Ann. Inst. Statist. Math.
Vol. 46, No. 2, 309–315 (1994)

AN EFFICIENT ESTIMATOR FOR THE EXPECTATION OF A BOUNDED FUNCTION UNDER THE RESIDUAL DISTRIBUTION OF AN AUTOREGRESSIVE PROCESS

W. WEFELMEYER

Mathematical Institute, University of Cologne, Weyertal 86, 50931 Cologne, Germany

(Received December 18, 1992; revised August 18, 1993)

Abstract. Consider a stationary first-order autoregressive process, with i.i.d. residuals following an unknown mean zero distribution. The customary estimator for the expectation of a bounded function under the residual distribution is the empirical estimator based on the estimated residuals. We show that this estimator is not efficient, and construct a simple efficient estimator. It is adaptive with respect to the autoregression parameter.

Key words and phrases: Autoregressive model, efficient estimator, empirical estimator, residual distribution.

1. Introduction

Let Q be a distribution on the real line with zero mean and positive and finite variance. We observe realizations $X_0, \dots, X_n$ from a stationary autoregressive process $X_i = \vartheta X_{i-1} + \varepsilon_i$, where the ε_i are i.i.d. with unknown distribution Q. A simple, but inefficient estimator for ϑ is the *least squares estimator*

$$\hat{\vartheta}_n = \sum_{i=1}^{n} X_{i-1}X_i \Big/ \sum_{i=1}^{n} X_{i-1}^2.$$

The unobserved residuals ε_i are estimated by $\hat{\varepsilon}_{ni} = X_i - \hat{\vartheta}_n X_{i-1}$. A natural estimator for the expectation $Ef(\varepsilon)$ of a bounded function f is the *empirical estimator* based on the estimated residuals,

$$\hat{E}_n f = n^{-1} \sum_{i=1}^{n} f(\hat{\varepsilon}_{ni}).$$

For fixed $t \in \mathbb{R}$ and $f(x) = I(x \le t)$ we obtain an estimator for the residual distribution function $F(t) = Q(-\infty, t]$,

$$\hat{F}_n(t) = n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}_{ni} \le t).$$

The estimator $\hat{F}_n$ is well studied. For the stationary case, with $|\vartheta| < 1$, a functional central limit theorem is proved by Boldin (1982, 1983). For moving average models see Boldin (1989). He requires a uniformly bounded second derivative of F. His result is generalized by Kreiss (1991) to linear processes with coefficients depending on a finite-dimensional parameter. Koul and Leventhal (1989) treat the explosive autoregressive process, with $|\vartheta| > 1$. They get by with a uniformly bounded first derivative of F. For related results under weak smoothness assumptions on F we refer to Koul (1991, 1992).

Here we restrict attention to the stationary case, with $|\vartheta| < 1$. First we show that $\hat{E}_n f$ is not efficient. This is not due to the fact that an inefficient estimator, the least squares estimator, is used for ϑ. Rather, the reason is that $\hat{E}_n f$ ignores that Q has mean zero. The estimator $\hat{E}_n f$ can be improved by adding noise,

$$E_n^* f = \hat{E}_n f - \hat{a}_n n^{-1} \sum_{i=1}^n \hat{\varepsilon}_{ni} \qquad \text{with}$$

$$\hat{a}_n = \sum_{i=1}^n \hat{\varepsilon}_{ni} f(\hat{\varepsilon}_{ni}) \Big/ \sum_{i=1}^n \hat{\varepsilon}_{ni}^2.$$

Here $\hat{a}_n$ estimates $a = E\varepsilon f(\varepsilon)/E\varepsilon^2$, and $n^{-1}\sum \hat{\varepsilon}_{ni}$ estimates zero and may be considered as noise. The asymptotic distribution of $\hat{E}_n f$ has variance $E(f(\varepsilon) - Ef(\varepsilon))^2$. The asymptotic distribution of $E_n^* f$ has variance

$$E(f(\varepsilon) - Ef(\varepsilon))^2 - (E\varepsilon f(\varepsilon))^2/E\varepsilon^2.$$

This is strictly smaller than $E(f(\varepsilon) - Ef(\varepsilon))^2$ unless ε and $f(\varepsilon)$ are uncorrelated. In particular, an improved estimator for the distribution function $F(t)$ is

$$F_n^*(t) = \hat{F}_n(t) - \hat{a}_n(t) n^{-1} \sum_{i=1}^n \hat{\varepsilon}_{ni} \qquad \text{with}$$

$$\hat{a}_n(t) = \sum_{i=1}^n \hat{\varepsilon}_{ni} I(\hat{\varepsilon}_{ni} \le t) \Big/ \sum_{i=1}^n \hat{\varepsilon}_{ni}^2.$$

It is easy to check that $E_n^* f$ is asymptotically optimal among estimators of the form $\hat{E}_n f - \hat{b}_n n^{-1} \sum_{i=1}^n \hat{\varepsilon}_{ni}$ with $\hat{b}_n$ a consistent estimator of some functional of Q. In Section 2 we prove, for smooth f, that $E_n^* f$ is efficient in the much larger class of *regular* estimators.

If ϑ is known, observing the X_i is equivalent to observing the ε_i. Moreover, one can replace the estimated residuals $\hat{\varepsilon}_{ni}$ by the true residuals ε_i. The problem then reduces to estimating $Ef(\varepsilon)$ efficiently on the basis of i.i.d. observations ε_i with $E\varepsilon = 0$. It can be solved with the methods described in Pfanzagl and Wefelmeyer (1982) and Bickel *et al.* (1993). The main technical difficulty in the present setting lies in showing that observing the $\hat{\varepsilon}_{ni}$ is asymptotically equivalent to observing the ε_i.

Boldin (1982, 1983, 1989), Koul and Leventhal (1989), Koul (1991, 1992) and Kreiss (1991) consider the residual distribution function, i.e. step functions

$f(x) = I(x \leq t)$, and need smoothness of the distribution function. We consider smooth functions f and get by without smoothness of the distribution function. This version is appropriate for the application we have in mind; see Wefelmeyer (1992). Two questions remain open. Under which conditions is our improved estimator $F_n^*(t)$ efficient? Over which classes of functions f does one obtain a functional version of our result?

The present result is an example for efficient estimation in models with certain restrictions. Here the restriction is zero mean of the residual distribution. Such a restriction is also common in regression models. We note that $E_n^* f$ is not a one-step improvement of $\hat{E}_n f$ as discussed, e.g., by Bickel (1982) and Schick (1987) for the i.i.d. case. For a general theory of one-step estimators under restrictions in the i.i.d. case see Koshevnik (1992).

2. Result

First we determine an asymptotic variance bound for estimators of $Ef(\varepsilon)$ when ϑ is *known*. Then we find an estimator which attains the bound for all ϑ. Such an estimator is called *adaptive*. Its existence implies that the variance bound for known ϑ equals the variance bound for unknown ϑ. We expect adaptivity to be symmetric in ϑ and Q. In fact, there exist estimators for ϑ which attain, for all Q, the asymptotic variance bound for the model in which ϑ varies and Q is known; see Kreiss (1987).

Fix ϑ with $|\vartheta| < 1$ and a distribution Q on the real line with zero mean and positive and finite variance. For known ϑ, a local model around Q is introduced as follows. Set

$$H = \{h : \mathbb{R} \to \mathbb{R} \text{ bounded}, \; Eh(\varepsilon) = 0, \; E\varepsilon h(\varepsilon) = 0\}.$$

For $h \in H$ define Q^{nh} by

$$Q^{nh}(dx) = (1 + n^{-1/2} h(x)) Q(dx).$$

Then Q^{nh} is a probability measure for large n, and the residual has again mean zero under Q^{nh},

$$E^{nh}\varepsilon = E\varepsilon + n^{-1/2} E\varepsilon h(\varepsilon) = 0.$$

Let P_n denote the joint distribution of $X_0, \ldots, X_n$ if Q is true. Define P_n^{nh} accordingly. It is easy to check *local asymptotic normality*,

$$\log dP_n^{nh}/dP_n = n^{-1/2} \sum_{i=1}^{n} h(X_i - \vartheta X_{i-1}) - \frac{1}{2} Eh(\varepsilon)^2 + o_{P_n}(1) \qquad \text{and}$$

$$n^{-1/2} \sum_{i=1}^{n} h(X_i - \vartheta X_{i-1}) \Rightarrow N_h \qquad \text{under } P_n,$$

where N_h is normal with mean zero and variance $Eh(\varepsilon)^2$. For details see Huang (1986) or Kreiss (1987).

Now we describe an asymptotic variance bound for estimators of the expectation $Ef(\varepsilon)$ and characterize efficient estimators. The norm $(Eh(\varepsilon)^2)^{1/2}$ in local asymptotic normality provides an inner product $Eh(\varepsilon)k(\varepsilon)$ on $L_2(Q)$. We can express differences $E^{nh}f(\varepsilon) - Ef(\varepsilon)$ in terms of this inner product,

$$n^{1/2}(E^{nh}f(\varepsilon) - Ef(\varepsilon)) = Eh(\varepsilon)f(\varepsilon).$$

The right side is a linear functional on H; the function f is a gradient. Let $\overline{H}$ denote the closure of H in $L_2(Q)$. The gradient f may be replaced by its projection into $\overline{H}$, the *canonical gradient*. We show that the latter is given by the function

$$g(x) = f(x) - Ef(\varepsilon) - ax \tag{2.1}$$

with

$$a = E\varepsilon f(\varepsilon)/E\varepsilon^2.$$

To see this, note that g is a gradient since for $h \in H$,

$$Eh(\varepsilon)g(\varepsilon) = Eh(\varepsilon)f(\varepsilon) - Eh(\varepsilon)Ef(\varepsilon) - aE\varepsilon h(\varepsilon) = Eh(\varepsilon)f(\varepsilon),$$

and that $g \in \overline{H}$ since

$$\begin{aligned} Eg(\varepsilon) &= Ef(\varepsilon) - Ef(\varepsilon) - aE\varepsilon = 0, \\ E\varepsilon g(\varepsilon) &= E\varepsilon f(\varepsilon) - E\varepsilon Ef(\varepsilon) - aE\varepsilon^2 = 0. \end{aligned}$$

An estimator T_n for $Ef(\varepsilon)$ is called *regular* at Q with *limit* L if, for $h \in H$,

$$n^{1/2}(T_n - E^{nh}f(\varepsilon)) \Rightarrow L \qquad \text{under } P_n^{nh}.$$

By the convolution theorem, an asymptotic variance bound for such an estimator is given by the squared length of the canonical gradient. More precisely, $L = M + N$ in distribution, where M is independent of N, and N is normal with mean zero and variance equal to

$$Eg(\varepsilon)^2 = E(f(\varepsilon) - Ef(\varepsilon))^2 - (E\varepsilon f(\varepsilon))^2/E\varepsilon^2.$$

The convolution theorem justifies calling an estimator *efficient* for $Ef(\varepsilon)$ at Q if its limit distribution under P_n is N. It is well known that an estimator T_n for $Ef(\varepsilon)$ is regular and efficient at Q if and only if it is asymptotically linear with influence function equal to the canonical gradient,

$$n^{1/2}(T_n - Ef(\varepsilon)) = n^{-1/2}\sum_{i=1}^{n} g(X_i - \vartheta X_{i-1}) + o_{P_n}(1). \tag{2.2}$$

We use this characterization to prove our main result. A convenient reference for the characterization is Greenwood and Wefelmeyer (1990).

A different local model is introduced in Huang (1986) and Kreiss (1987). It does not take into account that the residuals have mean zero. Hence the condition

$E\varepsilon h(\varepsilon) = 0$ is missing in the definition of H, the relation $E^{nh}\varepsilon = 0$ does not hold, and the local model is not contained in the given model. With this local model the canonical gradient is $f(x) - Ef(\varepsilon)$, and the empirical estimator $\hat{E}_n f$ would be declared efficient. We have seen that it can be improved. The mistake is harmless for estimating ϑ. In fact, because of adaptivity, it suffices then to consider local models with Q *fixed* and ϑ varying. Local asymptotic normality for such models is proved in Akahira (1976) and Akritas and Johnson (1982).

THEOREM 2.1. *Let f be bounded with uniformly continuous and bounded derivative. Then the estimator*

$$E_n^* f = \hat{E}_n f - \hat{a}_n n^{-1} \sum_{i=1}^n \hat{\varepsilon}_{ni}, \qquad \textit{with}$$

$$\hat{a}_n = \sum_{i=1}^n \hat{\varepsilon}_{ni} f(\hat{\varepsilon}_{ni}) \Big/ \sum_{i=1}^n \hat{\varepsilon}_{ni}^2,$$

is regular and efficient for $Ef(\varepsilon)$ at Q.

3. Proof

In this section, all sums extend over i from 1 to n, as before. We must show that (2.2), with g given in (2.1), holds for $T_n = E_n^* f$. This follows if we prove

$$n^{1/2}(\hat{E}_n f - Ef(\varepsilon)) = n^{-1/2} \sum (f(\varepsilon_i) - Ef(\varepsilon)) + o_{P_n}(1), \tag{3.1}$$

$$n^{-1/2} \sum \hat{\varepsilon}_{ni} = n^{-1/2} \sum \varepsilon_i + o_{P_n}(1), \tag{3.2}$$

$$\hat{a}_n = a + o_{P_n}(1). \tag{3.3}$$

Fix $\delta > 0$. Since $\hat{\vartheta}_n$ is $n^{1/2}$-consistent, there exists $c > 0$ such that

$$P_n\{|\hat{\vartheta}_n - \vartheta| > n^{-1/2} c\} \le \delta.$$

Since X_i is square-integrable,

$$P_n \left\{ \max_{0 \le i \le n} |X_i| > \delta c^{-1} n^{1/2} \right\} = o(1).$$

Hence we may restrict attention to $X_0, \ldots, X_n$ with

$$|\hat{\vartheta}_n - \vartheta| \le n^{-1/2} c, \qquad |X_i| \le \delta c^{-1} n^{1/2}.$$

(i) To prove (3.1), choose $\tilde{\varepsilon}_{ni}$ between ε_i and $\hat{\varepsilon}_{ni}$ such that

$$f(\hat{\varepsilon}_{ni}) = f(\varepsilon_i) + (\hat{\varepsilon}_{ni} - \varepsilon_i) f'(\tilde{\varepsilon}_{ni}). \tag{3.4}$$

Write

$$n^{1/2}(\hat{E}_n f - Ef(\varepsilon)) = n^{-1/2}\sum(f(\varepsilon_i) - Ef(\varepsilon)) \\ - n^{1/2}(\hat{\vartheta}_n - \vartheta)n^{-1}\sum X_{i-1}f(\tilde{\varepsilon}_{ni}).$$

The least squares estimator is $n^{1/2}$-consistent, $n^{1/2}(\hat{\vartheta}_n - \vartheta) = O_{P_n}(1)$. Use (3.4) to write

$$n^{-1}\sum X_{i-1}f'(\tilde{\varepsilon}_{ni}) = n^{-1}\sum X_{i-1}f'(\varepsilon_i) + n^{-1}\sum(f'(\tilde{\varepsilon}_{ni}) - f'(\varepsilon_i)).$$

Relation (3.1) follows if this is of order $o_{P_n}(1)$. The second right-hand term is $o_{P_n}(1)$ since f' is uniformly continuous and

$$|\tilde{\varepsilon}_{ni} - \varepsilon_i| \le |\hat{\varepsilon}_{ni} - \varepsilon_i| \le |\hat{\vartheta}_n - \vartheta||X_{i-1}| \le \delta,$$

with δ arbitrarily small. The first right-hand term is $o_{P_n}(1)$ by the ergodic theorem since

$$EX_0 f'(X_1 - \vartheta X_0) = EX_0 Ef'(\varepsilon) = 0.$$

(ii) To prove (3.2), write

$$n^{-1/2}\sum(\hat{\varepsilon}_{ni} - \varepsilon_i) = -n^{1/2}(\hat{\vartheta}_n - \vartheta)n^{-1}\sum X_{i-1}.$$

The least squares estimator is $n^{1/2}$-consistent, and the ergodic theorem and $EX_0 = 0$ imply $n^{-1}\sum X_{i-1} = o_{P_n}(1)$. Hence (3.2) holds.

(iii) To prove (3.3), note that $\hat{a}_n$ is a ratio of two averages. Consider first the numerator. Write

$$(3.5) \qquad n^{-1}\sum(\hat{\varepsilon}_{ni}f(\hat{\varepsilon}_{ni}) - \varepsilon_i f(\varepsilon_i)) \\ = n^{-1}\sum(\hat{\varepsilon}_{ni} - \varepsilon_i)f(\hat{\varepsilon}_{ni}) + n^{-1}\sum\varepsilon_i(f(\hat{\varepsilon}_{ni}) - f(\varepsilon_i)).$$

The function f is bounded, say by c. Hence the first right-hand term in (3.5) is bounded by $c|\hat{\vartheta}_n - \vartheta|n^{-1}\sum|X_{i-1}|$. This is of order $o_{P_n}(1)$ since $\hat{\vartheta}_n$ is consistent and $n^{-1}\sum|X_{i-1}| = E|X_0| + o_{P_n}(1)$ by the ergodic theorem. Use (3.4) to write the second right-hand term in (3.5) as

$$(\hat{\varepsilon}_{ni} - \varepsilon_i)n^{-1}\sum\varepsilon_i f'(\tilde{\varepsilon}_{ni}).$$

The function f' is bounded, say by c. Hence the second right-hand term in (3.5) is bounded by $\delta c n^{-1}\sum|\varepsilon_i|$ and is therefore of order $o_{P_n}(1)$ by the ergodic theorem since δ is arbitrarily small. We obtain that (3.5) is $o_{P_n}(1)$, and by the ergodic theorem,

$$(3.6) \qquad n^{-1}\sum\hat{\varepsilon}_{ni}f(\hat{\varepsilon}_{ni}) = n^{-1}\sum\varepsilon_i f(\varepsilon_i) + o_{P_n}(1) \\ = E\varepsilon f(\varepsilon) + o_{P_n}(1).$$

Consider now the denominator of $\hat{a}_n$. Write

$$n^{-1}\sum(\hat{\varepsilon}_{ni}^2-\varepsilon_i^2)=n^{-1}\sum(\hat{\varepsilon}_{ni}-\varepsilon_i)(\hat{\varepsilon}_{ni}+\varepsilon_i)$$
$$=-(\hat{\vartheta}_n-\vartheta)n^{-1}\sum X_{i-1}(2X_i+(\hat{\vartheta}_n+\vartheta)X_{i-1}).$$

This is bounded by $\delta 2(1+|\vartheta|)n^{-1}\sum|X_{i-1}|$ and hence is of order $o_{P_n}(1)$ by the ergodic theorem since δ is arbitrarily small. Again by the ergodic theorem,

$$n^{-1}\sum\hat{\varepsilon}_{ni}^2=n^{-1}\sum\varepsilon_i^2+o_{P_n}(1)=E\varepsilon^2+o_{P_n}(1). \tag{3.7}$$

Since $E\varepsilon^2$ is positive, relations (3.6) and (3.7) imply (3.3), and we are done.

Acknowledgements

I would like to thank the referees for several corrections and additional references.

References

Akahira, M. (1976). On the asymptotic efficiency of estimators in an autoregressive process, *Ann. Inst. Statist. Math.*, **28**, 35–48.

Akritas, M. G. and Johnson, R. A. (1982). Efficiencies of tests and estimators for p-order autoregressive processes when the error distribution is nonnormal, *Ann. Inst. Statist. Math.*, **34**, 579–589.

Bickel, P. J. (1982). On adaptive estimation, *Ann. Statist.*, **10**, 647–671.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation in Semiparametric Models*, Johns Hopkins University Press, Baltimore.

Boldin, M. V. (1982). Estimation of the distribution of noise in an autoregression scheme, *Theory Probab. Appl.*, **27**, 866–871.

Boldin, M. V. (1983). Testing hypotheses in autoregression schemes by the Kolmogorov and ω^2 criteria, *Soviet Math. Dokl.*, **28**, 550–553.

Boldin, M. V. (1989). On testing hypotheses in the sliding average scheme by the Kolmogorov-Smirnov and ω^2 tests, *Theory Probab. Appl.*, **27**, 866–871.

Greenwood, P. E. and Wefelmeyer, W. (1990). Efficiency of estimators for partially specified filtered models, *Stochastic Process. Appl.*, **36**, 353–370.

Huang, W.-M. (1986). A characterization of limiting distributions of estimators in an autoregressive process, *Ann. Inst. Statist. Math.*, **38**, 137–144.

Koshevnik, Y. (1992). Efficient estimation for restricted nonparametric models, preprint.

Koul, H. L. (1991). A weak convergence result useful in robust autoregression, *J. Statist. Plann. Inference*, **29**, 291–308.

Koul, H. L. (1992). *Weighted Empiricals and Linear Models*, IMS Lecture Notes—Monograph Series, Vol. 21, Hayward, California.

Koul, H. L. and Leventhal, S. (1989). Weak convergence of the residual empirical process in explosive autoregression, *Ann. Statist.*, **17**, 1784–1794.

Kreiss, J.-P. (1987). On adaptive estimation in autoregressive models when there are nuisance parameters, *Statist. Decisions*, **5**, 59–76.

Kreiss, J.-P. (1991). Estimation of the distribution function of noise in stationary processes, *Metrika*, **38**, 285–297.

Pfanzagl, J. and Wefelmeyer, W. (1982). *Contributions to a General Asymptotic Statistical Theory*, *Lecture Notes in Statistics*, **13**, Springer, New York.

Schick, A. (1987). A note on the construction of asymptotically linear estimators, *J. Statist. Plann. Inference*, **16**, 89–105 (correction: *ibid.* (1989). **22**, 269–270).

Wefelmeyer, W. (1992). Quasi-likelihood models and optimal inference, preprint.