Interval Estimation for a First-Order Positive Autoregressive Process

Wei-Cheng Hsiao
Institute of Statistics, National Tsing Hua University,
Hao-Yun Huang
Department of Economics, National Taiwan University,
and
Ching-Kang Ing
Institute of Statistics, National Tsing Hua University

February 21, 2018

Abstract

We are interested in constructing confidence intervals for the autoregressive (AR) coefficient of a first-order AR model with i.i.d. positive errors via an extreme value estimate (EVE). We assume that the error distribution has a density function $f_\varepsilon(x)$ behaving like $b_{1,0}x^{\alpha_0-1}$ as $x \to 0$, where $b_{1,0}$ and α_0 are unknown positive constants. These specifications imply that the EVE has a limiting distribution depending on $b_{1,0}$ and α_0 from which only an infeasible interval estimate can be obtained. To alleviate this difficulty, we introduce a novel procedure to estimate these two constants and establish the desired consistency. This consistency result enables us not only to gain a better understanding of the underlying error distribution, but also to construct a feasible, asymptotically valid confidence interval of the AR coefficient, without resorting to a bootstrap procedure described in Datta and McCormick (1995). The performance of the proposed interval estimate is further illustrated through simulation studies and real data analysis.

Keywords: Confidence intervals; extreme-value estimates; positive autoregressive processes; regular variation indices

---

Ing's research was supported in part by the Science Vanguard Research Program of the Ministry of Science and Technology (MoST), Taiwan.

1 Introduction

Suppose that the data are generated from the first-order autoregressive (AR(1)) model:

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

where $0 \leq \rho < 1$ and ε_t are i.i.d. positive errors satisfying

$$\lim_{x \to 0} \frac{f_\varepsilon(x)}{b_{1,0} x^{\alpha_0 - 1}} = 1, \text{ for some } \alpha_0 > 0 \text{ and } b_{1,0} > 0, \tag{2}$$

with $f_\varepsilon(x)$ denoting the density function of ε_1, and

$$\mathrm{E}(\varepsilon_1^{\beta_0}) < \infty, \text{ for some } \beta_0 > \alpha_0. \tag{3}$$

Model (1) with ε_t obeying (2) and (3) 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). Note that assumption (2) is quite flexible for positive errors although it is violated when the distribution of ε_1 has a point mass at zero or has support whose lower limit is greater than zero. Assumption (3) is, in general, difficult to justify. However, it is easily fulfilled when the distribution of ε_1 has finite moments of all positive orders.

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

$$\tilde{\rho}_n = \sum_{i=2}^{n} (y_{i-1} - \bar{y}_{n-1})(y_i - \bar{y}) / \sum_{i=2}^{n} (y_{i-1} - \bar{y}_{n-1})^2, \tag{4}$$

where $\bar{y}_t = t^{-1} \sum_{i=1}^{t} y_i$. However, when (2) is assumed, LSE may not be efficient and other estimation procedures are needed. 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). Another option that can be applied to more general situations is the extreme value estimate (EVE),

$$\hat{\rho}_n = \min_{1 \leq i \leq n-1} y_{i+1}/y_i, \tag{5}$$

which 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). It is shown in Corollary 2.4 of Davis and McCormick (1989) that the limiting distribution of $\hat{\rho}_n$ satisfies

$$\lim_{n \to \infty} P\{(b_{1,0} \mathrm{E}(y_1)^{\alpha_0} / \alpha_0)^{1/\alpha_0} n^{1/\alpha_0} (\hat{\rho}_n - \rho) > t\} = \exp\{-t^{\alpha_0}\}, \tag{6}$$

provided (2) and (3) hold true. Since $\tilde{\rho}_n$ is $\sqrt{n}$-consistent (see (19)), it is clear from (6) that when $\alpha_0 < 2$ ($\alpha_0 > 2$), 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 $\tilde{\rho}_n$. Ing and Yang (2014) also explored the asymptotic behaviors of the mean squared prediction error (MSPE) of the EV predictor, $\hat{y}_{n+1} = \hat{\mu}_n + \hat{\rho}_n y_n$, of y_{n+1}, under (1)–(3) with $\beta_0 \geq 2$, 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 $\mu = \mathrm{E}(\varepsilon_1)$ based on $\hat{\rho}_n$. Chan, Ing and Zhang (2017) recently generalized Ing and Yang's MSPE result to a general near-unit root model, which is model (1) with $\rho = \rho_n = 1 - b/n^{\beta_1}$, $0 < \beta_1 \leq 1$, and $b > 0$. This generalization allows 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.

Interval estimation of ρ based on $\hat{\rho}_n$ has also been considered by Datta and McCormick (1995). A major difficulty of constructing a confidence interval for ρ via (6) is that α_0 and $b_{1,0}$ appear in the normalizing constant and α_0 also appears in the limit. While the kernel method of density estimation can be applied to the AR residuals to obtain estimates of $b_{1,0}$ and α_0, such estimates may be seriously biased when $0 < \alpha_0 \leq 1$ because the underlying density function is nonzero or even has a pole at the origin; see Marron and Ruppert (1994). In fact, some sophisticated kernel estimation algorithms have been proposed by Marron and Ruppert (1994) to reduce the boundary bias. However, consistency of the resultant estimates of $b_{1,0}$ and α_0 based on the AR residuals still seems difficult to establish when only (2) is assumed. To bypass this difficulty, Datta and McCormick (1995) suggested an asymptotically pivotal quantity based on $\hat{\rho}_n$ and adopted a bootstrap procedure to consistently estimate its limiting distribution, thereby leading to an asymptotically valid confidence interval for ρ.

In this paper, we propose novel consistent estimates of α_0 and $b_{1,0}$, which can be used in conjunction with (6) to construct an asymptotically valid confidence interval for ρ. Our approach offers several advantages compared to the method proposed by Datta and McCormick (1995). To start with, it is less computationally intensive than the bootstrap procedure of Datta and McCormick (1995). Next, it delivers consistent estimates of α_0 and $b_{1,0}$ from which a better understanding of the error distribution can be gained. The consistent estimates of α_0 and $b_{1,0}$ also help decide whether or not the EVE is better than the LSE; see Section 3 for further details. Finally, the simulation study given in Section 4.1 reveals that our method has a better finite sample performance in terms of percentage coverage, in particular, when α_0 is relatively small. The rest of the paper is organized as follows. We introduce a procedure for estimating $b_{0,1}$ and α_0 in Section 2. Section 3 states the consistency property of the proposed procedure and provides an asymptotic valid confidence interval for ρ. A method to select the better estimate between $\hat{\rho}_n$ and $\tilde{\rho}_n$ is presented at the end of Section 3. In Section 4.1, we compare our method with Datta and McCormick's (1995) method through several simulation examples. In particular, their actual coverage rates for nominal 95% confidence intervals are compared. The performance of these two methods is also illustrated via real data analysis in Section 4.2. We conclude in Section 5. All proofs are deferred to the appendix.

2 Estimation of α_0 and $b_{1,0}$

We take a somewhat nonstandard approach to estimate α_0 and $b_{1,0}$. Note that (2) yields

$$\lim_{n\to\infty} P(n^{1/\alpha_0}\varepsilon_{(1)} \le x) = 1 - \exp(-(b_{1,0}/\alpha_0)x^{\alpha_0}), \tag{7}$$

where $\varepsilon_{(j)}$ is the jth order statistic of $\{\varepsilon_1, \ldots, \varepsilon_n\}$. The right-hand side of (7) is a Weibull distribution whose density function is given by

$$f_{(1)}(x) = (\alpha_0/\eta_0)x^{\alpha_0-1}\exp(-x^{\alpha_0}/\eta_0), \tag{8}$$

where α_0 and $\eta_0 = \alpha_0/b_{1,0}$ are shape and scale parameters, respectively.

Divide the AR residuals, $\hat{\varepsilon}_i = y_i - \hat{\rho}_n y_{i-1}, i = 2, \ldots, n$, into m subgroups, $\{\hat{\varepsilon}_j, j \in I_i\}, i = 1, \ldots, m$, where $m \approx n^{1-\theta}$, with $0 < \theta < 1/2$, and $I_i = \{(i-1)(n-1)/m+2, \ldots, i(n-1)/m+1\}$. Here, $a_n \approx b_n$ means $a_n/b_n \to 1$ as $n \to \infty$, and m and $n_1 = (n-1)/m$ are assumed to be positive integers for the sake of simplicity of exposition. Define $\hat{\varepsilon}_i^* = \min_{j\in I_i}\hat{\varepsilon}_j$. Since it is expected that $n_1^{1/\alpha_0}\hat{\varepsilon}_i^*, i = 1, \ldots, m$, are asymptotically independent, we are motivated by (7) and (8) to estimate α_0 through maximizing the average log-likelihood function,

$$\begin{aligned} l(\alpha, \eta) = l(\alpha, \eta|\hat{w}_1, \ldots, \hat{w}_m) &= \frac{1}{m}\sum_{i=1}^m \log f_{(1)}(\hat{w}_i) \\ &= \log\frac{\alpha}{\eta} + (\alpha-1)\frac{1}{m}\sum_{i=1}^m \log \hat{w}_i - \frac{1}{m\eta}\sum_{i=1}^m \hat{w}_i^\alpha, \end{aligned} \tag{9}$$

over $\alpha \in \mathbf{H}_1 = [\underline{\alpha}, \bar{\alpha}]$ and $\eta \in \mathbf{H}_2 = [\underline{\eta}, \bar{\eta}]$, where $\underline{\alpha}$ ($\underline{\eta}$) and $\bar{\alpha}$ ($\bar{\eta}$) are known lower and upper bounds for α_0 (η_0), and $\hat{w}_i = n_1^\xi \hat{\varepsilon}_i^\circ$, with $-\infty < \xi < \infty$, $\hat{\varepsilon}_i^\circ = \max\{\hat{\varepsilon}_i^*, n^{-\theta_0}\}$, and $\theta_0 > \theta/\underline{\alpha}$.

It is worth mentioning that $\hat{w}_i$ is used in (9) instead of $n_1^{1/\alpha_0}\hat{\varepsilon}_i^*$ because (i) the latter is practically inaccessible; and (ii) there are some extremely small $\hat{\varepsilon}_i^*$ leading to very large $|\log\hat{\varepsilon}_i^*|$. Moreover, while n_1^ξ may not be a proper normalizing constant for $\hat{\varepsilon}_i^\circ$, one of the key observations obtained in this work is that for any $-\infty < \xi < \infty$,

$$\hat{\alpha}(\hat{w}_1, \ldots, \hat{w}_m) = \hat{\alpha}(w_1^*, \ldots, w_m^*), \tag{10}$$

where $w_i^* = n_1^{1/\alpha_0}\hat{\varepsilon}_i^\circ$,

$$(\hat{\alpha}(\hat{w}_1, \ldots, \hat{w}_m), \hat{\eta}(\hat{w}_1, \ldots, \hat{w}_m)) = (\hat{\alpha}, \hat{\eta}) = \arg\max_{\mathbf{v}\in\mathbf{H}} l(\alpha, \eta|\hat{w}_1, \ldots, \hat{w}_m),$$

and

$$(\hat{\alpha}(w_1^*, \ldots, w_m^*), \hat{\eta}(w_1^*, \ldots, w_m^*)) = (\hat{\alpha}^*, \hat{\eta}^*) = \arg\max_{\mathbf{v}\in\mathbf{H}} l(\alpha, \eta|w_1^*, \ldots, w_m^*),$$

with $\mathbf{v} = (\alpha, \eta)'$ and $\mathbf{H} = \mathbf{H}_1 \times \mathbf{H}_2$. To see this, note that for a given α, $l(\alpha, \eta | \hat{w}_1, \ldots, \hat{w}_m)$ and $l(\alpha, \eta | w_1^*, \ldots, w_m^*)$ are maximized by choosing

$$\eta = \eta_\xi(\alpha) = m^{-1} \sum_{i=1}^{m} \hat{w}_i^\alpha \text{ and } \eta = \eta^*(\alpha) = m^{-1} \sum_{i=1}^{m} w_i^{*\alpha}, \tag{11}$$

respectively. Moreover, straightforward calculations yield

$$\begin{aligned} & l(\alpha, \eta_\xi(\alpha) | \hat{w}_1, \ldots, \hat{w}_m) - l(\alpha, \eta^*(\alpha) | w_1^*, \ldots, w_m^*) \\ & = \log \eta^*(\alpha) - \log \eta_\xi(\alpha) + (\alpha - 1) \frac{1}{m} \sum_{i=1}^{m} (\log \hat{w}_i - \log \hat{w}_i^*) \\ & = \log n_1^{\alpha/\alpha_0} - \log n_1^{\xi\alpha} + (\alpha - 1)(\log n_1^\xi - \log n_1^{1/\alpha_0}) \\ & = (1/\alpha_0 - \xi) \log n_1, \end{aligned}$$

which is independent of α. Therefore, $l(\alpha, \eta_\xi(\alpha) | \hat{w}_1, \ldots, \hat{w}_m)$ and $l(\alpha, \eta^*(\alpha) | w_1^*, \ldots, w_m^*)$ are maximized by the same value of α, and hence (10) follows.

On the other hand, (11) yields $\hat{\eta} = m^{-1} \sum_{i=1}^{m} \hat{w}_i^{\hat{\alpha}}$ and $\hat{\eta}^* = m^{-1} \sum_{i=1}^{m} w_i^{*\hat{\alpha}}$, which are not identical unless $\xi = 1/\alpha_0$. To remedy this difficulty, we suggest estimating η_0 using

$$\tilde{\eta} = m^{-1} \sum_{i=1}^{m} (n_1^{1/\hat{\alpha}} \hat{\varepsilon}_i^\circ)^{\hat{\alpha}} = \frac{n_1}{m} \sum_{i=1}^{m} (\hat{\varepsilon}_i^\circ)^{\hat{\alpha}}. \tag{12}$$

Finally, $b_{1,0}$ is estimated by

$$\hat{b}_{1,0} = \frac{\hat{\alpha}}{\tilde{\eta}}. \tag{13}$$

The proposed estimate $(\hat{\alpha}, \tilde{\eta}, \hat{b}_{1,0})$ has the advantage of being easy to implement. The consistency property of $(\hat{\alpha}, \tilde{\eta}, \hat{b}_{1,0})$ is discussed in the next section.

3 Consistency of $(\hat{\alpha}, \tilde{\eta}, \hat{b}_{1,0})$ and Asymptotic Valid Confidence Intervals for ρ

In this section, we need an assumption slightly stronger than (2),

$$\frac{f_\varepsilon(x)}{b_{1,0} x^{\alpha_0 - 1}} = 1 + o(x^\nu) \text{ for some } \nu > 0 \text{ as } x \to 0. \tag{14}$$

In addition, we set $\xi = 1/2$ (although other choices of ξ will lead to the same $\hat{\alpha}$) and assume $n_1 = \lfloor n^\theta \rfloor$, where $\lfloor a \rfloor$ denotes the largest integer $\leq a$. The next theorem asserts that $(\hat{\alpha}, \tilde{\eta}, \hat{b}_{1,0})$ possesses consistency in estimating $(\alpha_0, \eta_0, b_{1,0})$.

Theorem 3.1 *Assume* (1), (3) *and* (14). *Then, there exists* $\kappa > 0$, *depending on* θ, $\underline{\alpha}$, α_0, β_0, *and* ν, *such that*

$$\hat{\alpha} = \alpha_0 + O_p(n^{-\kappa}),\ \tilde{\eta} = \eta_0 + O_P(n^{-\kappa}\log n),\ \hat{b}_{1,0} = b_0 + O_P(n^{-\kappa}\log n). \tag{15}$$

One of the most interesting applications of Theorem 3.1 is that it can be used in conjunction with (6) to provide an asymptotically valid confidence interval for ρ. More specifically, denote by $Z\%$ the confidence level of interest, where $0 < Z < 100$. Let $0 < Z_1 < Z_2 < 100$ satisfy $Z_2 - Z_1 = Z$ and define

$$\begin{aligned} t_R &= (-\log Z_1\%)^{1/\alpha_0}, \\ t_L &= (-\log Z_2\%)^{1/\alpha_0}, \\ A_n &= \left(\frac{m_{\alpha_0}}{\eta_0}\right)^{1/\alpha_0} n^{1/\alpha_0}, \end{aligned}$$

where $m_{\alpha_0} = \mathrm{E}(y_1^{\alpha_0})$. Then, (6) implies that

$$(\hat{\rho}_n - t_R/A_n, \hat{\rho}_n - t_L/A_n) \tag{16}$$

is a $Z\%$ asymptotic confidence interval of ρ. Since (16) is infeasible, we replace t_R, t_L and A_n by their estimates

$$\begin{aligned} \hat{t}_R &= (-\log Z_1\%)^{1/\hat{\alpha}}, \\ \hat{t}_L &= (-\log Z_2\%)^{1/\hat{\alpha}}, \\ \hat{A}_n &= \left(\frac{\hat{m}_{\alpha_0}}{\tilde{\eta}}\right)^{1/\hat{\alpha}} n^{1/\hat{\alpha}}, \end{aligned}$$

where $\hat{m}_{\alpha_0} = n^{-1}\sum_{i=1}^n y_i^{\hat{\alpha}}$, and obtain the following feasible confidence interval of ρ,

$$(\hat{\rho}_n - \hat{t}_R/\hat{A}_n, \hat{\rho}_n - \hat{t}_L/\hat{A}_n). \tag{17}$$

The next theorem states the asymptotic validity of (17).

Theorem 3.2 *Assume* (1), (3) *and* (14). *Then,*

$$\lim_{n\to\infty} P(\rho \in (\hat{\rho}_n - \hat{t}_R/\hat{A}_n, \hat{\rho}_n - \hat{t}_L/\hat{A}_n)) = Z\%. \tag{18}$$

Equation (18) follows immediately from (3), (6), (14), and (15). We skip the details.

Theorem 3.1 can also be applied to selecting the better estimate between EVE, $\hat{\rho}_n$, and the LSE, $\tilde{\rho}_n$, when (3) holds with $\beta_0 > 2$. It is shown in Section 2 of Ing and Yang (2014) that

$$n^{1/2}(\tilde{\rho}_n - \rho) \Rightarrow \mathbf{W}_L, \tag{19}$$

where $\Rightarrow$ denotes convergence in distribution and $\mathbf{W}_L$ is a normal distribution with mean 0 and variance $1-\rho^2$. In view of (6) and (19), we conclude that $\tilde{\rho}_n$ is better (worse) than $\hat{\rho}_n$ if $\alpha_0 > 2$ (< 2). If $\alpha_0 = 2$, we deduce from (6) that

$$n^{1/2}(\hat{\rho}_n - \rho) \Rightarrow \mathbf{W}_E,$$

where $\mathbf{W}_E$ is a Weibull distribution with shape parameter α_0 and scale parameter η_0/m_{α_0}. Denote by μ and σ^2 the mean and the variance of ε_1. Then

$$\mathrm{E}(\mathbf{W}_E^2) = \frac{2}{b_{1,0}\left(\frac{\sigma^2}{1-\rho^2} + \frac{\mu^2}{(1-\rho)^2}\right)},$$

which is smaller than (larger than, identical to) $\mathrm{E}(\mathbf{W}_L^2) = 1-\rho^2$ when

$$b_{1,0} > \ (<, =)\, T_{\rho,\mu,\sigma^2} \equiv \frac{2(1-\rho)}{(1+\rho)\mu^2 + (1-\rho)\sigma^2}.$$

The above discussion suggests the following "oracle" estimate

$$\begin{aligned} \check{\rho}_n^{\mathbf{ora}} &= \hat{\rho}_n(I_{\{\alpha_0<2\}} + I_{\{\alpha_0=2, b_{1,0}>T_{\rho,\mu,\sigma^2}\}}) + \tilde{\rho}_n(I_{\{\alpha_0>2\}} + I_{\{\alpha_0=2, b_{1,0}<T_{\rho,\mu,\sigma^2}\}}) \\ &\quad + \hat{\rho}_n I_{\{\alpha_0=2, b_{1,0}=T_{\rho,\mu,\sigma^2}\}}. \end{aligned} \tag{20}$$

To implement (20), we first estimate T_{ρ,μ,σ^2} via

$$\hat{T}_{\rho,\mu,\sigma^2} = \frac{2(1-\hat{\rho}_n)}{(1+\hat{\rho}_n)\hat{\mu}_n^2 + (1-\hat{\rho}_n)\hat{\sigma}_n^2},$$

where $\hat{\mu}_n$ is defined in Section 1 and $\hat{\sigma}_n^2 = (1-\hat{\rho}_n^2)n^{-1}\sum_{i=1}^n (y_i - \bar{y}_n)^2$, and then mimic $\check{\rho}_n^{\mathbf{ora}}$ using

$$\begin{aligned} \hat{\rho}_n^{\mathbf{ora}} &= \hat{\rho}_n(I_{\{\hat{\alpha}_n<2-a_n\}} + I_{\{|\hat{\alpha}_n-2|\le a_n, \hat{b}_{1,0}>\hat{T}_{\rho,\mu,\sigma^2}-b_n\}}) \\ &+ \tilde{\rho}_n(I_{\{\hat{\alpha}_n>2+a_n\}} + I_{\{|\hat{\alpha}_n-2|\le a_n, \hat{b}_{1,0}\le\hat{T}_{\rho,\mu,\sigma^2}-b_n\}}), \end{aligned}$$

where $\{a_n\}$ and $\{b_n\}$ are sequences of positive numbers satisfying $a_n \to 0$, $b_n \to 0$, $n^{-\varsigma}/a_n \to 0$, and $n^{-\varsigma}/b_n \to 0$ for any $\varsigma > 0$. By making use of Theorem 3.1, (3) with $\beta_0 > 2$, and (6), it can be shown that $P(|\hat{\alpha}_n - \alpha_0| \ge a_n) = o(1)$, $P(|\hat{b}_{1,0} - b_{1,0}| \ge b_n/2) = o(1)$, and $P(|\hat{T}_{\rho,\mu,\sigma^2} - T_{\rho,\mu,\sigma^2}| \ge b_n/2) = o(1)$. Hence the desired property,

$$\lim_{n\to\infty} P(\hat{\rho}_n^{\mathbf{ora}} = \check{\rho}_n^{\mathbf{ora}}) = 1, \tag{21}$$

follows.

Before closing this section, we mention that Ing and Yang (2014) have proposed a somewhat different approach to choose between $\hat{\rho}_n$ and $\tilde{\rho}_n$. They suggested selecting the estimate having the smaller accumulated prediction error, and proved that their method has a property similar to (21). A comparison of the finite sample performance of $\hat{\rho}_n^{\mathbf{ora}}$ and their method would be interesting future work.

4 Numerical Illustrations

In Section 4.1, we present some simulation results in support of Theorems 3.1 and 3.2. In Section 4.2, the performance of the proposed estimates is illustrated through analyzing two real datasets.

4.1 Simulation Studies

We generate n observations from the AR(1) process

$$y_t = \rho y_{t-1} + \varepsilon_t, \tag{22}$$

where $n = 200$ or 400, $0 < \rho < 1$, and ε_t are i.i.d exponential random variables whose density function is given by

$$f(x) = \lambda \exp(-\lambda x) I_{\{x \geq 0\}}, \lambda > 0, \tag{23}$$

which is a special case of (2) with $\alpha_0 = 1$ and $b_{1,0} = \lambda$. In finite sample situations, the performance of our estimate, $(\hat{\alpha}, \tilde{\eta}, \hat{b}_{1,0})$, of $(\alpha_0, \eta_0, b_{1,0})$ can vary drastically according to the values of θ and θ_0, recalling that $\lfloor n^\theta \rfloor$ is the number of observations in each subgroup and $n^{-\theta_0}$ is the threshold value used to replace $\hat{\varepsilon}_i^* = \min_{j \in I_i} \hat{\varepsilon}_j$ when $\hat{\varepsilon}_i^* < n^{-\theta_0}$. Therefore, we suggest a procedure to choose (θ, θ_0) from the confidence interval perspective. Throughout this section, the confidence level is set to $Z\% = 95\%$.

Consider all grid-point pairs (θ, θ_0) in $\mathcal{A} \times \mathcal{B}$, where $\mathcal{A} = \mathcal{B} = \{0.2, 0.22, \ldots, 0.8\}$. Divide $y_1, \ldots, y_n$ into $K = 20$ equal subgroups chronologically. For a given (θ, θ_0), we apply $(\hat{\alpha}, \tilde{\eta}, \hat{b}_{1,0})$ and (17), with $Z_1 = 4.9$ and $Z_2 = 99.9$, to each subgroup separately, and denote the confidence interval obtained from the ith subgroup by $\hat{I}(i, \theta, \theta_0)$. We then compute the ratio,

$$R(\theta, \theta_0) = \frac{\sharp\{i : 1 \leq i \leq K, \hat{\rho}_n \in \hat{I}(i, \theta, \theta_0)\}}{K}, \tag{24}$$

and choose the pair $(\hat{\theta}, \hat{\theta}_0)$ minimizing

$$|R(\theta, \theta_0) - Z\%|, \tag{25}$$

noting that $\hat{\rho}_n$ is the EVE derived from the entire sample $y_1, \ldots, y_n$. With the chosen $(\hat{\theta}, \hat{\theta}_0)$, we compute $(\hat{\alpha}, \tilde{\eta}, \hat{b}_{1,0})$ and (17) again using the entire sample and the same Z, Z_1 and Z_2. If there is more than one $(\hat{\theta}, \hat{\theta}_0)$ that minimizes (25), the one leads to the shortest confidence interval is picked. In the sequel, this procedure is referred to as EV method.

Let $(\hat{\rho}_n^{(l)}, \hat{\alpha}^{(l)}, \tilde{\eta}^{(l)}, \hat{t}_R^{(l)}, \hat{t}_L^{(l)}, \hat{A}_n^{(l)})$ denote the estimate of $(\rho, \alpha_0, \eta_0, t_R, t_L, A_n)$ obtained in the lth, $1 \leq l \leq 100$, simulation run using the EV method. Since $(\hat{\rho}_n^{(l)} - \hat{t}_R^{(l)}/\hat{A}_n^{(l)}, \hat{\rho}_n^{(l)} -$

Table 1: The values of $R_i, i = 1, \ldots, 4$, $\text{MSE}_i, i = 1, 2$, and $CR_i, i = 1, 2$ under model (22) with error satisfying (23) ($n = 200$)

ρ	λ	MSE_1	R_1	MSE_2	R_2	R_3	R_4	CR_1	CR_2
0.3	0.5	1.93×10^{-4}	0.954	3.58×10^{-7}	0.998	1.829	0.497	0.92	0.93
	1	5.25×10^{-5}	0.957	6.55×10^{-9}	0.979	1.273	0.633	0.97	0.90
	1.5	3.07×10^{-5}	0.945	5.54×10^{-9}	0.960	1.168	0.649	0.92	0.85
0.5	0.5	9.89×10^{-5}	0.981	2.88×10^{-8}	0.999	1.597	0.536	0.94	0.92
	1	3.97×10^{-5}	0.949	7.74×10^{-9}	0.960	1.005	0.682	0.92	0.91
	1.5	1.73×10^{-5}	0.963	2.32×10^{-9}	0.969	1.195	0.616	0.96	0.93
0.8	0.5	1.30×10^{-5}	0.996	9.15×10^{-9}	0.999	1.595	0.535	0.96	0.85
	1	1.29×10^{-5}	0.985	2.57×10^{-9}	0.989	1.412	0.616	0.95	0.90
	1.5	2.16×10^{-5}	0.994	1.31×10^{-8}	0.999	1.688	0.533	0.95	0.93

Table 2: The values of $R_i, i = 1, \ldots, 4$, $\text{MSE}_i, i = 1, 2$, and $CR_i, i = 1, 2$ under model (22) with error satisfying (23) ($n = 400$)

ρ	λ	MSE_1	R_1	MSE_2	R_2	R_3	R_4	CR_1	CR_2
0.3	0.5	1.02×10^{-5}	0.942	8.25×10^{-10}	0.952	1.025	0.682	0.93	0.92
	1	6.90×10^{-6}	0.954	6.59×10^{-11}	0.969	1.012	0.842	0.93	0.91
	1.5	1.00×10^{-5}	0.951	4.51×10^{-10}	0.961	1.125	0.668	0.95	0.92
0.5	0.5	2.13×10^{-6}	0.988	2.67×10^{-11}	0.990	1.092	0.853	0.97	0.92
	1	3.51×10^{-6}	0.977	5.83×10^{-11}	0.982	1.073	0.835	0.96	0.92
	1.5	3.55×10^{-6}	0.977	3.19×10^{-11}	0.985	1.072	0.855	0.94	0.90
0.8	0.5	6.64×10^{-7}	0.969	1.14×10^{-11}	0.974	1.068	0.812	0.94	0.90
	1	7.49×10^{-7}	0.969	7.72×10^{-12}	0.971	1.046	0.834	0.95	0.90
	1.5	5.88×10^{-7}	0.989	4.82×10^{-12}	0.990	1.080	0.861	0.96	0.90

$\hat{t}_L^{(l)}/\hat{A}_n^{(l)}$) is used to approximate the ideal (infeasible) confidence interval, $(\hat{\rho}_n^{(l)} - t_R/A_n, \hat{\rho}_n^{(l)} -$

Table 3: The values of $R_i, i = 1, \ldots, 4$, $\mathrm{MSE}_i, i = 1, 2$, and $CR_i, i = 1, 2$ under model (22) with error satisfying (26) ($n = 200$)

ρ	α	MSE_1	R_1	MSE_2	R_2	R_3	R_4	CR_1	CR_2
0.3	0.8	4.69×10^{-5}	0.948	6.02×10^{-9}	0.940	1.533	0.428	0.93	0.90
	1.4	4.26×10^{-5}	1.001	3.07×10^{-8}	0.996	0.986	1.062	0.91	0.88
	1.8	2.63×10^{-4}	1.040	1.89×10^{-7}	0.995	0.781	1.413	0.77	0.91
0.5	0.8	3.65×10^{-5}	0.959	2.60×10^{-9}	0.970	1.528	0.435	0.92	0.85
	1.4	1.26×10^{-4}	0.979	5.72×10^{-7}	0.999	1.411	0.757	0.96	0.93
	1.8	4.83×10^{-5}	1.001	1.78×10^{-7}	0.999	0.983	1.081	0.86	0.90
0.8	0.8	6.60×10^{-6}	0.947	7.45×10^{-10}	0.950	1.490	0.396	0.89	0.84
	1.4	5.40×10^{-6}	1.000	6.69×10^{-9}	0.999	1.024	1.002	0.91	0.89
	1.8	9.11×10^{-6}	1.000	7.53×10^{-8}	0.999	0.987	1.073	0.87	0.84

Table 4: The values of $R_i, i = 1, \ldots, 4$, $\mathrm{MSE}_i, i = 1, 2$, and $CR_i, i = 1, 2$ under model (22) with error satisfying (26) ($n = 400$)

ρ	α	MSE_1	R_1	MSE_2	R_2	R_3	R_4	CR_1	CR_2
0.3	0.8	1.57×10^{-5}	0.958	1.24×10^{-10}	0.970	1.219	0.540	0.94	0.91
	1.4	1.52×10^{-5}	0.996	1.10×10^{-8}	0.999	1.047	0.988	0.91	0.89
	1.8	9.83×10^{-5}	1.023	2.24×10^{-8}	0.996	0.961	1.110	0.81	0.88
0.5	0.8	9.65×10^{-6}	0.974	8.22×10^{-11}	0.980	1.288	0.636	0.96	0.86
	1.4	8.33×10^{-6}	0.999	1.15×10^{-9}	0.999	0.968	1.128	0.93	0.91
	1.8	3.15×10^{-5}	1.006	1.71×10^{-8}	0.999	1.054	0.950	0.85	0.89
0.8	0.8	1.12×10^{-6}	0.989	4.08×10^{-12}	0.990	1.256	0.568	0.95	0.90
	1.4	1.91×10^{-6}	0.999	4.88×10^{-10}	0.999	1.021	1.049	0.93	0.90
	1.8	8.31×10^{-6}	0.999	4.52×10^{-8}	0.999	1.019	1.059	0.89	0.88

t_L/A_n), we consider the following performance measures:

$$
\begin{aligned}
R_1 &= \sum_{l=1}^{100} \frac{\hat{\rho}_n^{(l)} - \hat{t}_R^{(l)}/\hat{A}_n^{(l)}}{\hat{\rho}_n^{(l)} - t_R/A_n}, \\
R_2 &= \sum_{l=1}^{100} \frac{\hat{\rho}_n^{(l)} - \hat{t}_L^{(l)}/\hat{A}_n^{(l)}}{\hat{\rho}_n^{(l)} - t_L/A_n}, \\
R_3 &= \sum_{l=1}^{100} \frac{\hat{\alpha}^{(l)}}{\alpha_0},\ R_4 = \sum_{l=1}^{100} \frac{\tilde{\eta}^{(l)}}{\eta_0}, \\
\mathrm{MSE}_1 &= \sum_{l=1}^{100} (\hat{t}_R^{(l)}/\hat{A}_n^{(l)} - t_R/A_n)^2, \\
\mathrm{MSE}_2 &= \sum_{l=1}^{100} (\hat{t}_L^{(l)}/\hat{A}_n^{(l)} - t_L/A_n)^2.
\end{aligned}
$$

We also compute the actual coverage rates,

$$CR_1 = \frac{1}{100}\sum_{l=1}^{100} I_{\{\rho\in(\hat{\rho}_n^{(l)} - \hat{t}_R^{(l)}/\hat{A}_n^{(l)},\ \hat{\rho}_n^{(l)} - \hat{t}_L^{(l)}/\hat{A}_n^{(l)})\}},$$

for the proposed nominal 95% confidence levels. For the purpose of comparison, the actual coverage rates, CR_2, of the bootstrap procedure introduced in Theorem 3.2 of Datta and McCormick (1995) (hereafter D-M Method) are also obtained. The Z_2 and Z_1 for CR_2 are set to 99.9 and 4.9, too. For both EV and D-M methods, the commonly used $Z_2 = 97.5$ and $Z_1 = 2.5$ often lead to relatively unsatisfactory performance in terms of accuracy of actual coverage, and hence are not recommended.

The above performance measures are summarized in Table 1 ($n = 200$) and Table 2 ($n = 400$) for $(\rho, \lambda) \in \{0.3, 0.5, 0.8\} \times \{0.5, 1.0, 1.5\}$. Tables 1 and 2 show that in each (ρ, λ) combination, R_1 and R_2 are close to 1 and MSE_1 and MSE_2 are close to 0. While most values of R_3 in Table 1 are notably larger than 1, all R_3's approach 1 as n grows to 400; see Table 2. The values of R_4 in Table 1, falling between 0.49 and 0.69, are distinctly less than 1. However, they increase to around 0.85 as n increases to 400, except for the cases of $(\rho, \lambda) = (0.3, 0.5)$ and (0.3, 1.5), whose $R4$ values are slightly less than 0.7. In fact, our unreported simulation results reveal that all R_4 values are very close to 1 as n grows 1000. The actual coverage rates, CR_1, of the EV method are quite close to the nominal 95% confidence level in all cases. On the other hand, CR_2 seems slightly smaller than 95%.

We also evaluate the performance of the EV and D-M methods under (22) with gamma error whose probability density function is given by

$$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x) I_{\{x\geq 0\}}, \alpha > 0, \tag{26}$$

where $\Gamma(\cdot)$ denotes the gamma function and $\alpha \in \{0.8, 1.4, 1.8\}$. Under (22) and (26), we report the values of $R_i, i = 1, \ldots, 4$, $\text{MSE}_i, i = 1, 2$, and $CR_i, i = 1, 2$ in Table 3 ($n = 200$) and Table 4 ($n = 400$) based on 100 replications. Like Tables 1 and 2, Tables 3 and 4 also show that R_1 and R_2 are close to 1 and MSE_1 and MSE_2 are close to 0 in all cases. The values of R_3 (R_4), ranging from 0.78 to 1.54 (0.40 to 1.42), appear to be unstable when $n = 200$. However, R_3 (R_4), falling between 0.96 and 1.29 (0.54 and 1.13), becomes much more stable as n grows to 400. For all n and ρ considered in this example, CR_1 (CR_2) is generally closer to 95% than CR_2 (CR_1) when $\alpha = 0.8, 1.4$ ($\alpha = 1.8$).

4.2 Real Data Analysis

We analyze two positive-valued time series with sample sizes of $n = 82$ and 141 to demonstrate the usefulness of the EV method. The first 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. In view of Example 6.3 of Wei (2006), we only use the latest 82 data points,

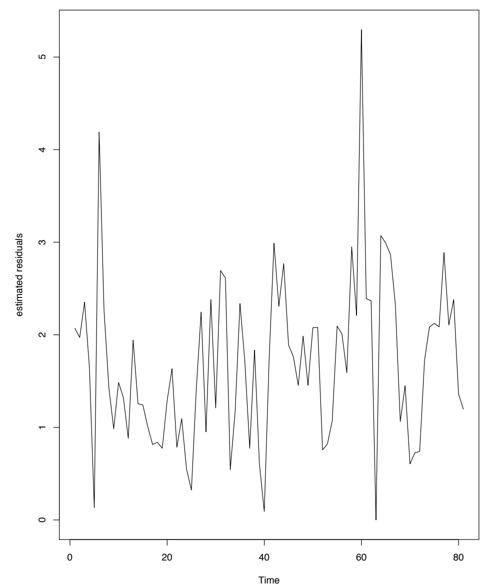

Figure 1: The EV residual plot for the blowfly data.

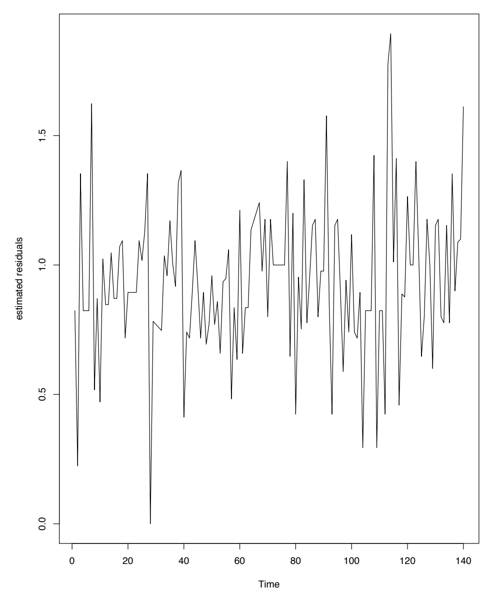

Figure 2: The EV residual plot for the viscosity data.

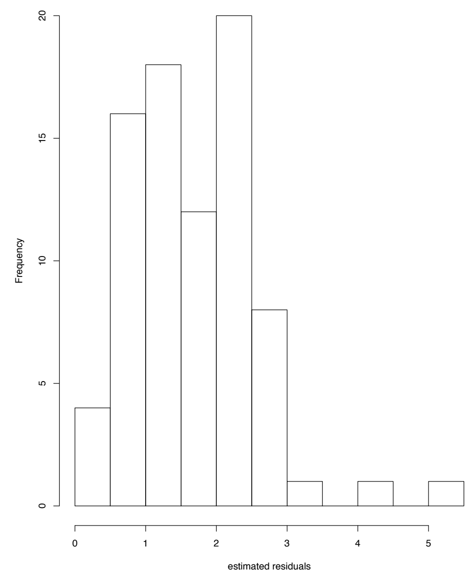

Figure 3: The histogram of the EV residuals for the blowfly data.

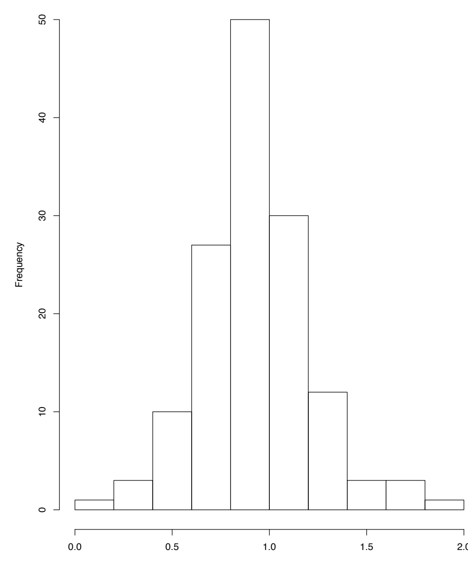

Figure 4: The histogram of the EV residuals for the viscosity data.

which are expressed in thousands in our analysis. It is mentioned in Ing and Yang (2014) that model (1) fits these data quite well; see Figure 1 for the corresponding EV residual plot. The second series is the viscosity data considered in Datta and McCormick (1995). Following Datta and McCormick (1995), we fit model (1) to the data subtracted by 1. The resultant EV residuals are plotted in Figure 2.

For both datasets, we provide confidence intervals of ρ with level 95% using the EV method, the D-M method, and the one developed from (19),

$$\big(\tilde{\rho}_n - 1.96\sqrt{\frac{1-\tilde{\rho}_n^2}{n}}, \tilde{\rho}_n + 1.96\sqrt{\frac{1-\tilde{\rho}_n^2}{n}}\big),$$

which is referred to as LS method. Note that for the EV method, the K in (24) is set to 10 instead of 20. This is because the sample sizes of these two datasets are relatively small, and hence a large K will make the sample size in each subgroup too small to produce reliable estimates.

Our analysis is summarized in Table 5. For the blowfly data, the upper half of Table 5 shows that $\hat{\rho}_n$ is notably smaller than $\tilde{\rho}_n$, but marginally included by the LS interval, whose length is much larger than that of the EV and D-M intervals. The estimated regular variation index 1.14 is only slightly larger than 1. This result appears to coincide with the histogram of the EV residuals for the blowfly data (see Figure 3), which exhibits a non-negligible frequency near the origin. The performance of the EV method is similar to that of the D-M method.

For the viscosity data, the lower half of Table 5 shows that the confidence interval provided by the D-M method is (0.81135, 0.86162). The EV interval is shifted to the right, and is a bit narrower. The LS interval is still much wider than the EV and D-M intervals in this dataset. It is worth mentioning that the $\hat{\alpha}$ obtained in this series is 2.77, suggesting that LSE is more efficient than the EVE. While this value of $\hat{\alpha}$ yields $n^{1/\hat{\alpha}} < n^{1/2}$, the length of the EV interval

$$\frac{(3.016^{1/\hat{\alpha}} - 0.001^{1/\hat{\alpha}})(\frac{\tilde{\eta}}{\hat{m}_{\alpha_0}})^{1/\hat{\alpha}}}{n^{1/\hat{\alpha}}} \tag{27}$$

is still smaller than that of the LS interval

$$\frac{3.92(1-\tilde{\rho}^2)^{1/2}}{n^{1/2}} \tag{28}$$

because the numerator of (27) is much smaller than that of (28). Unlike the histogram presented in Figure 3, the histogram of the EV residuals for the viscosity data has a very low frequency near the origin; see Figure 4. These two figures together explain why the $\hat{\alpha}$ in the first series is much smaller than in the second one.

5 Concluding Remarks

We have considered interval estimation for the AR(1) model with positive error satisfying (3) and (14). Instead of using the bootstrap scheme suggested by Datta and McCormick

Table 5: Point and Interval Estimation of ρ in the Blowfly and Viscosity Data

	Point Estimates of ρ	Interval Estimates of ρ	$\hat{\alpha}$
The Blowfly data			
EV Method	0.59951	(0.57486, 0.59948)	1.14
LS Method	0.73485	(0.58805, 0.88165)	
D-M Method	0.59951	(0.57559, 0.59490)	
The Viscosity data			
EV Method	0.88235	(0.84262, 0.88016)	2.77
LS Method	0.88155	(0.80363, 0.95948)	
D-M Method	0.88235	(0.81135, 0.86162)	

(1995), we took the approach of directly estimating the regular variation index α_0 and the associated scale parameter $\eta_0 = \alpha_0/b_{1,0}$. Our estimates are easy to compute and allow us to gain a better understanding of the behavior of the error distribution near the origin, thereby leading to a procedure for selecting the better estimate between the EVE and the LSE.

Our point estimates of α_0 and $b_{1,0}$ and interval estimate of ρ can be extended to the first-order moving-average (MA) model with positive error,

$$y_t = \varepsilon_t + \theta^{(0)}\varepsilon_{t-1},$$

when $0 \le \theta^{(0)} < 1$ is estimated by a method (denoted by $\hat{\theta}$) introduced in Section 2 of Feigin, Kratz and Resnick (1996). This is mainly because $n^{1/\alpha_0}(\hat{\theta} - \theta^{(0)})$ also has a limit Weibull distribution. For positive-valued stationary AR(p) models with $p > 1$, Feigin and Resnick (1994) provided a linear programming approach to estimate the AR coefficients. Since the rate of convergence of their estimate is still n^{1/α_0}, we believe that the corresponding AR residuals can be used in conjunction with our approach to deliver consistent estimates of α_0 and $b_{1,0}$. However, the limit distribution of their estimate, depending on the unknown distribution of p consecutive observations in a very complicated way, has no closed-form expression. Therefore, our interval estimate is no longer directly applicable in this case. To resolve this difficulty, one may construct interval estimates of the AR coefficients based on the aforementioned estimates of α_0 and $b_{1,0}$ and the *empirical* distribution of p consecutive observations. The asymptotic analysis of such an approach, however, is sufficiently complicated to warrant a separate investigation, and is left as future work.

A Proof of Theorem 3.1

Throughout this section, we shall assume (1), (3) and (14) hold. We start by proving several useful lemmas.

Lemma A.1 *Define* $\varepsilon_i^* = \min_{j\in I_i} \varepsilon_j$ *and* $A_i = \{\frac{(\hat{\rho}_n-\rho)\max_{j\in I_i} y_{j-1}}{\varepsilon_i^*} \leq \delta\}$ *for some* $0 < \delta < 1/2$. *Then*

$$\max_{1\leq i\leq m} P(A_i^c) = O(n^{-\eta_1}), \tag{A.1}$$

and

$$\max_{1\leq i\leq m} \mathrm{E}\left[\left(\frac{(\hat{\rho}_n-\rho)\max_{j\in I_i} y_{j-1}}{\varepsilon_i^*}\right)^s\right] = O\left(n^{-[\frac{1}{\alpha_0}-\theta(\frac{1}{\alpha_0}+\frac{1}{\beta_0})]s}\right), \tag{A.2}$$

where $0 < \eta_1 < \frac{1/\alpha_0}{1/\alpha_0+1/\beta_0} - \theta$, $0 < \theta < 1/2$ *is defined in Theorem 3.1, and* $0 < s < \alpha_0/2$.

PROOF. Theorem 1 of Ing and Yang (2014) ensures that for any $q > 0$,

$$\mathrm{E}(n^{1/\alpha_0}(\hat{\rho}_n - \rho))^q = O(1). \tag{A.3}$$

By (3), (14), and (A.3), one has for any $1 \leq i \leq m$,

$$\begin{aligned} P(A_i^c) &\leq P(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* \leq n_1^{-\frac{\eta_1}{\theta\alpha_0}}) + P(\max_{j\in I_i} y_{j-1} > n_1^{\frac{1}{\beta_0}+\frac{\eta_1}{\theta\beta_0}}) + P(n^{(\eta_1+\theta)(\frac{1}{\alpha_0}+\frac{1}{\beta_0})}(\hat{\rho}_n-\rho) > \delta) \\ &= O(n^{-\eta_1}) + O(n^{-[\frac{1}{\alpha_0}-(\eta_1+\theta)(\frac{1}{\alpha_0}+\frac{1}{\beta_0})]q}) = O(n^{-\eta_1}), \end{aligned}$$

where the last equality follows by taking q large enough. This completes the proof of (A.1).

To prove (A.2), we first choose $\eta_2 > 0$ such that $(2+\eta_2)s < \alpha_0$. By (3) and (14), it is not difficult to show that

$$\max_{1\leq i\leq m} \mathrm{E}(n_1^{-1/\beta_0} \max_{j\in I_i} y_j)^{\beta^*} = O(1) \text{ and } \max_{1\leq i\leq m} \mathrm{E}(n_1^{1/\alpha_0}\varepsilon_i^*)^{-\alpha^*} = O(1), \tag{A.4}$$

where $0 < \beta^* < \beta_0$ and $0 < \alpha^* < \alpha_0$. It follows from (A.3), (A.4), and Hölder's inequality that

$$\begin{aligned} &\max_{1\leq i\leq m} \mathrm{E}\left[\left(\frac{(\hat{\rho}_n-\rho)\max_{j\in I_i} y_{j-1}}{\varepsilon_i^*}\right)^s\right] = \max_{1\leq i\leq m} \mathrm{E}\left[\frac{(n^{\frac{1}{\alpha_0}+\frac{1}{\beta_0}}(\hat{\rho}_n-\rho)n_1^{-\frac{1}{\beta_0}}\max_{j\in I_i} y_{j-1})^s}{(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*)^s}\right] \\ &\leq \left\{\mathrm{E}\left[n^{-[\frac{1}{\alpha_0}-\theta(\frac{1}{\alpha_0}+\frac{1}{\beta_0})]}n^{\frac{1}{\alpha_0}}(\hat{\rho}_n-\rho)\right]^{s\frac{2+\eta_2}{\eta_2}}\right\}^{\frac{\eta_2}{2+\eta_2}} \\ &\times \left\{\max_{1\leq i\leq m} \mathrm{E}\left(n_1^{-\frac{1}{\beta_0}}\max_{j\in I_i} y_{j-1}\right)^{s(2+\eta_2)}\right\}^{\frac{1}{2+\eta_2}} \times \left\{\max_{1\leq i\leq m} \mathrm{E}\left(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\right)^{-s(2+\eta_2)}\right\}^{\frac{1}{2+\eta_2}} \\ &= O\left(n^{-[\frac{1}{\alpha_0}-\theta(\frac{1}{\alpha_0}+\frac{1}{\beta_0})]s}\right), \end{aligned}$$

which yields the desired conclusion (A.2).

Lemma A.2 *There exists some* $\kappa_1 > 0$ *such that*

$$\sup_{\mathbf{v}\in\mathbf{H}} \Big| l(\alpha,\eta|w_i^*, i=1,\cdots,m) - l(\alpha,\eta|n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*, i=1,\cdots,m)\Big| = O_p(n^{-\kappa_1}), \tag{A.5}$$

recalling that $w_i^* = n_1^{\frac{1}{\alpha_0}}\hat{\varepsilon}_i^{\circ}$.

PROOF. Straightforward calculations imply that the left-hand side of (A.5) is bounded above by

$$\begin{aligned} & C\Big\{\Big|\frac{1}{m}\sum_{i=1}^m (\log\hat{\varepsilon}_i^{\circ} - \log\varepsilon_i^*)\, I_{A_i}\Big| + \Big|\frac{1}{m}\sum_{i=1}^m (\log\hat{\varepsilon}_i^{\circ} - \log\varepsilon_i^*)\, I_{A_i^c}\Big| \\ & + \sup_{\alpha\in\mathbf{H}_1}\Big|\frac{1}{m}\sum_{i=1}^m \big[\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\alpha} - \big(w_i^*\big)^{\alpha}\big] I_{A_i}\Big| + \sup_{\alpha\in\mathbf{H}_1}\Big|\frac{1}{m}\sum_{i=1}^m \big[\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\alpha} - \big(w_i^*\big)^{\alpha}\big] I_{A_i^c}\Big|\Big\} \\ & \equiv \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)}, \end{aligned} \tag{A.6}$$

where C here and hereafter stands for a generic positive constant independent of the sample size n. Define $B_i = \{\varepsilon_i^* \le n_1^{-\frac{1}{\alpha_0}-\frac{s_1}{\alpha_0\theta}}\}$, where $0 < s_1 < \alpha_0\theta_0 - \theta$ and the positivity of $\alpha_0\theta_0 - \theta$ is ensured by $\theta_0 > \theta/\underline{\alpha}$. Then,

$$\begin{aligned} \text{(I)} \le & \Big|\frac{1}{m}\sum_{i=1}^m \big(\log n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* - \log n_1^{\frac{1}{\alpha_0}}\hat{\varepsilon}_i^{\circ}\big) I_{A_i} I_{B_i}\Big| \\ & + \Big|\frac{1}{m}\sum_{i=1}^m \big(\log n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* - \log n_1^{\frac{1}{\alpha_0}}\hat{\varepsilon}_i^{\circ}\big) I_{A_i} I_{B_i^c}\Big| \equiv \text{(a)} + \text{(b)}. \end{aligned} \tag{A.7}$$

By observing

$$\begin{aligned} & \hat{\varepsilon}_i^* \ge (1-\delta)\varepsilon_i^* \ge 0 \text{ on } A_i, \\ & \hat{\varepsilon}_i^* \ge (1-\delta)n^{-\frac{1}{\alpha_0}(\theta+s_1)} > n^{-\theta_0} \text{ on } B_i^c \cap A_i, \\ & \max_{1\le i\le m} P(B_i) \le n_1 P(\varepsilon_1 \le n_1^{-\frac{1}{\alpha_0}-\frac{s_1}{\alpha_0\theta}}) \le C n_1^{-\frac{s_1}{\theta}} = O(n^{-s_1}), \\ & \varepsilon_i^* - \hat{\varepsilon}_i^* \le (\hat{\rho}_n - \rho)\max_{j\in I_i} y_{j-1}, \end{aligned} \tag{A.8}$$

and making use of (A.2), it can be shown that for any $0 < s_1^* < s_1$ and $0 < \nu < \alpha_0^* < \alpha_0$ satisfying $\nu/\alpha_0^* = 1 - (s_1^*/s_1)$,

$$\begin{aligned} \mathrm{E}((\text{a})) \le & \frac{1}{m}\sum_{i=1}^m \mathrm{E}\Big[\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{-\nu} I_{B_i}\Big] + O(n^{-s_1}\log n) \\ \le & \frac{1}{m}\sum_{i=1}^m \Big[E\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{-\alpha_0^*}\Big]^{\nu/\alpha_0^*} P^{\frac{\alpha_0^*-\xi}{\alpha_0^*}}(B_i) + O(n^{-s_1}\log n) = O(n^{-s_1^*}), \end{aligned} \tag{A.9}$$

and for any $0 < s_2 < \min\{\alpha_0/2, 1\}$,

$$\begin{aligned} \mathrm{E(b)} &= \frac{1}{m}\sum_{i=1}^{m} \mathrm{E}\Big\{ \log\big(1 + (n_1^{1/\alpha_0}\varepsilon_i^* - n_1^{1/\alpha_0}\hat{\varepsilon}_i^*)/n_1^{1/\alpha_0}\hat{\varepsilon}_i^*\big) I_{A_i}\Big\} \\ &\leq \frac{1}{m}\sum_{i=1}^{m} \mathrm{E}\Big\{ (n_1^{1/\alpha_0}\varepsilon_i^* - n_1^{1/\alpha_0}\hat{\varepsilon}_i^*) I_{A_i}/[(1-\delta)n_1^{1/\alpha_0}\varepsilon_i^*]\Big\} \\ &\leq \frac{1}{(1-\delta)m}\sum_{i=1}^{m} \mathrm{E}\Big\{ [(\hat{\rho}_n - \rho)\max_{j\in I_i} y_{j-1}/\varepsilon_i^*] I_{A_i}\Big\} \\ &= O\left(n^{-[\frac{1}{\alpha_0} - \theta(\frac{1}{\alpha_0}+\frac{1}{\beta_0})]s_2}\right). \end{aligned} \tag{A.10}$$

In view of (A.7), (A.9) and (A.10),

$$\mathrm{(I)} = O_p(n^{-\kappa_1^*}), \tag{A.11}$$

where $\kappa_1 = \min\{[\alpha_0^{-1} - \theta(\alpha_0^{-1} + \beta_0^{-1})]s_2, s_1^*\}$.

To deal with (II), note that

$$\mathrm{(II)} \leq \Big|\frac{1}{m}\sum_{i=1}^{m} \log n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* I_{A_i^c}\Big| + \Big|\frac{1}{m}\sum_{i=1}^{m} \log n_1^{\frac{1}{\alpha_0}}\hat{\varepsilon}_i^\circ I_{A_i^c}\Big| \equiv \mathrm{(c)} + \mathrm{(d)}$$

By (A.1), $\hat{\varepsilon}_i^\circ \geq n^{-\theta_0}$,

$$\max_{1\leq i\leq m} \mathrm{E}(n_1^{1/\alpha_0}\varepsilon_i^*)^q = O(1) \text{ for any } q > 0, \tag{A.12}$$

and some algebraic manipulations, it can be shown that $\mathrm{E(c)} = O(n^{-\eta^*})$ and $\mathrm{E(d)} = O(n^{-\eta^*})$ for some $0 < \eta^* < \eta_1$. As a result,

$$\mathrm{(II)} = O_p(n^{-\eta^*}). \tag{A.13}$$

To deal with (III), note that for any $\alpha \in \mathbf{H}_1$,

$$\begin{aligned} &\frac{1}{m}\sum_{i=1}^{m} \big(w_i^*\big)^{\alpha} I_{A_i^c}\big(I_{\hat{\varepsilon}_i^* \leq n^{-\theta_0}} + I_{\hat{\varepsilon}_i^* > n^{-\theta_0}}\big) \\ &\leq \frac{1}{m}\sum_{i=1}^{m} \big(n_1^{\frac{1}{\alpha_0}} n^{-\theta_0}\big)^{\underline{\alpha}} I_{A_i^c} + \frac{1}{m}\sum_{i=1}^{m} \big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\alpha} I_{A_i^c} \\ &\leq \frac{n_1^{-\underline{\alpha}(\frac{\theta_0}{\theta} - \frac{1}{\alpha_0})}}{m}\sum_{i=1}^{m} I_{A_i^c} + \frac{1}{m}\sum_{i=1}^{m} \big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\bar{\alpha}} I_{A_i^c} I_{n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* \geq 1} \\ &+ \frac{1}{m}\sum_{i=1}^{m} \big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\underline{\alpha}} I_{A_i^c} I_{n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* < 1}. \end{aligned} \tag{A.14}$$

Let $0 < \eta_1^* < \eta_1$ and $q = (1 - \eta_1^*/\eta_1)^{-1}$. Then, by (A.14), (A.12), and (A.1),

$$\mathrm{E}\Big\{ \sup_{\alpha \in \mathbf{H}_1} \frac{1}{m} \sum_{i=1}^{m} \big(w_i^*\big)^{\alpha} I_{A_i^c} \Big\}$$
$$\leq n_1^{-\underline{\alpha}(\frac{\theta_0}{\theta} - \frac{1}{\alpha_0})} P(A_i^c) + \frac{1}{m} \sum_{i=1}^{m} \Big[\mathrm{E}\big(n_1^{\frac{1}{\alpha_0}} \varepsilon_i^*\big)^{\underline{\alpha} q}\Big]^{\frac{1}{q}} P^{\frac{q-1}{q}}(A_i^c) + P(A_i^c) = O(n^{-\eta_1^*}),$$

yielding

$$\sup_{\alpha \in \mathbf{H}_1} \frac{1}{m} \sum_{i=1}^{m} \big(w_i^*\big)^{\alpha} I_{A_i^c} = O_p(n^{-\eta_1^*}).$$

Similarly,

$$\sup_{\alpha \in \mathbf{H}_1} \frac{1}{m} \sum_{i=1}^{m} \big(n_1^{\frac{1}{\alpha_0}} \varepsilon_i^*\big)^{\alpha} I_{A_i^c} = O_p(n^{-\eta_1^*}).$$

Consequently,

$$(\mathrm{III}) = O_p(n^{-\eta_1^*}). \tag{A.15}$$

By Taylor's theorem, one has for any $\alpha \in \mathbf{H}_1$ and $i = 1, \ldots, m$,

$$(n_1^{\frac{1}{\alpha_0}} \varepsilon_i^*)^{\alpha} - (n_1^{\frac{1}{\alpha_0}} \hat{\varepsilon}_i^*)^{\alpha} = \alpha (n_1^{\frac{1}{\alpha_0}} X_{i,\alpha})^{\alpha - 1} (n_1^{\frac{1}{\alpha_0}} \varepsilon_i^* - n_1^{\frac{1}{\alpha_0}} \hat{\varepsilon}_i^*), \tag{A.16}$$

where $X_{i,\alpha} \in [\hat{\varepsilon}_i^*, \varepsilon_i^*]$. Armed with (A.16) and the first and the last equations of (A.8), it follows that for any $\alpha \in \mathbf{H}_1$,

$$\Big| \frac{1}{m} \sum_{i=1}^{m} \Big[\big(n_1^{\frac{1}{\alpha_0}} \varepsilon_i^*\big)^{\alpha} - \big(w_i^*\big)^{\alpha}\Big] I_{A_i}\big(I_{\hat{\varepsilon}_i^* \leq n^{-\theta_0}} + I_{\hat{\varepsilon}_i^* > n^{-\theta_0}}\big) \Big|$$
$$\leq \frac{1}{m} \sum_{i=1}^{m} \Big[\big(n_1^{\frac{1}{\alpha_0}} (1-\delta)^{-1} n_1^{-\frac{\theta_0}{\theta}}\big)^{\underline{\alpha}} + \big(n_1^{\frac{1}{\alpha_0} - \frac{\theta_0}{\theta}}\big)^{\underline{\alpha}}\Big] + \frac{1}{m} \sum_{i=1}^{m} \Big[\big(n_1^{\frac{1}{\alpha_0}} \varepsilon_i^*\big)^{\alpha} - \big(n_1^{\frac{1}{\alpha_0}} \hat{\varepsilon}_i^*\big)^{\alpha}\Big] I_{A_i}$$
$$\leq O\Big(n_1^{-(\frac{\theta_0}{\theta} - \frac{1}{\alpha_0})\underline{\alpha}}\Big) + \frac{1}{m} \sum_{i=1}^{m} \alpha (n_1^{\frac{1}{\alpha_0}} X_{i,\alpha})^{\alpha - 1} \big(n_1^{\frac{1}{\alpha_0}} \varepsilon_i^* - n_1^{\frac{1}{\alpha_0}} \hat{\varepsilon}_i^*\big) I_{A_i}$$
$$\leq O\Big(n^{-\theta(\frac{\theta_0}{\theta} - \frac{1}{\alpha_0})\underline{\alpha}}\Big) + \frac{1}{m(1-\delta)} \sum_{i=1}^{m} \Big[\big(n_1^{\frac{1}{\alpha_0}} \varepsilon_i^*\big)^{\bar{\alpha}} + \big(n_1^{\frac{1}{\alpha_0}} \varepsilon_i^*\big) + 1\Big] \Big(\frac{(\hat{\rho}_n - \rho) \max_{j \in I_i} y_{j-1}}{\varepsilon_i^*}\Big) I_{A_i},$$

which, together with (A.2) and (A.12), yields for any $0 < s_2 < \min\{\alpha_0/2, 1\}$,

$$\mathrm{E}(\mathrm{IV}) = O\Big(n^{-\theta(\frac{\theta_0}{\theta} - \frac{1}{\alpha_0})\underline{\alpha}}\Big) + O\Big(n^{-[\frac{1}{\alpha_0} - \theta(\frac{1}{\alpha_0} + \frac{1}{\beta_0})] s_2}\Big).$$

Hence

$$(\mathrm{IV}) = O_p\Big(n^{-\kappa_1^{\circ}}\Big), \tag{A.17}$$

where $\kappa_1^{\circ} = \min\{\theta[(\theta_0/\theta) - (1/\alpha_0)]\underline{\alpha}, [\alpha_0^{-1} - \theta(\alpha_0^{-1} + \beta_0^{-1})] s_2\}$. Combining (A.6), (A.11), (A.13), (A.15), and (A.17) yields (A.5) with $\kappa_1 = \min\{\kappa_1^*, \kappa_1^{\circ}, \eta^*, \eta_1^*\}$.

Lemma A.3 *Let $\{w_t\}$ be a sequence of i.i.d. random variables defined on the probability space the same as that of $\{y_t\}$. Suppose the density function of w_1 satisfies (8). Then, there exists $\kappa_2 > 0$ such that*

$$\sup_{\mathbf{v}\in\mathbf{H}}\Big|l(\alpha,\eta|n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*, i=1,\cdots,m) - l(\alpha,\eta|w_i, i=1,\cdots,m)\Big| = O_p(n^{-\kappa_2}). \tag{A.18}$$

PROOF. A straightforward algebraic manipulation yields

$$\begin{aligned}
&\sup_{\mathbf{v}\in\mathbf{H}}\Big|l(\alpha,\eta|n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*, i=1,\cdots,m) - l(\alpha,\eta|w_i, i=1,\cdots,m)\Big| \\
&\le C\Big|\frac{1}{m}\sum_{i=1}^m\big(\log n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* - \log w_i\big)\Big| + \sup_{\alpha\in\mathbf{H}_1}\Big|\frac{1}{m}\sum_{i=1}^m\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^\alpha - \big(w_i\big)^\alpha\Big| \\
&\le C\Big\{\sup_{\alpha\in\mathbf{H}_1}\Big|\mathrm{E}\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^\alpha - \mathrm{E}w_i^\alpha\Big| + \Big|\mathrm{E}\big(\log n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big) - \mathrm{E}\big(\log w_i\big)\Big| \\
&+ \sup_{\alpha\in\mathbf{H}_1}\Big|\frac{1}{m}\sum_{i=1}^m\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^\alpha - \mathrm{E}\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^\alpha\Big| + \sup_{\alpha\in\mathbf{H}_1}\Big|\frac{1}{m}\sum_{i=1}^m w_i^\alpha - \mathrm{E}w_i^\alpha\Big| \\
&+\Big|\frac{1}{m}\sum_{i=1}^m\log w_i - \mathrm{E}(\log w_i)\Big| + \Big|\frac{1}{m}\sum_{i=1}^m\log n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* - \mathrm{E}\big(\log n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)\Big|\Big\} \\
&\equiv C\{(\mathrm{A})+(\mathrm{B})+(\mathrm{C})+(\mathrm{D})+(\mathrm{E})+(\mathrm{F})\},
\end{aligned} \tag{A.19}$$

We first show that

$$(\mathrm{A}) = O(n^{-\theta\min\{\nu/\alpha_0,1\}}). \tag{A.20}$$

Let $0 < \xi < (\underline{\alpha}/\alpha_0)\min\{1/2, \nu/(\bar{\alpha}+\nu)\}$, where ν is defined in (14). Then,

$$\begin{aligned}
\Big|\mathrm{E}\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^\alpha - \mathrm{E}w_i^\alpha\Big| &\le \Big|\int_0^{n_1^\xi} P\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* > t^{\frac{1}{\alpha}}\big) - \exp\big(-\frac{b_{1,0}}{\alpha_0}t^{\frac{\alpha_0}{\alpha}}\big)dt\Big| \\
&+ \int_{n_1^\xi}^\infty P\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* > t^{\frac{1}{\alpha}}\big)dt + \int_{n_1^\xi}^\infty \exp\big(-\frac{b_{1,0}}{\alpha_0}t^{\frac{\alpha_0}{\alpha}}\big)dt,
\end{aligned} \tag{A.21}$$

Since (14) implies that for any small $\varepsilon > 0$, there exists $\delta > 0$ such that for all $0 < x < \delta$,

$$b_{1,0}x^{\alpha_0-1} - \varepsilon b_{1,0}x^{\alpha_0+\nu-1} \le f_\varepsilon(x) \le b_{1,0}x^{\alpha_0-1} + \varepsilon b_{1,0}x^{\alpha_0+\nu-1}, \tag{A.22}$$

one has for $0 < t \le n_1^\xi$,

$$\left(1 - \big(\frac{b_{1,0}t^{\frac{\alpha_0}{\alpha}}}{\alpha_0 n_1} + \frac{\varepsilon b_{1,0}t^{\frac{\alpha_0+\nu}{\alpha}}}{(\alpha_0+\nu)n_1^{1+\frac{\nu}{\alpha_0}}}\big)\right)^{n_1} \le P\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* > t^{\frac{1}{\alpha}}\big) \le \left(1 - \big(\frac{b_{1,0}t^{\frac{\alpha_0}{\alpha}}}{\alpha_0 n_1} - \frac{\varepsilon b_{1,0}t^{\frac{\alpha_0+\nu}{\alpha}}}{(\alpha_0+\nu)n_1^{1+\frac{\nu}{\alpha_0}}}\big)\right)^{n_1},$$

yielding

$$\Big|P\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* > t^{\frac{1}{\alpha}}\big) - \exp\big(-\frac{b_{1,0}}{\alpha_0}t^{\frac{\alpha_0}{\alpha}}\big)\Big| \le \exp\big(-\frac{b_{1,0}}{\alpha_0}t^{\frac{\alpha_0}{\alpha}}\big)\max\left\{\frac{t^{\frac{\alpha_0+\nu}{\alpha}}}{n_1^{\frac{\nu}{\alpha_0}}}, \frac{t^{\frac{2\alpha_0}{\alpha}}}{n_1}\right\}.$$

As a result,

$$\sup_{\alpha\in\mathbf{H}_1}\Big|\int_0^{n_1^{\xi}} P\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* > t^{\frac{1}{\alpha}}\big) - \exp\big(-\frac{b_{1,0}}{\alpha_0}t^{\frac{\alpha_0}{\alpha}}\big)dt\Big| = O(n^{-\theta\min\{1,\nu/\alpha_0\}}). \tag{A.23}$$

A simple calculation gives

$$\sup_{\alpha\in\mathbf{H}_1}\int_{n_1^{\xi}}^{\infty}\exp\big(-\frac{b_{1,0}}{\alpha_0}t^{\frac{\alpha_0}{\alpha}}\big)dt = o(n^{-\theta\min\{1,\nu/\alpha_0\}}). \tag{A.24}$$

Let $q>1$ and δ_1 be a small positive number. Then for all large n, there exist $c_1, c_2 > 0$ such that

$$\begin{aligned}\int_{n_1^{\xi}}^{\infty} P\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* > t^{\frac{1}{\alpha}}\big)dt &\le \int_{n_1^{\xi}}^{(\delta_1 n_1^{\frac{1}{\alpha_0}})^{\alpha}} P^{n_1}\Big(\varepsilon_1 > \frac{t^{\frac{1}{\alpha}}}{n_1^{\frac{1}{\alpha_0}}}\Big)dt \\ &+ \int_{(\delta_1 n_1^{\frac{1}{\alpha_0}})^{\alpha}}^{\infty} P^{\frac{q\bar{\alpha}}{\beta_0}}\Big(\varepsilon_1^{\beta_0} > \frac{t^{\frac{\beta_0}{\alpha}}}{n_1^{\frac{\beta_0}{\alpha_0}}}\Big)P^{n_1-\frac{q\bar{\alpha}}{\beta_0}}\big(\varepsilon_1 > \delta_1\big)dt \\ &\le \int_{n_1^{\xi}}^{\infty}\exp\big(-c_1 t^{\frac{\alpha_0}{\bar{\alpha}}}\big)dt + C\big(1-c_2\delta_1^{\alpha_0}\big)^{n_1-\frac{q\bar{\alpha}}{\beta_0}} n_1^{\frac{q\bar{\alpha}}{\alpha_0}}\int_1^{\infty} t^{-q}dt,\end{aligned}$$

yielding

$$\sup_{\alpha\in\mathbf{H}_1}\int_{n_1^{\xi}}^{\infty} P\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* > t^{\frac{1}{\alpha}}\big)dt = o(n^{-\theta\min\{1,\nu/\alpha_0\}}). \tag{A.25}$$

Consequently, (A.20) follows from (A.21) and (A.23)–(A.25). Moreover, by (A.22) and an argument similar to that used to prove (A.20), it can be shown that

$$(\mathrm{B}) = O(n^{-\theta\min\{\nu/\alpha_0,1\}}). \tag{A.26}$$

By an argument given in Chan and Ing (2011) (which uses an integral norm to dominate the supremum norm), one obtains

$$\begin{aligned}&\mathrm{E}\Big\{\sup_{\alpha\in\mathbf{H}_1}\Big(\frac{1}{m}\sum_{i=1}^{m}\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\alpha} - \mathrm{E}\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\alpha}\Big)^2\Big\} \\ &\le C\left\{\sup_{\alpha\in\mathbf{H}_1}\mathrm{E}\Big(\frac{1}{m}\sum_{i=1}^{m}\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\alpha}\log n_1^{\frac{1}{\alpha_0}}\varepsilon_i^* - \mathrm{E}\big[\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\alpha}\log n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big]\Big)^2\right. \\ &\left.+\sup_{\alpha\in\mathbf{H}_1}\mathrm{E}\Big\{\Big(\frac{1}{m}\sum_{i=1}^{m}\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\alpha} - \mathrm{E}\big(n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*\big)^{\alpha}\Big)^2\Big\}\right\}.\end{aligned} \tag{A.27}$$

Since ε_i^* are i.i.d. and

$$\sup_{\alpha\in\mathbf{H}_1} \mathrm{E}\left[\left(n_1^{\frac{1}{\alpha_0}}\varepsilon_1^*\right)^\alpha \log n_1^{\frac{1}{\alpha_0}}\varepsilon_1^*\right]^2 < \infty \text{ and } \sup_{\alpha\in\mathbf{H}_1} \mathrm{E}\left(n_1^{\frac{1}{\alpha_0}}\varepsilon_1^*\right)^{2\alpha} < \infty,$$

(A.27) leads to

$$(\mathrm{C}) = O_p(m^{-1/2}) = O_p(n^{-(1-\theta)/2}). \tag{A.28}$$

Similarly, it can be shown that

$$(\mathrm{D}) = O_p(n^{-(1-\theta)/2}). \tag{A.29}$$

Moreover, it is easy to see that

$$(\mathrm{E}) = O_p(n^{-(1-\theta)/2}) \text{ and } (\mathrm{F}) = O_p(n^{-(1-\theta)/2}), \tag{A.30}$$

which, together with (A.19), (A.20), (A.26), (A.28), and (A.29), implies (A.18) holds with $\kappa_2 = \min\{(1-\theta)/2, \theta\min\{\nu/\alpha_0, 1\}\}$.

Lemma A.4 *Let $\delta_n = n^{-\kappa_3}$, where $0 < \kappa_3 < \frac{\min\{\kappa_1,\kappa_2\}}{2}$. Then*

$$\lim_{n\to\infty} P\Big(\sup_{\mathbf{v}\in\mathbf{H}-B_{\delta_n}(\mathbf{v}_0)} l(\alpha,\eta|w_i^*, i=1,\ldots,m) < l(\alpha_0,\eta_0|w_i^*, i=1,\ldots,m)\Big) = 1, \tag{A.31}$$

where $\mathbf{v}_0 = (\alpha_0,\eta_0)^\top$ and $B_\delta(\mathbf{v}_0) = \{\mathbf{v} : \|\mathbf{v}-\mathbf{v}_0\| < \delta\}$ with $\|\cdot\|$ denoting the Euclidean norm. Therefore,

$$\|(\hat{\alpha}^* - \alpha_0, \hat{\eta}^* - \eta_0)^\top\| = O_p(n^{-\kappa_3}). \tag{A.32}$$

PROOF. We first show that for any $\delta > 0$,

$$P\Big(\sup_{\mathbf{v}\in\mathbf{H}-B_\delta(\mathbf{v}_0)} l(\alpha,\eta|w_i^*, i=1,\cdots,m) - l(\alpha_0,\eta_0|w_i^*, i=1,\cdots,m) \geq 0\Big) \to 0. \tag{A.33}$$

Since the event on the left-hand side of (A.33) implies

$$\begin{aligned}
&2\sup_{\mathbf{v}\in\mathbf{H}}\Big|l(\alpha,\eta|w_i^*, i=1,\cdots,m) - l(\alpha,\eta|n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*, i=1,\cdots,m)\Big| \\
&+ 2\sup_{\mathbf{v}\in\mathbf{H}}\Big|l(\alpha,\eta|n_1^{\frac{1}{\alpha_0}}\varepsilon_i^*, i=1,\cdots,m) - l(\alpha,\eta|w_i, i=1,\cdots,m)\Big| \\
&\geq l(\alpha_0,\eta_0|w_i, i=1,\cdots,m) - \sup_{\mathbf{v}\in\mathbf{H}-B_\delta(\mathbf{v}_0)} l(\alpha,\eta|w_i^*, i=1,\cdots,m),
\end{aligned}$$

in view of Lemmas A.2 and A.3, (A.33) follows from

$$P\Big(l(\alpha_0,\eta_0|w_i, i=1,\cdots,m) - \sup_{\mathbf{v}\in\mathbf{H}-B_\delta(\mathbf{v}_0)} l(\alpha,\eta|w_i^*, i=1,\cdots,m) > c_\delta\Big) \to 1, \tag{A.34}$$

as $n \to \infty$, where c_δ is some positive constant depending on δ.

Define

$$g(\alpha, \eta) = \log \frac{\alpha}{\eta} + \alpha \mathrm{E}(\log w_i) - \frac{\eta_0^{\frac{\alpha}{\alpha_0}} \Gamma(\frac{\alpha}{\alpha_0} + 1)}{\eta},$$

noting that $\mathrm{E}(w_i^m) = \eta_0^{\frac{m}{\alpha_0}} \Gamma(\frac{m}{\alpha_0} + 1)$ and $\mathrm{E}(\log w_i) = \frac{1}{\alpha_0}(-\gamma + \log \eta_0)$, in which $\gamma \sim 0.5772$ is the Euler-Mascheoni constant. Then,

$$\begin{aligned} & l(\alpha_0, \eta_0 | w_i, i = 1, \cdots, m) - l(\alpha, \eta | w_i^*, i = 1, \cdots, m) \\ & = (\alpha_0 - \alpha) \frac{1}{m} \sum_{i=1}^m (\log w_i - \mathrm{E}(\log w_i)) - \frac{1}{m} \sum_{i=1}^m \left(\frac{w_i^{\alpha_0}}{\eta_0} - 1 \right) \\ & + \frac{1}{m} \sum_{i=1}^m \left(\frac{w_i^\alpha}{\eta} - \frac{\mathrm{E}(w_i^\alpha)}{\eta} \right) + g(\alpha_0, \eta_0) - g(\alpha, \eta). \end{aligned} \tag{A.35}$$

Some elementary but tedious analysis shows that (i) (α_0, η_0) is the only point satisfying $\partial g(\alpha, \eta)/\partial \alpha = 0$ and $\partial g(\alpha, \eta)/\partial \eta = 0$, (ii) (α_0, η_0) is the unique maximizer of $g(\alpha, \eta)$, and (iii) there exists small positive constant ι such that for all $\mathbf{v} \in B_\iota(\mathbf{v}_0)$,

$$g(\alpha_0, \eta_0) - g(\alpha, \eta) \geq c_\iota \|\mathbf{v} - \mathbf{v}_0\|^2, \tag{A.36}$$

where c_ι is some positive constant depending on ι. Since $g(\alpha, \eta)$ is differentiable on $\mathbf{H}$, properties (i) and (ii) above ensure that for any $\delta > 0$, there is $\bar{c}_\delta > 0$ such that

$$g(\alpha_0, \eta_0) - \sup_{\mathbf{v} \in \mathbf{H} - B_\delta(\mathbf{v}_0)} g(\alpha, \eta) \geq \bar{c}_\delta \tag{A.37}$$

By an argument similar to that used to prove (A.29), one obtains

$$\sup_{\mathbf{v} \in \mathbf{H}} \left| \frac{1}{m} \sum_{i=1}^m \frac{w_i^\alpha}{\eta} - \frac{\mathrm{E}(w_i^\alpha)}{\eta} \right| = O_p(n^{-\frac{1-\theta}{2}}). \tag{A.38}$$

Combining (A.38), (A.37), (A.35), and (A.30) yields that (A.34) holds with $c_\delta = \bar{c}_\delta/2$. Thus, (A.33) is proved.

Recall the definition of ι in property (iii). With the help of (A.33), (A.31) is guaranteed by

$$P\Big(\sup_{\mathbf{v} \in B_\iota(\mathbf{v}_0) - B_{\delta_n}(\mathbf{v}_0)} l(\alpha, \eta | w_i^*, i = 1, \ldots, m) \geq l(\alpha_0, \eta_0 | w_i^*, i = 1, \ldots, m) \Big) = o(1),$$

which is, in turn, implied by (A.38), (A.36), (A.35), (A.30), and Lemmas A.2 and A.3. This completes the proof of the lemma.

We are now ready to prove Theorem 3.1.

PROOF OF THEOREM 3.1. It is mentioned in Section 2 that $\hat{\alpha} = \hat{\alpha}^*$. By (A.32),

$$\hat{\alpha} - \alpha_0 = O_p(n^{-\kappa_3}). \tag{A.39}$$

In addition, since $\hat{\eta}^* - \tilde{\eta} = \hat{\eta}^*(1 - n_1^{1-\hat{\alpha}/\alpha_0})$, it follows from (A.32) that

$$\tilde{\eta} - \eta_0 = O_p(n^{-\kappa_3} \log n),$$

which, together with (A.39), leads to

$$\hat{b}_{1,0} - b_{1,0} = O_p(n^{-\kappa_3} \log n).$$

Consequently, (15) holds with $\kappa = \kappa_3$.

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.

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., ING, C.-K. AND ZHANG, RM (2017). Nearly unstable processes: a prediction perspective. *Statistica Sinica*, to appear.

DATTA, S. AND MCCORMICK, W. P. (1995). Bootstrap inference for a first-order autoregression with positive innovations. *Journal of the American Statistical Association* **90**, 1289–1300.

FEIGIN, P. D AND RESNICK, S. I. (1994). Limit distributions for linear programming time series estimators. *Stochastic Processes and their Applications* **51**, 135–165.

FEIGIN, P. D., KRATZ, M. F., AND RESNICK, S. I. (1996). Parameter estimation for moving averages with positive innovations. *Annals of Applied Probability* **6**, 1157–1190.

DAVIS, R. AND MCCORMICK, W. P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Processes and their Applications* **31**, 237–250.

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

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.

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

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.

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

Wei, W. W. S. (2006). Time series analysis: univariate and multivariate methods. *Boston, MA: Pearson*.