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

Chairman, Examining Committee

Henry L. Parsons
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

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., $t = 1,2,\ldots,mn$.

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). $(Z_j^{(i)} = \Sigma_{k=1}^{i} X_{jk})$.

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

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

$$\text{var}(Y_j) = \Sigma_{i=1}^{m} \text{var}(X_{ji}) + 2 \sum_{\substack{i,k \\ i<k}} \text{cov}(X_{ji}, X_{jk})$$

$$= \Sigma_{i=1}^{m} \text{var}(X_{ji})$$

$$= m\, \sigma_{\epsilon}^{2} ,$$

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.,

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

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.,

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

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

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

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

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

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.

$$\mathbf{U}_j = \begin{bmatrix} \mu_{j\,1} \\ \mu_{j\,2} \\ \vdots \\ \mu_{j\,12} \end{bmatrix} \quad \mathbf{Y}_j = \begin{bmatrix} Y_j^{(0)} \\ Y_j^{(1)} \\ \vdots \\ Y_j^{(11)} \end{bmatrix} \quad \mathbf{X}_j = \begin{bmatrix} x_{j\,1} \\ x_{j\,2} \\ \vdots \\ x_{j\,12} \end{bmatrix} \quad \mathbf{E}_j = \begin{bmatrix} \epsilon_{j\,1} \\ \epsilon_{j\,2} \\ \vdots \\ \epsilon_{j\,12} \end{bmatrix}$$

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

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

The matrix model for the j-th period is

$$\mathbf{Y}_j = \mathbf{M}\mathbf{X}_j + \mathbf{N}\mathbf{U}_j + \mathbf{N}\mathbf{E}_j .$$

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

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

$$\hat{\mathbf{Y}}_j = \mathbf{M}\mathbf{X}_j + \mathbf{N}\hat{\mathbf{U}}_j .$$

Proof. $\mathcal{E}_G(\mathbf{Y}_j) = \mathcal{E}_G(\mathbf{M}\mathbf{X}_j + \mathbf{N}\mathbf{U}_j + \mathbf{N}\mathbf{E}_j)$

$$= \mathbf{M}\mathbf{X}_j + \mathbf{N}\mathbf{U}_j ,$$

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

In these expectations $G = \{X_{j1},\ldots,X_{jm}\}$. In the following, $H = \{X_{11},\ldots,X_{j-1,m}\}$.

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 $\mathcal{E}_G(Y_j)$. ∎

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

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

Proof. $\mathrm{Var}[Y_j^{(i)}] = \mathrm{Var}[\Sigma_{k=1}^{i} x_k + \Sigma_{k=i+1}^{m} \mu_k + \Sigma_{k=i+1}^{m} \epsilon_{jk}]$

$= \mathrm{Var}(\Sigma_{k=i+1}^{m} \epsilon_{jk}) = \Sigma_{k=i+1}^{m} \mathrm{Var}(\epsilon_{jk})$

$= (m-i)\sigma_\epsilon^2$.

Assume $i < i'$.

$$\mathrm{Cov}(Y_j^{(i)}\ Y_j^{(i')}) =$$

$$\mathcal{E}[(Y_j^{(i)} - \mathcal{E}(Y_j^{(i)}))(Y_j^{(i')} - \mathcal{E}(Y_j^{(i')})] =$$

$$\mathcal{E}\{[\Sigma_{k=i+1}^{m} \epsilon_{jk}] \cdot [\Sigma_{k'=i'+1}^{m} \epsilon_{jk'}]\} =$$

$$\mathcal{E}\{[\Sigma_{k=i+1}^{i'} \epsilon_{jk} + \Sigma_{k=i'+1}^{m} \epsilon_{jk}] \cdot [\Sigma_{k'=i'+1}^{m} \epsilon_{jk}]\} =$$

$$\mathcal{E}\{\Sigma_{k=i+1}^{i'} \epsilon_{jk} \cdot \Sigma_{k'=i'+1}^{m} \epsilon_{jk'}\}$$

$$+ \mathcal{E}\{\Sigma_{k=i'+1}^{m} \epsilon_{jk} \cdot \Sigma_{k'=i'+1}^{m} \epsilon_{jk'}\} =$$

$$\Sigma_{k=i+1}^{i'} \Sigma_{k'=i'+1}^{m} \mathcal{E}(\epsilon_{jk})\mathcal{E}(\epsilon_{jk'})$$

$$+ \mathcal{E}\{\Sigma_{k=i'+1}^{m} \epsilon_{jk}^2 + 2\Sigma_{\substack{kk'=i'+1 \\ k<k'}}^{m} \epsilon_{jk'}\epsilon_{jk}\} =$$

$$0 + \Sigma_{k=i'+1}^{m} \mathcal{E}(\epsilon_{jk}^2) + 0 =$$

$$(m-i')\sigma_\epsilon^2 .$$

Thus, ❋ is as shown. ❙

For m=12, ❋ is

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

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., $\mathbf{U}_j$'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

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

and base the estimates on a regression equation of the form

$$\mathbf{Y} = \mathbf{TB} + \mathbf{S},$$

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where $s^2_{Y.T}$ is the error variance obtained from the regression analysis and $\mathbf{A} = (1\ (n+1)\ \ldots\ (n+1)^p)$. 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,

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

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 ($i = 1,2,\ldots,m$), the model used will be

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

$$j = 1,\ldots,n$$

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

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

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

$$\beta_{i0}, \beta_{i1}, \ldots, \beta_{ip}, \text{ and } \sigma^2_{i\eta},$$

denoted by

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

Using the matrices

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

$$\mathbf{M}_0^{(i)} = [1 \ (i+nm) \ (i+nm)^2 \ \ldots \ (i+nm)^p] \quad \text{and}$$

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

the model is

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

giving

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

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

$$\hat{X}_{n+1,i} = \mathbf{M}_0^{(i)} \cdot \hat{\mathbf{B}}^{(i)} .$$

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

$$\hat{\sigma}_i^2 = S_i^2(1 + \mathbf{M}_0^{(i)'}(\mathbf{M}^{(i)'}\mathbf{M}^{(i)})^{-1}\mathbf{M}_0^{(i)})$$

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,...,m-1), the estimate of the periodic cumulative total is

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

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

$$v_{(k)}^2 = \widehat{\mathrm{Var}}[\Sigma_{i=k+1}^{m}(\hat{X}_{n+1,i} - X_{n+1,i})]$$

$$\doteq \Sigma_{i=k+1}^{m} \hat{\sigma}_i^2 .$$

Let $\mathbf{\phi}_i$ be a matrix of dimension (1 X i) whose elements are all zero, and $\mathbf{1}_i$ a matrix of dimension (1 X i) whose elements are all one; also note

$$\mathbf{X}'_{n+1} = (x_{n+1,1}, \cdots, x_{n+1,m})$$

and

$$\hat{\mathbf{U}}'_{n+1} = (\hat{X}_{n+1,1}, \cdots, \hat{X}_{n+1,m}) \cdot$$

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

$$\hat{Y}^{(0)}_{n+1} = \mathbf{\phi}_m \cdot \mathbf{X}_{n+1} + \mathbf{1}_m \cdot \hat{\mathbf{U}}_{n+1} ,$$

i.e., the estimated total at the beginning of the period is the sum of the estimates of the subperiodic observations. For $k > 0$,

$$\hat{Y}^{(k)}_{n+1} = [\mathbf{1}_k \,\vdots\, \mathbf{\phi}_{m-k}] \cdot \mathbf{X}_{n+1} + [\mathbf{\phi}_k \,\vdots\, \mathbf{1}_{m-k}] \cdot \hat{\mathbf{U}}_{n+1} ;$$

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-α by

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

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,

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

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

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

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

$$\mathbf{X}_j = (x_{j1}, x_{j2}, \ldots, x_{jm})'$$

$$\mathbf{M}_j = \left[\begin{array}{c:c} \mathbf{T}_j & \begin{matrix}\mathbf{I}_{m-1} \\ \hdashline \mathbf{\emptyset}_{m-1}\end{matrix} \end{array}\right] .$$

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

$$\mathbf{X}_1 = \mathbf{M}_1 \mathbf{B} + \mathbf{E}_1 = \left[\begin{array}{c:c} \mathbf{T}_1 & \begin{matrix}\mathbf{I}_{m-1} \\ \hdashline \mathbf{\emptyset}_{m-1}\end{matrix} \end{array}\right] \mathbf{B} + \mathbf{E}_1$$

$$= \left[\begin{array}{cccc:ccccc} 1 & 1 & \ldots & 1 & 1 & 0 & \ldots & 0 \\ 1 & 2 & \ldots & 2^p & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & m-1 & \ldots & (m-1)^p & 0 & 0 & \ldots & 1 \\ \hdashline 1 & m & \ldots & m^p & 0 & 0 & \ldots & 0 \end{array}\right] \left[\begin{array}{c} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \\ \hdashline \beta_{p+1} \\ \vdots \\ \beta_{m+p-1} \end{array}\right] + \left[\begin{array}{c} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1m-1} \\ \epsilon_{1m} \end{array}\right] .$$

Similarly, for j=6,

$$\mathbf{X}_6 = \left[\begin{array}{cccc:ccccc} 1 & 5m+1 & \cdots & (5m+1)^p & 1 & 0 & \cdots & 0 \\ 1 & 5m+2 & \cdots & (5m+2)^p & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 6m-1 & \cdots & (6m-1)^p & 0 & 0 & \cdots & 1 \\ 1 & 6m & \cdots & (6m)^p & 0 & 0 & \cdots & 0 \end{array}\right] \left[\begin{array}{c} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \\ \hdashline \beta_{p+1} \\ \vdots \\ \beta_{m+p-1} \end{array}\right] + \left[\begin{array}{c} \epsilon_{61} \\ \epsilon_{62} \\ \vdots \\ \epsilon_{6m-1} \\ \epsilon_{6m} \end{array}\right].$$

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

$$\mathbf{X}_1 = \left[\begin{array}{cccccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & 9 & 0 & 0 & 1 \\ 1 & 4 & 16 & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{array}\right] + \left[\begin{array}{c} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \end{array}\right]$$

or,

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

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,

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

where j is the period and k is the subperiod.

The model for all the data is therefore,

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} = \begin{bmatrix} M_1 \\ M_2 \\ \cdot \\ \cdot \\ \cdot \\ M_n \end{bmatrix} \mathbf{B} + \begin{bmatrix} E_1 \\ E_2 \\ \cdot \\ \cdot \\ \cdot \\ E_n \end{bmatrix}$$

$$= \mathbf{M}\,\mathbf{B} + \mathbf{E} .$$

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

$$\mathbf{M}_{Oi} = (1 \ (nm+i) \ \ldots \ (nm+i)^p \mathbf{C}_i)$$

where $\mathbf{C}_i$ is a matrix of dimension 1 X (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,...,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

$$\hat{X}_{n+1,i} = \mathbf{M}_{Oi}\hat{\mathbf{B}} \quad \text{for } i = 0,1,\ldots,m-1$$

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

$$\hat{\sigma}_i^2 = S_{X.M}^2 \ (1 + \mathbf{M}'_{Oi}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}_{Oi}).$$

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,

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

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 $(\hat{Y}_{n+1}^{(k)} - Y_{n+1})$,

$$\hat{\sigma}_{ih} = \widehat{Cov}[(\hat{X}_{n+1,i} - X_{n+1,i}),(\hat{X}_{n+1,h} - X_{n+1,h})]$$

$$= S_{X.M}^{2} \; \mathbf{M}_{0i}' \; (\mathbf{M}'\mathbf{M})^{-1} \; \mathbf{M}_{0h} \; .$$

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

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

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,...,m-1, regress the models

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

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

the j-th period.

Let

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{B}^{(i)} = \begin{bmatrix} \beta_0^{(i)} \\ \beta_1^{(i)} \\ \vdots \\ \beta_{p+1}^{(i)} \end{bmatrix} \qquad \mathbf{E}^{(i)} = \begin{bmatrix} \nu_1^{(i)} \\ \nu_2^{(i)} \\ \vdots \\ \nu_n^{(i)} \end{bmatrix}$$

and

$$\mathbf{M}^{(i)} = \begin{bmatrix} 1 & i & \dots & i^p & Z_1^{(i)} \\ 1 & m+i & \dots & (m+i)^p & Z_2^{(i)} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & km+i & \dots & (km+i)^p & Z_n^{(i)} \end{bmatrix}.$$

where $k = n-1$. Then,

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

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

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

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

$$\hat{Y}_{n+1}^{(i)} = \mathbb{M}_i \hat{\mathbb{B}}^{(i)} = [1 \ (nm+i) \ \ldots \ (nm+i)^p \ Z_{n+1}^{(i)}] \hat{\mathbb{B}}^{(i)} .$$

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

$$v_{(i)}^2 = \frac{(m-i)}{i} S_{(i)}^2 + S_{(i)}^2 \mathbb{M}_i' (\mathbb{M}^{(i)'} \mathbb{M}^{(i)})^{-1} \mathbb{M}_i)$$

$$+ S_{(i)}^2 - 2 \hat{\mathbb{B}}^{(i)} S_{(i)}^2$$

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

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

where $S_{Y\cdot T}^2$ is the residual mean square error which estimates $m\,\sigma_\epsilon^2$, provided the model is correct. Note that $\sigma_{Y\cdot T}^2 \geq m\,\sigma_\epsilon^2$ 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

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

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

$$\Sigma_{i=k+1}^{m} S_{X \cdot M}^2 \left(1 + \mathbf{M}_{Oi}' (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}_{Oi}\right) +$$

$$\Sigma_{i=k+1}^{m} \Sigma_{\substack{h=k+1 \\ i \neq h}}^{m} S_{X \cdot M}^2 (\mathbf{M}_{Oi}' (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}_{Oh}).$$

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

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

where S_i^2 is the residual mean square error.