *Then*

$$\lim_{n\to\infty} \mathrm{E}[n^3(\tilde{\rho}_n - \rho_n)^2] = \sigma_{1,b}^2 \tag{2.12}$$

It is interesting to point out that $\lim_{b\to 0}\sigma_{1,b}^2 = 12\sigma^2/\mu^2$, which is exactly the limiting variance of $n^{3/2}(\tilde{\rho}_n - 1)$ derived in the unit-root model; see Chan (1989) and Ing and Yang (2014). Based on (2.4) and Theorem 2.1, the next theorem gives an asymptotic expression for MSPE_A. Define $L_3(b) = b^{-4}[1 - \exp(-b) - b\exp(-b)]^2$ and $M'_{\alpha,b} = M_{\alpha,b}/\mu^\alpha$.

Theorem 2.3. *Assume (1.7) with* $0 < b < \infty$. *Suppose* ε_1 *obeys* $\mathrm{E}\varepsilon_1^\kappa < \infty$ *for some* $\kappa > 2$. *Then,*

$$\mathrm{MSPE}_A - \sigma^2 = n^{-2/\alpha}\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{cM'_{\alpha,b}}\Big)^{\frac{2}{\alpha}} L_3(b) + \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{-2/\alpha}\}), \tag{2.13}$$

yielding

$$\lim_{n\to\infty} n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2) = R_A(\alpha, b), \tag{2.14}$$

where

$$R_A(\alpha, b) = \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{cM'_{\alpha,b}}\Big)^{\frac{2}{\alpha}} L_3(b) I(\alpha \ge 2) + \sigma^2 I(\alpha \le 2),$$

and the dependence of $R_A(\alpha, b)$ *on* c *is suppressed to simplify notation.*

Note that in view of the proof of Theorem 2.3, the cross-product term, $2\mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})[(1-\rho_n)\bar{y} - \mu + z_n]\}$, in the expectation on the right-hand side of (2.4) is asymptotically negligible compared to the corresponding squared terms $\mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})\}^2$ and $\mathrm{E}\{(1-\rho_n)\bar{y} - \mu + z_n\}^2$. Therefore, the asymptotic behavior of $\mathrm{MSPE}_A - \sigma^2$ is mainly determined by the last two expectations. Since $(1-\rho_n)\bar{y} - \mu + z_n = (n-1)^{-1}\sum_{j=1}^{n-1}\eta_{j+1}$, we have $\mathrm{E}\{(1-\rho_n)\bar{y} - \mu + z_n\}^2 = \sigma^2/(n-1)$. In addition, $(y_n - \bar{y})^2/n^2$ converges in probability to $\mu^2 L_3(b)$; see (A.16). From this and (2.8), it is expected that $\mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})\}^2$ is of order $n^{-2/\alpha}$, and the details are presented in the proof of Theorem 2.3. As a result, the normalization factor $n^{\min\{1,2/\alpha\}}$ is needed.

Equipped with (2.5) and Theorem 2.2, we have the following asymptotic expression for MSPE_B.

Theorem 2.4. *Assume (1.7) with* $0 < b < \infty$. *Suppose* ε_1 *satisfies (2.11) and* $\mathrm{E}\varepsilon_1^s < \infty$ *for some* $s > 12$. *Then*

$$\lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2) = R_B(b) = \Big\{1 + \frac{L_3(b)}{I_1(b) - I_2(b)}\Big\}\sigma^2. \tag{2.15}$$

We conclude this section with a few remarks for Theorems 2.3 and 2.4.

Remark 1. By (37) of Ing and Yang (2014), for $0 \le \rho < 1$, we have

$$\mathrm{MSPE}_A - \sigma^2 = n^{-2/\alpha}\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big\{\frac{\alpha}{cM_\alpha(\rho)}\Big\}^{\frac{2}{\alpha}}\frac{\sigma^2}{1-\rho^2} + \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{-2/\alpha}\}), \tag{2.16}$$

yielding

$$\lim_{n\to\infty} n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2) = R_A^o(\alpha,\rho), \tag{2.17}$$

where

$$R_A^o(\alpha,\rho) = \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big\{\frac{\alpha}{cM_\alpha(\rho)}\Big\}^{\frac{2}{\alpha}}\frac{\sigma^2}{1-\rho^2}I(\alpha\ge 2) + \sigma^2 I(\alpha\le 2).$$

In addition, (42) of Ing and Yang (2014) implies that for $\rho = 1$,

$$n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2) = R_A^o(\alpha,1) = \frac{1}{4}\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big\{\frac{\alpha(\alpha+1)}{c}\Big\}^{\frac{2}{\alpha}} I(\alpha\ge 2) + \sigma^2 I(\alpha\le 2). \tag{2.18}$$

Therefore, (1.11) follows from (2.17), (2.18) and

$$\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big\{\frac{\alpha}{cM_\alpha(\rho)}\Big\}^{\frac{2}{\alpha}}\frac{\sigma^2}{1-\rho^2} \to 0 \text{ as } \rho\to 1.$$

On the other hand, in view of

$$\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{cM'_{\alpha,b}}\Big)^{\frac{2}{\alpha}} L_3(b) \to \begin{cases} \frac{1}{4}\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big\{\frac{\alpha(\alpha+1)}{c}\Big\}^{\frac{2}{\alpha}}, \text{ as } b\to 0\,, \\ 0, \text{ as } b\to\infty\,, \end{cases}$$

the discrepancy between $R_A^o(\alpha,1)$ and $R_A^o(\alpha,\rho)$ in (1.11) can be connected by $R_A(\alpha,b)$ in the sense of (1.12).

Remark 2. While

$$\lim_{b\to 0} R_B(b) = R_B^o(1) = 4\sigma^2, \tag{2.19}$$

$\lim_{b\to\infty} R_B(b)$ is not equivalent to $R_B^o(\rho)$ when ρ increases to 1. More specifically,

$$\sigma^2 = \lim_{b\to\infty} R_B(b) \ne \lim_{\rho\to 1} R_B^o(\rho) = 2\sigma^2, \tag{2.20}$$

recalling that $R_B^o(\rho) = 2\sigma^2$ for all $0 \le \rho < 1$. Equation (2.20) seems to suggest that $\rho_n = 1 - b/n$ converges to unity too fast, and hence $\lim_{b\to\infty} R_B(b)$ does not align with $\lim_{\rho\to 1} R_B^o(\rho)$. This motivates one to ask if there exists a $\beta\in(0,1)$ such that $\lim_{b\to 0} R_B(b) = 4\sigma^2$ and $\lim_{b\to\infty} R_B(b) = 2\sigma^2$, provided ρ_n is replaced by $1-(b/n^\beta)$. This question will be discussed in Section 3.

Remark 3. Theorems 2.3 and 2.4 imply that when $\alpha < 2$ ($\alpha > 2$), the EV (LS) predictor is better (worse) than the LS (EV) predictor in the sense that the convergence rate of $\mathrm{MSPE}_A - \sigma^2$ ($\mathrm{MSPE}_B - \sigma^2$)

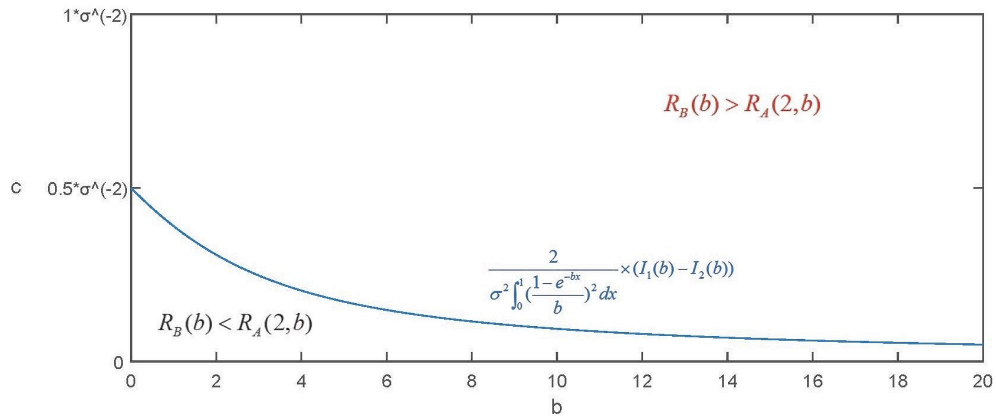

Figure 1: The graph of $c = h_\sigma(b)$.

is faster than $\text{MSPE}_B - \sigma^2$ ($\text{MSPE}_A - \sigma^2$). However, when $\alpha = 2$, $\text{MSPE}_A - \sigma^2$ and $\text{MSPE}_B - \sigma^2$ share the same rate of convergence, and EV (LS) predictor is more efficient than the LS (EV) predictor if and only if $R_A(\alpha, b) < R_B(b)$ ($R_A(\alpha, b) > R_B(b)$). In fact, it is not difficult to see that $R_A(\alpha, b) < R_B(b)$ or $R_A(\alpha, b) > R_B(b)$ depends on whether c is greater or smaller than the threshold function $h_\sigma(b) = 2(I_1(b) - I_2(b))[(\sigma^2/b^2)\int_0^1 (1 - \exp(-bx))^2 dx]^{-1}$. Moreover, $\lim_{b\to 0} h_\sigma(b) = 1/(2\sigma^2)$ is exactly the threshold value in the case of $\rho = 1$ that partitions c into $c > 1/(2\sigma^2)$, in which the EV predictor dominates the LS predictor; and $0 < c < 1/(2\sigma^2)$, in which the latter predictor becomes more appealing, see Section 3 of Ing and Yang (2014). Additionally, $h_\sigma(b)$ is a decreasing function of b, suggesting that the advantage of the EV predictor over the LS one is more evident as the underlying process becomes "less non-stationary". It is also interesting to point out that in the case $0 \le \rho < 1$, $h_{\mu,\sigma}(\rho) = 2(1-\rho)\{(1+\rho)\mu^2 + (1-\rho)\sigma^2\}^{-1}$ is the threshold function playing the same role as $h_\sigma(b)$ in the near unit-root case (see Section 3 of Ing and Yang (2014)), and $h_{\mu,\sigma}(\rho)$ and $h_\sigma(b)$ coincide in the limit in the sense that $\lim_{\rho\to 1} h_{\mu,\sigma}(\rho) = \lim_{b\to\infty} h_\sigma(b) = 0$. The above discussion is illustrated graphically by Figure 1. Finally, we mention that in view of Theorems 2.1 and 2.2, the rankings of $\hat{\rho}_n$ and $\tilde{\rho}_n$ in terms of MSE are exactly the same as those of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ in terms of MSPE.

Before closing this section, we remark that when data are generated from model (1.1) with ρ fixed but close to 1, $R_A(\alpha, n(1-\rho))$ and $R_B(n(1-\rho))$ can also be used in place of $R_A^o(\alpha, \rho)$ and $R_B^o(\rho)$ to approximate $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ and $n(\text{MSPE}_B - \sigma^2)$. These types of approximations, varying with the sample size n, are referred to as the finite sample approximations of $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ and $n(\text{MSPE}_B - \sigma^2)$, or finite sample corrections of $R_A^o(\alpha, \rho)$ and $R_B^o(\rho)$. As shown in Section 3.2, the performance of $R_A(\alpha, n(1-\rho))$ and $R_B(n(1-\rho))$ substantially improves upon that of $R_A^o(\alpha, \rho)$ and $R_B^o(\rho)$ when $n(1-\rho)$ is small to moderate.

3 General Near Unit-Root Models

In this section, we extend the results in Section 2 to the general near unit-root case $\rho_n = 1 - b/n^\beta$, where $0 < \beta < 1$ and $0 < b < \infty$. One major motive for this extension is to alleviate the difficulty of $R_A(\alpha, n(1-\rho))$ ($R_B(n(1-\rho))$) and $R_A^o(\alpha,\rho)$ ($R_B^o(\rho)$) in approximating $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ ($n(\text{MSPE}_B - \sigma^2)$) when $1-\rho$ is small but $n(1-\rho)$ becomes relatively large, see Tables 1–6 for details. In Section 3.1, we derive asymptotic expressions for the MSPE_A and MSPE_B in the general near unit-root model. As shown in Section 3.2, these expressions lead to alternative finite sample approximations of $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ and $n(\text{MSPE}_B - \sigma^2)$ that can indeed achieve the desired goal.

3.1 Asymptotic Theories

We start by exploring the asymptotic distribution and the MSE of $\hat{\rho}_n$.

Theorem 3.1. *Assume* (1.7) *with* $\rho_n = 1 - b/n^\beta$, *where* $0 < \beta < 1$ *and* $0 < b < \infty$. *Suppose that* ε_1 *satisfies* $\text{E}\varepsilon_1^\kappa < \infty$ *for some* $\kappa > 0$. *Then*

$$\text{E}\{n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n)\}^q < \infty, \text{ for any } q > 0. \tag{3.1}$$

Further, if

$$\text{E}(\varepsilon_1^{q_1}) < \infty, \text{ for some } q_1 > 2/\beta, \tag{3.2}$$

then

$$\lim_{n\to\infty} P\{(c/\alpha)^{1/\alpha}(\mu/b)n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n) > t\} = \exp\{-t^\alpha\},\ t > 0\,, \tag{3.3}$$

and

$$\lim_{n\to\infty} \text{E}[(c/\alpha)^{1/\alpha}(\mu/b)n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n)]^2 = \Gamma((\alpha+2)/\alpha). \tag{3.4}$$

Theorem 3.1 reveals that the convergence rate of $\hat{\rho}_n - \rho_n$ is slower in the general near unit-root case than in the near unit-root one. This is because in the former case, the denominator, y_i, in $\hat{\rho}_n - \rho_n = \min_{1\le i\le n-1} \varepsilon_{i+1}/y_i$ has order of magnitude $n^\beta(1-\rho_n^i)$, which is dominated by the order of magnitude of y_i in the near unit-root case.

Theorem 3.2. *Assume* (1.7) *with* $\rho_n = 1 - b/n^\beta$, *where* $0 < \beta < 1$ *and* $0 < b < \infty$. *Let* ε_1 *satisfy* $\text{E}(\varepsilon_1^{2+\tau}) < \infty$ *for some* $\tau > 0$. *Then,*

$$\sqrt{k_n}(\tilde{\rho}_n - \rho_n) \xrightarrow{d} N(0, \sigma^2_{\beta,b}), \tag{3.5}$$

where $k_n = n^{1+\beta}$ *if* $0 < \beta \le 1/2$ *and* $n^{3\beta}$ *if* $\beta > 1/2$, *and*

$$\sigma^2_{\beta,b} = \begin{cases} \{(2b^3\sigma^2)^{-1}\mu^2 I(\beta = 1/2) + (2b)^{-1}\}^{-1}, & \text{if } 0 < \beta \le 1/2\,, \\ 2b^3\sigma^2/\mu^2\,, & \text{if } 1/2 < \beta < 1. \end{cases}$$

Moreover, if $\mathrm{E}\varepsilon_1^s < \infty$ *for some* $s > 10$ *and* (2.11) *holds, then*

$$\lim_{n\to\infty} \mathrm{E}[k_n(\tilde{\rho}_n - \rho_n)^2] = \sigma^2_{\beta,b}. \tag{3.6}$$

It is shown in the proof of Theorem 3.2 that $\tilde{\rho}_n - \rho_n = \sum_{i=2}^n (y_{i-1} - \bar{y})\eta_i / \sum_{i=2}^n (y_{i-1} - \bar{y})^2 = \sum_{i=2}^n (a_{n,i-1} + \xi_{n,i-1})\eta_i / \sum_{i=2}^n (a_{n,i-1} + \xi_{n,i-1})^2$, where $a_{n,i}$ and $\xi_{n,i}$, respectively, denote the deterministic and random components of $y_i - \bar{y}$. Moreover, the order of magnitude of $(\tilde{\rho}_n - \rho_n)^2$ is determined by $(\sum_{i=2}^n a^2_{n,i-1})^{-1}$ for $1/2 < \beta < 1$, $(\sum_{i=2}^n \xi^2_{n,i-1})^{-1}$ for $0 < \beta < 1/2$, and $(\sum_{i=2}^n a^2_{n,i-1} + \xi^2_{n,i-1})^{-1}$ for $\beta = 1/2$, and the growth rates of $\sum_{i=2}^n a^2_{n,i-1}$ and $\sum_{i=2}^n \xi^2_{n,i-1}$ are $n^{3\beta}$ and $n^{1+\beta}$, respectively. This clarifies why the exponent of k_n needs to be changed from $1+\beta$ to 3β after $\beta > 1/2$.

With the help of Theorems 3.1 and 3.2, the next two theorems provide asymptotic expressions for MSPE_A and MSPE_B in the general near unit-root model.

Theorem 3.3. *Assume* (1.7) *with* $\rho_n = 1 - b/n^\beta$, *where* $0 < \beta < 1$ *and* $0 < b < \infty$. *Suppose that* ε_1 *obeys* (3.2). *Then the following conclusions hold.*

(a) For $0 < \beta < 2/3$,

$$\mathrm{MSPE}_A - \sigma^2 = n^{-\beta-\frac{2}{\alpha}} \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \frac{\sigma^2 b}{2\mu^2} + \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{-\beta-\frac{2}{\alpha}}\}). \tag{3.7}$$

(b) For $2/3 < \beta < 1$,

$$\mathrm{MSPE}_A - \sigma^2 = n^{2\beta-2-\frac{2}{\alpha}} \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \frac{1}{b^2} + \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{2\beta-2-\frac{2}{\alpha}}\}). \tag{3.8}$$

(b) For $\beta = 2/3$,

$$\begin{aligned}\mathrm{MSPE}_A - \sigma^2 &= n^{-\frac{2}{3}-\frac{2}{\alpha}} \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \Big\{\frac{\sigma^2 b}{2\mu^2} + \frac{1}{b^2}\Big\} \\ &+ \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{-\frac{2}{3}-\frac{2}{\alpha}}\}).\end{aligned} \tag{3.9}$$

It is worth mentioning that the first terms on the right-hand sides of (3.7) and (3.8), respectively, are asymptotically equivalent to those on the right-hand sides of (2.16) and (2.13) with ρ and b replaced by ρ_n and $n(1-\rho_n)$, where ρ_n is defined in Theorem 3.3. Moreover, the first term on the right-hand side of (3.9) is the sum of the first terms on right-hand sides of (3.7) and (3.8). These features reveal that when ρ_n converges to 1 at a rate slower than $n^{-2/3}$, the asymptotic behavior of MSPE_A is governed by the stationary formula (2.16). But when ρ_n converges to 1 at a rate faster than $n^{-2/3}$, the asymptotic behavior of MSPE_A is governed by the near unit-root formula (2.13). At the critical point $\beta = 2/3$, however, MSPE_A is related to both (2.13) and (2.16). In fact, one can combine (3.7)–(3.9) into

$$\begin{aligned}&\mathrm{MSPE}_A - \sigma^2 \\ &= n^{-2/\alpha} \Big\{\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \Big[\frac{\sigma^2 b}{2\mu^2 n^\beta} + \frac{1}{n^2 (b/n^\beta)^2}\Big]\Big\} + \frac{\sigma^2}{n} \\ &+ o(\max\{n^{-1}, n^{-\beta-\frac{2}{\alpha}}, n^{2\beta-2-\frac{2}{\alpha}}\}),\end{aligned} \tag{3.10}$$

where $0 < \beta < 1$. Equation (3.10) leads to another finite sample approximation of $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$; see (3.17) in Section 3.2. As will be shown in Section 3.2, (3.17) serves as a satisfactory complement to $R_A^o(\alpha,\rho)$ and $R_A(\alpha, n(1-\rho))$ when the latter two approximations perform poorly.

Theorem 3.4. *Assume* (1.7) *with* $\rho_n = 1 - b/n^\beta$*, where* $0 < \beta < 1$ *and* $0 < b < \infty$*. Suppose that* ε_1 *satisfies the same assumptions as in Theorem 2.4. Then*

$$\Lambda_1(\beta,b) \equiv \lim_{n\to\infty} n(\text{MSPE}_B - \sigma^2) = \begin{cases} 2\sigma^2\,, & 0 < \beta < 1/2\,, \\ \sigma^2\{1 + \left(\frac{b^2\sigma^2}{\mu^2 + b^2\sigma^2}\right)\}\,, & \beta = 1/2\,, \\ \sigma^2\,, & 1/2 < \beta < 1\,. \end{cases} \tag{3.11}$$

In view of (2.15), (2.19) and (2.20), Theorem 3.4 can be succinctly summarized to include the case $\beta = 1$ as follows.

$$\Lambda_1(\beta,b) = \begin{cases} 2\sigma^2\,, & 0 < \beta < 1/2 \text{ and } 0 < b < \infty\,, \\ \sigma^2\{1 + \left(\frac{b^2\sigma^2}{\mu^2 + b^2\sigma^2}\right)\}\,, & \beta = 1/2 \text{ and } 0 < b < \infty\,, \\ \sigma^2\,, & 1/2 < \beta < 1 \text{ and } 0 < b < \infty\,, \\ R_B(b)\,, & \beta = 1 \text{ and } 0 < b < \infty\,, \\ 4\sigma^2\,, & \beta = 1 \text{ and } b = 0. \end{cases} \tag{3.12}$$

Note that $\Lambda_1(\beta,b) = 2\sigma^2, 0 < \beta < 1/2$ is designated as the "stationary state", while $\Lambda_1(1,0) = 4\sigma^2$ is designated as the "unit-root state". Moreover, $\Lambda_1(\beta,b)$ with $1/2 < \beta < 1$ is designated as the "intermediate state" because its value, σ^2, is different from the values of the unit-root and the stationary states.

Equation (3.12) also reveals that there are two critical points, $\beta = 1/2$ and $\beta = 1$, as far as asymptotic predictions are concerned. At the critical point $\beta = 1$ that separates the unit-root and intermediate states, we have

(i) $\lim_{b\to 0} \Lambda_1(1,b) = 4\sigma^2$, which is the value of the unit-root state,

(ii) $\lim_{b\to\infty} \Lambda_1(1,b) = \sigma^2$, which is the value of the intermediate state.

At the critical point $\beta = 1/2$ that separates the stationary and intermediate states, we have

(i) $\lim_{b\to 0} \Lambda_1(\frac{1}{2},b) = \sigma^2$, which is the value of the intermediate state,

(ii) $\lim_{b\to\infty} \Lambda_1(\frac{1}{2},b) = 2\sigma^2$, which is the value of the stationary state.

Graphically, these critical phenomena are depicted in Figure 2.

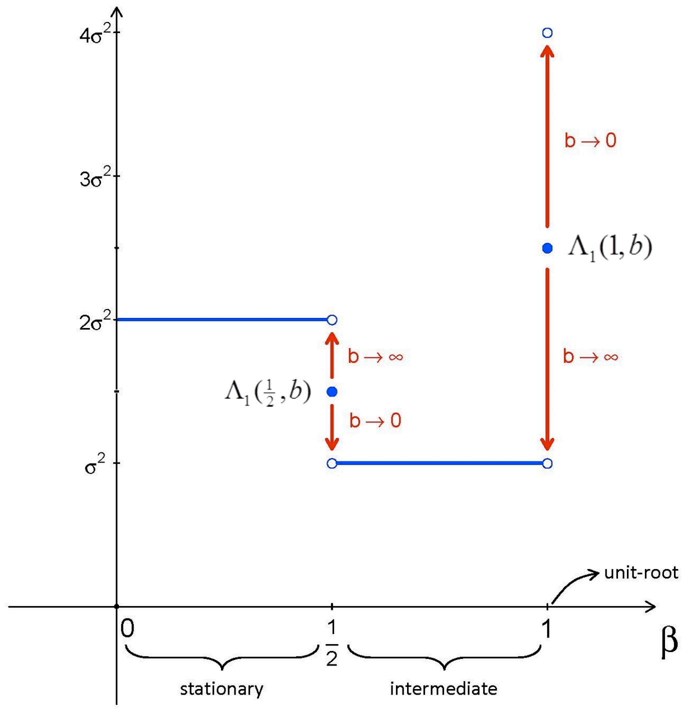

Figure 2: Theorem 3.4: the vertical axis denotes the value of $\Lambda_1(\beta, b)$. Note that the points associated with $\Lambda_1(1/2, b)$ and $\Lambda_1(1, b)$ are not necessarily in the middle. They are only used to illustrate $\Lambda_1(1/2, b)$ ($\Lambda_1(1, b)$) decreases (increases) to σ^2 ($4\sigma^2$) as $b \to 0$, and increases (decreases) to $2\sigma^2$ (σ^2) as $b \to \infty$.

One important conclusion deduced from the above discussion is that due to the existence of the intermediate state, there exists no $0 < \beta \le 1$ such that $\Lambda_1(\beta, b)$ simultaneously satisfies $\lim_{b\to 0} \Lambda_1(\beta, b) = 4\sigma^2$ and $\lim_{b\to\infty} \Lambda_1(\beta, b) = 2\sigma^2$, implying that the discontinuity between the unit-root and stationary states of the MSPE_B cannot be connected by a general near unit-root model. In contrast, we also note that this discontinuity in MSPE_B can indeed be connected through"two" general near unit-root models whose β values correspond to the aforementioned two critical points, 1 and 1/2, as illustrated in Figure 2. Moreover, when data are generated from model (1.1) with fixed ρ, the critical point $\beta = 1/2$ inspires another finite sample approximation, $\Lambda_1(1/2, n^{1/2}(1-\rho))$, of $n(\mathrm{MSPE}_B - \sigma^2)$ that is expected to surpass $\Lambda_1(1, n(1-\rho)) = R_B(n(1-\rho))$ when $n(1-\rho)$ becomes relatively large. The performance of $\Lambda_1(1/2, n^{1/2}(1-\rho))$ is also demonstrated in Section 3.2.

3.2 Simulations

In this section, we propose finite sample approximations of $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ and $n(\mathrm{MSPE}_B - \sigma^2)$ (finite sample corrections of $R^o_A(\alpha, \rho)$ and $R^o_B(\rho)$) based on Theorems 2.3, 2.4, 3.3 and 3.4, and investigate their performance via simulated data generated from model (1.1), with $\rho \in \{0.86, 0.9, 0.95, 0.975, 0.99\}$, $n \in \{100, 200, 500, 1000, 3000, 6000, 10000\}$, and ε_t obeying the Beta$(\alpha, 1)$ distribution with $\alpha \in \{1.5, 2, 4\}$. Note that the performance of our finite sample corrections in the case of $0 < \alpha < 1$ is largely similar to that in the case of $\alpha = 1.5$. The details are skipped here. For a given triple (n, ρ, α), let $y_1^{(j)}, \dots, y_n^{(j)}$ denote the data generated in the j-th simulation, where $1 \leq j \leq 5000$. We also generate $y_{n+1}^{(j)}$ and compute the empirical estimate of $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$,

$$R_{A_n} = n^{\min\{1,\frac{2}{\alpha}\}} \left\{ \frac{1}{5000} \sum_{j=1}^{5000} (y_{n+1}^{(j)} - \hat{y}_{n+1}^{(j)})^2 - \sigma^2 \right\}, \tag{3.13}$$

and that of $n(\text{MSPE}_B - \sigma^2)$,

$$R_{B_n} = n \left\{ \frac{1}{5000} \sum_{j=1}^{5000} (y_{n+1}^{(j)} - \tilde{y}_{n+1}^{(j)})^2 - \sigma^2 \right\}, \tag{3.14}$$

where $\hat{y}_{n+1}^{(j)}$ and $\tilde{y}_{n+1}^{(j)}$, respectively, denote the EV and LS predictors calculated in the j-th simulation based on $y_1^{(j)}, \dots, y_n^{(j)}$.

While (1.6) suggests that for a sufficiently large n, $R_A^o(\alpha, \rho)$ and $R_B^o(\rho)$ should give good approximations of $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ (or R_{A_n}) and $n(\text{MSPE}_B - \sigma^2)$ (or R_{B_n}), the results are often unsatisfactory in the scenarios considered. In view of Theorems 2.3 and 2.4, when $n(1-\rho)$ is small, it seems reasonable to use $R_A(\alpha, n(1-\rho))$ and $R_B(n(1-\rho))$ as finite sample corrections for $R_A^o(\alpha, \rho)$ and $R_B^o(\rho)$, where $R_A(\alpha, b)$ and $R_B(b)$ are defined in (2.14) and (2.15), and $n(1-\rho)$ can be viewed as a measure of how far the underlying model deviates from the unit-root model.

On the other hand, when $1-\rho$ is small but $n(1-\rho)$ becomes relatively large, both $R_A(\alpha, n(1-\rho))$ and $R_A^o(\alpha, \rho)$ may perform inferiorly. Therefore, we suggest an alternative approximation of $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ based on (3.10), which is derived from the general near unit-root model. More specifically, ignoring the smaller order terms in (3.10), we obtain from the equation that $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ is approximately equal to

$$\left\{ \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \Big[\frac{\sigma^2 b}{2\mu^2 n^\beta} + \frac{1}{n^2(b/n^\beta)^2} \Big] \right\} + \frac{\sigma^2}{n^{1-2/\alpha}} \tag{3.15}$$

for $\alpha \geq 2$, and

$$\sigma^2 + n^{1-2/\alpha} \left\{ \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \Big[\frac{\sigma^2 b}{2\mu^2 n^\beta} + \frac{1}{n^2(b/n^\beta)^2} \Big] \right\} \tag{3.16}$$

for $0 < \alpha < 2$. Replacing b/n^β in (3.15) and (3.16) by $1-\rho$, one gets an alternative to $R_A(\alpha, n(1-\rho))$:

$$R_A^*(\alpha, n, 1-\rho) = \begin{cases} R_1^*(\alpha, \rho) + R_2^*(\alpha, n(1-\rho)) + \frac{\sigma^2}{n^{1-2/\alpha}}\,, & \alpha \geq 2\,, \\ \sigma^2 + n^{1-2/\alpha} R_1^*(\alpha, \rho) + n^{1-2/\alpha} R_2^*(\alpha, n(1-\rho))\,, & 0 < \alpha < 2, \end{cases} \tag{3.17}$$

where

$$R_1^*(\alpha, \rho) = \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{2/\alpha} \frac{\sigma^2(1-\rho)}{2\mu^2},$$

and

$$R_2^*(\alpha, n(1-\rho)) = \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{2/\alpha} \frac{1}{n^2(1-\rho)^2}.$$

One of the most appealing features of (3.17) is that it depends on the parameters, β and b, in the general near unit-root model only through $1-\rho$, and hence can be implemented without facing the identifiability issue associated with β and b.

Similarly, both $R_B(\alpha, n(1-\rho))$ and $R_B^o(\alpha,\rho)$ may perform poorly for small $1-\rho$ and for relatively large $n(1-\rho)$. We therefore consider the following alternative,

$$\Lambda_1(1/2, n^{1/2}(1-\rho)) = \sigma^2\left\{1+\left(\frac{n(1-\rho)^2\sigma^2}{\mu^2+n(1-\rho)^2\sigma^2}\right)\right\}, \tag{3.18}$$

which is $\Lambda_1(1/2,b)$ with b replaced by $n^{1/2}(1-\rho)$. Since $\Lambda_1(1/2,b)$ falls between $(\sigma^2, 2\sigma^2)$ and can fill the gap between $2\sigma^2 = R_B^o(\alpha,\rho)$ and $\sigma^2 = \lim_{b\to\infty} R_B(\alpha,b)$ in the sense that $\lim_{b\to\infty}\Lambda_1(1/2,b) = 2\sigma^2$ and $\lim_{b\to 0}\Lambda_1(1/2,b) = \sigma^2$, it is expected that $\Lambda_1(1/2,n^{1/2}(1-\rho))$ will provide a more satisfactory approximation of $n(\text{MSPE}_B - \sigma^2)$ when ρ is too small (large) for $R_B(n(1-\rho))$ ($R_B^o(\alpha,\rho)$) to do a good job. We emphasize that the b in $\Lambda_1(1/2,b)$ is replaced by $n^{1/2}(1-\rho)$ instead of $n(1-\rho)$ because $\Lambda_1(1/2,b)$ is derived from the general near unit-root model with $\beta = 1/2$.

For notational simplicity, define

$$\begin{aligned} R_A^{(1)} &= R_A^o(\alpha,\rho),\ R_A^{(2)} = R_A(\alpha, n(1-\rho)),\ R_A^{(3)} = R_A^*(\alpha,n,1-\rho),\\ R_B^{(1)} &= R_B^o(\rho),\quad R_B^{(2)} = R_B(n(1-\rho)),\quad R_B^{(3)} = \Lambda_1(1/2,n^{1/2}(1-\rho)). \end{aligned}$$

The degree of closeness between $R_A^{(i)}$ and R_{A_n} and that between $R_B^{(i)}$ and R_{B_n} are assessed by

$$P_A^{(i)} = \frac{\min\{R_A^{(i)}, R_{A_n}\}}{\max\{R_A^{(i)}, R_{A_n}\}} \text{ and } P_B^{(i)} = \frac{\min\{R_B^{(i)}, R_{B_n}\}}{\max\{R_B^{(i)}, R_{B_n}\}},\ i=1,2,3.$$

Clearly, $0 \le P_A^{(i)}, P_B^{(i)} \le 1$ and a larger value represents a better performance. The values of R_{A_n} (R_{B_n}) and $P_A^{(i)}(P_B^{(i)}), i=1,2,3$, are summarized in Tables 1–3 (Tables 4–6) for $\alpha = 1.5, 2$ and 4, respectively. To better explain these tables, the corresponding value of $n(1-\rho)$, which is denoted by b, is also included.

Table 1 shows that when $\alpha = 1.5$, $P_A^{(1)} = P_A^{(2)}$ for each (n,ρ), which is due to $R_A^{(1)} = R_A^{(2)}$ for each (n,ρ). On the other hand, $P_A^{(3)}$ is notably larger than $P_A^{(1)} = P_A^{(2)}$ for $n = 100$ or $1 \le n(1-\rho) \le 5$ although $P_A^{(i)}, i=1,2,3$, have similar values otherwise. These facts reveal that while $n^{1-2/\alpha}(R_1^*(\alpha,\rho)+R_1^*(\alpha,n(1-\rho)))$ in $R_A^{(3)}$ is asymptotically negligible compared to the corresponding leading term, $\sigma^2 = R_A^{(1)} = R_A^{(2)}$, in the case of $\alpha < 2$, it can help improve the finite sample performance of this leading term. For any fixed ρ, all $P_A^{(i)}, i=1,\ldots,3$, gradually approach 1 as n increases, which coincides with the asymptotic result established for the stationary case, namely, the first relation of (1.6). Table 2 shows that when $\alpha = 2$ and $1 \le n(1-\rho) \le 5$, $P_A^{(2)}$ usually has the highest value among $P_A^{(i)}, i=1,\ldots,3$, except for the case of $(n,\rho) = (200, 0.99)$, where $P_A^{(3)} = 0.9747 > P_A^{(2)} = 0.9144 > P_A^{(1)} = 0.1786$. For $5 < n(1-\rho) \le 28$, $P_A^{(3)}$ appears to dominate its competitors with the exception of $(n,\rho) = (1000, 0.99)$, in which $P_A^{(2)}$ is slightly larger than $P_A^{(3)}$. One explanation of this result is that when $n(1-\rho)$ becomes larger, $R_1^*(\alpha,\rho)$ and $R_2^*(\alpha,n(1-\rho))$ in $R_A^{(3)}$ can complement each other and *jointly* provide a good approximation of R_{A_n}, which is further improved via $\sigma^2/n^{1-2/\alpha}$. For $28 < n(1-\rho) \le 1400$, $P_A^{(1)}$ and $P_A^{(3)}$ behave quite similarly and are usually significantly larger than $P_A^{(2)}$ except for $\rho \ge 0.975$. For any fixed ρ, $P_A^{(1)}$ has an obvious tendency to increase to 1 as n grows from 100 to 10000. This is also in agreement with the first relation of (1.6). Like Table 2, Table 3 ($\alpha = 4$) also shows that $P_A^{(2)}$ usually dominates $P_A^{(i)}, i = 1$ and 3, when $1 \le n(1-\rho) \le 5$. However, an exception happens in the case of $(n, \rho) = (100, 0.975)$, where $P_A^{(3)}$ is ranked first. Note that in this range of $n(1-\rho)$, $P_A^{(1)}$ in Table 3 is less than 0.02 and somewhat smaller than $P_A^{(1)}$ in Table 2. This is because for $\alpha = 2$, $R_A^{(1)}$ contains σ^2, which is non-negligible compared to R_{A_n}. As a result, σ^2 keeps $P_A^{(1)}$ bounded away from zero although the other component of $R_A^{(1)}$ is extremely small. However, for $\alpha = 4$ and small $n(1-\rho)$, $R_A^{(1)}$ is approximately equal to $R_1^*(\alpha, \rho)$, which is substantially smaller than R_{A_n}. The advantage of $R_A^{(3)}$ is more evident in Table 3 since $P_A^{(3)}$ is noticeably larger than $P_A^{(1)}$ and $P_A^{(2)}$ for almost all $10 \le n(1-\rho) \le 1400$. When $70 \le n(1-\rho) \le 1400$, $P_A^{(2)}$ in Table 3 is less than 0.05 and distinctively smaller than $P_A^{(2)}$ in Table 2. We explain this by noting that for $\alpha = 2$ and large $n(1-\rho)$, σ^2 in $R_A^{(2)}$ is quite close to R_{A_n}, and hence induces a relatively large $P_A^{(2)}$ although the other component of $R_A^{(2)}$ is very small. For $\alpha = 4$ and large $n(1-\rho)$, however, $R_A^{(2)}$ is approximately equal to $R_2^*(\alpha, n(1-\rho))$, which is exceedingly small compared to R_{A_n}. For any fixed ρ, Table 3 shows that $P_A^{(1)}$ still possesses a clear upward trend. This phenomenon can once again be explained by the first relation of (1.6). However, because the convergence rate of $\text{MSPE}_A - \sigma^2$ is much slower in $\alpha = 4$ than in $\alpha \le 2$, $P_A^{(1)}$ may not be very close to 1 even when $n = 10000$.

It is shown in Table 4 that $P_B^{(2)} = \max_{1 \le i \le 3} P_B^{(i)}$ for $1 \le n(1-\rho) \le 12.5$, and $P_B^{(3)} = \max_{1 \le i \le 3} P_B^{(i)}$ for $12.5 < n(1-\rho) \le 300$, with the exception of $n(1-\rho) = 30$, where $P_B^{(2)}$ is slightly larger than $P_B^{(3)}$. For $300 < n(1-\rho) \le 1400$, $P_B^{(1)}$ and $P_B^{(3)}$ have similar values and are much larger than $P_B^{(2)}$. Table 4 also reveals that for $\rho = 0.99$, R_{B_n} decreases from 0.217 to 0.0761 as n grows from 100 to 3000, and then slowly increases to 0.08 as n increases to 10000. When ρ gets smaller, this feature of R_{B_n}, first decreasing and then increasing as n grows, becomes more evident. For example, R_{B_n} with $\rho = 0.975$ decreases from 0.158 to 0.084 as n grows from 100 to 1000, and then increases to 0.110 as n increases to 10000. Moreover, for $\rho = 0.95$ (0.9, 0.86), R_{B_n} decreases from 0.117 (0.103, 0.109) to 0.086 (0.101, 0.108) as n grows from 100 to 500 (200, 200), and then increases to 0.136 (0.139, 0.134) as n increases to 10000. The decreasing part of R_{B_n} can be explained by $R_B^{(2)}$, which decreases to the value of intermediate state, $\sigma^2 = 0.0686$, as $n(1-\rho)$ (or, equivalently, n) increases. The increasing part of R_{B_n}, however, is more in line with $R_B^{(3)}$, which increases to $2\sigma^2$ after leaving the intermediate state. Tables 5 and 6 share similar features as Table 4. Moreover, as α grows, the areas of $n(1-\rho)$ for which $R_B^{(2)}$ works best and $R_B^{(3)}$ outperforms $R_B^{(1)}$ tend to expand. In Figure 3, we give the plots of time series realizations generated from model (1.1), with $\rho = 0.95$ and ε_t obeying Beta(1.5, 1), Beta(2, 1) and Beta(4, 1) distributions. This figure shows that the nonstationary feature of these series becomes more evident as α increases, which may partly explain why $R_B^{(2)}$ and $R_B^{(3)}$ play increasingly essential roles in approximating R_{B_n} when α becomes larger.

As a final remark, we note that Tables 1–3 (Tables 4–6) together portrait situations where $R_A^{(i)}, i = 2, 3$ ($R_B^{(i)}, i = 2, 3$) can approximate R_{A_n} (R_{B_n}) better than $R_A^{(1)}$ ($R_B^{(1)}$). This information in conjunction with suitable estimators of b, c and α enables one to construct a data-driven procedure for estimating $n^{\min\{1, 2/\alpha\}}(\text{MSPE}_A - \sigma^2)$ and $n(\text{MSPE}_B - \sigma^2)$ in the near unit-root region. The details are deferred to the supplementary document.

Table 1: *The values of* R_{An} *and* $P_A^{(i)}, i = 1, 2, 3$, *under the* Beta(1.5, 1) *noise.* $P_A^{(1)}$ *(*$P_A^{(2)}$, $P_A^{(3)}$*) is colored in red (blue, green) when it is equal to* $\max_{1 \le i \le 3} P_A^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{An}	0.0770	0.0786	0.0917	0.1349	0.2150
	$P_A^{(1)}$	0.8909	0.8715	0.7470	0.5078	0.3186
	$P_A^{(2)}$	0.8909	0.8715	0.7470	0.5078	0.3186
	$P_A^{(3)}$	0.9519	0.9361	0.8729	0.8170	0.6608
$n = 200$	b	28	20	10	5.0	2.0
	R_{An}	0.0707	0.0741	0.0743	0.0866	0.1299
	$P_A^{(1)}$	0.9703	0.9244	0.9219	0.7910	0.5273
	$P_A^{(2)}$	0.9703	0.9244	0.9219	0.7910	0.5273
	$P_A^{(3)}$	0.9881	0.9584	0.9633	0.8914	0.9212
$n = 500$	b	70	50	25	12.5	5.0
	R_{An}	0.0718	0.0713	0.0682	0.0719	0.0792
	$P_A^{(1)}$	0.9555	0.9607	0.9956	0.9527	0.8649
	$P_A^{(2)}$	0.9555	0.9607	0.9956	0.9527	0.8649
	$P_A^{(3)}$	0.9833	0.9826	0.9809	0.9720	0.9433
$n = 1000$	b	140	100	50	25	10
	R_{An}	0.0690	0.0707	0.0699	0.0709	0.0729
	$P_A^{(1)}$	0.9938	0.9689	0.9799	0.9661	0.9396
	$P_A^{(2)}$	0.9938	0.9689	0.9799	0.9661	0.9396
	$P_A^{(3)}$	0.9833	0.9861	0.9897	0.9738	0.9585
$n = 3000$	b	420	300	150	75	30
	R_{An}	0.0700	0.0691	0.0670	0.0693	0.0676
	$P_A^{(1)}$	0.9785	0.9913	0.9781	0.9884	0.9868
	$P_A^{(2)}$	0.9785	0.9913	0.9781	0.9884	0.9868
	$P_A^{(3)}$	0.9942	0.9971	0.9724	0.9913	0.9839
$n = 6000$	b	840	600	300	150	60
	R_{An}	0.0688	0.0661	0.0702	0.0694	0.06853
	$P_A^{(1)}$	0.9956	0.9649	0.9757	0.9870	0.9994
	$P_A^{(2)}$	0.9956	0.9649	0.9757	0.9870	0.9994
	$P_A^{(3)}$	0.9913	0.9565	0.9800	0.9903	0.9989
$n = 10000$	b	1400	1000	500	250	100
	R_{An}	0.0681	0.0693	0.0702	0.0702	0.0671
	$P_A^{(1)}$	0.9927	0.9885	0.9757	0.9757	0.9795
	$P_A^{(2)}$	0.9927	0.9885	0.9757	0.9757	0.9795
	$P_A^{(3)}$	0.9825	0.9970	0.9805	0.9786	0.9777

Table 2: *The values of R_{An} and $P_A^{(i)}, i = 1,2,3$, under the* Beta(2,1) *noise.* $P_A^{(1)}$ *(*$P_A^{(2)}$*,* $P_A^{(3)}$*) is colored in red (blue, green) when it is equal to* $\max_{1\le i\le 3} P_A^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{An}	0.0756	0.0812	0.1307	0.2644	0.5096
	$P_A^{(1)}$	0.8580	0.7635	0.4491	0.2160	0.1101
	$P_A^{(2)}$	0.8103	0.8288	0.8256	0.8718	0.9241
	$P_A^{(3)}$	0.9180	0.8843	0.7550	0.8211	0.4824
$n = 200$	b	28	20	10	5.0	2.0
	R_{An}	0.0685	0.0690	0.0745	0.1198	0.3141
	$P_A^{(1)}$	0.9465	0.8986	0.7879	0.4766	0.1786
	$P_A^{(2)}$	0.8302	0.8435	0.9034	0.9007	0.9144
	$P_A^{(3)}$	0.9573	0.9319	0.9218	0.8106	0.9747
$n = 500$	b	70	50	25	12.5	5.0
	R_{An}	0.0658	0.0623	0.0600	0.0673	0.1158
	$P_A^{(1)}$	0.9864	0.9952	0.9783	0.8484	0.4845
	$P_A^{(2)}$	0.8479	0.8973	0.9533	0.9331	0.9318
	$P_A^{(3)}$	0.9803	0.9984	0.9953	0.9438	0.8305
$n = 1000$	b	140	100	50	25	10
	R_{An}	0.0641	0.0619	0.0595	0.0588	0.0696
	$P_A^{(1)}$	0.9880	0.9984	0.9866	0.9711	0.8060
	$P_A^{(2)}$	0.8675	0.8982	0.9395	0.9728	0.9669
	$P_A^{(3)}$	0.9960	0.9999	0.9929	0.9986	0.9508
$n = 3000$	b	420	300	150	75	30
	R_{An}	0.0657	0.0602	0.0605	0.0593	0.0566
	$P_A^{(1)}$	0.9861	0.9710	0.9708	0.9621	0.9914
	$P_A^{(2)}$	0.8445	0.9216	0.9188	0.9386	0.9989
	$P_A^{(3)}$	0.9775	0.9753	0.9704	0.9648	0.9890
$n = 6000$	b	840	600	300	150	60
	R_{An}	0.0677	0.0610	0.0594	0.0576	0.0587
	$P_A^{(1)}$	0.9575	0.9837	0.9880	0.9905	0.9566
	$P_A^{(2)}$	0.8199	0.9095	0.9344	0.9639	0.9508
	$P_A^{(3)}$	0.9491	0.9882	0.9870	0.9910	0.9613
$n = 10000$	b	1400	1000	500	250	100
	R_{An}	0.0652	0.0633	0.0588	0.0562	0.0584
	$P_A^{(1)}$	0.9938	0.9794	0.9982	0.9842	0.9606
	$P_A^{(2)}$	0.8512	0.8767	0.9438	0.9875	0.9503
	$P_A^{(3)}$	0.9862	0.9764	0.9980	0.9836	0.9637

Table 3: *The values of* R_{An} *and* $P_A^{(i)}, i = 1, 2, 3,$ *under the* Beta(4, 1) *noise.* $P_A^{(1)}$ *(*$P_A^{(2)}$*,* $P_A^{(3)}$*) is colored in red (blue, green) when it is equal to* $\max_{1\le i\le 3} P_A^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{An}	0.0109	0.0158	0.0493	0.1444	0.3148
	$P_A^{(1)}$	0.2522	0.1203	0.0183	0.0035	0.0006
	$P_A^{(2)}$	0.4488	0.6266	0.8621	0.9252	0.9371
	$P_A^{(3)}$	0.8966	0.8465	0.7918	0.9963	0.3540
$n = 200$	b	28	20	10	5.0	2.0
	R_{An}	0.0062	0.0067	0.0134	0.0474	0.1808
	$P_A^{(1)}$	0.4440	0.2836	0.0672	0.0105	0.0011
	$P_A^{(2)}$	0.1894	0.3433	0.7388	0.8966	0.9591
	$P_A^{(3)}$	0.9033	0.8876	0.8709	0.7973	0.8084
$n = 500$	b	70	50	25	12.5	5.0
	R_{An}	0.0041	0.0035	0.0037	0.0081	0.0454
	$P_A^{(1)}$	0.6729	0.5429	0.2432	0.0617	0.0044
	$P_A^{(2)}$	0.0486	0.1143	0.4054	0.7654	1.0000
	$P_A^{(3)}$	0.9654	0.9695	0.9550	0.9044	0.8111
$n = 1000$	b	140	100	50	25	10
	R_{An}	0.0035	0.0029	0.0022	0.0029	0.0112
	$P_A^{(1)}$	0.7818	0.6551	0.4091	0.1724	0.0179
	$P_A^{(2)}$	0.0129	0.0345	0.1364	0.5172	0.8839
	$P_A^{(3)}$	0.9923	0.9580	0.9640	0.9389	0.8830
$n = 3000$	b	420	300	150	75	30
	R_{An}	0.0032	0.0024	0.0015	0.0011	0.0017
	$P_A^{(1)}$	0.8550	0.7910	0.5997	0.4194	0.1079
	$P_A^{(2)}$	0.0015	0.0040	0.0252	0.1436	0.5942
	$P_A^{(3)}$	0.9553	0.9598	0.9209	0.9940	0.9641
$n = 6000$	b	840	600	300	150	60
	R_{An}	0.0031	0.0023	0.0013	0.0008	0.0008
	$P_A^{(1)}$	0.8979	0.8313	0.7151	0.5316	0.2298
	$P_A^{(2)}$	0.0004	0.0010	0.0074	0.0451	0.3105
	$P_A^{(3)}$	0.9555	0.9442	0.9677	0.9628	0.9605
$n = 10000$	b	1400	1000	500	250	100
	R_{An}	0.0030	0.0022	0.0011	0.0007	0.0005
	$P_A^{(1)}$	0.9000	0.8636	0.8545	0.6571	0.3600
	$P_A^{(2)}$	0.0002	0.0040	0.0031	0.0202	0.1800
	$P_A^{(3)}$	0.9506	0.9608	0.9217	0.9428	0.9260

Table 4: *The values of* R_{Bn} *and* $P_B^{(i)}, i = 1, 2, 3$, *under the* Beta(1.5, 1) *noise.* $P_B^{(1)}$ *(*$P_B^{(2)}$, $P_B^{(3)}$*) is colored in red (blue, green) when it is equal to* $\max_{1 \le i \le 3} P_B^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{Bn}	0.1086	0.1028	0.1167	0.1584	0.2170
	$P_B^{(1)}$	0.7920	0.7498	0.8512	0.8655	0.6318
	$P_B^{(2)}$	0.7366	0.8327	0.9443	0.9836	0.9894
	$P_B^{(3)}$	0.8030	0.7737	0.6142	0.4379	0.3165
$n = 200$	b	28	20	10	5.0	2.0
	R_{Bn}	0.1084	0.1007	0.0919	0.1130	0.1688
	$P_B^{(1)}$	0.7901	0.7345	0.6703	0.8242	0.8122
	$P_B^{(2)}$	0.6815	0.7557	0.9314	0.9752	0.9819
	$P_B^{(3)}$	0.9029	0.8687	0.8110	0.6209	0.4077
$n = 500$	b	70	50	25	12.5	5.0
	R_{Bn}	0.1192	0.1096	0.0860	0.0868	0.1112
	$P_B^{(1)}$	0.8692	0.7994	0.6273	0.6331	0.8111
	$P_B^{(2)}$	0.5921	0.6515	0.8663	0.9401	0.9910
	$P_B^{(3)}$	0.9498	0.9308	0.9506	0.8343	0.6224
$n = 1000$	b	140	100	50	25	10
	R_{Bn}	0.1291	0.1209	0.0976	0.0835	0.0899
	$P_B^{(1)}$	0.9411	0.8818	0.7119	0.6090	0.6562
	$P_B^{(2)}$	0.5390	0.5782	0.7316	0.8922	0.9522
	$P_B^{(3)}$	0.9500	0.9390	0.9292	0.9085	0.7770
$n = 3000$	b	420	300	150	75	30
	R_{Bn}	0.1323	0.1257	0.1133	0.0900	0.0761
	$P_B^{(1)}$	0.9647	0.9166	0.8264	0.6564	0.5550
	$P_B^{(2)}$	0.5201	0.5488	0.6125	0.7822	0.9645
	$P_B^{(3)}$	0.9939	0.9904	0.9612	0.9624	0.9497
$n = 6000$	b	840	600	300	150	60
	R_{Bn}	0.1329	0.1306	0.1214	0.0977	0.0766
	$P_B^{(1)}$	0.9691	0.9523	0.8854	0.7126	0.5587
	$P_B^{(2)}$	0.5172	0.5267	0.5683	0.7103	0.9255
	$P_B^{(3)}$	0.9903	0.9932	0.9832	0.9942	0.9870
$n = 10000$	b	1400	1000	500	250	100
	R_{Bn}	0.1340	0.1392	0.1356	0.1096	0.0804
	$P_B^{(1)}$	0.9771	0.9849	0.9890	0.7873	0.5864
	$P_B^{(2)}$	0.5126	0.4935	0.5073	0.6304	0.8694
	$P_B^{(3)}$	0.9899	0.9606	0.9236	0.9656	0.9893

Table 5: *The values of* R_{Bn} *and* $P_B^{(i)}, i = 1,2,3$, *under the* Beta(2,1) *noise.* $P_B^{(1)}$ *(*$P_B^{(2)}$*,* $P_B^{(3)}$*) is colored in red (blue, green) when it is equal to* $\max_{1\le i\le 3} P_B^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{Bn}	0.0824	0.0793	0.0956	0.1285	0.1736
	$P_B^{(1)}$	0.7414	0.7138	0.8605	0.8646	0.6399
	$P_B^{(2)}$	0.7868	0.8752	0.9341	0.9821	0.9983
	$P_B^{(3)}$	0.8068	0.7784	0.5987	0.4356	0.3204
$n = 200$	b	28	20	10	5.0	2.0
	R_{Bn}	0.0828	0.0757	0.0728	0.0891	0.1370
	$P_B^{(1)}$	0.7454	0.6814	0.6553	0.8020	0.8109
	$P_B^{(2)}$	0.7223	0.8151	0.9533	0.9978	0.9835
	$P_B^{(3)}$	0.8916	0.8806	0.8080	0.6331	0.4065
$n = 500$	b	70	50	25	12.5	5.0
	R_{Bn}	0.0920	0.0824	0.0666	0.0694	0.0880
	$P_B^{(1)}$	0.8282	0.7417	0.5995	0.6247	0.7921
	$P_B^{(2)}$	0.6215	0.7015	0.9054	0.9524	0.9854
	$P_B^{(3)}$	0.9363	0.9335	0.9468	0.8306	0.6352
$n = 1000$	b	140	100	50	25	10
	R_{Bn}	0.0942	0.0900	0.0719	0.0651	0.0716
	$P_B^{(1)}$	0.8478	0.8101	0.6472	0.5860	0.6445
	$P_B^{(2)}$	0.5983	0.6289	0.8039	0.9263	0.9693
	$P_B^{(3)}$	0.9914	0.9602	0.9566	0.9166	0.7854
$n = 3000$	b	420	300	150	75	30
	R_{Bn}	0.1061	0.1027	0.0888	0.0699	0.0600
	$P_B^{(1)}$	0.9550	0.9249	0.7994	0.6291	0.5406
	$P_B^{(2)}$	0.5260	0.5441	0.6338	0.8165	0.9908
	$P_B^{(3)}$	0.9843	0.9673	0.9280	0.9456	0.9582
$n = 6000$	b	840	600	300	150	60
	R_{Bn}	0.1164	0.1041	0.0924	0.0739	0.0620
	$P_B^{(1)}$	0.9538	0.9375	0.8318	0.6656	0.5588
	$P_B^{(2)}$	0.4780	0.5350	0.6050	0.7612	0.9256
	$P_B^{(3)}$	0.9234	0.9961	0.9930	0.9908	0.9571
$n = 10000$	b	1400	1000	500	250	100
	R_{Bn}	0.1063	0.1059	0.0994	0.0805	0.0653
	$P_B^{(1)}$	0.9567	0.9531	0.8946	0.7245	0.5877
	$P_B^{(2)}$	0.5189	0.5250	0.5603	0.6956	0.8667
	$P_B^{(3)}$	0.9758	0.9897	0.9823	0.9928	0.9453

Table 6: *The values of* R_{Bn} *and* $P_B^{(i)}, i = 1, 2, 3,$ *under the* Beta(4, 1) *noise.* $P_B^{(1)}$ *(*$P_B^{(2)}$*,* $P_B^{(3)}$*) is colored in red (blue, green) when it is equal to* $\max_{1\le i\le 3} P_B^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{Bn}	0.0348	0.0343	0.0444	0.0625	0.0854
	$P_B^{(1)}$	0.6526	0.6435	0.8330	0.8528	0.6241
	$P_B^{(2)}$	0.8939	0.9708	0.9640	0.9696	0.9778
	$P_B^{(3)}$	0.8241	0.8085	0.6067	0.4277	0.3123
$n = 200$	b	28	20	10	5.0	2.0
	R_{Bn}	0.0340	0.0335	0.0336	0.0431	0.0666
	$P_B^{(1)}$	0.6371	0.6285	0.6304	0.8086	0.8003
	$P_B^{(2)}$	0.8452	0.8836	0.9911	0.9930	0.9970
	$P_B^{(3)}$	0.8944	0.8572	0.8098	0.6219	0.4007
$n = 500$	b	70	50	25	12.5	5.0
	R_{Bn}	0.0370	0.0336	0.0309	0.0322	0.0425
	$P_B^{(1)}$	0.6929	0.6304	0.5797	0.6041	0.7974
	$P_B^{(2)}$	0.7429	0.8244	0.9353	0.9844	1.0000
	$P_B^{(3)}$	0.9296	0.9304	0.9057	0.8388	0.6287
$n = 1000$	b	140	100	50	25	10
	R_{Bn}	0.0407	0.0376	0.0303	0.0303	0.0344
	$P_B^{(1)}$	0.7638	0.7054	0.5685	0.5685	0.6454
	$P_B^{(2)}$	0.6641	0.7234	0.9142	0.9538	0.9680
	$P_B^{(3)}$	0.9497	0.9178	0.9631	0.9024	0.7784
$n = 3000$	b	420	300	150	75	30
	R_{Bn}	0.0462	0.0439	0.0341	0.0295	0.0286
	$P_B^{(1)}$	0.8674	0.8239	0.6394	0.5549	0.5374
	$P_B^{(2)}$	0.5791	0.6109	0.7924	0.9256	0.9967
	$P_B^{(3)}$	0.9857	0.9439	0.9680	0.9662	0.9417
$n = 6000$	b	840	600	300	150	60
	R_{Bn}	0.0483	0.0480	0.0383	0.0313	0.0283
	$P_B^{(1)}$	0.9074	0.9003	0.7197	0.5876	0.5318
	$P_B^{(2)}$	0.5523	0.5572	0.6993	0.8622	0.9725
	$P_B^{(3)}$	0.9914	0.9520	0.9618	0.9657	0.9630
$n = 10000$	b	1400	1000	500	250	100
	R_{Bn}	0.0513	0.0495	0.0410	0.0323	0.0281
	$P_B^{(1)}$	0.9624	0.9287	0.7692	0.6060	0.5272
	$P_B^{(2)}$	0.5205	0.5397	0.6529	0.8321	0.9679
	$P_B^{(3)}$	0.9829	0.9731	0.9822	0.9961	0.9869

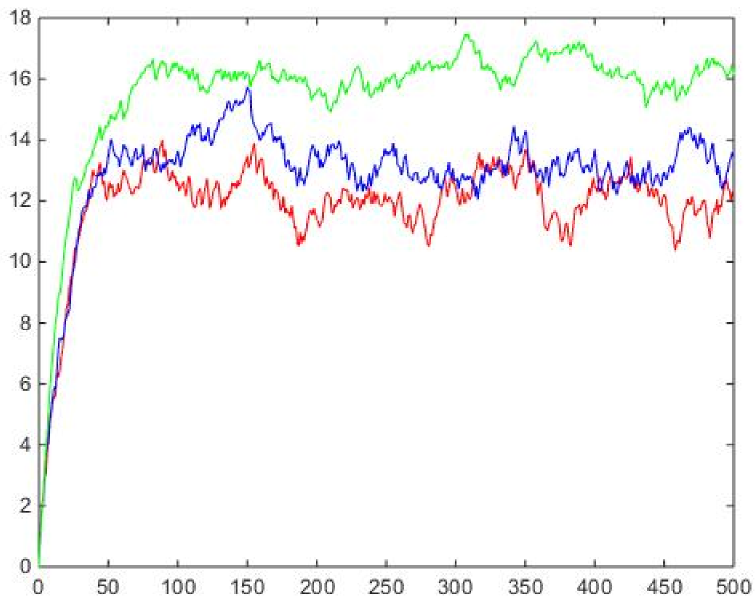

Figure 3: Plots of time series realizations generated from model (1.1), where $\rho = 0.95$ and ε_t has a Beta(α, 1) distribution, with $\alpha = 4$ (green line), 2 (blue line) and 1.5 (red line).

$\square$

4 Concluding Remarks

By deriving asymptotic expressions for the MSPEs of the EV and LS predictors under near unit-root and general near unit-root models, this paper provides a deeper appreciation for the performance of the LS and EV predictors in the near unit-root region. In particular, our analysis reveals that the expressions derived for the LS predictor for the critical points, $\beta = 1$ and $\beta = 1/2$, not only jointly connect the discontinuities in $\lim_{n\to\infty} n(\text{MSPE}_B - \sigma^2)$, but also combine their strengths to yield finite sample approximations of $n(\text{MSPE}_B - \sigma^2)$ that perform satisfactorily in the near unit-root region. Moreover, the expressions derived for the EV predictor for $\beta = 1$ and $0 < \beta < 1$ also lead to finite sample approximations of $n(\text{MSPE}_A - \sigma^2)$ that can achieve a similar goal. Finally, we mention that the results established in this paper and in Ing and Yang (2014) can be unified as follows:

$$\begin{aligned}
&\text{MSPE}_A - \sigma^2 \\
&= n^{-2/\alpha}\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \Big\{ \Big[\frac{\sigma^2 b}{2\mu^2 n^\beta} + \frac{1}{n^2(b/n^\beta)^2}\Big] I_{\{0<\beta<1, 0<b<\infty\}} \\
&+ \Big(\frac{1}{M'_{\alpha,b}}\Big)^{2/\alpha} L_3(b) I_{\{\beta=1, 0<b<\infty\}} + \Big(\frac{1}{M_\alpha(1-b)}\Big)^{2/\alpha} \frac{\sigma^2}{b(2-b)} I_{\{\beta=0, 0<b\le 1\}} + \frac{(\alpha+1)^{2/\alpha}}{4} I_{\{\beta=1, b=0\}} \Big\} \qquad (4.19) \\
&+ \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{-\beta-\frac{2}{\alpha}}, n^{2\beta-2-\frac{2}{\alpha}}\}),
\end{aligned}$$

and

$$\begin{aligned} &\text{MSPE}_B - \sigma^2 \\ &= n^{-1}\left\{\Lambda_1(\beta,b)I_{\{0<\beta\le 1,0<b<\infty\}} + 2\sigma^2 I_{\{\beta=0,0<b\le 1\}} + 4\sigma^2 I_{\{\beta=1,b=0\}}\right\} + o(n^{-1}). \end{aligned} \tag{4.20}$$

Equations (4.19) and (4.20) provide a more comprehensive perspective on the performance of the EV and LS predictors and may facilitate broader applications.

Acknowledgments

We would like to thank the associate editor and two anonymous reviewers for their insightful and constructive comments, which greatly improve the presentation of this paper. Ing's research was supported by the Science Vanguard Research Program, Ministry of Science and Technology, Taiwan. Zhang's research was supported by the National Natural Science Foundation of China under Grant No. 11371318, Zhejiang Provincial Natural Science Foundation of China under Grant No. R16A010001, and the Fundamental Research Funds for the Central Universities, Ministry of Education, China.

References

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

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

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **63**, 167–241.

Bell, C. B. and Smith, E. P. (1986). Inference for non-negative autoregressive schemes. *Communications in Statistics: Theory and Methods*, **15**, 2267–2293.

Chan, N. H. (2006). Inference for time series and stochastic processes. *Statistica Sinica* **16**, 683–696.

Chan, N. H. (2009). Time series with roots on or near the unit circle. In Andersen, T.G., Davis, R.A., Kreiss, J.P. and Mikosch, T. (eds.), *Handbook of Financial Time Series*, 695–707. Springer-Verlag, New York.

Chan, N. H. and Ing, C.-K. (2011). Uniform moment bounds of Fisher's information with applications to time series. *Annals of Statistics*, **39**, 1526-1550.

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

Datta, D. and McCormick, W. P. (1995). Bootstrap inference for a first-order autoregression with positive innovations. *Journal of the American Statistical Association*, **90**, 1289–1300.

Davis, R. A. and McCormick, W. P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Processes and Their Applications*, **31**, 237–250.

Findley, D. F. and Wei, C.-Z. (2002). AIC, overfitting principles, and the boundedness of moments of inverse matrices for vector autoregressions and related models. *Journal of Multivariate Analysis*, **83**, 415V-450.

Finkenstädt, B. and Rootzén, H. (eds.) (2004). *Extreme Values in Finance, Telecommunications and the Environment*. Chapman and Hall, New York.

Gaver, D. P. and Lewis, P. A. W. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability*, **12**, 727–745.

Hsiao, W.-C., Huang, H.-Y. and Ing, C.-K. (2017). Estimation and prediction for positive autoregressive processes. Technical Report, Institute of Statistics, National Tsing Hua University.

Ing, C.-K. and Yang, C.-Y. (2014). Predictor selection for positive autoregressive processes. *Journal of the American Statistical Association*, **109**, 243–253.

Lawrance, A. J. and Lewis, P. A. W. (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables (with Discussion). *Journal of the Royal Statistical Society, Ser. B*, **47**, 165–202.

Marron, J. S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society, Ser. B*, **56**, 653–671.

Nielsen, B. and Shephard, N. (2003). Likelihood analysis of a first-order autoregressive model with exponential innovations. *Journal of Time Series Analysis*, **24**, 337–344.

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

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

Sarlak, N. (2008). Annual streamflow modelling with asymmetric distribution function. *Hydrological Processes*, **22**, 3403–3409.

Smith, R. L. (1994). Nonregular regression. *Biometrika*, **81**, 173–243.

Wei, C.-Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with application to time series. *Annals of Statistics*, **15**, 1667–1682.

Wei, C.-Z. (1992). On predictive least squares principles. *Annals of Statistics*, **20**, 1–42.

Whittle, P. (1963). *Prediction and Regulation: by Linear Least-Square Methods.* Van-Nostrand, Princeton, New Jersey.

Yu, S.-H., Lin, C.-C. and Cheng, H.-W. (2012). A note on mean squared prediction error under the unit root model with deterministic trend. *Journal of Time Series Analysis*, **33**, 276–286.

Supplement to "Nearly Unstable Processes: A Prediction Perspective"

by Ngai Hang Chan, Ching-Kang Ing, and Romgmao Zhang

A Proofs of the Theorems in Sections 2 and 3.1

Before proceeding with the proofs, we would like to first point out the key difference between our asymptotic framework and those in Chan and Wei (1987), Phillips (1987) and Phillips and Magdalinos (2007). Let $\rho_n = 1 - b/n^{\beta}$, with $0 < \beta \leq 1$ and $0 < b < \infty$, and $y_t^* = y_t - n^{\beta}\mu/b$. Then, model (1.1) can be expressed as

$$y_t^* = \rho_n y_{t-1}^* + \eta_t, \tag{A.1}$$

which is an AR(1) model driven by a zero-mean white-noise process $\{\eta_t\}$. While the asymptotic behavior of the LSE under (A.1) has been extensively studied by the aforementioned authors, their results rely heavily on the initial condition $y_0^* = O_p(1)$, which is obviously violated by our initial condition $y_0 = 0$, leading to

$$y_0^* = -n^{\beta}\mu/b. \tag{A.2}$$

With the initial condition like (A.2), most existing results established for the LSE under model (A.1) are no longer applicable. As shown for the rest of this section, our asymptotic analysis is similar to that adopted by Ing and Yang (2014). However, substantial efforts are needed to deal with the critical behavior of the EV and LS predictors exhibited in the near unit-root region.

Proof of Theorem 2.1. We first prove (2.6). By $\exp(-b)\sum_{j=1}^{t}\varepsilon_j \leq y_t = \sum_{j=0}^{t-1}\rho_n^j\varepsilon_{t-j} \leq \sum_{j=1}^{t}\varepsilon_j$ and (19) of Ing and Yang (2014), which shows that for any $q > 0$,

$$\mathrm{E}\left\{n^{1+1/\alpha}\min_{2\leq t\leq n}\left(\varepsilon_t/\sum_{j=1}^{t-1}\varepsilon_j\right)\right\}^q < \infty, \tag{A.1}$$

the desired conclusion (2.6) follows. The proof of (2.7) is similar to that of (7) of Ing and Yang (2014). The details are omitted. Finally, (2.8) follows directly from (2.6) and (2.7). □

Proof of Theorem 3.1. Equation (3.1) can be shown by an argument similar to that used to prove (2.6). We thus skip the details. For (3.3), it suffices to show that

$$n^{1/\alpha+\beta}\min_{\nu_n\leq i\leq n}(\varepsilon_i/y_{i-1}) - n^{1/\alpha+\beta}\min_{2\leq i\leq n}(\varepsilon_i/y_{i-1}) = o_p(1) \tag{A.2}$$

and for any $t > 0$,

$$\lim_{n\to\infty} P\{(c/\alpha)^{1/\alpha}(\mu/b)n^{1/\alpha+\beta}\min_{\nu_n\leq i\leq n}(\varepsilon_i/y_{i-1}) > t\} = \exp\{-t^{\alpha}\}, \tag{A.3}$$

where $v_n \asymp n^{\theta}$ for some $\beta < \theta < 1$.

To show (A.2), note first that

$$n^{1/\alpha+\beta} \min_{v_n \le i \le n} (\varepsilon_i/y_{i-1}) - n^{1/\alpha+\beta} \min_{2\le i\le n} (\varepsilon_i/y_{i-1}) \le n^{1/\alpha+\beta} (\min_{\nu_n \le i \le n} \varepsilon_i/y_{i-1}) I_{A_n},$$

where $A_n = \{\min_{\nu_n \le i \le n} \varepsilon_i/y_{i-1} > \min_{2\le i\le \nu_n} \varepsilon_i/y_{i-1}\}$. Let $q_n = s_n^{1/2} n^{\beta}/\nu_n^{\beta}$ in which s_n satisfies $s_n \nu_n^{1/\alpha}/n^{1/\alpha} = o(1)$ and $s_n \to \infty$. Then, by (3.1), (1.2), the weak law of large number and Chebyshev's inequality, one has for any $\epsilon > 0$,

$$\begin{aligned}
& P(n^{1/\alpha+\beta}(\min_{\nu_n \le i \le n} \varepsilon_i/y_{i-1}) I_{A_n} > \epsilon) \le P(A_n) \\
\le\ & P(\min_{\nu_n \le i \le n} \varepsilon_i/y_{i-1} > s_n^{-1/2} q_n^{-1} \nu_n^{-\beta-(1/\alpha)}) + P(\max_{2\le i\le \nu_n} y_{i-1} \ge q_n \nu_n^{\beta}) \\
+\ & P(\min_{2\le i \le \nu_n} \varepsilon_i < s_n^{-1/2} \nu_n^{-1/\alpha}) \\
=\ & O\left(\frac{s_n^{1/2} q_n \nu_n^{1/\alpha+\beta}}{n^{1/\alpha+\beta}} \right) + o(1) + 1 - \left(1 - \frac{C}{\nu_n s_n^{\alpha/2}}\right)^{\nu_n} = o(1),
\end{aligned}$$

where C is some positive constant independent of n. Thus, (A.2) is proved.

To show (A.3), one can use (3.2) and Lemma 2 of Wei (1987) to obtain

$$\max_{\nu_n \le i \le n} |y_{i-1} - \mathrm{E} y_{i-1}| = O_p(n^{1/q_1+\beta/2}). \tag{A.4}$$

In addition, there exists $c_1 > 0$ such that for all $\nu_n \le i \le n$,

$$\mathrm{E} y_{i-1} = \mu n^{\beta}(1-\rho^i)/b \ge c_1 n^{\beta}, \tag{A.5}$$

By (A.4), (A.5), (1.2) and $1/q_1 + \beta/2 < \beta$, it holds that

$$\begin{aligned}
& \lim_{n\to\infty} P\{(c/\alpha)^{1/\alpha} (\mu/b) n^{1/\alpha+\beta} \min_{\nu_n \le i \le n} (\varepsilon_i/y_{i-1}) > t\} \\
&= \lim_{n\to\infty} P\left\{ (\mu/b)(c/\alpha)^{1/\alpha} n^{1/\alpha+\beta} \min_{\nu_n\le i\le n} \left(\frac{\varepsilon_i}{\mu n^{\beta}(1-\rho^i)/b} \right) > t \right\} = \exp\{-t^{\alpha}\},
\end{aligned}$$

which completes the proof of (A.3). Finally, (3.4) is a immediate consequence of (3.1) and (3.3). □

Proofs of Theorems 2.2 and 3.2. Define

$$a_{n,i} = \left(\frac{1-\rho_n^i}{1-\rho_n} - \frac{1}{(n-1)} \sum_{j=1}^{n-1} \frac{1-\rho_n^{j-1}}{1-\rho_n} \right) \mu,$$

and $\xi_{n,i} = \sum_{j=0}^{i-1} \rho^j \eta_{i-j} - (n-1)^{-1} \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \rho^j \eta_{i-j}$. Then

$$\begin{aligned}
\tilde{\rho}_n - \rho_n &= \sum_{i=2}^{n} (y_{i-1} - \bar{y}) \eta_i / \sum_{i=2}^{n} (y_{i-1} - \bar{y})^2 \\
&= \sum_{i=2}^{n} (a_{n,i-1} + \xi_{n,i-1}) \eta_i / \sum_{i=2}^{n} (a_{n,i-1} + \xi_{n,i-1})^2.
\end{aligned} \tag{A.6}$$

Straightforward calculations yield for $0 < \beta \leq 1$,

$$\lim_{n\to\infty} \frac{1}{n^{3\beta}} \sum_{i=2}^{n} a_{n,i-1}^2 = \frac{\mu^2}{2b^3} I(0 < \beta < 1) + \mu^2 (I_1(b) - I_2(b)) I(\beta = 1), \tag{A.7}$$

and

$$\sum_{i=2}^{n} \mathrm{E}\xi_{n,i-1}^2 = \sigma^2 (2b)^{-1} n^{1+\beta} (1 - \exp(-bn^{1-\beta}))(1 + o(1)). \tag{A.8}$$

In addition, for $0 < \beta < 1$, we have

$$\frac{1}{n^{1+\beta}} \sum_{i=2}^{n} \xi_{n,i-1}^2 - \frac{1}{n^{1+\beta}} \sum_{i=2}^{n} (\sum_{j=0}^{i-1} \rho^j \eta_{i-j})^2 = o_p(1), \tag{A.9}$$

and

$$\frac{1}{n^{1+\beta}} \sum_{i=2}^{n} (\sum_{j=0}^{i-1} \rho^j \eta_{i-j})^2 \xrightarrow{p} \frac{\sigma^2}{2b}. \tag{A.10}$$

Moreover, for $\beta = 1/2$,

$$\frac{1}{n^{3/2}} \sum_{i=2}^{n} (a_{n,i-1} + \xi_{n,i-1})^2 = \frac{1}{n^{3/2}} \sum_{i=2}^{n} (a_{n,i-1}^2 + \xi_{n,i-1}^2) + o_p(1) = \frac{\sigma^2}{2b} + \frac{\mu^2}{2b^3} + o_p(1). \tag{A.11}$$

By (A.6)–(A.11) and the martingale central limit theorem (see, e.g., Theorem 3.2 of Phillips and Magdalinos (2007)), (2.10) and (3.5) follow.

Set $k_n = n^3$ for $\beta = 1$. By (2.11), $\mathrm{E}\varepsilon_1^s < \infty$ for some $s > 10$, and an argument similar to that used to prove Lemma 2 of Yu, Lin and Cheng (2012), we have $\mathrm{E}|\sqrt{k_n}(\tilde{\rho}_n - \rho_n)|^\gamma < \infty$ for some $\gamma > 2$, and hence $\{k_n(\tilde{\rho}_n - \rho_n)^2\}$ is uniformly integrable. This together with (2.10) (resp. (3.5)) yields (2.12) (resp. (3.6)). □

Proofs of Theorems 2.3 and 3.3. It follows from (2.4) and $\mathrm{E}[(1 - \rho_n)\bar{y} - \mu + z_n]^2 = \mathrm{E}[(n - 1)^{-1} \sum_{j=2}^{n} \eta_j]^2 = \sigma^2/(n-1)$ that

$$\begin{aligned} \mathrm{MSPE}_A - \sigma^2 &= \sigma^2/(n-1) + \mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})\}^2 \\ &+ 2\mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})[((1 - \rho_n)\bar{y} - \mu) + z_n]\} \end{aligned} \tag{A.12}$$

To deal with the second term on the right-hand side of (A.12), we express $y_n - \bar{y}$ as

$$\begin{aligned} y_n - \bar{y} &= \left(\sum_{j=0}^{n-1} \rho^j - \frac{1}{n-1} \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \rho^j \right) \mu + \left(\sum_{j=0}^{n-1} \rho^j \eta_{n-j} - \frac{1}{n-1} \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \rho^j \eta_{i-j} \right) \\ &:= X_{1,n} + X_{2,n}. \end{aligned} \tag{A.13}$$

Some algebraic manipulations give

$$\lim_{n\to\infty} \frac{X_{1,n}^2}{n^{4\beta-2}} = \frac{\mu^2}{b^4} I(0 < \beta < 1) + \mu^2 L_3(b) I(\beta = 1) \tag{A.14}$$

and

$$\mathrm{E}X_{2,n}^2 = \frac{\sigma^2}{2b} n^\beta (1 + o(1)) I(0 < \beta < 1) + O(n) I(\beta = 1). \tag{A.15}$$

Combining (A.13)–(A.15) yields for $2/3 < \beta \leq 1$,

$$\frac{(y_n - \bar{y})^2}{n^{4\beta - 2}} \xrightarrow{p} \frac{\mu^2}{b^4} I(2/3 < \beta < 1) + \mu^2 L_3(b) I(\beta = 1), \tag{A.16}$$

and for $0 < \beta \leq 2/3$,

$$\lim_{n\to\infty} n^{-\beta} \mathrm{E}(y_n - \bar{y})^2 = \frac{\sigma^2}{2b} I(0 < \beta < 2/3) + \frac{\mu^2}{b^4} I(\beta = 2/3). \tag{A.17}$$

By the moment conditions on ε_1 and a straightforward calculation, it follows that for $0 < \beta \leq 2/3$ there exists $2/3 < \zeta < 1$ for which

$$n^{-\beta/2}(y_n - \bar{y}) = n^{-\beta/2}(X_{1,n} + \sum_{j=0}^{n^\zeta - 1} \rho^j \eta_{n-j}) + r_{1,n},$$

where $r_{1,n}$ satisfies $\mathrm{E}|r_{1,n}|^{q_1} = o(1)$ for some $q_1 > 2$, and

$$\lim_{n\to\infty} \mathrm{E}[n^{-\beta}(X_{1,n} + \sum_{j=0}^{n^\zeta - 1} \rho^j \eta_{n-j})^2] = \lim_{n\to\infty} \mathrm{E}[n^{-\beta}(y_n - \bar{y})^2].$$

In addition, by (3.1) and an argument similar to that used to prove (A.2), we obtain

$$n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n) = n^{\beta+1/\alpha}\big(\min_{2\leq i\leq n-n^\zeta} \frac{\varepsilon_i}{y_{i-1}}\big) + r_{2,n},$$

where $r_{2,n}$ satisfies $\mathrm{E}|r_{2,n}|^q = o(1)$ for any $q > 0$, and

$$\lim_{n\to\infty} \mathrm{E}\big[n^{\beta+1/\alpha}\big(\min_{2\leq i\leq n-n^\zeta} \frac{\varepsilon_i}{y_{i-1}}\big)\big]^2 = \lim_{n\to\infty} \mathrm{E}[n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n)]^2.$$

These facts and the independence between $n^{-\beta/2}\sum_{j=0}^{n^\zeta-1} \rho^j \eta_{n-j}$ and $n^{\beta+1/\alpha}(\min_{2\leq i\leq n-n^\zeta} \varepsilon_i/y_{i-1})$ yield for $0 < \beta \leq 2/3$

$$\begin{aligned} &\lim_{n\to\infty} \mathrm{E}\{n^{-\beta}(y_n - \bar{y})^2[n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n)]^2\} \\ &= \lim_{n\to\infty} \mathrm{E}\{n^{-\beta}(y_n - \bar{y})^2\} \lim_{n\to\infty} \mathrm{E}\{[n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n)]^2\}. \end{aligned} \tag{A.18}$$

Now, by (A.16)–(A.18), (2.6), (2.8), (3.1), (3.4) and the moment conditions imposed on ε_1, it holds that for $2/3 < \beta \leq 1$,

$$\begin{aligned} &\lim_{n\to\infty} \mathrm{E}[n^{-\beta+1/\alpha+1}(\hat{\rho}_n - \rho_n)(y_n - \bar{y})]^2 \\ &= \Gamma\big(\frac{\alpha+2}{\alpha}\big)\big(\frac{\alpha}{c}\big)^{2/\alpha} b^{-2} I(2/3 < \beta < 1) + \Gamma\big(\frac{\alpha+2}{\alpha}\big)\big(\frac{\alpha}{cM'_{\alpha,b}}\big)^{2/\alpha} L_3(b) I(\beta = 1), \end{aligned} \tag{A.19}$$

and for $0 < \beta \le 2/3$,

$$\lim_{n\to\infty} \mathrm{E}[n^{1/\alpha+\beta/2}(\hat\rho_n - \rho_n)(y_n - \bar y)]^2 = \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{2/\alpha}\left\{\frac{\sigma^2 b}{2\mu^2} I(0<\beta\le 2/3) + \frac{1}{b^2} I(\beta = 2/3)\right\}. \tag{A.20}$$

To deal with the third term on the right-hand side of (A.12), we obtain from an argument similar to that used to prove (2.9) in the supplementary document for Ing and Yang (2014) that for $2/3 < \beta \le 1$,

$$\mathrm{E}\{(\hat\rho_n - \rho_n)(y_n - \bar y)[((1-\rho_n)\bar y - \mu) + z_n]\} = o(\max\{n^{-1}, n^{2\beta-2-2/\alpha}\}), \tag{A.21}$$

and for $0 < \beta \le 2/3$,

$$\mathrm{E}\{(\hat\rho_n - \rho_n)(y_n - \bar y)[((1-\rho_n)\bar y - \mu) + z_n]\} = o(\max\{n^{-1}, n^{-\beta-2/\alpha}\}). \tag{A.22}$$

Consequently, the desired conclusions (2.13) and (3.7)–(3.9) are ensured by (A.12), and (A.19)–(A.22). □

Proofs of Theorems 2.4 and 3.4. It follows from (2.5) that

$$\begin{aligned} &\mathrm{E}(y_{n+1} - \tilde y_{n+1})^2 - \sigma^2 \\ = \;& \frac{\sigma^2}{n-1} + \mathrm{E}\{(y_n - \bar y)(\tilde\rho_n - \rho_n)\}^2 + \frac{2}{n-1}\mathrm{E}\{(\sum_{i=2}^n \eta_i)(y_n - \bar y)(\tilde\rho_n - \rho_n)\}. \end{aligned} \tag{A.23}$$

By (A.7)–(A.11), Theorems 2.2 and 3.2, (2.11), $\mathrm{E}\varepsilon_1^s < \infty$ for some $s > 12$, and an argument similar to that used to prove Lemma 2 of Yu, Lin and Cheng (2012), we obtain, after some tedious algebraic manipulations,

$$\mathrm{E}\{(\sum_{i=2}^n \eta_i)(y_n - \bar y)(\tilde\rho_n - \rho_n)\} = o(1), \tag{A.24}$$

and

$$\lim_{n\to\infty} \frac{n}{\sigma^2}\mathrm{E}\{(y_n - \bar y)(\tilde\rho_n - \rho_n)\}^2 = \begin{cases} 1 & 0<\beta<1/2\,, \\ \frac{b^2\sigma^2}{\mu^2+b^2\sigma^2} & \beta = 1/2\,, \\ 0 & 1/2<\beta<1\,, \\ \frac{L_3(b)}{I_1(b)-I_2(b)} & \beta = 1. \end{cases} \tag{A.25}$$

Consequently, Theorems 2.4 and 3.4 are guaranteed by (A.23)–(A.25).

B The implementation of finite sample approximations

B.1 Rules of thumb developed from Tables 1–6

With the help of Tables 1–3 (Tables 4–6), we offer a simple rule for choosing a better approximation of $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ ($n(\mathrm{MSPE}_B - \sigma^2)$) from $R_A^{(2)}$ and $R_A^{(3)}$ ($R_B^{(2)}$ and $R_B^{(3)}$) when $100 \le n \le 1000$,

$1 \leq n(1-\rho) = b \leq 140$ and $1.5 \leq \alpha \leq 4$. According to Tables 1–3, we first introduce Rule I for approximating $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$:

Rule I.

1. Choose $R_A^{(3)}$ if $1.5 \leq \alpha < 2$.
2. Choose $R_A^{(2)}$ if $2 \leq \alpha \leq 4$ and $1 \leq b \leq 5$.
3. Choose $R_A^{(3)}$ if $2 \leq \alpha \leq 4$ and $5 < b \leq 140$.

Although there are a few cases where Rule I leads to a $P_A^{(i)}$ (defined in Section 3.2) slightly smaller than $\max_{1\leq i\leq 3} P_A^{(i)}$, the rule has the advantage of easy implementation, which is practically appealing. In the same spirit, we propose using Rule II (according to Tables 4–6) for approximating $n(\text{MSPE}_B - \sigma^2)$:

Rule II.

1. Choose $R_B^{(2)}$ if $1.5 \leq \alpha \leq 2$ and $1 \leq b \leq 12.5$.
2. Choose $R_B^{(3)}$ if $1.5 \leq \alpha \leq 2$ and $12.5 < b \leq 140$.
3. Choose $R_B^{(2)}$ if $2 < \alpha \leq 4$ and $1 \leq b \leq 25$.
4. choose $R_B^{(3)}$ if $2 < \alpha \leq 4$ and $25 < b \leq 140$.

Note that Rules I and II can be further refined by checking a more dense grid of n, b and α, which is not pursued here. In Section B.2, we provide reliable estimators, $\hat{\rho}_n^*$, $\hat{b}_n^*$, $\hat{\alpha}_n^*$, $\hat{c}_n^*$, $\hat{\mu}_n^*$ and $\hat{\sigma}_n^{2^*}$, of ρ, b, α, c, μ and σ^2. With these estimators, Rules I and II can be implemented in practice via replacing b, α, $R_A^{(i)}, i=2,3$ and $R_B^{(i)}, i=2,3$ therein by $\hat{b}_n^*$, $\hat{\alpha}_n^*$, $\hat{R}_A^{(i)}, i=2,3$ and $\hat{R}_B^{(i)}, i=2,3$, where

$$\hat{R}_A^{(2)} = \begin{cases} \Gamma\left(\frac{\hat{\alpha}_n^*+2}{\hat{\alpha}_n^*}\right)\left(\frac{\hat{\alpha}_n^*}{\hat{c}_n^* \hat{M}'_{\hat{\alpha}_n^*,\hat{b}_n^*}}\right)^{2/\hat{\alpha}_n^*} L_3(\hat{b}_n^*) + {\hat{\sigma}_n^*}^2 I(\hat{\alpha}_n^* = 2) & \hat{\alpha}_n^* \geq 2\,, \\ {\hat{\sigma}_n^*}^2 & \hat{\alpha}_n^* < 2, \end{cases} \tag{B.1}$$

in which $\hat{M}'_{\hat{\alpha}_n^*,\hat{b}_n^*}$ is $M'_{\alpha,b}$ with α and b replaced by $\hat{\alpha}_n^*$ and $\hat{b}_n^*$, respectively,

$$\hat{R}_A^{(3)} = \begin{cases} \hat{R}_1^* + \hat{R}_2^* + \frac{{\hat{\sigma}_n^*}^2}{n^{1-2/\hat{\alpha}_n^*}} & \hat{\alpha}_n^* \geq 2\,, \\ n^{1-2/\hat{\alpha}_n^*}(\hat{R}_1^* + \hat{R}_2^*) + {\hat{\sigma}_n^*}^2 & \hat{\alpha}_n^* < 2, \end{cases} \tag{B.2}$$

with

$$\hat{R}_1^* = \Gamma((\hat{\alpha}_n^*+2)/\hat{\alpha}_n^*)(\hat{\alpha}_n^*/\hat{c}_n^*)^{2/\hat{\alpha}_n^*}[{\hat{\sigma}_n^*}^2(1-\hat{\rho}_n^*)]/2{\hat{\mu}_n^*}^2\,,$$

and

$$\hat{R}_2^* = \Gamma((\hat{\alpha}_n^*+2)/\hat{\alpha}_n^*)(\hat{\alpha}_n^*/\hat{c}_n^*)^{2/\hat{\alpha}_n^*}[n(1-\hat{\rho}_n^*)]^{-2},$$

$$\hat{R}_B^{(2)} = \left\{ 1 + \frac{L_3(\hat{b}_n^*)}{I_1(\hat{b}_n^*) - I_2(\hat{b}_n^*)} \right\} \hat{\sigma}_n^{*2}, \tag{B.3}$$

and

$$\hat{R}_B^{(3)} = \left\{ 1 + \frac{\hat{b}_n^{*2} \hat{\sigma}_n^{*2}}{n\hat{\mu}_n^{*2} + \hat{b}_n^{*2} \hat{\sigma}_n^{*2}} \right\} \hat{\sigma}_n^{*2}. \tag{B.4}$$

B.2 Estimation of unknown parameters in Rules I and II

In this section, we address the problem of estimating the unknown parameters in Rules I and II. Suppose first that $\alpha > 2$ or $\alpha \leq 2$ is known a priori. Then, according to Theorems 1 and 2 and Remark 3, it is reasonable to estimate ρ by

$$\hat{\rho}_n^* = \begin{cases} \hat{\rho}_n & \alpha \leq 2\,, \\ \tilde{\rho}_n & \alpha > 2. \end{cases} \tag{B.5}$$

By virtue of (B.5), it is natural to estimate μ and σ^2 by $\hat{\mu}_n^* = n^{-1}\sum_{t=1}^{n-1}(y_{t+1} - \hat{\rho}_n^* y_t)$ and $\hat{\sigma}_n^{*2} = n^{-1}\sum_{t=1}^{n-1}(y_{t+1} - \hat{\mu}_n^* - \hat{\rho}_n^* y_t)^2$. In addition, $b = n(1-\rho)$ can be consistently estimated by $\hat{b}_n^* = n(1-\hat{\rho}_n^*)$, in view of Theorems 2.1 and 2.2. The performance of $\hat{b}_n^*$ is demonstrated via the empirical estimate,

$$\hat{\mathrm{E}}\left(\frac{\hat{b}_n^* - b}{b} \right) = \frac{1}{5000} \sum_{i=1}^{5000} \frac{\hat{b}_n^*(i) - b}{b},$$

of the relative bias $\mathrm{E}[(\hat{b}_n^* - b)/b]$, based on the data generated from 5000 simulation runs of model (1.1) with Beta(α,1) error, where $\rho \in \{0.86, 0.9, 0.95, 0.975, 0.99\}$, $\alpha \in \{1, 1.5, 2, 2.5, 3.5, 4\}$, and $\hat{b}_n^*(i)$ is $\hat{b}_n^*$ obtained in the ith simulation. Since our study is meant to be illustrative rather than exhaustive, we only focus on the sample size $n = 10000$. The results are summarized in Table 7. It is shown in Table 7 that all values of $\hat{\mathrm{E}}[(\hat{b}_n^* - b)/b]$ are quite close to 0, and $|\hat{\mathrm{E}}[(\hat{b}_n^* - b)/b]|$ is clearly smaller in the case of $\alpha < 2$ than in the case of $\alpha \geq 2$. This latter feature coincides with the fact that the convergence rate of $\hat{\rho}_n$ in the case of $\alpha < 2$ is faster than $\tilde{\rho}_n$.

Estimating c and α is much more involved than estimating b. While it seems feasible to perform kernel density estimation based on the AR residuals, $\hat{\varepsilon}_i = y_i - \hat{\rho}_n^* y_{i-1}$, to estimate c and α, the usual kernel estimators can be seriously biased when $0 < \alpha \leq 1$ because the corresponding density function is nonzero or even has a pole at the origin; see Marron and Ruppert (1994). Indeed, Marron and Ruppert (1994) suggested some sophisticated kernel estimation algorithms to reduce the boundary bias. However, consistency of the resulting estimators of c and α still seems difficult to establish when only (1.2) is assumed. In this connection, we also mention that a similar difficulty arises in constructing a confidence interval for ρ based on (1.5), in which α and c appear in the normalizing constant and α also appears in the limit. To bypass this difficulty, Datta and McCormick (1995) proposed an asymptotically pivotal quantity based on $\hat{\rho}_n$ and adopted a bootstrap procedure to consistently estimate the limit distribution of the proposed pivotal quantity.

Table 7: The values of $\hat{\mathrm{E}}[(\hat{b}_n^* - b)/b]$, with $n = 10000$, under model (1.1) with Beta$(\alpha, 1)$ errors.

	$\rho(b)$				
α	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002
1.5	-0.0003	-0.0003	-0.0003	-0.0003	-0.0003
2	-0.014	-0.013	-0.013	-0.013	-0.013
2.5	0.001	0.004	0.003	0.004	0.004
3	0.003	0.004	0.004	0.003	0.003
4	0.001	0.002	0.005	0.004	0.002

Table 8: The values of $\hat{\mathrm{E}}(\hat{\alpha}_n^*(m) - \alpha)$, with $n = 10000$, under model (1.1) with Beta$(\alpha, 1)$ errors.

		$\rho(b)$				
α	m	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	250	0.022	0.023	0.018	0.022	0.023
1.5	250	-0.013	-0.009	-0.006	-0.007	-0.009
2	500	-0.040	-0.040	-0.040	-0.040	-0.043
2.5	250	0.068	0.107	0.095	0.109	0.099
3	250	0.098	0.101	0.088	0.097	0.095
4	250	0.097	0.101	0.148	0.128	0.114

Here, we take a somewhat nonstandard approach to estimate α and c. Note that (1.2) yields

$$\lim_{n\to\infty} P(n^{1/\alpha}\varepsilon_{(1)} > x) = \exp(-(c/\alpha)x^{\alpha}),$$

where $\varepsilon_{(j)}$ is the jth order statistic of $\{\varepsilon_1, \ldots, \varepsilon_n\}$, and hence $n^{1/\alpha}\varepsilon_{(1)}$ has the limiting Weibull density,

$$f_{(1)}(x) = \frac{\alpha}{\lambda^{\alpha}} x^{\alpha-1}\exp(-(x/\lambda)^{\alpha}), \tag{B.6}$$

with shape parameter α and scale parameter $\lambda = (\alpha/c)^{1/\alpha}$. This motivates the following procedure for estimating α and c:

1. Produce the AR residuals: $\hat{\varepsilon}_{i+1} = y_{i+1} - \hat{\rho}^*_n y_i, i = 1, \ldots, n-1$.

2. Divide $\{1, \ldots, n\}$ into m subgroups, $\{1, \ldots, n_1\}, \ldots, \{n_{m-1}+1, \ldots, n_m\}$, where $n_i = \lfloor (n-1)/m \rfloor$ or $\lfloor (n-1)/m \rfloor + 1$ with $\lfloor a \rfloor$ denoting the largest integer $\leq a$.

3. Let $\hat{\varepsilon}_{(1)}(j)$ denote the smallest positive value among $\{\hat{\varepsilon}_{n_{j-1}+1}, \ldots, \hat{\varepsilon}_{n_j}\}, j = 1, \ldots, m$.

4. Use the Weibull density (B.6) and $n_1^{1/\alpha}\hat{\varepsilon}_{(1)}(1), \ldots, n_m^{1/\alpha}\hat{\varepsilon}_{(1)}(m)$ to construct the maximum likelihood estimate $(\hat{\alpha}^*_n(m), \hat{\lambda}^*_n(m))$ of (α, λ).

5. Estimate c by $\hat{c}^*_n(m) = \hat{\alpha}^*_n(m)/(\hat{\lambda}^*_n(m))^{\hat{\alpha}^*_n(m)}$.

Under the stationary model (1.1), Hsiao, Huang, and Ing (2017) established the consistency of $(\hat{\alpha}_n(m), \hat{c}_n(m))$ regardless of whether $\alpha \leq 2$ or $\alpha > 2$, where $(\hat{\alpha}_n(m), \hat{c}_n(m))$ is $(\hat{\alpha}^*_n(m), \hat{c}^*_n(m))$ with $\hat{\varepsilon}_{i+1}$ replaced by the EV residual $y_{i+1} - \hat{\rho}_n y_i$, $m \to \infty$ and $n/m \to \infty$. This result enables them to asymptotically correctly identify the better estimator between $\hat{\rho}_n$ and $\tilde{\rho}_n$ in a data-driven fashion. The consistency of $(\hat{\alpha}^*_n(m), \hat{c}^*_n(m))$ under the near unit-root model (1.7) can also be established by an argument similar to that used in Hsiao, Huang and Ing (2017). The details, however, are not pursued here. In Table 8, the empirical estimate,

$$\hat{\mathrm{E}}(\hat{\alpha}^*_n(m) - \alpha) = \frac{1}{5000}\sum_{i=1}^{5000}(\hat{\alpha}^*_{n,i}(m) - \alpha),$$

of the bias of $\hat{\alpha}^*_n(m)$, $\mathrm{E}(\hat{\alpha}^*_n(m) - \alpha)$, is presented under the same scenarios as those in Table 7, where $\hat{\alpha}^*_{n,i}(m)$ is $\hat{\alpha}^*_n(m)$ obtained in the ith simulation. The tuning parameter m is set to 250 and 500 in our study. However, only the smaller one between $\hat{\mathrm{E}}(\hat{\alpha}^*_n(250) - \alpha)$ and $\hat{\mathrm{E}}(\hat{\alpha}^*_n(500) - \alpha)$ is reported in Table 8. It remains for future research to choose m such that the resultant $\hat{\alpha}^*_n(m)$ has a better finite sample performance. With the same m as in Table 8, we present the empirical estimate, $\hat{\mathrm{E}}(\hat{c}^*_n(m) - c)$, of $\mathrm{E}(\hat{c}^*_n(m) - c)$ in Table 9. Table 8 reveals that $\hat{\alpha}^*_n(m)$ appears to be a reliable estimate of α because all values of $|\hat{\mathrm{E}}(\hat{\alpha}^*_n(m) - \alpha)|$ are small. On the other hand, we notice that $|\hat{\mathrm{E}}(\hat{\alpha}^*_n(m) - \alpha)|$ is larger in $\alpha > 1.5$ than $\alpha = 1.5$, which may be attributed to a slower convergence rate of $\hat{\rho}^*_n$ in the former case. In addition, perhaps due to a positive value of the density function at the origin, the performance of $\hat{\alpha}^*_n(m)$ in the case of $\alpha = 1$ also looks inferior to that in the case of $\alpha = 1.5$, although $\hat{\rho}^*_n$ in the former case has a faster convergence rate.

Table 9: The values of $\hat{\mathrm{E}}(\hat{c}_n^*(m) - c)$, with $n = 10000$ and the same m as those in Table 8, under model (1.1) with Beta$(\alpha, 1)$ errors.

	$\rho(b)$				
α	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	0.152	0.152	0.151	0.160	0.146
1.5	0.033	0.059	0.049	0.071	0.055
2	0.097	0.109	0.111	0.105	0.077
2.5	0.484	0.596	0.698	0.703	0.539
3	0.517	0.451	0.388	0.473	0.455
4	0.591	0.578	0.565	0.560	0.551

Table 10: The values of $\hat{\mathrm{E}}(\hat{c}_n^*(m, r) - c)$, with $n = 10000$ and the same m as those in Table 8, under model (1.1) with Beta$(\alpha, 1)$ errors.

		$\rho(b)$				
α	r	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	25	0.014	0.011	-0.006	0.006	0.014
1.5	5	-0.005	-0.013	-0.031	0.015	0.013
2	5	0.023	0.018	0.021	0.015	0.005
2.5	20	-0.023	0.082	0.093	0.133	0.062
3	15	0.077	0.069	0.021	0.066	0.060
4	15	-0.046	-0.023	0.117	0.050	0.088

Table 9 shows that the performance of $\hat{c}_n^*(m)$ is in general unsatisfactory. In particular, all values of $\hat{\mathrm{E}}(\hat{c}_n^*(m) - c)$ are positive and are considerably larger than 0 for $\alpha > 2$. Taking a closer look at

$$\hat{c}_n^*(m) = \frac{\hat{\alpha}_n^*(m)}{m^{-1} \sum_{i=1}^m n_i [\hat{\varepsilon}_{(1)}(i)]^{\hat{\alpha}_n^*(m)}}, \tag{B.7}$$

we found that a non-negligible portion of $\{n_i[\hat{\varepsilon}_{(1)}(i)]^{\hat{\alpha}_n^*(m)}\}$ concentrates near 0. As a result, the denominator on the right-hand side of (B.7) tends to underestimate $\lambda^\alpha = \alpha/c$, and hence $\hat{c}_n^*(m)$ tends to overestimate c, as observed in Table 9. To remedy this difficulty, we suggest an alternative, $\hat{c}_n^*(m, r)$, which is $\hat{c}_n^*(m)$ with the denominator replaced by the sample mean from the highest $(1 - r)\%$ of the elements of $\{n_i[\hat{\varepsilon}_{(1)}(i)]^{\hat{\alpha}_n^*(m)}, i = 1, \ldots m\}$. Under the same simulation setting as Table 9, we compute the empirical estimate, $\hat{\mathrm{E}}(\hat{c}_n^*(m, r) - c)$, of $\mathrm{E}(\hat{c}_n^*(m, r) - c)$, and report the smallest one among $\hat{\mathrm{E}}(\hat{c}_n^*(m, r) - c), r = 5, 10, 15, 20, 25, 30$; see Table 10. Table 10 shows that all values of $|\mathrm{E}(\hat{c}_n^*(m, r) - c)|$ are not distant from 0, and clearly smaller than $|\hat{\mathrm{E}}(\hat{c}_n^*(m) - c)|$.

We now return to the more practical situation where $\alpha > 2$ or $\alpha \leq 2$ is unknown. In this case, we suggest the following rule:

Table 11: The values of F, with $n = 10000$, under model (1.1) with Beta$(\alpha, 1)$ errors.

	$\rho(b)$				
α	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	1.000	1.000	1.000	1.000	1.000
1.5	1.000	1.000	1.000	1.000	1.000
2	0.978	0.984	0.980	0.982	0.986
2.5	0.922	0.930	0.938	0.926	0.908
3	1.000	0.996	0.998	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000

Rule III.

1. Judge $\alpha > 2$ if $\hat{\alpha}_n(m) - \xi > 2$,

2. Judge $\alpha \leq 2$ if $\hat{\alpha}_n(m) - \xi \leq 2$,

where $\hat{\alpha}_n(m)$ is defined previously and ξ is a prescribed small positive number. In Table 11, with the same scenarios as those in Table 7, we report the percentage, F, of Rule III (with $m = 500$ and $\xi = 0.14$) making correct judgements, where

$$F = \sharp(\{i : 1 \leq i \leq 5000, I(\hat{\alpha}_{n,i}(500) - 0.14 > 2) = I(\alpha > 2)\})/5000,$$

$n = 10000$ and $\hat{\alpha}_{n,i}(500)$ denotes $\hat{\alpha}_n(500)$ obtained in the ith simulation. Note that $\xi = 0.14$ is an approximation of $2\sigma_{\text{MLE}}/\sqrt{500}$, where σ^2_{MLE} is the limiting variance of the MLE of α of the Weibull density (A.6) calculated at $\alpha = 2$ and $\lambda = 1$. As shown in Table 11, all values of F are near 1, in particular when $\alpha < 2$ or $\alpha > 2.5$. This result implies that Rule III provides a reliable decision about whether or not $\alpha > 2$, thereby allowing one to carry out the aforementioned estimates of α and c in practice.

Finally, we want to reiterate that this section is exploratory in nature, and there remain a number of unsettled issues (e.g., the choices of m and r in $\hat{\alpha}^*_n(m)$ and $\hat{c}^*_n(m, r)$) worthy of further investigation. On the other hand, our simulation study suggests that the notoriously difficult problem of estimating α and c in the distribution of ε_t can be somewhat alleviated through the proposed estimates, $\hat{\alpha}^*_n(m)$ and $\hat{c}^*_n(m, r)$, provided m and r are properly given.