

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
30 March 1999

SUMMARY

This paper derives the exact distribution of the maximum likelihood estimator of a first order linear autoregression with exponential innovations. We 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; Exponential innovations; Likelihood; Non-regular asymptotics.

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

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad \text{for } t = 1, \dots, T, \quad (1)$$

where 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. More general versions of this type of model have been discussed by, for example, Wold (1948) as a model for a sequence of waiting times. Recently Barndorff-Nielsen and Shephard (1999) have constructed a general continuous time versions of this type of model in order to provide a basis for volatility processes in financial economics.

For $0 \leq \rho < 1$ the process given by (1) can be given a stationary initial distribution. The stationary distribution is non-standard. To over-come this difficulty the related 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. By considering the simpler model with exponential innovations it is possible to demonstrate that the least squares estimator for the autoregression parameter can have rather unfortunate properties despite being consistent and asymptotically normal distributed whereas likelihood inference leads to nicer, but non-standard, properties.

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

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

We will show that the distribution is remarkably simple whatever the true value of ρ . In particular $\hat{\rho}$ is always upward biased. A special case is when $\rho < 1$, for then the estimator will be 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.

One of the implications of our results will be that in large samples, the least squares estimator will fall in the range of the parameter space where the likelihood function is exactly zero with probability approaching a half. Further, the residuals from a least squares fit will often be negative.

It is perhaps not surprising that the distribution theory of the maximum likelihood estimator of the parameters of this model is based on order statistics for it echoes familiar results on the estimation of parameters which control the support of the data. However, that these effects appear in the estimator of the autocorrelation parameter ρ is noteworthy.

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 $1(\cdot)$ 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*

$$P(\hat{\rho} - \rho > x) = \exp \left\{ -\frac{xX_0A_T}{\lambda} - \sum_{t=1}^{T-1} \log(1 + xA_t) \right\}, \quad (3)$$

where $x > 0$ 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, $\rho \leq 1$,

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

The super consistency for the stationary case was proposed for the related non-linear exponential autoregressive models by Raftery (1985). 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

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

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 $-\log LR$ is asymptotically exponential and $-2 \log LR$ 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. However, from a theoretical viewpoint some interesting insights are provided. In addition to those mentioned above it is worth noting that despite having standard properties the least squares estimator is poorly behaved.

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 $\tilde{\rho} = \sum_{t=1}^T (X_{t-1} - \bar{X})X_t / \sum_{t=1}^T (X_{t-1} - \bar{X})^2$. Further, the central limit theorem for martingale difference sequences, see Brown (1971), implies that $\{\sum_{t=1}^T (X_{t-1} - \bar{X})^2\}^{1/2}(\tilde{\rho} - \rho)$ 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.

APPENDIX: PROOFS OF THE THEOREMS

Proof of Theorem 1.

An induction argument is used. For $T = 1$ the estimator $\hat{\rho}$ is only well-defined for $X_0 > 0$ and the result follows from the equation

$$P(\varepsilon_1/X_0 > x) = P(\varepsilon_1 > xX_0) = P(\varepsilon_1 > xX_0|X_0).$$

For general T successive conditioning gives

$$\begin{aligned} P(\hat{\rho} - \rho > x) &= E1\left(\min_{2 \leq t \leq T} \frac{\varepsilon_t}{X_{t-1}} > x\right) 1(\varepsilon_1 > xX_0) \\ &= E1(\varepsilon_1 > xX_0) E\left\{1\left(\min_{2 \leq t \leq T} \frac{\varepsilon_t}{X_{t-1}} > x\right)\middle|X_1\right\}. \end{aligned}$$

The induction assumption, that

$$P\left(\min_{2 \leq t \leq T} \frac{\varepsilon_t}{X_{t-1}} > x \middle| X_1\right) = \exp\left\{-\frac{xX_1A_{T-1}}{\lambda} - \sum_{t=1}^{T-2} \log(1 + xA_t)\right\},$$

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

$$\begin{aligned} &E\left[1(\varepsilon_1 > xX_0) \exp\left\{-\frac{xX_1A_{T-1}}{\lambda} - \sum_{t=1}^{T-2} \log(1 + xA_t)\right\}\right] \\ &= \exp\left\{-\frac{x\rho X_0A_{T-1}}{\lambda} - \sum_{t=1}^{T-2} \log(1 + xA_t)\right\} E\left\{1(\varepsilon_1 > xX_0) \exp\left(-\frac{x\varepsilon_1A_{T-1}}{\lambda}\right)\right\}. \end{aligned}$$

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 $T \rightarrow \infty$

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

It then follows that

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

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

$$P(\hat{\rho} - 1 > x) = (1 + x)^{-T(T-1)/2} \exp\left[-xX_0 \left\{(1 + x)^T - 1\right\}/\lambda\right].$$

Thus

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

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 $0 \leq \rho < 1$. 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, $-\log LR$ 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, $-\log LR$ is asymptotically equivalent to $T(T - 1)/2(\hat{\rho} - 1)$.

For case $\rho > 1$. The theorem follows by proving that

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

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 $Z = X_0 + \sum_{s=1}^{\infty} \rho^{-s} \varepsilon_s$ 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

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

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 $t \geq T_0$ then $Z_t/Z > 1 - \delta_2$. Therefore

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

This expression can be bounded further by

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

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

REFERENCES

- 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.
- 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.
- PHILLIPS, P.C.B. & SOLO, V. (1992). Asymptotics for linear processes. *Ann. Statist.* **20**, 971-1001.
- RAFTERY, A.E. (1985). Discussion of Lawrance and Lewis (1985). *J. R. Statist. Soc. B* **47**, 185-87.
- WHITE, J.S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29**, 1188-97.
- WOLD, H. (1948) On stationary point processes and Markov chains, *Skand. Aktuar.*, **31**, 229-40.