1-2015

Linear programming-based estimators in nonnegative autoregression

Daniel P. A. PREVE

Singapore Management University, dpreve@smu.edu.sg

Citation

PREVE, Daniel P. A.. Linear programming-based estimators in nonnegative autoregression. (2015). *Journal of Banking and Finance*. 61, (2), S225-S234. Research Collection School Of Economics.

Available at: https://ink.library.smu.edu.sg/soe_research/2330

LINEAR PROGRAMMING-BASED ESTIMATORS IN NONNEGATIVE AUTOREGRESSION

DANIEL PREVE†
FEBRUARY 19, 2014

Abstract. This note studies robust estimation of the autoregressive (AR) parameter in a nonlinear, nonnegative AR model. It is shown that a linear programming estimator (LPE), considered by Nielsen and Shephard (2003) among others, remains consistent under severe model misspecification. Consequently, the LPE can be used to seek sources of misspecification and to isolate certain trend, seasonal or cyclical components. Simple and quite general conditions under which the LPE is strongly consistent in the presence of heavy-tailed, serially correlated, heteroskedastic disturbances are given, and a brief review of the literature on LP-based estimators in nonnegative autoregression is presented. Finite-sample properties of the LPE are investigated in a small scale simulation study.

1. Introduction

In the last decades, nonlinear and nonstationary time series analysis have gained much attention. This attention is mainly motivated by evidence that many real life time series are non-Gaussian with a structure that evolves over time. For example, many economic time series are known to show nonlinear features such as cycles, asymmetries, time irreversibility, jumps, thresholds, heteroskedasticity and combinations thereof. This note considers robust estimation in a (potentially) misspecified nonlinear, nonnegative autoregressive model, that may be a useful tool for describing the behavior of a broad class of nonnegative time series.

For nonlinear time series models it is common to assume that the disturbances are i.i.d. with zero-mean and finite variance. Recently, however, there has been considerable interest in nonnegative models. See, e.g., Abraham and Balakrishna (1999), Engle (2002), Tsai and Chan (2006), Lanne (2006) and Shephard and Sheppard (2010). The motivation to consider such models comes from the need to account for the nonnegative nature of certain time series. Examples from finance include variables such as prices, bid-ask spreads, trade volumes, trade durations and realized volatilities and bipower variations (Barndorff-Nielsen and Shephard, 2004). This note considers a nonlinear, nonnegative autoregressive model driven by nonnegative disturbances. More specifically, it considers robust estimation of the AR parameter β in the autoregression

$$y_t = \beta f(y_{t-1}, \ldots, y_{t-s}) + u_t, \tag{1}$$

Key words and phrases. robust estimation, linear programming estimator, strong convergence, nonlinear autoregression, heterogeneity, serially correlated disturbances, mixture model, heavy-tailed disturbances.

† City University of Hong Kong. Address correspondence to Daniel Preve, Department of Economics and Finance, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong; e-mail: pdapreve@cityu.edu.hk.
The author thanks Jun Yu, Marcelo C. Medeiros, Bent Nielsen and Peter C.B. Phillips for helpful comments on an early version of this note, and conference participants at the 6th Annual SoFiE conference (2013, Singapore) and seminar participants at Singapore Management University and National University of Singapore for their comments and suggestions. The author is responsible for any remaining errors. Partial research support from the Jan Wallander and Tom Hedelius Research Foundation (grant P 2006-0166:1) and the Sim Kee Boon Institute for Financial Economics, and the Institute's Centre for Financial Econometrics, at Singapore Management University is gratefully acknowledged.

with nonnegative (possibly) misspecified disturbances u_t. Potential distributions for u_t include lognormal, gamma, uniform, Weibull, inverse Gaussian, Pareto and mixtures of them. In some applications, robust estimation of the AR parameter is of interest in its own right. One example is point forecasting, as described in Preve et al. (2010). Another is seeking sources of model misspecification. In recognition of this fact, this note focuses explicitly on the robust estimation of β in (1). If the function f is known, a natural estimator for β given the sample $y_1, \ldots, y_n$ and the nonnegativity of the disturbances is

$$\hat{\beta}_n = \min \left\{ \frac{y_{s+1}}{f(y_s, \ldots, y_1)}, \ldots, \frac{y_n}{f(y_{n-1}, \ldots, y_{n-s})} \right\}. \tag{2}$$

This estimator has been used to estimate β in certain restricted first-order autoregressive, AR(1), models (e.g. Anděl, 1989b; Datta and McCormick, 1995; Nielsen and Shephard, 2003). An early reference of the autoregression in (1) is Bell and Smith (1986), who considers the linear AR(1) specification $f(y_{t-1}, \ldots, y_{t-s}) = y_{t-1}$ to model water pollution and the accompanying estimator in (2) for estimation.[1] The estimator in (2) can, under some additional conditions, be viewed as the solution to the linear programming problem of maximizing the objective function $g(\beta) = \beta$ subject to the $n - s$ linear constraints $y_t - \beta f(y_{t-1}, \ldots, y_{t-s}) \geq 0$ (cf. Feigin and Resnick, 1994). Because of this, we will refer to it as a LP-based estimator or LPE. As it happens, (2) is also the (on $y_1, \ldots, y_s$) conditional maximum likelihood estimator (MLE) for β when the disturbances are exponentially distributed (cf. Anděl, 1989a). What is interesting, however, is that $\hat{\beta}_n$ is a strongly consistent estimator of β for a wide range of disturbance distributions, thus the LPE is also a quasi-MLE.

In all of the above references the disturbances are assumed to be i.i.d.. To the authors knowledge, there has so far been no attempt to investigate the statistical properties of LP-based estimators in a *non* i.i.d. time series setting. This is the focus of the present study.

The remainder of this note is organized as follows. In Section 2 we give simple and quite general conditions under which the LPE is a strongly consistent estimator for the AR parameter, relaxing the assumption of i.i.d. disturbances significantly. In doing so, we also briefly review the literature on LP-based estimators in nonnegative autoregression. Section 3 reports the simulation results of a small scale Monte Carlo study investigating the finite-sample performance of the LPE. Mathematical proofs are collected in the Appendix. An extended Appendix available on request from the author contains some results mentioned in the text but omitted from the note to save space.

2. Theoretical Results

In economics, many time series models can be written in the form $y_t = \sum_{i=1}^{p} \beta_i f_i(y_{t-1}, \ldots, y_{t-s}) + u_t$. A recent example is Corsi's (2009) HAR model. In this section we focus on the particular case when $p = 1$ and the disturbances are nonnegative, m-dependent, heterogeneously distributed random variables.[2]

The nonlinear autoregressive model we consider is

$$\begin{cases} y_t = \beta f(y_{t-1}, \ldots, y_{t-s}) + u_t \\ u_t = \sigma_t \varepsilon_t, \quad t = s+1, s+2, \ldots \end{cases}$$

where the u_t are nonnegative random variables or disturbances. The model has two parts: The first part is a (potentially) nonlinear function of lagged values of y_t which, aside from the parameter β, is taken to be known. The second part, u_t, is taken to be unknown and potentially

[1] Bell and Smith (1986) refer to the LPE as a 'quick and dirty' nonparametric point estimator.

[2] A sequence $\varepsilon_1, \varepsilon_2, \ldots$ of random variables is said to be m-dependent if and only if ε_t and ε_{t+k} are pairwise independent for all $k > m$. In the special case when $m = 0$, m-dependence reduces to independence.

misspecified. In sum, the model combines a parametric part with a flexible, nonparametric part for the additive disturbance component. In some recent work, Preve et al. (2010) use similar types of autoregressive models to successfully forecast monthly S&P 500 realized volatility.

We now give simple and quite general conditions under which the LPE converges with probability one or almost surely (a.s.) to the AR parameter.

Condition 1. *The autoregression $\{y_t\}$ satisfies the stochastic difference equations*

$$y_t = \beta f(y_{t-1}, \ldots, y_{t-s}) + u_t, \quad t = s+1, s+2, \ldots$$

for some function $f : \mathbb{R}^s \to \mathbb{R}$ with AR parameter $\beta > 0$ and (a.s.) positive initial values $y_1, \ldots, y_s$. $\{u_t\}$ is a sequence of nonnegative, continuous random variables.

Condition 1 includes disturbance distributions supported on $[\eta, \infty)$, for any (unknown) nonnegative constant η, indicating that an intercept in the model is superflous. It also allows us to consider various mixture distributions that can account for data characteristics such as jumps.[3] The next condition concerns the mapping f, which allows for various lagged or seasonal model specifications.

Condition 2. *The mapping f is known (measurable and nonstochastic) and there exist constants $c > 0$ and $r \in \{1, \ldots, s\}$ such that $f(\mathbf{y}) = f(y_1, \ldots, y_r, \ldots, y_s) \geq cy_r$.*

The requirement that f dominates some hyperplane through the origin ensures the existence of a crude linear approximation of y_t in terms of lagged values of u_t at certain fixed instants of t. Conditions 1 and 2 combined ensure the nonnegativity of $\{y_t\}$. Condition 2 is met by elementary one-variable functions such as e^{y_s}, $\sinh y_s$ and any polynomial in y_s of degree higher than 0 with positive coefficients. Thus, in contrast to the setting of Anděl (1989b), we allow f to be non-monotonic.

Condition 3. *The disturbance at time t is given by*

$$u_t = \sigma_t \varepsilon_t, \quad t = s+1, s+2, \ldots$$

where $\{\sigma_t\}$ is a deterministic sequence of strictly positive reals of (possibly) unknown form, and $\{\varepsilon_t\}$ is a sequence of m-dependent, identically distributed, nonnegative continuous random variables. $m \in \mathbb{N}$ is finite and potentially unknown.

The σ_t are scaling constants which express the possible heteroskedasticity. The specification of the additive disturbance component can be motivated by the fact that it is common for the variance of a time series to change as its level changes. Since σ_t as well as the distribution and functional form of ε_t are unknown, the formulation is nonparametric. Condition 3 also allows for serially correlated disturbances. Such correlation arises if omitted variables included in u_t themselves are correlated over time.

Condition 4. *The time-varying 'volatility factor' σ_t satisfies*

$$\sigma \leq \frac{\sigma_{q(i)}}{\sigma_{q(i)+\triangle}} < \infty, \quad i = s+1, s+2, \ldots$$

for some $\sigma > 0$, where the integers $q(i)$, $\triangle$ are defined as $q(i) = (2i-2)(m+1)r + r + 1$, $\triangle = (m+1)r$.

[3]For instance, $u_t = (1 - b_t)u_{1t} + b_t u_{2t}$ where $\{b_t\}, \{u_{1t}\}$ and $\{u_{2t}\}$ are independent Bernoulli, lognormal and Pareto (potentially heavy-tailed) i.i.d. sequences, respectively.

Condition 4 ensures that σ_t (viewed as a function of time) does not increase, or decrease, too rapidly. The condition on the rate of change is very general and allows for various standard specifications, including abrupt breaks, smooth transitions, 'hidden' periodicities or combinations thereof, of the disturbance variance. Hence, the model is allowed to evolve over time.

The nonlinear, nonnegative autoregression implied by conditions 1–4 is flexible and nests several models in the related literature.[4] It is worth noting that, since $\hat{\beta}_n - \beta = R_n$ where $R_n = \min\,\{u_t/f(y_{t-1},\ldots,y_{t-s})\}_{t=s+1}^{n}$, the LPE is positively biased and stochastically nonincreasing in n under the conditions.

2.1. **Convergence.** Previous works focusing explicitly on the (stochastic) convergence of LP-based estimators in nonnegative autoregressions include Anděl (1989a), Anděl (1989b) and An (1992). LPEs are interesting as they can yield much more accurate estimates than traditional methods, such as conditional least squares (LS). See, e.g., Datta et al. (1998) and Nielsen and Shephard (2003). Like the LSE for β, the LPE is distribution free in the sense that its consistency does not rely on a particular distributional assumption for the disturbances. However, the LPE is sometimes superior to the LSE. For example, its rate of convergence can be faster than $\sqrt{n}$ even when $\beta < 1$.[5] For instance, in the linear AR(1) with exponential disturbances the (superconsistent) LPE converges to β at the rate of n. For another example, in contrast to the LSE, the consistency conditions of the LPE do not involve the existence of any higher order moments.

Here is our main result.

Proposition 1. *Under conditions 1–4, if* $P(c_1 < \varepsilon_t < c_2) < 1$ *for all* $0 < c_1 < c_2 < \infty$, *then* $\hat{\beta}_n \xrightarrow{a.s.} \beta$ *as* $n \to \infty$.

In other words, the LPE remains a consistent estimator for β if the i.i.d. disturbance assumption is significantly relaxed. The convergence is almost surely (and, hence, also in probability). Note that the additional condition of Proposition 1 is satisfied for any distribution with unbounded nonnegative support, and that the consistency conditions of the LPE do not involve the existence of any moments.[6] Hence, heavy-tailed disturbance distributions are also included.

2.2. **Distribution.** As aforementioned, the purpose of this note is not to derive the distribution of the LPE in our setting, but rather to highlight some of its robustness properties. Nevertheless, for completeness, we here mention some related distributional results. For the case with i.i.d. nonnegative disturbances several results are available: Davis and McCormick (1989) derive the limiting distribution of the LPE in a stationary AR(1) and Nielsen and Shephard (2003) derive the exact (finite-sample) distribution of the LPE in a AR(1) with exponential disturbances. Feigin and Resnick (1994) derive limiting distributions of LPEs in a stationary AR(p). Datta et al. (1998) establish the limiting distribution of a LPE in an extended nonlinear autoregression. The limited success of LPEs in applied work can be partially explained by the fact that their asymptotic distributions depend on the (in most cases) unknown distribution of the disturbances. To overcome this problem, Datta and McCormick (1995) and Feigin and Resnick (1997) consider bootstrap inference for linear autoregressions via LPEs. Some robustness properties and exact

[4] For example, Bell and Smith's model is obtained by choosing $f(\mathbf{y}) = y_1$, $\sigma_t = 1$ for all t, and $m = 0$.

[5] This occurs, under some additional conditions, when the exponent of regular variation of the disturbance distribution at 0 or ∞ is less than 2 (Davis and McCormick, 1989; Feigin and Resnick, 1992). The rate of convergence for the LSE is faster than $\sqrt{n}$ only when $\beta \geq 1$ (Phillips, 1987).

[6] As an extreme example, consider estimating β in the linear model $y_t = \beta y_{t-1} + u_t$ with independent, nonnegative stable disturbances $u_t \sim \mathcal{S}(a, b, c, d; 1)$, where the index of stability $a < 1$, the skewness parameter $b = 1$ and the location parameter $d \geq 0$ (cf. Lemma 1.10 in Nolan, 2012). In this case y_t also follows a stable distribution with index of stability a and, hence, no finite first moment for a suitable choice of y_1.

distributional results of the LPE in a *cross-sectional* setting were recently derived by Preve and Medeiros (2011).

3. Simulation Study

In this section we report simulation results concerning the estimation of β in the autoregression $y_t = \beta y_{t-1} + u_t$. We consider four different disturbance specifications, and the LPE and LSE as estimators for β. The LSE is used as a simple benchmark estimator as it is well known to be inconsistent for the AR parameter when the disturbances are serially correlated. We report the empirical bias and mean squared error (MSE) based on 100 000 simulated samples for different sample sizes n.

First we consider a data generating process (DGP) with linear disturbance specification

$$u_t = \epsilon_t + \sum_{i=1}^{m} \psi_i \epsilon_{t-i}.$$

The simulation results are shown in panel A of Table 1. The second specification is nonlinear and given by

$$u_t = \epsilon_t + \sum_{i=1}^{m} \psi_i \epsilon_t \epsilon_{t-i}.$$

The corresponding simulation results are shown in panel B of Table 1. For the last two specifications we consider DGPs with σ_t time-varying and follow the designs used in Section 5 of Phillips and Xu (2005). The first specification corresponds to the case of a single abrupt change of disturbance variance via

$$\sigma_t = \sqrt{\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbf{1}_{\{t \geq n\tau\}}},$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function and $\tau \in (0, 1)$. For the second specification heteroskedasticity is delivered by a smooth transition function of time, via

$$\sigma_t = \sqrt{\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\tfrac{t}{n}}.$$

The simulation results are shown in panels C and D of Table 1. For all four specifications, it is readily verified that the conditions of Proposition 1 are satisfied. Hence, the LPE is strongly consistent for β.

4. Conclusions

The focus in this note is on estimating the AR parameter in an autoregression driven by nonnegative disturbances using a LP-based estimator. In the previous literature the disturbances are assumed to be i.i.d.. Sometimes, this assumption may be considered too restrictive and one would like to relax it. In this note, we relax the i.i.d. assumption significantly by allowing for quite general kinds of dependencies between the disturbances and show that the LPE remains strongly consistent. In doing so, we also briefly review the literature on LP-based estimators in nonnegative autoregression. Because of its robustness properties, the LPE can be used to seek sources of misspecification in the disturbances of the initially specified model and to isolate certain trend, seasonal or cyclical components. Our simulation results indicate that the LPE can have very reasonable finite-sample properties.

TABLE 1. Simulation results: Each table entry, based on 100 000 simulated samples, reports the empirical bias/MSE of the LPE/LSE for β in $y_t = \beta y_{t-1} + u_t$ when the DGP is *Panel A)* $y_t = 0.8y_{t-1} + \epsilon_t + 0.75\epsilon_{t-1} + 0.25\epsilon_{t-2}$, *Panel B)* $y_t = 0.9y_{t-1} + \epsilon_t + 0.9\epsilon_t\epsilon_{t-1}$, *Panel C)* $y_t = 0.5y_{t-1} + \epsilon_t\sqrt{1 + (0.04 - 1)\mathbf{1}_{\{t \geq n/2\}}}$, *Panel D)* $y_t = 0.5y_{t-1} + \epsilon_t\sqrt{1 + 24t/n}$. For all four DGPs the ϵ_t are independent standard exponentially distributed.

	Panel A: Linear disturbances				*Panel B: Nonlinear disturbances*			
	LPE		LSE		LPE		LSE	
n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.036	0.002	0.085	0.010	0.002	0.000	-0.015	0.004
100	0.028	0.001	0.102	0.012	0.001	0.000	0.013	0.002
200	0.021	0.001	0.111	0.013	0.000	0.000	0.026	0.001

	Panel C: Abrupt change				*Panel D: Smooth transition*			
	LPE		LSE		LPE		LSE	
n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.015	0.000	0.181	0.045	0.023	0.001	0.021	0.030
50	0.009	0.000	0.214	0.051	0.011	0.000	0.060	0.016
100	0.005	0.000	0.227	0.054	0.005	0.000	0.082	0.013

References

Abraham, B. & N. Balakrishna (1999) Inverse Gaussian autoregressive models. *Journal of Time Series Analysis* 6, 605–618.

An, H.Z. (1992) Non-negative autoregressive models. *Journal of Time Series Analysis* 13, 283–295.

Anděl, J. (1989a) Non-negative autoregressive processes. *Journal of Time Series Analysis* 10, 1–12.

Anděl, J. (1989b) Nonlinear nonnegative AR(1) processes. *Communications in Statistics - Theory and Methods* 18, 4029–4037.

Barndorff-Nielsen, O.E. & N. Shephard (2004) Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics* 2, 1–37.

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

Corsi, F. (2009) A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.

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

Datta, S., G. Mathew, & W.P. McCormick (1998) Nonlinear autoregression with positive innovations. *Australian & New Zealand Journal of Statistics* 40, 229–239.

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

Engle, R. (2002) New frontiers for ARCH models. *Journal of Applied Econometrics* 17, 425–446.

Feigin, P.D. & S.I. Resnick (1992) Estimation for autoregressive processes with positive innovations. *Stochastic Models* 8, 479–498.

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

Feigin, P.D. & S.I. Resnick (1997) Linear programming estimators and bootstrapping for heavy tailed phenomena. *Advances in Applied Probability* 29, 759–805.

Ferguson, T. (1996) *A Course in Large Sample Theory*. Chapman & Hall, London.

Lanne, M. (2006) A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics* 4, 594–616.

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

Nolan, J.P. (2013) *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

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

Phillips, P.C.B. & K.L. Xu (2005) Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis* 27, 289–308.

Preve, D., A. Eriksson, & J. Yu (2010) Forecasting realized volatility using a nonnegative semiparametric model. Working Paper.

Preve, D. & M.C. Medeiros (2011) Linear programming-based estimators in simple linear regression. *Journal of Econometrics* 165, 128–136.

Shephard, N. & K. Sheppard (2010) Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 25, 197–231.

Tsai, H. & K.S. Chan (2006) A note on non-negative ARMA processes. *Journal of Time Series Analysis* 28, 350–360.

APPENDIX

The following lemmas are applied in the proof of Proposition 1.

Lemma 1. *Under conditions 1-4,* $R_n \xrightarrow{p} 0 \Rightarrow \hat{\beta}_n \xrightarrow{a.s.} \beta$.

Proof. We will use that $\hat{\beta}_n$ converges almost surely to β if and only if for every $\epsilon > 0$, $\lim_{n\to\infty} P(|\hat{\beta}_k - \beta| < \epsilon;\ k \geq n) = 1$ (Lemma 1 in Ferguson, 1996). Let $\epsilon > 0$ be arbitrary. Then,

$$P(|\hat{\beta}_k - \beta| < \epsilon;\ k \geq n) = P(|R_k| < \epsilon;\ k \geq n) = P(|R_n| < \epsilon) \to 1 \text{ as } n \to \infty.$$

The last equality follows since $\{R_k\}$ is stochastically decreasing, and the limit since $R_n \xrightarrow{p} 0$ by assumption. □

Lemma 2. *Under conditions 1-4,*

$$y_{lr+1} \geq (\beta c)^l y_1 + \sum_{j=0}^{l-1} (\beta c)^j \sigma_{(l-j)r+1} \varepsilon_{(l-j)r+1},$$

for $l = 1, \ldots, \lfloor \frac{n-1}{r} \rfloor$ *(a.s.), where* $\lfloor\cdot\rfloor$ *is the integer part function.*

Proof. We proceed with a proof by induction. Without loss of generality, suppose $y_i = 0$ for $i = (r+1-s), \ldots, 0$ if $r < s$. Since cy_r is dominated by $f(\mathbf{y})$ for $y_r > 0$, the assertion is true for $l = 1$. Suppose it is true for some positive integer k. Then, for $k+1$

$$\begin{aligned} y_{(k+1)r+1} &= \beta f(y_{(k+1)r}, \ldots, y_{kr+1}, \ldots, y_{(k+1)r+1-s}) + \sigma_{(k+1)r+1}\varepsilon_{(k+1)r+1} \\ &\geq \beta c y_{kr+1} + \sigma_{(k+1)r+1}\varepsilon_{(k+1)r+1} \geq (\beta c)^{k+1} y_1 + \sum_{j=0}^{k} (\beta c)^j \sigma_{(k-j+1)r+1}\varepsilon_{(k-j+1)r+1}, \end{aligned}$$

where the last inequality follows by the induction assumption. □

Lemma 3. *Let* v *and* w *be two nonnegative i.i.d. continuous random variables. Then, the following two statements are equivalent:*

(*i*) $P(v > \epsilon w) = 1$ *for some* $\epsilon > 0$,
(*ii*) *there exist* c_1 *and* c_2, $0 < c_1 < c_2 < \infty$, *such that* $P(c_1 < v < c_2) = 1$.

Proof. See p. 2291 in Bell and Smith (1986). □

Proof of Proposition 1. By the proof of Lemma 1, it is sufficient to show that $R_n \xrightarrow{p} 0$. Suppose that $n \geq q(s) + 3\triangle$ and let $\varepsilon > 0$ be given. Then

$$\begin{aligned} P(R_n > \varepsilon) &= P\big(u_t > \varepsilon f(y_{t-1}, \ldots, y_{t-s});\ t = s+1, \ldots, n\big) \\ &\leq P\big(u_{q(i)+\triangle} > \varepsilon f(y_{q(i)+\triangle-1}, \ldots, y_{q(i)+\triangle-s});\ i = s+1, \ldots, N\big), \end{aligned}$$

where

$$N(n) = \left\lfloor \frac{1}{2}\left(\frac{n-r-1}{\triangle} + 1\right)\right\rfloor.$$

Apparently, $N \in [s+1, n)$ and tends to infinity as $n \to \infty$. Furthermore, by Condition 2 and Lemma 2,

$$\begin{aligned} f(y_{lr+r}, \ldots, y_{lr+1}, \ldots, y_{lr+r+1-s}) &\geq c y_{lr+1} \\ &\geq c^{l+1}\beta^l y_1 + \sum_{j=0}^{l-1} c^{j+1}\beta^j \sigma_{(l-j)r+1}\varepsilon_{(l-j)r+1} \\ &\geq c^{j+1}\beta^j \sigma_{(l-j)r+1}\varepsilon_{(l-j)r+1}, \end{aligned}$$

for each $j \in \{0, \ldots, l-1\}$. Hence, for $l(i) = (2i-1)(m+1)$ and $j = m$ it is readily verified that

$$P(R_n > \varepsilon) \leq P(\sigma_{q(i)+\triangle}\varepsilon_{q(i)+\triangle} > \varepsilon c^{m+1}\beta^m \sigma_{q(i)}\varepsilon_{q(i)};\, i = s+1, \ldots, N).$$

Since the sequence $\varepsilon_{s+1}, \ldots, \varepsilon_n$ of disturbances is m-dependent, ε_t and ε_{t+k} are pairwise independent for all $k > m$. Let $w_i = \varepsilon_{q(i)+\triangle}/\varepsilon_{q(i)}$. Then $w_{s+1}, \ldots, w_N$ is a sequence of i.i.d. random variables, for which the numerator and denominator of each w_i are pairwise independent, and hence

$$\begin{aligned} P(R_n > \varepsilon) &\leq P\big(w_{s+1} > \tfrac{\sigma_{q(s+1)}}{\sigma_{q(s+1)+\triangle}}\varepsilon c^{m+1}\beta^m\big) \times \cdots \times \\ P\big(w_N &> \tfrac{\sigma_{q(N)}}{\sigma_{q(N)+\triangle}}\varepsilon c^{m+1}\beta^m\big) \leq P(\varepsilon_{q(s+1)+\triangle} > \epsilon\varepsilon_{q(s+1)})^{N-s} \end{aligned}$$

where $\epsilon = \sigma\varepsilon c^{m+1}\beta^m$. In view of Lemma 3 and the limiting behavior of $N(n)$ this implies that $P(|R_n| > \varepsilon) \to 0$ as $n \to \infty$. Since $\varepsilon > 0$ was arbitrary, R_n converges in probability to zero. □