Predictor selection for positive autoregressive processes

Ching-Kang Ing and Chiao-Yi Yang

Abstract

Let observations $y_1, \cdots, y_n$ be generated from a first-order autoregressive (AR) model with positive errors. In both the stationary and unit root cases, we derive moment bounds and limiting distributions of an extreme value estimator, $\hat{\rho}_n$, of the AR coefficient. These results enable us to provide asymptotic expressions for the mean squared error (MSE) of $\hat{\rho}_n$ and the mean squared prediction error (MSPE) of the corresponding predictor, $\hat{y}_{n+1}$, of y_{n+1}. Based on these expressions, we compare the relative performance of $\hat{y}_{n+1}$ ($\hat{\rho}_n$) and the least squares predictor (estimator) from the MSPE (MSE) point of view. Our comparison reveals that the better predictor (estimator) is determined not only by whether a unit root exists, but also by the behavior of the underlying error distribution near the origin, and hence is difficult to identify in practice. To circumvent this difficulty, we suggest choosing the predictor (estimator) with the smaller accumulated prediction error and show that the predictor (estimator) chosen in this way is asymptotically equivalent to the better one. Both real and simulated data sets are used to illustrate the proposed method.

KEY WORDS: Accumulated prediction error; Mean squared prediction error; Moment bound; Positive autoregressive model; Predictor selection; Unit root

---

Ching-Kang Ing is Research Fellow and Chiao-Yi Yang is Postdoctoral Researcher, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, 11529. Ching-Kang Ing is also Professor, Department of Economics, National Taiwan University, Taipei, Taiwan, 10055. Ing's research was supported in part by the National Science Council of Taiwan under grant NSC 101-2118-M-001-003 and Academia Sinica Investigator Award.

1. INTRODUCTION

Over the past few decades, modeling and estimation for positive-valued time series have attracted great interest in fields such as reliability theory, economics, finance, hydrology and meteorology; see, e.g., Gaver and Lewis (1980), Lawrance and Lewis (1985), Bell and Smith (1986), Sim (1987), Lewis, Mckenzie and Hugus (1989), Hutton (1990), Barndorff-Nielsen and Shephard (2001), Nielsen and Shephard (2003) and Sarlak (2008). Among the many positive-valued time series models proposed in the literature, the stationary positive AR(1) model,

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{1}$$

is one of the most popular, where $0 \leq \rho < 1$ is an unknown constant and ε_t's are i.i.d. positive random disturbances. If $\mu = \mathrm{E}(\varepsilon_t) < \infty$, then model (1) can be expressed as $y_t = \mu + \rho y_{t-1} + \delta_t$, where $\delta_t = \varepsilon_t - \mu$. To appreciate the practical relevance of model (1), note first that when the distribution of ε_1 is carefully specified, the sequence $\{y_t\}$ has a marginal exponential distribution (e.g., Gaver and Lewis 1980). By making use of this property, one can easily simulate queues with correlated service times which are useful for checking for the sensitivity of standard queuing results to departures from the independence. Model (1) has also found extensive applications in hydrological studies. For example, Bell and Smith (1986) analyzed two sets of pollution data from the Willamette River, Oregon, using model (1) with different positive errors, and Sarlak (2008) analyzed the annual streamflow data from Kizilirmak River, Turkey, showing that model (1) with a Weibull error distribution is more appropriate than a Gaussian one. In addition, models similar to (1) have been adopted by Barndorff-Nielsen and Shephard (2001) as components of their continuous time linear stochastic volatility models for financial assets.

Having observed $y_1, \ldots, y_n$, ρ can be estimated by the maximum likelihood estimator (MLE) when the parametric form of the error distribution is known. However, not only is the error distribution unknown, but the MLE is analytically difficult to work with; see Davis and McCormick (1989) for a related discussion. On the other hand, the extreme value estimator,

$$\hat{\rho}_n = \min_{1 \leq i \leq n-1} y_{i+1}/y_i, \tag{2}$$

which possesses consistency under rather mild assumptions (e.g., Bell and Smith 1986), is a good alternative to bypass these difficulties. By making use of point process techniques, Davis and McCormick (1989) further showed that the limiting distributions of $\hat{\rho}_n$ depend only on the local behaviors of the distribution of ε_1 near the origin. Specifically, if the probability density function (pdf) of ε_1, $f_{\varepsilon_1}(\cdot)$, satisfies

$$\lim_{x\downarrow 0} \frac{f_{\varepsilon_1}(x)}{cx^{\alpha-1}} = 1, \text{ for some unknown } \alpha > 0 \text{ and } c > 0, \tag{3}$$

then Corollary 2.4 (or Corollary 2.5) of Davis and McCormick (1989) yields that for $0 \le \rho < 1$,

$$\lim_{n\to\infty} \mathrm{P}\left((cM_\alpha/\alpha)^{1/\alpha} n^{1/\alpha}(\hat{\rho}_n - \rho) > t\right) = \exp(-t^\alpha), \tag{4}$$

where $M_\alpha = \mathrm{E}[(\sum_{j=0}^{\infty} \rho^j \varepsilon_{1-j})^\alpha]$ (see Section 2 below for more details).

Another commonly used estimator of ρ is the least squares estimator (LSE), $\tilde{\rho}_n$ [see (9) below], which also enjoys consistency under general error distributions. Moreover, as shown in Hamilton (1994), for $0 \le \rho < 1$, $n^{1/2}(\tilde{\rho}_n - \rho)$ has a limiting normal distribution. This, together with (4), reveals a special feature of $\hat{\rho}_n$ that its rate of convergence is faster than $\tilde{\rho}_n$ if $0 < \alpha < 2$, but slower if $\alpha > 2$. On the other hand, it seems difficult to use (4) to construct a confidence interval for ρ due to the unknown index of regular variation α, which appears in the normalizing constant and in the limit. To rectify this deficiency, Datta and McCormick (1995) proposed an asymptotically pivotal quantity based on $\hat{\rho}_n$ and adopted a bootstrap procedure to consistently estimate the limiting distribution of the proposed pivotal quantity, thereby leading to a totally nonparametric confidence interval for ρ.

While Davis and McCormick's (1989) results are profound, they preclude the unit root model, i.e., model (1) with $\rho = 1$, which is one of the most widely discussed nonstationary time series models in the case of zero-mean errors. In the case of positive errors, the unit root model has also found broad applications since it provides a convenient way to describe some economic, financial and epidemiological data that are always positive and fluctuate around an upward trend with variance increasing over time. See, for example, the natural logarithm of quarterly real GDP for the United States from 1947 to 1989 (Hamilton 1994, chap. 17) and the yearly cancer death rate in Pennsylvania between 1930 and 2000 (Wei 2006, chap. 6). In fact, when $\rho = 1$, the limiting distribution of $\hat{\rho}_n$ has been derived by Nielsen and Shephard (2003) under exponential innovations. More specifically, their Theorem 2 shows that if

$$f_{\varepsilon_1}(x) = \lambda^{-1}\exp(-x/\lambda), \tag{5}$$

where $x \geq 0$ and $\lambda > 0$ is an unknown scale parameter, then

$$\lim_{n\to\infty} \mathrm{P}\left(2^{-1}n^2(\hat{\rho}_n - \rho) > t\right) = \exp(-t). \tag{6}$$

However, (5) is quite restrictive compared to (3). Moreover, Nielsen and Shephard's approach, relying highly on the likelihood function associated with (5), seems difficult to extend to more general distributions. Therefore, the first goal of this article is to fill this gap by deriving the limiting distribution of $\hat{\rho}_n$ under model (1) with $\rho = 1$ and ε_1 satisfying (3). By making use of a somewhat direct approach given in the supplementary document, we obtain for $\mu < \infty$,

$$\lim_{n\to\infty} \mathrm{P}\left(\mu\{c/[\alpha(\alpha+1)]\}^{1/\alpha}n^{1+(1/\alpha)}(\hat{\rho}_n - \rho) > t\right) = \exp(-t^\alpha). \tag{7}$$

See Section 2 for more details. Since (5) yields $\mu = \lambda$, $c = \lambda^{-1}$ and $\alpha = 1$, (6) becomes an immediate consequence of (7). In addition, (7) and the fact that $n^{3/2}(\tilde{\rho}_n - \rho)$ has a limiting normal distribution if $\rho = 1$ [see Chan 1989 or (24) in Section 2] imply that $\hat{\rho}_n$ is better than $\tilde{\rho}_n$ if $0 < \alpha < 2$, and worse than $\tilde{\rho}_n$ if $\alpha > 2$, in terms of convergence speeds. This conclusion is exactly the same as the one drawn from the case of $0 \leq \rho < 1$.

The second goal of this paper is to provide asymptotic comparisons of the mean squared errors (MSEs), $\mathrm{MSE}_A = \mathrm{E}(\hat{\rho}_n - \rho)^2$ and $\mathrm{MSE}_B = \mathrm{E}(\tilde{\rho}_n - \rho)^2$. Such comparisons, allowing us to choose the more efficient estimator between $\hat{\rho}_n$ and $\tilde{\rho}_n$ (which are both consistent under general positive errors), are particularly relevant in situations where the distribution of ε_1 is unknown. However, since convergence in distribution does not imply convergence of moments, the intended goal *cannot* be achieved through comparing second moments of the limiting distributions of $\hat{\rho}_n$ and $\tilde{\rho}_n$. To overcome this difficulty, we establish moment bounds for $\hat{\rho}_n - \rho$ and $\tilde{\rho}_n - \rho$ after being multiplied by suitable normalizing constants. These moment bounds in conjunction with the limiting distributions of $\hat{\rho}_n$ and $\tilde{\rho}_n$ lead to asymptotic expressions for MSE_A and MSE_B, which in turn form the basis for our comparisons. Corresponding comparison results not only support the previous conclusion that $\hat{\rho}_n$ is better (worse) than $\tilde{\rho}_n$ if $\alpha < 2$ ($\alpha > 2$) from an alternative perspective, but they also reveal a quite subtle phenomenon in the critical case $\alpha = 2$. Whether $\hat{\rho}_n$ is better than $\tilde{\rho}_n$ depends on whether c [the other parameter in (3)] is larger than a threshold value, which varies drastically from the case of $0 \leq \rho < 1$ to the case of $\rho = 1$. Based on these comparisons, the rules, (R1) to (R4), for choosing the more efficient estimator between $\hat{\rho}_n$ and $\tilde{\rho}_n$ are established in Section 2.

One major purpose of time-series modeling is to make forecasts. However, unlike estimation problems under the positive AR(1) model, which have already attracted a lot of attention, prediction problems under this model are seldom discussed even in the stationary case. This motivates us to pursue the third goal of this paper: understanding the behaviors of the mean squared prediction errors (MSPEs) of the extreme value predictor, $\hat{y}_{n+1}$, and the least squares predictor, $\tilde{y}_{n+1}$, in situations where the data are generated from model (1) with $0 \leq \rho \leq 1$. Note that

$$\hat{y}_{n+1} = \hat{\mu}_n + \hat{\rho}_n y_n, \tag{8}$$

where $\hat{\mu}_n = (n-1)^{-1} \sum_{t=1}^{n-1} (y_{t+1} - \hat{\rho}_n y_t)$ is a natural estimate of μ based on $\hat{\rho}_n$, and

$$\tilde{y}_{n+1} = \tilde{\mu}_n + \tilde{\rho}_n y_n, \tag{9}$$

where $(\tilde{\mu}_n,\, \tilde{\rho}_n)^\top$ is the LSE of $(\mu,\, \rho)^\top$ satisfying $(\sum_{j=1}^{n-1} \mathbf{x}_j \mathbf{x}_j^\top)(\tilde{\mu}_n,\, \tilde{\rho}_n)^\top = \sum_{j=1}^{n-1} \mathbf{x}_j y_{j+1}$, with $\mathbf{x}_j = (1,\, y_j)^\top$. Moreover, assume that $\sigma^2 = \mathrm{var}(\varepsilon_1) < \infty$. Then, the MSPEs of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ are given by

$$\mathrm{MSPE}_A = \mathrm{E}(y_{n+1} - \hat{y}_{n+1})^2 = \sigma^2 + \mathrm{E}\{(\hat{\mu}_n - \mu) + (\hat{\rho}_n - \rho) y_n\}^2, \tag{10}$$

and

$$\mathrm{MSPE}_B = \mathrm{E}(y_{n+1} - \tilde{y}_{n+1})^2 = \sigma^2 + \mathrm{E}\{(\tilde{\mu}_n - \mu) + (\tilde{\rho}_n - \rho) y_n\}^2, \tag{11}$$

respectively. By making use of the moment bounds and limiting distributions established in Section 2, we obtain asymptotic expressions for MSPE_A and MSPE_B in Section 3. These expressions reveal that the normalizing constants of MSPE_A and MSPE_B differ markedly from those of MSE_A and MSE_B, but ordering of MSPE_A and MSPE_B is still asymptotically the same as for MSE_A and MSE_B. As a result, rules (R1) to (R4) can also be used to determine the more efficient predictor between $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$.

Unfortunately, (R1) to (R4), requiring knowledge of ρ and the behavior of the unknown error distribution near the origin, are rarely implemented in practice. Therefore, the last but possibly most important goal of this paper is to develop a *data-driven* method that can choose the more efficient predictor (estimator) between $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ ($\hat{\rho}_{n+1}$ and $\tilde{\rho}_{n+1}$) from the MSPE (MSE) point of view. To achieve this goal, in Section 4, we define the accumulated prediction errors (APEs) of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$, APE_A and APE_B, respectively, as

$$\text{APE}_A = \sum_{i=M_1}^{n-1} (y_{i+1} - \hat{y}_{i+1})^2, \tag{12}$$

and

$$\text{APE}_B = \sum_{i=M_2}^{n-1} (y_{i+1} - \tilde{y}_{i+1})^2, \tag{13}$$

where M_1 and M_2 are prescribed positive integers, and propose choosing the predictor (estimator) with the smaller APE. The whole selection scheme is summarized in three steps: (S1) to (S3). Instead of estimating ρ, c and α in (3) directly, (S1) to (S3) take the approach of letting the predictor's past performance speak for itself, which seems more closely related to the intrinsic nature of the underlying problem. In particular, by making use of this feature, we show in Theorem 7 that the predictor and the estimator selected by (S1) to (S3) are asymptotically equivalent to the ones selected by (R1) to (R4), thus ensuring the validity of (S1) to (S3).

In Section 5, the asymptotic results developed for (S1) to (S3) are illustrated by numerical simulations. The usefulness of (S1) to (S3) is also demonstrated by analyzing three time series datasets from studies in quality control, insect behavior and epidemiology. All these series are positive and exhibit significant AR(1) features. Our analysis shows that regardless of whether the series is stationary or nonstationary, (S1) to (S3) can always choose the predictor with the smaller *empirical* MSPE, which is a suitable surrogate for the *unobservable* MSPE. The proofs of the theoretical results presented in Section 4 are deferred to the Appendix, whereas those in Sections 2 and 3 are provided in the supplementary document in light of space constraint.

2. ASYMPTOTIC PROPERTIES OF $\hat{\rho}_n$

Our aim in this section is to pursue the first and the second goals mentioned in Section 1. To facilitate the exposition, we shall assume throughout the rest of this paper that $y_0 = 0$, which yields $y_t = \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j}$ for $t \geq 1$. We first derive the limiting distribution of $\hat{\rho}_n$, establish a bound for the qth moment of $n^{1/\alpha}(\hat{\rho}_n - \rho)$ with $q > 0$, and provide an asymptotic expression for $\text{MSE}_A = \text{E}(\hat{\rho}_n - \rho)^2$ in the case of $0 \leq \rho < 1$. Note that since $\hat{\rho}_n - \rho = \min_{2 \leq t \leq n} \varepsilon_t / y_{t-1}$, $\hat{\rho}_n$ is always positively biased.

Theorem 1. *Assume (1) with* $0 \leq \rho < 1$ *and (3). If*

$$\text{E}(\varepsilon_1^{q_1}) < \infty, \ \textit{for some } q_1 > 0, \tag{14}$$

then

$$\text{E}\{n^{1/\alpha}(\hat{\rho}_n - \rho)\}^q = O(1), \ \textit{for any } q > 0. \tag{15}$$

Moreover, if

$$\text{E}(\varepsilon_1^{q_2}) < \infty, \ \textit{for some } q_2 > \alpha, \tag{16}$$

then (4) follows, and

$$\text{E}(\hat{\rho}_n - \rho)^2 = n^{-2/\alpha}\big(\frac{\alpha}{cM_\alpha}\big)^{2/\alpha}\Gamma\big(\frac{\alpha+2}{\alpha}\big) + o(n^{-2/\alpha}), \tag{17}$$

where $\Gamma(\cdot)$ *denoting the gamma function.*

Remark 1. Equation (17) is an immediate consequence of (4) and (15). To see this, note first that (15) implies that $\text{E}\{n^{1/\alpha}(\hat{\rho}_n - \rho)\}^{2+\delta} = O(1)$ for some $\delta > 0$, which in turn entails the uniform integrability of $\{n^{2/\alpha}(\hat{\rho}_n - \rho)^2\}$. This latter property together with (4) yields

$$\lim_{n\to\infty} n^{2/\alpha}\text{E}(\hat{\rho}_n - \rho)^2 = \big(\frac{\alpha}{cM_\alpha}\big)^{2/\alpha}\int_0^\infty \exp(-t^{\alpha/2})dt = \big(\frac{\alpha}{cM_\alpha}\big)^{2/\alpha}\Gamma\big(\frac{\alpha+2}{\alpha}\big),$$

and hence (17) follows. Similarly, (4) and the uniform integrability of $\{n^{1/\alpha}(\hat{\rho}_n - \rho)\}$ lead to an asymptotic expression for the bias of $\hat{\rho}_n$,

$$\mathrm{E}(\hat{\rho}_n - \rho) = n^{-1/\alpha}\big(\frac{\alpha}{cM_\alpha}\big)^{1/\alpha}\Gamma\big(\frac{\alpha+1}{\alpha}\big) + o(n^{-1/\alpha}). \tag{18}$$

Remark 2. Assumption (14) is quite general because it holds even when the first moment of ε_1 does not exist. At first glance, (16) seems to be restrictive when α is large. This assumption, however, is indispensable in proving (4) and (17). Indeed, we have found that the rate of convergence of $\hat{\rho}_n$ can be faster than $n^{1/\alpha}$ when (16) is violated. While investigating the limiting distribution and the MSE of $\hat{\rho}_n$ in situations where (16) fails to hold is of theoretical interest, these types of problems require a separate treatment and are not pursued in this paper.

A unit-root counterpart of Theorem 1 is developed in the next theorem.

Theorem 2. *Assume (1) with $\rho = 1$ and (3). If (14) holds, then*

$$\mathrm{E}\{n^{1+(1/\alpha)}(\hat{\rho}_n - \rho)\}^q = O(1), \ \textit{for any } q > 0. \tag{19}$$

Moreover, if

$$\mathrm{E}(\varepsilon_1^{q_2}) < \infty, \ \textit{for some } q_2 > 1, \tag{20}$$

then (7) follows, and

$$\mathrm{E}(\hat{\rho}_n - \rho)^2 = n^{-2(1+\alpha^{-1})}\mu^{-2}\big\{\frac{\alpha(\alpha+1)}{c}\big\}^{2/\alpha}\Gamma\big(\frac{\alpha+2}{\alpha}\big) + o(n^{-2(1+\alpha^{-1})}). \tag{21}$$

Remark 3. In view of the argument given in Remark 1, it is not surprising that (21) can be obtained directly from (7) and (19). These two equations also yield an asymptotic expression for the bias of $\hat{\rho}_n$,

$$\mathrm{E}(\hat{\rho}_n - \rho) = n^{-(1+\alpha^{-1})}\mu^{-1}\big\{\frac{\alpha(\alpha+1)}{c}\big\}^{1/\alpha}\Gamma\big(\frac{\alpha+1}{\alpha}\big) + o(n^{-(1+\alpha^{-1})}), \tag{22}$$

in the case of $\rho = 1$.

Remark 4. Condition (20), precluding only some heavy-tailed distributions, seems flexible enough to accommodate a wide range of applications. The appearance of μ in (7) and (21) also suggests that (20) is difficult to be weakened. In fact, when (20) is violated, one can construct examples showing that the rate of converge of $\hat{\rho}_n$ to 1 is faster than $n^{1+(1/\alpha)}$. However, the details are beyond the scope of this paper and will be reported elsewhere.

We now have achieved the first goal of this paper through Theorem 2. To attain the second goal, we need asymptotic expressions for $\mathrm{MSE}_B = \mathrm{E}(\tilde{\rho}_n - \rho)^2$ in addition to those for MSE_A given in Theorems 1 and 2. Before proceeding further, it is worthwhile to investigate the convergence rate of $\tilde{\rho}_n$. By Chan (1989), it follows that for $\mu > 0$ and $0 \le \rho < 1$,

$$(n^{1/2}(\tilde{\mu}_n - \mu), n^{1/2}(\tilde{\rho}_n - \rho))^\top \Rightarrow \mathbf{Z}_1, \tag{23}$$

and for $\mu > 0$ and $\rho = 1$,

$$(n^{1/2}(\tilde{\mu}_n - \mu), n^{3/2}(\tilde{\rho}_n - \rho))^\top \Rightarrow \mathbf{Z}_2, \tag{24}$$

where $\Rightarrow$ denotes convergence in distribution and $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are bivariate normal distributions with mean vectors zero and covariance matrices

$$A = \begin{pmatrix} 1 & \frac{\mu}{1-\rho} \\ \frac{\mu}{1-\rho} & (\frac{\mu}{1-\rho})^2 + \frac{\sigma^2}{1-\rho^2} \end{pmatrix}^{-1} \sigma^2, \text{ and } B = \begin{pmatrix} 1 & \frac{\mu}{2} \\ \frac{\mu}{2} & \frac{\mu^2}{3} \end{pmatrix}^{-1} \sigma^2,$$

respectively. Unlike the normalizing constants in (4) and (7), those in (23) and (24) are independent of α. Moreover, (4), (7), (23) and (24) indicate that in both stationary and unit root cases, the convergence rate of $\hat{\rho}_n$ is faster than that of $\tilde{\rho}_n$ if $0 < \alpha < 2$, and is slower if $\alpha > 2$. In the critical case of $\alpha = 2$, both estimators share the same convergence rate, which is $n^{-1/2}$ when $0 \le \rho < 1$, and $n^{-3/2}$ when $\rho = 1$.

In the following, we shall provide moment bounds for $n^{1/2}(\tilde{\rho}_n - \rho)$ with $0 \le \rho < 1$, and $n^{3/2}(\tilde{\rho}_n - \rho)$ with $\rho = 1$. These bounds can be used in conjunction with (23) and (24) to yield the desired asymptotic expressions for MSE_B. By an argument similar to that used by Yu, Lin and Cheng (2012), it can be shown that if

$$\mathrm{E}|\delta_1|^s < \infty, \text{ for some } s > 10, \tag{25}$$

and if there exist positive numbers K, η_1, η_2 and M such that for all $m \ge M$ and $|x - y| \le \eta_1$, the distribution function of $m^{-1/2}\sum_{t=1}^{m}(\varepsilon_t - \mu)$, $F_m(\cdot)$, satisfies

$$|F_m(x) - F_m(y)| \le K|x - y|^{\eta_2}, \tag{26}$$

then

$$\mathrm{E}|n^{1/2}(\tilde{\rho}_n - \rho)|^{\gamma_1} = O(1),\ 0 \le \rho < 1, \tag{27}$$

and

$$\mathrm{E}|n^{3/2}(\tilde{\rho}_n - \rho)|^{\gamma_1} = O(1),\ \rho = 1, \tag{28}$$

where γ_1 is some positive number greater than 2. Combining (27) and (28) with (23) and (24) gives

$$\lim_{n\to\infty} n\mathrm{E}(\tilde{\rho}_n - \rho)^2 = 1 - \rho^2,\ 0 \le \rho < 1, \tag{29}$$

and

$$\lim_{n\to\infty} n^3\mathrm{E}(\tilde{\rho}_n - \rho)^2 = \frac{12\sigma^2}{\mu^2},\ \rho = 1. \tag{30}$$

It is clear from (17), (21), (29) and (30) that from the MSE point of view, $\hat{\rho}_n$ is again better (worse) than $\tilde{\rho}_n$ when $0 < \alpha < 2$ ($\alpha > 2$). Moreover, since (17) and (21) imply that for $\alpha = 2$, $\lim_{n\to\infty} n\mathrm{E}(\hat{\rho}_n - \rho)^2 = 2\{c[\sigma^2(1-\rho^2)^{-1} + \mu^2(1-\rho)^{-2}]\}^{-1}$ if $0 \le \rho < 1$, and $\lim_{n\to\infty} n^3\mathrm{E}(\hat{\rho}_n - \rho)^2 = 6/(c\mu^2)$ if $\rho = 1$, one gets from these identities and (29) and (30) that for $\alpha = 2$ and $0 \le \rho < 1$,

$$\lim_{n\to\infty} n\mathrm{E}(\hat{\rho}_n - \rho)^2 < \lim_{n\to\infty} n\mathrm{E}(\tilde{\rho}_n - \rho)^2 \text{ if and only if } c > \frac{2(1-\rho)}{(1+\rho)\mu^2 + (1-\rho)\sigma^2}, \tag{31}$$

and for $\alpha = 2$ and $\rho = 1$,

$$\lim_{n\to\infty} n^3\mathrm{E}(\hat{\rho}_n - \rho)^2 < \lim_{n\to\infty} n^3\mathrm{E}(\tilde{\rho}_n - \rho)^2 \text{ if and only if } c > 1/(2\sigma^2). \tag{32}$$

To conclude, the above comparison suggests the following rules for choosing the better estimator (in terms of MSE) between $\hat{\rho}_n$ and $\tilde{\rho}_n$:

(R1) Choose $\hat{\rho}_n$ if $0 < \alpha < 2$;

(R2) Choose $\tilde{\rho}_n$ if $\alpha > 2$;

(R3) For $\alpha = 2$ and $0 \le \rho < 1$, choose $\hat{\rho}_n$ if $c > 2(1-\rho)\{(1+\rho)\mu^2 + (1-\rho)\sigma^2\}^{-1}$, $\tilde{\rho}_n$ if $c < 2(1-\rho)\{(1+\rho)\mu^2+(1-\rho)\sigma^2\}^{-1}$, and either $\hat{\rho}_n$ or $\tilde{\rho}_n$ if $c = 2(1-\rho)\{(1+\rho)\mu^2+(1-\rho)\sigma^2\}^{-1}$;

(R4) For $\alpha = 2$ and $\rho = 1$, choose $\hat{\rho}_n$ if $c > 1/(2\sigma^2)$, $\tilde{\rho}_n$ if $c < 1/(2\sigma^2)$, and either $\hat{\rho}_n$ or $\tilde{\rho}_n$ if $c = 1/(2\sigma^2)$.

However, since ρ, α, c, μ and σ^2 are unknown, (R1) to (R4) seem to be practically irrelevant. In Section 4, we shall resolve this difficulty using a data-driven method based on the APE.

3. MEAN SQUARED PREDICTION ERROR

3.1. The MSPE of $\hat{y}_{n+1}$

Throughout this section it will be assumed that $0 < \sigma^2 = \text{var}(\varepsilon_1) < \infty$. Recall the definition of $\hat{y}_{n+1}$ given in (8). In the case of $0 \le \rho < 1$, straightforward calculations yield

$$y_{n+1} - \hat{y}_{n+1} = \delta_{n+1} - \{(\hat{\rho}_n - \rho)(y_n - \mu_y) + [(1-\rho)\bar{y}_{n-1} - \mu] - (\hat{\rho}_n - \rho)(\bar{y}_{n-1} - \mu_y) + z_n\}, \quad (33)$$

where $\mu_y = \mu/(1-\rho), \bar{y}_n = n^{-1}\sum_{j=1}^n y_j$ and $z_n = (y_n - y_1)/(n-1)$. Since δ_{n+1} is independent of $\hat{\rho}_n, y_n, \bar{y}_{n-1}$ and z_n, it follows from (33) that the MSPE of $\hat{y}_{n+1}$, MSPE_A, obeys,

$$\text{MSPE}_A - \sigma^2 = \text{E}\{(\hat{\rho}_n - \rho)(y_n - \mu_y) + [(1-\rho)\bar{y}_{n-1} - \mu] - (\hat{\rho}_n - \rho)(\bar{y}_{n-1} - \mu_y) + z_n\}^2. \quad (34)$$

Moreover, it is shown in the supplementary materials that (i) the orders of the magnitude of z_n and $(\hat{\rho}_n - \rho)(\bar{y}_{n-1} - \mu_y)$ are negligible compared to that of $(1-\rho)\bar{y}_{n-1} - \mu$, (ii) $n^{(1/2)+(1/\alpha)}[(1-\rho)\bar{y}_{n-1} - \mu](\hat{\rho}_n - \rho)(y_n - \mu_y)$ has an asymptotic mean of zero, and (iii) $n^{1/\alpha}(\hat{\rho}_n - \rho)$ and $y_n - \mu_y$ are asymptotically independent. These facts, together with (15) and (34), lead to

$$\text{MSPE}_A - \sigma^2 = \text{E}(y_n - \mu_y)^2 \text{E}(\hat{\rho}_n - \rho)^2 + \text{E}[(1-\rho)\bar{y}_{n-1} - \mu]^2 + o(\max\{n^{-1}, n^{-2/\alpha}\}). \quad (35)$$

Relation (35) points out that to obtain $\text{MSPE}_A - \sigma^2$ from the MSE of $\hat{\rho}_n$, one needs not only to multiply the latter by an adjustment factor, $\text{E}(y_n - \mu_y)^2$, which converges to $\sigma^2/(1-\rho^2)$ as $n \to \infty$, but also to add the MSE of $(1-\rho)\bar{y}_{n-1}$, which is a root-n consistent "semi-population" estimator of μ. Based on (35), the next theorem provides an asymptotic expression for the MSPE_A in the case of $0 \le \rho < 1$.

Theorem 3. *Assume (1) with* $0 \le \rho < 1$ *and (3). Suppose*

$$\mathrm{E}(\varepsilon_1^{q_2}) < \infty, \text{ for some } q_2 > \max\{\alpha, 2\}. \tag{36}$$

Then,

$$\mathrm{MSPE}_A - \sigma^2 = n^{-2/\alpha} L_0(\alpha, c, \rho, \sigma^2) + \sigma^2 n^{-1} + o(\max\{n^{-1}, n^{-2/\alpha}\}), \tag{37}$$

where

$$L_0(\alpha, c, \rho, \sigma^2) = \Gamma((\alpha+2)/\alpha)\{\alpha/(cM_\alpha)\}^{2/\alpha}\{\sigma^2/(1-\rho^2)\}. \tag{38}$$

When $\rho = 1$, an alternative decomposition of $y_{n+1} - \hat{y}_{n+1}$ is required: $y_{n+1} - \hat{y}_{n+1} = \delta_{n+1} - \{(\hat{\rho}_n - 1)(y_n - \bar{y}_{n-1}) + (z_n - \mu)\}$, which implies

$$\mathrm{MSPE}_A - \sigma^2 = \mathrm{E}((\hat{\rho}_n - 1)(y_n - \bar{y}_{n-1}) + (z_n - \mu))^2. \tag{39}$$

It is easy to see that $\lim_{n\to\infty} n\mathrm{E}(z_n - \mu)^2 = \sigma^2$. In addition, we show in the supplementary materials that $\mathrm{E}\{(\hat{\rho}_n - 1)(y_n - \bar{y}_{n-1})(z_n - \mu)\} = o(\max\{n^{-1}, n^{-2/\alpha}\})$, and $n^{1+(1/\alpha)}(\hat{\rho}_n - 1)$ and $(y_n - \bar{y}_{n-1})/n$ are asymptotically independent. These facts, together with (39) and (19), yield

$$\mathrm{MSPE}_A - \sigma^2 = \mathrm{E}(y_n - \bar{y}_{n-1})^2 \mathrm{E}(\hat{\rho}_n - 1)^2 + \mathrm{E}(z_n - \mu)^2 + o(\max\{n^{-1}, n^{-2/\alpha}\}). \tag{40}$$

The major difference between the right-hand sides of (35) and (40) is that the adjustment factor associated with the MSE of $\hat{\rho}_n$, $\mathrm{E}(y_n - \bar{y}_{n-1})^2$, of the latter grows at the rate n^2 whereas the former is bounded by a finite constant independent of n. In addition, the second terms on the right-hand sides of (40) and (35) are MSEs of two different consistent "estimators" of μ, namely, z_n and $(1-\rho)\bar{y}_{n-1}$, respectively. Note however that the MSE of z_n in the case of $\rho = 1$ and that of $(1-\rho)\bar{y}_{n-1}$ in the case of $0 \le \rho < 1$ are both asymptotically equivalent to σ^2/n. With the help of (40), we develop a unit-root counterpart of Theorem 3 in Theorem 4.

Theorem 4. *Assume (1) with* $\rho = 1$ *and (3). Suppose*

$$\mathrm{E}(\varepsilon_1^{q_2}) < \infty, \text{ for some } q_2 > 2. \tag{41}$$

Then,

$$\text{MSPE}_A - \sigma^2 = n^{-2/\alpha} L_1(\alpha, c) + \sigma^2 n^{-1} + o(\max\{n^{-1}, n^{-2/\alpha}\}), \tag{42}$$

where

$$L_1(\alpha, c) = 4^{-1}\Gamma((\alpha+2)/\alpha)\{\alpha(\alpha+1)/c\}^{2/\alpha}. \tag{43}$$

Equations (35), (37), (40) and (42) reveal that $\text{MSPE}_A - \sigma^2$ is mainly contributed by the MSE of a root-n consistent "estimator" of μ when $0 < \alpha < 2$, by the MSE of $\hat{\rho}_n$ times an adjustment factor when $\alpha > 2$, and by the sum of both MSEs when $\alpha = 2$. Equations (17), (21), (37) and (42) further indicate that while $\rho = 1$ can improve the convergence rate of the MSE of $\hat{\rho}_n$, there is no corresponding effect on the MSPE of $\hat{y}_{n+1}$. To explain this, we note that as shown in (35) and (40), the contribution of the MSE of $\hat{\rho}_n$ to $\text{MSPE}_A - \sigma^2$ needs to be adjusted for the variability of y_n, viz $\text{E}(y_n - \bar{y}_{n-1})^2$ when $\rho = 1$ and $\text{E}(y_n - \mu_y)^2$ when $0 \le \rho < 1$. Since $\text{E}(y_n - \mu_y)^2 = O(1)$ in the case of $0 \le \rho < 1$ and $\text{E}(y_n - \bar{y}_{n-1})^2 \sim n^2$ in the case of $\rho = 1$, the gain of the unit root in reducing the MSE of $\hat{\rho}_n$ is completely eliminated by the variability of y_n. Finally, we remark that the moment condition (36) [(41)] in Theorem 3 (Theorem 4) appears to be reasonable because $\sigma^2 < \infty$ and (16) [(20)] are needed in the proof of (37) [(42)].

3.2. The MSPE of $\tilde{y}_{n+1}$

In this section, we investigate the MSPE of $\tilde{y}_{n+1}$ under model (1). Note first that

$$y_{n+1} - \tilde{y}_{n+1} = \delta_{n+1} - \mathbf{x}_n^\top(\tilde{\mu}_n - \mu, \tilde{\rho}_n - \rho)^\top, \tag{44}$$

where $\mathbf{x}_n$ is defined after (9). Assume that (26) is true and (25) holds with $s > 12$. Then, by an argument similar to that used in Yu, Lin and Cheng (2012), one has for $0 \le \rho < 1$,

$$n^{1/2}\mathbf{x}_n^\top(\tilde{\mu}_n - \mu, \tilde{\rho}_n - \rho)^\top \Rightarrow \mathbf{x}^\top \mathbf{Z}_1, \tag{45}$$

where $\mathbf{x}$ has the same distribution as that of $(1, \sum_{j=0}^{\infty} \rho^j \varepsilon_{1-j})^\top$ and is independent of $\mathbf{Z}_1$ defined in Section 2, and for some $\eta > 0$,

$$\text{E}|n^{1/2}\mathbf{x}_n^\top(\tilde{\mu}_n - \mu, \tilde{\rho}_n - \rho)^\top|^{2+\eta} = O(1). \tag{46}$$

Combining (44)-(46) yields

$$\lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2) = 2\sigma^2. \tag{47}$$

For $\rho = 1$, it can similarly be shown that

$$n^{1/2}\mathbf{x}_n^\top(\tilde{\mu}_n - \mu, \tilde{\rho}_n - \rho)^\top \Rightarrow (1,\mu)\mathbf{Z}_2, \tag{48}$$

where $\mathbf{Z}_2$ is defined in Section 2, and

$$\lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2) = \mathrm{E}((1,\mu)\mathbf{Z}_2)^2 = 4\sigma^2. \tag{49}$$

Equations (47) and (49) indicate that the convergence rate of $\mathrm{MSPE}_B - \sigma^2$ and its corresponding limiting value, in sharp contrast to those of $\mathrm{MSPE}_A - \sigma^2$, are independent of the distributional properties of ε_1. Indeed, using arguments similar to those presented in Ing (2001) and Yu, Lin and Cheng (2012), we can link the second-order MSPE of the least squares predictor to the Fisher information matrix, $\sum_{j=1}^{n-1}\mathbf{x}_j\mathbf{x}_j^\top$, and express (47) and (49) in a unified way:

$$\lim_{n\to\infty}\frac{n(\mathrm{MSPE}_B - \sigma^2)}{\sigma^2} = \mathrm{plim}_{n\to\infty}\frac{\log\det(\sum_{j=1}^{n-1}\mathbf{x}_j\mathbf{x}_j^\top)}{\log n}, \tag{50}$$

noting that $\mathrm{plim}_{n\to\infty}(\log\det(\sum_{j=1}^{n-1}\mathbf{x}_j\mathbf{x}_j^\top)/\log n) = 2$ if $0 \le \rho < 1$, and 4 if $\rho = 1$. Since the right-hand side of (50) corresponds only to the *order of growth* of the determinant of the Fisher information matrix, it provides an explanation of why (i) $\lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2)$ is independent of the distributional properties of ε_1, and (ii) $\lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2)$ in the case of $\rho = 1$ is always larger than that in the case of $0 \le \rho < 1$.

According to (47), (49), (37) and (42), the ordering of MSPE_A and MSPE_B is asymptotically the same as for MSE_A and MSE_B. Therefore, with $\hat{\rho}_n$ and $\tilde{\rho}_n$ replaced by $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$, respectively, (R1) to (R4) can also be used to determine the more efficient predictor between $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$. [In the sequel, these *predictor* selection rules are still called "(R1) to (R4)".] Unfortunately, as noted at the end of Section 2, (R1) to (R4) cannot be implemented in practice unless ρ, α, c, μ and σ^2 are available.

4. PREDICTOR SELECTION BASED ON THE APE

The goal of this section is to choose the asymptotically more efficient predictor (estimator) between $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ ($\hat{\rho}_n$ and $\tilde{\rho}_n$) in a data-driven fashion. Instead of estimating the unknown parameters in (R1) to (R4) and then implementing these rules based on the estimated parameters, we suggest calculating APE_A and APE_B (defined in Section 1) and then choosing the predictor (estimator) with the smaller APE value. Through investigation of the asymptotic behaviors of APE_A and APE_B, we show in Theorems 5 and 6 below that the ordering of APE_A and APE_B is asymptotically equivalent to that of MSPE_A and MSPE_B, which provides the theoretical underpinning for the proposed predictor (estimator) selection method. Note that the asymptotic properties of APE based on least squares predictors have been extensively studied; see, among many others, Wei (1987, 1992), Hemerly and Davis (1989), Speed and Yu (1993), and Ing (2004, 2007). The major technical tools used in these papers are some recursive formulas developed for least squares estimates; see, for example, (2.8) of Wei (1987) and (3.10) of Ing (2004). While these formulas can be directly applied to APE_B, their extensions to APE_A seem difficult to obtain because, unlike $\tilde{\rho}_n$, $\hat{\rho}_n$ has no obvious recursive nature. We therefore take a quite different approach to dealing with the asymptotic properties of APE_A, as detailed in the Appendix. Let

$$L(f_{\varepsilon_1}(\cdot), \alpha, \rho) = \sigma^2 I_{\{0<\alpha\leq 2, 0\leq\rho\leq 1\}} + L_0(\alpha, c, \rho, \sigma^2) I_{\{\alpha\geq 2, 0\leq\rho<1\}} + L_1(\alpha, c) I_{\{\alpha\geq 2, \rho=1\}}, \tag{51}$$

where $L_0(\alpha, c, \rho, \sigma^2)$ and $L_1(\alpha, c)$ are given in (38) and (43), respectively, and define

$$L(\sigma^2, \rho) = 2\sigma^2 I_{\{0\leq\rho<1\}} + 4\sigma^2 I_{\{\rho=1\}}. \tag{52}$$

Theorem 5. *Assume (1) with $0 \leq \rho \leq 1$, (3) and (36). Let M_1 in APE_A be any integer ≥ 2 and remain fixed as n increases. Then, for $0 < \alpha \leq 2$,*

$$\text{APE}_A = \sum_{t=M_1}^{n-1} \delta_{t+1}^2 + L(f_{\varepsilon_1}(\cdot), \alpha, \rho) \log n + o_p(\log n), \tag{53}$$

where $L_0(\alpha, c, \rho, \sigma^2)$ and $L_1(\alpha, c)$ are defined in (38) and (43), respectively, and for $\alpha > 2$,

$$\text{APE}_A = \sum_{t=M_1}^{n-1} \delta_{t+1}^2 + C_n, \tag{54}$$

where C_n satisfies $\lim_{n\to\infty} \mathrm{P}(C_n/\log n > K) = 1$ *for any $K < \infty$.*

Theorem 6. *Assume (1) with $0 \leq \rho \leq 1$, (3) and (41). Let M_2 in APE_B be the first integer j such that the LSE $(\tilde{\mu}_j, \tilde{\rho}_j)^\top$ is uniquely defined. Then,*

$$\mathrm{APE}_B = \sum_{t=M_2}^{n-1} \delta_{t+1}^2 + L(\sigma^2, \rho) \log n + o_p(\log n). \tag{55}$$

It is worth pointing out that the coefficient associated with the $\log n$ term in (53) is exactly the same as the one associated with the n^{-1} term in (37) if $0 \leq \rho < 1$, and (42) if $\rho = 1$. The same correspondence emerges between (55) and (47) when $0 \leq \rho < 1$, and (55) and (49) when $\rho = 1$. These coincidences, together with (54), yield (i) for $\alpha > 2$, $\lim_{n\to\infty} \mathrm{P}(\mathrm{APE}_B < \mathrm{APE}_A) = 1$, (ii) for $0 < \alpha < 2$, $\lim_{n\to\infty} \mathrm{P}(\mathrm{APE}_A < \mathrm{APE}_B) = 1$, (iii) for $\alpha = 2$, $0 \leq \rho < 1$ and $c < 2(1-\rho)\{(1+\rho)\mu^2 + (1-\rho)\sigma^2\}^{-1}$, $\lim_{n\to\infty} \mathrm{P}(\mathrm{APE}_B < \mathrm{APE}_A) = 1$, (iv) for $\alpha = 2$, $0 \leq \rho < 1$ and $c > 2(1-\rho)\{(1+\rho)\mu^2 + (1-\rho)\sigma^2\}^{-1}$, $\lim_{n\to\infty} \mathrm{P}(\mathrm{APE}_A < \mathrm{APE}_B) = 1$, (v) for $\alpha = 2$, $\rho = 1$ and $c < 1/(2\sigma^2)$, $\lim_{n\to\infty} \mathrm{P}(\mathrm{APE}_B < \mathrm{APE}_A) = 1$, and (vi) for $\alpha = 2$, $\rho = 1$ and $c > 1/(2\sigma^2)$, $\lim_{n\to\infty} \mathrm{P}(\mathrm{APE}_A < \mathrm{APE}_B) = 1$. Combining (i) to (vi) leads to the following sample counterpart of (R1) to (R4):

(S1) Choose $\hat{y}_{n+1}$ $(\hat{\rho}_n)$ if $\mathrm{APE}_A < \mathrm{APE}_B$.

(S2) Choose $\tilde{y}_{n+1}$ $(\tilde{\rho}_n)$ if $\mathrm{APE}_A > \mathrm{APE}_B$.

(S3) Choose either $\hat{y}_{n+1}$ or $\tilde{y}_{n+1}$ (either $\hat{\rho}_n$ or $\tilde{\rho}_n$) if $\mathrm{APE}_A = \mathrm{APE}_B$.

(The issue of determining M_1 and M_2 for APE_A and APE_B from a finite-sample point of view will be addressed in the next section.)

The following theorem, which is an immediate consequence of Theorems 5 and 6, confirms the validity of (S1) to (S3). Let $\check{y}_{n+1}(S)$ and $\check{y}_{n+1}(R)$ [$\check{\rho}_n(S)$ and $\check{\rho}_n(R)$] denote the predictors (estimators) selected by (S1) to (S3) and (R1) to (R4), respectively.

Theorem 7. *Assume that the assumptions of Theorems 5 and 6 hold. Then,*

$$\lim_{n\to\infty} \mathrm{P}(\check{y}_{n+1}(S) = \check{y}_{n+1}(R)) = 1 \textit{ and } \lim_{n\to\infty} \mathrm{P}(\check{\rho}_n(S) = \check{\rho}_n(R)) = 1.$$

Some remarks regarding Theorem 7 are in order.

Remark 5. Theorem 7 reveals that (S1) to (S3) can ultimately choose the more efficient predictor (estimator) between $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ ($\hat{\rho}_n$ and $\tilde{\rho}_n$) regardless of whether $0 \leq \rho < 1$ or $\rho = 1$. However, this goal is not directly relevant to unit root tests because, as shown in Section 3 (Section 2), the relative performance of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ ($\hat{\rho}_n$ and $\tilde{\rho}_n$) is determined not only by ρ, but also by α and c. Moreover, while it is possible to estimate c and α from some kernel density estimators of $f_{\varepsilon_1}(x)$ based on the AR residuals, asymptotic behaviors of the resultant estimators of c and α are usually difficult to derive, particularly when $\alpha \in (0, 1]$ (which yields that $f_{\varepsilon_1}(x)$ is nonzero or has a pole at the origin).

Remark 6. Consider a trend stationary model,

$$y_t = \rho y_{t-1} + \beta + \gamma t + \delta_t, \tag{56}$$

where $0 \leq \rho < 1$, $\gamma > 0$ and β is large enough such that y_t's are always positive. Model (56) can be expressed as

$$y_t = \beta^* + \gamma^* t + \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}, \tag{57}$$

where $\beta^* = (\beta - \mu)/(1-\rho) - \gamma\rho/(1-\rho)^2$ and $\gamma^* = \gamma/(1-\rho)$. Unlike the variance of model (1) with $\rho = 1$, which grows to infinity as t does, the variance of model (57) does not vary with t. On the other hand, a linear time trend is a dominant feature shared by both models. As a result, when data are truly generated from model (56), $\hat{\rho}_n$ and $\tilde{\rho}_n$ [defined in (2) and (9)] become upwardly biased estimators of ρ. Moreover, the unit root tests (like those of Dickey and Fuller 1981) for $\rho = 1$ and $\gamma = 0$ in (56) tend to be biased towards accepting $\rho = 1$ even if $\rho < 1$. Fortunately, this dilemma is not an issue of overriding concern from a prediction point of view. More specifically, if it is unknown whether the data are generated from model (1) with $0 \leq \rho \leq 1$ or model (56) and the goal is to choose the best predictor among $\hat{y}_{n+1}$, $\tilde{y}_{n+1}$ and y^*_{n+1} in terms of MSPE, where y^*_{n+1} is the least squares predictor of y_{n+1} based on (56) and $y_1, \ldots y_n$, then one only needs to compute APE_A, APE_B and $\text{APE}_C = \sum_{i=M_3}^{n-1}(y_{i+1} - y^*_{i+1})^2$, and choose the predictor with the smallest APE value, where M_3 is the first integer j such that the LSE associated with y^*_{j+1} is uniquely defined. In fact, by an argument similar to that used to prove Theorems 5 and 6, it can be shown that the predictor chosen in this manner is asymptotically equivalent to the best one among $\hat{y}_{n+1}$, $\tilde{y}_{n+1}$ and y^*_{n+1}.

Remark 7. It is also possible to improve prediction performance of (S1) to (S3) through the following modified procedure: (i) (S1) to (S3) are used to choose the better predictor between $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ and the better estimator between $\hat{\rho}_n$ and $\tilde{\rho}_n$, (ii) the residuals produced by the better estimator are approximated by a parametric family of distributions and with this family of distributions, a new estimator of ρ and a new predictor of y_{n+1}, denoted by $\dot{\rho}_n$ and $\dot{y}_{n+1}$, respectively, are derived, (iii) the APE associated with $\dot{y}_{n+1}$, say APE_D, is computed and the final predictor is $\dot{y}_{n+1}$ if $\mathrm{APE}_D < \min\{\mathrm{APE}_A, \mathrm{APE}_B\}$, and is the one chosen in (i) otherwise. An investigation of the extent to which this modified procedure performs satisfactorily would be interesting future work.

5. NUMERICAL STUDIES

We now present some simulations in support of the theoretical results established for (S1) to (S3). The usefulness of (S1) to (S3) is also illustrated by analyzing three time series data sets.

5.1 Finite sample performance of (S1) to (S3)

We generate $n \in \{200, 500, 1000, 2000\}$ observations from model (1), where $\rho \in \{0.2, 0.5, 0.8, 1\}$ and ε_t's are i.i.d. Gamma(α, θ) or Beta(α, θ) distributions, with $\alpha \in \{1, 2, 4\}$ and $\theta \in \{1, 2, 6\}$. The pdfs of Gamma(α, θ) and Beta(α, θ) distributions are given by $\Gamma^{-1}(\alpha)x^{\alpha-1}\theta^{-\alpha}\exp(-x/\theta)$ and $\Gamma(\alpha+\theta)\Gamma^{-1}(\alpha)\Gamma^{-1}(\theta)x^{\alpha-1}(1-x)^{\theta-1}$, respectively. For each combination of ρ's and distributions, we ran 500 simulations and recorded the ratios, $F_n^{(500)} = (1/500)\sum_{l=1}^{500} I_{\{\breve{y}^{(l)}_{n+1}(S)=\breve{y}_{n+1}(R)\}}$, in Tables 1 to 3, where $\breve{y}^{(l)}_{n+1}(S)$ denotes the predictor selected by (S1) to (S3) in the lth simulation run. Obviously, a larger $F_n^{(500)}$ means a better performance of (S1) to (S3). In fact, since $F_n^{(500)}$ is an empirical estimate of $\mathrm{P}(\breve{y}_{n+1}(S) = \breve{y}_{n+1}(R))$, Theorem 7 suggests that $F_n^{(500)}$ will be close to 1 when n is large. The M_1 and M_2 in APE_A and APE_B are both set to 20. As will be seen later, these specifications can lead to quite satisfactory performances except in some difficult cases. To help illustrate the finite-sample performance of (S1) to (S3), the empirical estimates of $n^{b(\alpha)}(\text{MSPE}_A - \sigma^2)$, denoted by $\hat{S}_{A,n}$, are also obtained for each coefficient/distribution combination, where $b(\alpha) = 1$ if $0 < \alpha \leq 2$ and $2/\alpha$ if $\alpha > 2$. The closeness of $\hat{S}_{A,n}$ to its limiting value, namely $L(f_{\varepsilon_1}(\cdot), \alpha, \rho)$ given in (51), is measured by the ratio $R_{A,n} = \hat{S}_{A,n}/L(f_{\varepsilon_1}(\cdot), \alpha, \rho)$.

Tables 1 and 2 summarize $F_n^{(500)}$ for $\alpha = 1$ and $\alpha = 2$, respectively. The performance of (S1) to (S3) under these settings is affected mainly by

$$P(f_{\varepsilon_1}(\cdot), \alpha, \rho) = \frac{\max\{L(f_{\varepsilon_1}(\cdot), \alpha, \rho), L(\sigma^2, \rho)\}}{\min\{L(f_{\varepsilon_1}(\cdot), \alpha, \rho), L(\sigma^2, \rho)\}},$$

and $R_{A,200}$, where $L(\sigma^2, \rho)$ is defined in (52). Note that $P(f_{\varepsilon_1}(\cdot), \alpha, \rho)$ measures the "degree of distinguishability" between the prediction capabilities of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$, and a large (small) value of $P(f_{\varepsilon_1}(\cdot), \alpha, \rho)$ represents that it is easy (difficult) to tell the difference between the two predictors. On the other hand, $R_{A,200}$ is related to the convergence rate of $n\{\text{E}(y_{n+1} - \hat{y}_{n+1})^2 - \sigma^2\}$. A large $|R_{A,200} - 1|$ suggests that the behaviors of the squared prediction errors, $(y_{i+1} - \hat{y}_{i+1})^2$, are relatively difficult to control for small or even moderate i. These uncontrollable early errors can actually dominate APE_A if the sample size n is not large enough (see Section 5 of Wei (1992) for a related discussion), and make the ordering of APE_A and APE_B different from that of $L(f_{\varepsilon_1}(\cdot), \alpha, \rho)$ and $L(\sigma^2, \rho)$, thereby deteriorating the performance of (S1) to (S3).

To get a better understanding of the role played by $R_{A,200}$ and $P(f_{\varepsilon_1}(\cdot), \alpha, \rho)$ in affecting the finite sample performance of (S1) to (S3), we classify the pairs, $(R_{A,200}, P(f_{\varepsilon_1}(\cdot), \alpha, \rho))$, into six categories: (G, S), (G, M), (G, L), (B, S), (B, M) and (B, L), where $R_{A,200} \in$ G if $|R_{A,200} - 1| < 0.1$, $R_{A,200} \in$ B if $|R_{A,200} - 1| \geq 0.1$, $P(f_{\varepsilon_1}(\cdot), \alpha, \rho) \in$ S if $P(f_{\varepsilon_1}(\cdot), \alpha, \rho) < 1.4$, $P(f_{\varepsilon_1}(\cdot), \alpha, \rho) \in$ M if $1.4 \leq P(f_{\varepsilon_1}(\cdot), \alpha, \rho) < 1.9$, and $P(f_{\varepsilon_1}(\cdot), \alpha, \rho) \in$ L if $P(f_{\varepsilon_1}(\cdot), \alpha, \rho) > 1.9$. Each coefficient/distribution combination belongs to one category, which is also recorded in Tables 1 and 2. According to the above discussion, (S1) to (S3) should perform well in category (G, L) or (G, M), but poorly in category (B, S) or (B, M). Our simulation results confirm this conjecture. Specifically, when $n = 200$ and $(R_{A,200}, P(f_{\varepsilon_1}(\cdot), \alpha, \rho)) \in$ (G, L), (S1) to (S3) choose $\check{y}_{n+1}(R)$ about 76%-92% of the time for $0 < \rho < 1$, and over 93% of the time for $\rho = 1$. This ratio exceeds 93% in both stationary and unit root cases as n increases to 2000. In the (G, M) category, (S1) to (S3) perform slightly worse, but they can still identify $\check{y}_{n+1}(R)$ between 81%−93% of the time based on 2000 observations. On the other hand, (S1) to (S3) in the (B, M) category only select $\check{y}_{n+1}(R)$ for 30%-80% of the simulations even when $n = 2000$ (noting that there is no coefficient/distribution combination falling into the (B, S) category). The performance of (S1) to (S3) in the (G, S) category is also poor since the corresponding ratio for identifying $\check{y}_{n+1}(R)$ falls between 47% and 67% when $n = 200$, and between 53% and 67% when $n = 2000$. In contrast, (S1) to (S3) give quite satisfactory results in the (B, L) category. In particular, when $n = 2000$, they choose $\check{y}_{n+1}(R)$ about 81%-92% of the time in the stationary case, with the percentage exceeding 96% in the unit root case.

It is worth mentioning that a low percentage of identifying $\check{y}_{n+1}(R)$ in the (G, S) category does not necessarily lead to serious repercussions because the prediction capabilities of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ are very similar in this category. On the other hand, improving the performance of (S1) to (S3) in the (B, M) category seems to be a valid goal to pursue. One natural way to tackle this problem is to pick a larger M_1 for APE_A such that the effects of "early uncontrollable errors" (which are reflected by a large value of $|R_{A,200} - 1|$) are diminished, thereby turning "B" into "G". Indeed, our (unreported) simulation results show that under $\rho = 0.8$ and $\varepsilon_1 \sim \beta(2, \theta)$, $F_{2000}^{(500)}$, with $\theta = 1$, increases from 0.3 to 0.7 as M_1 and M_2 increase to 200; $F_{2000}^{(500)}$, with $\theta = 2$, increases from 0.546 to 0.78 as M_1 and M_2 increase to 80; and $F_{2000}^{(500)}$, with $\theta = 6$, increases from 0.732 to 0.866 as M_1 and M_2 also increase to 80. These findings suggest that in the (B, M) category, the choice of M_1 and M_2 is quite relevant from a finite sample point of view.

Table 3 summarizes the values of $F_n^{(500)}$ for $\alpha = 4$. Note that the ratio $P(f_{\varepsilon_1}(\cdot), \alpha, \rho)$ is no longer meaningful because $\text{MSPE}_A - \sigma^2$ and $\text{MSPE}_B - \sigma^2$ have different convergence rates when $\alpha > 2$. On the other hand, $R_{A,200}$ still has an important impact on the finite sample performance of (S1) to (S3), and hence the values of $R_{A,200}$ are included in Table 3. The performance of (S1) to (S3) is satisfactory under Beta$(4, \theta)$ errors. For example, in the case of $0 < \rho < 1$, they can identify $\check{y}_{n+1}(R)$ between 60%-99% of the time when $n = 200$, and 81%-

100% of the time when $n = 2000$. (S1) to (S3) deliver particularly good results under the unit root model with Beta$(4, \theta)$ errors since they choose $\check{y}_{n+1}(R)$ in all simulations from $n = 200$ to $n = 2000$. The performance of (S1) to (S3) under Gamma$(4, \theta)$ errors, however, depends on the value of $R_{A,200}$, and becomes worse as $R_{A,200}$ becomes larger. More specifically, in the case of $R_{A,200} \leq 2.11$, (S1) to (S3) identify $\check{y}_{n+1}(R)$ between 67% and 99% of the simulations when $n = 500$, and between 81% and 100% of the simulations when $n = 2000$. In contrast, if $R_{A,200} \geq 2.4$, (S1) to (S3) only select $\check{y}_{n+1}(R)$ between 51% and 66% of the time even when $n = 2000$. Since under Gamma(4, θ) errors, not only $R_{A,200}$'s but also $R_{A,2000}$'s are significantly larger than 1, it seems difficult to enhance the performance of (S1) to (S3) through choosing a larger M_1 or M_2. Nevertheless, we have found that the frequency of correct identification of (S1) to (S3) in the case of $R_{A,200} \geq 2.4$ can improve to 75%, provided n increases to 20000.

5.2 Data Analysis

We analyze three positive-valued time series with sample sizes of $n = 45$, 82 and 71 to demonstrate the usefulness of our predictor selection method. The first series, reported on page 134 of Burr (1976), consists of 45 daily average number of defects per truck found in the final inspection at the end of the assembly line of a truck manufacturing plant. Example 6.1 of Wei (2006) indicates that a stationary AR(1) model may be suitable for this series. The second series is a laboratory series of blowfly data taken from Nicholson (1950). A fixed number of adult blowflies with balanced sex ratios were kept in a cage and given a fixed amount of food daily. The blowfly population was then enumerated every other day for approximately two years, giving a total of 364 observations. Following Example 6.3 of Wei (2006), we only use the latest 82 data points in our analysis. It is shown in this example that a stationary AR(1) model also fits these data quite well. The third series is the yearly cancer death rate (per 100,000 population) of Pennsylvania between 1930 and 2000 published in the 2000 Pennsylvania Vital Statistics Annual Reports by the Pennsylvania Department of Health. Example 6.5 of Wei (2006) shows that while this series is clearly nonstationary with an increasing trend, it still exhibits first-order autoregressive behavior. Plots of the data and the sample autocorrelation and the partial autocorrelation functions (ACFs and PACFs) given in Figure 1 also demonstrate the appropriateness of AR(1) models for each series. Note that these ACFs and PACFs are computed in terms of correlations rather than covariances because, as shown in Nielsen (2006), the latter may be inappropriate for non-stationary time series. In view of the features of these series, it would be interesting to ask whether (S1) to (S3) can choose the better predictor between $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ for each series.

To investigate this question, we split each series, $\{y_1, \cdots, y_n\}$, into two parts. The first part contains the first 90% of the data points, $\{y_1, \cdots, y_{T_n}\}$, which are used for deriving $\text{APE}_A(T_n) = \sum_{i=M_1}^{T_n-1}(y_{i+1} - \hat{y}_{i+1})^2$ and $\text{APE}_B(T_n) = \sum_{i=M_2}^{T_n-1}(y_{i+1} - \tilde{y}_{i+1})^2$, and implementing (S1) to (S3) based on $\text{APE}_A(T_n)$ and $\text{APE}_B(T_n)$, where $T_n = \lfloor n \times 0.9 \rfloor$ with $\lfloor x \rfloor$ denoting the largest integer $\leq x$. The second part contains $n - T_n$ data points, which are reserved for calculating empirical MSPEs (EMSPEs), $\text{EMSPE}_A = (n - T_n)^{-1} \sum_{i=T_n}^{n-1}(y_{i+1} - \hat{y}_{i+1})^2$ and $\text{EMSPE}_B = (n - T_n)^{-1} \sum_{i=T_n}^{n-1}(y_{i+1} - \tilde{y}_{i+1})^2$. In light of the simulation results given in Section 5.1, both M_1 and M_2 in $\text{APE}_A(T_n)$ and $\text{APE}_B(T_n)$ are set to 20 for all three datasets. If the sign of $\text{APE}_A(T_n) - \text{APE}_B(T_n)$ is consistent with that of $\text{EMSPE}_A - \text{EMSPE}_B$, then (S1) to (S3) will choose the predictor with the smaller EMSPE. Since EMSPE_A and EMSPE_B are suitable surrogates for the unobservable MSPE_A and MSPE_B and since these types of errors are frequently used in practice to make forecast evaluations (Clark and West 2004), (S1) to (S3) are considered satisfactory if the "sign consistency" phenomenon mentioned above is observed in each time series.

Our calculations show that in the first series $\text{APE}_A(T_n) = 4.852 < \text{APE}_B(T_n) = 5.062$ and $\text{EMSPE}_A = 0.0145 < \text{EMSPE}_B = 0.0148$, in the second series $\text{APE}_A(T_n) = 4.901 \times 10^7 > \text{APE}_B(T_n) = 4.894 \times 10^7$ and $\text{EMSPE}_A = 386349.8 > \text{EMSPE}_B = 360928.1$, and in the third series $\text{APE}_A(T_n) = 334.3 < \text{APE}_B(T_n) = 348.7$ and $\text{EMSPE}_A = 12.986 < \text{EMSPE}_B = 17.023$. Therefore, the sign consistency phenomenon appears in each series, in spite of the sign being negative in the first and third series and positive in the second. In summary, our data analysis reveals that (S1) to (S3) can successfully combine the strengths of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$, thereby producing better prediction results.

APPENDIX: PROOFS OF THEOREMS 5 and 6

To prove Theorem 5, we need the following lemma, whose proof is given in the supplementary document. This lemma provides sufficient conditions for a sequence of random variables $\{d_n\}$ to satisfy

$$\text{plim}_{n\to\infty}(\log n)^{-1}\sum_{i=1}^{n} d_i = S, \tag{A.1}$$

where plim denotes the probability limit and S is some constant.

Lemma 1. *Let* $\{d_n\}$ *be a sequence of random variables. Assume* (i) *there exist a positive integer* M *and a positive number* $\bar{U}$ *such that for all* $n \geq M$*,* $\mathrm{E}(n^2 d_n^2) \leq \bar{U}$*;* (ii) $\lim_{n\to\infty}\mathrm{E}(nd_n) = S$*; and* (iii) *there exist* $0 < \eta < 1$ *and a sequence of positive numbers* $\{a_n\}$*, with* $\lim_{n\to\infty} a_n = \infty$ *and* $\log a_n/\log n = o(1)$*, such that* $\max_{M\leq i\leq n/a_n,\, i\eta a_n\leq j\leq n}|\mathrm{E}\left(Q_iQ_j\right)| = o(1)$*, where* $Q_i = \{id_i - \mathrm{E}(id_i)\}$*. Then (A.1) follows.*

We are now ready to prove Theorem 5. We first focus on the case of $\rho = 1$. By Chow (1965) and straightforward calculations, $\sum_{i=2}^{n-1}(y_{i+1}-\hat{y}_{i+1})^2 = \sum_{i=2}^{n-1}\delta_{i+1}^2 + \sum_{i=2}^{n-1}(y_{i+1}-\hat{y}_{i+1}-\delta_{i+1})^2(1+o(1)) + O(1)$ a.s., and $\sum_{i=2}^{n-1}(y_{i+1}-\hat{y}_{i+1}-\delta_{i+1})^2 = \sum_{i=2}^{n-1}(z_i-\mu)^2 + \sum_{i=2}^{n-1}(\hat{\rho}_1-1)^2(y_i-\bar{y}_{i-1})^2 + 2\sum_{i=2}^{n-1}(\hat{\rho}_1-1)(y_i-\bar{y}_{i-1})(z_i-\mu)$. In addition, Theorem 2 of Wei (1987) (with $\mathbf{x}_i = 1, \mathbf{b}_i = z_i$ and $\boldsymbol{\beta} = \mu$) implies $\sum_{i=2}^{n-1}(z_i-\mu)^2 = \sigma^2\log n + o_p(\log n)$. In view of these identities, Theorem 5 follows if one can show that for $\alpha = 2$,

$$G_{1,n} + G_{2,n} = 3(2c)^{-1}\log n + o_p(\log n), \tag{A.2}$$

for $0 < \alpha < 2$,

$$G_{1,n} + G_{2,n} = o_p(\log n), \tag{A.3}$$

and for $\alpha > 2$,

$$\text{plim}_{n\to\infty}(G_{1,n} + G_{2,n})/\log n = \infty, \tag{A.4}$$

where $G_{1,n} = \sum_{i=2}^{n-1}(\hat{\rho}_i-1)^2(y_i-\bar{y}_{i-1})^2$ and $G_{2,n} = 2\sum_{i=2}^{n-1}(\hat{\rho}_i-1)(y_i-\bar{y}_{i-1})(z_i-\mu)$.

We start by proving the most challenging result (A.2). Let M be the smallest positive integer such that for all $i \geq M$, $\mathrm{E}(i^{3/2}(\hat{\rho}_i-1))^4 < \infty$ [noting that the existence of such an M

is guaranteed by (19)]. It is not difficult to see that (A.2) is ensured by

$$\sum_{i=M}^{n-1} (\hat{\rho}_i - 1)^2 (y_i - \bar{y}_{i-1})^2 = 3(2c)^{-1} \log n + o_p(\log n), \tag{A.5}$$

and

$$\sum_{i=M}^{n-1} (\hat{\rho}_i - 1)(y_i - \bar{y}_{i-1})(z_i - \mu) = o_p(\log n). \tag{A.6}$$

Applying (19), (36), Hölder's inequality and Chebyshev's inequality, one has

$$\sum_{i=M}^{n-1} (\hat{\rho}_i - 1)^2 (y_i - \bar{y}_{i-1})^2 = (\mu^2/4) \sum_{i=M}^{n-1} \{i^{3/2}(\hat{\rho}_i - 1)\}^2/i + o_p(\log n). \tag{A.7}$$

Let $d_i = \{i^{3/2}(\hat{\rho}_i - 1)\}^2/i$, $S_i = \mathrm{E}(id_i)$ and $Q_i = id_i - S_i$. By (19), (36) and (21), it holds that $\mathrm{E}(n^2 d_n^2) = O(1)$ and $\lim_{n\to\infty} S_n = 6/(c\mu^2)$. Hence conditions (i) and (ii) of Lemma 1 are satisfied. If we can further show that condition (iii) of Lemma 1 holds, then

$$\sum_{i=M}^{n-1} \{i^{3/2}(\hat{\rho}_i - 1)\}^2/i = \mu^{-2} c^{-1} 6 \log n + o_p(\log n), \tag{A.8}$$

which, together with (A.7), leads to (A.5).

Let $\lim_{n\to\infty} a_n = \infty, \log a_n / \log n = o(1)$ and $0 < \eta < 1$. Also choose a sequence of positive numbers $\{g_n\}$ satisfying $\lim_{n\to\infty} g_n = \infty$ and $g_n/a_n = o(1)$. Define $Q_{j,i} = jd_{j,i} - S_j$ and $\tilde{Q}_{j,i} = j\tilde{d}_{j,i} - S_j$, where $d_{j,i} = \min_{ig_n \le s \le j} (j\varepsilon_s / y_{s-1})^2$ and $\tilde{d}_{j,i} = \min_{ig_n \le s \le j} (j\varepsilon_s / y_{s-1,i})^2$, with $y_{l,i} = \sum_{r=0}^{l-i-1} \varepsilon_{l-r}$. By (19), (36), Hölder's inequality and some algebraic manipulations, one obtains

$$\max_{M \le i \le n/a_n,\, \eta i a_n \le j \le n} |\mathrm{E}\{Q_i(Q_j - Q_{j,i})\}| = o(1),$$

and

$$\max_{M \le i \le n/a_n,\, \eta i a_n \le j \le n} |\mathrm{E}\{Q_i(Q_{j,i} - \tilde{Q}_{j,i})\}| = o(1).$$

In addition, $\max_{M \le i \le n/a_n,\, \eta i a_n \le j \le n} |\mathrm{E}(Q_i \tilde{Q}_{j,i})| = o(1)$ follows from $\mathrm{E}(Q_i) = 0$ and independence between Q_i and $\tilde{Q}_{j,i}$. As a result, condition (iii) of Lemma 1 holds true. This completes the proof of (A.8) and therefore of (A.5). To deal with (A.6), let $d_i = (\hat{\rho}_i - 1)(y_i - \bar{y}_{i-1})(z_i - \mu)$. It is shown in the supplementary document that $\mathrm{E}(nd_n) = o(1)$. Using this equation in conjunction with (19), (36), Hölder's inequality and an argument similar to that used to prove (A.8), one obtains (A.1) with $S = 0$, which results in (A.6). Consequently, (A.2) is proved.

By making use of (19), it is not difficult to verify (A.3). We therefore skip the details. To show (A.4), note first that $G_{1,n} \geq (\hat{\rho}_n - 1)^2 \sum_{i=2}^{n-1} (y_i - \bar{y}_{i-1})^2$. This inequality and Theorem 2 together imply that for any $0 < \epsilon_1 < 1$, there exists $\epsilon_2 > 0$ such that

$$\liminf_{n\to\infty} \mathrm{P}(G_{1,n}/n^{1-(2/\alpha)} > \epsilon_2) \geq 1 - \epsilon_1. \tag{A.9}$$

In addition, $G_{2,n} = o_p(n^{1-(2/\alpha)})$ follows from (36), (19) and Hölder's inequality. Combining this with (A.9) leads to (A.4). As a result, the proof of Theorem 5 for $\rho = 1$ is complete. The proof of Theorem 5 for $0 \leq \rho < 1$ is similar to that for $\rho = 1$, and is thus omitted. Finally, we note that Theorem 6 is an immediate consequence of Theorem 2 of Wei (1987).

SUPPLEMENTARY MATERIALS

Supplementary materials contain the proofs of Theorems 1-4 and Lemma 1.

ACKNOWLEDGMENTS

We would like to thank the handling editor, the associate editor and two anonymous reviewers for their insightful and constructive comments, which greatly improve the presentation of this paper.

REFERENCES

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), "Infrence for non-negative autoregressive schemes," *Communications in Statistics: Theory and Methods*, **15**, 2267-2293.

Burr, I. W. (1976), *Statistical Quality Control Methods*, New York: Marcel Dekker.

Chan, N. H. (1989), "Asymptotic inference for unstable autoregressive time series with drifts, " *Journal of Statistical Planning and Inference*, **23**, 301-312.

Chow, Y. S. (1965), "Local convergence of martingales and the law of large numbers," *The Annals of Mathematical Statistics*, **36**, 552-558.

Clark, T. E. and West, K. D. (2006), "Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis," *Journal of Econometrics*, **135**, 155-186.

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.

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

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

Hamilton, J. D. (1994), *Time series analysis*, Princeton University Press.

Hemerly, E. M. and Davis, M. H. A. (1989), "Strong consistency of the pls criterion for order determination of autoregressive processes," *The Annals of Statistics*, **17**, 941-946.

Hutton, J.L. (1990), "Non-negative time series models for dry river flow," *Journal of Applied Probability*, **27**, 171-182.

Ing, C.- K. (2001), "A note on mean-squared prediction errors of the least squares predictors in random walk models," *Journal of Time Series Analysis*, **22**, 711-724.

———— (2004), "Selecting optimal multistep predictors for autoregressive processes of unknown order," *The Annals of Statistics*, **32**, 693-722.

———— (2007), "Accumulated prediction errors, information criteria and optimal forecasting for autoregressive time series," *The Annals of Statistics*, **35**, 1238-1277.

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.

Lewis, P.A.W., McKenzie, E. and Hugus, D. K. (1989), "Gamma processes," *Communications in Statistics: Stochastic Models*, **5**, 1-30.

Nicholson, A. J. (1950), "Population oscillations caused by competition for food," *Nature*, **165**, 476-477.

Nielsen, B. (2006), "Correlograms for non-stationary autoregressions," *Journal of the Royal Statistical Society, Ser. B*, **68**, 707-720.

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.

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

Sim, C. H. (1987), "A mixed gamma ARMA(1,1) model for river flow time series," *Water Resources Research*, **23**, 32-36.

Speed, T. P. and Yu, B. (1993), "Model selection and prediction: normal regression," *Annals of the institute of statistical mathematics*, **45**, 35-54.

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

——— (1992), "On predictive least squares principles," *The Annals of Statistics*, **20**, 1-42.

Wei, W. W. S. (2006), *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.), Boston: Pearson Addison Wesley.

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.

Table 1: $F_n^{(500)}$ and the categories of $(R_{A,200}, P(f_{\varepsilon_1}(\cdot), \alpha, \rho))$ under model (1) with Gamma(1,θ) or Beta$(1,\theta)$ errors

		Gamma(1, θ) errors					$\beta(1,\theta)$ errors				
ρ	θ	n =200	n =500	n =1000	n =2000	Category	n =200	n =500	n =1000	n =2000	Category
0.2	1	0.904	0.948	0.960	0.954	(G,L)	0.906	0.944	0.954	0.982	(G,L)
	2	0.916	0.922	0.954	0.970	(G,L)	0.912	0.934	0.968	0.966	(G,L)
	6	0.906	0.932	0.958	0.988	(G,L)	0.896	0.934	0.968	0.970	(G,L)
0.5	1	0.916	0.948	0.968	0.986	(G,L)	0.880	0.936	0.960	0.976	(G,L)
	2	0.900	0.938	0.964	0.976	(G,L)	0.882	0.942	0.950	0.984	(G,L)
	6	0.898	0.946	0.972	0.978	(G,L)	0.892	0.956	0.968	0.966	(G,L)
0.8	1	0.764	0.846	0.914	0.938	(G,L)	0.626	0.662	0.778	0.812	(B,L)
	2	0.818	0.858	0.912	0.956	(G,L)	0.668	0.772	0.854	0.866	(B,L)
	6	0.816	0.854	0.912	0.938	(G,L)	0.740	0.840	0.870	0.920	(B,L)
1.0	1	0.986	1.000	1.000	1.000	(G,L)	0.932	0.974	0.988	0.998	(B,L)
	2	0.988	0.998	1.000	1.000	(G,L)	0.972	0.998	0.998	1.000	(G,L)
	6	0.996	0.996	0.998	1.000	(G,L)	0.984	0.998	1.000	1.000	(G,L)

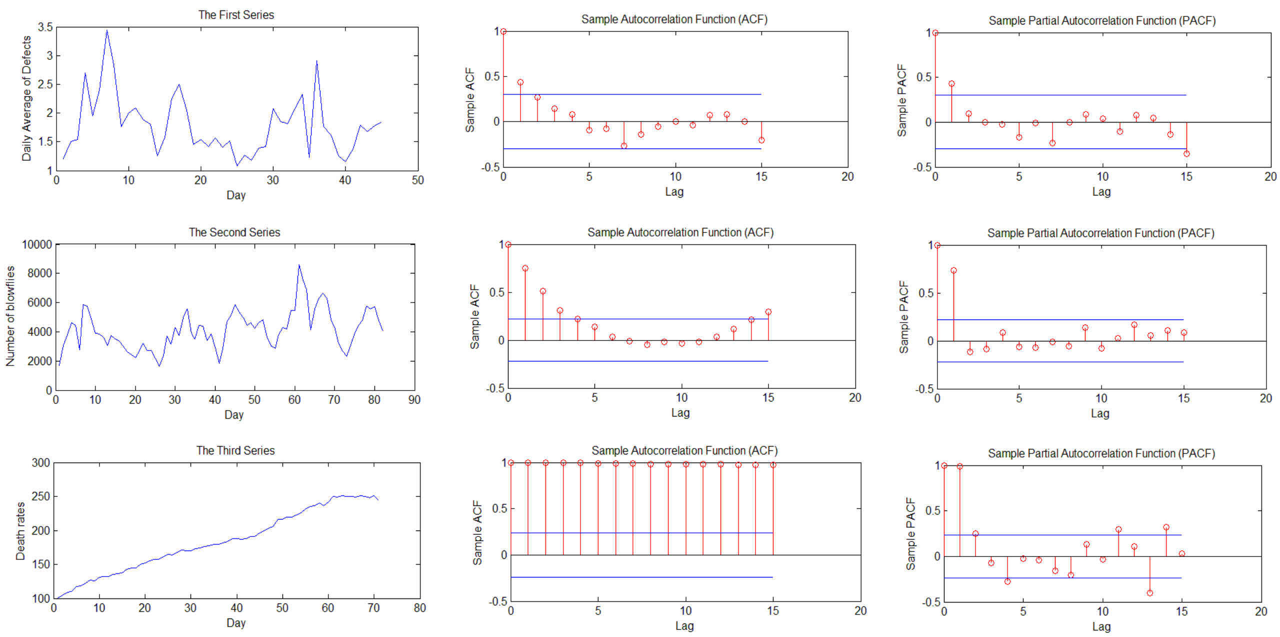
Table 2: $F_n^{(500)}$ and the categories of $(R_{A,200}, P(f_{\varepsilon_1}(\cdot), \alpha, \rho))$ under model (1) with Gamma(2,θ) or Beta$(2,\theta)$ errors

		Gamma(2, θ) errors					$\beta(2,\theta)$ errors				
ρ	θ	n =200	n =500	n =1000	n =2000	Category	n =200	n =500	n =1000	n =2000	Category
0.2	1	0.752	0.824	0.850	0.882	(G,M)	0.472	0.534	0.526	0.530	(G,S)
	2	0.790	0.840	0.850	0.880	(G,M)	0.670	0.636	0.648	0.666	(G,S)
	6	0.816	0.818	0.828	0.850	(G,M)	0.758	0.776	0.812	0.816	(G,M)
0.5	1	0.836	0.888	0.906	0.932	(G,M)	0.594	0.636	0.658	0.652	(G,S)
	2	0.832	0.884	0.882	0.914	(G,M)	0.738	0.750	0.752	0.814	(G,M)
	6	0.824	0.886	0.902	0.930	(G,M)	0.782	0.820	0.858	0.884	(G,M)
0.8	1	0.706	0.754	0.786	0.812	(B,L)	0.210	0.244	0.258	0.300	(B,M)
	2	0.666	0.694	0.812	0.818	(B,L)	0.358	0.454	0.504	0.546	(B,M)
	6	0.668	0.730	0.798	0.820	(B,L)	0.542	0.628	0.664	0.732	(B,M)
1.0	1	0.892	0.938	0.954	0.960	(B,L)	0.938	0.964	0.976	0.994	(G,L)
	2	0.888	0.936	0.950	0.966	(B,L)	0.730	0.766	0.772	0.798	(B,M)
	6	0.874	0.930	0.968	0.972	(B,L)	0.652	0.668	0.714	0.724	(B,M)

Table 3: $F_n^{(500)}$ and $R_{A,200}$ under model (1) with Gamma(4,θ) or Beta$(4,\theta)$ errors

		Gamma(4, θ) errors					$\beta(4,\theta)$ errors				
ρ	θ	n =200	n =500	n =1000	n =2000	$R_{A,200}$	n =200	n =500	n =1000	n =2000	$R_{A,200}$
0.2	1	0.564	0.676	0.760	0.840	2.090	0.950	0.976	0.994	0.998	1.070
	2	0.578	0.672	0.752	0.850	2.050	0.890	0.976	0.982	1.000	1.210
	6	0.578	0.710	0.750	0.810	2.110	0.764	0.838	0.922	0.962	1.510
0.5	1	0.452	0.494	0.572	0.660	2.410	0.900	0.912	0.964	0.992	1.190
	2	0.438	0.462	0.586	0.620	2.450	0.784	0.888	0.932	0.958	1.380
	6	0.428	0.486	0.524	0.624	2.430	0.604	0.726	0.778	0.854	1.750
0.8	1	0.600	0.578	0.512	0.574	4.800	0.992	0.986	0.992	0.994	2.530
	2	0.572	0.566	0.538	0.514	4.850	0.942	0.944	0.950	0.968	2.740
	6	0.598	0.558	0.534	0.524	4.870	0.780	0.792	0.778	0.810	3.640
1.0	1	0.922	0.986	0.996	1.000	1.600	1.000	1.000	1.000	1.000	1.030
	2	0.908	0.976	0.994	1.000	1.600	1.000	1.000	1.000	1.000	1.150
	6	0.914	0.972	1.000	0.994	1.600	1.000	1.000	1.000	1.000	1.340

Figure 1: Plots of the data used in Section 5.2 and their sample ACFs and PACFs



PREDICTOR SELECTION FOR POSITIVE AUTOREGRESSIVE PROCESSES
SUPPLEMENTARY MATERIALS

1 PROOFS OF THEOREMS 1 and 2

Since (4) can be directly deduced from Corollary 2.5 of Davis and McCormick (1989) and since the proofs of (17) and (21) have been given in Remarks 1 and 3, respectively, we only need to show that (7), (15) and (19) hold true. We start by proving (19), which is ensured by

$$\mathrm{E}(n^{1+(1/\alpha)} \min_{\lfloor n\delta_0 \rfloor+1\leq i\leq n} \varepsilon_i/y_{i-1})^q = O(1), \text{ for any } q>0, \tag{1.1}$$

where $0<\delta_0<1$. Because y_t's are monotonically non-decreasing and $y_{\lfloor n\delta_0\rfloor}$ is independent of $\varepsilon_i, i\geq \lfloor n\delta_0\rfloor+1$, the left-hand side of (1.1) is bounded by

$$\mathrm{E}(n^{1/\alpha} \min_{\lfloor n\delta_0 \rfloor+1\leq i\leq n} \varepsilon_i)^q \mathrm{E}(y_{\lfloor n\delta_0\rfloor}/n)^{-q} := (\mathrm{I})\times(\mathrm{II}). \tag{1.2}$$

Let m be any positive integer greater than q/α and $u_n = \lfloor n/m\rfloor - 1$. Then, it follows that $y_n = \sum_{j=1}^n \varepsilon_j \geq \sum_{i=0}^{u_n}\sum_{j=1}^m \varepsilon_{im+j}$. This and the convexity of $x^{-q}, x\geq 0$, yield

$$\mathrm{E}(y_n/n)^{-q} \leq C m^q u_n^{-1} \sum_{i=0}^{u_n} \mathrm{E}(\sum_{j=1}^m \varepsilon_{im+j})^{-q}, \tag{1.3}$$

where here and hereafter C denotes a generic positive constant whose value is independent of n and may vary at different occurrences. By (3), there is a sufficiently large constant K such that

$$\begin{aligned} \mathrm{E}(\sum_{j=1}^m \varepsilon_j)^{-q} &= \int_0^\infty P(\sum_{j=1}^m \varepsilon_j < t^{-1/q})dt \\ &\leq K + \int_K^\infty P^m(\varepsilon_1 < t^{-1/q})dt \leq K + C(c/\alpha)^m \int_K^\infty t^{-\alpha m/q}dt \leq C, \end{aligned} \tag{1.4}$$

where the second relation is ensured by the independence between ε_t's and the third and last ones are due to (3) and $m > q/\alpha$. Equation (1.4) and an i.i.d. assumption on ε_t's imply that for each $0 \leq i \leq u_n$, $\mathrm{E}(\sum_{j=1}^{m} \varepsilon_{im+j})^{-q} \leq C$, which in conjunction with (1.3) gives $\mathrm{E}(y_n/n)^{-q} = O(1)$. Thus we obtain

$$(\mathrm{II}) = O(1). \tag{1.5}$$

It follows from (3) that there exist $\delta > 0$ and $0 < \underline{c} \leq c < \bar{c}$ such that

$$\underline{c}x^{\alpha-1} \leq f_{\varepsilon_1}(x) \leq \bar{c}x^{\alpha-1}, \text{ for all } 0 < x < \delta^{1/q}. \tag{1.6}$$

Let $q^* > \max\{1, q/q_1\}$ and $l > qq^*q_1/\alpha(q_1q^* - q)$. Then, (3), (1.6), (14) and Chebyshev's inequality yield

$$\begin{aligned}
(\mathrm{I}) \leq & \int_0^{\delta n^{q/\alpha}} P(n^{1/\alpha} \min_{\lfloor n\delta_0 \rfloor + 1 \leq i \leq n} \varepsilon_i > t^{1/q}) dt + \int_{\delta n^{q/\alpha}}^{n^l} P(n^{1/\alpha} \min_{\lfloor n\delta_0 \rfloor + 1 \leq i \leq n} \varepsilon_i > t^{1/q}) dt \\
+ & \int_{n^l}^{\infty} P(n^{1/\alpha} \min_{\lfloor n\delta_0 \rfloor + 1 \leq i \leq n} \varepsilon_i > t^{1/q}) dt \leq \int_0^{\delta n^{q/\alpha}} P^{n(1-\delta_0)}(\varepsilon_1 > t^{1/q} n^{-1/\alpha}) dt \\
+ & \int_{\delta n^{q/\alpha}}^{n^l} P^{n(1-\delta_0)}(\varepsilon_1 > \delta^{1/q}) dt + \int_{n^l}^{\infty} P^{q^*}(n^{q_1/\alpha} \varepsilon_1^{q_1} > t^{q_1/q}) dt \\
\leq & \int_0^{\infty} \exp\{-(\underline{c}/\alpha) t^{\alpha/q} (1-\delta_0)\} dt + \int_{\delta n^{q/\alpha}}^{n^l} \exp\{-(\underline{c}/\alpha) \delta^{\alpha/q} n (1-\delta_0)\} dt \\
+ & (\mathrm{E}(\varepsilon_1^{q_1}))^{q^*} n^{q_1 q^*/\alpha} \int_{n^l}^{\infty} t^{-q^* q_1/q} dt = O(1).
\end{aligned} \tag{1.7}$$

Combining (1.7), (1.5) and (1.2) gives the desired conclusion (19).

Next, we prove (7), which is ensured by

$$(n^{1+(1/\alpha)} \min_{\nu_n \leq i \leq n} \varepsilon_i/y_{i-1}) - (n^{1+1/\alpha} \min_{2 \leq i \leq n} \varepsilon_i/y_{i-1}) = o_p(1), \tag{1.8}$$

and for any $t > 0$,

$$\lim_{n\to\infty} P(n^{1+(1/\alpha)} \min_{\nu_n \leq i \leq n} \varepsilon_i/y_{i-1} > t) = \exp(-\mu^\alpha t^\alpha c/\{\alpha(\alpha+1)\}), \tag{1.9}$$

where $\nu_n = n^\theta$ for some $1/2 < \theta < 1$. It is easy to see that

$$\text{the left-hand side of (1.8) } \leq n^{1+(1/\alpha)}(\min_{\nu_n \leq i \leq n} \varepsilon_i/y_{i-1}) I_{A_n}, \tag{1.10}$$

where $A_n = \{\min_{\nu_n \le i \le n} \varepsilon_i/y_{i-1} > \min_{2\le i<\nu_n} \varepsilon_i/y_{i-1}\}$. Let g_n satisfy $g_n\nu_n^{1+(1/\alpha)}/n^{1+(1/\alpha)} = o(1)$ and $g_n \to \infty$. Then, by (3), (19), (20), the weak law of large number, Chebyshev's inequality and independence between ε_t's, one has for any $\epsilon > 0$,

$$\begin{aligned}
& P(n^{1+(1/\alpha)}(\min_{\nu_n\le i\le n} \varepsilon_i/y_{i-1})I_{A_n} > \epsilon) \le P(A_n) \\
\le\ & P(\min_{\nu_n\le i\le n} \varepsilon_i/y_{i-1} > g_n^{-1}\nu_n^{-1-(1/\alpha)}) + P(\max_{2\le i\le \nu_n} y_{i-1} \ge g_n^{1/2}\nu_n) + P(\min_{2\le i\le\nu_n} \varepsilon_i < g_n^{-1/2}\nu_n^{-1/\alpha}) \\
=\ & O(g_n\nu_n^{1+(1/\alpha)}/n^{1+(1/\alpha)}) + o(1) + 1 - \{1 - C/(\nu_n g_n^{\alpha/2})\}^{\nu_n} = o(1). \qquad (1.11)
\end{aligned}$$

In view of (1.10) and (1.11), (1.8) follows. Let $Z_n = \min_{\nu_n\le i\le n} n\varepsilon_i/\{(i-1)\mu - b_n\}$, $Z_n^* = \min_{\nu_n\le i\le n} n\varepsilon_i/\{(i-1)\mu + b_n\}$ and $B_n = \{\max_{\nu_n\le i\le n} |y_{i-1} - (i-1)\mu| \le b_n\}$, where $b_n = n^{\theta'}$ with $1/2 < \theta' < \theta$ (noting that $(i-1)\mu - b_n > 0$ for all $\nu_n \le i \le n$ and all large n). According to Corollary 11.2.1 of Chow and Teicher (1997) and (20),

$$P(B_n^c) = o(1). \qquad (1.12)$$

Since on the set B_n, $\varepsilon_i/\{(i-1)\mu + b_n\} \le \varepsilon_i/y_{i-1} \le \varepsilon_i/\{(i-1)\mu - b_n\}$ for all $\nu_n \le i \le n$, it holds that

$$n^{1/\alpha}Z_n^* I_{B_n} \le n^{1+(1/\alpha)}(\min_{\nu_n\le i\le n} \varepsilon_i/y_{i-1})I_{B_n} \le n^{1/\alpha}Z_n I_{B_n}. \qquad (1.13)$$

Moreover, by (3) and a straightforward calculation, one has for any $t > 0$,

$$\lim_{n\to\infty} P(n^{1/\alpha}Z_n > t) = \lim_{n\to\infty} P(n^{1/\alpha}Z_n^* > t) = \exp(-\mu^\alpha t^\alpha c/\{\alpha(\alpha+1)\}). \qquad (1.14)$$

Consequently, (1.9) follows from (1.12)-(1.14). Thus, the proof of (7) is complete.

While the proof of (15) is similar in spirit to that of (19), a nontrivial modification is needed because y_t is no longer monotonically non-decreasing when $0 \le \rho < 1$. Let δ be defined as in (1.6), $0 < \theta < 1$, $q^* > \max\{q/(q_1(1-\theta)), q/(\theta\alpha), 1\}$ and $l > q_1 q^* q/\{\alpha[(1-\theta)q_1 q^* - q]\}$. Then, for $0 \le \rho < 1$,

$$\mathrm{E}\{n^{1/\alpha}(\hat\rho_n - \rho)\}^q \le \text{(III)}+\text{(IV)}+\text{(V)}, \qquad (1.15)$$

where (III) $= \int_0^{\delta n^{q/\alpha}} P(n^{1/\alpha}\min_{2\le i\le n} \varepsilon_i/y_{i-1} > t^{1/q})dt$, (IV) $= \int_{\delta n^{q/\alpha}}^{n^l} P(n^{1/\alpha}\min_{2\le i\le n} \varepsilon_i/y_{i-1} > t^{1/q})dt$ and (V) $= \int_{n^l}^{\infty} P(n^{1/\alpha}\min_{2\le i\le n} \varepsilon_i/y_{i-1} > t^{1/q})dt$. Let $0 < \eta < 1$ satisfy $P(\varepsilon_1 \ge \eta) >$

0. Then, it follows from (1.6) and $\min_{2\le i\le n}\varepsilon_i/y_{i-1} \le \min_{1\le i\le \lfloor n/2\rfloor}\varepsilon_{2i}/\varepsilon_{2i-1}$ that (III) $\le \int_0^{\delta n^{q/\alpha}} P^{\lfloor n/2\rfloor}(\varepsilon_2/\varepsilon_1 > n^{-1/\alpha}t^{1/q})dt \le \int_0^{\delta n^{q/\alpha}}\{1-P(\varepsilon_1\ge\eta)P(\varepsilon_2\le n^{-1/\alpha}t^{1/q}\eta)\}^{\lfloor n/2\rfloor}dt = O(1)$. Similarly, (IV) $= O(1)$. In addition, (V) $\le \int_{n^l}^{\infty}\{P(\varepsilon_2/\varepsilon_1 > n^{-1/\alpha}t^{1/q}, \varepsilon_1 > t^{-\theta/q}) + P(\varepsilon_1 \le t^{-\theta/q})\}^{q^*}dt \le C\int_{n^l}^{\infty}\{P(\varepsilon_2^{q_1} > n^{-q_1/\alpha}t^{(1-\theta)q_1/q}) + P(\varepsilon_1 \le t^{-\theta/q})\}^{q^*}dt = O(1)$. The bounds for (III), (IV) and (V) in conjunction with (1.15) lead to the desired conclusion (15).

2 PROOFS OF THEOREMS 3 and 4

We begin by proving Theorem 3. It is easy to see that

$$\lim_{n\to\infty} n\mathrm{E}[(1-\rho)\bar{y}_{n-1}-\mu]^2 = \sigma^2. \tag{2.1}$$

Making use of (15), (36) and Hölder's inequality, it follows that

$$\mathrm{E}\{(\hat{\rho}_n-\rho)^2(\bar{y}_{n-1}-\mu_y)^2\} = O(n^{-1-(2/\alpha)}), \tag{2.2}$$

and

$$\mathrm{E}(z_n^2) = O(n^{-2}). \tag{2.3}$$

In view of (34) and (2.1)-(2.3), it remains to show that

$$\lim_{n\to\infty} n^{2/\alpha}\mathrm{E}\{(\hat{\rho}_n-\rho)^2(y_n-\mu_y)^2\} = L_0(\alpha,c,\rho,\sigma^3), \tag{2.4}$$

and

$$\mathrm{E}\{n^{(1/2)+(1/\alpha)}(\hat{\rho}_n-\rho)(y_n-\mu_y)[(1-\rho)\bar{y}_{n-1}-\mu]\} = o(1). \tag{2.5}$$

To prove (2.4), note first that a standard truncation argument (see, e.g., Section 7.3 of Brockwell and Davis, 1987), together with Theorem 4.2 of Billingsley (1968) and (4), yields

$$n^{1/\alpha}(\hat{\rho}_n-\rho)(y_n-\mu_y) \Rightarrow S_1S_2, \tag{2.6}$$

where S_1 has the same distribution as the limiting distribution of $n^{1/\alpha}(\hat{\rho}_n-\rho)$, which is given in (4), S_2 is distributed as that of $\sum_{j=0}^{\infty}\rho^j\delta_{1-j}$, and S_1 and S_2 are independent. By (2.6), the continuous mapping theorem, and $\mathrm{E}|n^{1/\alpha}(\hat{\rho}_n - \rho)(y_n - \mu_y)|^{\xi} = O(1)$ for some $\xi > 2$ (which is ensured by (15), (36) and Hölder's inequality), the desired conclusion (2.4) follows. Finally, applying a similar truncation argument, we get $\mathrm{E}\{n^{(1/2)+(1/\alpha)}(\hat{\rho}_n - \rho)(y_n - \mu_y)[(1-\rho)\bar{y}_{n-1} - \mu]\} = \mathrm{E}\{n^{(1/2)+(1/\alpha)}(\hat{\rho}_n - \rho)[(1-\rho)\bar{y}_{n-1} - \mu]\}\mathrm{E}(y_n - \mu_y) + o(1) = o(1)$. Thus, (2.5) holds and the proof is complete.

Next, we prove Theorem 4. It is easy to see that

$$\lim_{n\to\infty} n\mathrm{E}(z_n - \mu)^2 = \sigma^2. \tag{2.7}$$

Since $(y_n - \bar{y}_{n-1})/n = \mu/2 + o_p(1)$, $n^{1+(1/\alpha)}(\hat{\rho}_n - 1)$ and $(y_n - \bar{y}_{n-1})/n$ are asymptotically independent. This, together with Theorem 2, the continuous mapping theorem, Hölder's inequality and (41), yields

$$\mathrm{E}(n^{2/\alpha}(\hat{\rho}_n - 1)^2(y_n - \bar{y}_{n-1})^2) = L_1(\alpha, c). \tag{2.8}$$

According to (39), (2.7) and (2.8), (42) follows if

$$\mathrm{E}((\hat{\rho}_n - 1)(y_n - \bar{y}_{n-1})(z_n - \mu)) = o(\max\{n^{-1}, n^{-2/\alpha}\}). \tag{2.9}$$

By applying (41), (19) and Hölder's inequality, it is readily seen that (2.9) holds for $0 < \alpha < 2$ and $\alpha > 2$. The proof of (2.9) in the critical case $\alpha = 2$, however, is nontrivial and may be of independent interest.

Let ν_n, B_n, b_n, Z_n and Z_n^* be defined as in Section 1. Equations (1.2), (1.5) and (1.7) yield

$$\mathrm{E}(n^{3/2} \min_{\nu_n \le i \le n} \varepsilon_i/y_{i-1})^q = O(1) \text{ for any } q > 0. \tag{2.10}$$

In addition, (41) implies $\mathrm{E}|\{(y_n - \bar{y}_{n-1})/n\} - (\mu/2)|^{q_2} = o(1)$. According to this, (19), (1.8), (1.12), (41), (2.10) and Hölder's inequality, (2.9) with $\alpha = 2$ follows from

$$\mathrm{E}(\mathrm{I}) := \mathrm{E}\{(n^{3/2} \min_{\nu_n \le i \le n} \varepsilon_i/y_{i-1})(n^{-1/2}\sum_{j=2}^{n} \delta_j)I_{B_n}\} = o(1). \tag{2.11}$$

Note that (I) is bounded below and above by $n^{1/2}Z_n(n^{-1/2}\sum_{j=2}^n \delta_j)I_{B_n} - |n^{-1/2}\sum_{j=2}^n \delta_j|n^{1/2}(Z_n - Z_n^*)$ and $n^{1/2}Z_n^*(n^{-1/2}\sum_{j=2}^n \delta_j)I_{B_n} + |n^{-1/2}\sum_{j=2}^n \delta_j|n^{1/2}(Z_n - Z_n^*)$, respectively. Therefore, (2.11) holds if

$$\mathrm{E}\{|n^{-1/2}\sum_{j=2}^{n}\delta_j|n^{1/2}(Z_n - Z_n^*)\} = o(1), \tag{2.12}$$

$$\mathrm{E}\{n^{1/2}Z_n^*(n^{-1/2}\sum_{j=2}^{n}\delta_j)I_{B_n}\} = o(1), \tag{2.13}$$

and

$$\mathrm{E}\{n^{1/2}Z_n(n^{-1/2}\sum_{j=2}^{n}\delta_j)I_{B_n}\} = o(1). \tag{2.14}$$

By an argument similar to that used to prove (19), it can be shown that for any $q > 0$,

$$\mathrm{E}(n^{1/2}Z_n)^q = O(1) \text{ and } \mathrm{E}(n^{1/2}Z_n^*)^q = O(1). \tag{2.15}$$

Using (2.15), $0 \leq n^{1/2}(Z_n - Z_n^*) \leq Cn^{1/2}Z_n b_n/(\mu\nu_n)$, (41) and the Cauchy-Schwarz inequality, one obtains $\mathrm{E}\{|n^{-1/2}\sum_{j=2}^{n}\delta_j|n^{1/2}(Z_n - Z_n^*)\} \leq C(b_n/\nu_n)\mathrm{E}(|n^{-1/2}\sum_{j=2}^{n}\delta_j|n^{1/2}Z_n) = o(1)$, which gives (2.12). To show (2.13), note that

$$\begin{aligned} & |\mathrm{E}\{n^{1/2}Z_n^*(n^{-1/2}\sum_{j=2}^{n}\delta_j)I_{B_n}\}| \leq |\mathrm{E}\{n^{1/2}Z_n^*(n^{-1/2}\sum_{j=2}^{n}\delta_j)\}| + |\mathrm{E}\{n^{1/2}Z_n^*(n^{-1/2}\sum_{j=2}^{n}\delta_j)I_{B_n^c}\}| \\ = \; & |\mathrm{E}\{n^{1/2}Z_n^*(n^{-1/2}\sum_{j=2}^{n}\delta_j)\}| + o(1), \end{aligned}$$

where the identity is ensured by (2.15), (41), (1.12) and Hölder's inequality. Moreover, some tedious calculations yield that for all $\nu_n \leq j \leq n$ and some $\theta > 1/2$, $|\mathrm{E}(n^{1/2}Z_n^*\delta_j)| \leq Cn^{-\theta}$, and hence $\mathrm{E}\{n^{1/2}Z_n^*(n^{-1/2}\sum_{j=2}^{n}\delta_j)\} = o(1)$. Consequently, (2.13) follows. The proof of (2.14) is similar to that of (2.13). We omit the details.

3 PROOF OF LEMMA 1

Since by (ii),

$$(\log n)^{-1}\sum_{i=M}^{n} d_i = (\log n)^{-1}\sum_{i=M}^{n}\frac{Q_i}{i} + (\log n)^{-1}\sum_{i=M}^{n}\frac{\mathrm{E}(idi)}{i} = V_n + S + o(1),$$

where $V_n = (\log n)^{-1} \sum_{i=M}^{n} Q_i/i$, it suffices for Lemma 1 to show that

$$\mathrm{E}(V_n^2) \to 0 \text{ as } n \to \infty. \tag{3.1}$$

It is clearly no loss of generality to assume that n/a_n and $\eta i a_n$ are positive integers. Hence

$$\begin{aligned} V_n^2 &= (\log n)^{-2} \sum_{i=M}^{n} \frac{Q_i^2}{i^2} + 2(\log n)^{-2} \sum_{i=M}^{n-1} \sum_{j=i+1}^{n} \frac{Q_i\, Q_j}{ij} \\ &= (\log n)^{-2} \sum_{i=M}^{n} \frac{Q_i^2}{i^2} + 2(\log n)^{-2} \sum_{i=M}^{n/a_n} \sum_{j=i+1}^{\eta i a_n - 1} \frac{Q_i\, Q_j}{i\, j} \\ &\quad + 2(\log n)^{-2} \sum_{i=M}^{n/a_n} \sum_{j=\eta i a_n}^{n} \frac{Q_i\, Q_j}{i\, j} + 2(\log n)^{-2} \sum_{i=(n/a_n)+1}^{n-1} \sum_{j=i+1}^{n} \frac{Q_i\, Q_j}{i\, j} \\ &:= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}. \end{aligned} \tag{3.2}$$

By making use of (i), the Cauchy-Schwarz inequality and $\log a_n / \log n \to 0$, it holds that

$$\mathrm{E}(\mathrm{I}) = C(1/\log n)^2 = o(1), \tag{3.3}$$

$$|\mathrm{E}(\mathrm{II})| \leq C(\log a_n / \log n) = o(1), \tag{3.4}$$

and

$$|\mathrm{E}(\mathrm{IV})| \leq C(\log a_n / \log n)^2 = o(1). \tag{3.5}$$

Moreover, (iii) and some algebraic manipulations yield

$$|\mathrm{E}(\mathrm{III})| \leq C \max_{M \leq i \leq n/a_n,\, i\eta a_n \leq j \leq n} |\mathrm{E}\,(Q_i Q_j)\,| = o(1). \tag{3.6}$$

Now, (3.1) follows directly from (3.2)-(3.6). Thus, the proof is complete.

REFERENCES

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Brockwell, P. J. and Davis, R. A. (1987), *Time Series: Theory and Methods*, New York: Springer.

Chow, Y. S. and Teicher, H. (1997), *Probability Theory: Independence, Interchangeability, Martingales* (2nd ed.), New York: Springer.

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.