Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER 503834
UNCLASSIFIED	

TITLE

AN ATTRITION FORECASTING MODEL: TECHNICAL SUMMARY OF DATA AND METHODOLOGY

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to: FF

DEPARTMENT OF NATIONAL DEFENCE
CANADA

OPERATIONAL RESEARCH AND ANALYSIS

DIRECTORATE OF SOCIAL AND ECONOMIC ANALYSIS

RESEARCH NOTE 4/94

AN ATTRITION FORECASTING MODEL: TECHNICAL SUMMARY OF DATA AND METHODOLOGY

by

B. SOLOMON

SEPTEMBER, 1994

OTTAWA, CANADA

National Defence Défense nationale

Operational Research and Analysis

CATEGORIES OF PUBLICATION

ORA Reports are the most authoritative and most carefully considered publications issued. They normally embody the results of major research activities or are significant works of lasting value or provide a comprehensive view on major defence research initiatives. ORA Reports are approved by DGOR and are subject to peer review.

ORA Project Reports record the analysis and results of studies conducted for specific sponsors. This category is the main vehicle to report completed research to the sponsors and may also describe a significant milestone in ongoing work. They are approved by DGOR and are subject to peer review. They are released initially to sponsors and may, with sponsor approval, be released to other agencies having an interest in the material.

Directorate Research Notes are issued by directorates. They are intended to outline, develop or document proposals, ideas, analysis or models which do not warrant more formal publication. They may record development work done in support of sponsored projects which could be applied elsewhere in the future. As such they help serve as the corporate scientific memory of the directorate.

DEPARTMENT OF NATIONAL DEFENCE

CANADA

OPERATIONAL RESEARCH AND ANALYSIS

DIRECTORATE OF SOCIAL AND ECONOMIC ANALYSIS

RESEARCH NOTE 4/94

AN ATTRITION FORECASTING MODEL: TECHNICAL SUMMARY OF DATA AND METHODOLOGY

by
B. SOLOMON

OTTAWA, CANADA **SEPTEMBER, 1994**

ABSTRACT / RÉSUMÉ

This note provides a technical discussion on the time series model employed in ORA Project Report 692 "An Attrition Forecasting Model ". Apart from the description of the model, the note also provides a listing of the data and sources used in the attrition study.

RÉSUMÉ

On trouve dans le présent rapport une analyse technique du modèle de série chronologique employé dans le compte rendu du projet 692 de RA Op intitulé "An Attrition Forecasting Model". À part la description du modèle, le rapport fournit aussi une liste des données et des sources utilisées dans l'étude sur l'attrition.

EXECUTIVE SUMMARY

This note is designed to accompany a preceding project report (ORA Project Report 692) on forecasting voluntary attrition in the Canadian Forces. This note includes a detailed discussion on time series modelling and a complete listing of the data and sources employed in the attrition study.

Since the model and the results are discussed fully in Project Report 692, this paper will not repeat the study findings. Instead, a technical summary of the methodology is detailed here to assist other researchers interested in replicating or critically examining the model and methodology. Given the likelihood of a follow up study on attrition, the note may also help as a modelling reference guide.

TABLE OF CONTENTS

INTRODUCTION

1. This note provides a technical discussion on the time series model employed in ORA Project Report 692 "An Attrition Forecasting Model". Apart from the description of the model, the note also provides a listing of the data and sources. The study is sponsored by the Directorate Establishment and Manpower Requirements (DEMR) under ORA activity number 45744.

2. Usually detailed discussions on data and methodology are relegated to the annexes of the main report. However, the sponsor or the sponsoring agency is often interested in the results and implications for policy or decision making process and consequently prefers the report to include as little technical detail as possible. For the analyst who wants to replicate the study or critically examine the results on the other hand, the reverse is required. Thus in order to satisfy both requirements the project is divided into two separate reports.

3. As indicated in Project Report 692, more studies on voluntary attrition are expected and this note is designed to serve as the methodological reference for subsequent studies. The rest of the paper consists of three parts. Section I presents the modelling process of univariate (single variable) Autoregressive Integrated Moving Average (ARIMA) models while section II deals with the multivariate version. In section III selected time series and econometric modelling problems and consequences are summarized. The final section presents the detailed data and sources.

I. MODELLING UNIVARIATE ARIMA PROCESSES.

4. Univariate Box-Jenkins (B-J from here on) is a time series modelling process which describes a single series as a function of its own past values. The purpose of the B-J process is to find the equation that reduces a time series with underlying structure to white noise (Box and Jenkins, 1970). Since the equation accounts for the predictable portion of the time series, it can be used to forecast future values of the series.

5. The modelling procedure itself is a three stage iterative process :

a. Identification: Choosing a tentative model form by examining a plot of the series and several key statistics (such as the autocorrelation and partial autocorrelation functions).

b. Estimation: Fitting an appropriate model through some non-linear estimation procedure to the time series under study.

c. Forecasting: Using the fitted model to predict future values of the time series.

6. The ARIMA models used for forecasting the time series are of the general multiplicative type (Box and Jenkins, 1970), that is:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D Z_t=\theta_q(B)\Theta_Q(B^s)a_t \qquad (1)$$

Where s denotes periodicity of the seasonal component (4 for quarters and 12 for months); B denotes the backward operator, i.e:

$BZ_t = Z_{t-1}$; $B^s Z_t = Z_{t-s}$;

$\nabla^d = (1-B)^d$ is the ordinary difference operator of order d; ${}^D\nabla_s = (1-B^s)^D$ is the seasonal difference operator of order D; $\phi_p(B)$ and $\Phi_p(Bs)$ are stationary autoregressive operators (they are polynomials in B of degree p and in B^s of degree P, respectively); $\theta_q(B)$ and $\Theta_Q(B^s)$ are invertible moving average operators (they are polynomials in B of degree q and in B^s of degree Q, respectively); a_t is a purely random process. The general multiplicative model (1) is said to be of order (p,d,q) $(P,D,Q)_s$.

Model Identification

7. The identification phase entails examining the time series in order to choose a tentative model form. There are several key statistical tools used during this phase. The two most important tools are the autocorrelation function and the partial autocorrelation function of a time series. The first step in identification is to make the series stationary. As explained in the accompanying Project Report 692, in a stationary series the mean and the variance are constant over time. This implies two types of methods for inducing stationarity. Applying the appropriate differencing factor to a series creates a mean stationary series. By applying the correct power transformation, a variance stationary series may be obtained.

8. The autocorrelations of a time series process provide an indication as to the appropriate level of differencing that is required. The need for a power transformation can be ascertained by examining plots of both the original series and the transformed series (Cryer, 1986). If the autocorrelation function starts out high and decays slowly, it usually implies the need for differencing. To determine the order of the differencing, the number of time periods between the relatively high autocorrelation is usually a good indicator (see figure 1 for example).

9. On the other hand if the series shows a variance that changes over time, transforming the original series may provide a stationary variance series.

The transformation can be obtained from a flexible family of transformations introduced by Box and Cox (1964,Vandaele,1983). For a given value of λ, the transformation:

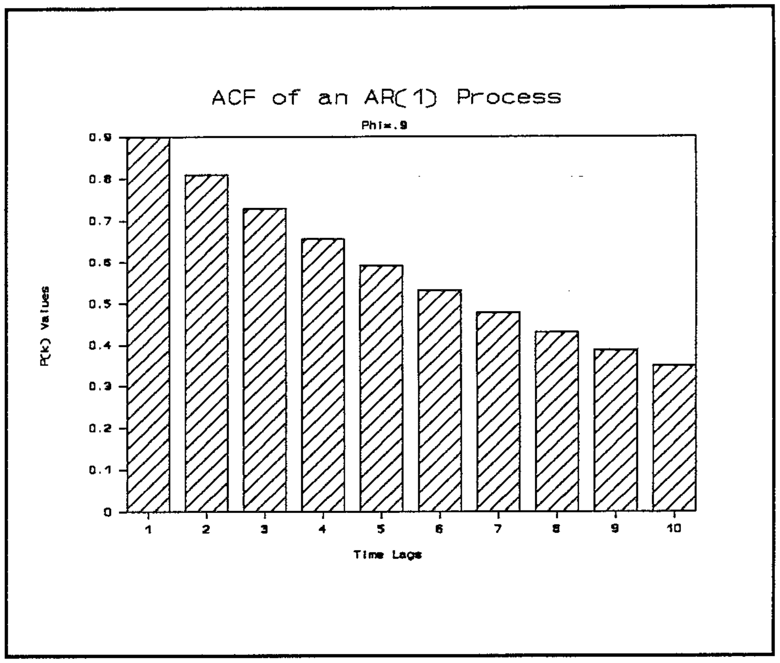

Figure 1 A Typical AR(1) Process

$$g(x)=\frac{x^{\lambda}-1}{\lambda} \quad \lambda\neq 0 \qquad (2)$$
$$=\log(x) \quad \lambda=0$$

note if λ equals 1, it implies the original series, a value of -1 implies the inverse (of the original series) and so on .

10. Box-Cox transformations for forecasting purposes have two uses (Box-Cox,1964). The first is to make a series variance stationary while the other is to assist in determining the relationship between two or more variables (for example, between dependent and explanatory variables). Consider

$$\underset{(\lambda,m)}{Z_t}=\frac{(Z_t+M)^{\lambda}-1}{\lambda} \qquad (3)$$

suppose m=0 then we have the following three cases:

Case a: $\lambda=0$

$$\rightarrow\frac{(Z_t)^0-1}{0}$$
$$\lim_{\lambda\to 0}\frac{\partial[(Z_t)^\lambda-1]}{\partial\lambda}\qquad (3a)$$
$$\mathit{Lim}_{\lambda\to 0}\frac{Z_t^\lambda \ln Z_t}{1}=\ln Z_t$$

(above derived using the L'Hôpital's rule)

Case b: $\lambda= 1$

$$\rightarrow\frac{(Z_t)-1}{1}\qquad (3b)$$
$$=Z_t-1$$

Case c: $\lambda= -1$

$$\rightarrow\frac{(Z_t)^{-1}-1}{-1}\qquad (3c)$$
$$=-(Z_t)^{-1}-1$$

The Box-Cox method is more accurate than just observing the plot of a series. If the type of software allows for Box-Cox test one can estimate a simple mean model with lambda determination option used (AFS, 1986,1990).

11. Once stationarity is obtained, the autocorrelation and partial autocorrelation functions of the transformed and stationary series (ARMA) will be studied so as to

determine the order of the autoregressive and/or moving average process. An estimate of the theoretical autocorrelation function is given by:

$$r_k = \frac{\sum_{t=k+1}^{n} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \tag{4}$$

where $\bar{x}$ is the sample mean. For uncorrelated observations, the variance of r_k is approximately:

$V(r_k) \sim = 1/n$ where n is number of observations.

For the general case, however, the Bartlett (Box-jenkins,1970; 34-35) approximation is used to calculate the standard deviation:

$$s(r_k) = \sqrt{\frac{(1 + 2\sum_{j=1}^{k-1} r_j)}{n}} \qquad \forall r_j > k-1 \tag{4a}$$

The theoretical autocorrelation is also calculated as the ratio of the covariance and variance of the time series at specific lags. The following are three simple ARMA models and their theoretical acf.

CASE 1 AR(1)

$e_t = \phi e_{t-1} + n_t$ where n_t is white noise

$E(n_t) = 0$

$E(n_t)^2 = \sigma^2$

$E(n_{t-i} n_{t-j}) = 0 \ \forall \ i \neq j$

$$e_t = \phi(\phi e_{t-2}+n_{t-1})+n_t$$
$$= \phi^2(\phi e_{t-3}+n_{t-2})+\phi n_{t-1}+n_t$$
$$= \phi^3(\phi e_{t-4}+n_{t-3})+\phi^2 n_{t-2}+\phi n_{t-1}+n_t$$
$$= \ldots$$
$$= \ldots.$$
$$= \ldots.$$
$$= \phi^j e_{t-j} + n_t + \phi n_{t-1} + \phi^2 n_{t-2} + \ldots. + \phi^{j-1} n_{t-j+1},$$

and for $j \to \infty$,

$$e_t = n_t + \phi n_{t-1} + \phi^2 n_{t-2} + \ldots$$

$$E(e_t) = E(n_t+\phi n_{t-1}+\phi^2 n_{t-2}+\ldots) = 0$$

$$E(e_t)^2 = E(n_t+\phi n_{t-1}+\phi^2 n_{t-2}+\ldots)^2$$
$$= E(n_t)^2+\phi^2 E(n_{t-1})^2+\phi^4 E(n_{t-2})^2\ldots$$
$$= \sigma^2 + \phi^2\sigma^2 + \ldots.$$ since $E(n_{t-i}n_{t-j})=0$ all cross terms are ignored.
$$= \sigma^2(1+\phi^2+\phi^4+\ldots)$$

$$E(e_t)^2 = \frac{\sigma^2}{1-\phi^2}$$

$$E(e_t e_{t-1}) = E(n_t + \phi n_{t-1} + \phi^2 n_{t-2}+\ldots)(n_{t-1} + \phi n_{t-2} + \phi^2 n_{t-3}\ldots)$$
$$= \phi\sigma^2 + \phi^3\sigma^2 + \phi^5\sigma^2 + \ldots$$
$$= \phi\sigma^2 (1 + \phi^2 + \phi^4 + \ldots)$$
$$= \frac{\phi\sigma^2}{1-\phi^2}$$

$$\rho_1 = \frac{(e_t e_{t-1})}{(e_t)^2} = \frac{\phi\sigma^2/1-\phi^2}{\sigma^2/1-\phi^2} = \phi$$

$$E(e_t e_{t-k}) = E(n_t + \phi n_{t-1} + \ldots + \phi^k n_{t-k} + \phi^{k+1} n_{t-k+1} \ldots)(n_{t-k} + \phi n_{t-k+1} + \ldots)$$

$$= (\phi^k\sigma^2 + \phi^{k+1}\phi\sigma^2 + \phi^{k+2}\phi^2\sigma^2 + \ldots)$$

$$= \phi^k\sigma^2(1 + \phi^2 + \phi^4 + \ldots)$$

$$= \frac{\phi^k\sigma^2}{1-\phi^2}$$

$$\rho_k = \frac{E(e_t e_{t-k})}{E(e_t)^2} = \frac{\phi^k\sigma^2/1-\phi^2}{\sigma^2/1-\phi^2} = \phi^k$$

CASE 2 MA(1)

$e_t = n_t + \theta n_{t-1}$ (again n_t is white noise)

$E(e_t) = 0$

$$E(e_t)^2 = E(n_t + \theta n_{t-1})^2$$

$$= \sigma^2 + \theta^2\sigma^2 \quad \text{(all cross terms are zero)}$$

$$= \sigma^2 (1 + \theta^2)$$

$$E(e_t e_{t-1}) = E(n_t + \theta n_{t-1})(n_{t-1} + \theta n_{t-2})$$

$$= \theta\sigma^2$$

$$\rho_1 = \frac{(e_t e_{t-1})}{(e_t)^2} = \frac{\theta\sigma^2}{\sigma^2(1+\theta^2)}$$

$$\rho_1 = \frac{\theta}{(1+\theta^2)}$$

$$E(e_t e_{t-k}) = E(n_t + \theta n_{t-1})(n_{t-k} + \theta n_{t-k+1})$$

$$= 0$$

$$\rho_k \quad = 0$$

CASE 3 ARMA(1,1)

$$e_t = \phi e_{t-1} + n_t + \theta n_{t-1}$$

$$= \phi(e_{t-2} + n_{t-1}) + n_t + \theta n_{t-1}$$

$$\phi^2(e_{t-3} + n_{t-2}) + n_t + (\phi+\theta) n_{t-1}$$

....

....

$$= \phi^j e_{t-j} + n_t + (\theta+\phi) n_{t-1} + \phi(\theta+\phi) n_{t-2} + \phi^2(\theta+\phi) n_{t-3} + .. + \phi^{j-2} n_{t-j+1},$$

again for j -> ∞,

$$e_t = n_t + (\theta+\phi) n_{t-1} + \phi(\theta+\phi) n_{t-2} + \phi^2(\theta+\phi) n_{t-3} + ...$$

$$E(e_t) = 0$$

$$E(e_t)^2 = E(n_t + (\theta+\phi) n_{t-1} + \phi(\theta+\phi) n_{t-2} + ...)^2$$

$$= \sigma^2 [\ 1 + (\theta+\phi)^2 + \phi^2(\theta+\phi)^2 + \phi^4(\theta+\phi)^2 + ...]$$

$$= \sigma^2\{\ 1 + (\theta+\phi)^2 [\ 1 + \phi^2 + \phi^4 + \phi^6 + ...]\}$$

$$= \sigma^2 \{\ 1 + [(\theta+\phi)^2/(1-\phi^2)]\}$$

$$E(e_t)^2 = \frac{(1+ \theta^2 + 2\phi\theta)\sigma^2}{1-\phi^2}$$

$$E(e_t e_{t-1}) = E(n_t + (\theta+\phi) n_{t-1} + \phi(\theta+\phi) n_{t-2} + \ldots)(n_{t-1} + (\theta+\phi) n_{t-2} + \ldots)$$

$$= [(\theta+\phi)\sigma^2 + \phi(\theta+\phi)^2\sigma^2 + \phi^3(\theta+\phi)^2\sigma^2 + \ldots]$$

$$= \sigma^2(\theta+\phi)\ [1 + \phi(\theta+\phi) + \phi^3(\theta+\phi) + \ldots]$$

$$= \sigma^2(\theta+\phi)\{1 + \phi(\theta+\phi)[1 + \phi^2 + \phi^4 + \ldots]\}$$

$$= \sigma^2(\theta+\phi)\{1 + [\phi(\theta+\phi)/(1-\phi^2)]\}$$

$$= \frac{(1+\phi\theta)(\phi+\theta)\sigma^2}{1-\phi^2}$$

$$\rho_1 = \frac{(1+\phi\theta)(\phi+\theta)\sigma^2}{1-\phi^2} \cdot \frac{(1-\phi^2)}{(1+\theta^2+2\phi\theta)\sigma^2}$$

$$= \frac{(1+\phi\theta)(\phi+\theta)}{(1+\theta^2+2\phi\theta)}$$

$$E(e_t e_{t-k}) = (\ldots + \phi^{k-1}(\theta+\phi) n_{t-k} + \phi^k(\theta+\phi) n_{t-k+1} \ldots)(\ldots + n_{t-k} + (\phi+\theta) n_{t-k+1} + \phi(\phi+\theta) n_{t-k+2} \ldots)$$

$$= \phi^{k-1}(\theta+\phi)\sigma^2 + \phi^k(\theta+\phi)^2\sigma^2 + \phi^{k+2}(\theta+\phi)^2 + \ldots$$

$$= \phi^{k-1}(\theta+\phi)\sigma^2[1 + \phi(\phi+\theta) + \phi^3(\theta+\phi)^2 + \ldots]$$

$$= \phi^{k-1}(\theta+\phi)\sigma^2\{1 + \phi(\phi+\theta)[1 + \phi^2 + \phi^4 + \ldots]\}$$

$$= \phi^{k-1}(\theta+\phi)\sigma^2\{1 + [\phi(\phi+\theta)/(1-\phi^2)]\}$$

$$= \frac{\sigma^2 \ (\phi^{k-1})\ (1+\theta\phi)\ (\theta+\phi)}{1-\phi^2}$$

$$\rho_k = \frac{\sigma^2 \ (\phi^{k-1})\ (1+\theta\phi)\ (\theta+\phi)}{1-\phi^2} \cdot \frac{1-\phi^2}{(1+\ \theta^2\ +2\phi\theta)\,\sigma^2}$$

$$= \frac{(\phi^{k-1})\ (1+\theta\phi)\ (\theta+\phi)}{(1+\ \theta^2\ +2\phi\theta)}$$

12. The correlation between two random variables, in some cases, is due to the correlation of these two variables to the same third variable. To adjust for this correlation the partial autocorrelation function (pacf) is used. The pacf essentially measures the additional correlations between two lags after adjusting for the intermediate lags.

13. To calculate the sample pacf one can fit autoregressive models of increasing order; the estimate of the last coefficient(ϕ) of each model is the sample pacf.

$$\Phi_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\Phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\Phi}_{k-1,j} r_j} \qquad (5)$$

Model Estimation

14. From the identification stage a tentative ARIMA model is specified for the data generating process on the basis of the estimated autocorrelation and partial autocorrelation (Box and Jenkins, 1970). The following are some possible results:

a. For an MA(q) process the autocorrelation $\rho_k = 0$ for $k>q$ and the partial autocorrelations taper off. To determine a cut off point to the autocorrelation function the sample autocorrelations are used.

b. For an AR(p) the partial autocorrelations $\phi_{kk} = 0$ for $k>p$ and the autocorrelations taper off. A cutoff point of the partial autocorrelation function may be determined by comparing the estimates with T, since $(1/T)^{.5}$ is the approximate standard deviation of the estimators ϕ_{kk} for $k>p$.

c. If neither the autocorrelation nor the partial autocorrelations have a cutoff point, an ARIMA model may be adequate. The AR and the MA degree have to be inferred from the particular pattern of the autocorrelation and partial autocorrelation.

15. Once the identification is completed, a non-linear least squares or a maximum likelihood estimation based on the Marquardt Algorithm, is employed (Box-Jenkins,1970,pp 504 - 505). A pure AR(1) process sometimes known as a "random walk" model can be estimated by linear methods. The non-linear estimation whether by minimizing least squares or maximizing a likelihood function, makes an appropriate computer software necessary to lessen the labour input and computational time.

16. The next step after estimation is diagnostic checking. These tests are for necessity, invertibility and sufficiency. Each parameter included in the model should be statistically significant (necessary) and each factor must be invertible. In addition, the residuals from the estimated models should be white noise (model sufficiency).

17. The test for necessity is performed by examining the T-ratios for the individual parameter estimates. Parameters with non-significant coefficients may be deleted from the model in order to have a parsimonious model. Invertibility is determined by extracting the roots from each factor in the model. All the roots must lie outside of the unit circle.

18. If one of the factors is non-invertible, then the model must be adjusted. The appropriate adjustment is dictated by the type of the factor that is non-invertible. For example, a non-invertible autoregressive factor usually indicates under-differencing, while a non-invertible moving average factor may indicate over differencing. A non-invertible moving average factor could also represent the presence of a deterministic factor. Since the model fixup is not really clear-cut, the overall model must be considered when adjusting for non-invertibility.

To illustrate consider the following examples:

i) $\phi_1 = 0.8 \quad \phi_2 = -0.15$

$$(1 - 0.8B + 0.15B^2) = (1 - .5B)(1 - .3B) = 0$$

The characteristic roots in this case are all greater than one (i.e, they are outside the unit circle) thus the stationarity condition is satisfied.

ii) $\phi_1 = 1.5 \quad \phi_2 = -.5$

$$(1 - 1.5B + 0.5B^2) = (1 - B)(1 - 0.5B)$$

This example has one root at 1 and consequently the stationarity condition is not satisfied[1].

19. The residuals are tested for white noise by studying the autocorrelation and partial autocorrelation of the residuals. Furthermore, a test statistics Q or "portmanteau test" is performed on the residuals autocorrelations of all lags. If the model is misspecified or inappropriately fitted, the Q test tends to be inflated (Box-Jenkins, 1970; Vandaele, 1983).

Recall that $\rho_k = \gamma_k / \gamma_0 \quad \forall k\ 0,1,2,...$

For an ARMA (p,q) process:

$$(\phi)Z_t = (\theta)n_t \text{ thus,}$$

$$[1 + \alpha_1\phi + \alpha_2\phi^2 + ... + \alpha_p\phi^p]Z_t = [1 + \beta_1\theta + \beta_2\theta^2 + ... + \beta_q\theta^q)n_t$$

---

[1]Box and Jenkins (1970) provide the proof for the invertibility condition.

If this process is invertible and the estimates are significant then the residuals are estimated for white noise property:

$\hat{n}_t$ from $[\hat{}(\phi)/\hat{}(\theta)]$

$$r_k = \frac{(\Sigma^T_{t=k+1} \hat{n}_t \hat{n}_{t-k})}{(\Sigma^T_{t=1} \hat{n}_t)^2} \quad \forall\ k\ 1,2,...$$

According to Box and Pierce (1970):

$T\ \Sigma^n_{k=1}\ r_k^2 \sim X^2_{m-p-q}$

Ho: r_k is white noise when n_t is $n \sim (0,\sigma^2)$

Model Forecasting

20. The forecasting function of the general multiplicative model (1) can be expressed in different forms. For computational purpose, the difference equation form is the most useful. Thus, at time t+1 the ARIMA model (1) may be written:

$$Z_{t+v}=\Psi_1 Z_{t+v-1}+...\Psi_m Z_{t+v-m}-a_{t+v}-\Pi_1 a_{t+v-1}-...-\Pi_n a_{t+v-n} \quad (6)$$

where $m=(p+s.P+d+s.D)$ and $n=(q+s.Q)$; $\psi(B)=\phi_p(B^s)\nabla^d\nabla^D_s$ is the general autoregressive operator; $\pi(B)=\theta_q(B)\Theta_Q(B)$ is the general moving average operator. For example, if the ARIMA model is of order $(0,1,1)(0,1,1)_{12}$ the difference equation that generates the observations Z_{t+v}, is:

$$Z_{t+v}=Z_{t+v-1}+Z_{t+v-12}-Z_{t+v-13}+a_{t-v}-\theta a_{t+v-1}-\Theta a_{t+v-12}-\theta\Theta a_{t+v-13} \quad (6a)$$

Standing at origin t, to make a forecast $\hat{Z}_t(v)$ of Z_{t+v}, the conditional expectation of (6) is taken at time t with the following assumptions:

$E_t(Z_{t+j}) = Z_{t+j}, j \leq 0$; $E_t(Z_{t+j}) = Z\hat{}_t(j), j>0$

$E_t(a_{t+j}) = a_{t+j}, j \leq 0$; $E_t(a_{t+j}) = 0, j>0$

where $E_t(Z_{t+j})$ is the conditional expectation of Z_{t+j} taken at origin t. Thus, the forecasts $Z_t(v)$ for each lead time are computed from previous observed Z's, previous forecasts of Z's and current and previous random shocks a's. The unknown a's are replaced by zeroes.

21. In general, if the moving average operator $\pi(B) = \theta(B)\Theta(B^s)$is of degree (q+s.Q), the forecast equations for $Z_t(1),Z_t(2),\ldots,Z_t(q+s.Q)$ will depend directly on the a's but forecasts at longer lead times will not. The latter will receive indirectly, the impact of the a's by means of the previous forecasts. In effect, $Z_t(q+s.Q+1)$ will depend on the (q+s.Q) previous Z_t which in turn will depend on the a's.

22. From the view point of studying the nature of the forecasts, it is important to consider the explicit form of the forecasting function. For $v>n=(q+s.Q)$, the conditional expectation of (6) at time t is:

$$Z\hat{}_t(v)-\psi_1 Z\hat{}_t(v-1)-\ldots-\psi_m Z_t(v-m) = 0 \quad v>m$$

and the solution of this difference equation is:

$$Z\hat{}_t(v) = b_0(t)f_0(v) + b_1^{(t)}f_1(v) + \ldots + b_{m-1}^{(t)}f_{m-1}^{(v)} \quad v>n-m$$

This function is called the "eventual forecast function", eventual because when n>m, it supplies the forecasts only for lead times l>n-m.

23. In the above representation, $f_0(V)$, $f_1(V)$, $\ldots f_{m-1}(V)$ are functions of the lead time V and in general they include polynomials, exponentials, sines and cosines, and products of these functions. For a given origin t, the coefficients $b_j^{(t)}$ are constants applying for all lead time V but they change from one origin to the next, adapting themselves to the particular part of the series being considered. It is important to point out that it is the general autoregressive operator $\psi(B)$ defined above , which determines the mathematical form of the forecasts function, i.e., the

nature of the f's. In other words, it determines whether the forecasting function is to be a polynomial, a mixture of sines and cosines, a mixture of exponentials or some combinations of these functions.

Integration

24. The original variable Z_t and a differenced variable W_t are linked deterministically by the differencing operator $(1\text{-}B)^d$.

$$W_t = (1\text{-}B)^d Z_t$$

This relationship between Z_t and W_t is very important because, after building an ARIMA model for the stationary series W_t, we often want to forecast the original nonstationary series Z_t.

Suppose d = 1 then,

$$Z_t = (1\text{-}B)^{-1} W_t$$

$(1\text{-}B)^{-1}$ can be written as infinite series:

$$(1 + B + B^2 + B^3 + \dots)$$

$$Z_t = (1 + B + B^2 + B^3 + \dots) W_t$$

$$= W_t + W_{t-1} + W_{t-2} + \dots$$

$$= {}^t\Sigma_{i=\infty}\, W_i$$

Since the Z's are sums of the W's, we can get to Z by integrating.

II. MULTIVARIATE TIME SERIES MODELLING

25. The multivariate time series model is also a three stage iterative operation of identification, estimation (and diagnostic check) and forecasting. In multivariate time series models the independent variable (a.k.a. input series) is **prewhitened** by fitting a univariate ARIMA model. Similarly the dependent variable (output series) is also "prewhitened" by fitting the same AR and MA factors as the input series.

26. While in a univariate time series model the autocorrelation and partial autocorrelation functions determine the appropriate factors for estimation, in a multivariate model the crosscorrelation function is used. This function determines the interrelationship between the input and output series. A typical multivariate time series model is specified symbolically as:

$$\nabla^{y} Y_t = \alpha_0 + \frac{\gamma_1(B)}{\delta_1(B)}(x_{t-b1})\nabla^{x1} + .. + \frac{\theta(B)}{\phi(B)} A_t^{*} \qquad (12)$$

where:

- Y_t is the dependent variable at time t and ∇^{y} is the ordinary difference operator,
- α_0 denotes the deterministic trend,
- X_t is the independent input series and γ, δ represent the numerator and denominator factor(s) of the independent series, particularly, γ represents polynomial lagged independent variables while δ represents lagged polynomial of the dependent variable,
- B denotes the backward operator, i.e $BZ_t = Z_{t-1}$; $B^{s}Z_t = Z_{t-s}$ while b represents the pure delay or lag on Xs,

A_t is a random process with $\phi(B)$ and $\Phi(B)$ representing the stationary autoregressive and moving average (polynomials in B) operators. The '*' indicates that the error term may need to be differenced to make it stationary.

27. For a single input variable equation, equation (12) can be rewritten as:

$$\frac{\phi_x(B)}{\theta_x(B)}Y_t=\frac{w(B)\phi_x(B)}{\theta_x(B)}X_t+\frac{\theta_x(B)}{\phi_x(B)}n_t \tag{13}$$

or denoting the left hand side as β_t and the input series as α_t

$$\beta_t=w(t)\alpha_t+\epsilon_t$$

Taking the expectation operator across and assuming α and ϵ are independent:

$$\begin{gathered} E[\alpha_{t-k}\beta_t]=w_0E(\alpha_{t-k}\alpha_t)+w_1E(\alpha_{t-k}\alpha_{t-1}+\ldots+E(\alpha_{t-k}\epsilon_t) \\ C_{\alpha\beta}(k)=w_kC_{\alpha\alpha}(t-k)+0 \\ w_k=\frac{C_{\alpha\beta}(k)}{C_{\alpha\alpha}(t-k)} \end{gathered} \tag{14}$$

Thus by substituting sample values, the impulse response weight can be derived as the crosscorrelation between α and β multiplied by the standard deviation of the β series and divided by the standard deviation of the α series.

28. Like the univariate case, it is possible to use the Box-Pierce Chi-square test to determine if the set of autocorrelations from the residuals are significantly different from zero. The degrees of freedom for the test depend on the values of the time lag in the independent series, the parameter associated with the impact of

the independent series on the dependent and finally the parameter associated with the impact of past history of the dependent variable on itself.

Outlier Detection

29. Time series data are often influenced by isolated events such as strikes, earthquakes and etc. Often such events are modelled as "dummy" variables with a series of zeros and one(s) for the time period(s) of the isolated event. Since most of these events may or may not be identifiable to the modeller, a theory-free detection algorithm can also be applied. The algorithm begins by first fitting a univariate ARIMA model to the series divided by a series of regressions at each time period to test the hypothesis that there is an intervention. After identifying an outlier the residuals are modified and the test will resume until all outliers are uncovered.

30. The automatic outlier detection algorithm is obviously better than the theory-based method for cases when the modeller does not have apriori knowledge of when the event may have occurred. In addition, if the outlier causes a structural shift (such as institutional legislations) and if the impact of the event takes a longer time lag to affect the behaviour, then the modeller may bias the effect of the outlier by choosing the day the event occurred as the break in the series. On the other hand, theory-free methods have the "potential for finding spurious significance" during the testing of the hypothesis (AFS, 1986;1990).

31. Some forecasters (see AFS manual, 1990) argue that fitting a univariate ARIMA model does not necessarily imply that the series is homogeneous (the error term of the series is random about a constant mean). In such situations, the identification of the outliers become dubious due to the recursive process. The suggested solution is the testing of a more rigorous specification, i.e., the mean of of the errors must be near zero for all time sections (AFS, 1990).

III. SOME NOTES ON ECONOMETRIC AND TIME SERIES MODELS.

32. The main advantages of econometric and multivariate time series (MTS) models are their ability to deal with interdependencies. Often econometric models are used to assess the impact of various policy scenarios on aggregate economic variables such as the GDP and investment. Likewise a system of equations designed to explain voluntary attrition may also provide a good simulation exercise on various economic and military policies on attrition.

33. However, if the objective is forecasting rather than explaining, such complicated models may not provide as good a forecast given the time and resources cost of building such models. Furthermore, the specification and identification of a multi-variable system entails systematic errors that may be difficult to detect. As summarized in Solomon (1991), various economists and econometricians have criticised large scale econometric models as restrictive (when setting coefficients),
arbitrary (during the sorting of variables into exogenous and endogenous categories), and limited (when specifying and testing the orders of serial correlation and cross serial correlation of the disturbance terms).

34. For example, the test statistics used to examine serial correlations of the error terms are for first-order autocorrelation, a common but not necessarily the only order of autocorrelation found in economic time series. As shown in Solomon (1991), temporal aggregation can cause higher order autoregressive or moving average processes on any time series with white noise or first order error component due to aggregation of data into larger time intervals. For illustration assume a simple one equation with a lagged dependent variable and a white noise random error.

$$Y_t = \alpha + \beta y_{t-1} + e_t$$

$e_t = n_t$ (white noise)

Substituting the error term

$$Y_t = \beta^2 Y_{t-2} + (\alpha + \alpha\beta) + (n_t + \beta n_{t-1})$$

$$Y_{2,t} = Y_t + y_{t-1} \qquad \text{for } t=2, 4, 6 \ldots$$

$$= \beta^2(Y_{t-2} + Y_{t-3}) + 2\alpha(1+\beta) + n_t + (1+\beta)n_{t-1} + \beta n_{t-2}$$

$$= \beta^2 Y_{2,t-2} + 2\alpha(1+\beta) + e_{2,t}$$

$$e_{2,t} = n_t + (1+\beta)n_{t-1} + \beta n_{t-2}$$

$$E(e^2_{2,t}) = \sigma^2[1 + (1+\beta)^2 + \beta^2]$$

$$E(e_{2,t}e_{2,t-2}) = \beta\sigma^2$$

$$E(e_{2,t}e_{2,t-s}) = 0 \qquad \text{for } s = 2, 3, 4, \ldots$$

Apart from a second order moving average error term, the coefficient of the lagged dependent variable is squared. If the theoretical specification of the lagged variable was negative initially, the temporal aggregation may incorrectly specify a positive relationship instead.

35. In multivariate time series models the dynamic relationship between dependent and independent variables are examined more fully than is available in most regression based econometric models. Furthermore, unlike the univariate time series models, multivariate models can provide us with information regarding the effect of certain policy shock such as government expenditure on the forecasts beyond the sample period. Although the MTS class of models are quite broad, they are restricted to stationary time series or those series that exhibit stationarity after differencing. Since the prediction of future values are constrained to be linear functions of the observations, we also have to assume linearity of the models. Such restriction, often sufficient approximation of reality, does not cover every aspect of real life situation and thus a broader class of models should be entertained.

36. The assumption of linearity is often relaxed in most case studies by applying the Box-Cox transformation to the model. Since stationarity is an important factor in MTS modelling, increasing trends are usually differenced. Such trends may be caused by systematic components that consequently, should be explained appropriately through some explicit economic theory.

37. While the statistical tools such as the autocorrelation, crosscorrelation and partial autocorrelation functions help the researcher in choosing a tentative model for estimation and forecasting, most of the time he or she has to rely on personal (subjective) judgements. Unless the researcher has extensive knowledge and experience in time series modelling, such judgements can be inaccurate and may lead to inferior forecast values. It is desirable, therefore, for the researcher to seek expert opinion from other researchers who have done more work on that particular time series or related data. Furthermore, fitting a series of simple ARIMA and seasonal ARIMA models and selecting the one with desirable statistics is a good starting point if the identification stage gives inadequate information.

38. As verified by the preceding and other studies, time series models produce relatively superior forecast values than competing regression-based models. Subsequently applications of time series models have extended into econometric modelling realms as substitutes for large econometric models (such as Vector Autoregressive models) or as a complement to regression models through the modelling of the stochastic term (Pindyck and Rubenfeld,1981). Both of these applications require extensive theoretical exposition and empirical verification to include in this paper, thus, interested readers are referred to Pindyck and Rubenfeld (1981), Newbold (1983) and Priestley (1988) for a detailed discussion.

IV. DATA SUMMARY

39. The following is the listing of the variables presented in the subsequent pages.:

Year	Calendar year
ur	The unemployment rate
vaoff	Voluntary attrition (officers)
vancm	Voluntary attrition (NCMs)
RGDP	Real (constant $) GDP
DRGDP	Growth in real GDP
RGDPMAN	Real GDP in the manufacturing sector
RGDPCOM	Real GDP in the communication sector
RGDPBSR	Real GDP in the business service sector
EMPMAN	Employment in the manufacturing sector
EMPTCOM	Employment in the communication and transportation sector
EMPSERV	Employment in the service sector
indprd	Industrial production
cur	Capacity utilization rates (manufacturing)
awhm	Average weekly hours (manufacturing)
lincm	Labour income (manufacturing)
linct	Labour income growth
ur1524	Unemployment rate (15-24 year olds)

ENGINEE	Engineering group (officers)	GENERAL	Generals
		OPERATI	Operational (officers)
OTHERS	Other occupations (officers)	SUPPORT	Support (Officers)
SPECIALI	Specialists (officers)		

AIR OPS	Air operations (NCMs)	AIR TECH	Air technicians (NCMs)
COMBAT	Combat pers. (NCMs)	DENTAL	Dental assistants (NCMs)
COMMUNI	Communications (NCMs)	LEME	Land electrical and Mechanical eng.(NCMs)
LOGISTIC	Logistics (NCMs)	MEDICAL	Medical related (NCMs)
MIL-ENG	Military Engineer (NCMs)	OTHERS	Other Occ. (NCMs)
SEA OPS	Sea operations (NCMs)	SUPPORT	Support (NCMs)

Source:

Economic data from the CANSIM database, Statistics Canada (951 8116)
Attrition statistics from DEMR

Year	ur	vaoff	vancm	RGDP	DRGDP	RGDPMAN	RGDPCOM
1973	5.5	76	516	326848	.	67865	5295
1974	5.3	751	5190	341235	4.40174	69907	5980
1975	6.9	675	4973	350113	2.601726	65250	6748
1976	7.1	493	3857	371688	6.162296	70037	7430
1977	8.1	518	3698	385122	3.614322	72578	7884
1978	8.3	448	3327	402737	4.573875	75884	8638
1979	7.4	474	4027	418328	3.871261	78731	9486
1980	7.5	502	4432	424537	1.484242	75203	10379
1981	7.5	555	4019	440127	3.672236	77972	11186
1982	11	375	2417	425970	-3.21657	67921	11177
1983	11.8	255	1445	439448	3.164073	72311	11460
1984	11.2	330	2000	467167	6.307686	81622	12016
1985	10.5	349	2415	489437	4.767032	86218	12700
1986	9.5	391	2311	505666	3.315851	86849	13311
1987	8.8	445	2746	526730	4.165595	91025	14204
1988	7.8	542	3344	552958	4.979401	95643	15299
1989	7.5	607	3756	566486	2.446479	95830	16964
1990	8.1	569	3150	565576	-0.16064	90947	18287
1991	10.3	464	2636	556029	-1.68801	84929	19025
1992	11.3	432	1811	560048	0.722804	85362	19464

Year	RGDPBSR	EMPMAN	EMPTCOM	EMPSERV	indprd	cur	awhm
1973	64994	1927	775	2290	95741.6	86.8	39.56
1974	68844	1978	791	2389	97602.7	85.8	38.83
1975	71494	1871	812	2520	90428.3	77.5	38.56
1976	74681	1921	824	2573	96505.6	80.6	38.65
1977	76700	1888	819	2695	99750.2	81.6	38.65
1978	79507	1956	859	2812	103213.5	84.2	38.69
1979	81232	2071	903	2954	108195.9	85.4	38.79
1980	85589	2111	906	3096	104513.6	79.9	38.5
1981	89960	2124	911	3262	106673.8	79.9	38.56
1982	89843	1928	882	3273	96204.4	68.2	37.6
1983	90105	1879	865	3395	102435.6	71.6	38.24
1984	93784	1954	852	3458	114882.7	79.7	38.4
1985	96303	1960	876	3630	121272.9	83.1	38.62
1986	100363	1989	891	3765	120356.4	82.2	38.41
1987	104254	2018	899	3918	126226	83.3	38.71
1988	109588	2104	904	4062	132918.4	83.1	38.91
1989	113709	2126	961	4150	132729.4	81.1	38.68
1990	115853	2001	951	4299	128551.3	77.3	38.23
1991	114454	1865	916	4376	123847.8	73.7	37.82
1992	114401	1788	922	4408	124356.1	74.8	38.25

Year	lincm	linct	ur1524	ENGINEE	GENERAL	OPERATI	OTHERS
1973	15669	15.3	9.6	16	1	35	5
1974	18211	19.3	9.3	146	9	263	151
1975	20038	16.7	12	125	10	191	136
1976	22848	15.9	12.7	83	10	120	125
1977	24770	10.9	14.4	71	8	100	169
1978	27459	8.9	14.5	83	2	140	72
1979	31114	12.6	12.9	78	3	139	73
1980	34341	13	13.2	117	2	148	44
1981	38835	15.5	13.2	171	1	153	56
1982	38944	6.8	18.7	113	1	88	26
1983	40860	4.8	19.8	73	0	43	27
1984	44498	7.7	17.8	80	0	79	32
1985	47969	7.8	16.4	87	0	88	22
1986	50814	6.7	15.1	83	1	120	32
1987	54397	8.8	13.7	98	1	160	27
1988	59131	9.7	12	130	0	187	53
1989	62702	7.9	11.3	134	0	209	37
1990	62377	5.3	12.8	126	1	198	35
1991	60744	2.8	16.2	116	1	142	27
1992	60653	2.7	17.8	137	0	103	14

Year	SPECIALI	SUPPORT	AIR OPS	AIR TECH	COMBAT	COMMUNI	DENTAL
1973	6	13	17	61	34	30	6
1974	104	78	161	489	926	514	19
1975	115	98	120	559	936	529	10
1976	86	69	108	416	768	374	21
1977	92	78	86	464	667	358	15
1978	81	70	90	462	519	273	13
1979	107	74	111	619	656	294	17
1980	110	81	110	612	802	359	22
1981	97	77	110	529	661	365	18
1982	99	48	56	287	379	202	16
1983	81	31	44	151	219	123	11
1984	101	38	36	210	408	158	11
1985	81	71	82	313	402	245	17
1986	89	66	53	356	348	221	12
1987	96	63	75	374	472	235	17
1988	96	76	94	440	667	291	35
1989	118	109	88	532	613	359	26
1990	99	110	81	374	531	289	30
1991	92	86	65	281	569	242	19
1992	94	84	57	94	466	149	13

Year	LEME	LOGISTIC	MEDICAL	MIL - ENG	OTHERS	SEA OPS	SUPPORT
1973	23	151	9	20	95	8	62
1974	199	916	129	247	881	183	526
1975	238	805	140	285	680	172	499
1976	221	667	80	246	447	160	349
1977	183	664	95	226	485	170	285
1978	175	612	84	175	485	165	274
1979	219	691	100	223	558	223	316
1980	255	706	108	251	626	236	345
1981	220	669	99	255	507	254	332
1982	118	435	61	165	359	126	213
1983	66	289	52	92	159	91	148
1984	112	364	72	105	163	169	192
1985	139	408	78	128	91	275	237
1986	117	392	61	121	96	280	254
1987	115	472	110	125	56	404	291
1988	155	571	114	150	75	410	342
1989	209	639	146	202	78	520	344
1990	163	507	80	226	74	476	319
1991	149	421	93	135	38	344	280
1992	97	278	77	71	31	265	213

Model: MODEL1
Dependent Variable: ENG

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	5966.58160	2983.29080	6.051	0.0118
Error	15	7395.91840	493.06123		
C Total	17	13362.50000			

Root MSE	22.20498	R-square	0.4465
Dep Mean	105.83333	Adj R-sq	0.3727
C.V.	20.98109		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	164.502389	28.99454152	5.674	0.0001
LUR1524	1	-6.068198	2.05564034	-2.952	0.0099
TTRND	1	2.631830	1.04242249	2.525	0.0233

Durbin-Watson D	1.534
(For Number of Obs.)	18
1st Order Autocorrelation	0.195

Model: MODEL2
Dependent Variable: OPN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	32268.92590	10756.30863	51.674	0.0001
Error	14	2914.18521	208.15609		
C Total	17	35183.11111			

Root MSE	14.42762	R-square	0.9172
Dep Mean	133.77778	Adj R-sq	0.8994
C.V.	10.78476		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	138.559563	30.31469263	4.571	0.0004
LUR1524	1	-8.265275	1.56849524	-5.270	0.0001
CGDPCOM	1	906.530343	135.40307091	6.695	0.0001
TTRND	1	4.846030	0.72830993	6.654	0.0001

Durbin-Watson D	1.682
(For Number of Obs.)	18
1st Order Autocorrelation	0.120

Model: MODEL3
Dependent Variable: OTHR

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	26073.97477	8691.32492	17.761	0.0001
Error	14	6850.96968	489.35498		
C Total	17	32924.94444			

Root MSE	22.12137	R-square	0.7919
Dep Mean	55.94444	Adj R-sq	0.7473
C.V.	39.54167		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	246.865595	35.34621138	6.984	0.0001
LUR1524	1	-10.276662	2.73831955	-3.753	0.0021
CGDPM	1	283.043889	114.61462543	2.470	0.0270
TTRND	1	-4.615143	1.10570917	-4.174	0.0009

Durbin-Watson D	1.958
(For Number of Obs.)	18
1st Order Autocorrelation	-0.020

Model: MODEL4
Dependent Variable: SPEC

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	532.09620	266.04810	2.605	0.1069
Error	15	1531.90380	102.12692		
C Total	17	2064.00000			

Root MSE	10.10579	R-square	0.2578
Dep Mean	96.33333	Adj R-sq	0.1588
C.V.	10.49044		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	124.251741	13.19580647	9.416	0.0001
LUR1524	1	-2.133804	0.93554961	-2.281	0.0376
TTRND	1	0.231336	0.47442052	0.488	0.6329

Durbin-Watson D	2.308
(For Number of Obs.)	18
1st Order Autocorrelation	-0.187

Model: MODEL5
Dependent Variable: SUPOF

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	4069.24269	4069.24269	19.120	0.0005
Error	16	3405.25731	212.82858		
C Total	17	7474.50000			

Root MSE	14.58865	R-square	0.5444
Dep Mean	73.83333	Adj R-sq	0.5159
C.V.	19.75889		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	155.112778	18.90363688	8.205	0.0001
LUR1524	1	-5.714961	1.30698754	-4.373	0.0005

Durbin-Watson D	1.020
(For Number of Obs.)	18
1st Order Autocorrelation	0.420

Model: MODEL6
NOTE: No intercept in model. R-square is redefined.
Dependent Variable: AIROPS

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	128157.91904	42719.30635	310.448	0.0001
Error	15	2064.08096	137.60540		
U Total	18	130222.00000			

Root MSE	11.73053	R-square	0.9841
Dep Mean	81.44444	Adj R-sq	0.9810
C.V.	14.40311		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
LUR	1	-9.269821	1.41963514	-6.530	0.0001
DRGDP	1	4.039473	1.15746390	3.490	0.0033
LCUR	1	1.872913	0.14835326	12.625	0.0001

Durbin-Watson D	2.279
(For Number of Obs.)	18
1st Order Autocorrelation	-0.173

Model: MODEL7
Dependent Variable: AIRTECH

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	331881.21855	110627.07285	26.343	0.0001
Error	14	58793.72589	4199.55185		
C Total	17	390674.94444			

Root MSE	64.80395	R-square	0.8495
Dep Mean	392.94444	Adj R-sq	0.8173
C.V.	16.49189		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-883.819204	401.04784122	-2.204	0.0448
DRGDP	1	19.898203	6.48678968	3.067	0.0084
LUR	1	-36.216910	11.78426554	-3.073	0.0083
LCUR	1	19.170057	4.18616188	4.579	0.0004

Durbin-Watson D	1.171
(For Number of Obs.)	18
1st Order Autocorrelation	0.332

Model: MODEL8
Dependent Variable: ARMS

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	424306.56362	212153.28181	25.406	0.0001
Error	15	125257.93638	8350.52909		
C Total	17	549564.50000			

Root MSE	91.38123	R-square	0.7721
Dep Mean	560.16667	Adj R-sq	0.7417
C.V.	16.31322		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	1286.890998	110.71756470	11.623	0.0001
LUR	1	-93.137310	13.16883583	-7.073	0.0001
DRGDP	1	25.953037	9.03017848	2.874	0.0116

Durbin-Watson D	1.523
(For Number of Obs.)	18
1st Order Autocorrelation	0.159

Model: MODEL9
Dependent Variable: COMM

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	147710.97448	49236.99149	33.991	0.0001
Error	14	20279.46997	1448.53357		
C Total	17	167990.44444			

Root MSE	38.05961	R-square	0.8793
Dep Mean	281.44444	Adj R-sq	0.8534
C.V.	13.52295		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	247.779722	235.53691974	1.052	0.3106
DRGDP	1	14.186627	3.80971621	3.724	0.0023
LUR	1	-45.562191	6.92094389	-6.583	0.0001
LCUR	1	4.818127	2.45854877	1.960	0.0702

Durbin-Watson D	1.658
(For Number of Obs.)	18
1st Order Autocorrelation	0.041

Model: MODEL10
Dependent Variable: DENT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	303.78618	151.89309	4.716	0.0258
Error	15	483.15827	32.21055		
C Total	17	786.94444			

Root MSE	5.67543	R-square	0.3860
Dep Mean	17.94444	Adj R-sq	0.3042
C.V.	31.62781		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	28.197333	6.88487791	4.096	0.0010
TTRND	1	0.814959	0.29276799	2.784	0.0139
LUR	1	-2.190034	0.88991463	-2.461	0.0265

Durbin-Watson D	1.785
(For Number of Obs.)	18
1st Order Autocorrelation	-0.031

Model: MODEL11
Dependent Variable: LEMC

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	36525.19190	18262.59595	20.137	0.0001
Error	15	13603.75254	906.91684		
C Total	17	50128.94444			

Root MSE	30.11506	R-square	0.7286
Dep Mean	163.94444	Adj R-sq	0.6924
C.V.	18.36906		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	375.534844	36.48742876	10.292	0.0001
DRGDP	1	7.962959	2.97593245	2.676	0.0173
LUR	1	-27.251070	4.33984400	-6.279	0.0001

Durbin-Watson D 1.278
(For Number of Obs.) 18
1st Order Autocorrelation 0.354

Model: MODEL12
Dependent Variable: LOGIS

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	390048.90297	130016.30099	68.599	0.0001
Error	14	26534.20814	1895.30058		
C Total	17	416583.11111			

Root MSE	43.53505	R-square	0.9363
Dep Mean	532.77778	Adj R-sq	0.9227
C.V.	8.17133		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	332.672856	269.42243510	1.235	0.2372
LUR	1	-70.077310	7.91662538	-8.852	0.0001
LCUR	1	9.091578	2.81224785	3.233	0.0060
DRGDP	1	26.786837	4.35780097	6.147	0.0001

Durbin-Watson D 1.241
(For Number of Obs.) 18
1st Order Autocorrelation 0.369

Model: MODEL13
Dependent Variable: MED

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3435.91889	3435.91889	7.173	0.0165
Error	16	7664.08111	479.00507		
C Total	17	11100.00000			

Root MSE	21.88618	R-square	0.3095
Dep Mean	91.66667	Adj R-sq	0.2664
C.V.	23.87584		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	161.190914	26.46642368	6.090	0.0001
LUR	1	-8.094673	3.02237199	-2.678	0.0165

Durbin-Watson D	1.717
(For Number of Obs.)	18
1st Order Autocorrelation	0.111

Model: MODEL14
Dependent Variable: MENG

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	59046.76565	29523.38282	40.234	0.0001
Error	15	11006.84546	733.78970		
C Total	17	70053.61111			

Root MSE	27.08855	R-square	0.8429
Dep Mean	176.72222	Adj R-sq	0.8219
C.V.	15.32832		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	456.653972	32.82051058	13.914	0.0001
LUR	1	-35.009624	3.90369782	-8.968	0.0001
DRGDP	1	7.358989	2.67685682	2.749	0.0149

Durbin-Watson D	1.295
(For Number of Obs.)	18
1st Order Autocorrelation	0.309

Model: MODEL15
Dependent Variable: NCMOTH

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	810267.28155	405133.64078	55.096	0.0001
Error	15	110297.82956	7353.18864		
C Total	17	920565.11111			

Root MSE	85.75074	R-square	0.8802
Dep Mean	278.22222	Adj R-sq	0.8642
C.V.	30.82095		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	954.700963	104.02435215	9.178	0.0001
LUR	1	-37.793211	13.44581468	-2.811	0.0132
TTRND	1	-33.512100	4.42346271	-7.576	0.0001

Durbin-Watson D	0.844
(For Number of Obs.)	18
1st Order Autocorrelation	0.530

Model: MODEL16
Dependent Variable: SEAOPS

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	4485.39963	2242.69981	0.133	0.8761
Error	15	252160.60037	16810.70669		
C Total	17	256646.00000			

Root MSE	129.65611	R-square	0.0175
Dep Mean	263.33333	Adj R-sq	-0.1135
C.V.	49.23650		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	275.099879	157.09143732	1.751	0.1003
DRGDP	1	-6.470659	12.81245411	-0.505	0.6209
LUR	1	0.755536	18.68458138	0.040	0.9683

Durbin-Watson D	0.424
(For Number of Obs.)	18
1st Order Autocorrelation	0.772

Model: MODEL17
Dependent Variable: SUPNCM

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	88187.72772	22046.93193	16.856	0.0001
Error	13	17003.88339	1307.99103		
C Total	17	105191.61111			

Root MSE	36.16616	R-square	0.8384
Dep Mean	290.72222	Adj R-sq	0.7886
C.V.	12.44011		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	191.171864	255.63577736	0.748	0.4679
DRGDP	1	11.030566	3.99987850	2.758	0.0163
LUR	1	-40.523287	10.70050208	-3.787	0.0023
LINCT	1	-4.750628	3.53820055	-1.343	0.2024
LCUR	1	5.810717	2.34989364	2.473	0.0280

Durbin-Watson D	0.913
(For Number of Obs.)	18
1st Order Autocorrelation	0.313

REFERENCES

A.F.S. (1990), Autobox: Version 3.0, Hatboro, Pennsylvania.

A.F.S. (1986), **User's Manual for Autobox Software.** Hatboro, Pennsylvania.

Box, G.E.P. and D.R. Cox (1964), "An Analysis of Transformations", Journal Royal Statistical Society, Series B, 26, pp. 211-243.

Box, G.E.P. and G. Jenkins (1970, 1976), **Time Series Analysis: Forecasting and Control.**, 2nd Ed., Holdenday, San Francisco.

Box, G.E.P. and D.A.Pierce (1970), "Distribution of Residual Autocorrelations in Autoregressive moving average Time Series Models", Journal of American Statistical Association, 65, pp. 1509-1526.

Cryer, J.D. (1986), **Time Series Analysis.** Duxbury Press, Boston.

Newbold, Paul (1983), "ARIMA Model Building and the Time Series Analysis Approach to Forecasting." Journal of Forecasting, 2: 23-35.

Pindyck, R.S. and D.L. Rubinfeld (1981), **Econometric Models and Economic Forecasts** (2nd Ed.) McGraw-Hill, New York.

Priestley, M.B. (1988), "Current Developments in Time Series Modelling", Journal of Econometrics, Annals 1988-1,37: 67-86.

S.A.S. (1985), **SAS User's Guide: ETS** (ver.5), SAS Institute Inc. Cary, North Carolina.

Solomon, B. (1991), "Alternative Modelling Strategies for the Canadian Economy : Macroeconometric Methodology.", DSEA Staff Note 5/91.

Vandaele, W. (1983), **Applied Time Series and Box-Jenkins Models** Academic press, New York.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared e.g. Establishment Sponsoring a contractor's report, or tasking agency, are entered in Section 8).

   DIRECTORATE OF SOCIAL AND ECONOMIC ANALYSIS

2. SECURITY CLASSIFICATION (overall security classification of the document, including special warning terms if applicable)

   UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title)

   AN ATTRITION FORECASTING MODEL: TECHNICAL SUMMARY OF DATA AND METHODOLOGY

4. AUTHORS (last name, first name, middle initial)

   SOLOMON, B.

5. DATE OF PUBLICATION (month Year of Publication of document)

   SEPTEMBER, 1994

6a. NO OF PAGES (total containing information. Include Annexes, Appendices, etc.)

   41

6b. NO OF REFS (total cited in document)

   12

7. DESCRIPTIVE NOTES (the category of document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

   DSEA RESEARCH NOTE

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address).

   DIRECTORATE OF ESTABLISHMENT AND MANPOWER REQUIREMENTS, OTTAWA

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)

   ACTIVITY 45744

9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR's document number (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)

10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification.)

   (X) Unlimited distribution
   ( ) Distribution limited to defence departments and defence contractors: further distribution only as approved
   ( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
   ( ) Distribution limited to government departments and agencies; further distribution only as approved
   ( ) Distribution limited to defence departments; further distribution only as approved
   ( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the test is bilingual).

This report provides a technical discussion on the time series model employed in ORA Project Report 692 "An Attrition Forecasting Model". Apart from the description of the model, the report also provides a listing of the data and sources used in the attrition study.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified . If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

UNIVARIATE TIME SERIES METHODS
TRANSFER FUNCTION, MULTIVARIATE TIME SERIES, MARIMA MODELS
REGRESSION ANALYSIS
TEMPORAL AGGREGATION
AUTOCORRELATION, PARTIAL AUTOCORRELATION FUNCTIONS
CROSS CORRELATION, OUTLIER, INTERVENTION DETECTION

Canada

S03634