MONASH 15/90

MONASH

UNIVERSITY

AUSTRALIA

MAXIMUM LIKELIHOOD ESTIMATION:

A PREDICTION ERROR APPROACH

Ralph D. Snyder

Working Paper No. 15/90

November 1990

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 86746 971 4

MAXIMUM LIKELIHOOD ESTIMATION:
A PREDICTION ERROR APPROACH

Ralph D. Snyder

Working Paper No. 15/90

November 1990

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS, COMMERCE & MANAGEMENT

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

ABSTRACT

In this paper the problem of computing maximum likelihood estimates of the parameters of linear statistical models is considered. The proposed approach relies on the prediction error decomposition of the likelihood function. A distinctive feature is that the required prediction errors are obtained using conventional linear least squares methods rather than the more usual Kalman filter. More specifically, it is shown that the orthogonalization procedure based on fast Givens transformations, used to obtain the triangular representation of the normal equations, automatically yields the one-step ahead prediction errors and their mean squared errors without additional side calculations.

Keywords: maximum likelihood estimation; fast Givens transformations; Kalman filter; time series analysis.

INTRODUCTION

Linear statistical models often contain parameters which cannot be estimated directly with linear least squares methods. In the case of a linear regression with moving average disturbances, for example, it is necessary to evaluate the likelihood function for a sequence of trial values of the moving average parameters. Often each trial is complicated by the need to invert the associated variance matrix. A common strategy is to convert the model into its dynamic linear form and then to employ the Kalman filter to effectively do this task in conjunction with the so-called prediction error decomposition of the likelihood function (Jones, 1980). In this paper it is shown that identical results can also be obtained with conventional linear least squares methods. More specifically, it is demonstrated that the one-step ahead prediction errors and their mean squared errors required for the evaluation of the likelihood function are automatically obtained as part of the calculations associated with the standard orthogonalization procedure based on fast Givens transformations (Gentleman, 1973).

LINEAR LEAST SQUARES ESTIMATION

The linear statistical model considered in this paper involves a random n-vector x with a diffuse probability distribution. A random m-vector y is dependent on x according to the linear relationship:

$$y = Ax + u, \tag{1a}$$

A being a fixed matrix and u a vector of independent, normally distributed disturbances with zero mean and a diagonal variance matrix V. The vector x cannot be directly observed. So the problem is to utilize the observed value of y together with the relationship (1) to infer a value for x. Given that this value is not unique when $m \geq n$, the problem is supplemented with the quadratic loss function

$$s(x) = (y - Ax)' V^{-1} (y - Ax) \tag{1b}$$

for use as a criterion for estimating x.

A problem of this kind possesses a structure that is not well suited for computing the best estimate of x and so it is a common practice to transform it into into a more amenable form. There are in fact an infinite number of problems which are equivalent to the original problem. Pre-multiplication of (1a) by an arbitrary non-singular matrix P yields a new problem

$$\bar{y} = \bar{A}x + \bar{u} \tag{2a}$$

with $\bar{y} = Py$, $\bar{A} = PA$, $\bar{u} = Pu$ and where the variance of $\bar{u}$ is $\bar{V} = PVP'$. It is readily established

that the new quadratic loss function

$$s(x) = (\bar{y}-\bar{A}x)'\bar{V}^{-1}(\bar{y}-\bar{A}x) \qquad (2b)$$

is equivalent to (1b) and so both problems have the same solution.

A transformation is of little value unless it results in a new problem with a more amenable computational structure. To this end we consider only transformations with the following properties:

(C1) P is orthogonal in the generalized sense that the new covariance matrix $\bar{V}=PVP'$ is diagonal;

(C2) the new matrix $\bar{A}$ is unit, upper trapezoidal ie.

$$\bar{A} = \begin{bmatrix} R \\ O \end{bmatrix}$$

where R is a unit, upper triangular matrix;

(C3) the $(m-n)\times(m-n)$ sub-matrix P_{22} in the partition

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

is unit lower triangular and the elements of P_{21} and P_{22} are independent of m.

The properties C1 and C2 are common to all orthogonalization procedures used to obtain linear least squares estimates and are well known in one form or another eg.

see Ansley (1985). Property C3, and its implications for maximum likelihood estimation to be outlined below, appears to have been ignored.

In the appendix it is shown that the orthogonalization procedure based on fast Givens transformations (Gentleman, 1973), when applied to each row of the data matrix in succession, possesses all three of the above properties. The transformed problem then has the form

$$\begin{aligned}\overline{y}^{[1]} &= Rx + \overline{u}^{[1]} \\ \overline{y}^{[2]} &= \qquad \overline{u}^{[2]}\end{aligned} \tag{3a}$$

involving the components of the partitioned vectors:

$$\overline{y} = \begin{bmatrix} \overline{y}^{[1]} \\ \overline{y}^{[2]} \end{bmatrix} \qquad \overline{u} = \begin{bmatrix} \overline{u}^{[1]} \\ \overline{u}^{[2]} \end{bmatrix}.$$

The associated loss function is

$$s(x) = (\overline{y}^{[1]} - Rx)' \, \overline{V}_1^{-1} \, (\overline{y}^{[1]} - Rx) + \overline{y}^{[2]\prime} \, \overline{V}_2^{-1} \, \overline{y}^{[2]} \tag{3b}$$

where $\overline{V}_1$ and $\overline{V}_2$ are diagonal sub-matrices of V. The best estimate $\hat{x}$ of x must therefore satisfy the convenient triangular equation system

$$R\hat{x} = \bar{y}^{[1]}$$

with a minimum loss of

$$s(\hat{x}) = \bar{y}^{[2]\prime}\bar{V}_2\,\bar{y}^{[2]}.$$

Although these results are well known, matters may be taken further by providing an interpretation for $\bar{y}^{[2]}$ and $\bar{V}_2$. The triangularity condition C3 and the fact that $\bar{y}$ is a transformation of y, imply that

$$\bar{y}_i = y_i + \sum_{j=1}^{i-1} p_{ij} y_j \qquad (i = n+1 \ldots m)$$

Let

$$\hat{y}_i = -\sum_{j=1}^{i-1} p_{ij} y_j \tag{4}$$

Because, by condition C3, the p_{ij} do not change as the sample size m increases, the $\hat{y}_i$ in (4) may be viewed as one-step ahead linear predictors of the y_i and the $\bar{y}_i = y_i - \hat{y}_i$ are then the associated one-step ahead prediction errors. Furthermore, by the orthogonality condition (a), the one-step ahead prediction errors are statistically independent. This is sufficient to guarantee that the $\hat{y}_i$ are minimum mean squared error linear predictors.

LIKELIHOOD FUNCTION

Both A and V potentially depend on a vector of auxiliary parameters θ which must themselves be estimated. Because the distribution of x is undefined, the likelihood function of model (1) does not exist. However, in the equivalent model (3a), the $(m-n)$-vector $\bar{y}^{[2]}$ does not depend on x. Its density is given by

$$f(\bar{y}^{[2]}) = (2\pi \prod_{i=n+1}^{m} \bar{v}_i)^{-1/2} \exp(-\sum_{i=n+1}^{m} \bar{y}_i^2/(2\bar{v}_i))$$

When treated as a function of θ, this forms a marginal likelihood function (Kalbfleisch and Sprott, 1970) and corresponds to the so-called 'prediction error decomposition of the likelihood function' (Schweppe, 1965) normally used in conjunction with the Kalman filter (Jones, 1980) in maximum likelihood applications. Thus the prediction errors and the associated mean squared errors obtained as a by-product of the fast Givens orthogonalization procedure can be used to evaluate the likelihood function for trial values of θ.

CONCLUDING REMARKS

The method outlined in this paper can be applied to any dynamic, linear time series model. Duncan and Horn (1972) have shown that the successive equations of any dynamic linear model can be stacked to give an equivalent linear model of the form (1). Although it may be quite large in size, Page and Saunders (1977) have shown that the resulting problem can be solved efficiently with orthogonalization methods such as Givens transformations which exploit the sparsity of the associated design matrix. Their approach for incorporating prior information may also be adapted to the context of this paper so as to circumvent the limitations of the non-informative prior assumption. Thus, not only can the method be employed instead of the Kalman filter, it may be used to estimate stationary as well as non-stationary time series.

To utilize Givens transformations, it was necessary to assume that the disturbances in the model are statistically independent. This may appear to be unduly restrictive. However, where this condition does not prevail, an equivalent problem satisfying this assumption can always be constructed. The basic strategy is to rewrite the disturbances as a linear function of a vector of zero mean random variables ϵ as follows:

$$u = L\epsilon .$$

One possibility, if necessary, is to use the lower triangular matrix of the Cholesky factorization of V as the matrix L in this relationship. Then the vector ϵ can be absorbed into the vector x to yield a new, equivalent problem with the required structure, despite the fact that the associated disturbances are now identically equal to zero. As such the framework applies to all linear models and is therefore quite general.

APPENDIX

The purpose of this appendix is to show that when fast Givens transformations are used for the linear least squares calculations then the associated transformation matrix satisfies condition C3. The method of proof is inductive, beginning with the assumption that an $m \times m$ matrix P conforming to the three conditions has already been created with fast Givens transformations and applied to an $m \times n$ design matrix A to create an upper trapezoidal matrix according to the relationship

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} A = \begin{bmatrix} R \\ O \end{bmatrix} \tag{A1}$$

where the sub-matrix P_1 has n rows and R is a commensurate unit upper triangular matrix. Now assume that a new row represented by an n-vector a must be appended to the design matrix. The situation may then be summarized by the equation:

$$\begin{bmatrix} P_1 & O \\ P_2 & O \\ O & 1 \end{bmatrix} \begin{bmatrix} A \\ a' \end{bmatrix} = \begin{bmatrix} R \\ O \\ a' \end{bmatrix} \tag{A2}$$

where the matrix on the right hand size can be viewed as a reduced design matrix. It is now necessary to convert this reduced design matrix to unit upper trapezoidal form by eliminating the elements in the new row. This can be done with a series of fast Givens transformations (Gentleman, 1973) which operate only on the the rows of the triangular sub-matrix and the new row of the reduced design matrix. Accordingly, these transformations are equivalent to an application of a matrix with the structure depicted in the following equation

$$\begin{bmatrix} B & O & c \\ O & I & O \\ d' & O & 1 \end{bmatrix} \begin{bmatrix} R \\ O \\ a' \end{bmatrix} = \begin{bmatrix} \tilde{R} \\ O \\ O \end{bmatrix} \tag{A3}$$

B being a square n-dimensional matrix, c and d being n-vectors, and $\tilde{R}$ the new unit triangular matrix. Thus, on combining all the transformations into a single matrix, the following is obtained:

$$P^{new} = \begin{bmatrix} B & O & c \\ O & I & O \\ d' & O & 1 \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} & O \\ P_{21} & P_{22} & O \\ O & O & 1 \end{bmatrix} = \begin{bmatrix} BP_{11} & BP_{12} & c \\ P_{21} & P_{22} & O \\ d'P_{11} & d'P_{12} & 1 \end{bmatrix}$$

From this it follows that

$$P_{21}^{new} = \begin{bmatrix} P_{21} \\ d'P_{11} \end{bmatrix}$$

and

$$P_{22}^{new} = \begin{bmatrix} P_{22} & 0 \\ d'P_{12} & 1 \end{bmatrix}.$$

Apart from the new row, the elements of P_{21} and P_{22} remain unchanged. Furthermore, the new unit column for P_{22} ensures that its unit lower triangular structure is preserved. Both these findings suffice to establish that the condition C3 holds in general.

REFERENCES

Ansley, C.F. (1985), Quick proofs of some regression theorems via the QR algorithm, *The American Statistician*, 39, 55-59.

Duncan, D.B. and S.D. Horn (1972), Linear dynamic recursive estimation from the viewpoint of regression analysis, *Journal of the American Statistical Association*, 67, 815-821.

Gentleman, W.M. (1973), Least squares computations by Givens transformations without square roots, *Inst. Math. Applics.*, 12, 329 - 336.

Jones, R.H. (1980), Maximum likelihood fitting of ARMA models with independent regressors, *Technometrics*, 22, 389-395.

Kalbfleisch, J.D. and D.A. Sprott (1970), Applications of likelihood methods to models involving large numbers of parameters, *Journal of the Royal Statistical Society B*, 32, 175-192.

Page, C.C. and M.A. Saunders (1977), Linear least squares estimation of discrete linear dynamic systems using orthogonal transformations, *SIAM Journal of Numerical Analysis*, 15, 180-193.

Schweppe, F.C. (1965) Evaluation of likelihood functions for Guassian signals, *IEEE Trans. Inform. Theory*, IT-11, 61-70.