Global Research Unit
Working Paper #2016-001

Linear Programming-based Estimators In Nonnegative Autoregression

Daniel Preve
City University of Hong Kong

LINEAR PROGRAMMING-BASED ESTIMATORS IN NONNEGATIVE AUTOREGRESSION

DANIEL P. A. PREVE[†]

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. Comments and suggestions from two anonymous referees are also gratefully acknowledged. 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).[1] This note considers a nonlinear, nonnegative autoregressive model driven by nonnegative errors. More specifically, it considers robust estimation of the AR parameter β in the autoregression

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

with nonnegative (possibly) misspecified errors 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. (2015). 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 $y_1, \ldots, y_n$ of size n and the nonnegativity of the errors is

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

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.[2] 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 errors 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 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

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

[2] Bell 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.[3] 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. The case when $p = 1$ is special in our setting as the linear programming problem of maximizing the objective function $g(\beta_1, \ldots, \beta_p) = \sum_{i=1}^{p} \beta_i$ subject to the $n - s$ linear constraints

$$y_t - \sum_{i=1}^{p} \beta_i f_i(y_{t-1}, \ldots, y_{t-s}) \geq 0$$

(cf. Feigin and Resnick, 1994) then has an explicit solution. This simplifies the statistical analysis of the LPE. In general ($p > 1$), one has to rely on numerical methods.

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 η, 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

[3] The 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 $s = 22$.

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 $f(\mathbf{x}) = f(x_1, \ldots, x_r, \ldots, x_s) \geq cx_r$ when all of its arguments are nonnegative.*

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^{x_s}, $\sinh x_s$ and any polynomial in x_s of degree higher than 0 with positive coefficients.[4] 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.[5] For example, finite-order moving average (MA) specifications (Section 3.2.2). The σ_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 μ_t, σ_t and ε_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 $0 \leq \overline{\mu} < \infty$ 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 μ_t and σ_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).[6]

2.2. **Finite-Sample Theory.** The nonlinear, nonnegative autoregression implied by assumptions 1–4 is flexible and nests several specifications in the related literature.[7] 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.

---

[4]An interesting example of a multi-variable function f is given by the AR index process considered by Im et al. (2006) for which $f(\mathbf{x}) = x_1 + \cdots + x_s$ or, equivalently, $f(y_{t-1}, \ldots, y_{t-s}) = y_{t-1} + \cdots + y_{t-s}$. The AR index models of order 1, 5 and 22 all can be viewed as special cases of Corsi's (2009) HAR model.

[5]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.

[6]It is important to note that μ_t and σ_t are allowed to be *degenerate* random variables (i.e. deterministic).

[7]For example, Bell and Smith's specification is obtained by choosing $f(\mathbf{x}) = x_1$ 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..

2.3. **Asymptotic Theory.**

2.3.1. *Convergence.* Previous works focusing explicitly on the strong 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 β, the LPE is distribution-free in the sense that 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 $\beta < 1$.[8] For instance, in the linear AR(1) with exponential errors 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.

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 β in (1), i.e. $\hat{\beta}_n$ converges to β a.s. as n tends to infinity, if either (i) $P(c_1 < \varepsilon_t < c_2) < 1$ for all $0 < c_1 < c_2 < \infty$ and $\mu_t = 0$ for all t, or (ii) $P(\varepsilon_t < c_3) < 1$ for all $0 < c_3 < \infty$.*

In other words, the LPE remains a consistent estimator for β 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 (sufficient, but not necessary), and that the consistency conditions of the LPE do not involve the existence of any moments.[9] 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).

---

[8]This occurs, under some additional conditions, when the exponent of regular variation of the error 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).

[9]As an extreme example, consider estimating $0 < \beta < 1$ in the linear specification $y_t = \beta y_{t-1} + u_t$ with independent, nonnegative stable errors $u_t \sim \mathcal{S}(a, b, c, d; 1)$, 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, 2015). 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 (cf. the Monte Carlo experiment with Lévy distributed errors in Section 3.1.2).

3. Simulation Results

In this section we report simulation results concerning the estimation of the AR parameter β in the nonnegative autoregression $y_t = \beta f(y_{t-1}, \ldots, y_{t-s}) + u_t$, $u_t = \mu_t + \sigma_t \varepsilon_t$, 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 $s = 1, 4$.[10] Thus, in the simulations the data generating process (DGP) is

$$y_t = \beta y_{t-s} + u_t, \quad \text{with} \quad u_t = \mu_t + \sigma_t \varepsilon_t, \quad t = s+1, s+2, \ldots \tag{3}$$

and the LPE is

$$\hat{\beta}_{LP} = \min \left\{ \frac{y_{s+1}}{y_1}, \ldots, \frac{y_n}{y_{n-s}} \right\} = \beta + \min \left\{ \frac{u_{s+1}}{y_1}, \ldots, \frac{u_n}{y_{n-s}} \right\}. \tag{4}$$

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 β 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 β given a sample $y_1, \ldots, y_n$ of size $n > s$ is

$$\hat{\beta}_{LS} = \frac{\sum_{t=s+1}^{n} (y_t - \bar{y}_+)(y_{t-s} - \bar{y}_-)}{\sum_{t=s+1}^{n} (y_{t-s} - \bar{y}_-)^2} = \beta + \frac{\sum_{t=s+1}^{n} (u_t - \bar{u}_+)(y_{t-s} - \bar{y}_-)}{\sum_{t=s+1}^{n} (y_{t-s} - \bar{y}_-)^2}, \tag{5}$$

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 β 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 μ_t, σ_t and ε_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 β. 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. All the reported experiments share a common initial state of the generator for pseudo-random number generation, use initial values $y_1 = \sum_{i=0}^{100\,000} \beta^i u_{1-is}$ as an approximation of $\sum_{i=0}^{\infty} \beta^i u_{1-is}$, and were carried out using Matlab. The initial values μ_{s+1} and σ_{s+1} for the experiments in sections 3.2.1, 3.2.4 and 3.3.1.–3.3.2 were obtained similarly.[11]

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.

[10]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.

[11]Matlab code for the Monte Carlo experiments of this section is available through the authors webpage at http://www.researchgate.net\profile\Daniel_Preve.

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.[12] 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. 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. The simulation results show that the accuracy of the LPE can be quite remarkable. For example, when $n = 25$ and the errors are exponentially distributed the results are close to those expected by the limit theory. The magnitudes of the bias and MSE of the LSE on the other hand are nearly 5 and 40 times as large, respectively.

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.[13] 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.[14] 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). The simulation results show that LPE performs well, particularly in small samples. The bias and MSE of the LSE on the other hand can be severe even in moderate samples, as illustrated in Panel E.

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)$\times$(1,0)$_4$.[15] These three experiments illustrate that the bias and MSE of the LPE for a fixed n, viewed as a function of β, is stochastically decreasing in the AR parameter. It can be shown that this property holds under fairly general conditions on f in Assumption 2.[16]

3.2. **Serially Correlated Errors.** Panels I–N of Table 1 report simulation results when $y_t = \beta y_{t-1} + u_t$, 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

[12]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).

[13]Pareto and Lévy pseudo-random numbers were generated using the inversion method and via Problem 1.17 in Nolan (2015), respectively.

[14]Cf. Corollary 1 in the EA.

[15]More generally, it is not difficult to show that (Proposition 3 in the EA) the LPE in (4) can consistently estimate the AR parameter of a nonnegative covariance stationary SARMA(0,q)$\times$(1,Q)$_4$ process under the assumptions of Theorem 1.

[16]See Proposition 4 in the EA.

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.[17]

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

$$u_t = \sigma_t \varepsilon_t$$
$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots$$

with i.i.d. ε_t, that is, a MEM(p,q). Here μ_t in Assumption 3 is zero for all t and m is also zero. 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

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

with independent identically beta distributed ε_t (beta distributed with both shape parameters equal to 2 and, hence, with about 0.5 symmetric common density), and $\alpha_1 = 0.2$, $\beta_1 = 0.75$, $\alpha_0 = 1 - \alpha_1 - \beta_1$. 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). It can be shown that Assumption 4 is satisfied for this case with $\overline{\mu} = 0$, $\underline{\sigma} = 0.2$ and $\overline{\sigma} = 1$.

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

$$u_t = \varepsilon_t = \epsilon_t + \sum_{i=1}^{q} \psi_i \epsilon_{t-i}, \quad t = 0, \pm 1, \pm 2, \dots$$

with i.i.d. ϵ_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 $q = 1, 2$.[18] The DGPs for panels J and K are

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

respectively, where the ϵ_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. Although the LPE is strongly consistent in this case whenever q is finite, extended simulations not reported in the note indicate that its convergence can be slow for large values of q.

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

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

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

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

respecively, where the ϵ_t are i.i.d. inverse Gaussian with mean and variance both equal to 1.

[17] Recall that the variance of an estimator is equal to the difference of its MSE and its squared bias.

[18] For example, finite sums of finite-order MA processes driven by i.i.d. disturbances are m-dependent.

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

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

with i.i.d. ε_t. Here the pth-order AR specification for μ_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

$$y_t = 0.75 y_{t-1} + \mu_t + \varepsilon_t, \quad \mu_t = 0.25\mu_{t-1} + \epsilon_t,$$

with standard exponential ε_t (mean and variance both equal to 1) and i.i.d. on $(0, 25)$ uniform ϵ_t, mutually independent of the ε_t. It is not difficult to show that Assumption 4 is satisfied for this case with $\overline{\mu} = 100/3$ and $\underline{\sigma} = \overline{\sigma} = 1$.

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.

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

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

where $\mathbf{1}_{\{\cdot\}}$ 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

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

with $\alpha_1 = 2.3543$, $\alpha_2 = \alpha_1/2$, and i.i.d. truncated normal ε_t (normal distribution with mean 2.3263 and variance 1, truncated at zero).[19] 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 $\{n_1 + 1, \ldots, n-1\}$.[20]

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

$$\sigma_t = \mathbf{1}_{\{t \leq n_1\}} + \sum_{i=1}^{b-1} \alpha_i \mathbf{1}_{\{n_i < t \leq n_{i+1}\}} + \alpha_b \mathbf{1}_{\{t > n_b\}},$$

and multiplicative errors $u_t = \sigma_t \varepsilon_t$. We can further allow for jumps by letting the i.i.d. ε_t have a k-component mixture distribution, with each component representing a different jump size

[19] Here $\mu_t = \alpha_1 \mathbf{1}_{\{n_1 < t \leq n_2\}} + \alpha_2 \mathbf{1}_{\{t > n_2\}}$, $\sigma_t = 1$ and $m = 0$. Other types of breaks are considered in the EA.

[20] 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.

(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

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

with i.i.d. ε_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. The purpose of the study is twofold: First, it illustrates how the robustness of the LPE can be capitalized on to estimate and forecast a simple semiparametric model in an environment where traditional estimators, like the LSE, are likely to be inconsistent. The model combines a parametric component, taking into account the most recent observation, with a nonparametric component - forecasted by a moving median similar to the moving averages frequently used in technical analysis, taking into account possible omitted variables, measurement errors and structural breaks. Second, it also illustrates that volatility proxy forecasts of a parsimonious (single parameter) model estimated by the LPE can be at least as accurate as those of a rather involved (multiple parameter) ARFIMA benchmark model. The ARFIMA is probably the most commonly used model for forecasting volatility proxies. See, for example, Andersen et al. (2003), Koopman et al. (2005), Lanne (2006), Corsi (2009) and the references therein.

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 components) with measurement errors.[21] Also, the LPE of β 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.

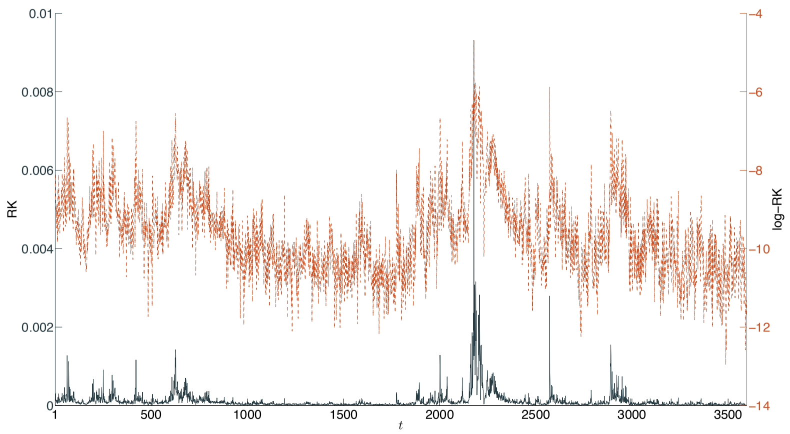
Following Andersen et al. (2003), we use a fractionally-integrated long-memory Gaussian autoregression of order five (corresponding to five days or one trading week) for the daily *logarithmic* volatility proxy as our benchmark model. We use the realized kernel (RK) of Barndorff-Nielsen et al. (2008) 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

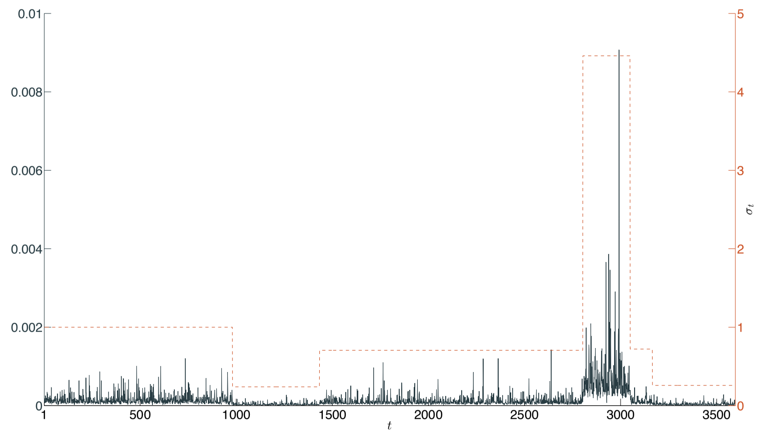
[21] If the observable process $\{y_t\}$ is a noisy proxy for an underlying latent process $\{x_t\}$ then measurement errors will influence the dynamics of $\{y_t\}$ and may conceal the persistence of $\{x_t\}$. In this case the robustness properties of the LPE can be useful. Consider Example 1 in Hansen 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 $y_t = x_t + \xi + w_t$. Thus $y_t = \beta y_{t-1} + u_t$, where $u_t = (1-\beta)(\delta + \xi) + v_t + w_t - \beta w_{t-1}$. Here the LPE consistently estimates β, the persistence parameter, under suitable conditions for the supports of the independent zero-mean disturbances v_t and w_t (ensuring that $\{x_t\}$ and $\{y_t\}$ both are nonnegative processes) when $0 < \beta < 1$ and the bias ξ is positive. The LSE of β, however, is inconsistent (Hansen and Lunde, 2014).

Table 1. Each table entry, based on 1 000 000 simulated samples, reports the empirical bias/mean squared error of the linear programming (LP) and ordinary least squares (LS) estimators for the AR parameter β in the nonnegative autoregression $y_t = \beta y_{t-s} + u_t$. Different sample sizes (n), AR parameters (β), lags (s), and specifications for u_t are considered. See Section 3 for details.

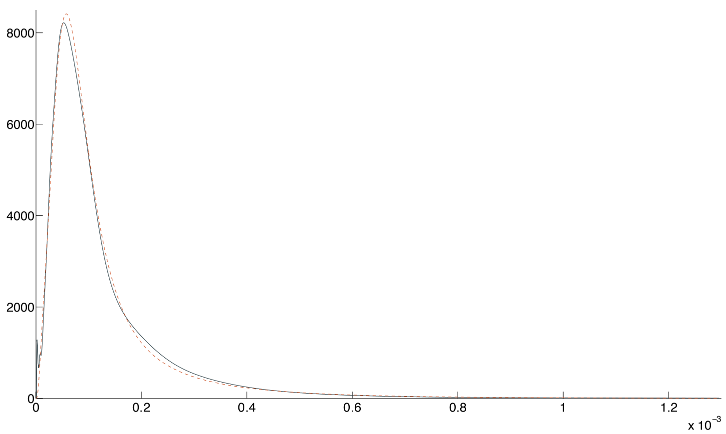
	$\hat{\beta}_{LP}$		$\hat{\beta}_{LS}$		$\hat{\beta}_{LP}$		$\hat{\beta}_{LS}$		$\hat{\beta}_{LP}$		$\hat{\beta}_{LS}$		$\hat{\beta}_{LP}$		$\hat{\beta}_{LS}$	
n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
			Panel A (uniform)				*Panel B (exponential)*		*Panel C (Weibull)*				*Panel D (Pareto)*			
25	0.038	0.003	−0.105	0.048	0.020	0.001	−0.098	0.040	0.095	0.011	−0.103	0.046	0.082	0.009	−0.086	6.842
50	0.019	0.001	−0.051	0.020	0.010	0.000	−0.049	0.017	0.067	0.006	−0.051	0.019	0.053	0.004	−0.041	0.600
100	0.010	0.000	−0.025	0.009	0.005	0.000	−0.025	0.008	0.048	0.003	−0.025	0.009	0.033	0.002	−0.020	0.020
200	0.005	0.000	−0.013	0.004	0.003	0.000	−0.012	0.004	0.034	0.001	−0.013	0.004	0.020	0.001	−0.010	0.003
400	0.003	0.000	−0.006	0.002	0.001	0.000	−0.006	0.002	0.024	0.001	−0.006	0.002	0.012	0.000	−0.005	0.002
800	0.001	0.000	−0.003	0.001	0.001	0.000	−0.003	0.001	0.017	0.000	−0.003	0.001	0.007	0.000	−0.003	0.001
1600	0.001	0.000	−0.002	0.001	0.000	0.000	−0.002	0.001	0.012	0.000	−0.002	0.001	0.004	0.000	−0.001	0.000
			Panel E (Lévy)				*Panel F* (SARMA, $\beta = 0.25$)		*Panel G (SARMA, $\beta = 0.5$)*				*Panel H (SARMA, $\beta = 0.75$)*			
25	0.003	0.000	−517.081	1.491×10^{11}	0.149	0.028	−0.065	0.050	0.103	0.013	−0.085	0.049	0.053	0.004	−0.091	0.042
50	0.001	0.000	254.149	1.237×10^{11}	0.101	0.013	−0.035	0.021	0.070	0.006	−0.048	0.020	0.036	0.002	−0.056	0.016
100	0.000	0.000	−97.272	5.013×10^{9}	0.071	0.006	−0.018	0.010	0.048	0.003	−0.025	0.009	0.025	0.001	−0.031	0.006
200	0.000	0.000	−6.934	2.131×10^{7}	0.050	0.003	−0.009	0.005	0.034	0.002	−0.012	0.004	0.017	0.000	−0.016	0.003
400	0.000	0.000	−1.524	2.336×10^{6}	0.035	0.002	−0.004	0.002	0.024	0.001	−0.006	0.002	0.012	0.000	−0.008	0.001
800	0.000	0.000	−0.003	58.263	0.025	0.001	−0.002	0.001	0.017	0.000	−0.003	0.001	0.009	0.000	−0.004	0.001
1600	0.000	0.000	−0.008	42.893	0.017	0.000	−0.001	0.001	0.012	0.000	−0.002	0.001	0.006	0.000	−0.002	0.000
			Panel I (MEM(1,1))				*Panel J (MA(1))*		*Panel K (MA(2))*				*Panel L (nonlinear I)*			
25	0.104	0.014	−0.020	0.037	0.066	0.005	0.043	0.013	0.089	0.009	0.100	0.019	0.072	0.006	0.081	0.030
50	0.073	0.007	0.054	0.020	0.053	0.003	0.084	0.011	0.074	0.006	0.132	0.020	0.055	0.003	0.126	0.026
100	0.052	0.003	0.094	0.017	0.045	0.002	0.103	0.012	0.063	0.004	0.147	0.023	0.043	0.002	0.151	0.027
200	0.036	0.002	0.114	0.017	0.038	0.002	0.113	0.013	0.054	0.003	0.154	0.024	0.034	0.001	0.164	0.029
400	0.026	0.001	0.124	0.017	0.033	0.001	0.117	0.014	0.048	0.002	0.158	0.025	0.027	0.001	0.171	0.031
800	0.018	0.000	0.130	0.018	0.028	0.001	0.120	0.015	0.042	0.002	0.160	0.026	0.023	0.001	0.175	0.031
1600	0.013	0.000	0.132	0.018	0.025	0.001	0.121	0.015	0.037	0.001	0.161	0.026	0.019	0.000	0.177	0.032
			Panel M (nonlinear II)				*Panel N (omitted variables)*		*Panel O (breaks)*				*Panel P (breaks & jumps)*			
25	0.078	0.007	0.104	0.035	0.074	0.006	−0.022	0.020	0.170	0.036	0.201	0.064	0.080	0.009	0.123	0.057
50	0.061	0.004	0.154	0.033	0.062	0.004	0.038	0.008	0.127	0.022	0.257	0.080	0.069	0.006	0.189	0.072
100	0.049	0.003	0.180	0.037	0.052	0.003	0.066	0.007	0.092	0.013	0.285	0.091	0.058	0.004	0.225	0.081
200	0.040	0.002	0.194	0.040	0.044	0.002	0.079	0.007	0.063	0.007	0.299	0.098	0.048	0.003	0.244	0.085
400	0.033	0.001	0.202	0.042	0.038	0.002	0.085	0.008	0.042	0.004	0.306	0.101	0.039	0.002	0.253	0.085
800	0.028	0.001	0.207	0.043	0.033	0.001	0.088	0.008	0.027	0.002	0.310	0.103	0.032	0.001	0.257	0.084
1600	0.024	0.001	0.209	0.044	0.029	0.001	0.090	0.008	0.016	0.001	0.312	0.104	0.026	0.001	0.259	0.082



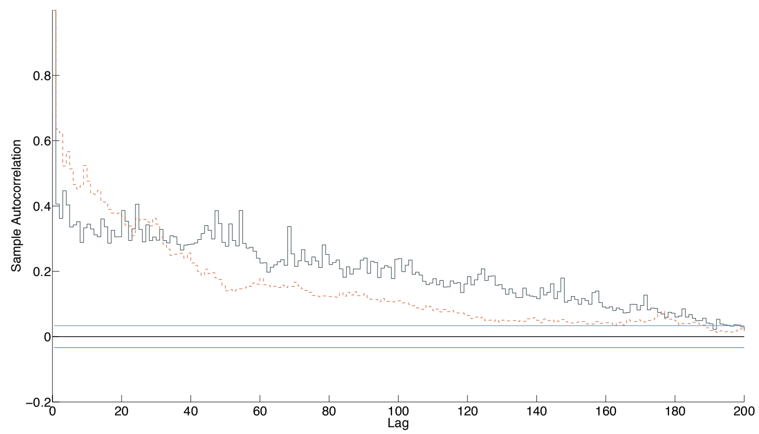
(A) Daily S&P 500 realized kernel (RK, solid line) and logarithmic realized kernel (log-RK, dashed line).



(B) Sample path of a nonnegative autoregression (solid line), with parameters calibrated to the S&P 500 RK data, and its structural breaks (dashed line).



(C) Kernel density estimate (solid line) and fitted two component, lognormal mixture density (dashed line) for the LPE residuals of the initial sample used in the S&P 500 RK forecasting exercise.



(D) Sample autocorrelations for the S&P 500 RK data in Figure 1a (dashed line) and the associated simulated data in Figure 1b (solid line).

Figure 1. Results for the daily S&P 500 realized kernel data.

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

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

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 of volatility proxies (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

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

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 2. Forecasting performance (one-day-ahead forecasts).

	S&P 500		*NASDAQ-100*		*DJIA*	
Model	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Log-ARFIMA(5,d,0)	4.573×10^{-8}	-8.676	1.863×10^{-8}	-8.699	4.483×10^{-8}	-8.671
NNAR	4.434×10^{-8}	-8.639	1.853×10^{-8}	-8.676	4.548×10^{-8}	-8.627

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.[22] The results for NASDAQ-100 and DJIA are similar.[23]

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

[22]We 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.

[23]Specifically, 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. We emphasize that the purpose of the study is *not* to advocate the superiority of the NNAR model, 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.

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 misspecifications. 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 show 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. Extended simulations not reported in the note indicate that the LPE works best when the probability that ε_t in Assumption 3 is near zero (cf. the first condition of Theorem 1), or is relatively large (cf. the second condition of Theorem 1), is not close to zero. 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 1–2,* $R_n \xrightarrow{p} 0 \Rightarrow \hat{\beta}_n \xrightarrow{a.s.} \beta$.

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 the sequence $\{R_k\}$ of nonnegative random variables is stochastically decreasing, and the limit since $R_n \xrightarrow{p} 0$ by assumption. □

Lemma 2. *Under assumptions 1–2,*

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

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

Proof. We proceed with a proof by induction. By Assumption 1, with $t = r + s$,

$$y_{r+s} = \beta f(y_{r+s-1}, \ldots, y_r) + u_{r+s}. \tag{6}$$

By Assumption 2, with $x_1 = y_{(r+s)-1}, \ldots, x_r = y_{(r+s)-r}, \ldots, x_s = y_{(r+s)-s}$,

$$f(y_{r+s-1}, \ldots, y_s, \ldots, y_r) \geq c y_s. \tag{7}$$

Equations (6) and (7) together imply that

$$y_{r+s} \geq c\beta y_s + u_{r+s}.$$

Hence the assertion is true for $l = 1$. Suppose it is true for some positive integer k. Then, for $k+1$

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

where the last inequality follows by the induction assumption. □

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

$$0 = P(w \geq \tfrac{\kappa+v}{\epsilon}) \geq P(\kappa + v \leq \delta,\, w > \tfrac{\delta}{\epsilon})$$

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

$$P(\kappa + v \leq \delta) P(w > \tfrac{\delta}{\epsilon}) = 0. \tag{8}$$

Clearly there exists some $\delta_0 > 0$ such that $P(\kappa + v \leq \delta_0) > 0$. By (8) we then must have

$$P(w > \tfrac{\delta_0}{\epsilon}) = 0.$$

Since v and w are identically distributed we can take $c_3 = \delta_0/\epsilon$. Conversely, if (ii) holds $P(\epsilon w < \epsilon c_3) = 1$ for all $\epsilon > 0$. At the same time $P(\kappa + v > \kappa) = 1$. 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 $\varepsilon > 0$ be given. By a series of inequalities, we will establish an upper bound for $P(|R_n| > \varepsilon)$ and then show that this bound tends to zero as $n \to \infty$. For ease of exposition, let $k = 2(m+1)r$ where m is as in Assumption 3. We begin by noting that for $n \geq k(s+1) + r + s$

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

where N is 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,

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

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

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

By assumptions 3 and 4, respectively,

$$\begin{aligned}
P(|R_n| > \varepsilon) &\le P(u_{ki+r+s} > \varepsilon c^{m+1}\beta^m u_{ki-mr+s};\ i = s+1, \ldots, N) \\
&= P(\mu_{ki+r+s} + \sigma_{ki+r+s}\varepsilon_{ki+r+s} > \varepsilon c^{m+1}\beta^m \mu_{ki-mr+s} + \varepsilon c^{m+1}\beta^m \sigma_{ki-mr+s}\varepsilon_{ki-mr+s};\ i = s+1, \ldots, N) \\
&\le P(\mu_{ki+r+s} + \sigma_{ki+r+s}\varepsilon_{ki+r+s} > \varepsilon c^{m+1}\beta^m \sigma_{ki-mr+s}\varepsilon_{ki-mr+s};\ i = s+1, \ldots, N) \\
&\le P(\overline{\mu} + \sigma_{ki+r+s}\varepsilon_{ki+r+s} > \varepsilon c^{m+1}\beta^m \sigma_{ki-mr+s}\varepsilon_{ki-mr+s};\ i = s+1, \ldots, N).
\end{aligned}$$

Moreover, by assumption 4,

$$\begin{aligned}
P(|R_n| > \varepsilon) &\le P(\overline{\mu} + \sigma_{ki+r+s}\varepsilon_{ki+r+s} > \varepsilon c^{m+1}\beta^m \sigma_{ki-mr+s}\varepsilon_{ki-mr+s};\ i = s+1, \ldots, N) \\
&= P(\overline{\mu}/\sigma_{ki+r+s} + \varepsilon_{ki+r+s} > \varepsilon c^{m+1}\beta^m (\sigma_{ki-mr+s}/\sigma_{ki+r+s})\varepsilon_{ki-mr+s};\ i = s+1, \ldots, N) \\
&\le P(\kappa + \varepsilon_{ki+r+s} > \epsilon\varepsilon_{ki-mr+s};\ i = s+1, \ldots, N),
\end{aligned}$$

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

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

Since the sequence $\varepsilon_{s+1}, \ldots, \varepsilon_n$ of errors is m-dependent, ε_t and ε_{t+k} are pairwise independent for all $k > m$. Let $\xi_i = \varepsilon_{ki+r+s}/\varepsilon_{ki-mr+s}$. Then $\xi_{s+1}, \ldots, \xi_N$ is a sequence of i.i.d. random variables, for which the numerator and denominator of each ξ_i are pairwise independent, and hence

$$P(|R_n| > \varepsilon) \le P(\xi_{s+1} > \epsilon) \times \cdots \times P(\xi_N > \epsilon) = P(\varepsilon_{k(s+1)+r+s} > \epsilon\varepsilon_{k(s+1)-mr+s})^{N-s}.$$

In view of Lemma 3 and the limiting behavior of 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. Similarly, for case (ii) where $\kappa > 0$ we have that

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

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

References

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

Andersen, T.G., T. Bollerslev, F.X. Diebold & P. Labys (2003) Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.

Andrews, B., M. Calder & R.A. Davis (2009) Maximum likelihood estimation for α-stable autoregressive processes. *The Annals of Statistics* 37, 1946–1982.

Andrews, B. & R.A. Davis (2013) Model identification for infinite variance autoregressive processes. *Journal of Econometrics* 172, 222–234.

Alexandridis, A.K. & A.D. Zapranis (2013) *Weather derivatives: modeling and pricing weather-related risk*. Springer, New York.

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.

Anděl, J. (1992) Nonnegative multivariate AR(1) processes. *Kybernetika* 28, 213–226.

Barndorff-Nielsen, O.E. & N. Shephard (2004) Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1–37.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde & N. Shephard (2008) Designing realized kernels to measure ex post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.

Bauwens, L. & P. Giot (2000) The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Économie et de Statistique* 60, 117–149.

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

Campbell, S.D. & F.X. Diebold (2005) Weather forecasting for weather derivatives. *Journal of the American Statistical Association* 100, 6–16.

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.

Diebold, F.X. & R.S. Mariano (1995) Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 134–145.

Doornik, J.A. (2009) *An Object-Oriented Matrix Programming Language Ox 6*. Timberlake Consultants Press, London.

Engle, R.F. & J.R. Russell (1998) Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66, 1127–1162.

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

Fama, E.F. (1965) The behavior of stock-market prices. *The Journal of Business* 38, 34–105.

Feigin, P.D., M.F. Kratz, & S.I. Resnick (1996) Parameter estimation for moving averages with positive innovations. *The Annals of Applied Probability* 6, 1157–1190.

Feigin, P.D. & S.I. Resnick (1992) Estimation for autoregressive processes with positive innovations. *Communications in Statistics - 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.

Granger, C.W.J. & P. Newbold (1986) *Forecasting Economic Time Series (2nd ed.)*. Academic Press, Orlando Florida.

Hansen, P.R. & A. Lunde (2014) Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory* 30, 60–93.

Heber, G., A. Lunde, N. Shephard, & K. Sheppard (2009) "Oxford-Man Institute's Realized Library v0.2", Oxford-Man Institute, University of Oxford.

Im, E.I., D.L. Hammes, & D.T. Wills (2006) Stationarity condition for AR index process. *Econometric Theory* 22, 164–168.

Koopman, S.J., B. Jungbacker, & E. Hol (2005) Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12, 445–475.

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

Ling, S. (2005) Self-weighted absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 381–393.

Mariano, R.S. & D. Preve (2012) Statistical tests for multiple forecast comparison. *Journal of Econometrics* 169, 123–130.

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. (2015) *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

Nouali, K. & H. Fellag (2005) Testing of the first-order autoregressive model with uniform innovations under contamination. *Journal of Mathematical Sciences* 131, 5657–5663.

Patton, A.J. (2011) Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.

Patton, A.J. & K. Sheppard (2009) Evaluating volatility and correlation forecasts, in T.G. Andersen, R.A. Davis, J.P. Kreiß and T. Mikosch (Eds.), *Handbook of Financial Time Series*. Springer, Berlin Heidelberg.

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

Preve, D., A. Eriksson, & J. Yu (2015) 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.

Seshadri, V. (1993) *The Inverse Gaussian Distribution - A Case Study in Exponential Families*. Clarendon Press, Oxford.

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.