

Working Paper 91 - 01
February 1991

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)

A NOTE ON LIKELIHOOD ESTIMATION OF
MISSING VALUES IN TIME SERIES

Daniel Peña and George C. Tiao*

Abstract

Missing values in time series can be treated as unknown parameters and estimated by maximum likelihood, or as random variables and predicted by the expectation of the unknown values given the data. The difference between these two procedures is illustrated by an example. It is argued that the second procedure is, in general, more relevant for estimating missing values in time series.

Key Words

ARIMA models; interpolation; Mean Square Error.

* Daniel Peña is Professor of Statistics, Universidad Carlos III de Madrid and Laboratorio de Estadística, ETSII, Universidad Politécnica de Madrid, and George C. Tiao is W. Allen Wallis Professor of Statistics, Graduate School of Business, University of Chicago.

The use of the maximum likelihood method to estimate missing observations or, in general, unobserved values of random variables is a controversial topic because different authors use different likelihoods to obtain the estimators (see Bayarri, DeGroot and Kadane, 1986 and the discussion of the paper, and Fuller, 1988).

To illustrate the problem in time series models, suppose that the time series $\{z_t\}$ follows the stationary first order autoregressive process

$$z_t = \phi z_{t-1} + a_t, \quad |\phi| < 1,$$

where the a_t 's are i.i.d. $N(0, \sigma^2)$. For simplicity, let us assume that ϕ and σ^2 are known. Suppose that out of n observations z_t , $t = 1, \dots, n$, the observation z_T is missing, $1 \leq T \leq n$. Then, denoting Z_n as the $n \times 1$ vector $Z_n = (z_1, \dots, z_n)'$ and $Z_{(T)}$ as the $(n-1) \times 1$ vector obtained from Z_n by dropping z_T , the joint density function of the available data $Z_{(T)}$ for given z_T is:

$$f(Z_{(T)}|z_T) = \frac{f(Z_n)}{f(z_T)} \quad (1)$$

where:

$$f(Z_n) = (2\pi\sigma^2)^{-\frac{1}{2}n} (1 - \phi^2) \exp\left\{-\frac{1}{2\sigma^2}[(1 - \phi^2)z_1^2 + \sum_{t=2}^n (z_t - \phi z_{t-1})^2]\right\} \quad (2)$$

and

$$f(z_T) = (2\pi\sigma^2)^{-\frac{1}{2}n} (1 - \phi^2) \exp\left\{-\frac{1}{2\sigma^2}(1 - \phi^2)z_T^2\right\}. \quad (3)$$

In (1), z_T is now an unknown parameter of the model for $Z_{(T)}$. The likelihood function of z_T can be written as

$$(z_T|Z_{(T)}) = (2\pi\sigma^2)^{-\frac{1}{2}(n-1)} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{t=1}^{T-1} (z_t - \phi z_{t+1})^2 + \sum_{t=T+1}^n (z_t - \phi z_{t-1})^2\right]\right\} \quad (4)$$

where it is understood that in the exponent a term disappears if the range of summation is not positive.

Hence, the maximum likelihood estimator of z_T takes the form:

$$\hat{z}_T = \begin{cases} \phi^{-1} z_{T+\delta}, & T = 1 \text{ or } n \\ (2\phi)^{-1} (z_{T+1} + z_{T-1}), & 1 < T < n \end{cases} \quad (5)$$

with $\delta = 1$ if $T = 1$ and $\delta = -1$ if $T = n$. It is easy to verify from (4) that the mean square error (MSE) of T is:

$$\text{MSE}(\hat{z}_T) = \begin{cases} \phi^{-2}\sigma^2, & T = 1 \text{ or } n \\ (2\phi^2)^{-1}\sigma^2, & 1 < T < n. \end{cases} \quad (6)$$

Several authors have computed least squares or "maximum likelihood" estimators by maximizing the joint density function $f(Z_n)$ in (2) with respect to missing observations for known data. (See, for instance, Brubacher and Tunnicliffe-Wilson, 1977). Then, in the present case, it is easy to verify that the estimator is given by

$$\hat{z}_T = \begin{cases} \phi z_{T+\delta}, & T = 1 \text{ or } n \\ (1 + \phi^2)^{-1}\phi(z_{T+1} + z_{T-1}), & 1 < T < n \end{cases} \quad (7)$$

However, this estimator \hat{z}_T cannot be called a maximum likelihood estimator because the function $f(Z_n)$ for z_T considered as an unknown parameter is not a joint density function, and, therefore, $f(Z_n)$ for unobserved z_T and known $Z_{(T)}$ is *not* a likelihood function as it is usually defined in standard texts.

To interpret the meaning of (7), let us consider z_T as a random variable following the probabilistic structure in (2). Then, the distribution of z_T given the data $Z_{(T)}$ is:

$$f(z_T|Z_{(T)}) = \frac{f(Z_n)}{f(Z_{(T)})} \quad (8)$$

where $f(Z_{(T)})$ can be obtained by integrating out z_T from $f(Z_n)$. As is well known (see e.g. Peña (1987)) the distribution in (8) is normal with:

$$\begin{aligned} E(z_T|Z_{(T)}) &= \hat{z}_T \\ \text{Var}(z_T|Z_{(T)}) &= \begin{cases} \sigma^2, & T = 1 \text{ or } n \\ (1 + \phi^2)^{-1}\sigma^2, & 1 < T < n. \end{cases} \end{aligned} \quad (9)$$

We see that the conditional expectation, $E(z_T|Z_{(T)})$, is equal to \hat{z}_T in (7); this is because $f(Z_n)$ is proportional to $f(z_T|Z_{(T)})$ and, for the present example, the mode and the mean of the distribution $f(z_T|Z_{(T)})$ are identical. More important, we see that $E(z_T|Z_{(T)})$, which is the minimum MSE estimator of z_T , can be very different from the maximum likelihood estimator \hat{z}_T in (5). Indeed, the MSE of \hat{z}_T in (6) always exceeds, and can be very much larger than, $\text{Var}(z_T|Z_{(T)})$ in (9), which is, of course, also the MSE of the estimator $E(z_T|Z_{(T)})$.

The difference between the two estimators, \tilde{z}_T and \hat{z}_T , is not surprising if we look at the problem from a Bayesian point of view. The estimator \tilde{z}_T is the mean (or mode) of the posterior distribution $f(z_T|Z_{(T)})$ in (8), which is proportional to the product of the likelihood function $(z_T|Z_{(T)})$ in (4) and the prior distribution $f(z_T)$ in (3). On the other hand, the estimator \hat{z}_T can be regarded as the mean (or mode) of a posterior distribution of z_T proportional to the product $(z_T|Z_{(T)})p_0(z_T)$, where $p_0(z_T)$ is a "locally uniform" or noninformative prior distribution (Box and Tiao, 1973). Thus, in the stationary case, $|\phi| < 1$, the two means can be very different because very different prior distributions are employed. This also explains the fact that when ϕ goes to 1 (the model approaches a nonstationary one) the difference between these two estimators goes to zero simply because in this case the prior distribution $f(z_T)$ also becomes nearly locally uniform.

It may be argued from a frequentist point of view that the optimal properties of \hat{z}_T and \tilde{z}_T in (6) and (9) respectively are not really comparable, because they are obtained under very different assumptions. For the maximum likelihood estimator \hat{z}_T the unknown observation z_T is regarded as a fixed parameter, and the $\text{MSE}(\hat{z}_T)$ in (6) is obtained under (or at least motivated by) such an assumption: while for the estimator \tilde{z}_T , the $\text{MSE}(\tilde{z}_T) = \text{Var}(z_T|Z_{(T)})$ in (9) is obtained when z_T is regarded as random following the structure in (2). Indeed, it can be verified from (4) that, for fixed z_T , the MSE of \tilde{z}_T is:

$$\text{MSE}^*(\tilde{z}_T) = \begin{cases} \phi^2\sigma^2 + (1 - \phi^2)^2 z_T^2, & T = 1 \text{ or } n \\ (1 + \phi^2)^{-2} \{2\phi^2\sigma^2 + (1 - \phi^2)^2 z_T^2\}, & 1 < T < n \end{cases} \quad (10)$$

so that for some values of z_T , $\text{MSE}^*(\tilde{z}_T)$ can be larger than $\text{MSE}(\hat{z}_T)$.

The point of the above discussion is to show that in estimating missing values in time series, the method of maximum likelihood can lead to results very different from those obtained by optimal prediction under stationary assumptions. Except for the initial value at $t = 0$, we do not think, however, that it is appropriate to treat missing *observations* as fixed parameters. This seems almost a contradiction in terms. In time series analysis we believe it is natural in most applications to regard the missing observations as random variables following the same probabilistic structure as the remaining ones, and hence adopt the conditional expectation or posterior mean as their optimal estimator.

REFERENCES

- Bayarri, M.J., DeGroot, M.H. and Kadane, J.B. (1986), "What is the likelihood function?", in *Proceedings of the Fourth Purdue Symposium on Statistical Decision Theory and Related Topics*, eds. S.S. Gupta and J. Berger, Springer-Verlag.
- Box, G.E.P. and Tiao, G.C. (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley.
- Brubacher, S.R. y Tunnicliffe-Wilson, G. (1976), "Interpolating Time Series with Application to the Estimation of Holiday Effects on Electricity Demand", *Appl. Statist.*, 25, 2, 107-116.
- Fuller, W.A. (1987), *Measurement Errors Models*, John Wiley and Sons.
- Peña, D. (1987), "Measuring the importance of outliers in ARIMA models", in *New Perspectives in Theoretical and Applied Statistics*, eds. M.L. Puri et al, Wiley, 109-118.