GLOBAL JOURNAL OF MATHEMATICAL SCIENCES VOL. 3, NO.1, 2004: 47 - 52
 ISSN 1596- 620

ESTIMATION OF IMPULSE RESPONSE FUNCTION BY ITERATIVELY APPLIED MULTIPLE REGRESSION

E. I. ENANG

(Received 30 January 2003; Revision accepted 21 September 2003)

ABSTRACT

In this paper the author uses iterative multiple regression and backward elimination process to determine the impulse response function coefficients of a given pair of input/output process. The computer-based solution is done with the help of a Pascal program, which organises the selection of the input variables with increasing time lag for the iterative solutions. The truncation point is determined by using the error square contributions of the individual input variables. The method proves stable in both numerical and statistical sense. There was no instability observed up to 80 input variables.

KEYWORDS: Iterative multiple regression, impulse response function, error square contribution, time series analysis, transfer function model

INTRODUCTION

This paper applies iterative multiple regression to identify the impulse response function $\{h_k\}_{k\in N}$ of the system defined by the input process$\{X_t\}_{t\in T}$ and output process $\{Y_t\}_{t\in T}$ described by the equation (1) :

$$Y_t = \sum_{k=0}^{\infty} h_k X_{t-k} + N_t \tag{1}$$

where N_t is uncorrelated with X_s for all $s\leq t$.

Both the input and output data were extracted from the yearly annual reports of the Calabar port of Nigerian Ports Authority covering the period from April, 1978 to December, 1997 (monthly data). The gross throughput (total inward plus outward traffic handled in a port in a month) in metric tones serves as the input process, while the total revenue collected (in Naira) serves as the output, both after appropriate transformation.

Box & Jenkins (1968, 1970) proposed a method for identifying a physical realisable linear system in the time domain, in the presence of added noise. The impulse response function is obtained by writing the least square equations for the impulse response function. According to Box and Jenkins, these equations, which do not in general provide efficient estimates, are cumbersome to solve and also require the knowledge of the truncation point K.

Newland (1975) proposed to revisit Yule-Walker equation and solved it directly since the increased computing power of modern computers removed Box & Jenkins objections. In line with the idea of recommendation, this paper gives least square estimates of the impulse response function. The problems of selecting a finite model is solved by checking the iteratively calculated least square errors as an alternative to the partial autocorrelation values. This therefore, eliminates Box & Jenkins objections about the unknown parameter K, and working with least square estimates may give a stability to the calculations due to their optimality. As a solution, a programme in Turbo Pascal 7, was developed to organise the data input and iterated solutions of the least square problems involved.

This work therefore proposes the estimation of the impulse response function by iteratively applying the least square approach. Here, the orthogonal projections of y_t onto the space spanned by $\{x_{t-s}\}_{0\leq s\leq k}$ for $0\leq k\leq \infty$ is taken after which the backward elimination procedure proposed by Draper & Smith (1968) is applied to the result of the iterative multiple regression procedure to eliminate insignificant lags.

METHODOLOGY

The Iterative Multiple Regression (IMR)

In a general study of contribution of individual variables, one must cover somehow the contribution of all variables.

E. I. ENANG, Department of Mathematics/Statistics and Computer Science, University of Calabar, Calabar, Nigeria.

The results about the correlation functions of stationary processes make it clear that the projection to the past always can be expressed in terms of the first l^*-1 elements of the past, where l^* may be finite or infinite. It is reasonable therefore, to organise the selection of variables into an iterative multiple regression schemes by starting with the first lag, and gradually increasing it. As a measurement, the average residual mean square, s^2, can be used to measure the impact of the variables.. When s^2 does not change in two subsequent solutions, that variable has no contribution to least square estimate of the variable on the right side. This means that either it is linearly dependent with the previous set of variables, or simply, the variable on the right side has no component in the direction represented by this vector. Hence, automatically, a stopping point. Since a real data will not be fully stationary, the trend of s^2 plotted against lag must be checked and where it turns horizontal gives the stopping point. The method however has to be used in combination with the autocorrelation function, or alternative statistical indices like partial autocorrelation function. The reason is that the process dynamism may operate with delayed parameters. In this case, the first b parameters from the past will not appear on the right side at all. Autocorrelation gives information, where the actual contributions of the past start. The partial autocorrelation also shows the nature of this dependency.

These reasoning lead to the following scheme of selecting the variables:

$s^2(1)$	X_{t-1}						
$s^2(2)$	X_{t-1}	X_{t-2}					
$s^2(3)$	X_{t-1}	X_{t-2}	X_{t-3}				
. . .	. .	. .	. .	. .			
. . .	. .	. .	. .	. .	. .		
$s^2(k)$	X_{t-1}	X_{t-2}	X_{t-3}	. .	X_{t-k}	X_{t-k-1}	X_{t-k-2}. .

where the first column represents the average residual mean square, s^2, to the multiple regression solution with the selected variables as indicated in the same line.

The error squares as calculated, with the differenced curve (which describes the contribution of each selected variable in the explanation of the variance of the right side), were represented. The latter advises on one hand where to stop, and on the other, suggests variables with negligible contribution for a possible after selection backward elimination procedure.

RESULTS AND DISCUSSION

The input process is transformed by taking its logarithm to remove the seasonal components after which a first order difference of the transformed input process results in a stationary process as shown in Fig. 1, hereafter, referred to as input process.

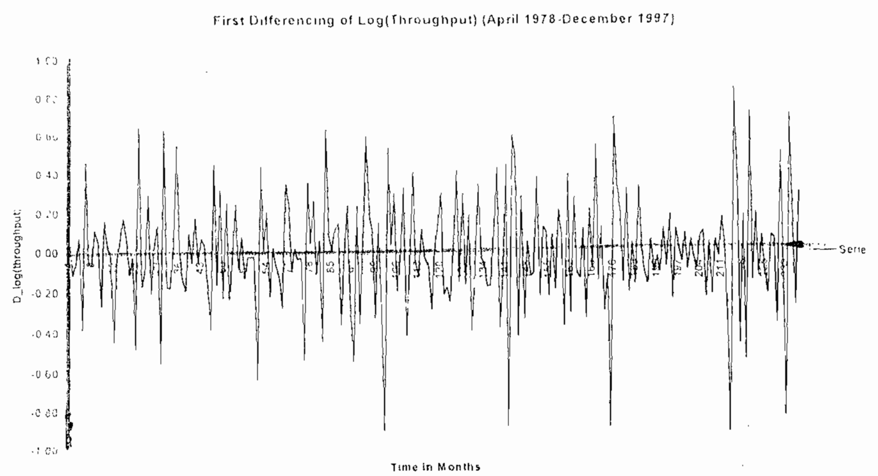

Fig. 1. Time plot of the differenced log input process.

Dividing the monthly revenue with the monthly consumer's price index (using the 1985 as base year) deflates the output series. The seasonal variation is removed from the deflated revenue by taking its logarithm after which a first order differencing of the log (deflated revenue) helps to get rid of the trend. The

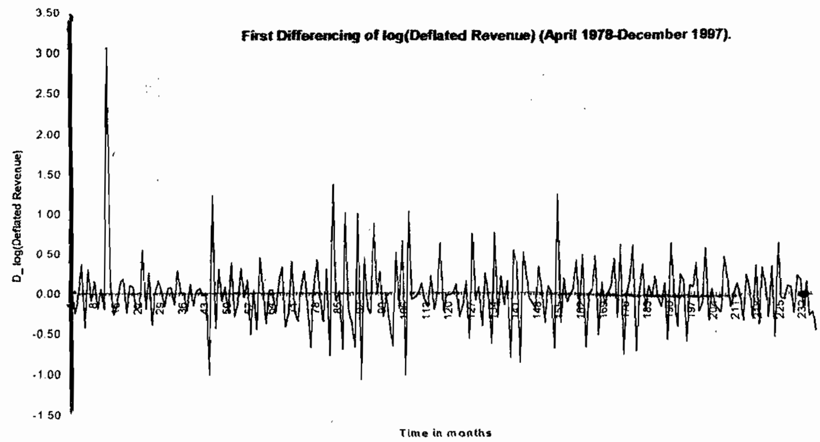

Fig. 2. Time plot of transformed output process.

resulting stationary process is shown in Fig. 2. Hereafter, referred to as the output process.

Impulse Response Function

The impulse response function coefficients (blue) with their error square contribution (red) obtained by direct least square approach are shown in Fig 3. Making use of the backward elimination procedure, it can be observed that only lags 0, 1 and 2 contribute to optimal least squares solution. The contributions from other lags are insignificant and should therefore be ignored.

The transfer function is given here as

$$y_t = 0.49423x_t - 0.031776x_{t-1} + 0.138762x_{t-2} + n_t.$$

where y_t and x_t are the transformed input and output processes respectively while n_t is the corresponding noise process.

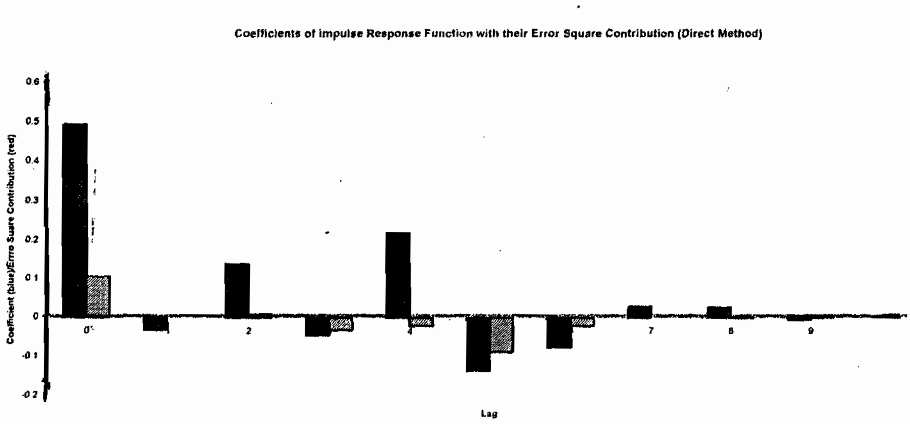

Fig. 3. Coefficients of impulse response function (series 1) with their error square contributions (series 2).

The Noise Process

The autocorrelation function of the noise process and the partial autocorrelation suggest an AR(6) process .

To select the best AR model, the iterative multiple regression procedure was used and the plot of the error squares contribution shows that an AR(2) process described by:

$$n_t = -0.665882n_{t-1} - 0.282361\, n_{t-2} + \varepsilon_t$$

where ε_t is the error in the noise process.

DIAGNOSTIC CHECKING

The residual autocorrelation function obtained with this method shows a very low correlation pattern, which suggests that the noise model is adequate, as confirmed by the Chi- square test of goodness-of-fit.

CONCLUSION

The least square equations were iteratively solved directly with increasing number of input variables for the impulse response function coefficients with the help of a computer. The selection of input variables was organised with the help of a Pascal program. Experience showed that using 233Mhz Pentium MMX processor, the calculations were fast up to 80 variables, and the system did not show any sign of unhealthy symptoms of cumulative numerical errors or other kinds of instability.

The results confirmed the expectations of Newland (1975), and the strategy to determine the truncation point seemed to be successful. The use of error square contributions of individual variables proved to be a good indicator to determine the truncation point.

REFERENCES

Akaike, 1974. A New Look at Statistical Model Identification, I. E. E. C Trans Automatic Control AC-19, pp. 716-723

Box, G. E. P and Jenkins, G. M., 1968. Some recent advances in forecasting and control, part 1, Applied Statistics, vol.17, pp 91-109

Box, G. E. P and Jenkins, G. M., 1970. Time Series Analysis, Forecasting and_Control, San Francisco: Holden Day (Revised edition published 1976)

Chartfield, C., 1979 Inverse autocorrelation. Journal of Royal Statistics Society (A) 142:.363-377

Chartfield, C., 1980 The Analysis of Time Series: An Introduction. (second edition) Chapman and Hall, London.

Draper, N. R and Smith, H., 1966 Applied Regression Analysis. John Wiley & Sons, Inc New York

Kalman, R., Falb, P. and Arbib, M., 1969. Topics in Mathematical System Theory. McGraw-Hill, New York.

Kendal, M..G., and Stuart, A., 1966. The Advanced Theory of Statistics, 3, London: Griffin.

Newland, D. E., 1975. Introduction to Random Vibration and Spectral Analysis. Longman, London.

Olisaoebe, E. U., 1991. Central Bank of Nigerian Economic & Financial Review *29*, (2)

Parsen, E., 1974. Some Recent advances in time Series modelling. Trans. IEEE on Automatic Control, 19: 723-730

Priestly, M. B., 1981. Spectral Analysis and Time Series. Academic press, London.

Shibata, R., 1976. Selection of the order of an AR model by AIC. Biometrica 63: 117-126.

APPENDIX

Iterative multiple regression:

This selects the first k variables for k=1,2,3, .. N, and solves the multiple regression problem with the given right side variable. The module tabulates the solution, the variance of the residuals and the variance of the residuals divided by the variance of the right side variable for each k. The Pascal code of the module is as follows:

```
Procedure iterate(Var Fvlim, Vlim, Llim: integer;  Var eps: real);
                              {Fvlim - the first independent variable to start
                               the analysis with
                               Vlim - last independent variable to use (Fvlim≤Vlim)
                               Llim - the number of lines to be considered
                               Eps -  precision for the variance of the residual, if
                               achieved, the program stops.}
Var
  Ncc, Nll, ll: integer;      {Ncc - actual number of variables used
                               Nll - the actual number of rows considered
                               ll - loop counter}
begin
  Ncc:= Fvlim-1;
  Nll:= Llim;
  stanerr:=1.0;
  clrscr;
  assign(repfil, rawdat+'.REP');
  rewrite(repfil);
  while (Ncc<=Vlim) and (stanerr>=eps) do begin   {The loop stops when Vlim is
                                                    reached}
       inc(ncc);                              {sets the number of variables to consider}
       for ll:=1 to Ncc do notes[ll]:= ll+1;  {Fills the column numbers to be used
                                                into the selection array}
       notes[Ncc+1]:=1;                       {puts the right side into the selection array}
       multiply(NLL, Ncc);             {Calculates A^TA with the selected right side
                                        matrix A}
       solve(Y, AO, Ncc);              {Solves the linear equation A^TAx =A^Ty by Gauss
                                        elimination method}
       standarderr(NLL, NCC);          {calculates statistical indices from residuals}
       Iteriport(Nll, Ncc);            {prepares report to file and as required}
  end;
  flush(repfil);
  close(repfil);
end;
```

The modules of matrix multiplication, implementations of Gauss elimination method, and the statistical and display modules can be made available on request.