Proceedings of the 1985 Winter Simulation Conference
D. Gantz, G. Blais, S. Solomon (eds.)

MICROCOMPUTER BASED INTERACTIVE ANALYSIS OF UNIVARIATE AND MULTIVARIATE ARIMA MODELS

Joseph B. Seif
Lubin Graduate School of Business
Pace University
New York, NY 10038

1. INTRODUCTION

The Box-Jenkins methodology is a model based approach to analyzing and forecasting time-series which has been especially successful in applications to short term forecasting. Many other time-series methods, such as the various forms of exponential smoothing are special cases of this model. The approach is exhaustively described in [1]. It presupposes a form known as the ARIMA model, short for AutoRegressive Integrated Moving Average. In its forecasting form, the forecasted value of the time-series is a finite linear combination of previous observations of the series, and of previous observations of a series of shock terms. The shock terms are completely random and are assumed to embody all the unknown factors that are relevant to the series of interest. As with many other mathematical models, the coefficients must be estimated by non-linear approximation methods. (An important exception is the class of multivariate models including regression, analysis of variance and discriminant analysis whose coefficients can be evaluated as closed form expressions in the data. This feature of these models is at least partly responsible for their great popularity.)

In this paper, the emphasis is mainly on the estimation of the coefficients of ARMA models by various non-linear approximation algorithms. A non-linear approximation method can be simply described as a trial and error approach to minimizing the error in a model. This error is usually taken as the sum-of-squares of the residuals, (the exact for of the error term is actually a function of the probabilistic assumptions about the model. (The residuals can be defined as the difference between the values that would have been forecasted by the model and the past values that were actually observed. In geometric terms this involves finding lowest point on an N-dimensional surface in where N is the number of coefficients to be estimated. For example, if two coefficients are to be estimated, the surface lies in 3-dimensional space with the x-axis and y-axis representing the coefficients and the z-axis representing the sum-of-squares of the residuals. Such methods start with an educated guess for the location of the minimum point and then use information about the shape of the surface to improve the value by finding the direction of steepest descent. The Marquardt algorithm [2] is advocated for estimating ARIMA models in [1] and we use it most often in the case of univariate series. We also use several other estimation methods in our programs .

We introduce another estimation method for ARIMA models based on a general estimation method proposed by Nelder and Mead [4] and brought to the attention of the author by the article [5]. It is based on the concept of a simplex in N-dimensional space named here the *simplicial method* so as not to associate it with the well known simplex method of linear programming. For example, in the case of the above mentioned two dimensional surface, the simplex is a triangle and the direction of descent is determined by a straight line from the highest point through the midpoint between the two other points. Though this may not be the direction of steepest descent, it can generally be expected to be a direction of descent and is far easier to determine than the direction of steepest descent obtained by calculating derivatives.

2. ARIMA Models and Box-Jenkins Method

A time-series is a series of values {x(t)} defined at equally spaced time intervals t = 0,1,2,... A time-series is an autoregressive process of order p, AR(p) if it depends on its previous values according to

$$x(t) = a(1)\,x(t-1) + a(2)\,x(t-2) + \ldots + a(p)x(t-p) + e(t)$$

where {e(t)} is a series of independent identically distributed normal random variates, called the shock series.

{x(t)} is called moving-average of order q, MA(q), if it satisfies the relationship

$$x(t) = e(t) - b(1)\,e(t-1) - b(2)e(t-2) - \ldots - b(q)x(t-q)$$

Combining these two forms gives the ARMA(p,q) model

$$x(t)=a(1)x(t-1) + a(2)x(t-2) + \ldots + a(p)x(t-p) + e(t)-b(1)e(t-1) - b(2)e(t-2) - \ldots - b(q)x(t-q)$$

The ARMA(p,q) models are usually assumed to be stationary, which means, essentially, that one piece of the series has the same behavior as another piece. In particular, the mean and variance of different sections of the series remain constant. This assumption is important for identification

of the series, i.e. determination of (p,q) order of the series. However, the assumption of stationarity excludes most commonly observed series in business and economics, such as any series exhibiting a growth trend, and hence the use of the ARIMA model. If {X(t)} is a series whose differences {X(t) - X(t-1)} or seasonal differences {X(t) - X(t-s)} form a stationary ARMA model, or if several applications of such differencing form an ARMA model, then {X(t)} is called an Integrated ARMA model or an ARIMA model. Actually, the definition of ARIMA is more general than described but this definition is sufficient for most applications. When the resulting ARMA model has been determined, then the original ARIMA model can be recovered for forecasting purposes.

The Box-Jenkins method consists of three stages: (1) identification, in which the series is transformed by differencing and(or other transformations until it is stationary, then an examination of the autocorrelation function and the partial autocorrelation function suggest the (p,q) order of the resulting ARMA series, (2) estimation of the coefficients (or parameters) ARMA series and (3) diagnostic checking of the residuals to ensure that all the structure in the series has been accounted for by the estimated series.

The preceding paragraphs describe the univariate model, however if let {x(t)} denote a time dependent vector, say of dimension m, and a and b above (m x m) matrices then we have described the multivariate ARMA model.

EXAMPLES UNIVARIATE MODELS

We shall first illustrate the Box-Jenkins approach on univariate simulated series. We start by simulating a sequence of independent normally distributed numbers with mean 0 and standard deviation 1 as the series e(t) above. We generate several terms more than the desired length of the x(t) series in order to minimize starting effects. Then using ( ) above we can generate the series x(t). The various types of models can perhaps best be illustrated by three 'canonical' examples, namely, the AR(1), MA(1) and ARMA(1,1) models. A plot of the series, the autocorrelation function (ACF) and partial autocorrelation function (PACF) is shown below for each of these models as realized with a generated series of 100 terms.

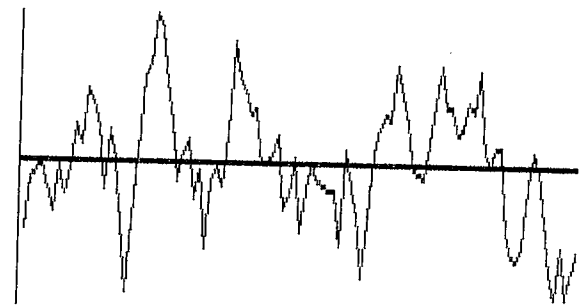

Figure 1. Generated AR(1) series, a(1)=0.8

*** AUTOCORRELATION FUNCTION ***

K		R(K)
0	**********	1
1	*******	.74
2	*****	.49
3	***	.27
4	*	.12
5		-.02
6	*	-.08
7	*	-.07
8		-.04
9		.01
10	*	.05
11	*	.07
12	*	.06
13	*	.05
14	*	.05
15		-.01

Table 1.a

*** PARTIAL AUTOCORRELATION ***

K		P(K)
1	*******	.74
2	*	-.11
3	*	-.11
4		0
5	*	-.14
6		.03
7	*	.09
8		-.01
9		.04
10		.03
11		-.04
12		-.01
13		.04
14		.01
15	*	-.12

Table 1.b

Tables 1 (a,b)
ACF and PACF of AR(1) series

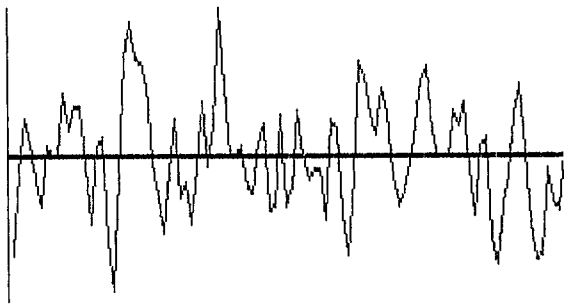

Figure 2. Generated MA(1) series, b(1)=-0.8

Tables 2 (a,b), ACF and PACF of MA(1) series

*** AUTOCORRELATION FUNCTION ***

K	R(K)
0	1
1	.5
2	-.03
3	-.08
4	-.09
5	-.17
6	-.18
7	-.08
8	-.03
9	0
10	.05
11	.06
12	.03
13	.09
14	.14
15	.01

Table 2.a

*** PARTIAL AUTOCORRELATION ***

K	P(K)
1	.5
2	-.37
3	.18
4	-.22
5	-.05
6	-.1
7	.05
8	-.1
9	.06
10	0
11	-.01
12	0
13	.14
14	0
15	-.07

Table 2.b

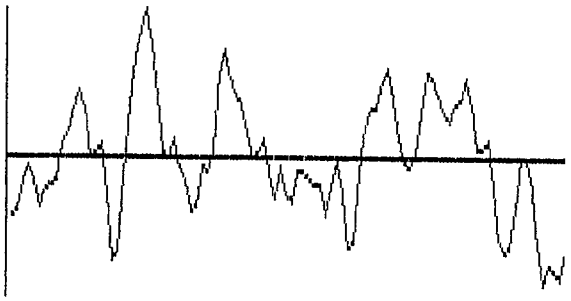

Figure 3. Generated ARMA(1,1) series, a(1)=)=0.8, b(1)=-0.8.

Tables 3 (a,b) ACF and PACF of ARMA(1,1) series

*** AUTOCORRELATION FUNCTION ***

K	R(K)
0	1
1	.84
2	.56
3	.32
4	.14
5	-.01
6	-.08
7	-.08
8	-.04
9	0
10	.04
11	.05
12	.05
13	.05
14	.04
15	0

Table 3.a

*** PARTIAL AUTOCORRELATION ***

K	P(K)
1	.84
2	-.53
3	.22
4	-.24
5	.04
6	.07
7	.03
8	0
9	.04
10	-.03
11	-.02
12	.04
13	.02
14	-.06
15	-.08

Table 3.b

The main tools for identification are the graph of the series, the sample autocorrelation function and the sample partial autocorrelation function.

From the graph of the series certain gross behavior may observed such as non-

stationarity due change in level or trend. Also as we see from the graphs in Figure 1, the AR series seems to have more 'memory' while the MA series looks more like white noise. (The graphs of these three series are quite similar as they were all generated from the same shock series.)

An MA(q) series has the property that theoretical autocorrelation function is zero for lags greater than q, and dually, an AR(p) series has its theoretical partial autocorrelations equal to zero for lags greater than p so that we can seek these properties in the corresponding sample functions to identify these particular types of series. A rough estimate for the standard error of the sample auto- and partial autocorrelations is 1/(square-root of n), where n is the length of the series so that a term is significantly different from zero as the 5% level only if its absolute value is greater than two standard errors i.e. > 2/(sqrt(n)). In practice, most (stationary, non-seasonal) series can be adequately modeled by ARMA(p,q) with p,q <= 2, and these examples are exhaustively discussed and graphed in many texts, e.g. [1], so that a direct comparison with these 'textbook' models can often go a long way toward identifying an adequate model.

Often the identification stage will present more than one possible model. In such cases two or more different models may be estimated and the 'best' one chosen according to criteria which can be examined at the estimation or diagnostic checking stages. Usually, such alternative models will yield barely indistiguishable forecasts.

3. ESTIMATION OF UNIVARIATE MODELS

This is the process of determining the actual coefficients of the model. For the ARMA estimation we discuss three methods: (1) the Marquardt [2] algorithm as presented in Box and Jenkins [1], (2) the simplicial iteration algorithm of Nelder and Mead[3], and (3) the algorithm of Spliid [5]. The Marquardt and simplicial algorithms are maximum likelihood methods. The Marquardt algorithm is extensively described in [1].

The algorithm if Spliid is described in [5] and compared with the Marquardt algorithm. Spliid calls this method "essentially a method of moments, although not in a traditional form." This method is usually even faster than the Marquardt algorithm. Also this method provides its own starting values at the first iteration and may be the best way of providing starting values when a maximum likelihood estimator is desired. Also this method works for the more general situation of multivariate series with exogenous variables and we use this method for the multivariate case.

In this paper, I will briefly describe the simplicial algorithm as, to my knowledge, this method has not been elsewhere discussed in the context of ARMA estimation, and present an example using all three estimation procedures.

MARQUARDT ALGORITHM

Given a tentative model and a trial set of coefficients for that model, the basic equation for the series and the observed data can be used to calculate the residual (or error) terms e(t). The maximum likelihood estimator of the coefficients is equivalent to those obtained by minimizing the sum of squares of the residuals (SSR):

$$SSR = \text{sum}[e(t)^2]$$

The algorithm of Marquardt [2] is propounded in [1] as an efficient non-linear estimation procedure. This method is an efficient estimation procedure whenever the function SSR (as a function of the parameters to be estimated) is sufficiently smooth and approximately quadratic near the minimum. This method also gives the variance-covariance matrix of the estimates so that significance of the coefficients can be assessed.

THE SIMPLICIAL ALGORITHM

The simplicial algorithm has as an advantage the property of being very general in its application so that it can be used for finding a (local) minimum of any function without any knowledge of the form of the derivatives of the function. Also the estimation process does not require storage of large matrices, matrix inversion of solutions of systems of equations. Under the stronger assumption that the response function is approximately quadratic near the minimum, we follow Nelder and Mead in fitting this quadratic (after computation of additional points) to estimate the variance covariance matrix of the parameter estimates, though this process does require matrix inversion.

A motivation for the application of this estimation algorithm is the desire for a self contained tool for the study of perhaps more general time series models than the Box-Jenkins ARIMA models in situations where there is not sufficient smoothness in the dependence on the parameters for the derivative methods to be applicable. A brief description of this method follows:

Let P[0], P[1], ..., P[m] be m+1 points in m-dimensional space, defining a simplex, f(P) the function we wish to minimize, in our case of an ARMA model the sum of squares of residual of the ARMA model with the components of P as its parameters. We arrange the points so that P[0] is the "worst " point i.e. f(P[0]) > f(P[i]) for i = 1,...,m.

Let F(p,q,...) be a function depending on several parameters, in the present case we are considering F as the sum of squares of the residuals of an ARMA model where p,q,... are tentative values of the autoregressive

and moving average parameters. If k parameters are to be fitted we consider a simplex in the k-dimensional parameter space formed by k+1 points. Each of these points has an associated value of F. At least one of these points will have a worst (i.e. largest) value of F. A reasonable and simple choice for for an improvement in this worst value is a point lying in the direction of the straight line through the center of the remaining points of the simplex. This is the basic idea of the algorithm that replaces the "steepest descent" concept of derivative methods. The new point is "better" than the "worst" then it replaces the "worst" point of the simplex and the process is repeated.

If P[center] is the midpoint of the non-worst points i.e.,

P[center]=(P[1]+...+ P[m])/m

then the new reflected point is given by

P[reflected]=P[0]+(1+a)(P[center]- P[0]),

where a is a positive constant known as the reflection coefficient. (If a=1, this corresponds to exact geometrical reflection). If the point obtained by reflection is better than the "best" point of the simplex, then a new point is tried which is yet farther from the worst point along the same straight line at a distance determined by an expansion coefficient. If the reflected point is worse than the worst point of the simplex, then this point is not accepted and we try a contracted new point given by

P[contracted]=P[0]+ b(P[center] - P[0]),

where the reflection coefficient 0 < b < 1. Reflection usually occurs as convergence is approached. If the reflected point is better than the worst, then it replaces the the worst point and the process continues. If the reflected point is worse than the worst then all points of the simplex are shrunk towards the best point i.e. each point, P, other than P[best] is replaced by the midpoint between P and P[best]. This process will always converge to some point which is a local minimum. A Pascal program for the general use of the simplicial algorithm is given in [2].

AN EXAMPLE OF UNIVARIATE ESTIMATION

An ARMA(1,1) series with 200 terms was generated with autoregressive coefficients ar(1)=0.8, and moving average coefficient ma(1)=-0.8 . This series was, estimated on a COMPAQ portable microcomputer with an 8087 math coprocessor.The estimations were performed without back-forecasting. Using the Marquardt algorithm with zero initial parameters convergence was achieved in 5 iterations and 7 seconds. With the simplicial algorithm, convergence was achieved in 18 iterations and 15 seconds.

		s.e.	t-value
ar[1]:	0.6892	0.0527	13.084
ma[2]:	-0.7554	0.0464	-16.295

squared residuals: 206.86
shock variance: 1.04475

Correlation Matrix
1.0000 0.2044
0.2044 1.0000

Table 4.a. Results of estimation with Marquardt algorithm

		s.e.	t
ar[1]:	0.6962	0.0525	13.2546
ma[2]:	-0.7502	0.0650	-11.5424

squared residuals = 206.857
shock variance= 1.045

Correlation Matrix
1.0000 0.3193
0.3193 1.0000

Table 4.b. Results of estimation with simplicial algorithm

Table 3.c Results of estimation with Spliid algorithm

		s.e.	t
ar[1]:	0.6601	0.0781	8.452
ma[2]:	-0.7142	0.0574	-12.432

squared residuals = 207.82
shock variance= 1.049

Correlation Matrix
1.0000 0.4538
0.4538 1.0000

Table 3.Results of estimation with Spliid algorithm

MULTIVARIATE MODELS

For multivariate models we currently restrict ourselves to estimation programs. See [6],[7] for identification techniques and references in them for maximum likelihood estimation. We use the method of Spliid which seems much simpler to program (though we have not actually programmed the maximum likelihood method). Spliid shows in [5] that his method is much faster for large models than maximum likelihood.

We generate an bivariate ARMA(1,1) series of 300 terms from a from a bivariate white noise series with covariance matrix:

$$\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$$

and AR(1) matrix, A, and MA(1) matrix, B:

$$A = \begin{bmatrix} .2 & .3 \\ .6 & -1.1 \end{bmatrix} \qquad B = \begin{bmatrix} -0.2 & -0.3 \\ -0.6 & 1.1 \end{bmatrix}$$

i.e. the vector shock series e(t) consists of two white noise series with variances 4 and 1 respectively, and covariance 1. Such a vector series is generated by computing a square root of the (positive definite) covariance matrix, S, and letting X = S z, where z is an uncorrelated bivariate (normal) series, each component having mean 0 and variance 1.

The series X is generated from

$$X(t)=A(1)X(t-1) + e(t) - B(1)e(t-1).$$

The estimation program, according to Spliid's algorithm, convergence in 7 iterations and gave the following estimates.

Estimated autoregressive matrix:

A(1):	0.14804	0.27480
	-0.66407	1.11103

Estimated moving average matrix:

B(1):	-0.35035	-0.15281
	0.44467	-0.81790

Residual error covariance matrix:

4.1678	0.9359
0.9359	1.0694

ARMA Estimates of generated series.

AN EXAMPLE: HOUSING STARTS AND MORTGAGE RATES

We examine the bivariate monthly series, from January 1976 to April 1985 (112 terms) with the components:

1. housing starts, thousands of units, seasonally adjusted.
2. common mortgage rate as a percent.

Differences are taken for both series and the resulting bivariate series denoted by X:

$$X = \begin{bmatrix} \text{1st difference of housing starts} \\ \text{1st common mortgage rates} \end{bmatrix}$$

Using the Spliid algorithm, we fit this series to a vector ARMA(3,1) model with the following coefficient matrices:

$$A(1) = \begin{bmatrix} -0.27 & -645.04 \\ -0.00 & 0.33 \end{bmatrix}$$

$$A(2) = \begin{bmatrix} -0.81 & -729.18 \\ -0.00 & 1.01 \end{bmatrix}$$

$$A(3) = \begin{bmatrix} -0.43 & 320.93 \\ 0.00 & -0.70 \end{bmatrix}$$

$$B(1) = \begin{bmatrix} -0.02 & -354.62 \\ -0.00 & 0.27 \end{bmatrix}$$

The estimated model is :

$$X(t)=A(1)X(t-1)+A(2)X(t-2)+A(3)X(t-3) + e(t) - B(1)e(t-1) \qquad (2)$$

As X is the differenced series, the original series, Y, can be determined by

$$Y(t) = X(t) + Y(t-1).$$

Also, for prediction purposes, e(t) is unknown and should be set to its expected value of 0, while past values of e can be calculated (2) and the known data.

The algorithm converged in 21 iterations. Though it is not clear whether or not this model is properly identified, an obvious point of interest in this model is that the (2,1) elements are estimated as 0 in all the coefficient matrices. This signifies that past values of Housing Starts have no predictive effect on Common Mortgage rate. Conversly, the large values of the (1,2) element indicate that Common Mortgage rates have a predictive effect on Housing Starts.

REFERENCES

1. Box, G.E.P. and Jenkins, G.M., Time Series Analysis, Holden-Day, Oakland, 1976, 575 pages.

2. Marquardt, D.W., (1963)" An Algorithm for Least-Squares Estimation of Non-Linear Parameters, " Jour. Society for Industrial and Applied Mathematics," 11, 1963, p 431 ff.

3. Nelder, J.A. and Mead, R., " A Simplex Method for Function Minimization," Computer Journal, 7, 1965, 308-313.

4. Caceci, M.S. and Cacheris, W.P., "Fitting Curves to Data," Byte, vol. 9. no. 5, May, 1984, 340-362

5. Spliid, H. "A Fast Estimation Method for the Vector Autoregressive Moving Average Model With Exogenous Variables," Jour. Amer. Stat. Assoc, 78, 384, 843-849.

6. Tiao, G.C. and Box, G.E.P., "Modeling Multiple Time Series With Applications," Jour. Amer. Stat. Assoc, vol. 76, no. 376, Dec., 1981, 802-816.

7. Tiao, G.C. and Tsay, R.S., "Multiple Time Series Modeling and Extended Sample Cross-Correlations," Jour. of Business & Economic Statistics, vol. 1, no. 1, Jan., 1983.

JOSEPH B. SEIF

I am an Associate Professor in the Department of Management Science, Lubin Graduate School of Business at Pace University.

I received a Ph.D. in Mathematics from the State University of New York at Stony Brook in 1972, with specialization in Applied Mathematics and Statistics.

I have published in the areas of differential equations, microcomputer applications and time series. My current interests are in microcomputers, time series analysis and survey analysis.