A Class of Transformations for Box–Jenkins Seasonal Models

By C. F. Ansley, W. A. Spivey and W. J. Wrobleski

The University of Michigan, Ann Arbor, Mich., USA

Summary

A numerical algorithm is developed for estimating the Box–Cox transformation parameter λ in a seasonal ARIMA model, jointly with other model parameters. The algorithm is easily implemented and requires only modest modification of existing Box–Jenkins computer programs. Use of the algorithm is illustrated in analysing a time series of Chatfield and Prothero.

Keywords: FORECASTING; SEASONAL TIME SERIES; BOX–JENKINS ANALYSIS; BOX–COX TRANS-FORMATIONS

1. Introduction

To apply the Box–Jenkins methodology to the analysis and forecasting of time series, a preliminary transformation of the data is often employed. Evidence suggests that the choice of this preliminary transformation is of critical importance, particularly for seasonal time series. Chatfield and Prothero (1973a) used a log transformation for this purpose and did not obtain a satisfactory forecasting model using Box–Jenkins methods. Several discussants of this paper, Box and Jenkins (1973), Harrison (1973) and Tunnicliffe Wilson (1973), suggested that a more flexible parametric family of transformations, introduced by Box and Cox (1964) for the fixed effects analysis of variance model, could be used to improve forecasting performance. This family is given by:

$$\left. \begin{aligned} y_\lambda(t) &= [\{y(t)\}^\lambda - 1]/\lambda, \quad \lambda \neq 0, \\ y_0(t) &= \ln y(t), \end{aligned} \right\} \tag{1}$$

where for our purposes $y(t)$, assumed to be positive, denotes a nonstationary time series and λ is the transformation parameter.

In using a maximum likelihood method for estimating λ, however, computational difficulties were experienced by Chatfield and Prothero (1973b), who indicated that over 40 minutes of computer time were required. This paper presents a numerical algorithm for approximate maximum likelihood estimation of the parameter λ jointly with the other parameters of a seasonal ARIMA model. The algorithm is computationally efficient, is easily implemented, and requires only a modest modification of existing Box–Jenkins computer programs. On the University of Michigan's Amdahl 470 computer, the algorithm requires only about 10 per cent more central processing unit time than Box–Jenkins estimation without the transformation option.

Before developing the algorithm, we give a brief development of the likelihood function associated with the Box–Cox transformation (1) for seasonal ARIMA time series. The use of the algorithm is then illustrated for the seasonal time series of Chatfield and Prothero (1973a).

2. Development of the Likelihood Function

We assume that for some value of the transformation parameter λ the transformed observations $y_\lambda(t)$ in (1) follow a seasonal $\text{ARIMA}(p, d, q) \times (P, D, Q)$ process. The random shocks $a(t)$

are assumed normal with mean zero and variance σ^2. It should be noted that (1) necessarily precludes negative values so that, strictly speaking, $a(t)$ can be only approximately normal. The notation of Box and Jenkins (1970) is adopted throughout this paper.

Since the $y_\lambda(t)$ form an ARIMA $(p, d, q) \times (P, D, Q)$ process, the backward differences,

$$w_\lambda(t) = \nabla^d \nabla_s^D y_\lambda(t) \tag{2}$$

form an ARIMA $(p, q) \times (P, Q)$ process. The joint probability density of $\mathbf{w}_\lambda = (w_\lambda(1), ..., w_\lambda(n))$ is (Box and Jenkins, 1970, p. 273):

$$f_n(\mathbf{w}_\lambda | \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \delta, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}n} |M_n|^{\frac{1}{2}} \exp(-S/2\sigma^2) \tag{3}$$

where

$$S = S(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \delta) = \sum_{t=-\infty}^{n} [E\{a(t) | \mathbf{w}_\lambda; \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \delta\}]^2, \tag{4}$$

and where $\sigma^2 M_n^{-1} = \sigma^2 M_n^{-1}(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Theta})$ is the $n \times n$ covariance matrix of the differenced series $w_\lambda(t)$.

We assume that the initial $d + sD$ observations $y(0), y(-1), ..., y(-d - sD + 1)$ are fixed, and we seek the joint probability density of the n most recent observations $\mathbf{y} = (y(n), ..., y(1))$ of the time series conditional on these values.

Two transformations must be considered in this derivation: the differencing transformation (2), whose Jacobian is one, and the Box–Cox transformation (1), whose Jacobian, assuming the initial $d + sD$ observations are fixed, is

$$J = J_n(\lambda, \mathbf{y}) = \prod_{t=1}^{n} \left| \frac{\partial y_\lambda(t)}{\partial y(t)} \right| = \prod_{t=1}^{n} \{y(t)\}^{\lambda-1}. \tag{5}$$

The conditional joint density of the original (untransformed) time series $y(1), ..., y(n)$ is therefore

$$h_n\{\mathbf{y} | y(0), ..., y(-d-sD+1); \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \delta, \sigma^2, \lambda\} = (2\pi\sigma^2)^{-\frac{1}{2}n} |M_n|^{\frac{1}{2}} \exp(-S/2\sigma^2)J. \tag{6}$$

From (6) the log likelihood is

$$L = \text{const} - \tfrac{1}{2}n \ln \sigma^2 + \tfrac{1}{2} \ln |M_n| - S/2\sigma^2 + \ln J. \tag{7}$$

3. A NUMERICAL ALGORITHM FOR PARAMETER ESTIMATION

Following Box and Jenkins (1970, p. 213) we will assume that for moderate and large values of n, the term $\tfrac{1}{2} \ln |M_n|$ in (7) is dominated by $S/2\sigma^2$. Discarding this term we approximate (7) by

$$L^* = \text{const} - \tfrac{1}{2}n \ln \sigma^2 - S/2\sigma^2 + \ln J, \tag{8}$$

and note that the solution to the normal equation obtained from (8) for σ^2 is

$$\hat{\sigma}^2 = S/n. \tag{9}$$

For fixed values of the remaining parameters we can maximize L^* over σ^2 by substituting (9) into (8), obtaining

$$L^*_{\max} = L^*_{\max}(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \delta, \lambda) = \text{const} - \tfrac{1}{2}n \ln S + \ln J. \tag{10}$$

Maximization of $L^*_{\max}$ remains complicated. Our approach is to re-express (10) as a monotonic function of a sum of squares, thus enabling the use of a nonlinear least-squares algorithm for numerical solution.

We simplify the notation by abbreviating $E\{a(t) | \mathbf{w}_\lambda; \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \delta, \lambda\}$ to $[a(t)]$ and $E\{w_\lambda(t) | \mathbf{w}_\lambda; \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \delta, \lambda\}$ to $[w_\lambda(t)]$. Numerical values for $[a(t)]$ in (4) are obtained from the difference equations

$$\phi(B)\Phi(B^s)[w_\lambda(t)] = \delta + \theta(B)\Theta(B^s)[a(t)]. \tag{11}$$

Following Box and Cox we now define an additional transformation

$$z_\lambda(t) = [w_\lambda(t)]/J^{1/n}, \quad t = 1, \dots, n; \tag{12}$$

outside the range $1, \dots, n$ we write

$$z_\lambda(t) = [w_\lambda(t)]/J^{1/n}. \tag{13}$$

If we now write

$$\hat{a}_z(t) = [a(t)]/J^{1/n}, \tag{14}$$

we can divide (11) by $J^{1/n}$ and obtain

$$\phi(B)\,\Phi(B^s)\,z_\lambda(t) = \delta_z + \theta(B)\,\Theta(B^s)\,\hat{a}_z(t) \tag{15}$$

where

$$\delta_z = \delta/J^{1/n}. \tag{16}$$

Given $z_\lambda(1), \dots, z_\lambda(n)$, it is clear that values of $\hat{a}_z(t)$ are obtained from (15) in exactly the same way that values of $[a(t)]$ are obtained from (11).

Using (14), the sum of squares S can be written

$$S = \sum_{t=-\infty}^{n} [a(t)]^2 = J^{2/n} \sum_{t=-\infty}^{n} \{\hat{a}_z(t)\}^2 = J^{2/n} S_z \tag{17}$$

where

$$S_z = \sum_{t=-\infty}^{n} \{\hat{a}_z(t)\}^2. \tag{18}$$

Finally, using (17) in (10), we obtain

$$L^*_{\max} = \text{const} - \tfrac{1}{2}n \ln(J^{2/n} S_z) + \ln J = \text{const} - \tfrac{1}{2}n \ln S_z. \tag{19}$$

The problem of maximizing (8) is now reduced to that of minimizing S_z in (18) with respect to $\phi, \theta, \Phi, \Theta, \delta_z,$ and λ. From the estimates $\hat{\delta}_z$ and $\hat{\lambda}$ we then obtain

$$\hat{\delta} = J^{1/n}(\hat{\lambda})\,\hat{\delta}_z. \tag{20}$$

We have developed an algorithm which involves the minimization of the nonlinear sum of squares function S_z in (18). Because any computer program for Box–Jenkins estimation will contain both a nonlinear least squares routine and a routine for solution of difference equations of the form (15), our algorithm can be incorporated in such a program with a minimum amount of modification. A flow chart for the major steps involved appears in the appendix.

The Box–Jenkins program used by the authors employs the nonlinear least squares algorithm of Marquardt (1963). This program was adapted as described above to minimize S_z simultaneously over all the parameters including λ, and has proven efficient over a wide range of ARIMA models. Our experience with a variety of time series has indicated that the sum of squares S_z is well behaved and approximately quadratic in λ.

4. AN EXAMPLE

We apply our results to the analysis of the sales data discussed by Chatfield and Prothero (1973a, b), in which monthly data for January 1965 through May 1971 are given. All 77 observations from this time series are used in this illustration.

(a) *Identification.* Identification of (p, d, q) and (P, D, Q) may change for various choices of the parameter λ. An effective strategy is to choose an initial value for λ which seems reasonable from an examination of the raw data, and, based on this initial transformation, to carry out the usual Box–Jenkins identification prodecure to find p, d, q, P, D, Q and initial estimates of the parameters.

Our experience with several time series indicates that the identification of (p, d, q) and (P, D, Q) is not affected by the initial choice of the parameter λ over a wide range of values. Moreover, except for the constant term, initial estimates of other model parameters appear to be relatively insensitive to the choice of λ.

Using an initial logarithmic transformation ($\lambda^0 = 0$), the Chatfield–Prothero series was tentatively identified as an ARIMA $(1, 1, 0) \times (0, 1, 1)$ model, with initial parameter estimates $\phi^0 = -\cdot6$ and $\Theta^0 = \cdot6$. This is the same class of model as that of Model A in Chatfield and Prothero (1973a).

(b) *Estimation.* Approximate maximum likelihood estimates and 95 per cent confidence intervals were obtained using our algorithm and are shown in Table 1. Standard errors were

<p align="center">TABLE 1</p>

<p align="center">*Parameter estimates, standard errors and confidence intervals*</p>

Parameter	Estimate	Standard error	Approximate (2 std error) 95% confidence interval
ϕ	$-\cdot506$	$\cdot110$	$(-\cdot726, -\cdot286)$
Θ	$\cdot799$	$\cdot052$	$(\cdot695, \cdot903)$
λ	$\cdot274$	$\cdot086$	$(\cdot102, \cdot446)$

calculated by inverting the estimated information matrix (see Box and Jenkins, 1970, p. 227). In addition, σ^2 was estimated for this model to be $\cdot490$. Note that for these data the confidence interval for λ does not contain 0.

Fig. 1 gives a plot of the log likelihood L (maximized over ϕ and Θ) against λ. This figure shows L is approximately quadratic near the maximum and that the function is well behaved and unimodal.
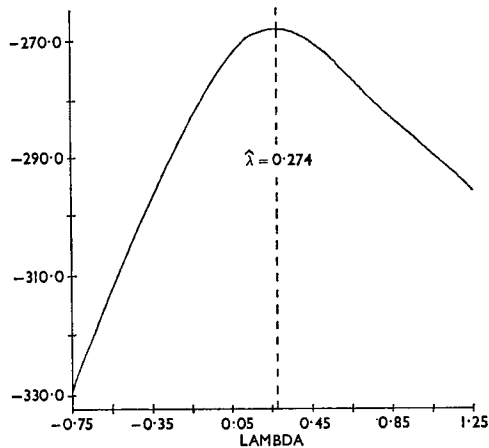

FIG. 1. Log likelihood against λ.

Residual autocorrelations were calculated for lags up to 36 and in only one case (lag 11) exceeded two standard errors. The autocorrelation for lag 24 equalled two standard errors. Cross-correlations between residuals and the series showed no irregularities. The Q statistic for the first 36 autocorrelations was not significant at the $\cdot05$ level.

The value $\hat{\lambda} = \cdot274$ is close to the value $\hat{\lambda} = \cdot24$ obtained by Chatfield and Prothero (1973b). We believe the difference is attributable to slight differences in the likelihood functions used.

Box and Jenkins (1973) and Tunnicliffe Wilson (1973) each suggested that use of the Box–Cox transformation resulted in improved forecasting for this time series. Although we have also noted improvements through its use in forecasting several other time series, the value in practice of the Box–Cox transformation for forecasting time series remains to be established. The ease of implementing the algorithm presented here should facilitate a thorough investigation of this issue.

APPENDIX

Flow Chart for Modification of Estimation Program

The two flow charts in Fig. 2 compare the major steps involved in standard Box–Jenkins estimation and in estimation of the transformation parameter λ simultaneously with other
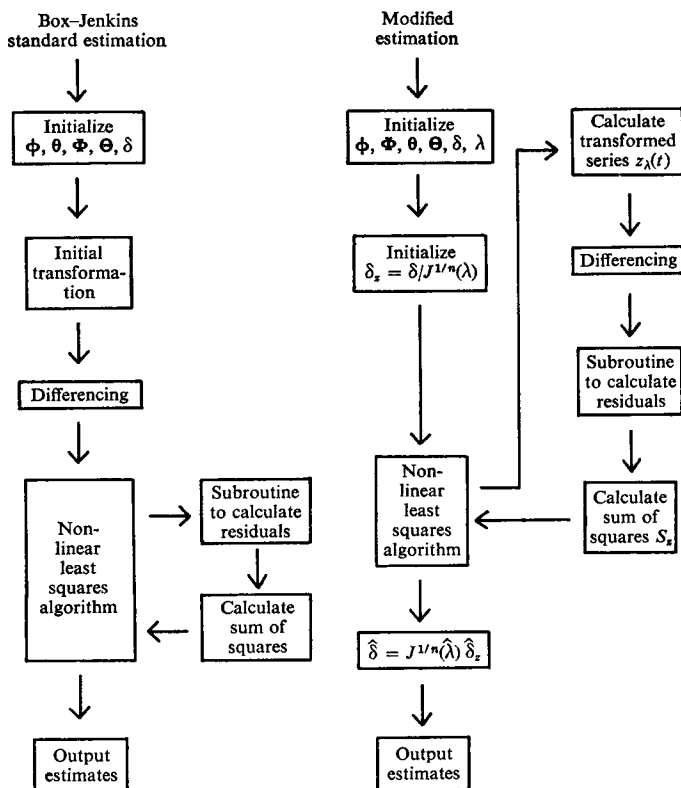

FIG. 2. Flow chart for modification of estimation program.

model parameters using the algorithm developed in this paper. The principal difference is that in estimating λ, the series must be transformed and differenced before each evaluation of the sum of squares within the nonlinear least squares algorithm.

In practice, the modified procedure may be made still more efficient by testing for a change in λ prior to transformation and differencing. The Marquardt algorithm, for example, calculates several values of S_z for a fixed value of λ at each iteration, corresponding to perturbations of the other model parameters. In this case, the previously calculated transformed and differenced series may be used without recalculation.

REFERENCES

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations. *J. R. Statist. Soc.* B, **26**, 211–243.
BOX, G. E. P. and JENKINS, G. M. (1973). Some comments on a paper by Chatfield and Prothero and on a
     review by Kendall. *J. R. Statist. Soc.* A, **136**, 337–345.
—— (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
CHATFIELD, C. and PROTHERO, D. L. (1973a). Box–Jenkins seasonal forecasting: problems in a case study.
     J. R. Statist. Soc. A, **136**, 295–315.
—— (1973b). A reply by Dr Chatfield and Dr Prothero. *J. R. Statist. Soc.* A, **136**, 345–352.
HARRISON, P. J. (1973). Discussion of a paper by Dr Chatfield and Dr Prothero. *J. R. Statist. Soc.* A, **136**,
     319–324.
MARQUARDT, D. W. (1963). An algorithm for least squares estimation of nonlinear parameters. *J. Soc. Ind.
     and Appl. Math.*, **11**, 431–441.
TUNNICLIFFE WILSON, G. (1973). Discussion of a paper by Dr Chatfield and Dr Prothero. *J. R. Statist. Soc.*
     A, **136**, 315–319.