LIKELIHOOD ANALYSIS OF A FIRST ORDER AUTOREGRESSIVE MODEL WITH EXPONENTIAL INNOVATIONS

By B. Nielsen & N. Shephard
Nuffield College, Oxford OX1 1NF, UK
bent.nielsen@nuf.ox.ac.uk
neil.shephard@nuf.ox.ac.uk
19 July 1999

Summary

This paper derives the exact distribution of the maximum likelihood estimator of a first order linear autoregression with an exponential disturbance term. We also show that even if the process is stationary, the estimator is T-consistent, where T is the sample size. In the unit root case the estimator is T^2-consistent, while in the explosive case the estimator is ρ^T-consistent. Further, the likelihood ratio test statistic for a simple hypothesis on the autoregressive parameter is asymptotically uniform for all values of the parameter.

Some key words: Autoregression; Exact distribution; Exponential innovations; Likelihood; Non-regular asymptotics; Stochastic volatility.

1. Introduction

This paper looks at the distributional behaviour of the maximum likelihood estimator of the non-negative first order linear autoregressive time series model

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad \text{for} \quad t = 1, ..., T, \tag{1}$$

where the initial value X_0 is fixed, $\{\varepsilon_t\}$ is a sequence of independent exponentially distributed random variables with common positive scale parameter λ, and the autoregression parameter ρ is non-negative. When $\rho < 1$, ρ has the interpretation of the first order autocorrelation of the series.

The main point of this paper is to derive the exact distribution of the maximum likelihood estimator, written $\widehat{\rho}$, of ρ, conditioning on some initial value X_0,

$$\widehat{\rho} = \min_{1 \leq t \leq T} \left(\frac{X_t}{X_{t-1}} \right). \tag{2}$$

We will show that the distribution is remarkably simple whatever the true value of ρ. In particular $\widehat{\rho}$ is always upward biased. When $\rho < 1$ the estimator is asymptotically exponential and T-consistent. This result extends to the case where $\rho = 1$. Then the distribution will again be exponential but this time the estimator will be T^2-consistent. These convergence rates are the squares of those obtained in the Gaussian autoregressive model, see White (1958). In the explosive case the estimator is ρ^T-consistent as in the Gaussian case although the limit distribution is more complicated. In all cases the likelihood ratio test statistic for $\rho = \rho_0$ is asymptotically uniform distributed. This is different from the Gaussian case where the test distribution depends on ρ_0.

A number of related results are known. For the stationary case Bell and Smith (1986) proved almost sure consistency for $\hat{\rho}$ whereas An and Huang (1993) showed that the consistency is faster than the usual $T^{1/2}$ consistency rate. In both cases more general error distributions were considered as well. Consistency results for a second order process were given by Anděl (1989). Anděl (1988) found the distribution of an approximation to $\hat{\rho}$. As mentioned above the process given by (1) can be given a stationary initial distribution, for $\rho < 1$. The stationary distribution is non-standard, see Matthai (1982) and Sim (1992). To overcome this deficiency some of the literature has focused on non-linear autoregressive processes in exponential variables, see for instance Lawrance and Lewis (1985). These are processes with exponential marginals whereas the innovations are mixtures of exponentials. Raftery (1980,1982) discussed almost sure consistency as well as T-consistency for the parameter having the role of ρ for such processes.

In some recent work, Barndorff-Nielsen and Shephard (1999) have used similiar types of models to (1) as components of their continuous time linear stochastic volatility models. In the very simplest case they model the sequence $\{y_1, ..., y_T\}$ as

$$y_t = \eta_t \sqrt{X_t},$$

where $\{\eta_t\}$ and $\{X_t\}$ are totally independent, $\{\eta_t\}$ are independently standard normal distributed and X_t is given by (1). An unfortunate feature of this model is that Bayesian estimation via Markov chain Monte Carlo will typically involve a step which samples from the conditional density $f(\rho|X_0, ..., X_T, \lambda)$ which shares the non-regular properties of the maximum likelihood estimator of (1).

2. Likelihood analysis

In this section the likelihood function of the model given by (1) is analysed. The distribution theory for inference on the autoregression parameter is given. All proofs are collected in the Appendix.

The likelihood function for the model given by (1) is

$$\lambda^{-T} \exp\left\{-\frac{1}{\lambda}\left(\sum_{t=1}^{T} X_t - \rho \sum_{t=1}^{T} X_{t-1}\right)\right\} 1\left(\min_{1 \leq t \leq T} \frac{X_t}{X_{t-1}} \geq \rho\right)$$

with the usual convention that division by zero gives infinity and $1(.)$ being an indicator function. The first term is increasing in ρ and thus for each value of λ the likelihood is maximised with respect to ρ by $\hat{\rho}$ given in (2). It follows that $\hat{\rho} - \rho = \min(\varepsilon_t / X_{t-1})$. The exact and, next, the asymptotic distributions of this estimator are given as follows.

Theorem 1: *The distribution of $\hat{\rho}$ is given by*

$$P\left(\hat{\rho} - \rho > x\right) = \exp\left\{-\frac{x X_0 A_T}{\lambda} - \sum_{t=1}^{T-1} \log\left(1 + x A_t\right)\right\}, \qquad (3)$$

where $x > 0$, X_0 is fixed *and A_t is the polynomial $A_t = \sum_{s=0}^{t-1} (x + \rho)^s$.*

Theorem 2: *The convergence rate for $\hat{\rho}$ is given by*

$$r_T = \begin{cases} T/(1-\rho) & \text{for } 0 \leq \rho < 1, \\ T(T-1)/2 & \text{for } \rho = 1, \\ \rho^T & \text{for } \rho > 1. \end{cases}$$

Further, for the non-explosive cases, $\rho \leq 1$,

$$P\left\{r_T\left(\hat{\rho}-\rho\right) > x\right\} \rightarrow \exp(-x), \quad \text{as} \quad T \rightarrow \infty.$$

The super consistency for the stationary case was proposed for the related non-linear exponential autoregressive models by Raftery (1980,1982), see also An and Huang (1993). It is worth noting that for the non-explosive cases the convergence rate is the square of the Gaussian case whereas it is the same for the explosive case, see White (1958).

Having derived the maximum likelihood estimator for $\hat{\rho}$ it is rather easy to see that the maximum likelihood estimator for λ is

$$\hat{\lambda} = \sum_{t=1}^{T} X_t - \hat{\rho} \sum_{t=1}^{T} X_{t-1},$$

and further that the likelihood ratio test statistic for $\rho = \rho_0$ is given by

$$Q = \left\{1 - (\hat{\rho}-\rho) \frac{\sum_{t=1}^{T} X_{t-1}}{\sum_{t=1}^{T}(X_t - \rho X_{t-1})}\right\}^T. \tag{4}$$

In Gaussian autoregressive models the corresponding test statistic has a non-degenerated limit distribution for all values of ρ_0 although its exact form is different for each of the stationary, the unit root and the explosive cases. This is even nicer for the considered exponential model.

THEOREM 3: *The distribution of the likelihood ratio statistic for $\rho = \rho_0$ given by (4) is asymptotically uniform. Consequently $-\log Q$ is asymptotically exponential and $-2\log Q$ is asymptotically χ^2 with two degrees of freedom.*

3. Discussion

In practice this type of model has some significant disadvantages for the estimator of ρ may be very sensitive to small amounts of mismeasurement. In the stochastic volatility context they are more compelling — for this sensitivity is removed by the addition of measurement noise. However, when we use simple Markov chain Monte Carlo methods to estimate them these issues resurface.

From a theoretical viewpoint some interesting insights are provided by the results in this paper. In addition to those mentioned above it is worth noting that despite having standard properties the least squares estimator is poorly behaved, since it will fall in the range of the parameter space where the likelihood function is exactly zero with probability approaching a half in large samples. Further, the residuals from a least squares fit will often be negative.

In the stationary case, $0 \leq \rho < 1$, the mean of the time series is positive. Thus ρ can be estimated consistently using the demeaned least squares estimator given by $\tilde{\rho} = \sum_{t=1}^{T}(X_{t-1}-\overline{X})X_t / \sum_{t=1}^{T}(X_{t-1}-\overline{X})^2$. Further, the central limit theorem for martingale difference sequences, see Brown (1971), implies that $\{\sum_{t=1}^{T}(X_{t-1}-\overline{X})^2\}^{1/2}(\tilde{\rho}-\rho)$ has a standard normal distribution. On the other hand the likelihood function is zero whenever ρ is larger than the maximum likelihood estimator, $\hat{\rho}$, which is T-consistent. Thus with probability approaching a half the least squares estimator leads to a zero likelihood.

Proof of Theorem 1.
An induction argument is used. Recall that X_0 is fixed. Thus, for $T = 1$ the estimator $\hat{\rho}$ is only well-defined for $X_0 > 0$ and the result follows from the equation

$$P\left(\varepsilon_1 / X_0 > x\right) = P\left(\varepsilon_1 > x X_0\right) = P\left(\varepsilon_1 > x X_0 | X_0\right).$$

For general T, first rewrite $P(\hat{\rho} - \rho > x)$ as

$$E 1\left(\min_{2 \leq t \leq T} \frac{\varepsilon_t}{X_{t-1}} > x\right) 1\left(\varepsilon_1 > x X_0\right).$$

Since X_0 is held fixed this can rewritten as

$$E 1\left(\varepsilon_1 > x X_0\right) E\left\{1\left(\min_{2 \leq t \leq T} \frac{\varepsilon_t}{X_{t-1}} > x\right) \middle| X_1\right\}.$$

The induction assumption, that

$$P\left(\min_{2 \leq t \leq T} \frac{\varepsilon_t}{X_{t-1}} > x \middle| X_1\right) = \exp\left\{-\frac{x X_1 A_{T-1}}{\lambda} - \sum_{t=1}^{T-2} \log\left(1 + x A_t\right)\right\},$$

then implies that $P(\hat{\rho} - \rho > x)$ equals

$$\begin{aligned} & E\left[1\left(\varepsilon_1 > x X_0\right) \exp\left\{-\frac{x X_1 A_{T-1}}{\lambda} - \sum_{t=1}^{T-2} \log\left(1 + x A_t\right)\right\}\right] \\ = \ & \exp\left\{-\frac{x \rho X_0 A_{T-1}}{\lambda} - \sum_{t=1}^{T-2} \log\left(1 + x A_t\right)\right\} E\left\{1\left(\varepsilon_1 > x X_0\right) \exp\left(-\frac{x \varepsilon_1 A_{T-1}}{\lambda}\right)\right\}. \end{aligned}$$

Finally (3) follows by computing the expectation and using the identity $A_T = 1 + (x + \rho) A_{T-1}$.

Proof of Theorem 2.
The case $0 \leq \rho < 1$. Choose $x = z/T$. For sufficiently large T the polynomials A_t are increasing and convergent and hence for $T \to \infty$

$$\sum_{t=1}^{T-2} \log\left(1 + x A_t\right) \approx \frac{z}{T} \sum_{t=1}^{T-2} A_t = \frac{z}{T} \sum_{t=1}^{T-2} \sum_{s=0}^{t-1} \left(\rho + \frac{z}{T}\right)^s \approx \frac{z}{1 - \rho}.$$

It then follows that

$$P\left(T \frac{\hat{\rho} - \rho}{1 - \rho} > z\right) \to \exp\left(-z\right).$$

The case $\rho = 1$. The polynomial A_T is given by $1 + x A_T = (1 + x)^T$ and hence

$$P\left(\hat{\rho} - 1 > x\right) = \left(1 + x\right)^{-T(T-1)/2} \exp\left[-x X_0 \left\{\left(1 + x\right)^T - 1\right\} / \lambda\right].$$

Thus

$$P\left\{\frac{T\left(T - 1\right)}{2} \left(\hat{\rho} - 1\right) > x\right\} \to \exp\left(-x\right).$$

The case $\rho > 1$. The polynomial A_T is of order ρ^T. Thus all T terms in the exponent of (3) are equally important.

Proof of Theorem 3.
In all three case it is used that the denominator, $\sum_{t=1}^T (X_t - \rho X_{t-1})/T = \sum_{t=1}^T \varepsilon_t/T$, in (4) converges to λ by the law of large numbers.
The case $0 \le \rho < 1$. The Law of Large Numbers for linear processes, see Phillips and Solo (1992), implies that the average of the observations $\sum_{t=1}^T X_{t-1}/T$ converges in probability to $\lambda/(1-\rho)$. Thus, $-\log Q$ is asymptotically equivalent to $T(\hat\rho - \rho)/(1-\rho)$.
The case, $\rho = 1$. Theorem 3.27 of Breiman (1968) implies that $\sum_{t=1}^T X_{t-1}$ normalised by $T(T-1)/2$ converges in probability to λ. Thus, $-\log Q$ is asymptotically equivalent to $T(T-1)/2(\hat\rho - 1)$.
For case $\rho > 1$. The theorem follows by proving that

$$T\left\{1 - Q^{1/T}\right\} = (\hat\rho - \rho)\frac{\sum_{t=1}^T X_{t-1}}{\frac{1}{T}\sum_{t=1}^T \varepsilon_t}$$

is asymptotically exponential. Theorem 2 of Lai and Wei (1983) implies that the normalised process $Z_t = \rho^{-t}X_t$ converges almost surely to $Z = X_0 + \sum_{s=1}^\infty \rho^{-j}\varepsilon_j$ which has a continuous distribution. Correspondingly, $(\rho - 1)\sum_{t=1}^T X_{t-1} = \rho^T Z_{T-1} - X_0 - \sum_{t=1}^{T-1}\varepsilon_t$, so that the Strong Law of Large Numbers implies that $\rho^{-T}\sum_{t=1}^T X_{t-1}$ converges almost surely to $Z/(\rho-1)$. Thus it suffices to proved that $H = \rho^T(\hat\rho - \rho)Z/\{\lambda(\rho - 1)\}$ converges in distribution to an exponential. Two cases are considered. First, since Z_{t-1} is an increasing, convergent process, then

$$\begin{aligned} P(H > x) &= E\prod_{t=1}^T 1\left\{\varepsilon_t > x\lambda(\rho-1)\rho^{t-1-T}Z_{t-1}/Z\right\} \\ &\ge E\prod_{t=1}^T 1\left\{\varepsilon_t > x\lambda(\rho-1)\rho^{t-1-T}\right\} \to \exp(-x). \end{aligned} \tag{5}$$

Secondly, since Z_t converges almost surely to Z Egoroff's Theorem implies that for any $\eta_1, \eta_2 > 0$ there exist a set Ω_1 with probability $P(\Omega_1) = 1 - \delta_1$ and a $T_0 > 0$ such that for $\omega \in \Omega_1$ and $t \ge T_0$ then $Z_t/Z > 1 - \delta_2$. Therefore

$$\begin{aligned} P(H > x) &= P(H > x, \Omega_1^c) + P(H > x, \Omega_1) \\ &\le P(\Omega_1^c) + P(H > x, \Omega_1) \\ &= \delta_1 + E\prod_{t=1}^T 1\left\{\varepsilon_t > x\lambda(\rho-1)\rho^{t-1-T}Z_{t-1}/Z\right\}1(\Omega_1) \end{aligned}$$

This expression can be bounded further by

$$\begin{aligned} P(H > x) &\le \delta_1 + E\prod_{t=T_0+1}^T 1\left\{\varepsilon_t > x\lambda(\rho-1)\rho^{t-1-T}(1-\delta_2)\right\} \\ &\to \delta_1 + (1-\delta_1)\exp\{-x(1-\delta_2)\}. \end{aligned} \tag{6}$$

Combination of (5) and (6) implies that $P(H > x) \to \exp(-x)$.

References

An, H.-Z. & Huang, F. (1993). Estimation for regressive and autoregressive models with non-negative residual errors. *J. Time Ser. Anal.* **14**, 179-191.

Anděl, J. (1988). On AR(1) Processes with exponential white noise. *Commun. Statist.-Theory. Meth.* **17**, 1481-1495.

Anděl, J. (1989). Non-negative autoregressive processes. *J. Time Ser. Anal.* **10**, 1-12.

Barndorff-Nielsen, O.E. & Shephard, N. (1999). Non-Gaussian OU based models and some of their uses in financial economics. Unpublished paper, Nuffield College, Oxford.

Bell, C.B. & Smith, E.P. (1986). Inference for non-negative autoregressive schemes. *Commun. Statist.-Theory. Meth.* **15**, 2267-2293.

Breiman, L. (1968). *Probability.* Addison-Wesley, Reading, Massachusetts.

Brown, B.M. (1971). Martingale central limit theorems. *Ann. Math. Stat.* **42**, 59-66.

Lai, T.L. & Wei, C.Z. (1983). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.* **13**, 1-23.

Lawrance, A.J. & Lewis, P.A.W. (1985). Modelling and residual analysis of non-linear autoregressive time series in exponential variables (with discussion). *J. R. Statist. Soc.* B **47**, 165-202.

Matthai, A.M. (1982). Storage capacity of a dam with gamma type inputs. *Ann. Inst. Statist. Math.* **34**, 591-597.

Phillips, P.C.B. & Solo, V. (1992). Asymptotics for linear processes. *Ann. Statist.* **20**, 971-1001.

Raftery, A.E. (1980). Estimation efficace pour un processus autoregressif exponentiel a densite discontinue. *Publ. Inst. Statist. Univ. Paris* **25**, 65-91.

Raftery, A.E. (1982). Generalised non-normal time series models. In O.D. Anderson (ed.). *Time Series Analysis: Theory and Practice 1.* North-Holland.

Sim, C.H. (1992). Point processes with correlated gamma interarrival times. *Statist. Probab. Lett.* **15**, 135-141.

White, J.S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29**, 1188-97.