



Estimation of cumulative totals
by Charles Edward Shaffer

A thesis submitted in partial fulfillment of the requirements for the degree of DOCTOR OF
PHILOSOPHY in Mathematics
Montana State University
© Copyright by Charles Edward Shaffer (1976)

Abstract:

In this thesis, the estimation of a cumulative total is investigated through the use of two groups of estimators. The first group is comprised of standard estimation procedures but with corrections in the standard deviations of the estimated cumulative totals. The second group involves ratio estimate procedures but again with corrections in the standard deviations. The estimators are then compared theoretically based on these variances and in the fourth chapter the estimators are compared through a monte-carlo study under various conditions.

ESTIMATION OF CUMULATIVE TOTALS

by

CHARLES EDWARD SHAFFER

A thesis submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Mathematics

Approved:

Robert D. Engle
Head, Major Department

K. F. Ziehl
Chairman, Examining Committee

Henry L. Persons
Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

July, 1976

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to his thesis advisor, Dr. Kenneth J. Tiahrt, for the guidance and the many helpful suggestions made during the preparation of this thesis.

Appreciation is also extended to Professors Martin A. Hamilton and Richard E. Lund who gave willingly of their time to aid in many areas.

The author is also very grateful to Professors Myron S. Henry, Franklin S. McFeely and Robert Sanks for serving on his graduate committee.

Finally, appreciation is extended to my wife, Linda, for her time and patience in typing this thesis and in helping proofread the final manuscript.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION	1
II. ESTIMATES - GROUP 1	6
Model	8
Method 1 - Prediction Based on Periodic Totals ..	16
Method 2 - Prediction Based on Subperiodic Ob- servations-Separate Estimates For Each Subperiod	20
Method 3 - Prediction Based on Subperiodic Ob- servations-Combined Estimates For All Sub- periods	26
Method 4 - Prediction Based on Cumulative Sub- periodic Data	33
Comparison	37
III. ESTIMATES - GROUP 2	40
Method 1 - Prediction Using Separate Subperiod Estimates Based on Fractions of the Total	43
Method 2 - Prediction Using All Subperiod Esti- mates Based on Fractions of the Total	59
Method 3 - Prediction Using Moving Totals of Consecutive Subperiod Estimates Based on Fractions of the Total	69
Comparison	76
IV. MONTE-CARLO STUDY	79
Table 1	81
Table 2	82
Table 3	83
Table 4	84
Table 5	85
Table 6	86
Table 7	87
Table 8	88
Table 9	89
Table 10	90
Table 11	91

CHAPTER	PAGE
Table 12	92
Table 13	93
Table 14	95
V. CONCLUSIONS	96
BIBLIOGRAPHY	100
APPENDIX	102

ABSTRACT

In this thesis, the estimation of a cumulative total is investigated through the use of two groups of estimators. The first group is comprised of standard estimation procedures but with corrections in the standard deviations of the estimated cumulative totals. The second group involves ratio estimate procedures but again with corrections in the standard deviations. The estimators are then compared theoretically based on these variances and in the fourth chapter the estimators are compared through a monte-carlo study under various conditions.

CHAPTER I

INTRODUCTION.

In many instances it is necessary to estimate the total number of occurrences of a given event which will occur in a future time period. Estimation of the number of passengers expected to arrive and depart from a particular airport during the next calendar year or the total assets of a banking institution at the end of their next fiscal year are two somewhat different examples for which related statistical techniques may be used.

In estimating airport traffic for a future year, one starts with a zero count and accumulates for subperiods (daily, weekly, or monthly) to finally obtain the cumulative total traffic at the end of the year. When considering bank assets, one starts with the total assets at the beginning of the fiscal year and again accumulates the increase in assets over each subperiod to arrive at the total value of all accumulated assets for the end of the year.

The purpose of this thesis is to present and compare several methods of estimating such cumulative totals. For ease of calculation, it is assumed that the subperiodic (e.g. monthly) observations are available at the end of

each subperiod and that these subperiodic observations are independent. If the observations are not independent, then the methods can be altered in an appropriate manner.

A great deal of work has been done on estimation for missing observations (see Hamilton [5] for a survey and bibliography of this work). Estimation of dependent random variables within the range of previously observed independent predictive variables using regression analysis is also a well-known and widely used technique. However, when the dependent variable to be estimated is based on values of the predictive variables outside the range of data which was used in forming the regression equation, the methods are not reliable. Draper and Smith [2, p. 6-7, 22] warn about the problems encountered in estimation of future observations. The major problem is the fact that the procedures usually used depend a great deal upon the assumption made; for example, no change in the trend or that the change in the trend is as predicted.

Since the data considered in this thesis will be cumulative totals, i.e., totals of subperiodic observations, these totals will be dependent from one subperiod to another. Consequently, standard regression methods can be used only after performing suitable transformations to

provide independence between the observations. This procedure alone will not utilize all of the information available. If one ignores the fact that subperiodic observations are available, the estimated value and the confidence interval for the predicted cumulative total may be obtained in the usual way. This procedure is given as the first method only to provide a reference point for those methods which follow. However, if the subperiodic observations are utilized and since the estimates are for cumulative totals, the confidence intervals should be the largest at the beginning of the period and gradually shorten as the period progresses until the interval becomes a point at the end of the period (since all the subperiodic observations have been recorded).

The procedures examined will use standard regression analysis to determine the estimates in those cases where the use of computer programs is necessary. These procedures have been divided into two groups; the methods in the first group utilize the data directly, while those in the second group involve ratios. The methods in the first group involve basic regression procedures with appropriate corrections in the variance of the estimators so that the confidence intervals shorten rather than lengthen as the

period progresses and subperiodic data becomes known.

Ratios are used in the second group because when variables are correlated (ratios are usually considered useful when correlations are greater than .5), the ratio estimate reduces the variance of the estimated total, [see Cochran, 1, p. 166]. In addition, the variables used in the ratio contain information about the trend and thus these estimates should not be as "sensitive" to changes in the trend. The variance for ratio estimates can only be approximated but the degree of the approximation has been studied and is referenced later. Again, the appropriate corrections are made in the estimated variance of the total so that the corresponding confidence intervals shorten rather than lengthen as the period progresses and subperiodic data becomes available.

The advantages and disadvantages of each procedure are examined and finally at the end of each group, the methods are compared from a theoretical standpoint.

In Chapter Four, the methods are compared by use of monte-carlo techniques. The procedures are initially compared under ideal situations where all the assumptions are met. In each case, the model for the data is presented as well as the distribution of the errors including

values of all necessary parameters. Comparisons are based upon the percentage of the confidence intervals which contain the actual cumulative total.

By varying the trend and the distribution of errors, comparisons can be made concerning the sensitivity of each model to different assumptions regarding these factors. In some cases, the model assumed was a continuous curve with the trend remaining constant or with the trend changing at the beginning of the current period; in other cases, the model contains a shift upwards (or downwards) in the trend, again at the beginning of the current period. All possible combinations could not be examined but several different combinations of trend and distribution of errors are analyzed under each method.

Finally, a comparison of the estimates is presented, based upon the monte-carlo study as opposed to the theoretical comparisons given earlier.

CHAPTER II

ESTIMATES - GROUP 1

Each of the methods of estimation in Group 1 utilize standard regression analysis to obtain estimates of the future observations for the $n+1$ period. Each method given in this chapter uses these estimates to form an estimate for the cumulative total for the $(n+1)$ st period and its associated confidence interval.

For all of the procedures described in Group 1, the basic model is the same. The cumulative totals are to be estimated for the $(n+1)$ st period. It will be assumed that there are n periods of past data, each containing m subperiods, and that data is available for all these subperiods. For example, periods could be years, while subperiods could be months or weeks. The following notation will be employed in describing the various models.

Definition of Symbols

t = index of observations, numbered in order from one for the observation from the first subperiod to mn for the observation from the last subperiod of the latest complete period, i.e., $t = 1, 2, \dots, mn$.

x_{ji} = observed value of the random variable X_{ji}
 ($= \mu_{ji} + \epsilon_{ji}$) for the i -th subperiod of
 the j -th period.

μ_{ji} = mean value for the i -th subperiod of the
 j -th period. (For the purpose of this
 thesis, the μ_{ji} 's are assumed to have a
 polynomial representation in t .)

ϵ_{ji} = error term for the i -th subperiod of the
 j -th period and it is assumed the ϵ_{ji} are
 identically and independently distributed
 with a mean of zero and variance of σ_{ϵ}^2 .

σ_{ϵ}^2 = variance for subperiodic observations.

$Z_j^{(i)}$ = periodic total for the j -th period at the
 end of the i -th subperiod; that is, the
 sum of the first i subperiodic observations
 in period j . (Usually i will be equal to
 m). ($Z_j^{(i)} = \sum_{k=1}^i X_{jk}$).

Y_j = cumulative total for period j ($Y_j = Z_j^{(m)}$).

$\hat{Y}_j^{(i)}$ = estimated periodic total for the j -th period
 based upon data for the first i subperiods
 of the j -th period ($i = 0, 1, \dots, m-1$). When
 $i = 0$, no data is available.

MODEL - GROUP 1

In deriving a suitable model, any properties that the estimates should have must be taken into consideration.

Assume the subperiodic observations are independently distributed random variables. Assume also, for ease of calculation, that the variances are the same for all subperiods.

Based upon the estimated periodic total, an "estimation interval" will be formed. Ideally, as additional subperiodic data becomes available, the estimation intervals will become shorter since only the portion of the cumulative total due to the remaining unobserved subperiods is unknown. Finally, after the last subperiodic observation in the present period becomes available, the cumulative total is known so the estimation interval should reduce to a point, the actual cumulative total for the just concluded period.

Ordinarily when $Y_j = \sum_{i=1}^m X_{ji}$, the variance of Y_j would be

$$\begin{aligned}
\text{var}(Y_j) &= \sum_{i=1}^m \text{var}(X_{ji}) + 2 \sum_{\substack{i,k \\ i < k}} \text{cov}(X_{ji}, X_{jk}) \\
&= \sum_{i=1}^m \text{var}(X_{ji}) \\
&= m \sigma_{\epsilon}^2,
\end{aligned}$$

since the subperiodic observations are assumed independent with equal variances. However, since the objective is to estimate the actual periodic cumulative total, (not the mean value for similar periods), subperiods for which observations have been realized should not enter into the variance of the actual cumulative total. Actual subperiodic observations will be used when available in calculating the total and these actual subperiodic observations contain "no error" relative to the actual cumulative total. The estimate of the cumulative total will be obtained by first estimating those subperiodic observations that are not available, then the estimate of the cumulative total is found by summing these estimates together with those subperiodic observations that are already available. Thus, the variance of the estimate of the cumulative total is due to the deviation of the estimated values from the actual values and not from the deviations of the actual values from population parameters.

Therefore, the model for the cumulative total before any subperiodic observations are available will be the sum of the mean and error terms for each subperiod, i.e.,

$$Y_j^{(0)} = \sum_{i=1}^m X_{ji}.$$

The model, after the observation for the first subperiod becomes available, will consist of the observed value for the first subperiod plus the sum of the mean and error terms for each of the remaining subperiods, i.e.,

$$Y_j^{(1)} = x_{j1} + \sum_{i=2}^m X_{ji}.$$

Continuing in this manner, replacing the mean and error terms for a subperiod with the observations as they become available, the model for the cumulative total at the end of k-th subperiod becomes

$$Y_j^{(k)} = \sum_{i=1}^k x_{ji} + \sum_{i=k+1}^m X_{ji}.$$

Finally, at the end of the last subperiod, the model for the cumulative total becomes

$$Y_j \equiv Y_j^{(m)} = \sum_{i=1}^m x_{jm}.$$

Notice that at the end of the last subperiod, no error remains in the cumulative total because at that time, the cumulative total is known exactly. Therefore, an estimate for the cumulative total after the end of the last subperiod is not required.

The following notation is required to express the model in matrix form. Matrices are denoted by bold face print and the case $m=12$ is shown.

$$\mathbf{U}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^{12} \end{bmatrix} \quad \mathbf{Y}_j = \begin{bmatrix} Y_j^{(0)} \\ Y_j^{(1)} \\ \vdots \\ Y_j^{(11)} \end{bmatrix} \quad \mathbf{X}_j = \begin{bmatrix} x_j^1 \\ x_j^2 \\ \vdots \\ x_j^{12} \end{bmatrix} \quad \mathbf{E}_j = \begin{bmatrix} \epsilon_j^1 \\ \epsilon_j^2 \\ \vdots \\ \epsilon_j^{12} \end{bmatrix}$$

[illegible]

$$N = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The matrix model for the j -th period is

$$Y_j = MX_j + NU_j + NE_j.$$

Let $\hat{U}_j$ be an unbiased estimate for U_j . Note $\hat{U}_j$ is a matrix of estimates of the subperiodic observations.

THEOREM 1. An unbiased estimate of $\mathcal{E}_G(Y_j)$ is given by

$$\hat{Y}_j = MX_j + N\hat{U}_j.$$

Proof. $\mathcal{E}_G(Y_j) = \mathcal{E}_G(MX_j + NU_j + NE_j)$

$$= MX_j + NU_j,$$

since $\mathcal{E}_G(E_j) = \emptyset$. Note, $\mathcal{E}_G(\epsilon_{ji}) = 0$ by assumption.

In these expectations $G = \{X_{ji}, \dots, X_{jm}\}$. In the following,

$$H = \{X_{11}, \dots, X_{j-1,m}\}.$$

Also, $\mathcal{E}_H(\hat{Y}_j) = \mathcal{E}_H(MX_j + N\hat{U}_j) = MX_j + NU_j$ because

$\mathcal{E}_H(\hat{U}_j) = U_j$. Thus, $\hat{Y}_j$ is an unbiased estimate of $\mathcal{E}_G(Y_j)$. $\square$

THEOREM 2. The variance-covariance matrix Σ for Y_j is

$$\Sigma = \sigma_\epsilon^2 \begin{bmatrix} m & m-1 & m-2 & m-3 & \dots & k & k-1 & \dots & 1 \\ m-1 & m-1 & m-2 & m-3 & \dots & k & k-1 & \dots & 1 \\ m-2 & m-2 & m-2 & m-3 & \dots & k & k-1 & \dots & 1 \\ m-3 & m-3 & m-3 & m-3 & \dots & k & k-1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ k & k & k & k & \dots & k & k-1 & \dots & 1 \\ k-1 & k-1 & k-1 & k-1 & \dots & k-1 & k-1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & \dots & 1 & 1 & \dots & 1 \end{bmatrix}.$$

$$\begin{aligned} \text{Proof. } \text{Var}[Y_j^{(i)}] &= \text{Var}[\sum_{k=1}^i x_k + \sum_{k=i+1}^m \mu_k + \sum_{k=i+1}^m \epsilon_{jk}] \\ &= \text{Var}(\sum_{k=i+1}^m \epsilon_{jk}) = \sum_{k=i+1}^m \text{Var}(\epsilon_{jk}) \\ &= (m-i)\sigma_\epsilon^2. \end{aligned}$$

Assume $i < i'$.

$$\begin{aligned}
 \text{Cov}(Y_j^{(i)}, Y_j^{(i')}) &= \\
 \mathcal{E}[(Y_j^{(i)} - \mathcal{E}(Y_j^{(i)}))(Y_j^{(i')} - \mathcal{E}(Y_j^{(i')}))] &= \\
 \mathcal{E}[\sum_{k=i+1}^m \epsilon_{jk} \cdot \sum_{k'=i'+1}^m \epsilon_{jk'}] &= \\
 \mathcal{E}[\sum_{k=i+1}^{i'} \epsilon_{jk} + \sum_{k=i'+1}^m \epsilon_{jk}] \cdot \sum_{k'=i'+1}^m \epsilon_{jk'} &= \\
 \mathcal{E}[\sum_{k=i+1}^{i'} \epsilon_{jk} \cdot \sum_{k'=i'+1}^m \epsilon_{jk'}] &+ \\
 + \mathcal{E}[\sum_{k=i'+1}^m \epsilon_{jk} \cdot \sum_{k'=i'+1}^m \epsilon_{jk'}] &= \\
 \sum_{k=i+1}^{i'} \sum_{k'=i'+1}^m \mathcal{E}(\epsilon_{jk}) \mathcal{E}(\epsilon_{jk'}) &+ \\
 + \mathcal{E}[\sum_{k=i'+1}^m \epsilon_{jk}^2 + 2 \sum_{\substack{k, k'=i'+1 \\ k < k'}}^m \epsilon_{jk} \epsilon_{jk'}] &= \\
 0 + \sum_{k=i'+1}^m \mathcal{E}(\epsilon_{jk}^2) + 0 &= \\
 (m-i') \sigma_{\epsilon}^2. &
 \end{aligned}$$

Thus, $\mathfrak{Z}$ is as shown. $\square$

For $m=12$, $\mathfrak{Z}$ is

$$\Sigma = \begin{bmatrix} 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 11 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 10 & 10 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 9 & 9 & 9 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 8 & 8 & 8 & 8 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 7 & 7 & 7 & 7 & 7 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 6 & 6 & 6 & 6 & 6 & 6 & 6 & 5 & 4 & 3 & 2 & 1 \\ 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 & 4 & 3 & 2 & 1 \\ 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 3 & 2 & 1 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \sigma_{\epsilon}^2$$

Several methods will be described to obtain estimates for the cumulative totals. The first method involves estimating the cumulative total directly while the remaining methods in this group involve estimating the unavailable subperiodic observations, i.e., u_j 's, and then forming an estimate of the cumulative total by summing observed and estimated subperiodic data.

The variance-covariance matrix for $\hat{Y}_j$ is presented separately for each of the methods considered because it is a function of the variance-covariance matrix for the estimates of the future observations.

Any model may be used for obtaining the estimates of future observations based on historic data. However, in the interest of simplicity, polynomial models have been used for each of the methods presented in this thesis.

ESTIMATES

Method 1 - Prediction Based On Periodic Totals

This is the only method of estimation for the cumulative total in Group 1 that does not involve estimates for the individual subperiodic observations. To predict Y_{n+1} , compute the cumulative totals for periods one through n

$$Y_j = \sum_{i=1}^m x_{ji},$$

and base the estimates on a regression equation of the form

$$Y = TB + S,$$

where

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & \dots & 1 \\ 1 & 2 & 4 & \dots & 2^k & \dots & 2^p \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 1 & n & n^2 & \dots & n^k & \dots & n^p \end{bmatrix},$$

$$B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix}.$$

That is, regress the totals versus a polynomial of degree p ($< n$) in periods (time),

$$Y_j = \beta_0 + \beta_1 j + \beta_2 j^2 + \dots + \beta_p j^p + \delta_j$$

where δ_j is the error term for the j -th period. Note

$$\delta_j = \sum_{i=1}^m \epsilon_{ji}.$$

From this regression an estimate of the trend and an estimate of the variance of the error term for the periodic total is obtained.

Based upon this regression, one obtains

$$\hat{\mathbf{B}} = (\mathbf{T}'\mathbf{T})^{-1} \mathbf{T}'\mathbf{Y} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ . \\ . \\ . \\ \hat{\beta}_p \end{bmatrix}.$$

Using this estimate of the β 's, the estimate of the periodic total for the coming period ($(n+1)$ -st period) is

$$\hat{Y}_{n+1}^{(0)} = \hat{\beta}_0 + \hat{\beta}_1(n+1) + \dots + \hat{\beta}_p(n+1)^p.$$

This estimate is the best linear unbiased estimate (BLUE) among all estimates based on the cumulative totals

[Graybill, 4] (under the previous assumptions and by assuming the error terms to be normally distributed). Therefore, an estimation interval can be obtained such that

$$\Pr_{X_{11} \dots X_{n+1, m}} [\hat{Y}_{n+1}^{(0)} - Z_{\alpha/2} s \leq Y_{n+1} \leq \hat{Y}_{n+1}^{(0)} + Z_{\alpha/2} s] \\ \approx 1 - \alpha.$$

In this expression, $Z_{\alpha/2}$ is the standard normal deviate such that

$$\Pr(Z \geq Z_{\alpha/2}) = \alpha/2, \text{ and}$$

$$s^2 = \text{var}(\hat{Y}_{n+1}^{(0)} - Y_{n+1}) = [1 + \mathbf{A}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{A}'] s_{Y.T}^2$$

where $s_{Y.T}^2$ is the error variance obtained from the regression analysis and $\mathbf{A} = (1 \ (n+1) \ \dots \ (n+1)^P)$. Z is used as an approximation of the student's t distribution. This type of approximation will be used in subsequent sections for the t and the Satterthwaite approximate t .

In subsequent methods, the estimation intervals of the type above will be called estimation intervals of level $1-\alpha$ and will be expressed in the form,

$$\hat{Y}_{n+1}^{(0)} \pm Z_{\alpha/2} \cdot (s).$$

This method is precisely identical to the manner in which one would predict a future observation based on ordinary regression techniques. Consequently, it is subject to all of the usual limitations on this type of prediction. These limitations include:

- 1) the prediction cannot be updated as subperiodic data becomes available,
- 2) it is dependent on a continuation of the assumed trend during the future period being predicted, and
- 3) the variance of the estimate does not decrease to zero as the subperiodic observations become available since they are not utilized in any way.

Method 2 - Prediction Based On Subperiodic Observations -
Separate Estimates for Each Subperiod

For this method, an estimate of the cumulative total is obtained based on individual estimates for each of the subperiods.

Since estimates for each of the subperiodic observations in the coming period are required, m regressions will be needed, one for each subperiod. For each subperiod ($i = 1, 2, \dots, m$), the model used will be

$$X_{ji} = \beta_{i0} + \beta_{i1}t + \dots + \beta_{ip}t^p + \eta_{ji}$$

$$j = 1, \dots, n$$

where η_{ji} is the error term. Note, η_{ji} may not equal ϵ_{ji} since η_{ji} is the failure of X_{ji} to follow the polynomial equation, not the error from the mean value and would be equal to ϵ_{ji} only if the assumed trend is correct. The values of t for consecutive observations included in the regression analysis for subperiod i will be

$$t = i, i+m, i+2m, \dots, i+(n-1)m.$$

From the regression run for subperiod i , one obtains estimates of

$\beta_{i0}, \beta_{i1}, \dots, \beta_{ip}$, and $\sigma_{i\eta}^2$,

denoted by

$\hat{\beta}_{i0}, \hat{\beta}_{i1}, \dots, \hat{\beta}_{ip}$, and $\hat{\sigma}_{i\eta}^2$.

Using the matrices

$$X^{(i)} = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix}, \quad B^{(i)} = \begin{bmatrix} \beta_{i0} \\ \beta_{i1} \\ \vdots \\ \beta_{ip} \end{bmatrix}, \quad E^{(i)} = \begin{bmatrix} \eta_{1i} \\ \eta_{2i} \\ \vdots \\ \eta_{ni} \end{bmatrix},$$

$$M_0^{(i)} = [1 \ (i+nm) \ (i+nm)^2 \ \dots \ (i+nm)^p] \quad \text{and}$$

$$M^{(i)} = \begin{bmatrix} 1 & i & i^2 & \dots & i^p \\ 1 & (i+m) & (i+m)^2 & \dots & (i+m)^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & i+(n-1)m & [i+(n-1)m]^2 & \dots & [i+(n-1)m]^p \end{bmatrix},$$

the model is

$$X^{(i)} = M^{(i)} B^{(i)} + E^{(i)},$$

giving

$$\hat{B}^{(i)} = (M^{(i)'} M^{(i)})^{-1} M^{(i)'} X^{(i)}.$$

The estimate for the i -th subperiod of the coming period is, therefore,

$$\hat{X}_{n+1,i} = M_0^{(i)} \cdot \hat{B}^{(i)}.$$

This method is the BLUE among all estimates formed from this data.

The estimate of the variance of $(\hat{X}_{n+1,i} - X_{n+1,i})$ is

$$\hat{\sigma}_i^2 = S_i^2 (1 + M_0^{(i)' (M^{(i)' M^{(i)})^{-1} M_0^{(i)})}$$

where S_i^2 is an estimate of residual error from the standard regression run for subperiod i .

Assuming subperiodic data is available for k subperiods of the $(n+1)$ -st period ($k = 0, 1, \dots, m-1$), the estimate of the periodic cumulative total is

$$\hat{Y}_{n+1}^{(k)} = \sum_{i=0}^k x_{n+1,i} + \sum_{i=k+1}^m \hat{X}_{n+1,i}$$

where, $x_{n+1,0} \equiv 0$.

As a direct result of Theorem 1, this estimate is unbiased. Also, the estimated variance of $(\hat{Y}_{n+1}^{(k)} - Y_{n+1})$ is

$$\begin{aligned} v_{(k)}^2 &= \widehat{\text{Var}}[\sum_{i=k+1}^m (\hat{X}_{n+1,i} - X_{n+1,i})] \\ &\doteq \sum_{i=k+1}^m \hat{\sigma}_i^2. \end{aligned}$$

Let ϕ_i be a matrix of dimension $(1 \times i)$ whose elements are all zero, and 1_i a matrix of dimension $(1 \times i)$ whose elements are all one; also note

$$x'_{n+1} = (x_{n+1,1}, \dots, x_{n+1,m})$$

and

$$\hat{u}'_{n+1} = (\hat{x}_{n+1,1}, \dots, \hat{x}_{n+1,m})$$

Then, for $k=0$ (before any subperiodic data is observed)

$$\hat{y}_{n+1}^{(0)} = \phi_m \cdot x_{n+1} + 1_m \cdot \hat{u}_{n+1},$$

i.e., the estimated total at the beginning of the period is the sum of the estimates of the subperiodic observations.

For $k > 0$,

$$\hat{y}_{n+1}^{(k)} = [1_k : \phi_{m-k}] \cdot x_{n+1} + [\phi_k : 1_{m-k}] \cdot \hat{u}_{n+1};$$

that is, the actual subperiodic observations are substituted for their estimates as they become available.

Estimation intervals can be formed for subperiod i of level $1-\alpha$ by

$$\hat{x}_{n+1}^{(i)} \pm (Z_{\alpha/2}) \hat{\sigma}_{(i)}.$$

These intervals may be used to see if the assumed trend has changed; that is, if $x_{n+1,i}$ is not contained in the interval, then the possibility that the assumed trend has changed is greatly increased.

The estimation interval of level $1-\alpha$ on the estimated cumulative total for period $(n+1)$ at the end of the k -th subperiod is

$$\hat{Y}_{n+1}^{(k)} \pm (Z_{\alpha/2})v_{(k)}.$$

Advantages of this method are:

- 1) estimates for the cumulative total can be updated as data becomes available for subperiods,
- 2) subperiods can be used to check for a change in the assumed trend,
- 3) the variance of the estimated cumulative total decreases as data becomes available for subperiods. (Note the variance of the estimated total would decrease to zero if an estimate were formed at the end of the last subperiod),

- 4) variances of the error terms for different subperiods do not need to be assumed equal; however, variances for the same subperiod in different periods must be equal.

However, a disadvantage of this method is that because of the number of parameters to be estimated, which depends on the degree of the fitted polynomial, a great deal of data would be needed to have sufficiently many degrees of freedom for the estimates of the variance. For instance, five periods of data would only give one degree of freedom in each error variance if a third degree model is used. If the variances of the error terms for different subperiods are assumed equal, then it is possible to combine all the m regression problems into one analysis [Zellner, 17]. This in effect increases the number of degrees of freedom in the estimated variance of the error term. However, the disadvantage of different trends for each of the subperiods remains.

Method 3 - Prediction Based on Subperiodic Observations -
Combined Estimates For All Subperiods

One of the disadvantages of the preceding method is that a large amount of data is necessary in order to obtain sufficient degrees of freedom for the estimates of the error variance. This third method utilizes the assumption of equal variances for all subperiods in order to increase degrees of freedom for error variance. This is accomplished by combining all the data into one regression run. In addition, the disadvantage of unequal trends for different subperiods encountered in Method 2 is overcome by allowing only one equation for the trend.

Define the following matrices where

n = number of periods of data available

m = number of subperiods per period,

$$T_j = \begin{bmatrix} 1 & [(j-1)m+1] & [(j-1)m+1]^2 & \dots & [(j-1)m+1]^p \\ 1 & [(j-1)m+2] & [(j-1)m+2]^2 & \dots & [(j-1)m+2]^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & [(j-1)m+m] & [(j-1)m+m]^2 & \dots & [(j-1)m+m]^p \end{bmatrix}$$

$$\phi_{m-1} = (0 \ 0 \ 0 \ \dots \ 0)$$

$$\mathbf{B} = (\beta_0, \beta_1, \dots, \beta_{m+p-1})'$$

$$\mathbf{E}_j = (\epsilon_{j1}, \epsilon_{j2}, \dots, \epsilon_{jm})'$$

$$\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jm})'$$

$$\mathbf{M}_j = \begin{bmatrix} \mathbf{T}_j & \begin{smallmatrix} \vdots & \mathbf{I}_{m-1} \\ \vdots & \mathbf{0}_{m-1} \end{smallmatrix} \end{bmatrix}.$$

Thus the model for the j -th period is

$$\mathbf{x}_j = \mathbf{M}_j \cdot \mathbf{B} + \mathbf{E}_j.$$

For the first period ($j=1$) the model is

$$\mathbf{x}_1 = \mathbf{M}_1 \mathbf{B} + \mathbf{E}_1 = \begin{bmatrix} \mathbf{T}_1 & \begin{smallmatrix} \vdots & \mathbf{I}_{m-1} \\ \vdots & \mathbf{0}_{m-1} \end{smallmatrix} \end{bmatrix} \mathbf{B} + \mathbf{E}_1$$

$$= \begin{bmatrix} 1 & 1 & \dots & 1 & \vdots & 1 & 0 & \dots & 0 \\ 1 & 2 & \dots & 2^p & \vdots & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & m-1 & \dots & (m-1)^p & \vdots & 0 & 0 & \dots & 1 \\ 1 & m & \dots & m^p & \vdots & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \\ \beta_{p+1} \\ \vdots \\ \beta_{m+p-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1m-1} \\ \epsilon_{1m} \end{bmatrix}.$$

Similarly, for $j=6$,

$$x_6 = \begin{bmatrix} 1 & 5m+1 & \dots & (5m+1)^p & 1 & 0 & \dots & 0 \\ 1 & 5m+2 & \dots & (5m+2)^p & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 6m-1 & \dots & (6m-1)^p & 0 & 0 & \dots & 1 \\ 1 & 6m & \dots & (6m)^p & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_p \\ \beta_{p+1} \\ \vdots \\ \beta_{m+p-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{61} \\ \epsilon_{62} \\ \vdots \\ \vdots \\ \epsilon_{6m-1} \\ \epsilon_{6m} \end{bmatrix}.$$

Specifically for the first of four subperiods ($m=4$) and assuming a quadratic model ($p=2$),

$$x_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & 9 & 0 & 0 & 1 \\ 1 & 4 & 16 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \end{bmatrix}$$

or,

$$\begin{aligned} x_{11} &= \beta_0 + \beta_1 + \beta_2 + \beta_3 + \epsilon_{11} \\ x_{12} &= \beta_0 + 2\beta_1 + 4\beta_2 + \beta_4 + \epsilon_{12} \\ x_{13} &= \beta_0 + 3\beta_1 + 9\beta_2 + \beta_5 + \epsilon_{13} \\ x_{14} &= \beta_0 + 4\beta_1 + 16\beta_2 + \epsilon_{14} \end{aligned}$$

Thus, β_0 , β_1 and β_2 are the coefficients for quadratic trend in index of time while β_3 , β_4 and β_5 allow for differences due to the subperiod (i.e., a subperiod effect). Therefore, in general for a quadratic model,

$$X_{jk} = \begin{cases} \beta_0 + \beta_1((m-1)j+k) + \beta_2((m-1)j+k)^2 + \beta_{2+k} + \epsilon_{jk} & \text{when } k = 1, \dots, m-1 \\ \beta_0 + \beta_1(mj) + \beta_2(mj)^2 + \epsilon_{jk} & \text{when } k = m \end{cases}$$

where j is the period and k is the subperiod.

The model for all the data is therefore,

$$\begin{aligned} \mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} &= \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{bmatrix} \mathbf{B} + \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix} \\ &= \mathbf{M} \mathbf{B} + \mathbf{E} . \end{aligned}$$

Based upon the overall regression, an estimate of $\mathbf{B}$ is

$$\hat{\mathbf{B}} = (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{X}$$

and an estimate of the error variance is

$$\hat{\sigma}_\epsilon^2 = S_{X.M}^2,$$

the residual mean square.

Let

$$M_{0i} = (1 \dots (nm+i) \dots (nm+i)^p C_i)$$

where C_i is a matrix of dimension $1 \times (m-1)$ with all elements equal to zero except for the element in the i -th column which is equal to 1 (for $i = 1, \dots, m-1$). For $i=m$, all elements are zero. Using these properties, the estimate of the observation for the i -th subperiod of the $(n+1)$ -st period is

$$\hat{X}_{n+1,i} = M_{0i} \hat{B} \quad \text{for } i = 0, 1, \dots, m-1$$

with an estimated variance for $(\hat{X}_{n+1,i} - X_{n+1,i})$ of

$$\hat{\sigma}_i^2 = S_{X.M}^2 (1 + M_{0i}' (M'M)^{-1} M_{0i}).$$

Based upon these estimates for the subperiodic observations, the estimates for the cumulative total at the end of the k -th subperiod have exactly the same form as those for Method 2. That is,

$$\hat{Y}_{n+1}^{(k)} = \sum_{i=0}^k x_{n+1,i} + \sum_{i=k+1}^m \hat{X}_{n+1,i},$$

where $x_{n+1,0} \equiv 0$. However, the covariance between

$$(\hat{X}_{n+1,i} - X_{n+1,i})$$

and

$$(\hat{X}_{n+1,h} - X_{n+1,h})$$

is not zero as in Method 2. Therefore, these covariances must be estimated in order to find an estimate of the variance of $(\hat{Y}_{n+1}^{(k)} - Y_{n+1})$,

$$\begin{aligned} \hat{\sigma}_{ih} &= \widehat{\text{Cov}}[(\hat{X}_{n+1,i} - X_{n+1,i}), (\hat{X}_{n+1,h} - X_{n+1,h})] \\ &= S_{X.M}^2 \mathbf{M}'_{0i} (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}_{0h}. \end{aligned}$$

Thus, the estimated variance of $(\hat{Y}_{n+1}^{(k)} - Y_{n+1})$ is

$$v_{(k)}^2 = \sum_{i=k+1}^m \hat{\sigma}_i^2 + \sum_{i=k+1}^m \sum_{\substack{h=k+1 \\ i \neq h}}^m \hat{\sigma}_{ih}.$$

Consequently, the same procedures can be used to form estimation intervals on the subperiodic observations and on the cumulative totals.

Advantages of this method are:

- 1) estimates of cumulative totals can be updated as data becomes available for subperiods,

- 2) subperiods can be used to check for changes in the assumed trend,
- 3) the variance of the estimated cumulative total decreases as data becomes available for subperiods,
- 4) increased degrees of freedom for residual error are obtained from the same amount of data as for Methods 1 and 2, and
- 5) different equations for the overall trend in each subperiod will not occur.

The main disadvantage of this method is its inability to correct for changes in the assumed trend.

Method 4 - Prediction Based on Cumulative Subperiodic Data

In some cases it may be inconvenient to observe the subperiodic observations; however, the cumulative totals at the end of each subperiod may be available. Although it would be possible to convert these cumulative subperiodic totals into the individual subperiodic observations, this would be time consuming especially if there is a large amount of data. Therefore, this procedure utilizes these cumulative totals (not the individual observations) as an independent variable in order to predict the total for the current period. In addition to removing the necessity of converting to individual subperiodic observations, this procedure uses information from previous subperiods, i.e., for subperiod two, the cumulative total at that time is composed of the observations for subperiods one and two.

Let $Z_j^{(i)}$ be the cumulative total for the j -th period through the end of the i -th subperiod.

Then for $i = 1, \dots, m-1$, regress the models

$$Y_j = \beta_0^{(i)} + \beta_1^{(i)}t + \dots + \beta_p^{(i)}t^p + \beta_{p+1}^{(i)}Z_j^{(i)} + v_j^{(i)}$$

where $t = 1, 2, \dots, n, \dots, nm$ is the index for the observations and $v_j^{(i)}$ is the error term for the i -th subperiod of

the j -th period.

Let

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad B^{(i)} = \begin{bmatrix} \beta_0^{(i)} \\ \beta_1^{(i)} \\ \vdots \\ \beta_{p+1}^{(i)} \end{bmatrix} \quad E^{(i)} = \begin{bmatrix} v_1^{(i)} \\ v_2^{(i)} \\ \vdots \\ v_n^{(i)} \end{bmatrix}$$

and

$$M^{(i)} = \begin{bmatrix} 1 & i & \dots & i^p & z_1^{(i)} \\ 1 & m+i & \dots & (m+i)^p & z_2^{(i)} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & km+i & \dots & (km+i)^p & z_n^{(i)} \end{bmatrix}$$

where $k = n-1$. Then,

$$Y = M^{(i)} B^{(i)} + E^{(i)}$$

and the estimate for $B^{(i)}$ is given by

$$\hat{B}^{(i)} = (M^{(i)}, M^{(i)})^{-1} M^{(i)}, Y$$

Therefore, based upon the regression at the end of the i -th subperiod, the estimate of the cumulative total for period $n+1$ is

$$\hat{Y}_{n+1}^{(i)} = M_i \hat{B}^{(i)} = [1 \ (nm+i) \ \dots \ (nm+i)^p \ Z_{n+1}^{(i)}] \hat{B}^{(i)}.$$

The estimated variance of $\hat{Y}_{n+1}^{(i)} - Y_{n+1}$ is

$$v_{(i)}^2 = \frac{(m-i)}{i} S_{(i)}^2 + S_{(i)}^2 M_i' (M^{(i)}, M^{(i)})^{-1} M_i \\ + S_{(i)}^2 - 2 \hat{B}^{(i)} S_{(i)}^2$$

where $S_{(i)}^2$ is the residual mean square error from regression run i .

The estimation interval of level α is given by

$$\hat{Y}_{n+1}^{(i)} \pm Z_{\alpha/2} v_{(i)}$$

where Z is the usual standard normal deviate.

Advantages of this method are:

- 1) the estimates of cumulative totals are updated as data becomes available for subperiods,

- 2) it is extremely easy to run regressions, since the set-up of the design matrix $M^{(i)}$ is straight forward, and
- 3) the variance of the estimated periodic total decreases as data becomes available for subperiods.

Disadvantages of this method are:

- 1) it is difficult to check if the assumed trend has changed, and
- 2) the degrees of freedom for estimated variance are reduced.

COMPARISON OF ESTIMATORS

One of the primary ways to compare methods of estimation is to compare their variances. The theoretical values of the variances for each of the estimators in this chapter will be examined.

Method 1

The estimated variance of $(\hat{Y}_{n+1}^{(0)} - Y_{n+1})$ as given in Method 1 on page 18 is

$$[1 + \mathbf{A} (\mathbf{T}' \mathbf{T})^{-1} \mathbf{A}'] S_{Y.T}^2$$

where $S_{Y.T}^2$ is the residual mean square error which estimates $m \sigma_e^2$, provided the model is correct. Note that $\sigma_{Y.T}^2 \geq m \sigma_e^2$ for Method 1 [Draper & Smith, 2].

Method 2

The estimated variance of $(\hat{Y}_{n+1}^{(k)} - Y_{n+1})$ for Method 2 given on page 22 is

$$\sum_{i=k+1}^m S_i^2 (1 + \mathbf{M}_0^{(i)'} (\mathbf{M}^{(i)'} \mathbf{M}^{(i)})^{-1} \mathbf{M}_0^{(i)})$$

where S_i^2 is the residual mean square error from the i -th regression run. Provided the model is correct, S_i^2 esti-

mates σ_ϵ^2 , otherwise S_i^2 estimates σ_ϵ^2 , which is greater than σ_ϵ^2 .

Method 3

The estimated variance of $(\hat{Y}_{n+1}^{(k)} - Y_{n+1})$ as given in Method 3 on page 31 is

$$\sum_{i=k+1}^m S_{X.M}^2 (1 + \mathbf{M}_{0i}' (\mathbf{M}' \mathbf{M})^{-1} \mathbf{M}_{0i}) +$$

$$\sum_{i=k+1}^m \sum_{h=k+1}^m S_{X.M}^2 (\mathbf{M}_{0i}' (\mathbf{M}' \mathbf{M})^{-1} \mathbf{M}_{0h}),$$

$$i \neq h$$

The theoretical mean square error is greater than or equal to σ_ϵ^2 depending on whether or not the model is the true model.

Method 4

The estimated variance of $(\hat{Y}_{n+1}^{(i)} - Y_{n+1})$ as given for Method 4 on page 35 is

$$\frac{m-1}{i} S_i^2 + S_i^2 \mathbf{M}_i' (\mathbf{M}^{(i)'} \mathbf{M}^{(i)})^{-1} \mathbf{M}_i$$

where S_i^2 is the residual mean square error.

Since each of the above expressions involve matrices, it is not feasible to compare them in general. However, when there are 12 subperiods per period and five periods of past data, then at the beginning of the sixth period, the theoretical variances are $26.2 \sigma_{\epsilon}^2$ for Method 1, $26.2 \sigma_{\epsilon}^2$ for Method 2, and $25.2 \sigma_{\epsilon}^2$ for Method 3. Since the theoretical variances for Method 4 depend upon the data itself, these variances were not calculated. At the end of the sixth subperiod of the present period, the theoretical variances are $12.6 \sigma_{\epsilon}^2$ for Method 2 and $9.9 \sigma_{\epsilon}^2$ for Method 3. Note that Method 1 cannot be updated and therefore is not changed. At the end of the 11th subperiod, the theoretical variances are $2.1 \sigma_{\epsilon}^2$ for Method 2 and $1.3 \sigma_{\epsilon}^2$ for Method 3. Thus at the end of each subperiod, Method 3 has a slightly smaller theoretical variance than does Method 2. In addition, Methods 2 and 3 always have theoretical variances no greater than that of Method 1.

CHAPTER III
ESTIMATES - GROUP 2

The estimates in this group will utilize ratio estimates comprised of the cumulative subtotals divided by the periodic totals. Ratio estimates are desirable when the correlation between the numerator and denominator is greater than five tenths. The ratio estimate has a larger variance than the regression estimate (Method 2-4 without any trend terms) unless the regression is a straight line through the origin because

$$\text{Var}(\hat{Y}_{\text{regression}}) = S_y^2(1-\rho^2),$$

while

$$\text{Var}(\hat{Y}_{\text{ratio}}) = S_y^2 + R^2 S_x^2 - 2R\rho S_y S_x.$$

Therefore,

$$\text{Var}(\hat{Y}_{\text{reg}}) < \text{Var}(\hat{Y}_{\text{ratio}})$$

provided

$$-\rho^2 S_y^2 < R^2 S_x^2 - 2R\rho S_y S_x$$

or

$$(\rho S_y - R S_x)^2 > 0,$$

which implies the relation between the two variables is a straight line through the origin [Cochran 1, p. 199-200].

In addition, information about the trend is contained in the observations and cumulative subtotals so no terms are needed in the "model" to explain the trend as in Chapter II. Unfortunately, ratio estimates are not unbiased. This bias will be examined in detail for each method considered.

Use of ratio estimates will reduce the variability because the observations as well as the cumulative subtotals are correlated with the totals for the period. When ratio estimates are used for estimating a total, the ratio is estimated first. Since this ratio has little variability, the total is estimated by multiplying the value of the denominator variable (assumed known) by the estimated ratio.

The model for the subperiodic observations will be

$$X_{ji} = \mu_i + \alpha_j + \epsilon_{ji} \text{ for } i = 1, 2, \dots, m \text{ and} \\ j = 1, 2, \dots, n+1.$$

where

- μ_i = mean value for the i -th subperiod.
- α_j = mean value for the j -th period.
- ϵ_{ji} = error term for the i -th subperiod of the j -th period and it is assumed the ϵ_{ji} are identically and independently distributed with a mean of zero and variance of σ_ϵ^2 .

σ_{ϵ}^2 = variance for subperiodic observations.

$Z_j^{(i)}$ = periodic total for the j -th period at the end of the i -th subperiod; that is, the sum of the first i subperiodic observations in period j
 $(Z_j^{(i)} = \sum_{k=1}^i X_{jk})$.

Y_j = cumulative total for period j
 $(Y_j = Z_j^{(m)} = \sum_{i=1}^m X_{ji})$.

$\hat{Y}_j^{(i)}$ = estimated periodic total for the j -th period based upon data for the first i subperiods of the j -th period ($i = 0, 1, \dots, m-1$).

Method 1 - Prediction Using Separate Subperiod Estimates
Based on Fractions of the Total.

Since the total for a period and the cumulative total through the i -th subperiod of that period are strongly correlated, these variables are used to form the ratio estimate. Define the fraction of the total observed at the end of the i -th subperiod as

$$R^{(i)} = \mathcal{E}(Z_{n+1}^{(i)}) / \mathcal{E}(Y_{n+1}).$$

This ratio is the reciprocal of the "standard" ratio estimate. The estimate of the total for the $(n+1)$ st period is the observed subtotal through subperiod i divided by the estimated fraction of the total observed at that time. The fraction of the total observed is estimated by

$$\hat{R}^{(i)} = \bar{Z}^{(i)} / \bar{Y}$$

where

$$\bar{Y} = \sum_{j=1}^n Y_j / n$$

and

$$\bar{Z}^{(i)} = \sum_{j=1}^n Z_j^{(i)} / n.$$

An estimate of the cumulative total for period $n+1$ at the end of subperiod i is given by

$$\hat{Y}_{n+1}^{(i)} = Z_{n+1}^{(i)} / \hat{R}^{(i)}.$$

The bias in this ratio estimate, $\hat{R}^{(i)}$, is now examined.

Since

$$Z_j^{(k)} = \sum_{i=1}^k \mu_i + k\alpha_j + \sum_{i=1}^k \epsilon_{ji}$$

and

$$Y_j = \sum_{i=1}^m \mu_i + m\alpha_j + \sum_{i=1}^m \epsilon_{ji}$$

it follows that

$$Y. = \sum_{j=1}^n Y_j = n \sum_{i=1}^m \mu_i + m \sum_{j=1}^n \alpha_j + \sum_{ij}^{mn} \epsilon_{ji}$$

and that

$$Z.^{(k)} = n \sum_{i=1}^k \mu_i + k \sum_{j=1}^n \alpha_j + \sum_{ij}^{kn} \epsilon_{ji}.$$

Therefore, the estimated ratio is given by

$$\hat{R}^{(k)} = \frac{n \sum_{i=1}^k \mu_i + k \sum_{j=1}^n \alpha_j + \sum_{ij}^{kn} \epsilon_{ji}}{n \sum_{i=1}^m \mu_i + m \sum_{j=1}^n \alpha_j + \sum_{ij}^{mn} \epsilon_{ji}}$$

while

$$R^{(k)} = \frac{\mathcal{E}(Z_{n+1}^{(k)})}{\mathcal{E}(Y_{n+1})} = \frac{\sum_{i=1}^k \mu_i + k\alpha_{n+1}}{\sum_{i=1}^m \mu_i + m\alpha_{n+1}}$$

Consequently, the expected value of the bias is

$$\mathcal{E}(\hat{R}^{(k)} - R^{(k)}) = \mathcal{E}\left(\frac{Z^{(k)}}{\bar{Y}} - R^{(k)}\right) = \mathcal{E}\left(\frac{Z^{(k)} - R^{(k)} \bar{Y}}{\bar{Y}}\right).$$

Provided $\left(\frac{1}{\bar{Y}}\right)$ and $(Z^{(k)} - R^{(k)} \bar{Y})$ are independent, the expected value of the bias is

$$\begin{aligned} \mathcal{E}\left(\frac{1}{\bar{Y}}\right) \mathcal{E}(Z^{(k)} - R^{(k)} \bar{Y}) &= \mathcal{E}\left(\frac{1}{\bar{Y}}\right) \mathcal{E}(Z^{(k)} - \frac{\mathcal{E}(Z_{n+1}^{(k)})}{\mathcal{E}(Y_{n+1})} \bar{Y}) = \\ \mathcal{E}\left(\frac{1}{\bar{Y}}\right) \mathcal{E}\left(\frac{Z^{(k)} \mathcal{E}(Y_{n+1}) - \mathcal{E}(Z_{n+1}^{(k)}) \bar{Y}}{\mathcal{E}(Y_{n+1})}\right) &= \\ \mathcal{E}\left(\frac{1}{\bar{Y}}\right) \frac{1}{\mathcal{E}(Y_{n+1})} \mathcal{E}(Z^{(k)} \mathcal{E}(Y_{n+1}) - \mathcal{E}(Z_{n+1}^{(k)}) \bar{Y}). \end{aligned}$$

If one can not assume the independence required, an alternative method of examining the bias is given later.

Assuming $\mathcal{E}\left(\frac{1}{\bar{Y}}\right) \doteq \frac{1}{\mathcal{E}(\bar{Y})}$, the expected value of the bias is

$$\begin{aligned} \frac{1}{\mathcal{E}(\bar{Y})} \cdot \frac{1}{\mathcal{E}(Y_{n+1})} [(\alpha_{n+1} - \sum_{i=1}^n \alpha_j) \cdot \\ (m \sum_{i=1}^k \mu_i - k \sum_{i=1}^m \mu_i)]. \end{aligned}$$

Since $\hat{Y}$ is an unbiased estimate of $\mathcal{E}(Y)$ and $\hat{Y}_{n+1}$ is an unbiased estimate of $\mathcal{E}(Y_{n+1})$, an estimate of the bias is

$$\frac{1}{\hat{Y}} - \frac{1}{\hat{Y}_{n+1}} \cdot [(\alpha_{n+1} - \sum_{i=1}^n \alpha_i) \cdot (m \sum_{i=1}^k \mu_i - k \sum_{i=1}^m \mu_i)]$$

An alternative method to find the expected value of the bias is to expand the estimated ratio, $\hat{R}^{(k)} = \bar{Z}^{(k)}/\bar{Y}$, in a Taylor Series about $\mathcal{G}(\bar{Z}^{(k)})/\mathcal{G}(\bar{Y})$, provided the Taylor Series expansion is convergent. Dropping all terms of order higher than two, the expected value of the bias is approximately

$$\begin{aligned} \mathcal{G}(\hat{R}^{(k)} - R^{(k)}) &\doteq (\sum_{i=1}^k \mu_i + k\bar{\alpha})/(\sum_{i=1}^m \mu_i + m\bar{\alpha}) - \\ &(\sum_{i=1}^k \mu_i + k\alpha_{n+1})/(\sum_{i=1}^m \mu_i + m\alpha_{n+1}) + \\ &(k\sigma^2/n)(\sum_{i=1}^k \mu_i + k\bar{\alpha})/(\sum_{i=1}^m \mu_i + m\bar{\alpha})^3 - \\ &(k\sigma^2/n)(1)/(\sum_{i=1}^m \mu_i + m\bar{\alpha})^2. \end{aligned}$$

In examining this expression, as well as the first one derived, the expected value of the bias is approximately zero under the following conditions:

1. If the μ_i 's are all approximately equal and if α_{n+1} is the average of $\alpha_1, \alpha_2, \dots, \alpha_n$; or,
2. if $k\sum_{i=1}^m \mu_i$ is approximately equal to $m\sum_{i=1}^k \mu_i$.

In addition, the bias decreases if $\hat{Y}_{n+1}$ is increased significantly. It is important to note that although different assumptions were made for each of the derivations and although different expressions for the expected value of the bias were obtained, in each case, the expected value of the bias is approximately zero under the same conditions.

Note that if the standard ratio estimate is used, the estimated bias has $Z_{n+1}^{(k)}$ in the denominator instead of $\hat{Y}_{n+1}$ and would therefore be larger. Thus, the bias is a function of $\hat{Y}_{n+1}$, the subperiod, k , and the number of subperiods, m , as well as a function of the μ 's and α 's. The expression for the bias as derived can be used to approximate the bias at the end of each subperiod.

In order to examine the behavior of this bias, the following example is presented. Example: Assume that data is available for three past periods and that each period had observations taken after each of three subperiods.

That is, $X_{ji} = \mu_i + \alpha_j + \epsilon_{ji}$ for $j = 1, 2, 3$ and $i = 1, 2, 3$.

Let $Z_j^{(k)} = \sum_{i=1}^k X_{ji}$, $\alpha_j = j$, $\mu_i = i+10$ and $\epsilon_{ji} = -1, 0, 1$ with equal probabilities. Assume the errors were all zero during the first three periods (ie., perfect knowledge).

Thus the following results could be obtained for the three subperiods in the 4th period ($j=4$).

Subperiod 1			Subperiod 2			Subperiod 3	
x_{41}	$\hat{Y}_4^{(1)}$	$\mathcal{E}(Y_4^{(1)})$	x_{42}	$\hat{Y}_4^{(2)}$	$\mathcal{E}(Y_4^{(2)})$	x_{43}	Y_4
14	45.2	47	15	45.1	46	$\begin{Bmatrix} 16 \\ 17 \\ 18 \end{Bmatrix}$	$\begin{Bmatrix} 45 \\ 46 \\ 47 \end{Bmatrix}$
			16	46.7	47	$\begin{Bmatrix} 16 \\ 17 \\ 18 \end{Bmatrix}$	$\begin{Bmatrix} 46 \\ 47 \\ 48 \end{Bmatrix}$
			17	48.2	48	$\begin{Bmatrix} 16 \\ 17 \\ 18 \end{Bmatrix}$	$\begin{Bmatrix} 47 \\ 48 \\ 49 \end{Bmatrix}$
15	48.5	48	15	46.7	47	$\begin{Bmatrix} 16 \\ 17 \\ 18 \end{Bmatrix}$	$\begin{Bmatrix} 46 \\ 47 \\ 48 \end{Bmatrix}$
			16	48.2	48	$\begin{Bmatrix} 16 \\ 17 \\ 18 \end{Bmatrix}$	$\begin{Bmatrix} 47 \\ 48 \\ 49 \end{Bmatrix}$
			17	49.8	49	$\begin{Bmatrix} 16 \\ 17 \\ 18 \end{Bmatrix}$	$\begin{Bmatrix} 48 \\ 49 \\ 50 \end{Bmatrix}$
16	51.7	49	15	48.2	48	$\begin{Bmatrix} 16 \\ 17 \\ 18 \end{Bmatrix}$	$\begin{Bmatrix} 47 \\ 48 \\ 49 \end{Bmatrix}$
			16	49.8	49	$\begin{Bmatrix} 16 \\ 17 \\ 18 \end{Bmatrix}$	$\begin{Bmatrix} 48 \\ 49 \\ 50 \end{Bmatrix}$
			17	51.3	50	$\begin{Bmatrix} 16 \\ 17 \\ 18 \end{Bmatrix}$	$\begin{Bmatrix} 49 \\ 50 \\ 51 \end{Bmatrix}$

$$\hat{R}(1) = \frac{x_{11} + x_{21} + x_{31}}{z_1^{(3)} + z_2^{(3)} + z_3^{(3)}} = \frac{39}{126} = .3095$$

$$\hat{R}(2) = \frac{z_1^{(2)} + z_2^{(2)} + z_3^{(2)}}{z_1^{(3)} + z_2^{(3)} + z_3^{(3)}} = \frac{81}{126} = .6429$$

Thus the expected value of the bias in the cumulative total at the end of subperiod 1, is given by

$$\mathcal{E}(\text{bias}) = \frac{1}{3}(45.2-47 + 48.5-48 + 51.7-49) = 1.4/3 = .47 .$$

Also, this expected bias at the end of subperiod 2 is

$$\mathcal{E}(\text{bias}) = \frac{1}{9}(45.1-46 + 46.7-47 + 48.2-48 + 46.7-47 + 48.2-48 + 49.8-49 + 48.2-48 + 49.8-49 + 51.3-50) = .22.$$

Note that if the errors were large relative to the observations, this bias would increase.

Procedures have been devised for removing the bias when ratio estimates are used. See, for example, Durbin [3], Rao [7], Rao and Webster [9], and Schucany, Gray and Owen [12]. The application of these various procedures to this type of estimation is a matter for further study.

The variance of the estimate of the cumulative total is approximately

$$\begin{aligned} \text{Var}(\hat{Y}_{n+1}^{(i)}) = & \\ \frac{1}{[R^{(i)}]^2} \{ \text{Var}(Z_{n+1}^{(i)}) - 2\mathcal{E}(Y_{n+1}) \text{Cov}(Z_{n+1}^{(i)}, \hat{R}^{(i)}) + & \\ [\mathcal{E}(Y_{n+1})]^2 \text{Var} \hat{R}^{(i)} \}. & \end{aligned}$$

For this thesis, it will be assumed that $Z_{n+1}^{(i)}$ is approximately equal to $\bar{Z}^{(i)}$. The approximation error in $\text{Var} \hat{R}^{(i)}$ based on these assumptions is less than 9% to terms in $1/n^2$. [Sukhatme, 15].

In the following procedures, use of the term approximate variance will refer to the degree of approximation resulting from the ratio estimate.

The theoretical values of

$$\begin{aligned} \text{Var } Z_k^{(i)}, \text{Var } Y_k, \text{ and} \\ \text{Cov}(Z_k^{(i)}, Y_k) \end{aligned}$$

are found from the assumption that

$$Z_k^{(i)} = \sum_{j=1}^i X_{kj}.$$

That is, the i -th subperiodic total is the sum of the observations through superperiod i of the k -th period. Since the subperiodic observations are assumed independent with common variance, σ_e^2 ;

$$\text{Var}(Z_k^{(i)}) =$$

$$\text{Var}(\sum_{j=1}^i X_{kj}) =$$

$$\sum_{j=1}^i \text{Var}(X_{kj}) =$$

$$\sum_{j=1}^i \sigma_\epsilon^2 = i \sigma_\epsilon^2 .$$

At first this appears contrary to the variance expressions derived in Chapter II. In Chapter II the variance of the predicted value involved errors for only the remaining subperiods since the observations are available for the first subperiods of the period. Here, however, the variances of the subtotals are desired for past periods, which indeed do involve errors for each observation in the subtotal.

Since

$$Y_k = Z_k^{(m)} ,$$

$$\text{Var } Y_k = m \sigma_\epsilon^2 .$$

In order to find the variance (approximate) of the ratio estimate, the covariance between the numerator and denominator of the ratio must be found; that is, the covariance between the subperiodic and periodic totals is needed.

THEOREM 3. $\text{Cov}(Z_k^{(i)}, Y_k) = i \sigma_\epsilon^2$ when the observations are independent identically distributed random variables.

$$\begin{aligned}
 &\text{Proof. } \text{Cov}(Z_k^{(i)}, Y_k) \\
 &= \text{Cov}(\sum_{j=1}^i X_{kj}, \sum_{j'=1}^m X_{kj'}) \\
 &= \mathcal{E}[\sum_{j=1}^i X_{kj} \sum_{j'=1}^m X_{kj'}] - \mathcal{E}[\sum_{j=1}^i X_{kj}] \mathcal{E}[\sum_{j'=1}^m X_{kj'}] \\
 &= \mathcal{E}[\sum_{j=1}^i [X_{kj}]^2 + \sum_{\substack{j=1 \\ j \neq j'}}^i \sum_{j'=1}^m X_{kj} X_{kj'}] -
 \end{aligned}$$

$$\begin{aligned}
 &\mathcal{E}[\sum_{j=1}^i X_{kj}] \mathcal{E}[\sum_{j'=1}^i X_{kj'} + \sum_{j'=i+1}^m X_{kj'}] \\
 &= \sum_{j=1}^i \mathcal{E}(X_{kj})^2 + \sum_{\substack{j=1 \\ j \neq j'}}^i \sum_{j'=1}^m \mathcal{E}(X_{kj}) \mathcal{E}(X_{kj'}) - \\
 &\sum_{j=1}^i \mathcal{E}(X_{kj}) \cdot (\sum_{j'=1}^i \mathcal{E}(X_{kj'}) + \sum_{j'=i+1}^m \mathcal{E}(X_{kj'})) \\
 &= \sum_{j=1}^i [\mathcal{E}(X_{kj})^2 - [\mathcal{E}(X_{kj})]^2] + \\
 &\sum_{\substack{j=1 \\ j \neq j'}}^i \sum_{j'=1}^m [\mathcal{E}(X_{kj}) \mathcal{E}(X_{kj'}) - \mathcal{E}(X_{kj}) \mathcal{E}(X_{kj'})] \\
 &= \sum_{j=1}^i \text{Var}(X_{kj}) + 0 = i \sigma_\epsilon^2
 \end{aligned}$$

which completes the proof. $\square$

The variance of the estimated total depends on the approximate variance of $\hat{R}^{(i)}$, given by

$$\text{Var}(\hat{R}^{(i)}) = \frac{1}{n[\mathcal{E}(Y_{n+1})]^2} [\text{Var}(Z_{n+1}^{(i)}) - 2R^{(i)}\text{Cov}(Z_{n+1}^{(i)}, Y_{n+1}) + [R^{(i)}]^2 \text{Var}(Y_{n+1})].$$

Using the theoretical values of $\text{Var}(Z_{n+1}^{(i)})$, $\text{Var}(Y_{n+1})$, and $\text{Cov}(Z_{n+1}^{(i)}, Y_{n+1})$ as previously found,

$$\text{Var}[\hat{R}^{(i)}] = \frac{\sigma_{\epsilon}^2}{n[\mathcal{E}(Y_{n+1})]^2} [i - 2R^{(i)}i + [R^{(i)}]^2 m].$$

THEOREM 4. The approximate variance of $\hat{Y}_{n+1}^{(i)} - Y_{n+1}$ is

$$\begin{aligned} & \frac{1}{[R^{(i)}]^2} \{ i \sigma_{\epsilon}^2 + [\mathcal{E}(Y_{n+1})]^2 \text{Var} \hat{R}^{(i)} \} \\ & + m \sigma_{\epsilon}^2 - \frac{2}{R^{(i)}} i \sigma_{\epsilon}^2. \end{aligned}$$

Proof. Since

$$\begin{aligned} \text{Var}(\hat{Y}_{n+1}^{(i)} - Y_{n+1}) &= \\ \text{Var}(\hat{Y}_{n+1}^{(i)} - 2 \text{Cov}(\hat{Y}_{n+1}^{(i)}, Y_{n+1}) + \text{Var}(Y_{n+1})), \end{aligned}$$

it is necessary to estimate

$$\text{Cov}(\hat{Y}_{n+1}^{(i)}, Y_{n+1}).$$

$$\begin{aligned} \text{Cov}(\hat{Y}_{n+1}^{(i)}, Y_{n+1}) &= \text{Cov}\left(\frac{Z_{n+1}^{(i)}}{\hat{R}^{(i)}}, Y_{n+1}\right) \\ &= \frac{1}{R^{(i)}} \text{Cov}(Z_{n+1}^{(i)}, Y_{n+1}) = \frac{1}{R^{(i)}} i \sigma_{\epsilon}^2. \end{aligned}$$

Also,

$$\begin{aligned} \text{Var}(\hat{Y}_{n+1}^{(i)}) &= \text{Var}\left(\frac{Z_{n+1}^{(i)}}{\hat{R}^{(i)}}\right) = \\ &= \frac{1}{[R^{(i)}]^2} [\text{Var}(Z_{n+1}^{(i)}) - 2 \mathcal{E}(Y_{n+1}) \text{Cov}(Z_{n+1}^{(i)}, \hat{R}^{(i)}) \\ &+ [\mathcal{E}(Y_{n+1})]^2 \cdot \text{Var}[\hat{R}^{(i)}]]. \end{aligned}$$

Note,

$$\text{Cov}(Z_{n+1}^{(i)}, \hat{R}^{(i)}) = 0$$

since $\hat{R}^{(i)}$ does not depend upon $Z_{n+1}^{(i)}$.

Therefore,

$$\begin{aligned} \text{Var}(\hat{Y}_{n+1}^{(i)}) &= \\ &= \frac{1}{[R^{(i)}]^2} [i \sigma_{\epsilon}^2 + [\mathcal{E}(Y_{n+1})]^2 \text{Var} \hat{R}^{(i)}]. \end{aligned}$$

Thus, the approximate variance of $\hat{Y}_{n+1}^{(i)} - Y_{n+1}$ is

$$v_{(i)}^2 = \left[\frac{1}{R^{(i)}} \right]^2 \{ 1 \sigma_\epsilon^2 + [\mathcal{E}(Y_{n+1})]^2 \text{Var } \hat{R}^{(i)} \} \\ + m \sigma_\epsilon^2 - \frac{2}{R^{(i)}} 1 \sigma_\epsilon^2$$

which completes the proof. $\square$

Unbiased estimates of $\text{Var}(Z_{n+1}^{(i)})$, $\text{Cov}(Z_{n+1}^{(i)}, Y_{n+1})$, and $\text{Var}(Y_{n+1})$ can be obtained from the data. Let $\hat{V}_{(i)}^2$ be the resulting estimate of $\text{Var}(\hat{Y}_{n+1}^{(i)} - Y_{n+1})$ obtained by substituting $\hat{R}^{(i)}$ for $R^{(i)}$ and $\bar{Y}$ for $\mathcal{E}(Y_{n+1})$.

The estimation interval of level α is given by

$$\hat{Y}_{n+1}^{(i)} \pm Z_{\alpha/2} \cdot \hat{V}_{(i)}$$

where $Z_{\alpha/2}$ is the standard normal deviate such that

$$\Pr(Z \geq Z_{\alpha/2}) = \alpha/2.$$

Note that near the end of the period, say at the end of the $(m-1)$ -st subperiod,

$$\hat{V}_{(m-1)}^2 = \left[\frac{\sigma_{\epsilon}^2}{\hat{R}^{(m-1)}} \right]^2 \left\{ m-1 + \frac{1}{n} [(m-1) - 2(m-1)\hat{R}^{(m-1)} + m(\hat{R}^{(m-1)})^2] \right\} + m\sigma_{\epsilon}^2 - \frac{2(m-1)\sigma_{\epsilon}^2}{\hat{R}^{(m-1)}}.$$

Since $\hat{R}^{(m-1)}$ will be very close to one, use of $\hat{R}^{(m-1)} = 1$ gives

$$\begin{aligned}\hat{V}_{(m-1)}^2 &= \sigma_{\epsilon}^2 \left\{ m-1 + \frac{1}{n} [(m-1) - (2m-2) + m] \right\} + m\sigma_{\epsilon}^2 - (2m-2)\sigma_{\epsilon}^2 \\ &= \sigma_{\epsilon}^2 + \frac{1}{n}\sigma_{\epsilon}^2;\end{aligned}$$

while for $\hat{R}^{(m-1)} = .9$,

$$\begin{aligned}\hat{V}_{(m-1)}^2 &= \frac{\sigma_{\epsilon}^2}{(.81)} \left\{ m-1 + \frac{1}{n} [m-1 - (2m-2)(.9) + m(.81)] \right\} \\ &\quad + m\sigma_{\epsilon}^2 - \frac{2m-2}{(.9)} \sigma_{\epsilon}^2 \\ &= \frac{\sigma_{\epsilon}^2}{(.81)} [.01m + .8] + \frac{\sigma_{\epsilon}^2}{n} [.01m + .8].\end{aligned}$$

In each case, $\hat{V}_{m-1}^2$ is approximately σ_e^2 provided m is not extremely large and n small.

For example, assuming $n=5$, $m=12$, and $\sigma_e^2=1$, the approximate variances when $\hat{R}^{(i)} = i/12$ are as tabled below.

i	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{V}_{(i)}^2$	158	72	43.2	28.8	20.2	14.4	8.7	7.2	4.4	2.88	1.3	0

Thus, the variance of the estimate decreases as the number of subperiods for which data is available increases.

Consequently, the associated estimation intervals will become narrower as the estimates are updated throughout the period.

Advantages of this method are:

- 1) changes in the assumed trend have less effect since information about the trend is contained in $Z_{n+1}^{(i)}$,
- 2) calculations to obtain estimates are easy to perform,
- 3) correlations between the periodic totals and subperiodic totals are used to decrease the variance of the

estimator, and

- 4) the ratio is between zero and one.

A disadvantage of this method is that there are too few degrees of freedom for the estimated variance of the estimator since the individual components $\text{Var}(Z_{n+1}^{(i)})$, $\text{Var}(Y_{n+1})$, and $\text{Cov}(Z_{n+1}^{(i)}, Y_{n+1})$ are individually estimated. Each of these estimates has $(n-1)$ degrees of freedom where n is the number of periods used to form the estimated ratio. Thus, the degrees of freedom for the variance of the estimator will be small when using only the data from a few periods.

Note that data for other subperiodic totals is not used even though this data also provides estimates of σ_ϵ^2 . Therefore, a method which utilizes data from all of the subtotals would provide more degrees of freedom for the variance of the estimator.

Method 2 - Prediction Using All Subperiod Estimates Based
On Fractions Of The Total

In order to increase the degrees of freedom for the estimate of the variance, above that obtained in Method 1, data for other subperiodic totals can be utilized. That is, all the subperiodic totals may be combined into one regression analysis. This will require that certain transformations of the data be performed, since the subperiodic totals are not independent within a given period.

The fraction of the total observed at the end of subperiod i will be regressed versus an equation of degree p in i , the index for subperiods, in order to estimate the expected fraction of the total observed by the end of subperiod i for the current period.

The model is

$$U_j^{(i)} = \frac{Z_j^{(i)}}{Y_j} = \beta_0 + \beta_1 i + \beta_2 i^2 \dots + \beta_p i^p + \eta_{ji}$$

where i denotes the subperiod ($i=1, \dots, m$), and η_{ji} is an error term.

After estimates of $\beta_0, \beta_1, \beta_2, \dots$, and β_p are obtained, the estimate of $U_j^{(i)}$ is

$$\hat{U}_j^{(i)} = \hat{\beta}_0 + \hat{\beta}_1(i) + \hat{\beta}_2 i^2 + \dots + \hat{\beta}_p i^p$$

where $\hat{U}_j^{(i)}$ is the estimated fraction of the cumulative total observed by the end of subperiod i .

Since $U_j^{(i)}$ is equivalent to the ratio estimate given for Method 1, when $n=1$, its approximate variance is of the same form; that is,

$$\begin{aligned} \text{Var}(U_j^{(i)}) &\doteq \frac{1}{[\mathcal{E}(Y_j)]^2} [\text{Var}(Z_j^{(i)}) + [\mathcal{E}(U_j^{(i)})]^2 \text{Var}(Y_j) \\ &\quad - 2\mathcal{E}[U_j^{(i)}] \text{Cov}(Z_j^{(i)}, Y_j)] \\ &= \frac{\sigma_\epsilon^2}{[\mathcal{E}(Y_j)]^2} [i + [\mathcal{E}(U_j^{(i)})]^2 m - 2\mathcal{E}[U_j^{(i)}] i] . \end{aligned}$$

It is necessary to assume that the $U_j^{(i)}$'s are independent in order to utilize standard regression programs. Consequently, the variance-covariance structure of the $U_j^{(i)}$'s must now be examined. Since the approximate variance is a function of $\mathcal{E}[U_j^{(i)}]$, the results will be dependent upon $\mathcal{E}[U_j^{(i)}]$. Let $\sigma_\epsilon^2 \mathbf{V}$ be the variance-covariance matrix of the $U_j^{(i)}$'s for one period.

THEOREM 5. For $i < i'$,

$$\begin{aligned} &\text{Cov}(U_j^{(i)}, U_j^{(i')}) \\ &\doteq \frac{\sigma_\epsilon^2}{[\mathcal{E}(Y_j)]^2} [i - \mathcal{E}[U_j^{(i)}] i' - \mathcal{E}[U_j^{(i')}] i + \mathcal{E}[U_j^{(i)} U_j^{(i')}] m] \end{aligned}$$

Proof. Applying the same techniques and assumptions as used in deriving the expected value of the bias in Method 1,

$$\begin{aligned}
 & \text{Cov}(U_j^{(i)}, U_j^{(i')}) \\
 &= \text{Cov}\left[\frac{Z_j^{(i)}}{Y_j}, \frac{Z_j^{(i')}}{Y_j}\right] \\
 &= \frac{1}{[\mathcal{E}(Y_j)]^2} \text{Cov}[(Z_j^{(i)} - \mathcal{E}[U_j^{(i)}]Y_j)(Z_j^{(i')} - \mathcal{E}[U_j^{(i')}]Y_j)] \\
 &= \frac{1}{[\mathcal{E}(Y_j)]^2} [\text{Cov}[Z_j^{(i)}Z_j^{(i')}] - \mathcal{E}[U_j^{(i)}]\text{Cov}(Y_jZ_j^{(i')}) \\
 &\quad - \mathcal{E}[U_j^{(i')}] \text{Cov}(Z_j^{(i)}Y_j) + \mathcal{E}[U_j^{(i)}]U_j^{(i')}] \text{Var}([Y_j]^2)] \\
 &= \frac{\sigma_\epsilon^2}{[\mathcal{E}(Y_j)]^2} [i - \mathcal{E}[U_j^{(i)}]i' - \mathcal{E}[U_j^{(i')}]i + \mathcal{E}[U_j^{(i)}]U_j^{(i')}]_m],
 \end{aligned}$$

which completes the proof. ■

Note the covariance is zero if the periods are different, and also if $i' = m = i$, then $U_j^{(i)} = U_j^{(i')} = 1$, so again, the expression above becomes zero. Since the

above expression depends upon $U_j^{(i)}$ and $U_j^{(i')}$ as well as i and i' , the covariance cannot be determined until estimates of $U_j^{(i)}$ and $U_j^{(i')}$ are calculated.

If the covariances are too large to be ignored, then the U 's must be transformed into independent "observations" in order to use standard regression programs. In this case, a non-singular matrix M can be found such that

$$M \cdot \begin{bmatrix} U_j^{(1)} \\ \vdots \\ U_j^{(m-1)} \end{bmatrix} = M \cdot U_j$$

has a variance-covariance matrix of

$$\sigma_\epsilon^2 MVM' = \sigma_\epsilon^2 \cdot I_{m-1}$$

Using this transformation, the model for the j -th period becomes

$$MU_j = MCB + MN_j$$

where

$$C = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 4 & & 2^p \\ & & \vdots & & \\ & & \vdots & & \\ 1 & m & m^2 & & m^p \end{bmatrix},$$

$$B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

and

$$N = \begin{bmatrix} \eta_{j1} \\ \vdots \\ \eta_{jm} \end{bmatrix}.$$

Thus, the overall model becomes

$$A = \begin{bmatrix} M & & & \\ & M & & \\ & & \ddots & \\ & & & \emptyset \\ & \emptyset & & \\ & & \ddots & \\ & & & M \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ \vdots \\ \vdots \\ \vdots \\ u_n \end{bmatrix} = M^* U^*$$

$$\begin{bmatrix} M & & & & \\ & M & & & \\ & & \ddots & & \\ & & & \ddots & \\ \emptyset & & & & M \end{bmatrix} \begin{bmatrix} C \\ \vdots \\ C \end{bmatrix} B + \begin{bmatrix} M & & & & \\ & M & & & \\ & & \ddots & & \\ & & & \ddots & \\ \emptyset & & & & M \end{bmatrix} \begin{bmatrix} N_1 \\ \vdots \\ N_n \end{bmatrix}$$

$$= M^* C^* B + M^* N = D B + E.$$

THEOREM 6. The estimate $\hat{B}$ for B obtained from the transformed model is the estimate desired in the original model.

Proof. The estimate of B from the transformed model is

$$\begin{aligned} \hat{B} &= (D' D)^{-1} D' A \\ &= (C^* M^* M^* C^*)^{-1} C^* M^* M^* U^*. \end{aligned}$$

Let

$$V^* = \begin{bmatrix} V & & & & \\ & V & & & \\ & & \ddots & & \\ & & & \ddots & \\ \emptyset & & & & V \end{bmatrix}.$$

Since

$$M V M' = I,$$

$$M^{-1} M V M' M'^{-1} = M^{-1} M'^{-1}, \text{ and}$$

$$V = (M M')^{-1},$$

thus

$$M' M = V^{-1}.$$

This implies that,

$$M^* V M^{*'} = \begin{bmatrix} M V M' & & & \\ & M V M' & & \\ & & \ddots & \\ & & & M V M' \end{bmatrix} = I_{n(m-1)}.$$

Therefore,

$$M^{*-1} M^* V M^{*'} M^{*-1} = M^{*-1} M^{*'} = (M^{*'} M^*)^{-1}$$

which implies $V^* = (M^{*'} M^*)^{-1}$

or $M^{*'} M^* = V^{*-1}.$

Thus,

$$\hat{B} = (C^{*'} V^{*-1} C^*)^{-1} C^{*'} V^{*-1} U^*$$

which is the standard estimate as given in Graybill [4]

when the data is correlated. I

After $\hat{\mathbf{B}}$, the estimate of $\mathbf{B}$, is found, one calculates the estimated fraction of the total observed for period $n+1$ at the end of the i -th subperiod by

$$\hat{U}_{n+1}^{(i)} = (1 \ i \ i^2 \ \dots \ i^p) \hat{\mathbf{B}} = \mathbf{c}_0 \hat{\mathbf{B}}.$$

As in Method 1, the estimated total is given by

$$\hat{Y}_{n+1}^{(i)} = \frac{Z_{n+1}^{(i)}}{\hat{U}_{n+1}^{(i)}}.$$

The estimated approximate variance of $\hat{U}_{n+1}^{(i)}$ is

$$S^2 \cdot (\mathbf{c}_0' (\mathbf{C}^*{}' \mathbf{V}^{*-1} \mathbf{C}^*)^{-1} \mathbf{c}_0)$$

where S^2 is the residual mean square error for the transformed model.

The approximate variance of $(\hat{Y}_{n+1}^{(i)} - Y_{n+1})$ (using the same procedure as used for Method 1) is

$$\text{Var}(\hat{Y}_{n+1}^{(i)} - Y_{n+1}) = V_{(i)}^2 =$$

$$\frac{1}{(\mathcal{E}[U_{n+1}^{(i)}])^2} \{i \sigma_{\epsilon}^2 + [\mathcal{E}(Y_{n+1})]^2 \text{Var } U_{n+1}^{(i)}\} + m \sigma_{\epsilon}^2 - \frac{2}{U_{(i)}} i \sigma_{\epsilon}^2$$

Note that if it is not necessary to transform the data, then $\mathbf{V}^{*-1} \propto \mathbf{I}$, and the estimated approximate variance of $\hat{U}_{n+1}^{(i)}$ is the usual estimate for uncorrelated data.

Based upon $\hat{V}_i$, the estimation interval of level α is

$$\hat{Y}_{n+1}^{(i)} \pm Z_{\alpha/2} \cdot \hat{V}_i.$$

The advantage of this method over Method 1 is the increased degrees of freedom in the residual mean square error.

The disadvantages are:

- 1) if the data is strongly correlated; that is if the $U_j^{(i)}$'s are strongly correlated, then it is necessary to transform the data using an $\mathbf{M}$ matrix,
- 2) calculations are much more difficult than for previous methods, especially in finding the estimated variance of the estimate,
- 3) the inverse of $\mathbf{V}^*$ must be found; note this is equivalent to finding the inverse of $\mathbf{V}$ and could be

difficult for a large number of subperiods per period, and.

- 4) an estimate of σ_{ϵ}^2 must be found; note that the residual mean square from the regression analysis does not estimate σ_{ϵ}^2 . An estimate can be found using Methods 2-2 or 2-3.

Method 3 - Prediction Using Moving Totals of Consecutive Subperiod Estimates Based on Fractions Of The Total

A disadvantage of the estimates formed in Methods 1 and 2 is that the variances are large for small values of i (the first sub-periods of the period of interest) and thus the estimates are not very reliable. That is, the estimates are affected a great deal by observations in early subperiods that deviate significantly from their expected values. One method of removing a lot of the variability is to use the sum of the observations for the last m subperiods, $W_j^{(i)}$, at the end of each subperiod i . Then the value of the ratio

$$U_j^{(i)} = W_j^{(i)} / Y_j$$

will be close to one for every i . Thus, the model is

$$U_j^{(i)} = \beta_0 + \beta_1 i + \beta_2 i^2 + \dots + \beta_p i^p + \eta_{ji}$$

where i denotes the subperiod ($i=1, \dots, m$) and η_{ji} is an error term.

Since the $U_j^{(i)}$'s are not independent, the theoretical variance-covariance matrix of the $U_j^{(i)}$'s must again be found so an appropriate transformation can be performed.

Using the same techniques as used in Method 2,

$$\begin{aligned}\text{Var}(U_j^{(i)}) &\doteq \frac{1}{[\mathcal{E}(Y_j)]^2} [\text{Var } W_j^{(i)} + [\mathcal{E}(U_j^{(i)})]^2 \text{Var}(Y_j) \\ &\quad - 2\mathcal{E}[U_j^{(i)}] \text{Cov}(W_j^{(i)}, Y_j)] \\ &= \frac{\sigma_\epsilon^2}{[\mathcal{E}(Y_j)]^2} [m + [\mathcal{E}(U_j^{(i)})]^2 m - 2\mathcal{E}[U_j^{(i)}]i].\end{aligned}$$

It is also necessary to determine covariances between U_j 's within the same period.

$$\text{THEOREM 7. } \text{Cov}(U_j^{(i)}, U_j^{(k)}) \doteq \frac{1}{[\mathcal{E}(Y_j)]^2} \cdot \sigma_\epsilon^2 \cdot$$

$$[\{ m - (k-i) \} - \mathcal{E}(U_j^{(i)})_k - \mathcal{E}(U_j^{(k)})_i + \mathcal{E}(U_j^{(i)} U_j^{(k)})_m].$$

Proof. Using the same techniques as used in Method 2,

$$\begin{aligned}\text{Cov}(U_j^{(i)}, U_j^{(k)}) &\doteq \frac{1}{[\mathcal{E}(Y_j)]^2} \text{Cov}[(W_j^{(i)} - \mathcal{E}(U_j^{(i)})Y_j)(W_j^{(k)} - \mathcal{E}(U_j^{(k)})Y_j)] \\ &= \frac{1}{[\mathcal{E}(Y_j)]^2} [\text{Cov}[W_j^{(i)} W_j^{(k)}] - \mathcal{E}(U_j^{(i)}) \text{Cov}(Y_j W_j^{(k)}) \\ &\quad - \mathcal{E}(U_j^{(k)}) \text{Cov}(W_j^{(i)} Y_j) + \mathcal{E}(U_j^{(i)} U_j^{(k)}) \text{Var}([Y_j]^2)].\end{aligned}$$

$$= \frac{\sigma_{\epsilon}^2}{[\mathcal{E}(Y_j)]^2} [\{m-(k-i)\} - \mathcal{E}(U_j^{(i)})_k - \mathcal{E}(U_j^{(k)})_i + \mathcal{E}(U_j^{(i)} U_j^{(k)})_m].$$

Note that

$$\text{Cov}(W_j^{(i)}, W_j^{(k)}) = \{m-(k-i)\} \sigma_{\epsilon}^2$$

because $W_j^{(i)}$ and $W_j^{(k)}$ have $\{m-(k-i)\}$ observations in common. The other covariance terms are as derived in Theorem 5; which completes the proof. $\square$

Unfortunately, the U_j 's are not uncorrelated for different periods; that is, the U_j 's can be correlated between adjacent periods.

THEOREM 8. For $j = j' + 1$ and $i < k < 2i$,

$$\begin{aligned} \text{Cov}(U_j^{(i)}, U_{j'}^{(k)}) &= \\ \frac{\sigma_{\epsilon}^2}{\mathcal{E}(Y_j Y_{j'})} [\{m-(k-i)\} - \mathcal{E}(U_j^{(i)})_{(m-k)} - \mathcal{E}(U_{j'}^{(k)})_{(m-i)}]. \end{aligned}$$

Proof. $\text{Cov}(U_j^{(i)}, U_{j'}^{(k)}) =$

$$\frac{1}{\mathcal{E}(Y_j Y_{j'})} \text{Cov}[(W_j^{(i)} - \mathcal{E}(U_j^{(i)}) Y_j) (W_{j'}^{(k)} - \mathcal{E}(U_{j'}^{(k)}) Y_{j'})]$$

by the same techniques as used in Method 2.

$$\begin{aligned}
&= \frac{1}{\mathcal{E}(Y_j Y_{j'})} \{ \text{Cov}(W_j^{(i)} W_{j'}^{(k)}) - \mathcal{E}(U_j^{(i)}) \text{Cov}(W_{j'}^{(k)} Y_j) \\
&\quad - \mathcal{E}(U_{j'}^{(k)}) \text{Cov}(W_j^{(i)} Y_{j'}) + \mathcal{E}(U_j^{(i)} U_{j'}^{(k)}) \text{Cov}(Y_j, Y_{j'}) \} \\
&= \frac{\sigma_\epsilon^2}{\mathcal{E}(Y_j Y_{j'})} [\{m - (k-i)\} - \mathcal{E}(U_j^{(i)})(m-k) - \mathcal{E}(U_{j'}^{(k)})(m-1) + 0],
\end{aligned}$$

which completes the proof. $\square$

All other covariances are zero since

$$W_j^{(i)}, W_{j'}^{(k)}, Y_j, \text{ and } Y_{j'}$$

would not have any observations in common for the remaining terms in the variance-covariance matrix.

After the variance-covariance matrix $\sigma_\epsilon^2 \mathbf{V}^*$ for the $U_j^{(i)}$'s has been derived, a non-singular matrix $\mathbf{P}^*$ can be found such that

$$\mathbf{P}^* \mathbf{V} \mathbf{P}^{*'} = \mathbf{I}_{n(m-1)},$$

in order to obtain independent "observations" for the regression analysis.

Using the notation of Method 2,

$$A = P^*U^* = P^*C^*B + P^*N^* = DB + E$$

where N^* is the matrix of error terms. Thus,

$$\hat{B} = (D'D)^{-1} D'A$$

by Theorem 6. Then the estimated ratio is

$$\hat{U}_{n+1}^{(i)} = (1 \ i \ \dots \ i^p) \hat{B} = C_0 \hat{B}$$

and the estimated total is

$$\hat{Y}_{n+1}^{(i)} = \frac{W_{n+1}^{(i)}}{\hat{U}_{n+1}^{(i)}}$$

The estimated approximate variance of $\hat{U}_{n+1}^{(i)}$ is

$$S^2 \cdot (C_0'(C^*{}' V^{*-1} C^*)^{-1} C_0)$$

where S^2 is the residual mean square error. Thus, the estimated approximate variance of $(\hat{Y}_{n+1}^{(i)} - Y_{n+1})$ is

$$\hat{V}_{(i)}^2 = \frac{[\bar{Y}]^2}{(\hat{U}_{n+1}^{(i)})^2} S^2 \cdot (C_0'(C^*{}' V^{*-1} C^*)^{-1} C_0) + \frac{m \hat{\sigma}_\epsilon^2}{[\hat{U}_{n+1}^{(i)}]^2} + m \hat{\sigma}_\epsilon^2 - \frac{2i}{\hat{U}_{n+1}^{(i)}} \hat{\sigma}_\epsilon^2$$

and the estimation intervals of level α are

$$\hat{Y}_{n+1}^{(i)} \pm Z_{\alpha/2} \cdot \hat{V}_{(i)}.$$

Advantages of this method are:

- 1) the number of degrees of freedom in the residual mean square error is the same as for Method 2, and
- 2) the estimates are not affected as greatly by observations that deviate significantly from their expected values as in Method 2, especially during the first few sub-periods.

The disadvantages are:

- 1) the data must be transformed whenever the correlations between $V_j^{(i)}$'s are large,
- 2) calculations are much more difficult than for previous methods,
- 3) the inverse of V^* must be found which will be more difficult than for Method 2,

- 4) the estimates will not be affected as much by observations significantly above their "expected values" when the increase is due to a change in the system instead of random error, i.e., it is less sensitive to actual changes in trend, and
- 5) an estimate of σ_{ϵ}^2 must be found; note that the residual mean square from the regression analysis does not estimate σ_{ϵ}^2 . An estimate can be found using Methods 2-2 or 2-3.

COMPARISON OF ESTIMATORS

As in Chapter II, the estimators will be compared by comparing their variances. Since all the variances in this group are approximate and depend upon the estimate of the appropriate ratio, the comparisons may not always indicate the "best" method for a particular set of data.

In order to compare Method 1 to Method 2, note that for Method 1,

$$V_{(i)}^2 = \frac{1}{[R(i)]^2} \{ i \sigma_{\epsilon}^2 + [\mathcal{E}(Y_{n+1})]^2 \text{Var } \hat{R}(i) \} \\ + m \sigma_{\epsilon}^2 - \frac{2}{R(i)} i \sigma_{\epsilon}^2 ,$$

while for Method 2,

$$V_{(i)}^2 = \frac{1}{\{\mathcal{E}[U_{n+1}^{(i)}]\}^2} \{ i \sigma_{\epsilon}^2 + [\mathcal{E}(Y_{n+1})]^2 \text{Var } U_{n+1}^{(i)} \} + \\ + m \sigma_{\epsilon}^2 - \frac{2}{\mathcal{E}[U_{n+1}^{(i)}]} i \sigma_{\epsilon}^2 .$$

These formulas for the variance of the estimated totals are quite similar; since Method 2 provides many more degrees of freedom for the variance estimate, much greater precision will be attained by the use of Method 2. This

greater precision is attained at the sacrifice of the additional computations required in performing the necessary transformations to attain independence.

In comparing Methods 2 and 3, it is possible to construct examples where each method gives a smaller variance. However, if the covariances are positive, then it appears that Method 3 gives the smaller variance.

For Method 2, note that for subperiods at the beginning of the period, the estimate $\hat{U}_{n+1}^{(i)}$ will be small; thus, dividing by the square of $\hat{U}_{n+1}^{(i)}$ will make the variances of the estimated total large.

However, for Method 3, the estimate $\hat{U}_{n+1}^{(i)}$ will be only slightly less than one, thus the variance of the estimated total will be less than the variance from Method 2 provided the estimates of $\text{Var}(U_{n+1}^{(i)})$ do not differ greatly for Methods 2 and 3.

Thus, Method 3 will usually provide smaller variances than Method 2 for subperiods at the beginning of the period.

The choice between these two methods at the end of subperiods from the middle to the end of the period would depend more upon knowledge of the situation than on the variances. If there is no reason to suspect a drastic change in the trend, Method 3 will generally provide a

better estimate since the estimate is not influenced as drastically by observations with a large random error.

But, if the trend has had a radical jump at the beginning of the period, then Method 2 would be better since data from the previous period would be misleading.

CHAPTER IV

MONTE-CARLO STUDY

Each of the methods of estimation for cumulative totals described in Chapters II and III which require the use of regression programs are compared by Monte-Carlo techniques.

For each data set, monthly observations were generated for five previous years using the past model. In addition, one hundred repetitions were generated for the twelve months of the current year; this data was generated under the future model (possibly a different model than used for the past data). In each case, the model(s) used for the past and future data are presented as well as all necessary parameters for the distribution of the random errors.

The methods of estimation are denoted by chapter and number. For example, 2-1 is the first method of Chapter II.

For Method 2-2, the same intercept was used for each subperiod and in Methods 3-2 and 3-3, the estimate of the yearly total as given by Method 2-1 was used in calculating the variance of the estimator.

The percentage of 95% estimation intervals (EI) as formed at the end of the i -th subperiod containing the actual cumulative total for the current year is presented

as well as the variance of the estimator. In addition, the percentage of the estimated totals falling within three percent of the actual total (PI) is given. In each table, the following format is used for each method:

EI PI

VARIANCE .

At the top of each table all necessary information about the model(s) for the past and future data are given along with the distribution of the error terms. The following notation will be used:

$\epsilon \sim N(0,100)$ for a normal distribution with
zero mean and a variance of
100, and

$\epsilon \sim U(20)$ for a uniform distribution with
a mean of zero and a half-width
of 20.

TABLE 1

Past Model: $Y = 100 + 2T + \epsilon$ for $T = 1, \dots, 60$
 Future Model: $Y = 100 + 2T + \epsilon$ for $T = 61, \dots, 72$
 Error Terms: $\epsilon \sim N(0, 100)$

i	Method 2-1	Method 2-2	Method 2-3	Method 2-4	Method 3-2	Method 3-3
0	92 92 1803	94 92 1822	92 92 1905			
1		92 92 1670	92 92 1684	99 69 5830	97 49 24649	99 65 3982
2		94 96 1518	94 96 1474	100 65 9831	96 69 8270	99 79 3315
3		92 94 1366	92 97 1276	100 91 3455	91 78 4446	100 99 2726
4		96 100 1215	97 100 1089	72 89 1315	87 82 2809	100 100 2210
5		90 97 1063	92 99 913	93 86 3275	84 83 1918	99 99 1766
6		92 99 911	97 100 748	91 90 2354	86 92 1355	95 99 1390
7		86 99 759	96 100 595	77 97 1108	90 98 970	89 97 1081
8		92 100 607	91 100 454	99 100 748	92 100 690	92 100 837
9		94 100 455	90 100 323	98 100 584	91 100 478	90 100 656
10		88 100 304	86 100 204	82 100 215	96 100 311	39 99 535
11		96 100 152	93 100 96	80 100 97	95 100 177	21 100 474

TABLE 2

Past Model: $Y = 100 + 2T + \epsilon$ for $T = 1, \dots, 60$
 Future Model: $Y = 220 + 3(T-60) + \epsilon$ for $T = 61, \dots, 72$
 Error Terms: $\epsilon \sim N(0, 100)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	73 91 1053	91 91 2375	97 91 2803		
1		92 92 2177	95 92 2481	97 56 23479	00 00 6864
2		91 91 1979	93 91 2172	88 46 8055	00 00 5351
3		76 84 1781	88 88 1879	74 46 4372	00 00 4105
4		73 84 1583	83 89 1604	61 44 2777	01 00 3091
5		60 80 1385	71 89 1345	45 45 1898	23 12 2286
6		56 88 1187	70 93 1102	47 65 1342	74 79 1663
7		48 84 990	57 93 877	48 87 961	93 100 1202
8		74 100 792	70 100 668	55 95 683	98 100 884
9		88 100 594	80 100 476	77 100 472	100 100 693
10		100 100 396	91 100 301	88 100 306	100 100 615
11		100 100 198	97 100 142	99 100 172	75 100 636

TABLE 3

Past Model: $Y = 100 + 2T + \epsilon$ for $T = 1, \dots, 60$
 Future Model: $Y = 220 + 4(T-60) + \epsilon$ for $T = 61, \dots, 72$
 Error Terms: $\epsilon \sim N(0, 100)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	27 04 4301	01 03 1732	07 04 2303		
1		03 05 1588	03 02 2036	95 48 22892	06 01 2831
2		01 02 1444	02 02 1782	61 21 7982	04 01 2482
3		00 02 1299	01 02 1542	34 13 4356	02 02 2156
4		00 00 1155	00 02 1316	17 09 2776	02 02 1852
5		00 05 1011	00 07 1103	17 20 1900	03 07 1570
6		00 16 866	00 16 905	15 33 1345	04 14 1308
7		09 58 722	04 36 720	23 57 963	14 54 1065
8		07 64 577	03 59 548	33 81 684	10 66 839
9		07 84 433	07 86 391	55 99 472	14 85 632
10		24 100 289	24 100 247	82 100 305	36 100 441
11		71 100 144	61 100 117	93 100 170	100 100 266

TABLE 4

Past Model: $Y = 100 + 2T + \epsilon$ for $T = 1, \dots, 60$
 Future Model: $Y = 120 + 2T + \epsilon$ for $T = 61, \dots, 72$
 Error Terms: $\epsilon \sim N(0, 100)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	00 00 538	00 00 2592	00 00 3010		
1		00 00 2376	00 00 2661	97 47 27466	02 00 4009
2		00 00 2160	00 00 2329	99 71 8800	03 00 3342
3		00 00 1944	00 00 2016	88 76 4633	01 00 2752
4		00 00 1728	00 01 1720	82 76 2900	05 01 2235
5		00 02 1512	00 02 1442	72 75 1966	08 13 1791
6		01 07 1296	01 06 1183	74 88 1388	06 17 1413
7		03 15 1080	03 21 941	67 96 995	04 31 1100
8		09 53 864	09 63 717	75 100 712	20 82 848
9		29 97 648	21 98 511	87 100 500	84 100 655
10		45 100 432	28 100 323	98 100 333	100 100 518
11		61 100 216	46 100 152	100 100 200	100 100 434

TABLE 5

Past Model: $Y = 100 + 3T + \epsilon$ for $T = 1, \dots, 60$
 Future Model: $Y = 280 + 2(T-60) + \epsilon$ for $T = 61, \dots, 72$
 Error Terms: $\epsilon \sim N(0, 100)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	40 22 3640	07 18 1977	14 22 2374		
1		07 20 1812	11 22 2098	71 07 41702	00 00 22070
2		06 16 1647	08 22 1837	93 51 11225	00 00 15369
3		06 26 1483	06 26 1589	98 89 5491	00 00 10341
4		15 58 1318	09 45 1356	99 99 3345	00 00 6665
5		14 74 1153	08 59 1137	98 99 2194	00 00 4082
6		25 94 988	13 82 932	99 100 1535	00 00 2388
7		47 97 824	20 93 742	96 100 1103	00 00 1420
8		46 99 659	23 98 565	99 100 799	00 00 1043
9		32 100 494	22 100 403	99 100 575	00 00 1149
10		52 100 329	37 100 254	100 100 403	00 00 1652
11		94 100 165	86 100 120	98 100 267	100 100 2477

TABLE 6

Past Model: $Y = 100 + 3T + \epsilon$ for $T = 1, \dots, 60$
 Future Model: $Y = 280 + 1(T-60) + \epsilon$ for $T = 61, \dots, 72$
 Error Terms: $\epsilon \sim N(0, 100)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	14 16 2531	10 17 2142	11 16 2193		
1		05 17 1963	05 18 1939	70 04 38064	00 00 5463
2		08 27 1785	07 26 1697	80 45 10681	00 00 4382
3		06 28 1606	04 27 1469	88 70 5311	00 00 3468
4		01 16 1428	01 21 1253	89 85 3225	00 00 2705
5		04 32 1249	01 32 1051	93 94 2152	00 00 2079
6		05 41 1071	01 36 862	90 100 1509	00 00 1578
7		04 64 892	01 62 685	91 100 1084	00 00 1191
8		17 93 714	05 92 522	93 100 784	00 00 907
9		42 100 535	24 100 372	94 100 562	00 00 718
10		23 100 357	15 100 235	84 100 390	00 00 616
11		52 100 178	35 100 111	81 100 255	00 98 591

TABLE 7

Past Model: $Y = 100 + 2T + \alpha_i + \epsilon$ for $T = 1, \dots, 60$

i	1	2	3	4	5	6	7	8	9	10	11	12
α_i	-5	-5	0	5	10	20	30	30	20	10	5	0

Future Model: $Y = 100 + 2T + \alpha_i + \epsilon$ for $T = 61, \dots, 72$
 Error Terms: $\epsilon \sim N(0, 100)$

i	Method 2-1		Method 2-2		Method 2-3		Method 3-2		Method 3-3	
0	39	96	97	96	97	96				
	123		2400		2107					
1			97	93	99	99	93	17	00	00
			2200		1862		46122		3726	
2			84	84	96	98	91	52	00	00
			2000		1630		11554		3133	
3			75	81	95	98	50	20	00	00
			1800		1411		5538		2610	
4			69	81	99	99	21	08	00	00
			1600		1204		3305		2152	
5			67	87	96	100	13	13	05	08
			1400		1009		2180		1753	
6			60	84	93	100	21	31	06	19
			1200		828		1519		1407	
7			79	98	92	100	63	82	28	62
			1000		658		1088		1111	
8			99	100	95	100	89	100	66	99
			800		502		786		860	
9			100	100	93	100	97	100	90	100
			600		357		563		651	
10			98	100	93	100	98	100	99	100
			400		226		393		479	
11			96	100	94	100	99	100	94	100
			200		107		259		343	

TABLE 8

Past Model: $Y = 100 + 2T + \varepsilon$ for $T = 1, \dots, 60$
 Future Model: $Y = 100 + 2T + \varepsilon$ for $T = 61, \dots, 72$
 Error Terms: $\varepsilon \sim U(20)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	100 81 15837	95 82 2688	96 81 3490		
1		98 95 2464	99 89 3085	79 18 35566	86 00 14104
2		100 99 2240	100 97 2701	88 59 11306	70 00 10663
3		99 98 2016	98 97 2337	95 78 5983	88 06 7839
4		100 100 1792	99 99 1995	94 80 3755	82 19 5576
5		96 97 1568	97 98 1673	89 78 2547	93 59 3818
6		94 100 1344	100 100 1371	86 87 1794	98 90 2522
7		88 98 1120	94 99 1091	86 95 1282	100 100 1643
8		89 98 896	95 100 831	88 94 910	99 100 7144
9		96 100 672	96 100 592	88 100 629	99 100 990
10		79 100 448	85 100 374	90 100 409	86 99 1149
11		72 100 224	87 100 177	99 100 232	82 99 1591

TABLE 9

Past Model: $Y = 100 + 2T + \epsilon$ for $T = 1, \dots, 60$
 Future Model: $Y = 100 + 2T + \epsilon$ for $T = 61, \dots, 72$
 Error Terms: $\epsilon \sim U(30)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	99 68 15176	93 66 6627	96 68 8011		
1		94 60 6074	96 60 7081	87 21 72039	51 00 96982
2		98 73 5522	98 69 6199	92 41 24211	75 00 66698
3		99 81 4970	99 76 5365	99 58 13110	76 00 44087
4		99 79 4418	99 77 4578	96 61 8323	75 00 27622
5		98 80 3866	99 85 3839	93 74 5678	89 02 16106
6		96 82 3313	96 84 3147	93 75 4007	96 34 8584
7		100 92 2761	95 88 2504	95 89 2857	100 96 4293
8		100 100 2209	99 97 1907	98 96 2018	80 67 2626
9		99 100 1657	97 99 1359	97 100 1379	16 02 3090
10		97 100 1104	94 100 858	94 100 877	03 00 5287
11		87 100 552	89 100 405	90 100 471	00 00 8891

TABLE 10

Past Model: $Y = 100 + 2T + \epsilon$ for $T = 1, \dots, 60$
 Future Model: $Y = 100 + 2T + \epsilon$ for $T = 61, \dots, 72$
 Error Terms: $\epsilon \sim U(40)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	98 68 13731	98 65 12547	100 68 15738		
1		97 65 11593	100 67 13911	100 09 137085	85 06 22005
2		98 72 10540	98 71 12178	99 46 44838	92 14 18184
3		100 76 9486	100 75 10539	95 42 23917	91 27 14882
4		100 75 8432	100 75 8993	95 50 15068	96 41 12054
5		100 73 7378	100 83 7541	91 62 10248	100 76 9657
6		100 87 6324	100 88 6183	93 69 7232	100 73 7658
7		99 89 5270	100 91 4918	95 72 5172	100 84 6023
8		100 95 4216	100 92 3747	93 79 3681	100 86 4722
9		97 88 3162	99 96 2670	93 85 2549	100 94 3730
10		91 86 2108	99 100 1686	98 98 1664	100 98 3017
11		100 100 1054	100 100 796	100 100 951	100 100 2564

TABLE 11

Past Model: $Y = 100 + 2T + \epsilon$ for $T = 1, \dots, 12$
 and for periods $i=1, \dots, 5$
 Future Model: $Y = 100 + 2T + \epsilon$ for $T = 1, \dots, 12$
 Error Terms: $\epsilon \sim N(0, 100)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	34 24 701	85 34 2362	75 24 2339		
1		92 37 2165	74 28 2064	98 27 19558	92 35 2392
2		92 44 1969	67 27 1803	86 20 7271	92 39 2148
3		98 66 1772	79 39 1557	71 18 4074	96 49 1910
4		98 84 1575	86 50 1325	60 19 2630	100 71 1678
5		98 82 1378	85 55 1108	55 26 1811	99 79 1452
6		97 78 1181	94 68 906	58 32 1284	98 87 1231
7		96 77 984	93 84 719	61 34 917	98 95 1017
8		97 90 787	92 87 546	64 52 647	99 96 809
9		96 94 591	96 96 387	73 73 440	97 96 606
10		99 100 394	88 98 244	85 94 276	100 99 409
11		98 100 197	90 100 114	95 100 144	98 100 218

TABLE 12

Past Model: $Y = 100 + (.75 + .25i)T + \epsilon$ for $T = 1, \dots, 12$
 for periods $i=1, \dots, 5$
 Future Model: $Y = 100 + 2.5T + \epsilon$
 Error Terms: $\epsilon \sim N(0, 100)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	41 17 1186	52 21 1562	48 18 1447		
1		48 14 1432	42 12 1328	91 18 17490	99 10 7436
2		43 17 1302	43 17 1209	75 16 6887	99 17 5985
3		34 11 1171	34 14 1090	62 11 3938	99 14 4700
4		35 11 1041	36 12 972	42 08 2566	89 10 3587
5		28 14 911	40 20 850	37 12 1776	88 19 2645
6		28 14 781	45 28 727	40 12 1262	78 18 1877
7		18 13 651	33 23 605	42 20 902	60 16 1287
8		34 24 521	40 36 484	49 31 635	43 12 876
9		40 48 390	47 59 362	52 55 430	43 26 645
10		44 73 260	54 86 241	70 90 267	25 16 598
11		54 100 130	64 100 121	85 100 134	85 46 737

TABLE 13

Past Model: $Y = 100 + (.75 + .25i)T + \varepsilon$ for $T = 1, \dots, 12$
 for periods $i=1, \dots, 5$
 Future Model: $Y = 100 + 1.5T + \varepsilon$
 Error Terms: $\varepsilon \sim N(0, 100)$

i	Method 2-1	Method 2-2	Method 2-3	Method 3-2	Method 3-3
0	72 01 5336	38 01 2868	47 01 3408		
1		49 04 2629	50 03 2997	97 30 18387	39 00 4084
2		55 05 2390	56 05 2608	91 28 7048	29 01 3427
3		67 10 2151	72 12 2243	85 30 3990	51 03 2839
4		61 09 1912	76 15 1901	81 35 2589	53 04 2316
5		55 11 1673	71 19 1582	74 38 1788	58 08 1858
6		47 08 1434	78 23 1287	77 45 1269	44 07 1462
7		42 07 1195	84 42 1014	78 58 907	26 04 1126
8		50 13 956	89 61 765	83 68 640	38 10 846
9		66 32 717	96 90 539	84 84 435	43 18 624
10		70 58 478	98 99 336	87 94 272	77 71 455
11		94 99 239	97 100 156	96 100 140	100 100 337

Since Methods 3-2 and 3-3 have not been studied in detail before and involve approximations, the estimation interval percentages were compared to the theoretical values for Method 3-2 by Monte-carlo techniques. Since Method 3-3 involves basically the same theory and because it did not perform nearly as well, a comparison was not made for Method 3-3. Since Method 3-2 should perform the best when each period is a repetition of the previous period, i.e.,

$$X_{ji} = \mu + \beta i + \epsilon_{ji} \quad \text{for } \begin{matrix} i = 1, 2, \dots, m \\ j = 1, 2, \dots, n \end{matrix}$$

the comparison involving generating five periods of past data and a period of current data, or a total of six periods of data under the above model for each of the 25 repetitions. Three regression models were used to attempt to fit the fractions; linear, quadratic and cubic.

Standard deviations of 10 and 30 were used. In addition, a combination model is listed for $\sigma = 10$ where the cubic model was used for the first five subperiods and the quadratic for the remaining subperiods. In each case, the proportion of the estimation intervals containing the actual total are presented.

TABLE 14

Percent of intervals which contained the actual total for 25 repetitions of generated data applying Method 3-2 based on Table 11.

i	Linear $\sigma=10$	Quad $\sigma=10$	Cubic $\sigma=10$	Combination $\sigma=10$	Linear $\sigma=10$	Quad $\sigma=30$
1	92	92	100	100	92	96
2	88	80	100	100	80	84
3	76	76	90	90	80	84
4	68	76	90	90	88	92
5	76	76	80	90	84	92
6	80	88	80	88	88	92
7	80	84	80	84	92	92
8	84	96	70	96	96	96
9	84	92	70	92	88	92
10	92	96	50	96	96	96
11	100	88	40	88	92	92

CHAPTER V

CONCLUSIONS

In this chapter, the various methods will be compared based upon the Monte-Carlo results. In Table 1, Method 2-3 performed the best, based upon the percentages of the estimates within three percent of the actual value. Method 2-2 did almost as well but the variances of the estimator were slightly larger.

However, Method 3-2 did quite well, especially after the end of the third subperiod. Method 3-3 had the same percentage of the estimates within three percent of the actual total, but the variances of the estimates were large. Due to the size of the variances in Method 2-4, this method was not considered for any other situations. Table 1, of course, represents the best possible situation; a normal distribution for the errors and no change in the trend.

In Table 2, the trend has changed in that the slope changes from 2 to 3. (Note all changes in the trend occur at the beginning of the current year). Here, Method 3-3 really does poorly as compared to the other methods. When the slope of the trend changes from 2 to 4 as in Table 3, Method 3-2 does better than the other methods when using any of the criteria in the table, except the variances.

When the trend changes by a jump, but no change in slope as in Table 4, again Method 3-2 performed better than any of the others, which is not surprising since it was designed for just this situation.

In Table 5, only the slope was changed from 3 to 2. After the end of the second subperiod, Method 3-2 gave the best results, although the variances were larger than for Method 2-2. Method 2-2 apparently underestimated the variances, as can be seen by the estimation interval percentages.

When the trend changed from 3 to 1 as in Table 6, the results were almost the same as for Table 5.

In Table 7, monthly effects, α_i , were added as given at the top of the table. Method 2-3 gave the highest percentage of the estimates within three percent of the total, but Method 2-2 did almost as well.

In Tables 8, 9 and 10, the distribution of the error terms was changed from the normal to a uniform distribution to see how sensitive the methods were to normality. Methods 2-2, 2-3 and 3-2 still performed quite well under the uniform distribution.

In Tables 11, 12 and 13, the intercept was assumed to be the same for each period but the slope was varied from

period to period in Tables 12 and 13 while it remained constant for Table 11. In Table 12, the slope continued to increase for the current period while in Table 13 the slope decreased for the current period. Although Method 3-3 performed extremely well when there were no changes in the slope as in Table 11, it was the poorest in Table 13. In Tables 12 and 13, Method 3-2 performed at least as well as the other methods and considerably better in some cases.

Table 14 shows that the performance of Method 3-2 depends a great deal upon the regression model used. If the "correct" model is used the percentages are fairly close to 95%. The average percentage for the combination model was 92% which is fairly close to 95%. One problem with the Monte-carlo study is that upon looking at the data, if two or three observations at the beginning deviate from the mean in the same direction, then this method interprets this as a change and assumes the remaining observations will also be above the mean; thus, this method is affected more than the others studied. However, if the first observations were significantly bigger than usual, it would be reasonable to assume a change in the trend. This problem in predicting a change when there was actually no change was the main reason the percentages were smaller

than 95% near the beginning of the period. Even in these cases, this method recovered fairly quickly under one of the methods. In general, the cubic was better at the beginning and the linear at the end of the period.

In terms of variances and accuracy, Method 3-3 did not perform as well as did the others except in cases where there were no changes in the model for the past and the future. For this reason and the fact the transformation matrix is extremely difficult to calculate, Method 3-3 is not recommended.

Overall, Method 3-2 performed quite well under all circumstances, especially after the second subperiod. Some of the other methods, especially 2-2 and 2-3, performed quite well under some circumstances but did poorly under others. Therefore, unless one knows that the trend fits one of the situations under which Methods 2-2 and 2-3 performed extremely well, Method 3-2 appears to be the best procedure to use because it did almost as well as any other procedure regardless of the changes in the trend or in the distribution of the errors.

BIBLIOGRAPHY

1. Cochran, William G., Sampling Techniques, 2nd Ed., New York: John Wiley and Sons, Inc., 1963.
2. Draper, N.R. and Smith, H., Applied Regression Analysis, New York: John Wiley and Sons, Inc., 1968.
3. Durbin, J., "Applications of Quenouille's Method of Bias Reduction to the Estimation of Ratios", Biometrika, 46:477-80, 1959.
4. Graybill, Franklin A., An Introduction to Linear Statistical Models, New York: McGraw-Hill Book Company, Inc., 1961.
5. Hamilton, Martin A., "Regression Analysis When There Are Missing Observations", Technical Report, No. 1-3-75, Statistical Laboratory, Montana State University, Bozeman, Montana, 1975.
6. Quenouille, M.H., "Notes on Bias in Estimation", Biometrika, 43:353-60, 1956.
7. Rao, J.N.K., "Estimation of Ratios by Quenouille's Method", Biometrika, 52:647-49, 1965.
8. Rao, J.N.K. and Biegler, L.D., "Monte-Carlo Study of Ratio Estimators", Sankhya, 29:47-56, 1967.
9. Rao, J.N.K. and Webster, J., "Bias Reduction in Estimation of Ratios", Biometrika, 53:571-77, 1966.
10. Rao, P.S.R.S., "Comparison of Four Ratio-Type Estimates Under a Model", Jr. of Amer. Statistical Assoc., 64:574-80, 1969.
11. Rao, P.S.R.S. and Rao, J.N.K., "Some Small Sample Results for Ratio Estimators", Annals of Math Stat, 41:1141-42, 1970.
12. Schucany, W.R., Gray, H.L. and Owen, D.B., "Bias Reduction in Estimation", Jr. of Amer. Statistical Assoc., 66:524-33, 1971.

13. Searles, S.R., Linear Models, New York: John Wiley and Sons, Inc., 1971.
14. Shorack, G.R., "Testing and Estimating Ratios of Scale Parameters", Jr. of Amer. Statistical Assoc., 64:997-1013, 1969.
15. Sukhatme, P.V., Sampling Theory of Surveys with Applications, Iowa State College Press, Ames, Iowa, 1954.
16. Tin, M., "Comparison of Some Ratio Estimators", Jr. of Amer. Statistical Assoc., 60:294-307, 1965.
17. Zellner, A., "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Agregation Bias", Jr. of Amer. Statistical Assoc., 57: 348-68, 1962.

APPENDIX

```

C*****
C      SET F:5/REG DATA;OUTIN
C      SET F:6/R;OUTIN;REL
C      SET F:7/O;OUTIN;REL
C      SET F:8/H;OUTIN;REL
C      SET F:9/RE;OUT
C      SET F:10/CURRENT DATA;OUTIN
C      SET F:12/TEMP DATA;OUTIN;REL
C      SET F:14/TEMP DATA;OUTIN;REL
C      SET F:15/CONTROL FILE;IN
C      SET F:108 CARD PUNCH
C*****
C      MULTIPLE REGRESSION PROGRAM

C      USES SUBROUTS: PARAM,IREAD,TRAN,CORRE,SELECT,MINT,COE
F,RES,LOCATE
C      CONTROL CARDS READ ON 105, DATA 105, ERROR MESSAGES O
UT ON 108,
C      LISTED OUTPUT ON 6--BUT DUPLICATED ON 9, UNITS 8 &
9 ARE MISC.
C
C      NCOPY=MAX COPIES OF OUTPUT, MAXM=MAX SIZE OF REGRESSI
ON,
C      MAXI=MAX SIZE OF INPUT & TRAN ARRAY, MAXT=MAX NO. O
F TRANSFORMS
C
C      DOUBLE PRECISION R,RX,XBAR,STD,BOB,DET
COMMON R,RX,XBAR,STD,BOB,TOB,SB
C
C      NCOPY=9;MAXM=58;MAXI=99;MAXT=50
C      FOLLOWING DIMENSIONS DEPEND ON: 1711=58*(58+1)/2,1653
=(58-1)*58/2,
C      114=58*2 WHERE MAXM=58;100=99+1 WHERE MAXI=99; AND
MAXT=50
C      DIMENSION R(1711),RX(1653),XBAR(58),STD(58),BOB(58),B
(114)
C      DIMENSION ISAVE(58),TOB(100),T(99),SB(99),ITC(5,50),C
C(50)
C      DIMENSION IHEAD(20),IPAR(20),IA(8),IF5(37)
C      BOB IS DUMMY DOUBLE PRECISION B ARRAY +1 -- JOB IS T
+ /
C      EQUIVALENCE (BOB(2),B),(TOB(2),T)

```

```

DIMENSION ICCI(24), ICPI(24), BB(0:24), BC(0:24)
DO 6 I=1,24
  ICCI(I)=0
6  ICPI(I)=0
  SIGMA=0
  XW=0
  K=0
  OUTPUT 'INPUT CODE FOR DISTRIBUTION OF ERRORS' ,
1 'NORMAL=1' , 'UNIFORM=2'
  INPUT NCO
  OUTPUT 'INPUT METHOD OF ESTIMATION'
  INPUT NM
  GO TO (1,2),NCO
1  OUTPUT 'INPUT STANDARD DEVIATION'
  INPUT SIGMA
  OUTPUT 'INPUT # OF SDEVS FROM MEAN'
  INPUT XW
  GO TO 5
2  OUTPUT ' INPUT HALF-WIDTH OF INTERVAL'
  INPUT XW
5  CONTINUE
  OUTPUT 'INPUT RANDOM NUMBER GENERATOR'
  INPUT ISEED
  DATA ENDF/'ENDF'/
  OUTPUT ' INPUT NO OF REPS OF PAST DATA'
  INPUT NR
  OUTPUT 'INPUT # OF REPS OF CURRENT DATA/REP OF PAST'
  INPUT NRR
  OUTPUT 'INPUT # OF SUBPERIODS, DEGREE OF TREND'
  INPUT MPER, NTREND
  OUTPUT 'INPUT Z VALUE FOR CI, % DEV OF MEAN'
  INPUT ZCI, ZPI
  OUTPUT 'INPUT NO OF PERIODS'
  INPUT NPER
  IF(NM.NE.4) IM=0; GO TO 8
  OUTPUT 'INPUT SUBPERIOD #'
  INPUT IM
8  CONTINUE
  IF(NM.NE.7.AND.NM.NE.8) GO TO 9

```

```

OUTPUT 'INPUT 1 TO OUTPUT MATRICES'
INPUT ND
OUTPUT 'INPUT YHAT, VARIANCE'
INPUT YHAT1, VARIANCE
9  CONTINUE
   OUTPUT 'FOR PAST DATA'
   OUTPUT 'INPUT BETA FOR I=0, NTREND'
   DO 7 I=0, NTREND
     INPUT BB(I)
7  CONTINUE
   OUTPUT 'INPUT EFFECTS FOR SUBPERIODS 0 THRU M-1'
   DO 3 I=1, MPER-1
     INPUT XX
     J=I+NTREND
3  BB(J)=XX
     J=NTREND+MPER
     BB(J)=0
   OUTPUT 'INPUT NCH USING CODE'
   OUTPUT '0 - CONTINUE SAME TREND'
   OUTPUT '1 - DUPLICATE TREND FOR EACH PERIOD'
   OUTPUT '2 - DIFFERENT TREND FOR EACH PERIOD'
   INPUT NCH
   OUTPUT 'FOR PRESENT DATA'
   OUTPUT 'INPUT BETA FOR I=0, NTREND'
   DO 7000 I=0, NTREND
     INPUT BC(I)
7000 CONTINUE
   OUTPUT 'INPUT EFFECTS FOR SUBPERIODS 0 THRU M-1'
   DO 8000 I=1, MPER-1
     INPUT XX
     J=I+NTREND
8000 BC(J)=XX
     J=NTREND+MPER
     BC(J)=0
     DO 100 IR=1, NR
       NQQ=0
       IF (IR.EQ.NR) NQQ=1
       K=KALPHA
       REWIND 5, 10, 12, 14, 15
       CALL CSGEN(NCO, SIGMA, XW, K, ISEED, NTREND, NPER, MPER, BB, N
CH, NRR, BC)
       IF (NM.EQ.7) CALL CSTR7(MPER, ND, ND)
       IF (NM.EQ.8) CALL CSTR8(MPER, ND, ND)

```

```

C      INITIALIZATION & PARAMETER SETTING
10  END FILE 7; REWIND 6,7; IEND=1; N=0; T(0)=0.0; B(0)=0.
    REWIND 15,9
    CALL ABORTSET (13S,1)
86  FORMAT(/1H1)
    CALL PARAM(1,10,MAXT,IHEAD,IPAR,ITC,CC,IF5,IF5)
    DO 11 I=1,MAXI
11  T(I)=0.
    INN=5; IF(IPAR(19).NE.0) INN=105
    K=IPAR(5); M=IPAR(6);          END FILE 7; REWIND 7,8
    IF (M.EQ.0) M=K
    IF ((M.GT.MAXM).OR.(K.GT.MAXI).OR.(M.LT.2).OR.(K.LT.2
)) GO TO 13
    IF (IPAR(13).EQ.0) IPAR(13)=1
    IF (IPAR(12).NE.0) IPAR(12)=1
    J=K; IF (M.GT.K) J=M
    WRITE (6,12) IF5,IPAR(12),(I,I=1,J)
12  FORMAT(' CARD FORMAT',17A4/T2,20A4// ' OBS. ',7X, 'VARIA
BLE NO. '/
    1' NO. ',N(90(T5,I7,4I12/)),(T5,I7,9I12))
C      READ OBSERVATIONS
    CALL ABORTSET (14S,1)
20  N=N+1; CALL IREAD (IEND,0,K,N,IF5,IF5,B,IPAR(18),INN)
    IF (IEND.EQ.2) N=N-1; GO TO 30
    0IF((N.LE.IP(13)).OR.(IP(13).LT.0)) WRITE (6,21) N
,IPAR(12),
    1(B(I),I=1,K)
21  FORMAT(/T2,I3,N(90(T5,5G12.4/)),(T5,10G12.4))
22  FORMAT(' TRF',N(90(T5,5G12.4/)),(T5,10G12.4))
    IF (IPAR(7).LE.0) GO TO 23
    CALL TRAN(IEND,IPAR(7),MAXI,ITC,CC,B,T)
    IF (IEND.EQ.3) GO TO 14
    DO 24 I=1,M
24  T(I)=B(I)
    IF (IEND.EQ.2) N=N-1; IEND=1; GO TO 20
    0IF ((N.LE.IP(13)).OR.(IP(13).LT.0)) WRITE(6,22) I
PAR(12),
    1(B(I),I=1,M)
23  IF (IPAR(9).NE.0) WRITE(8) (B(I),I=1,M)
    CALL CORRE(IEND,M,N,IPAR(12),XBAR,STD,R,B)
    GO TO 20
C
C      RUNNING OFF REGRESSIONS
30  IF (IPAR(14).NE.0) IEND=6; GO TO 33
    IF (IPAR(15).NE.0) IEND=4

```

```

33 CALL CORRE(IEND,M,N,IPAR(12),XBAR,STD,R,B)
   JR=IPAR(8)
   IF (JR.EQ.0) JR=1
   DO 31 I=1,JR
   CALL ABORTSET (38S,1);IEND=1
   CALL SELECT (IEND,IPAR(8),NRES,M,K,KY,ISAVE,IA,R,RX)
   WRITE (6,88) IHEAD,IA
88  FORMAT(20A4,T1,1H1/T2,8A4//)
   IF (IEND.EQ.3) GO TO 38
   CALL SYMINV(IEND,RX,K,DET)
   IF (IEND.GT.1) WRITE(6,41) ISAVE(KY+IEND); GO TO 31
41  FORMAT(' X''X MATRIX BECAME SINGULAR AT VARIABLE',I3)
   IF (IEND.EQ.3) GO TO 38
   DO 37 II=1,KY
   IF (II.GT.1) WRITE (6,88) IHEAD,IA
   CALL COEF(K,KY,II,N,IPAR,ISAVE,XBAR,STD,R,RX,BOB,SP,N
M,M,PER,
   INTREND,ZCI,ZPI,IM,YHATI,VARIANCE,NRR,ICCI,ICPI,NQQ)
   IF (NRES.NE.0) WRITE(6,88) IHEAD,IA ;
   1CALL RES (N,K,KY,II,M,IPAR(14),NRES,ISAVE,XBAR,STD,BO
B,T,RX,SP)
   37 CONTINUE; CALL PRINT(N,K,KY,IPAR,ISAVE,IHEAD,R,RX,STD,
T,IA)
   GO TO 31
38  WRITE (6,39)
39  FORMAT(' ERROR IN REGRESSION SELECTION')
31  CONTINUE
C
C  MULTIPLE COPIES & ERROR RECOVERY
C
17  WRITE(6,86);WRITE(6,88) ENDF;NC=IPAR(11)
   IF (NC.LE.0) NC=1
   IF (NC.GT.NCOPY) NC=NCOPY
   CALL ABORTSET (10S,1)
   DO 16 I=1,NC
   REWIND 6
18  READ(6,87) (T(J),J=1,31)
   IF (T(1).EQ.ENDF) GO TO 16
   WRITE(9,87) (T(J),J=1,31);GO TO 18
16  CONTINUE; GO TO 100
87  FORMAT(31A4)
13  WRITE(108,93) IHEAD;GO TO 19
93  FORMAT(20A4,T1,////' ERROR IN PARAMETERS, PROBLEM IGN
ORED')

```

```

14 WRITE(108,94) IHEAD
94 FORMAT(20A4,T1,////' ERROR IN TRANSFORMATION OR READ-
IN')
19 CALL IREAD(3,K,K,N,IF5,IF5,B,IPAR(18)); GO TO 100
100 CONTINUE
WRITE(108,101) NM
101 FORMAT(2/,20X,'METHOD',I2)
ICC=1;IIP=12
IF(NM.EQ.1) IIP=1
IF(NM.EQ.4) ICC=IM; IIP=IM
IF(NM.EQ.7.OR.NM.EQ.8) IIP=11
DO 103 IC=ICC,IIP
XICCI=ICCI(IC)
XICCI=XICCI/NR*(1./NRR)
YICPI=ICPI(IC)
YICPI=YICPI/NR*(1./NRR)
WRITE(108,102) IC, XICCI,YICPI
102 FORMAT(2X,'SUBPERIOD=',I3,3X,'CI FRACTION =',
1F6.4,4X,'PI FRACTION =',F6.4)
103 CONTINUE
END

```

C

```

SUBROUTINE PARAM (IO,IP,IT,IH,IPAR,ITC,CC,L,IF5)
DIMENSION IA(20),IPAR(1),IH(1),ITC(5,1),CC(1),L(21),I
F5(1),KARD(8)
DATA KARD/'F','L','P','R','T','0',' ','C'/
DO 1 I=1,37
IF (I.LE.20) IH(I)=KARD(7)
IF (I.LE.21) L(I)=KARD(7)
IF (I.LE.IP+10) IPAR(I)=0
1 IF5(I)=KARD(7);NHEAD=0
30 FORMAT(20A4,T1,A1,T6,A1,T2,4I1, T13,17A4)
31 FORMAT(T11,7(2A4,I2))
32 FORMAT(T11,35I2)
33 FORMAT(T6,I1,I2)
34 FORMAT(T11,5I2,F10.5)
10 READ(15,30,END=12) IA,J,K;REWIND 8;WRITE(8,30) IA;RE
WIND 8
IF (J.EQ.KARD(1)) GO TO 21
IF (J.EQ.KARD(3)) READ(8,32) (IPAR(I+10),I=1,IP); GO T
O 10
IF (J.EQ.KARD(1)) GO TO 21
IF ((J.EQ.KARD(2)).AND.(IO.EQ.2)) IPAR(9)=1; GO TO 22
IF ((J.EQ.KARD(4)).AND.(IO.EQ.1)) IPAR(8)=IPAR(8)+1;G
O TO 23

```

```

      0 IF ((J.EQ.KARD(5)).AND.(IPAR(7)+1.LE.IT)) K=IPAR(7)+1
;
      1WRITE (6,30) KARD(7),(IA(I),I=1,19); GO TO 24
      IF ((J.EQ.KARD(8)).AND.(IO.EQ.1)) GO TO 25
      NHEAD=NHEAD+1
      IF (NHEAD.EQ.1) READ(8,30) (IH(I),I=1,20)
      IF (NHEAD.LE.10) WRITE(6,30) IA,KARD(7)
      GO TO 10
21 READ(8,30) IA,J,K,(IPAR(I),I=1,4),          (IF5(I),I=1
,17)
      IPAR(5)=IPAR(1)*10+IPAR(2); IPAR(6)=IPAR(3)*10+IPAR(4
)
      IF (K.EQ.KARD(6).OR.K.EQ.KARD(7)) RETURN
      READ(15,30) (IF5(I),I=18,37); RETURN
22 READ(8,31) (L(I),L(I+7),L(I+14),I=1,7); GO TO 10
23 WRITE(7,30) IA;REWIND 8;READ(8,33) I,J;IF (J.NE.0) IP
AR(9)=J
      IF (I.NE.0)          READ(105,30) IA;WRITE(7,30) IA
      GO TO 10
24 IPAR(7)=K;READ(8,34) (ITC(I,K),I=1,5),CC(K); GO TO 10
25 WRITE (7,30) IA;REWIND 8;READ(8,33) I
26 I=I-1;IF (I.LT.0) GO TO 10
      READ (15,30) IA; WRITE (7,30) IA; GO TO 26
12 CALL EXIT
      END

C
C      NORMAL RANDOM VARIABLE
C
      FUNCTION RNORM(IY,XMEAN,SIGMA,XMIN,XMAX)
      CALL MYRAN(IY,RA)
      CALL MYRAN(IY,RB)
      V=(-2.0*ALOG(RA))*0.5*COS(6.283*RB)
      RNORM=V*SIGMA+XMEAN
      IF(RNORM-XMIN) 6,7,8
6      RNORM=XMIN
7      RETURN
8      IF(RNORM-XMAX) 7,7,9
9      RNORM=XMAX
      RETURN
      END

```

```

      SUBROUTINE CHOL(IEND,A,N,DET)
C      CHOLESKY DECOMPOSITION OF A POSITIVE DEFINITE MATRIX
      STORED AS
C      AN UPPER TRIANGLE, COLUMN BY COLUMN
      DOUBLE PRECISION A( 1),DET,SUM,PIVOT,DSQRT
      INTEGER SUB
      DET=1.0
      DO 25 I=1,N
      LOOP=I-1
      DO 20 J=I,N
      SUM=0.0
      IF(LOOP.LE.0) GO TO 10
      DO 5 K=1,LOOP
        5 SUM=SUM+A(SUB(K,I))*A(SUB(K,J))
      10 SUM=A(SUB(I,J))-SUM
      IF (I.NE.J) GO TO 15
C      TEST FOR FAILURE OF POSITIVE DEFINITENESS
      IF(SUM.LT.1.D-13) GO TO 30
C      TRANSFORM PIVOT
      PIVOT=DSQRT(SUM)
      A(SUB(I,J))=PIVOT
      DET=DET*PIVOT
      PIVOT=1.0/PIVOT
      GO TO 20
C      COMPUTE OFF-DIAGONAL TERMS
      15 A(SUB(I,J))=SUM*PIVOT
      20 CONTINUE
      25 CONTINUE
C      COMPUTE DET(A)=DET(L)*DET(L')
      DET=DET*DET
      RETURN
      30 DET=0.0
      IEND=I
      RETURN
      END

```

```

SUBROUTINE SYMINV( IEND, A, N, DET)
C   SUBROUTINE TO INVERT A POSITIVE DEFINITE MATRIX BY CH
OLESKY METHOD
      DOUBLE PRECISION A( 1), DET, SUM
      INTEGER SUB
C   FACTOR A INTO A=(LL')
      CALL CHOL( IEND, A, N, DET)
      IF (DET.EQ.0.0) GO TO 30
C   INVERT UPPER TRIANGULAR MATRIX (L'). THIS MUST BE DON
E FROM
C   COLUMN N,N-1.....1 SO AS NOT TO DESTROY NEEDED ELEMEN
TS.
      DO 15 L=1,N
      K=N+1-L
      A(SUB(K,K))=1.0/A(SUB(K,K))
      DO 10 M=L,N
      J=N+1-M
      IF(J.EQ.K) GO TO 10
      SUM=0.0
      J1=J+1
      DO 5 I=J1,K
5      SUM=SUM-A(SUB(J,I))*A(SUB(I,K))
      A(SUB(J,K))=SUM/A(SUB(J,J))
10     CONTINUE
15     CONTINUE
C   NOW A CONTAINS L TRANSPOSE INVERSE
C   MULTIPLY INVERTED LOWER TRIANGLES TOGETHER IN REVERSE
ORDER
      DO 25 I=1,N
      DO 25 J=I,N
      SUM=0.0
      DO 20 K=J,N
20      SUM=SUM+A(SUB(I,K))*A(SUB(J,K))
      A(SUB(I,J))=SUM
25     CONTINUE
30     RETURN
      END

```

```

INTEGER FUNCTION LOCATE(I,J)
IF ((I-J).LT.0) LOCATE=I+(J*J-J)/2; RETURN
LOCATE=J+(I*I-I)/2; RETURN
END

```

C
C
C

```

SUBROUTINE SELECT (IEND,L,N,M,K,KY,ISAVE,IA,R,RX)
DOUBLE PRECISION R(1),RX(1)
DIMENSION ISAVE(1),IA(8)
DATA IB,IR/'      R'/
DO 1,I=1,8
1  IA(I)=IB; ISAVE(1)=0; ISAVE(2)=0; KY=0; IJ=0; N=0
   IF (L.GT.0) READ(7,4) ID,J,N,KY,IA,(ISAVE(I),I=1,20)
   IF ((L.GT.0).AND.(ID.NE.IR)) IEND=3; RETURN
   IF ((L.GT.0).AND.(J.NE.0)) READ(7,5) (ISAVE(I),I=21,M
)
4  FORMAT(A1,T6,I1,2I2,7A4,A2,20I2)
5  FORMAT(40I2)
   IF (KY.LE.0) KY=1
   IF (KY.GT.(M-1)) IEND=3; RETURN
   IF (ISAVE(2).NE.0) GO TO 3
   IF (ISAVE(1).EQ.0) ISAVE(1)=1
   J=0
   DO 21 I=2,M
   J=J+1
   IF (J.EQ.ISAVE(1)) J=J+1
21  ISAVE(I)=J
3  K=M-KY
   DO 6 I=1,M
   IF (ISAVE(I).LE.0) K=I-KY-1; GO TO 7
   IF (ISAVE(I).GT.M) IEND=3; RETURN
6  CONTINUE
7  IF (K.LE.0) IEND=3; RETURN
   DO 22 I=1,K
   DO 22 J=1,I
   IJ=IJ+1
22  RX(IJ)=R(LOCATE(ISAVE(I+KY),ISAVE(J+KY))); RETURN
END

```

```

SUBROUTINE COEF(K,KY,II,N,IPAR,ISAVE,XBAR,STD,R,RX,B,
SE
1,NM,MPER,NTREND,ZCI,ZPI,IM,YHAT1,VARIANCE,NRR,ICCI,IC
PI,NQQ)
DOUBLE PRECISION XBAR,STD,R,RX,C,SME,B,BO
DIMENSION IPAR(1),ISAVE(1),XBAR(1),STD(1),R(1),B(1),R
X(1)
DIMENSION ICCI(1),ICPI(1)
C FIND REG COEF,R-SQUARED, INTERCEPT--PLACED IN B(K+1)
RM=0.0;BO=0.0;NY=ISAVE(II);NN=N-1;IF (IPAR(14).NE.0)
NN=N
F=0.0;SE=0.0;SME=0.0
DO 1 J=1,K
B(J)=0.0;NX=ISAVE(J+KY)
DO 2 I=1,K
2 B(J)=B(J)+R(LOCATE(NY,ISAVE(I+KY)))*RX(LOCATE(J,I))
RM=RM+B(J)*R(LOCATE(NY,NX));B(J)=B(J)*STD(NY)/STD(NX)
1 BO=BO+B(J)*XBAR(NX)
B(K+1)=XBAR(NY)-BO;IF (IPAR(14).NE.0) B(K+1)=0.0
C FIND SUM &MEAN OF SQUARES DUE TO REG, ERROR, TOTAL
SST=STD(NY)*STD(NY)*(N-1);SSR=RM*SST;SSE=SST-SSR;SMR=
SSR/K
IF ((SSE.LE.0.0).OR.(NN.LE.K)) SSE=0.0; GO TO 27
SME=SSE/(NN-K);F=SMR/SME;SE=DSQRT(SME)
C WRITE-OUT COEFFICIENTS, ETC.
27 WRITE (6,23)
DO 3 I=1,K
C=RX(I*(I+1)/2);L=ISAVE(KY+I);C=C*SME/(STD(L)*STD(L)
)/(N-1)
SB=DSQRT(DABS(C));PR=B(I)/DSQRT(C*(NN-K)+B(I)*B(I));T
=0.0
IF (SE.GT.0.0) T=B(I)/SB
3 WRITE (6,24) L,XBAR(L),PR,B(I),SB,T
WRITE (6,25) NY,XBAR(NY),B(K+1),RM,SE
IF (IPAR(14).NE.0) WRITE (6,28)
NE=NN-K;WRITE (6,26) K,SSR,SMR,F,NE,SSE,SME,N,NN,SST
230FORMAT(' VAR.',T13,'MEAN',T24,'PARTIAL',T36,'REGRESSI
ON',T49,'STD.
1 ERROR',T63,'COMPUTED'/' NO.',T25,'CORR.',T38,'COEF.
',T49,'REGR.
2COEF.',T63,'T VALUE')
24 FORMAT(T3,I2,3X,5G13.4)

```

```

250FORMAT('DEPENDENT'/T3,I2,3X,G13.4//T3,'INTERCEPT = ',
T21,G13.4/T3,
1'R SQUARED = ',T21,G13.4/T3,'STD.ERROR(SY.X) = ',T21,G1
3.4)
260FORMAT('//////// T17,'ANALYSIS OF VARIANCE FOR THIS
REGRESSION'/
1//T6'SOURCE',T26,'D.F.',T38,'S.S.',T51,'M.S.',T62,'F
VALUE'/' DUE
2 TO REGRESSION',T24,I5,T34,3G13.6/' RESIDUAL',T24,I5,
T34,2G13.6/
3' TOTAL ( N = ',I5,' )',T24,I5,T34,G13.6)
28 FORMAT ('/' NOTE THAT THE OPTION OF 'NO COMPUTER SU
PLIED INTERC
1EPT' WAS USED')
C CALCULATION OF CONTRASTS
CALL INTER(K,RX,B,SE,NE,XBAR,ISAVE,STD,KY,N,ICCI,ICP
I
1,NM,MPER,NTREND,ZCI,ZPI,IM,YHAT1,VARIANCE,NRR,NQQ)
DIMENSION IT(8),CON(58)
DATA IC,IR/'C R'/
JJ=0;IF (II.GE.2) GO TO 47
34 READ (7,31,END=5) I;BACKSPACE 7;L=K+1;IF (I.NE.IC) R
ETURN
31 FORMAT(A1,T6,I1,T11,7A4,A2,8F5.0)
DO 33 I=1,L
33 CON(I)=0.
READ (7,31,END=5) I,I,IT,(CON(J),J=1,8)
IF (I.EQ.1) READ (7,32,END=5) (CON(J),J=9,24)
IF (I.EQ.2) READ (7,32,END=5) (CON(J),J=9,40)
IF (I.EQ.3) READ (7,32,END=5) (CON(J),J=9,56)
IF (I.EQ.4) READ (7,32,END=5) (CON(J),J=9,58)
32 FORMAT(16F5.0)
47 CL=0.0;SL=0.0
DO 42 I=1,K
CL=CL+B(I)*CON(I+ 1)
DO 42 J=1,K
XX=RX(LOCATE(I,J))/STD(ISAVE(I+KY))/STD(ISAVE(J+KY))/
(N-1)
42 SL=SL+CON(I+ 1)*CON(J+ 1)*XX
CL=CL+CON(II)*XBAR(ISAVE(II))
SL=DSQRT((CON(II)*CON(II)/N +SL)*SME)
JJ=JJ+1; IF (JJ.EQ.1) WRITE (6,46)
46 FORMAT('CALCULATION OF CONTRASTS AND PREDICTED VALUE
S' /)

```

```

WRITE (6,43) IT, ISAVE(II),(ISAVE(I+KY),I=1,K)
WRITE (6,44) CON(II),(CON(I+1 ),I=1,K)
WRITE (6,45) CL,SL;IF (II.GE.2) RETURN
GO TO 34
43 FORMAT(/T2,8A4/' VARIABLES & CONTRAST WEIGHTS -- '/(16
,9I12))
44 FORMAT(10G12.4)
45 FORMAT('/CONTRAST = ',G12.4,10X,'STD. ERROR = ',G12.4)
5 RETURN
END

SUBROUTINE CSCONV(INN,IOUT,NOUT)
C
C PROGRAM TO PUT NEEDED INFORMATION ON DATA FILE
C
DIMENSION X(12),A(12,12),M(12),NT(12),Z(12)
10 DO 30 J=1,12
READ(INN,20,END=80) X(J),(A(J,I),I=1,12),M(J),NT(J)
20 FORMAT(F5.0,12I1,2I2)
IF(X(J).LT.0) STOP
30 CONTINUE
DO 40 I=1,12
Z(I)= 0
DO 40 J=1,I
40 Z(I)=Z(I) + X(J)
DO 60 I=1,11
IO=IOUT
45 WRITE(IO,50) X(I),(A(I,J),J=1,12),M(I),NT(I),Z(I+1)
1,Z(I),Z(12),I,I+1
50 FORMAT(15,12I1,2I2,3I5,2I2,T60,'1')
IF(NOUT.NE.1) GO TO 60
WRITE(14,50) X(I),(A(I,J),J=1,12),M(I),NT(I),Z(I+1)
1,Z(I),Z(12),I,I+1
60 CONTINUE
IO=IOUT
65 WRITE(IO,70) X(12),(A(12,J),J=1,12),M(12),NT(12),
1Z(12),Z(12)
70 FORMAT(15,12I1,2I2,5X,2I5,2X,' 1',T60,'1')
IF(NOUT.NE.1) GO TO 10
WRITE(14,70) X(12),(A(12,J),J=1,12),M(12),NT(12)
1,Z(12),Z(12)
GO TO 10
80 CONTINUE
RETURN
END

```

SUBROUTINE CSTR7(M,ND,NDD)

```

C
C   PROGRAM TO TRANSFORM DATA FOR METHOD 7
C   DIMENSION X(60),XT(60),S(1540),SS(1540)
C   1,XNT(12),NT(12),R(1541),RR(1541),YT(12)
C   2,SSS(11,11)
3     DO 4 I=1,M
4       XT(I)=0
5       DO 20 I=1,M
        READ(14,10,END=30) X(I)
10      FORMAT(T27,F5.0)
        XT(I)=XT(I)+X(I)
20     CONTINUE
        GO TO 5
30     DO 40 I=1,M-1
40     XT(I)=XT(I)/XT(M)
        N=M-1
        R(1541)=0
45     DO 55 I=1,M-1
        DO 50 J=1,M-1
            II=I
            IF(J.LT.I) II=J
            S(LOCATE(I,J))=(II-XT(I)*J-XT(J)*I+XT(I)*XT(J)*M)/X(M)
        )**2
        SSS(I,J)=S(LOCATE(I,J))
50     R(LND(I,J))=0
55     R(LND(I,I))=1
        IF(ND.NE.1) GO TO 60
        DO 57 I=1,M-1
57     WRITE(108,115) (S(LOCATE(I,J)),J=1,11)
60     RR(LND(1,1))=1
        SS(LOCATE(1,1))=S(LOCATE(1,1))
        DO 65 J=2,N
        RR(LND(1,J))=0
65     SS(LOCATE(1,J))=0
        RR(1541)=0
        DO 95 K=1,N-1
        DO 80 I=K+1,N
            XM=S(LOCATE(I,K))/S(LOCATE(K,K))
            DO 70 J=1,N
            RR(LND(I,J))=R(LND(I,J))-R(LND(K,J))*XM
70     SS(LOCATE(I,J))=S(LOCATE(I,J))-S(LOCATE(K,J))*XM
80     SS(LOCATE(I,K))=0
        DO 95 I=1,N

```

```

DO 95 J=1,I
S(LOCATE(I,J))=SS(LOCATE(I,J))
95  R(LND(I,J))=RR(LND(I,J))
DO 100 I=1,N
DO 100 J=1,I
100  R(LND(I,J))=R(LND(I,J))/
1SQRT(ABS(S(LOCATE(I,I))))
IF(ND.NE.1) GO TO 116
DO 110 I=1,N
110  WRITE(108,115) (R(LND(I,J)),J=1,11)
115  FORMAT(6G12.6)
116  CONTINUE
REWIND 14
117  READ(14,120,END=160) (NT(I),XT(I),I=1,M)
120  FORMAT(T18,I2,T27,F5.0)
IF(ND.EQ.1) WRITE(108,120) (NT(I),XT(I),I=1,M)
DO 130 I=1,M-1
130  XT(I)=XT(I)/XT(M)
DO 145 I=1,M-1
YT(I)=0
XB=0
XNT(I)=0.
DO 140 J=1,M-1
XNT(I)=XNT(I)+R(LND(I,J))*NT(J)
XB=XB+R(LND(I,J))
140  YT(I)=YT(I)+R(LND(I,J))*XT(J)
145  WRITE(5,150) YT(I),XNT(I),XB
150  FORMAT(3F26.8)
GO TO 117
160  CONTINUE
IF(ND.NE.1) GO TO 170
READ(10,120) (XT(I),I=1,M)
WRITE(108,120) (XT(I),I=1,M)
REWIND 10
170  CONTINUE
REWIND 5
REWIND 14
RETURN
END

```

SUBROUTINE CSTR8(M,ND,NDD)

C

C

PROGRAM TO GENERATE INDEPENDENT DATA FOR METHOD 8

DIMENSION X(60),XT(5),S(1540),SS(1540)

1,XXT(44),R(1541),RR(1541),YT(44),W(12)

2,NT(44),XNT(44)

DO 180 I=1,5

180 XT(I)=0.

DO 185 I=1,M

READ(14,183) X(I)

183 FORMAT(F5.0)

185 XT(I)=XT(I)+X(I)

DO 188 I=1,M

188 W(I)=0.

J=2

189 CONTINUE

DO 190 I=1,M

READ(14,183,END=200) X((J-1)*M+I)

190 XT(J)=XT(J)+X((J-1)*M+I)

J=J+1

GO TO 189

200 Z=XT(1)

J=J-1

KK=1

KKK=0

NY=J

DO 210 K=1,J-1

DO 210 I=1,M

Z=Z+X(I+M*K)-X(I+M*(K-1))

IF(I.EQ.M) GO TO 205

XXT(KK)=Z

KKK=KKK+1

IF(KKK.GT.11) KKK=1

NT(KK)=KKK

KK=KK+1

205 W(I)=W(I)+Z

210 CONTINUE

DO 220 I=1,M-1

220 W(I)=W(I)/W(M)

W(M)=W(M)/J

```

C
C   TO CALCULATE VAR-COV MATRIX
C
R(1541)=0.
DO 240 I=1,44
DO 230 JJ=1,I
S(LOCATE(I,JJ))=0
230 R(LND(I,JJ))=0
240 R(LND(I,I))=1
RR(LND(1,1))=1
DO 250 JJ=1,J-1
DO 250 II=1,M-1
DO 250 III=1,II
KV=ABS(II-III)
KI=LOCATE((JJ-1)*(M-1)+II,(JJ-1)*(M-1)+III)
S(KI)=(M-KV-W(II)*III-W(III)*II+W(II)*W(III)*M)
250 S(KI)=S(KI)/W(M)**2
DO 260 JJ=1,J-1
DO 260 II=1,M-1
DO 260 III=1,M-1
KI=(JJ-1)*(M-1)+II
KJ=JJ*(M-1)+III
KV=ABS(KI-KJ)
IF(KV.GE.11) GO TO 260
KV=ABS(II-III)
S(LOCATE(KI,KJ))=(M-KV-W(II)*(M-III)-W(III)*
1(M-II))/W(M)**2
260 CONTINUE
IF(ND.NE.1) GO TO 280
DO 270 II=1,44
DO 270 III=1,II
XX=S(LOCATE(II,III))
IF(XX.EQ.0) GO TO 270
WRITE(108,115) II,III,XX
270 CONTINUE
280 CONTINUE
SS(LOCATE(1,1))=S(LOCATE(1,1))
N=(J-1)*(M-1)
DO 65 J=2,N
RR(LND(1,J))=0
65 SS(LOCATE(1,J))=0

```

```

RR(1541)=0
DO 95 K=1,N-1
DO 80 I=K+1,N
DO 70 J=1,N
XM=S(LOCATE(I,K))/S(LOCATE(K,K))
RR(LND(I,J))=R(LND(I,J))-R(LND(K,J))*XM
SS(LOCATE(I,J))=S(LOCATE(I,J))-S(LOCATE(K,J))*XM
70 CONTINUE
80 SS(LOCATE(I,K))=0
DO 95 I=1,N
DO 95 J=1,I
S(LOCATE(I,J))=SS(LOCATE(I,J))
95 R(LND(I,J))=RR(LND(I,J))
DO 100 I=1,N
DO 100 J=1,I
100 R(LND(I,J))=R(LND(I,J))/
1 SQRT(ABS(S(LOCATE(I,I))))
IF(NDD.NE.1) GO TO 116
DO 116 I=1,N
DO 116 J=1,N
IF(R(LND(I,J)).EQ.0.) GO TO 116
WRITE(108,115) I,J,R(LND(I,J))
115 FORMAT(2I3,2X,E.3)
116 CONTINUE
KK=0
DO 130 I=2,NY
DO 120 J=1,M-1
120 XXT(KK+J)=XXT(KK+J)/XT(I)
130 KK=KK+M-1
DO 160 I=1,N
YT(I)=0.
XB=0
XNT(I)=0.
DO 140 J=1,N
XNT(I)=XNT(I)+R(LND(I,J))*NT(J)
XB=XB+R(LND(I,J))
140 YT(I)=YT(I)+R(LND(I,J))*XXT(J)
WRITE(5,150) YT(I),XNT(I),XB
150 FORMAT(3F26.8)
160 CONTINUE
REWIND 5
REWIND 14
RETURN
END

```

```

FUNCTION PRBF( DA, DB, FR)
PRBF=1.0
IF(DA*DB*FR.EQ.0.0) RETURN
IF(FR.LT.1.0) PRBF=.5; RETURN
A=DA
B=DB
F=FR
AA=2.0/(9.0*A)
BB=2.0/(9.0*B)
Z1=((1.0-BB)*F**0.333333-1.0+AA)/SQRT(BB*F**0.666667+
AA)
Z=ABS(Z1)
IF(B.LT.4.) Z=Z*(1.0+0.08*Z**4/B**3)
PRBF=0.5/(1.0+Z*(.196854+Z*(.115194+Z*(.000344+Z*.019
527))))**4
IF(Z1.LT.0.0) PRBF=1.0-PRBF
RETURN
END

```

C
C
C

```

SUBROUTINE IVAL(SB,RX,D,N,NP,VAR,KY,ISAVE,STD,NC)
DOUBLE PRECISION RX,STD
DIMENSION STD(1),XXI(365),ISAVE(1),RX(1),B(1),D(1)
NDF=N-1
XZ=0.
DO 20 I=1,NP
DO 20 J=1,I
X= RX(LOCATE(J,I))/STD(ISAVE(J+KY))/STD(ISAVE(I+KY))/
NDF
XXI(LOCATE(J,I))=X
20 CONTINUE
DO 25 I=1,NP
DO 25 J=1,NP
25 XZ=XZ+XXI(LOCATE(J,I))*D(I)*D(J)
IF(NC.EQ.7.OR.NC.EQ.8.OR.NC.EQ.4) GO TO 30
XZ=XZ+1
30 VAR=(SB**2)*XZ
RETURN
END

```

```

SUBROUTINE RPROB (IDT,TS,N,K)
C PROGRAM TO FIND P-VALUE(L.E.) FOR STANDARDIZED RESIDUAL
C
DOUBLE PRECISION S,SS,Z,Z2,C
PRECISION = 0.001
MAXTERM = 500
WRITE (6,2) IDT
2 FORMAT(/' FOR',I4,'TH T-RESIDUAL,')
T=TS
IF (N-K.LE.2) GO TO 8
IF (N.GT.100) GO TO 8
Z=SQRT(T*T/(N-K))
IF (Z.GE.1.0) GO TO 8
C
C FIND CONSTANT
C=1.0
H=(N-K)/2.0
IH=H+0.1
IH4=H-0.4
II=IH-1
IF (II.LE.0) GO TO 22
DO 21 I=1,II
21 C=C*(H-I)/(H-I-0.5)
22 IF (IH.EQ.IH4) C=C*0.5
IF (IH.NE.IH4) C=C/3.141593
RERROR2=(PRECISION/(2.0*C*N))**2
C
C EVALUATING THE INTEGRAL
S=Z
Z2=Z*Z
E=(N-K-3.0)/2.0
I=1
10 Z=-Z*E*Z2/I
SS=Z/(2*I+1)
S=S+SS
I=I+1
E=E-1.0
IF (SS.GT.0.0) GO TO 10
RERROR2=SS*SS
IF (I.GT.MAXTERM) GO TO 8
IF (RERROR2.GT.RERROR2) GO TO 10
ALPHA=N*(0.5-S*C)*2.0
IF (ALPHA.GT.0.3) GO TO 5
IF (ALPHA.LE. 0.0) GO TO 8
WRITE (6,11) ALPHA; RETURN

```

```

11 FORMAT(' 2-TAILED ALPHA(UPPPER BOUND) = ',F5.3)
5 WRITE (6,4);RETURN
4 FORMAT(' UPPER BOUND ON PROBABILITY EXCEEDS 0.30, ')
8 WRITE (6,7);RETURN
7 FORMAT(' PROBABILITY CANNOT BE EVALUATED.')
```

100 CONTINUE
END

C
C
C

SUBROUTINE IREAD(IEND,K,L,N,IF5,M,X,J,U)
INTEGER U
DIMENSION IF5(1),M(1),X(1)

C GIVES IEND=2 TO SIGNAL !EOD; CALL WITH IEND=3 IS TO D
UMP CARDS;
C K=NO. OF CLASS. VAR.; L=NO. OF REAL VAR.; M(K) & X(L)
ARE CLASS &
C REAL INPUT ARRAYS; N GIVES OBSERVATION COUNT ; IF5 IS
FORMAT
C

```

IF (IEND.EQ.3) GO TO 3
IF ((K*L).GT.0) READ(U,IF5,END=2) (M(I),I=1,K),(X(I),
I=1,L);RETURN
IF (K.GT.0) READ(U,IF5,END=2) (M(I),I=1,K);RETURN
IF (L.GT.0) READ(U,IF5,END=2) (X(I),I=1,L);RETURN
2 IEND=2; IF (J.EQ.0) RETURN
REWIND U; RETURN
3 WRITE(108,4) N
4 FORMAT('/' READ-IN TERMINATED ON CARD--,I7)
5 READ(U,6,END=2)
GO TO 5
6 FORMAT(' ')
END
```

```

SUBROUTINE INTER(KP, RX, B, SB, NE, XBAR, ISAVE, STD, KY, N, IC
, IP
1, NC, M, NP, Z, ZP, IMETHOD, YHAT1, VARIANCE, NR, NQQ)
DOUBLE PRECISION RX, B, STD, XBAR
DIMENSION XBAR(1), RX(1), B(1), ISAVE(1), STD(1)
DIMENSION X(-11:12), NT(12), XHAT(12), A(12,12), XCUM(12)
1, YLL(12), YUL(12), VARY(0:11), YHAT(0:11), D(144), U(11)
2, IC(1), E(144), COV(144), IP(1)
X(0)=0
YBAR=XBAR(ISAVE(KY))
NQP=0
GO TO (20, 40, 140, 240, 300, 300, 310, 310), NC
C
C USING PERIODIC TOTALS ONLY---METHOD 2-1
C
20 NE=N-1-NP
DO 25 I=1, 144
25 D(I)=0
DO 38 IR=1, NR
READ(10, 36) YA, NT(M)
36 FORMAT(11/, T27, F5.0, T20, I2)
YHAT1=B(NP+1)
D(NP+1)=1.
DO 37 I=1, NP
D(I)=(NT(M))**I
37 YHAT1=YHAT1+B(I)*D(I)
CALL IVAL(SB, RX, D, N, NP+1, VAR, KY, ISAVE, STD, NC)
IX=IC(1)
IY=IP(1)
CALL CINT(YHAT1, YA, Z, VAR, 0, IX, IR, IY, ZP, NQQ)
IC(1)=IX
IP(1)=IY
38 CONTINUE
YDIF=YHAT1-YA
WRITE(108, 39) YHAT1, YA, YDIF
39 FORMAT(' 0', 3F15.2)
RETURN

```

```

C
C      USING SUBPERIODIC OBSERVATIONS SEPARATELY---METHOD 2-
C
C
40      NE=N-1-M*NP
      DO 130 IR=1,NR
      READ(10,10) (X(I),(A(I,J),J=1,12),NT(I),XCUM(I),I=1,1
2)
10      FORMAT(F5.0,12I1,2X,I2,5X,F5.0)
      YACT = 0
      DO 50 I=1,M
      XHAT(I)=B(M*NP+1)
49      YACT=YACT+X(I)
50      CONTINUE
      VA=0
      DO 80 II=1,M
      DO 55 I=1,144
55      D(I)=0
      DO 60 J=1,NP
      XHAT(II)=XHAT(II)+B((II-1)*NP+J)*NT(II)**J
      D((II-1)*NP+J)=NT(II)**J
60      CONTINUE
      D(M*NP+1)=1.
      KPP=M*NP+1
65      CALL IVAL(SB,RX,D,N,KPP,VAR,KY,ISAVE,STD,NC)
      IY=IP(II+12)
      IX=IC(II+12)
      VA=VA+VAR
      CALL CINT(XHAT(II),X(II),Z,VAR,II,IX,IR,IY,ZP,NQP)
      IC(II+12)=IX
      IP(II+12)=IY
80      CONTINUE
      VAR=VA/M
      YHAT(0)=0
      DO 90 I=1,M
90      YHAT(0)=YHAT(0) + XHAT(I)
      DO 110 I=1,M-1
      YHAT(I)=0
      DO 100 J=1,I
100     YHAT(I)=YHAT(I)+X(J)
      DO 110 J=I+1,M
110     YHAT(I)=YHAT(I)+XHAT(J)
      DO 120 II=0,M-1
120     VARY(II)=(M-II)*VAR

```

```

DO 130 I=0,M-1
IX=IC(I+1)
IY=IP(I+1)
CALL CINT(YHAT(I),YACT,Z,VARY(I),I,IX,IR,IY,ZP,NQQ )
IC(I+1)=IX
IP(I+1)=IY
130 CONTINUE
DO 138 I=0,11
YDIF=YHAT(I)-YACT
WRITE(108,137) I,YHAT(I),YACT,YDIF
137 FORMAT(13,3F15.2)
138 CONTINUE
RETURN

C
C USING ALL SUBPERIODIC OBSERVATIONS SIMULTANEOUSLY---M
ETHOD 2-3
C
140 NE=N-1-NP-M+1
DO 230 IR=1,NR
READ(10,10) (X(I),(A(I,J)),J=1,12),NT(I),XCUM(I),I=1,1
2)
YACT=0
DO 142 I=1,M
142 YACT=YACT+X(I)
145 DO 190 II=1,M
XHAT(II)=B(NP+M)
IF(II.EQ.M) GO TO 150
XHAT(II)=XHAT(II)+B(II)
150 DO 160 K=1,NP
160 XHAT(II)=XHAT(II)+B(M-1+K)*NT(II)**K
DO 170 I=1,M*M
170 D(I)=0
D(NP+M)=1.
D(II)=1.
DO 180 K=1,NP
180 D(M-1+K)=NT(II)**K
KKP=NP+M
CALL IVAL(SB,RX,D,N,KKP,VARY(II),KY,ISAVE,STD,NC)
IX=IC(II+12)
IY=IP(II+12)
CALL CINT(XHAT(II),X(II),Z,VARY(II)
1,II,IX,IR,IY,ZP,NQP )
IC(II+12)=IX
IP(II+12)=IY
190 CONTINUE

```

```

      IF(IR.GT.1) GO TO 195
      DO 191 II=1,M*M
      E(II)=0
191    D(II)=0
      DO 194 II=1,M-1
      DO 194 J=II+1,M
      D(II)=1.
      E(J)=1.
      DO 193 K=1,NP
      D(M-1+K)=NT(II)**K
193    E(M-1+K)=NT(J)**K
      D(NP+M)=1.
      E(NP+M)=1.
      CALL COVAR(SB,RX,D,E,N,NP+M,CO,KY,
1 ISAVE,STD)
      COV(LOCATE(II,J))=CO
      D(II)=0
194    E(J)=0
195    CONTINUE
      DO 230 IJ=0,M-1
      YHAT(IJ)=0
      DO 200 I=IJ+1,M
200    YHAT(IJ)=YHAT(IJ)+XHAT(I)
      DO 210 I=0,IJ
210    YHAT(IJ)=YHAT(IJ)+X(I)
      VAR=0
      DO 220 II=IJ+1,M
220    VAR=VAR+VARY(II)
      IF(IJ.GE.(M-1)) GO TO 228
      DO 225 II=IJ+1,M-1
      DO 225 IK=II+1,M
225    VAR=VAR+2*COV(LOCATE(II,IK))
228    CONTINUE
      IX=IC(IJ+1)
      IY=IP(IJ+1)
      CALL CINT(YHAT(IJ),YACT,Z,VAR,IJ,IX,IR,IY,ZP,NQQ )
      IP(IJ+1)=IY
      IC(IJ+1)=IX
230    CONTINUE
      DO 235 I=0,11
      YDIF=YHAT(I)-YACT
      WRITE(108,137) I,YHAT(I),YACT,YDIF
235    CONTINUE
      RETURN

```

```

C
C   USING CUMULATIVE OBSERVATIONS SEPARATELY---METHOD 2-4
C
240  IJ=IMETHOD
      IX=0
      IY=0
      NE=N-NP
      DO 280 IR=1,NR
        READ(10,10) (X(I),(A(I,J),J=1,12),NT(I),XCUM(I),I=1,1
2)
          YACT=0
          DO 246 I=1,M
246    YACT = YACT + X(I)
          Y=B(NP+2)+B(1)*XCUM(IJ)
          DO 250 I=2,NP+1
250    Y=Y+B(I)*(NT(IJ))**(I-1)
          DO 260 I=1,144
260    D(I)=0
          D(1)=XCUM(IJ)
          D(NP+2)=1.
          DO 270 I=1,NP
270    D(I+1)=NT(IJ)**I
          KPP=NP+2
          CALL IVAL(SB,RX,D,N,KPP,VAR,KY,ISAVE,STD,NC)
          VAR=VAR+(M-IJ)/IJ*(SB**2)
          YHAT1=Y
          CALL CINT(Y,YACT,Z,VAR,IJ,IX,IR,IY,ZP,NQQ)
280    CONTINUE
          YDIF=YHAT1-YACT
          WRITE(108,137) IJ,YHAT1,YACT,YDIF
          RETURN
C
C
300  CONTINUE
      RETURN

```

C
 C USING PERIODIC FRACTIONS SIMULTANEOUSLY---METHODS 3-2
 & 3-3

C
 310 NE=N-1-NP
 YBAR1=YHAT1
 D(1)=1
 IF(NC.EQ.7) GO TO 315
 YA=0
 READ(14,313) (X(I),I=-11,0)
 313 FORMAT(47/,12(/,F5.0))
 DO 314 I=-11,0
 314 YA=YA+X(I)
 YAC=YA
 315 DO 330 IR=1,NR
 YA=YAC
 IF(IR.NE.1) NQQ=0
 READ(10,10) (X(I),(A(I,J),J=1,12),NT(I),XCUM(I),I=1,1
 2)
 YACT=0
 DO 316 I=1,M
 316 YACT=YACT+X(I)
 DO 330 IJ=1,11
 YA=YA+X(IJ)-X(IJ-12)
 U(IJ)=B(NP+1)
 DO 320 K=1,NP
 D(K)=IJ**K
 320 U(IJ)=U(IJ)+D(K)*B(K)
 D(NP+1)=1
 YHAT(IJ)=XCUM(IJ)/U(IJ)
 IF(NC.EQ.8) YHAT(IJ)=YA/U(IJ)
 SSB=SB*YBAR1/U(IJ)
 KPP=NP+1
 CALL IVAL(SSB,RX,D,N,KPP,VAR,KY,ISAVE,STD,NC)
 VAR=VAR+(M-IJ)*VARIANCE
 VAR1=VARIANCE
 KJ=IJ
 IF(NC.EQ.8) KJ=M
 VAR=VAR+KJ*VAR1/(U(IJ)*U(IJ))+IJ*VAR1-2/U(IJ)*IJ*VAR1
 IX=IC(IJ)
 IY=IP(IJ)
 CALL CINT(YHAT(IJ),YACT,Z,VAR,IJ,IX,IR,IY,ZP,NQQ)

```

IC(IJ)=IX
IP(IJ)=IY
SDEV=SQRT(VAR)
YL=YACT-YHAT(IJ)+Z*SDEV
YU=YHAT(IJ)+Z*SDEV-YACT
YDIF=YHAT(IJ)-YACT
WRITE(108,325) IJ,YHAT(IJ),YACT,YDIF,YL,YU,SDEV
325  FORMAT(I3,6F13.2)
330  CONTINUE
RETURN
END

C
C
C
SUBROUTINE CINT(YHAT1,Y,Z,VAR,IK,IC,IR,IY,ZP,NQQ)
SDEV=SQRT(ABS(VAR))
YL=YHAT1-Z*SDEV
YU=YHAT1+Z*SDEV
IF(YL.LT.Y.AND.YU.GT.Y) IC=IC+1
40  YL=Y*(1.-ZP)
YU=Y*(1.+ZP)
IF(YL.LT.YHAT1.AND.YU.GT.YHAT1) IY=IY+1;RETURN
RETURN
END

C
C
C
SUBROUTINE TO CALCULATE COVARIANCES

SUBROUTINE COVAR(SB,RX,D,E,N,NP,COV,KY,ISAVE,STD)
DOUBLE PRECISION RX,STD
DIMENSION STD(1),XXI(365),ISAVE(1),RX(1),B(1)
1, D(1),E(1)
NDF=N-1
XZ=0
DO 20 I=1,NP
DO 20 J=1,I
X=RX(LOCATE(J,I))/STD(ISAVE(J+KY))/
1STD(ISAVE(I+KY))/NDF
20 XXI(LOCATE(J,I))=X
DO 25 I=1,NP
DO 25 J=1,NP
25 XZ=XZ+XXI(LOCATE(J,I))*D(I)*E(J)
COV=(SB**2)*XZ
RETURN
END

```

```

SUBROUTINE RES(N,K,KY,II,M,IPAR14,NRES,ISAVE,XBAR,STD
,B,X,RX,SP)
DOUBLE PRECISION XBAR,STD,B,RX
DIMENSION ISAVE(1),XBAR(1),STD(1),B(1),X(1),RX(1)
NSE=1; IF (NRES.EQ.1) NSE=0
REWIND 8; ZN=0.0; ZD=0.0; L=ISAVE(II); WRITE (6,1) NSE; TM
AX=0.0
10 FORMAT(' RESIDUALS ARE-- '/' OBS.   Y VALUE       Y ESTI
MATE',
14X, 'RESIDUAL',N( 3X,'S.E.-RESIDUAL', 3X,'T-RESIDUAL
'))
2 FORMAT(T2,I4,5G13.4)
3 FORMAT ('0DURBIN WATSON',8X,F9.5)
DO 10 NN=1,N
READ (8) (X(I),I=1,M)
ID=0; SE=0.0; YE=0.0; IF (IPAR14.EQ.0) YE=XBAR(L)
DO 11 I=1,K
LI=ISAVE(KY+I)
IF (IPAR14.EQ.0) X(LI)=X(LI)-XBAR(LI)
IF (NRES.EQ.1) GO TO 11
IF (I.EQ.1) GO TO 12
DO 13 J=1,I-1
LJ=ISAVE(KY+J)
13 SE=SE+(X(LJ)*X(LI)*RX(ID+J)/STD(LJ)/STD(LI))*2.0
12 ID=ID+1
SE=SE+((X(LI)/STD(LI))**2)*RX(ID)
11 YE=YE+B(I)*X(LI); RE=X(L)-YE; SE=SE/(N-1)
IF (NRES.EQ.1) WRITE (6,2) NN,X(L),YE,RE; G
O TO 14
IF (IPAR14.EQ.0) SE=SE+(1.0/N)
SE=SP*SQRT(1.0-SE); T=RE/SE; WRITE (6,2) NN,X(L),YE,RE,
SE,T
IF (ABS(T).GT.TMAX) TMAX=ABS(T); IDT=NN
14 IF (NN.EQ.1) RE1=RE
ZN=ZN+((RE-RE1)*(RE-RE1)); ZD=ZD+RE*RE
10 RE1=RE
ZN=ZN/ZD; WRITE(6,3) ZN
IF (NRES.NE.1) CALL RPROB (IDT,TMAX,N,K)
RETURN
END

```

```

SUBROUTINE CSGEN(NC,SIGMA,XW,K,ISEED,NP,N,M,B,NCH,NR,
B0)
C THIS PROGRAM GENERATES DATA FOR COMPARISON OF
C REGRESSION ESTIMATES
C
  DIMENSION B(0:24), C(0:24), A(12), BO(0:24)
  INTEGER UL
  REAL MEAN,MEAN1,MIN,MAX
  X0=XW
  ICT=0
  DO 15 I=1,M-1
15  A(I)=0
  MN=M*N
  KK=0
  DO 31 IJ=1,N
  IF(NCH.EQ.1) KK=0
  DO 30 II=1,M
  MEAN = B(0)
  KK=KK+1
  DO 20 J=1,NP
  K=NP+II
20  MEAN=MEAN + B(J)*(KK**J)
  MEAN1=MEAN+B(K)
  A(II)=1
  N0=0;IR=1
  GO TO (21,22),NC
21  MIN=MEAN1-X0*SIGMA
  MAX=MEAN1+X0*SIGMA
  Y=RNORM(ISEED,MEAN1,SIGMA,MIN,MAX)
  GO TO 26
22  LL=MEAN1-XW
  UL=MEAN1+XW
  Y=UNFRM(LL,UL,ISEED)
26  CONTINUE
  WRITE(12,28) Y,(A(KI),KI=1,12),II,KK
28  FORMAT(15,12I1,2I2)
30  A(II)=0
31  IF(NCH.EQ.2) ICT=0; GO TO 32
  KK1=KK

```

```

C      GENERATION OF DATA FOR CURRENT PERIOD
C
32     IF(NCH.NE.2) GO TO 65
        ICT=ICT+1; IR=NR
        IF(ICT.GE.N) NCH=1; GO TO 65
        OUTPUT 'TYPE 1 IF NO CHANGE IN TREND'
        INPUT ND
        IF(ND.EQ.1) GO TO 50
35     OUTPUT ' INPUT DEGREE FOR TREND'
        INPUT NP
        OUTPUT 'INPUT ALPHA FOR I=0,NP'
        DO 40 I=0,NP
            INPUT X
40     C(I)=X
50     OUTPUT 'TYPE 1 IF NO CHANGE IN EFFECTS'
        INPUT ND
        IF(ND.EQ.1) GO TO 63
        OUTPUT 'INPUT EFFECTS FOR SUBPERIODS 1 THRU M-1'
        DO 60 I=1,M-1
            INPUT X
            J=I+NP
60     C(J)=X
            J=NP+M
            C(J)=0
            GO TO 70
63     DO 64 I=1,M-1
64     C(I)=B(I+NP)
            GO TO 70
65     DO 90 IR=1,NR
        NKK = KK1
        DO 66 I=0,24
66     C(I)= B0(I)
70     DO 90 II=1,M
        X = C(0)
        DO 80 J=1,NP
80     X=X+ C(J)*(II**J)
        K=NP+II
        X=X+ C(K)
        A(II)=1
        GO TO (81,82),NC
81     MIN=X-XQ*SIGMA
        MAX=X+XQ*SIGMA
        Y=RNORM(ISEED,X,SIGMA,MIN,MAX)
        GO TO 86

```

```

82  LL=X-XW
    LU=X+XW
    Y=UNFRM(LL,LU,ISEED)
86  CONTINUE
    INN=14
    NKK=NKK+1
    IF(NCH.EQ.2) INN=12
    WRITE(INN,87) Y,(A(IK),IK=1,12),II,NKK
87  FORMAT(15,12I1,2I2)
90  A(II)=0
    IF(NCH.EQ.2) MEAN=X;KK1=NKK;GO TO 32
    OUTPUT ISEED
    REWIND 12
    REWIND 14
    CALL CSCONV(14,10,0)
    REWIND 14
    IOUT=5
    CALL CSCONV(12,IOUT,1)
    REWIND 5
    REWIND 14
    REWIND 10
    REWIND 12
    RETURN
    END

```

```

C
C  UNIFORM RANDOM NUMBER ON INTERVAL 0,1
C

```

```

SUBROUTINE MYRAN(IY,YFL)
  IY=IY*65539
  IF(IY.LE.0) IY=IY+1+2147483647
  YFL=IY*.4656613E-09
  RETURN
  END

```

```

C
C  UNIFORM RANDOM NUMBER ON INTERVAL FROM A TO B
C

```

```

FUNCTION UNFRM(LL,LU,IY)
  A=LL
  B=LU
  CALL MYRAN(IY,RNUM)
  UNFRM=A+(B-A)*RNUM
  RETURN
  END

```

```

SUBROUTINE CORRE(IEND,M,N,IP,XBAR,STD,R,D)
DOUBLE PRECISION XBAR(1),STD(1),R(1),C
DIMENSION D(1)
IJ=0
IF (N.GT.1) GO TO 10
DO 1, I=1,M
XBAR(I)=D(I); STD(I)=0.
DO 1, J=1, I
IJ=IJ+1
1 R(IJ)=0.0; RETURN
10 IF (IEND.NE.1) GO TO 20
DO 11, I=1,M
D(I)=D(I)-XBAR(I); STD(I)=D(I)+STD(I)
DO 11, J=1, I
IJ=IJ+1
11 R(IJ)=R(IJ)+D(I)*D(J); RETURN
20 IF (IEND.EQ.4) WRITE (6,31) IP,(I,I=1,M)
IF (IEND.EQ.5) WRITE (6,32) (I,I=1,M)
31 FORMAT('////////' CORRELATION MATRIX IS--'/(N(90(5I12/)),
(10I12))
32 FORMAT(' VARIANCE-COVARIANCE MATRIS IS--'/(T2,I10,9I1
2))
DO 21 I=1,M
XBAR(I)=(STD(I)/N)+XBAR(I)
DO 21 J=1, I
IJ=IJ+1
R(IJ)=(R(IJ)-STD(I)*STD(J)/N)/(N-1)
IF (IEND.EQ.6) R(IJ)=R(IJ)+XBAR(I)*XBAR(J)*N/(N-1)
21 CONTINUE; II=0
DO 22 I=1,M
IS=II+1; II=II+I; STD(I)=DSQRT(DABS(R(II)))
DO 23 J=1, I
IJ=IS+J-1; C=STD(I)*STD(J)
IF ((IEND.NE.5).AND.(C.GT.0)) R(IJ)=R(IJ)/C
23 CONTINUE
IF ((IEND.EQ.4).OR.(IEND.EQ.5)) WRITE (6,33) I,IP,(R(
J),J=IS,II)
33 FORMAT(T2,I3, N(90(T5,5G12.4/)),(T5,10G12.4))
IF (M.GT.10) WRITE (6,33)
22 CONTINUE
0IF ((IEND.EQ.4).OR.(IEND.EQ.5))
1WRITE (6,40) (I,XBAR(I),STD(I),I=1,M)
40 FORMAT(/T8,'MEAN:',6X,'STD.DEV.'/(I4,2G12.4))
RETURN
END

```

```

SUBROUTINE PRINT(N,K,KY, IPAR, ISAVE, IHEAD, R, RX, STD, T
,IA)
  DOUBLE PRECISION R, RX, STD
  DIMENSION IPAR(1), ISAVE(1), IHEAD(20), R(1), RX(1), STD(1
), T(1), IA(8)
  DF=N-1; IF (IPAR(16).EQ.0) GO TO 20
  WRITE (6,1) IHEAD, IA; J=1; IF (IPAR(14).EQ.1) J=0
  NV=K+KY; WRITE(6,2) J, IPAR(12), (ISAVE(L), L=1, KY)
  DO 11 I=1, NV
    DO 12 J=1, I
      12 T(J)=R(LOCATE(ISAVE(J), ISAVE(I)))*STD(ISAVE(J))*STD(I
SAVE(I))*DF
      IF (I.EQ.KY+1) WRITE(6,4) IPAR(12), (ISAVE(L), L=1, NV)
      IF (KY+K.GT.IPAR(12)) WRITE (6,1)
      11 WRITE (6,3) ISAVE(I), IPAR(12), (T(L), L=1, I)
      20 IF (IPAR(17).EQ.0) GO TO 30
      WRITE (6,1) IHEAD; WRITE(6,5) IPAR(12), (ISAVE(L+KY), L=
1,K)
      DO 21 I=1, K
        DO 22 J=1, I
          22 T(J)=RX(LOCATE(J, I))/STD(ISAVE(J+KY))/STD(ISAVE(I+KY)
)/DF
          IF (K.GT.IPAR(12)) WRITE (6,1)
          21 WRITE (6,3) ISAVE(I+KY), IPAR(12), (T(L), L=1, I)
          30 IF (IPAR(20).EQ.0) RETURN
          WRITE (6,1) IHEAD; WRITE(6,6) IPAR(12), (ISAVE(L+KY), L=
1,K)
          DO 31 I=1, K
            DO 32 J=1, I
              32 T(J)=RX(LOCATE(J, I))/DSQRT(RX(LOCATE(J, J))*RX(LOCATE(
I, I)))
              IF (K.GT.IPAR(12)) WRITE (6,1)
              31 WRITE (6,3) ISAVE(I+KY), IPAR(12), (T(L), L=1, I); RETURN
              1 FORMAT(20A4, T1, 1H1/8A4//)
              2 FORMAT( ' Y''Y MATRIX', N(90(5I12/)), (10I12)
MEANS--')/
              1
              3 FORMAT(T2, I3, N(90(T5, 5G12.4/)), (T5, 10G12.4))
              4 FORMAT(/' X''Y AND X''X MATRICES'/N(90(5I12/)), (10I12
))
              5 FORMAT( ' INVERSE OF X''X'/N(90(5I12/)), (10I12))
              60 FORMAT(' INVERSE OF X''X IN CORRELATION FORM'/
              1
              N(90(5I12/)), (10I12))
            END

```

```

SUBROUTINE TRAN (IEND,KT,KV,IT,CC,D,E),
DIMENSION IT(5,1),CC(1),D(1),E(1)
DO 97,N=1,KT
M=IT(1,N);I=IT(2,N);J=IT(3,N);K=IT(4,N);L=IT(5,N);C=C
C(N)
91 IF ((I.GT.KV).OR.(J.GT.KV).OR.(K.GT.KV)) IEND=3;RETUR
N
GO TO (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,1
9),M
GO TO 97
1 D(I)=ALOG(D(J));GO TO 99
2 D(I)=SIN(D(J));GO TO 99
3 D(I)=COS(D(J));GO TO 99
4 D(I)=EXP(D(J));GO TO 99
5 D(I)=ATAN(D(J));GO TO 99
6 D(I)=D(J)*C;GO TO 99
7 D(I)=D(J)**C;GO TO 99
8 D(I)=D(J)*D(K);GO TO 99
9 D(I)=D(J)/D(K);GO TO 99
10 D(I)=D(J)/C;GO TO 99
11 D(I)=D(J)+D(K);GO TO 99
12 D(I)=D(J)-D(K);GO TO 99
13 D(I)=D(J)+C;GO TO 99
14 D(I)=C/D(J);GO TO 99
15 D(I)=ALOG10(D(J));GO TO 99
16 D(I)=D(J);GO TO 99
17 D(I)=D(J)+E(K);GO TO 99
18 D(I)=ASIN(D(J));GO TO 99
19 IF (D(I).LE.C) IEND=2;RETURN
99 L=L-1
IF (L.GT.0) I=I+IT(2,N-1);J=J+IT(3,N-1);K=K+IT(4,N-1)
;GO TO 91
97 CONTINUE;RETURN
END

```

```

INTEGER FUNCTION LND(I,J)
IF((I-J).LT.0) LND=1541;RETURN
LND=J+(I*I-I)/2;RETURN
END

```

C
C
C
C

```

      INTEGER FUNCTION SUB(I,J)
C      TO CALCULATE THE SUBSCRIPT OF ELEMENT A(I,J) FOR A SY
MMETRIC
C      MATRIX STORED COLUMN BY COLUMN (UPPER TRIANGLE ONLY)
IN A ONE
C      DIMENSIONAL ARRAY
      SUB=I+(J*J-J)/2
      RETURN
END

```

MONTANA STATE UNIVERSITY LIBRARIES

3 1762 10011224 0

DATE	ISSUED TO

D370
 Sh 13
 cap 2