LINEAR PROGRAMMING-BASED ESTIMATORS IN NONNEGATIVE AUTOREGRESSION

DANIEL P. A. PREVE†
JANUARY 30, 2015

Abstract. This note studies robust estimation of the autoregressive (AR) parameter in a nonlinear, nonnegative AR model driven by nonnegative errors. 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 test for, and seek sources of, misspecification when a pure autoregression cannot satisfactorily describe the data generating process, 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 serially dependent, non-identically distributed or otherwise misspecified errors 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 an extensive simulation study covering a wide range of model misspecifications. A small scale empirical study, employing a volatility proxy to model and forecast latent daily return volatility of three major stock market indexes, illustrates the potential usefulness of the LPE.

JEL classification. C13, C14, C22, C51, C58.

Key words and phrases. Robust estimation, linear programming estimator, strong convergence, nonlinear nonnegative autoregression, dependent non-identically distributed errors, heavy-tailed errors.

† 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, Rickard Sandberg, Kevin Sheppard 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), the 2014 Asian Meeting of the Econometric Society (Taipei), the 1st Conference on Recent Developments in Financial Econometrics and Applications (2014, Geelong), 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. The work described in this note was substantially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (project no. CityU198813). 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 also gratefully acknowledged.
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 behaviour of a broad class of nonnegative time series.

For nonlinear time series models it is common to assume that the errors 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 absolute or squared returns, bid-ask spreads, trade volumes, trade durations, and standard volatility proxies such as realized variance, realized bipower variation (Barndorff-Nielsen and Shephard, 2004) or realized kernel (Barndorff-Nielsen et al., 2008). This note considers a nonlinear, nonnegative autoregressive model driven by nonnegative errors. More specifically, it considers robust estimation of the AR parameter $\beta$ in the autoregression

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

with nonnegative (possibly) misspecified errors $u_t$. Potential distributions for $u_t$ include log-normal, 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. (2015). Another is seeking sources of model misspecification. In recognition of this fact, this note focuses explicitly on the robust estimation of $\beta$ in (1). If the function $f$ is known, a natural estimator for $\beta$ given the sample $y_1, \ldots, y_n$ of size $n$ and the nonnegativity of the errors is

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

This estimator has been used to estimate $\beta$ in certain restricted first-order autoregressive, AR(1), models (e.g. Anděl, 1986b; 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. 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 $y_1, \ldots, y_n$) conditional maximum likelihood estimator (MLE) for $\beta$ when the errors are exponentially distributed (cf. Anděl, 1989a). What is interesting, however, is that $\hat{\beta}_n$ is a strongly consistent estimator of $\beta$ for a wide range of error distributions, thus the LPE is also a quasi-MLE (QMLE).

In all of the above references the errors 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 $\hat{\beta}_n$.

---

1Another example is temperature, which can be used for pricing weather derivatives (e.g. Campbell and Diebold, 2005; Alexandris and Zapranis, 2013).

2Bell and Smith (1986) refer to the LPE as a ‘quick and dirty’ nonparametric point estimator.
in a non i.i.d. time series setting. This is the focus of the present note. In that sense, the note can be viewed as a companion note to Preve and Medeiros (2011) in which the authors establish statistical properties of a LPE in a non i.i.d. cross-sectional setting. Estimation of time series models with dependent, non-identically distributed errors is important for two reasons: First, the assumption of independent, identically distributed errors is a serious restriction. In practice, possible causes for non i.i.d. or misspecified errors include omitted variables, measurement errors and regime changes. Second, traditional estimators, like the least squares estimator, may be inconsistent when the errors are misspecified. In some applications the errors may also be heavy-tailed. The main theoretical contribution of the note is to provide conditions under which the LPE in (2) is consistent for the unknown AR parameter in (1) when the errors are serially dependent, non-identically distributed and heavy-tailed.

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. errors significantly. In doing so, we also briefly review the literature on LP-based estimators in nonnegative autoregression. Section 3 reports the simulation results of an extensive Monte Carlo study investigating the finite-sample performance of the LPE and at the same time illustrating its robustness to various types of model misspecification. Section 4 reports the results of a small scale empirical study, and Section 5 concludes. Mathematical proofs are collected in the Appendix. An extended Appendix (EA) 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 finance, 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 errors are nonnegative, serially correlated, possibly heterogeneously distributed and heavy-tailed random variables.

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

Assumption 1. The autoregression $\{y_t\}$ is given by

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

for some function $f : \mathbb{R}^s \to \mathbb{R}$, AR parameter $\beta > 0$, and (a.s.) positive initial values $y_1, \ldots, y_s$. The errors $u_t$ driving the process are nonnegative random variables.

Assumption 1 includes error distributions supported on $[\eta, \infty)$, for any unknown nonnegative constant $\eta$, indicating that an intercept in the process is superfluous (Section 3.1.2). It also allows us to consider various mixture distributions that can account for data characteristics such as jumps (Section 3.3.2). The next assumption concerns the potentially multi-variable function $f$, which allows for various lagged or seasonal specifications (Section 3.1.3).

Assumption 2. The function $f : \mathbb{R}^s \to \mathbb{R}$ is known (measurable and nonstochastic), and there exist constants $c > 0$ and $r \in \{1, \ldots, s\}$ such that $f(x) = f(x_1, \ldots, x_r, \ldots, x_s) \geq cx_r$.

---

3The HAR model of Corsi can be written as $y_t = \sum_{i=1}^{3} \beta_i f_i(y_{t-1}, \ldots, y_{t-22}) + u_t$, where $f_1(y_{t-1}, \ldots, y_{t-22}) = y_{t-1}$, $f_2(y_{t-1}, \ldots, y_{t-22}) = y_{t-2} + \cdots + y_{t-5}$, $f_3(y_{t-1}, \ldots, y_{t-22}) = y_{t-6} + \cdots + y_{t-22}$ and $y_t$ is the realized volatility over day $t$. Here $p = 3$ and $s = 22$. 


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 \( t \). Assumptions 1 and 2 combined ensure the nonnegativity of \( \{y_t\} \), indicating that the process may be used to model durations, volatility proxies, and so on. Assumption 2 is, for instance, met by elementary one-variable functions such as \( e^{\psi x} \), \( \sinh x \), and any polynomial in \( x \) of degree higher than 0 with positive coefficients. Thus, in contrast to Anděl (1989b), we allow \( f \) to be non-monotonic.

**Assumption 3.** The error at time \( t \) is given by

\[
u_t = \mu_t + \sigma_t \varepsilon_t, \quad t = s + 1, s + 2, \ldots \]

where \( \{\mu_t\} \) and \( \{\sigma_t\} \) are discrete-time processes, and \( \{\varepsilon_t\} \) is a sequence of \( m \)-dependent, identically distributed, nonnegative continuous random variables. The order, \( m \), of the dependence is finite.

Assumption 3 allows for different kinds of \( m \)-dependent error specifications, with \( m \in \mathbb{N} \) potentially unknown.\(^4\) For example, finite-order moving average (MA) specifications (Section 3.2.2). The \( \sigma_t \) of (possibly) unknown form are scaling variates, which express the possible heteroskedasticity. The specification of the additive error 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 the forms and distributions of \( \mu_t, \sigma_t \) and \( \varepsilon_t \) are taken to be unknown, the formulation is nonparametric. Assumption 3 also allows for more general forms of serially correlated errors (Section 3.2). Such correlation arises if omitted variables included in \( u_t \) themselves are correlated over time, or if \( y_t \) is measured with error (Section 4).

**Assumption 4.** There exist constants \( 0 \leq \underline{\sigma} < \infty \) and \( 0 < \underline{\sigma} \leq \overline{\sigma} < \infty \) such that \( P(0 \leq \mu_t \leq \overline{\mu}) = 1 \) and \( P(\underline{\sigma} \leq \sigma_t \leq \overline{\sigma}) = 1 \) for all \( t \).

Assumption 4 ensures that \( \{\mu_t\} \) and \( \{\sigma_t\} \) in Assumption 3 are bounded in probability. The bounds (and the forms) for \( \mu_t \) and \( \sigma_t \) are not required to be known. The assumption is quite general and allows for various standard specifications, including structural breaks, Markov switching, thresholds, smooth transitions, ‘hidden’ periodicities or combinations thereof, of the error mean and variance (Section 3.3).\(^5\)

### 2.2. Finite-Sample Theory.

The nonlinear, nonnegative autoregression implied by assumptions 1–4 is flexible and nests several specifications in the related literature.\(^6\) It is worth noting that, since \( \hat{\beta}_n - \beta = R_n \) where \( R_n = \min \{u_{s+1}/f(y_{s}, \ldots, y_1), \ldots, u_n/f(y_{n-1}, \ldots, y_{n-s})\} \), the LPE is positively biased and stochastically decreasing in \( n \) under the assumptions. Moreover, it is not difficult to show that the LP residuals \( \hat{u}_t = y_t - \hat{\beta}_n f(y_{t-1}, \ldots, y_{t-s}) \), by construction, are nonnegative.

#### 2.3. Asymptotic Theory.

##### 2.3.1. Convergence.

Previous works focusing explicitly on the (stochastic) convergence of LP-based estimators in nonnegative autoregressions include Anděl (1989a), Anděl (1989b) and An (1992). These 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 \( \beta \), the LPE is distribution-free in the sense that

---

4 A sequence \( \varepsilon_1, \varepsilon_2, \ldots \) of random variables is said to be \( m \)-dependent if and only if \( \varepsilon_t \) and \( \varepsilon_{t+k} \) are pairwise independent for all \( k > m \). In the special case when \( m = 0 \), \( m \)-dependence reduces to independence.

5 It is important to note that \( \mu_t \) and \( \sigma_t \) are allowed to be degenerate random variables (i.e. deterministic).

6 For example, Bell and Smith’s specification is obtained by choosing \( f(x) = x_1 \) or, equivalently, \( f(y_{t-1}, \ldots, y_{t-s}) = y_{t-1}, \mu_t = 0 \) and \( \sigma_t = 1 \) for all \( t \), and \( m = 0 \). Note that in this case the errors are i.i.d..
its consistency does not rely on a particular distributional assumption for the errors. However, the LPE is sometimes superior to the LSE. For example, its rate of convergence can be faster than $\sqrt{n}$ even when $\beta < 1$.\footnote{This occurs, under some additional conditions, when the exponent of regular variation of the error distribution at 0 or $\infty$ 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).} For instance, in the linear AR(1) with exponential errors the (superconsistent) LPE converges to $\beta$ 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.

The following theorem is the main theoretical contribution of the note. It provides conditions under which the LPE is strongly consistent for the unknown AR parameter.

**Theorem 1.** Suppose that assumptions 1–4 hold. Then the LPE or QMLE in (2) is strongly consistent for $\beta$ in (1), i.e. $\hat{\beta}_n$ converges to $\beta$ a.s. as $n$ tends to infinity, if either (i) $P(c_1 < \epsilon_t < c_2) < 1$ for all $0 < c_1 < c_2 < \infty$ and $\mu_1 = 0$ for all $t$, or (ii) $P(\epsilon_t < c_3) < 1$ for all $0 < c_3 < \infty$.

In other words, the LPE remains a consistent estimator for $\beta$ if the i.i.d. error assumption is significantly relaxed. The convergence is almost surely (and, hence, also in probability). Note that the additional condition of Theorem 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.\footnote{As an extreme example, consider estimating $0 < \beta < 1$ in the linear specification $y_t = \beta y_{t-1} + u_t$, $u_t = \mu_1 + \sigma_1 \epsilon_t$, considered in sections 1–2. The purpose of the simulations is to see how the LPE and a benchmark estimator perform under controlled circumstances when the data generating process is known.} Hence, heavy-tailed error distributions are also included (Section 3.1.2).

### 2.3.2. Distribution

As aforementioned, the purpose of this note is not to derive the distribution of the LPE in our (quite general) 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 errors 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 errors. 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 errors. 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 distributional results of the LPE in a *cross-sectional* setting were recently derived by Preve and Medeiros (2011).

### 3. Simulation Results

In this section we report simulation results concerning the estimation of the AR parameter $\beta$ in the nonnegative autoregression $y_t = \beta f(y_{t-1}, \ldots, y_{t-s}) + u_t$, $u_t = \mu_1 + \sigma_1 \epsilon_t$, considered in sections 1–2. The purpose of the simulations is to see how the LPE and a benchmark estimator perform under controlled circumstances when the data generating process is known.
For ease of exposition, we let \( f(y_{t-1}, \ldots, y_{t-s}) = y_{t-s} \) and \( s = 1, 4 \).\(^9\) Thus, in the simulations the data generating process (DGP) is

\[
y_t = \beta y_{t-s} + u_t, \quad \text{with} \quad u_t = \mu_t + \sigma_t \varepsilon_t,
\]

and the LPE is

\[
\hat{\beta}_{LP} = \min \left\{ \frac{y_{s+1}}{y_1}, \ldots, \frac{y_n}{y_{n-s}} \right\} = \beta + \min \left\{ \frac{u_{s+1}}{y_1}, \ldots, \frac{u_n}{y_{n-s}} \right\}.
\]

In this case, whenever the errors \( u_t \) are believed to be serially uncorrelated and to satisfy the usual moment conditions, a natural benchmark for the LPE of \( \beta \) is the corresponding ordinary least squares estimator, \( \hat{\beta}_{LS} \), which also is a distribution-free estimator. If \( \{y_t\} \) is generated by (3) and \( \{u_t\} \) is a sequence of random variables with common finite mean, the LSE for \( \beta \) given a sample \( y_1, \ldots, y_n \) of size \( n > s \) is

\[
\hat{\beta}_{LS} = \frac{\sum_{t=s+1}^{n}(y_t - \bar{y}_+)(y_{t-s} - \bar{y}_-)}{\sum_{t=s+1}^{n}(y_t - \bar{y}_+)^2} = \beta + \frac{\sum_{t=s+1}^{n}(u_t - \bar{u}_+)(y_{t-s} - \bar{y}_-)}{\sum_{t=s+1}^{n}(y_t - \bar{y}_+)^2},
\]

where

\[
\bar{y}_+ = \frac{1}{n-s} \sum_{t=s+1}^{n} y_t, \quad \text{and} \quad \bar{y}_- = \frac{1}{n-s} \sum_{t=s+1}^{n} y_{t-s}.
\]

By the second equality in (5) it is clear that the LSE, like the LPE, can be decomposed into two parts: the true (unknown) value \( \beta \) and a stochastic remainder term, indicating that \( \hat{\beta}_{LS} \) may be asymptotically biased. For instance, if the errors \( u_t \) are serially correlated.

Table 1 shows simulation results for various specifications of \( \mu_t, \sigma_t \) and \( \varepsilon_t \) in (3). The assumptions of Theorem 1 are satisfied for all of these specifications, hence, the LPE (but not necessarily the LSE) is consistent for \( \beta \). Our simulations try to answer two questions: First, how well does the LPE perform when the estimated model is misspecified? Second, how well does a traditional estimator, the LSE, perform by comparison? We report the empirical bias and mean squared error (MSE) of the LPE and LSE based on 1 000 000 simulated samples for different sample sizes \( n \). Each table entry is rounded to three decimal places.\(^10\)

3.1. I.I.D. Errors. Panels A–H of Table 1 report simulation results when the errors \( u_t \) in (3) are independent, identically distributed. In all eight experiments, the bias and MSE of the LPE is quite reasonable. The LSE also performs reasonably well, but often has a larger bias and a much larger MSE.

3.1.1. Light-Tailed Errors. The LPE can be comparatively accurate also when there is no model misspecification (i.e. when model and DGP coincide). To illustrate this, we first consider three different light-tailed distributions for the errors. Panels A–C report simulation results when \( y_t = 0.5y_{t-1} + u_t \) and the \( u_t \) are i.i.d. uniform, exponential, and Weibull.\(^11\) In all experiments the estimated model is an AR(1). In the first experiment (Panel A) the error at time \( t \) has a standard uniform distribution. Here the distribution of \( y_t \) conditional on \( y_{t-1} \) is also uniform.

\(^9\)Sample paths of some of the processes considered in this study and additional supporting simulation results can be found in Section 2 of the EA.

\(^10\)MATLAB code for the Monte Carlo experiments of this section is available through the authors webpage at http://www.researchgate.net/profile/Daniel_Preve.

\(^11\)Nonnegative first-order autoregressions with uniformly distributed errors have been considered by Nouali and Fellag (2005) and Bell and Smith (1986), and with exponential errors by Nielsen and Shephard (2003), among others. Exponential and Weibull errors are popular in the autoregressive conditional duration (ACD) literature initiated by Engle and Russell (1998).
In the second experiment (Panel B) the errors are standard exponential, and in the third (Panel C) Weibull distributed with scale parameter 1 and shape parameter 2.

3.1.2. Heavy-Tailed Errors. Panels D–E report simulation results when \( y_t = 0.5y_{t-1} + u_t \) and the \( u_t \) are i.i.d. Pareto and Lévy, respectively.\(^\text{12}\) In both experiments the estimated model is an AR(1). In the first experiment (Panel D) the error is Pareto distributed with scale parameter 1 and shape parameter 1.25 and, hence, with support \([1, \infty)\). The Pareto distribution is one of the simplest distributions with heavy-tails. Here the errors have finite mean, 5, but infinite variance. In the second experiment (Panel E) the error is Lévy distributed with location parameter 0 and scale parameter 1.\(^\text{13}\) The Lévy distribution is a member of the class of stable distributions, that allow for asymmetry and heavy-tails. Here the errors have infinite mean (but finite median) and variance. AR(MA) processes with infinite variance have been used by Fama (1965) and others to model stock market prices. See also Ling (2005), Andrews et al. (2009), and Andrews and Davis (2013).

3.1.3. Seasonal Autoregression. Panels F–H report simulation results when \( y_t = \beta y_{t-4} + u_t \), the AR parameter \( \beta = 0.25, 0.5, 0.75 \), the errors \( u_t \) are i.i.d. Weibull with scale parameter 1 and shape parameter 2, and the (correctly specified) estimated model is a SARMA(0,0)×(1,0)\(^\text{4}\).\(^\text{14}\) These three experiments illustrate that the bias and MSE of the LPE for a fixed \( n \), viewed as a function of \( \beta \), is stochastically decreasing in the AR parameter. It can be shown that this property holds under fairly general conditions on \( f \) in Assumption 2.\(^\text{15}\)

3.2. Serially Correlated Errors. Panels I–N of Table 1 report simulation results when \( y_t = \beta y_{t-4} + u_t \), and the \( u_t \) are serially correlated. In all six experiments the estimated model is an AR(1). To investigate the sensitivity of the LPE to serially correlated errors, we consider four different specifications for \( u_t \): a multiplicative specification belonging to the MEM family of Engle (2002), a MA specification, a nonlinear specification, and an omitted variables specification. In all experiments the bias and MSE of the LPE vanishes rather quickly. In contrast, the bias of the LSE does not vanish and its MSE is quite substantial even for large samples.\(^\text{16}\)

3.2.1. MEM Errors. First we consider the multiplicative error specification

\[
u_t = \sigma_t \varepsilon_t
\]

\[
\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots
\]

with i.i.d. \( \varepsilon_t \), that is, a MEM\((p,q)\). This specification has the same structure as the ACD model of Engle and Russell (1998) for trade durations. Panel I reports simulation results for the case \( p = q = 1 \). The DGP is

\[
y_t = 0.5y_{t-1} + \sigma_t \varepsilon_t, \quad \sigma_t = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \beta_1 \sigma_{t-1},
\]

with independent identically beta distributed \( \varepsilon_t \) (beta distributed with both shape parameters equal to 2 and, hence, with about 0.5 symmetric common density), and \( \alpha_1 = 0.2, \beta_1 = 0.75,\)
α_0 = 1 - α_1 - β_1, which ensures that Assumption 4 is satisfied. For these values of α_1, β_1 and α_0 the autocorrelation function of \{u_t\} decays slowly (cf. Bauwens and Giot, 2000, p. 124).

3.2.2. MA Errors. Next we consider the linear m-dependent error specification

\[ u_t = \epsilon_t + \sum_{i=1}^{q} \psi_i \epsilon_{t-i}, \quad t = 0, \pm 1, \pm 2, \ldots \]

with i.i.d. \( \epsilon_t \), that is, a MA(q). Here \( \mu_t = 0, \sigma_t = 1 \) and \( m = q \). Panels J–K report simulation results for this case, which may be considered as a basic omitted variables specification, with \( q = 1, 2 \). The DGPs for panels J and K are

\[ y_t = 0.75y_{t-1} + \epsilon_t + 0.75\epsilon_{t-1}, \quad \text{and} \quad y_t = 0.75y_{t-1} + \epsilon_t + 0.75\epsilon_{t-1} + 0.5\epsilon_{t-2}, \]

respectively, where the \( \epsilon_t \) are i.i.d. inverse Gaussian with mean and variance both equal to 1. The inverse Gaussian distribution (Seshadri, 1993) has previously been considered by Abraham and Balakrishna (1999) for the error term in a nonnegative first-order autoregression.

3.2.3. Nonlinear Specification. The third specification we consider is a nonlinear m-dependent error specification

\[ u_t = \epsilon_t = \epsilon_t + \sum_{i=1}^{m} \psi_i \epsilon_{t-i}, \]

with i.i.d. \( \epsilon_t \). Panels L–M report simulation results for this case, with \( m = 1, 2 \). The DGPs for panels L and M are

\[ y_t = 0.5y_{t-1} + \epsilon_t + 0.75\epsilon_{t-1}, \quad \text{and} \quad y_t = 0.5y_{t-1} + \epsilon_t + 0.75\epsilon_{t-1} + 0.5\epsilon_{t-2}, \]

respectively, where the \( \epsilon_t \) are i.i.d. inverse Gaussian with mean and variance both equal to 1.

3.2.4. Omitted Variables. Last we consider the linear error specification

\[ u_t = \mu_t + \epsilon_t \]

\[ \mu_t = \sum_{i=1}^{p} \alpha_i \mu_{t-i} + \epsilon_t, \quad t = 0, \pm 1, \pm 2, \ldots \]

with i.i.d. \( \epsilon_t \). Here the \( p \)-th order AR specification for \( \mu_t \) may be considered to represent one or more omitted variables, \( \sigma_t = 1 \), and \( m = 0 \). Panel N reports simulation results for the case \( p = 1 \). The DGP is

\[ y_t = 0.75y_{t-1} + \mu_t + \epsilon_t, \]

\[ \mu_t = 0.25\mu_{t-1} + \epsilon_t, \]

with standard exponential \( \epsilon_t \) (mean and variance both equal to 1) and i.i.d. on \((0,25)\) uniform \( \epsilon_t \), mutually independent of the \( \epsilon_t \).

3.3. Structural Breaks. Finally we investigate the sensitivity of the LPE to an unknown number of unknown breaks in the mean. This is of interest as such structural breaks are well known to be able to reproduce the slow decay frequently observed in the sample autocorrelations of financial variables such as volatility proxies and absolute stock returns. The simulation results are reported in Panels O–P of Table 1. In both experiments the estimated model is an AR(1). Once again the bias and MSE of the LPE vanishes rather quickly, whereas the bias of the LSE does not vanish and its MSE is quite substantial even for large samples.

---

\(^{17}\)For example, finite sums of finite-order MA processes driven by i.i.d. disturbances are m-dependent.
3.3.1. Random Breakdates. An autoregression with \( b \geq 1 \) structural breaks in the sample and breakdates \( n_1 < \cdots < n_b \) can be specified using

\[
\mu_t = \sum_{i=1}^{b-1} \alpha_i 1_{\{n_i < t \leq n_{i+1}\}} + \alpha_b 1_{\{t > n_b\}},
\]

where \( 1_{\{\cdot\}} \) is the indicator function. Panel O reports simulation results for a nonnegative autoregression with \( b = 2 \) structural breaks in the sample and random breakdates \( n_1 < n_2 \). The DGP is

\[
y_t = 0.5y_{t-1} + \alpha_1 1_{\{n_1 < t \leq n_2\}} + \alpha_2 1_{\{t > n_2\}} + \varepsilon_t,
\]

with \( \alpha_1 = 2.3543 \), \( \alpha_2 = \alpha_1/2 \), and i.i.d. truncated normal \( \varepsilon_t \) (normal distribution with mean 2.3263 and variance 1, truncated at zero).\(^{18}\) The random breakdate \( n_1 \) has a discrete uniform distribution on \( \{1, \ldots, n - 2\} \), and the on \( n_1 \) conditional distribution of \( n_2 \) is discrete uniform on \( \{n_1 + 1, \ldots, n - 1\} \).\(^{19}\)

3.3.2. Random Breakdates & Occasional Jumps. Alternatively, we can specify an autoregression with \( b \) structural breaks using

\[
\sigma_t = \sum_{i=1}^{b-1} \alpha_i 1_{\{n_i < t \leq n_{i+1}\}} + \alpha_b 1_{\{t > n_b\}},
\]

and multiplicative errors \( u_t = \sigma_t \varepsilon_t \). We can further allow for jumps by letting the i.i.d. \( \varepsilon_t \) have a \( k \)-component mixture distribution, with each component representing a different jump size (e.g. small, medium, large). Panel P report simulation results for a nonnegative autoregression with \( b = 5 \) structural breaks in the sample and random breakdates \( n_1 < \cdots < n_5 \). The DGP is

\[
y_t = \beta y_{t-1} + \left( 1_{\{t \leq n_1\}} + \sum_{i=1}^{4} \alpha_i 1_{\{n_i < t \leq n_{i+1}\}} + \alpha_5 1_{\{t > n_5\}} \right) \varepsilon_t,
\]

with i.i.d. \( \varepsilon_t \) having a 2-component lognormal mixture distribution. Once more, the conditional distribution of the breakdate \( n_i \) is discrete uniform. For this experiment, the parameters of the DGP were calibrated using the S&P 500 realized kernel data considered in Section 4. Figures 1b and 1c show that this fairly simple process is able to reproduce some of the features of the S&P 500 data.

4. Empirical Results

This section reports the results of a small scale empirical study employing the LPE. We emphasize that the purpose of the study is not to advocate the superiority of the nonnegative, semiparametric, autoregressive (NNAR) model used, but rather to illustrate that its forecasting performance, presumably due to the robustness of the LPE, can match that of a commonly used benchmark model. We use a volatility proxy to model and forecast latent daily return volatility of three major stock market indexes. The Standard & Poor’s 500 (S&P 500), the NASDAQ-100, and the Dow Jones Industrial Average (DJIA). We employ the LPE as the observable process is nonnegative, persistent and likely to be nonlinear (changes in index

\(^{18}\)Here \( \mu_t = \alpha_1 1_{\{n_1 < t \leq n_2\}} + \alpha_2 1_{\{t > n_2\}} \), \( \sigma_t = 1 \) and \( m = 0 \). Other types of breaks are considered in the EA.

\(^{19}\)Note that the results for different sample sizes are not directly comparable for this (and the following) experiment, as the supports for the breakdate variables involve \( n \).
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>100</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>50</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>25</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Table 1:** Each table entry based on 1,000,000 simulated samples. Empirical bias/mean squared error of the linear programming (LP) estimators.
Figure 1. Results for the daily S&P 500 realized kernel data.
components) with measurement errors. Also, the LPE of $\beta$ avoids the need for estimating potential additional parameters, which may prove useful. Two models are considered: a simple nonnegative, semiparametric, autoregressive (NNAR) model and a parametric benchmark model.

Following Andersen et al. (2003), we use a fractionally-integrated long-memory Gaussian autoregression of order five (corresponding to five days or one trading week) for the daily logarithmic volatility proxy as our benchmark model. We use the realized kernel (RK) of Barndorff-Nielsen et al. (2008) as a volatility proxy for latent volatility. This proxy is known to be robust to certain types of market microstructure noise. Daily RK data over the period 3 January 2000 through 3 June 2014 for the three assets was obtained from the Oxford-Man Institute’s Realized Library v0.2 (Heber et al., 2009). Figure 1a shows the S&P 500 data, which will be used for illustration in the remainder of this section.

We estimate the ARFIMA($5,d,0$) for log-RK using the Ox language of Doornik (2009) and compute bias corrected forecasts for raw RK, due to the data transformation (Granger and Newbold, 1986, p. 311). We fit the NNAR model $y_t = \beta y_{t-1} + u_t$ using the LPE and calculate the, by construction, nonnegative LP residuals

$$\tilde{u}_t = y_t - \hat{\beta}_{LP} y_{t-1}.$$  

Ideally, we want to allow for an unknown number of unknown breaks in the mean as such structural breaks are able to reproduce the slow decay observed in the sample autocorrelations (cf. Section 3.3). Due to the robustness of the LPE, simple semiparametric forecasts in the presence of breaks can be obtained by applying a one-sided moving average, or moving median, to the LP residuals. Motivated by the five trading days ARFIMA specification, and the several large observations in the sample, as a simple one-day-ahead semiparametric forecast we take

$$\hat{y}_{n+1|n} = \hat{\beta}_{LP} y_n + \hat{u}_n,$$

where $\tilde{u}_n$ is the sample median of the last five LP residuals.

For the S&P 500 data, we use the period Jan 3, 2000–Dec 31, 2003 (985 observations) to initialize the forecasts, and the remaining 2612 observations to compare the forecasts. We use the recursive scheme, where the size of the sample used for parameter estimation grows as we make forecasts for successive observations.

<table>
<thead>
<tr>
<th>Model</th>
<th>S&amp;P 500 MSE</th>
<th>S&amp;P 500 QLIKE</th>
<th>NASDAQ-100 MSE</th>
<th>NASDAQ-100 QLIKE</th>
<th>DJIA MSE</th>
<th>DJIA QLIKE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-ARFIMA($5,d,0$)</td>
<td>$4.573 \times 10^{-8}$</td>
<td>-8.676</td>
<td>$1.863 \times 10^{-8}$</td>
<td>-8.699</td>
<td>$4.483 \times 10^{-8}$</td>
<td>-8.671</td>
</tr>
<tr>
<td>NNAR</td>
<td>$4.434 \times 10^{-8}$</td>
<td>-8.639</td>
<td>$1.853 \times 10^{-8}$</td>
<td>-8.676</td>
<td>$4.548 \times 10^{-8}$</td>
<td>-8.627</td>
</tr>
</tbody>
</table>

20If the observable process $\{y\}$ is a noisy proxy for an underlying latent process $\{x\}$ then measurement errors will influence the dynamics of $\{y\}$ and may conceal the persistence of $\{x\}$. In this case the robustness properties of the LPE can be useful. Consider Example 1 in Hansen and Lunde (2014) for illustration: With a slightly different notation, the latent variable is $x_t = \beta x_{t-1} + (1-\beta)\delta + v_t$ and the observable noisy, possibly biased, proxy is $y_t = x_t + \xi + \omega_t$. Thus $y_t = \beta y_{t-1} + u_t$, where $u_t = (1-\beta)(\delta + \xi) + v_t + \omega_t - \beta\omega_{t-1}$. Here the LPE consistently estimates $\beta$, the persistence parameter, under suitable conditions for the supports of the independent zero-mean disturbances $v_t$ and $\omega_t$ (ensuring that $\{x\}$ and $\{y\}$ both are nonnegative processes) when $0 < \beta < 1$ and the bias $\xi$ is positive. The LSE of $\beta$, however, is inconsistent (Hansen and Lunde, 2014).
We consider MSE and QLIKE as these loss functions are known to be robust to the use of noisy volatility proxies (Patton and Sheppard, 2009; Patton, 2011). The results for the three assets are shown in Table 2. One might, of course, also wonder if the observed differences under MSE and QLIKE in Table 2 are statistically significant or not. One way to address this question is to employ the Diebold and Mariano (1995, DM) test for equal predictive accuracy. For the S&P 500 data, the DM t-statistics under MSE and QLIKE loss are −0.009 and 0.148, respectively, indicating that the two models have equal predictive accuracy. The results for NASDAQ-100 and DJIA are similar.

5. Conclusions and Future Work

The focus in this note is on robust estimation of the AR parameter in a nonlinear, nonnegative AR model driven by nonnegative errors using a LPE or QMLE. In the previous literature the errors are assumed to be i.i.d.. Many times 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 serially correlated, heterogeneously distributed, heavy-tailed errors and give simple conditions under which the LPE is strongly consistent for these types of model mis-specifications. 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 errors of the estimated model and to isolate certain trend, seasonal or cyclical components. In addition, the observed difference between the LPE and a traditional estimator, like the LSE, can form the basis of a test for model misspecification. Our simulation results indicate that the LPE can have very reasonable finite-sample properties, and that it can be a strong alternative to the LSE when a purely autoregressive process cannot satisfactorily describe the data generating process. Our empirical study used a nonnegative semiparametric autoregressive model, estimated by the LPE, to successfully forecast latent daily return volatility of three major stock market indexes.

Some extensions may also be possible. First, a natural question is whether the established robustness generalizes to the LP estimators for extended nonnegative autoregressions described in Feigin and Resnick (1994) and Datta et al. (1998). Second, it would be interesting to see if the results of the note generalize to the multivariate setting described in Anděl (1992). These extensions will be explored in later studies.

APPENDIX

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

**Lemma 1.** Under assumptions 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 $\beta$ 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,

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

---

21We could, of course, also fit one or more parametric models to the LP residuals. For example the MEM model considered in Section 3, or a nonnegative MA model (Feigin et al., 1996). For the case with three or more forecasting models a multivariate version of the DM test can be used (Mariano and Preve, 2012). Here we restrict ourselves to two models for simplicity.

22Specifically, the DM t-statistics under MSE and QLIKE loss are −0.002 and 0.129 for NASDAQ-100, and 0.010 and 0.133 for DJIA.
The last equality follows since \( \{R_k\} \) is stochastically decreasing, and the limit since \( R_n \xrightarrow{p} 0 \) by assumption.

\[ y_l r_s \geq (c\beta)^l y_s + \sum_{j=0}^{l-1} (c\beta)^j u_{(l-j)r+s} \]

for \( l = 1, 2, \ldots \) (a.s.).

**Lemma 2.** Under assumptions 1–3,

\[ y_{(k+1)r+s} = \beta f y_{(k+1)r+s-\delta}, \ldots, y_{kr+s}, \ldots, y_{(k+1)r} + u_{(k+1)r+s} \geq c\beta y_{kr+s} + u_{(k+1)r+s} \]

where the last inequality follows by the induction assumption.

\[ (c\beta)^{k+1} y_s + \sum_{j=0}^{k} (c\beta)^j u_{(k+1-j)r+s} \]

**Proof.** We proceed with a proof by induction. Since \( cy_s \) is dominated by \( f(y_{r+s-1}, \ldots, y_s, \ldots, y_r) \), the assertion is true for \( l = 1 \). Suppose it is true for some positive integer \( k \). Then, for \( k + 1 \)

\[ y_{(k+1)r+s} \geq (c\beta)^{k+1} y_s + \sum_{j=0}^{k} (c\beta)^j u_{(k+1-j)r+s} \]

\[ \geq (c\beta)^{k+1} y_s + \sum_{j=0}^{k} (c\beta)^j u_{(k+1-j)r+s} \]

**Lemma 3.** Let \( v \) and \( w \) be i.i.d. nonnegative 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).

**Lemma 4.** Let \( v \) and \( w \) be i.i.d. nonnegative continuous random variables, and let \( \kappa > 0 \). Then the following two statements are equivalent:

(i) \( P(\kappa + v > \epsilon w) = 1 \) for some \( \epsilon > 0 \),

(ii) there exists \( c_3 \), \( 0 < c_3 < \infty \), such that \( P(v < c_3) = 1 \).

**Proof.** If (i) holds, a geometric argument shows that\[ 0 = P(w \geq \frac{6+\epsilon}{\epsilon}) \geq P(\kappa + v \leq \delta, w > \frac{6}{\epsilon}) \]

for any \( \delta > 0 \). By independence it follows that

\[ P(\kappa + v \leq \delta) P(w > \frac{6}{\epsilon}) = 0. \]  

Clearly there exists some \( \delta_0 > 0 \) such that \( P(\kappa + v \leq \delta_0) > 0 \). By (6) we then must have

\[ P(w > \frac{6}{\epsilon}) = 0. \]

Thus, \( c_3 = \delta_0/\epsilon \) is a possible value. Conversely, if (ii) holds \( 1 = P(\epsilon w < \epsilon c_3) = P(\kappa + v > \kappa) \) for all \( \epsilon > 0 \). Hence, \( P(\kappa + v > \epsilon w) = 1 \) for all \( 0 < \epsilon < \kappa/c_3 \).

**Proof of Theorem 1.** In view of Lemma 1 it is sufficient to show that \( R_n \xrightarrow{p} 0 \) as the sample size \( n \) tends to infinity. Let \( \epsilon > 0 \) be given. By a series of inequalities, we will establish an upper bound for \( P(|R_n| > \epsilon) \) and then show that this bound tends to zero as \( n \to \infty \). For ease of exposition, let \( k = 2(m+1)r \). We begin by noting that for \( n \geq k(s+1) + r + s \)

\[ P(|R_n| > \epsilon) = P(R_n > \epsilon) = P(u_t > \epsilon f(y_{t-1}, \ldots, y_{t-s}); t = s + 1, \ldots, n) \]

\[ \leq P(u_{ki+r+s} > \epsilon f(y_{ki+r+s-1}, \ldots, y_{ki+r}); i = s + 1, \ldots, N), \]
where $N = \lfloor (n - r - s)/k \rfloor$, the integer part of $(n - r - s)/k$. Here the first equality follows as $R_n$ is nonnegative a.s. under assumptions 1–2. Apparently, $s + 1 \leq N < n$ and tends to infinity as $n \to \infty$. Let $l = ki/r$. By Assumption 2 and Lemma 2, respectively,

$$f(y_{lt+r+s-1}, \ldots, y_{lt+s}, \ldots, y_{lt+r}) \geq c y_{lt+s}$$

$$\geq c^{l+1} \beta^l y_{lt+s} + \sum_{j=0}^{l-1} c^{j+1} \beta^j u_{(l-j)r+s}$$

$$\geq c^{l+1} \beta^l u_{(l-j)r+s}$$

for each $j \in \{0, \ldots, l-1\}$. Hence, for $j = m$ it is readily verified that

$$P(|R_n| > \epsilon) \leq P(u_{ki+r+s} > \epsilon c^{m+1} \beta^m u_{ki-mr+s}; i = s+1, \ldots, N).$$

By assumptions 3 and 4, respectively,

$$P(|R_n| > \epsilon) \leq P(u_{ki+r+s} > \epsilon c^{m+1} \beta^m u_{ki-mr+s}; i = s+1, \ldots, N)$$

$$= P(\mu_{ki+r+s} + \sigma_{ki+r+s} \epsilon_{ki+r+s} > \epsilon c^{m+1} \beta^m \mu_{ki-mr+s} + \epsilon c^{m+1} \beta^m \sigma_{ki-mr+s} \epsilon_{ki-mr+s}; i = s+1, \ldots, N)$$

$$\leq P(\mu_{ki+r+s} + \sigma_{ki+r+s} \epsilon_{ki+r+s} > \epsilon c^{m+1} \beta^m \sigma_{ki-mr+s} \epsilon_{ki-mr+s}; i = s+1, \ldots, N)$$

$$\leq P(\mu_{ki+r+s} + \sigma_{ki+r+s} \epsilon_{ki+r+s} > \epsilon c^{m+1} \beta^m \sigma_{ki-mr+s} \epsilon_{ki-mr+s}; i = s+1, \ldots, N).$$

Moreover, by assumption 4,

$$P(|R_n| > \epsilon) \leq P(\mu_{ki+r+s} \epsilon_{ki+r+s} > \epsilon c^{m+1} \beta^m \sigma_{ki-mr+s} \epsilon_{ki-mr+s}; i = s+1, \ldots, N)$$

$$= P(\mu_{ki+r+s} \epsilon_{ki+r+s} > \epsilon c^{m+1} \beta^m (\sigma_{ki-mr+s} / \sigma_{ki+r+s}) \epsilon_{ki-mr+s}; i = s+1, \ldots, N)$$

$$\leq P(\mu_{ki+r+s} \epsilon_{ki+r+s} > \epsilon c^{m+1} \beta^m \epsilon_{ki-mr+s}; i = s+1, \ldots, N),$$

where $\kappa = \mu / \sigma$ and $\epsilon = \epsilon c^{m+1} \beta^m (\sigma / \sigma)$. We first consider case (i) of the theorem. Since $\mu_t = 0$ for all $t$ we can take $\mu = 0$, which gives $\kappa = 0$ and

$$P(|R_n| > \epsilon) \leq P(\epsilon_{ki+r+s} \epsilon_{ki-mr+s}; i = s+1, \ldots, N).$$

Since the sequence $\epsilon_{s+1}, \ldots, \epsilon_n$ of errors is $m$-dependent, $\epsilon_t$ and $\epsilon_{t+k}$ are pairwise independent for all $k > m$. Let $\zeta_{i} = \epsilon_{ki+r+s} / \epsilon_{ki-mr+s}$. Then $\zeta_{s+1}, \ldots, \zeta_N$ is a sequence of i.i.d. random variables, for which the numerator and denominator of each $\zeta_i$ are pairwise independent, and hence

$$P(|R_n| > \epsilon) \leq P(\zeta_{s+1} > \epsilon) \times \cdots \times P(\zeta_N > \epsilon) = P(\epsilon_{k(s+1)+r+s} > \epsilon \epsilon_{k(s+1)-mr+s})^{N-s}.$$

In view of Lemma 3 and the limiting behavior of $N$ this implies that $P(|R_n| > \epsilon) \to 0$ as $n \to \infty$. Since $\epsilon > 0$ was arbitrary, $R_n$ converges in probability to zero. Similarly, for case (ii) where $\kappa > 0$ we have that

$$P(|R_n| > \epsilon) \leq P(\kappa + \epsilon_{k(s+1)+r+s} > \epsilon \epsilon_{k(s+1)-mr+s})^{N-s}.$$

In view of Lemma 4 this also implies that $R_n$ converges in probability to zero. \qed

**References**


