TIME SERIES METHODS AND REPEATED
SAMPLE SURVEYS

Stephen John Haslett

Submitted for the degree of
Doctor of Philosophy
in Mathematics
at the
Victoria University of Wellington

June 1986

Bir katredir ancak aldığım hep
Derya yine durmada lebalep

The things I've chosen are a drop, no more;
The undiminished sea still crowds the shore.

Ziya Pasha

ABSTRACT

When applied to a sequence of repeated surveys, the traditional sample survey estimators of means or totals for one time period only, fail to take advantage of any time series structure. Such structure may result from correlation between successive responses for resampled individuals, or from time series properties in the parameters of interest.

Historically, the initial published papers on time series improvement of repeated sample survey estimates allowed only the first possibility, treating the sum over the population of the individual responses as fixed; individual responses were seen as having stochastic properties only with respect to the sampling scheme. The alternative and later development allowed that both individual responses and their sum have stochastic properties with respect to a superpopulation from which the population of individual responses are drawn. Superpopulations allowed the application of mainstream time series techniques, including signal extraction and stochastic least squares, to repeated sample survey data. These developments in their historical perspective are the topic of Chapter 1.

Superpopulation models may also be applied to sample surveys from a single time period, and superpopulation and design properties of the one period linear non-homogeneous sample survey estimator form the topic of Chapter 2; this estimator is sufficiently general to subsume almost all single period non-informative sample survey estimators, and Chapter 2 allows systematisation of a wide range of previously disparate results. This linear estimator may also be extended beyond one time period to include the known estimators for repeated surveys, and this topic, together with a consideration of the effects of data aggregation on non-stochastic

and stochastic least squares, is the subject of Chapter 3.

Given the central role of the general linear model, and the time series nature of repeated surveys, projection and parameter updating formulae for linear models should form an integral part of repeated survey analysis. The correlation of sample survey errors however, invalidates the formulae appropriate to the known iid error case, and Chapters 4 and 5 develop the general formulae to allow correlated error structure. Chapter 4 considers parameter vectors of fixed length, as for example, for polynomial models, and provides formulae for estimating the length of the parameter vector, and for calculating independent recursive residuals and cusums when further data are added to the model. Chapter 5 considers updating and projection formulae in a wider context, and allows that the parameter vector may be stochastic or non-stochastic and that its length may increase with additional data; it consequently provides a general extension of the Kalman filter to the case of coloured noise over time.

The paucity of suitable data has limited data analysis to that contained in Chapter 6, where a simulation study and an analysis of medical data gauge the efficacy of polynomial models in time with multiple observations per time point and autocorrelated errors. The formulae of Chapter 4 allow testing for the constancy of the regression relationships over time.

The appendix details SAS computer programs for fitting the polynomial models of Chapter 6.

ACKNOWLEDGEMENTS

This thesis continues ideas first nurtured during my three years at the Sample Design Division of the Department of Statistics, Wellington, and developed during my time at the Institute of Statistics and Operations Research at Victoria University. To my colleagues at both institutions, particularly Mr. Vic Duoba, Dr. Peter Thomson, Mr. Ross Renner, Professor David Vere-Jones, and also to Professor Alastair Scott of Auckland University, I owe considerable gratitude. Thanks are also due to Joy, Marian, Athena and Shelley for transforming my sometimes minuscule handwriting into legible typeface, and to Dr. Tim Maling of the Wellington Clinical School for making available his propranolol trial data. But principally, I would thank Margie, without whose love and unstinting support, this thesis would still lie unfinished.

TABLE OF CONTENTS

	<u>Page</u>
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
LIST OF MAIN NOTATION	ix
CHAPTER 1 REPEATED SAMPLE SURVEYS - THE HISTORICAL CONTEXT	1
1. Introduction	2
2. The Traditional Approach	5
3. The Time Series Approach, with $\tilde{Y}$ Non-stochastic	6
4. The Time Series Approach, with $\tilde{Y}$ Stochastic	32
CHAPTER 2 THE LINEAR NON-HOMOGENEOUS SAMPLE SURVEY ESTIMATOR	65
1. Introduction	66
2. Terminology	67
3. Expectation, Bias, Variance, and Mean Square Error of $\hat{\bar{Y}}$	69
4. Unbiased Estimation of the Variance of $\hat{\bar{Y}}$	77
5. The True and Estimated Variance of Traditional Linear Sample Estimators	80
6. Conclusion	88
CHAPTER 3 THE LINEAR NON-HOMOGENEOUS SAMPLE SURVEY ESTIMATOR IN A TIME SERIES CONTEXT	89
1. Introduction	90
2. Notation	90
3. Expectation, Bias, Variance and Mean Square Error	93
4. The Effect of Data Aggregation on Parameter Estimates from General Linear Models	96
5. Previous Published Results and the lnhssts Estimator	101

	<u>Page</u>
CHAPTER 4 RECURSIVE ESTIMATION OF THE GENERAL LINEAR MODEL FOR A PARAMETER VECTOR OF FIXED LENGTH	110
1. Introduction	111
2. Definitions and Notation	112
3. Updating Formulae and Recursive Residuals	115
4. Cusum Procedures	124
CHAPTER 5 UPDATING GENERAL LINEAR MODELS TO INCLUDE ADDITIONAL DATA AND/OR REGRESSORS	126
1. Introduction	127
2. Model 1 : Additional Regressors	130
3. Model 2 : Additional Data	133
4. Model 3(a) : Additional Data and Regressors, $X_{12} \neq 0$	135
5. Model 3(b) : Additional Data and Regressors, $X_{12} = 0$	144
6. Structural Parallels in the Updating Equations	148
7. Updating Algorithms and a Generalised Kalman Filter	149
CHAPTER 6 POLYNOMIAL MODELS FOR REPEATED SAMPLE SURVEY DATA	153
1. Introduction	154
2. The Model	156
3. A Simulation Study	161
4. An Example : The Effect of Propranolol on Heartrate	170
APPENDIX 1 THE POLYMACR SAS MACROS	180
General Comments	181
1. Macro INITPOL	185
2. Macro POLSTAT	199
3. Macro POLPROJ	202
4. Macro POLUPDT	209

	<u>Page</u>
5. Macro ESTRCOV	217
Programs	223
REFERENCES	272

LIST OF MAIN NOTATION

N	population size
i	$1 < i \leq N$ subscript denoting the elements of the population
$\mathcal{U}$	$\{u_i : i = 1, 2, \dots, N\}$ denotes the N population units
P	sample selection procedure
t	time; generally $t = 1, 2, \dots, h$
s	the set of sampled units
$\underline{s} = (s(1), s(2), \dots, s(h))$	the h sets of sampled units for times 1 to h
$\mathcal{S}$	set of all possible samples s for one time period only
S	set of all possible samples $\underline{s} = (s(1), s(2), \dots, s(h))$ from h time periods
$p(s)$	probability of drawing the sample, s
$p(\underline{s})$	probability of drawing the sample, $\underline{s} = (s(1), s(2), \dots, s(h))$
E	design expectation, i.e. p -expectation (except in Chapters 4 and 5 where E may denote joint design-superpopulation expectation)
B	design bias
V	design variance
ξ	joint distribution of $Y_1, Y_2, \dots, Y_N$ or $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_N$
ϵ	superpopulation expectation, i.e. ξ expectation
β	superpopulation bias
$\mathcal{V}$	superpopulation variance
u	number of unmatched units between two consecutive time periods, denoted $u_t = u(t)$ at time t if $u_t \neq u$
m	number of matched units between two consecutive time periods, denoted $m_t = m(t)$ at time t if $m_t \neq m$
n	sample size, denoted $n(t) = n_t$ at time t , if $n(t) \neq n$ necessarily; $u + m = n$ where appropriate; cf. Chapter 4, p.113

$\pi = m/n$	proportion of overlap between consecutive samples
π_i	inclusion probability for unit i in sample, s
π_{ij}	joint inclusion probability of units i and j in the sample s
p_i	probability of drawing unit i at any draw for a with replacement design with equivalent draws
$W_{i(s)}$	random variable with respect to the sampling scheme associated with the i^{th} population unit
$W_{i(\underline{s})}$	$h \times h$ matrix, random with respect to the sampling scheme for the i^{th} population unit, in a series of h repeated samples, $\underline{s}$
y	variable of interest
$\{y_i : i \in s\}$	set of sampled values of variable of interest
Y	population quantity of interest, possibly random with respect to a superpopulation
$\bar{Y}$	population mean, possibly random with respect to a superpopulation for which $\varepsilon(\bar{Y}) = \bar{\theta}$
$Y(t)$	population quantity of interest at time t , almost invariably a mean
$\underline{Y} = \underline{Y}^{(h)} = (Y(1), Y(2), \dots, Y(h))^T = (\tau(1), \tau(2), \dots, \tau(h))^T = \underline{\tau}_{(h)} = \underline{\tau}$ $\underline{\theta} = \varepsilon(\underline{Y}) = (\theta(1), \theta(2), \dots, \theta(h))^T$. Occasionally denoted $\bar{\underline{\theta}}$ to emphasise that it is a vector of superpopulation means	
$\hat{Y}$	estimate or estimator of Y , usually of linear non-homogeneous form
$\hat{Y}(t)$	traditional estimate or estimator of Y at time t , i.e. based only on data from time t
$\hat{\underline{Y}}_{(h)} = (\hat{Y}(1), \hat{Y}(2), \dots, \hat{Y}(h))^T$ [c.f. $\underline{Y}_{(h)}$ the vector of individual responses to time h]	
$\hat{\bar{Y}}$	linear non-homogeneous estimator of $\bar{Y}$; the corresponding estimate is $\hat{\bar{y}}$

$\hat{\bar{Y}}(t)$ linear non-homogeneous estimator of $\bar{Y}(t)$, not necessarily a traditional estimator, c.f. $\hat{Y}(t)$

$\hat{\bar{Y}} = (\hat{\bar{Y}}(1), \hat{\bar{Y}}(2), \dots, \hat{\bar{Y}}(h))^T$

Y_i random variable associated with the i^{th} population unit

$Y_i(t)$ random variable associated with the i^{th} population unit at time t

$\underline{Y}_i = (Y_i(1), Y_i(2), \dots, Y_i(h))^T$

$Y_{ij}(t)$ the realisation of this random variable is the value of y for the j^{th} drawn population unit, u_i

$\hat{Y}_j(t)$ elementary estimate or estimator of $Y(t)$ at time t , i.e. unbiased traditional estimator based on a subsample at time t

$\hat{\underline{Y}}_G$ vector of elementary estimates or estimators ($rx1$) from h time periods, $r \geq h$

$\hat{\bar{Y}}(t)_u$ elementary mean estimate at time t for units unmatched with those from time $t-1$ or $t+1$ as appropriate

$\hat{\bar{Y}}(t)_m$ as above, except for matched units

$\hat{\bar{Y}}(t)'$ mean of units matched with units in sample at time $t-1$

$\hat{\bar{Y}}(t)''$ mean of units unmatched with units in sample at time $t-1$

$\hat{\bar{Y}}(t)'''$ mean of units matched with units in sample at time $t+1$

$\underline{Y}(t)$ vector of survey responses at time t

$\underline{Y}(t) = (\underline{Y}(1)^T, \underline{Y}(2)^T, \dots, \underline{Y}(t)^T)^T = (Y_{i_1(1)}, Y_{i_2(1)}, \dots, Y_{i_{n(1)}(1)}, Y_{i_1(2)}, \dots, Y_{i_{n(2)}(2)}, \dots, Y_{i_{n(t)}(t)})^T$ the vector of survey responses to time t

$\underline{Y}(t)_m$ vector of survey responses matched with those at time $(t-1)$ or $(t+1)$ as appropriate

$\underline{Y}(t)_u$ as above, except for unmatched units

$\underline{Y}(t) = (\underline{Y}_{t-1}^T, \underline{Y}(t)_m^T - \rho \underline{Y}(t-1)_m^T, \underline{Y}(t)_u^T)^T$

$$\underline{y}_t = (\hat{\bar{Y}}(t)_u, \hat{\bar{Y}}(t)_m - \rho \hat{\bar{Y}}(t-1)_m)^T$$

$\hat{Y}^{(t')}(t)$ best linear unbiased predictor of $\hat{Y}(t)$ given data to time $t' < t$

$\hat{\tau}_0^{(t')}(t)$ best linear unbiased estimator of $Y(t) = \tau(t)$ given data to time t' assuming $Y(t)$ is non-stochastic. $\hat{\tau}_0^{(t)}(t) = \hat{\tau}_0(t)$

$\hat{\tau}_0^{(t')}(1,t) = (\hat{\tau}_0^{(t')}(1), \hat{\tau}_0^{(t')}(2), \dots, \hat{\tau}_0^{(t')}(t))^T$ with $\hat{\tau}_0^{(t)}(1,t) = \hat{\tau}_0^{(t)}$ and $\hat{\tau}_0^{(h)} = \hat{\tau}_0$

$\hat{\tau}^{(t')}(t)$ best linear unbiased estimator of $\tau(t) = Y(t)$ given data to time t' assuming $\tau(t)$ stochastic. $\hat{\tau}^{(t)}(t) = \hat{\tau}(t)$

$\hat{\tau}^{(t')}(1,t) = (\hat{\tau}^{(t')}(1), \hat{\tau}^{(t')}(2), \dots, \hat{\tau}^{(t')}(t))^T$ with $\hat{\tau}^{(t)}(1,t) = \hat{\tau}^{(t)}$ and $\hat{\tau}^{(h)} = \hat{\tau}$

$$\hat{\tau}_t = (\hat{\tau}^{(t)}(t-1), \hat{\tau}^{(t)}(t))^T$$

$$\tau_t = (\tau(t-1), \tau(t))^T = (Y(t-1), Y(t))^T$$

$\hat{\tau}(t)^{(P)}$ Patterson's (1950) estimator at time t

$\hat{\tau}(t)^{(B,S)}$ Blight and Scott's (1973) estimator at time t

$\hat{\tau}(t)^{(S,S)}$ Scott and Smith's (1974) estimator at time t

$$\nabla \hat{\tau}(t)^{(x)} = \hat{\tau}(t)^{(x)} - \hat{\tau}^{(h)}(h-1)^{(x)}, \text{ where } (x) = (P), (B,S) \text{ or } (S,S)$$

VR signal to noise variance ratio

$e(t)$ sample survey error at time t

$e_i(t)$ sample survey error for respondent i at time t

$\underline{e}$ vector of elementary estimate errors from times 1 to h . On occasions there may be only one elementary estimate per time period, or, when sampling is simple random, each unit's response may be considered an elementary estimate.

$$\Sigma E(\underline{e} \underline{e}^T)$$

$\Sigma_{tt'}$ (t, t') th block submatrix of Σ

var variance with respect to the survey design (except in Chapter 5 where it may denote joint design-superpopulation variance)

var	variance with respect to the superpopulation
ρ	correlation between individual responses one time period apart under a first order autoregressive assumption
$\sigma_i^2 = \epsilon(Y_i^2) - (\epsilon(Y_i))^2$	
$\sigma_{ij} = \epsilon(Y_i Y_j) - \epsilon(Y_i) \epsilon(Y_j)$	
$\Sigma_{ij} = \epsilon(\underline{Y}_i \underline{Y}_j^T) - \epsilon(\underline{Y}_i) \epsilon(\underline{Y}_j)^T$	
S^2	traditional sampling variance $S^2 = \frac{1}{N} \left[\sum_{i=1}^N (Y_i - \bar{Y})^2 \right]$
$T = var(\underline{Y})$ where $\underline{Y} = (Y(1), \dots, Y(h))^T$	
$\underline{\epsilon}_{(h)}$	vector of errors associated with $\underline{Y}_{(h)}$
$\Sigma_{(h)} = var \underline{\epsilon}_{(h)}$	
λ	correlation between consecutive population means $Y(t)$ and $Y(t+1)$ for a first order autoregressive process, under a superpopulation model
σ_α^2	innovation variance for the above AR(1) process
$\{\epsilon(t)\}$	innovation process for $\hat{Y}(t)$ and $\hat{\tau}^{(t)}(t+l)$ when expressed in moving average form
σ_ϵ^2	variance of $\epsilon(t)$
$\sigma_{h h-1}^2 = \epsilon E(Y(h) - \hat{\tau}^{(h-1)}(h))^2$	
$\{b(t)\}$	innovation process for $e(t)$ represented in ARMA (autoregressive moving average) form
σ_b^2	variance of $b(t)$
$C_{\phi\psi}(\ell)$	covariance between two stationary random variables ϕ, ψ at lag ℓ . $C_{\phi\phi}(\ell) = C_\phi(\ell)$
$\hat{C}_{\phi\psi}(\ell)$	estimate of $C_{\phi\psi}(\ell)$

$$g_{\phi\psi}(z) = \sum_{-\infty}^{\infty} c_{\phi\psi}(\ell) z^{\ell}$$

$$\Lambda = E\varepsilon(\hat{\tilde{y}}_{(h)} - E\varepsilon(\hat{\tilde{y}}_{(h)}))(\hat{\tilde{y}}_h - E\varepsilon(\hat{\tilde{y}}_{(h)}))^T$$

K_t is the number of elementary estimates at time t ; if independent of t , $K_t = K$; $\sum K_t$ for times 1 through h is denoted r in Chapter 1 and N_h in Chapter 4

$\psi_j(t)$ denotes the j^{th} orthogonal polynomial at time t

$\{\psi_0(t), \psi_1(t), \dots, \psi_{p-1}(t)\}$ denotes the set of such polynomials to order $p-1$;

$$\text{Note: } \sum_{\ell=1}^L \psi_j(t_{\ell})\psi_{j'}(t_{\ell}) = \begin{cases} 0 & \text{if } j \neq j' \\ 1 & \text{if } j = j' \end{cases}$$

$\tilde{y}_h$ is the data for a polynomial model (Chapter 6); $\tilde{y}_h = \tilde{y}_{(h)}$ when individual responses are being analysed, else $\tilde{y}_h$ has as components elementary estimates $\{\hat{y}_j(t)\}$

$\hat{\tilde{y}}_i$ has as component elements either $y_i(t)$ with i fixed if data for a polynomial model are individual responses, or $\hat{y}_i(t)$ if they are elementary estimates from the i^{th} elementary group

$\tilde{y}$ is a vector of dependent variables in a polynomial model

N' is the number of data points at each time in a random coefficient polynomial regression model

Notes:

- Equations referenced within the chapter where they are defined, are referred to by equation number within section only (e.g. (3.2) refers to equation 2 in section 3 of the current chapter). Equations mentioned in the current chapter, but defined in other chapters are further prefaced by number of the chapter in which they are defined.

2. Section numbers within chapters do not include the current chapter number.
3. Figures and Tables are separately sequenced to include chapter number, followed by relevant section number, followed by number within section.

CHAPTER 1

REPEATED SAMPLE SURVEYS - THE HISTORICAL CONTEXT

SUMMARY

In the historical development of parameter estimation for repeated, probability based, sample surveys, two distinct paths have been followed. In the first of these, the parameters of interest $\underline{y} = Y(1), Y(2), \dots, Y(h)$ are treated as fixed constants. The earliest of the papers on repeated sample survey estimation were based on this premise. Following some introduction, the development of estimates for $\underline{y}$, based on this fixed parameter assumption, are discussed in Section 3. The final section of the chapter, Section 4, follows the alternative line, that $\underline{y}$ is a random parameter vector with stochastic properties with respect to some superpopulation, and details the development of this approach.

* * * * *

1. INTRODUCTION

Probability based sample surveys are routinely used in economics and social science, by industry, government and research institutions to assess the characteristics of populations to desired accuracy with cost kept as near as practicable to a minimum. Such sample surveys are often repeated at regular intervals with the aim of monitoring not only the level of some parameter of interest but also its change in the interval between surveys. With regular repetition of surveys additional accuracy can be gained from the time series structure of the survey responses or of the parameter of interest, when compared with the conventional or traditional sample survey approach of separate parameter estimation at each time period.

We consider the traditional sample survey approach in more detail. In mathematical terms, we desire a good estimate of a population quantity Y using the variable y . A sampling frame is constructed from the members of the population and denoted $\{i : 1 \leq i \leq N\}$ where N is the population size, and each population element is labelled with a unique number i . A random sample s (not necessarily a simple random sample) of size n is selected from the sampling frame using the selection procedure P , where s is an element of the set of all possible samples, S . For each element in the sample ($i \in s$), the associated value of the variable y is determined. Using the set $\{y_i : i \in s\}$ of such sample values, together with an estimator $\hat{Y}$, we determine our estimate of Y , namely $\hat{y}$. (In what follows we may often not distinguish the estimator from the estimate, denoting both by $\hat{Y}$ to avoid some notational complexity.)

Designing a traditional sample survey is then the problem of finding a good sample design, $(P, \hat{Y})$, so that the sampling distribution of $\hat{Y}$ has a mean equal, or close to, Y , and its mean square error, $MSE(\hat{Y})$ is smaller than other practicable sample designs.

$$\text{MSE}(\hat{Y}) = E(\hat{Y} - Y)^2 \quad (1.1)$$

$$= E(\hat{Y} - E(\hat{Y}))^2 + (E(\hat{Y}) - Y)^2$$

$$= \text{variance}(\hat{Y}) + [\text{bias}(\hat{Y})]^2$$

where E denotes expectation in the discrete sense, i.e.

$$E(\hat{Y}) = \sum_{s \in S} \hat{Y} p(s)$$

$$E(\hat{Y}^2) = \sum_{s \in S} \hat{Y}^2 p(s)$$

and $p(s)$ is the probability of drawing the sample s .

To measure the level of Y at two different times, say t_1 and t_2 , or the change in Y between t_1 and t_2 , using a repeated sample survey and traditional sample survey estimators, let $Y(t_1)$, $Y(t_2)$ denote Y and $\hat{Y}(t_1)$, $\hat{Y}(t_2)$ denote the estimate $\hat{Y}$ at times t_1 and t_2 respectively. We determine $\hat{Y}(t_1)$ by the procedure above carried out at time t_1 . To determine $\hat{Y}(t_2)$ we update the sampling frame if necessary by removing population "deaths" and adding "births" between t_1 and t_2 , then reapply the selection procedure P and the estimator $\hat{Y}$.

The traditional sample survey estimates of $Y(t_1)$, $Y(t_2)$ and $[Y(t_2) - Y(t_1)]$ are then $\hat{Y}(t_1)$, $\hat{Y}(t_2)$ and $\hat{Y}(t_2) - \hat{Y}(t_1)$ respectively. Note that for any time t we may write

$$\hat{Y}(t) = Y(t) + e(t). \quad (1.2)$$

where $Y(t)$ is the population parameter of interest at time t
 $\hat{Y}(t)$ is the conventional sample survey estimator at time t
 $e(t)$ is the sample survey error at time t .

The time t is usually viewed as discrete, incrementing by one with each repetition of the sample survey providing the repeated sample surveys are

equally spaced.

For the conventional sample survey design $\hat{Y}(t)$ and $e(t)$ are random variables, and $Y(t)$ a fixed constant. $\hat{Y}(t)$ is, however, random only via the randomisation involved in the selection procedure P ; the values for individual population elements are not themselves realisations of random variables.

Given one time period only there is an alternative formulation of the sample survey design problem. This is outlined for example by Royall and Herson (1973) (a) (b). It involves the use of a superpopulation model, and what Royall and Herson call the "prediction approach". A superpopulation model assumes that the population values $\{y_i : i = 1, 2, \dots, N\}$ (not all of which are known, except when a census is conducted) are themselves realisations of the random variables $\{Y_i : i = 1, 2, \dots, N\}$, i.e. that the conventional sample survey population is itself a simple random sample of size N from a hypothetical superpopulation. Then if, for example, Y is a population total, our problem is one of estimating Y by predicting $\sum_{i \notin s} Y_i$ given $\{y_i : i \in s\}$. One corollary of this approach is that we can now consider situations where Y is itself a random variable rather than a fixed constant, with an expectation with respect to the superpopulation denoted by $\epsilon(Y) = \theta$.

Given the population has been selected from the superpopulation our sample survey design $(P, \hat{Y})$ proceeds much as before. The exception is that the estimator $\hat{Y}$ may now incorporate auxiliary information via a superpopulation model, that stipulates a predictor of Y_i for $i \notin s$. The conventional view of sample survey design can be considered to be a degenerate superpopulation model, where the superpopulation size is N , and $Y = \epsilon(Y) = \theta$. Superpopulation models for single surveys are in general a wider class of models than those considered here. Interested

readers are referred, for example, to Royall and Herson (1973(a):881), Cochran (1977:158) and to Chapter 2.

The distinction between conventional and superpopulation models can be extended to repeated surveys, and in the next three sections three distinct historical approaches to the improvement of sample survey estimates, in the sense of minimum variance of the corresponding estimators are discussed. Particularly, traditional sample survey design is considered in Section 2 and relaxation of assumptions so as to include estimators which make explicit use of non current data collected since the survey design, is discussed in Sections 3 and 4. These ideas are further developed in Chapter 3.

2. THE TRADITIONAL APPROACH

The first approach remains within the framework of traditional sample surveys, and may be applied to both single and repeated surveys. It has involved development of better selection procedures, P , and of better estimators, $\hat{Y}$, leading to enhancement of the survey design $(P, \hat{Y})$. For repeated surveys there is a constraint to such traditional developments. Estimators may use only respondent information for the current period, i.e. $\{y_i : i \in s\}$, together possibly with base period information known at the time of sample design (e.g. respondent information from a previous census). Information collected in the current period may not be used to improve past estimates. Nor is any use made of sample information, collected at sampling times between sample design and the current period, to determine the current estimate, except where imputation of current missing respondent information is involved. Examples of traditional estimators include the simple expansion estimator which uses only current information, and the ratio estimator which in addition makes use of base period information.

The constraint for traditional estimators also applies to traditional selection procedures. Selection procedures are determined at time of sample design either making use of base period estimation (e.g. stratified random sampling, for the determination of stratum boundaries) or not (e.g. simple random sampling). The selection procedure is not modified when the sample survey is repeated and, properly speaking, if there were such modification, we would not be dealing with the repetition of a sample survey of fixed design $(P, \hat{Y})$, but with a series of estimates $\hat{Y}(t)$, each using a different survey design $(P(t), \hat{Y}(t))$.

The traditional sample survey design strategy for a single survey, which may be in a repeated series, is expounded in the many texts on sample survey design (e.g. Hansen, Hurwitz and Madow, Vol I and II (1953) or Cochran (1977)). A brief but comprehensive overview is also provided in Chapter 2.

3. THE TIME SERIES APPROACH, WITH $\underline{Y}$ NON-STOCHASTIC

Methods of improving estimates of non-stochastic population means, totals and change between two periods have been available for some time for repeated overlapping sample surveys (e.g. Jessen (1942), Patterson (1950)). The quantities of interest $\underline{Y} = (Y(1), Y(2), \dots, Y(h))^T$ are non-stochastic, and we may again speak of their being derived from a degenerate superpopulation in which $\underline{Y} = \varepsilon(\underline{Y}) = \underline{\theta} = (\theta(1), \theta(2), \dots, \theta(h))^T$, where ε denotes expectation with respect to the superpopulation, as defined rigorously in Chapters 2 and 3. For later notational convenience, we also define $\underline{\tau} = (\tau(1), \tau(2), \dots, \tau(h))^T = \underline{Y}$ and the best linear unbiased estimator of $\underline{Y}$ when $\underline{Y}$ is non-stochastic, by $\hat{\underline{\tau}}_0$. Equation (1.2) again holds for $t = 1, 2, \dots, h$.

The assumptions of the conventional sample design are relaxed to allow that individual respondent information only, from previous periods, may conform to some time series structure. There is a consequent requirement that sample surveys be overlapping. Respondent information $\{y_i(t)\}$ but not conventional sample survey parameters $\{Y(t)\}$ have time series structure, and the respondent information $\{y_i(t)\}$ has time series structure only through associated random variables $\{Y_i(t)\}$. The best or minimum variance linear unbiased estimator of $Y(t)$, namely $\hat{\tau}_0(t)$, by incorporating this respondent time series structure, is improved in the sense of having lower variance than the corresponding traditional estimator $\hat{Y}(t)$. Given h periods of data, note that this improved estimator of $Y(t)$ is no longer a function defined only on the sample at one time period t . It is instead defined on all samples taken between the two fixed time periods, 1 and h . $\hat{\tau}_0(t)$ will be denoted $\hat{\tau}_0^{(h)}(t)$ where necessary, to indicate that it is derived from h periods of data.

Let an elementary estimator $\hat{Y}_j(t)$ of $Y(t)$ be an unbiased estimator of $Y(t)$, which makes use of survey data from time t only, i.e. $\hat{Y}_j(t)$ is a traditional sample survey estimator at time t . We use the subscript j to denote that there may be more than one such estimator at time t , usually obtained by suitable subdivision of the sample, from time t , and that $\hat{Y}_j(t)$ is the j^{th} of these. Elementary estimators arise, for example, from sampling units grouped as matched and unmatched between two given time periods, and given h periods of data we may extend our definition of design expectation, E , in equation (1.1), so that

$$E[\hat{Y}_j(t)] = \sum_{\underline{s} \in S} \hat{Y}_j(t) p(\underline{s}(t)) \quad (3.1)$$

Here $\underline{s} = (s(1), s(2), \dots, s(h))$ denoting the entire set of samples between

$t = 1$ and h , belongs to the revised sample space S so that $s(t)$ is the sample at time t .

Individual responses may be regarded as elementary estimates, if the individual label i carries no information about the associated value y_i . Although this is, strictly speaking, a superpopulation concept directly related to exchangeability of survey responses, Cassel, Sarndal and Wretman (1977:84) note in defining exchangeability that this may approximate what many traditional practitioners have in mind when they consider simple random sampling to be appropriate.

Given the definition of equation (3.1) one form of the unbiased property of elementary estimates is

$$E[\hat{Y}_j(t)] = Y(t) \quad (3.2)$$

for $j = 1, 2, \dots, K_t$ and $t = 1, 2, \dots, h$.

Let $\hat{\underline{Y}}_G$ and $\underline{\tau}$ (i.e. $\underline{Y}$) be vectors with components $\hat{Y}_j(t)$ and $\tau(t)$ (i.e. $Y(t)$) respectively, and $\underline{e}$ the vector of sampling errors for each of the $\hat{Y}_j(t)$. Then we can write

$$\hat{\underline{Y}}_G = X\underline{\tau} + \underline{e} \quad (3.3)$$

where $\hat{\underline{Y}}_G$ is $(r \times 1)$, $\underline{\tau}$ is $(h \times 1)$, $\underline{e}$ is $(r \times 1)$, X is an $(r \times h)$ design matrix of zeros and ones containing a single one in each row and K_t ones in the t^{th} column, and $r = \sum K_t$ with the sum being over the h time periods $t = 1, 2, \dots, h$.

Since the $\{\hat{Y}_j(t)\}$ are unbiased, $E(\underline{e}) = 0$. We let the variance covariance matrix of the elementary estimates be denoted Σ .

The central result in the improvement of sample survey estimates using restricted time series techniques is then simply a result in least squares, namely that the minimum variance (or best) linear unbiased estimator (b.l.u.e.) of $\underline{y} = \underline{\tau}$ is

$$\hat{\underline{\tau}}_0 = (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} \underline{X}^T \underline{\Sigma}^{-1} \hat{\underline{y}}_G \quad (3.4)$$

(h×1) (h×r) (r×r) (r×h) (h×r) (r×r) (r×1)

with $\text{var}(\hat{\underline{\tau}}_0) = (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1}$ (see, for example, Rao (1973)).

The best linear unbiased estimator of $[Y(t) - Y(t-1)]$, the change in Y between $t-1$ and t , is $[\hat{\tau}_0^{(h)}(t) - \hat{\tau}_0^{(h)}(t-1)]$. This is essentially the general result derived by Gurney and Daly (1965). In their paper an additional assumption was made that an elementary estimate at time t was made up using all those elements in the sample at time t , which had entered and were later to leave the sample together. This led to simplifications in the structure of $\underline{\Sigma}$. Their paper also completely generalised the assumption of its predecessors (e.g. Jessen (1942), Patterson (1950)) of an exponential correlation structure defined by

$$\text{corr}(Y_i(t), Y_{i'}(t')) = \begin{cases} \rho^{|t-t'|} & \text{if } i = i' \\ 0 & \text{otherwise} \end{cases} \quad (3.5)$$

This assumption is equivalent to

$$Y_i(t) - Y(t) = \rho(Y_i(t-1) - Y(t-1)) + \eta_i(t) \quad (3.6)$$

for each respondent, i , i.e. an autoregressive process of order one, AR(1), for each respondent.

The other major assumption in Gurney and Daly (1965) and also in their forerunners, Jessen (1942), Yates (1949), Patterson (1950), Cochran

(1963), Eckler (1955), Raj (1964, 1965), and Singh (1968) is that of infinite population size, N , at each time t . Limited work has been done to allow for finite population size by Rao and Graham (1964), making use of earlier work on variance estimation by Hansen, Hurwitz, Nisselson and Steinberg (1955). The immediate effect of assuming finite population sizes and sampling without replacement is to induce correlation between sample responses. When population size, N , is large comparative to sample size, n , the effect is usually sufficiently small to be considered insignificant. Further detail is given later in this section.

The pioneer work for improvement of sample survey estimators with $\tilde{Y}$ nonstochastic, was that of Jessen (1942), and involved the improvement of the estimate of $Y(t)$, in this case the population mean, by incorporation of information collected at time $t-1$. The samples at each of $t-1$ and t were divided into two parts; the two parts consisting of matched and unmatched units, with each matched unit by definition occurring in both of the samples, each unmatched unit in one sample only. Jessen assumed simple random sampling, a sample of size n at both time t and time $t-1$, common sampling variance, infinite population size, and an exponential correlation structure for responses at the two consecutive time periods. The extension of the results to allow for unequal sample sizes and/or variances presents no complications. Although Jessen's results were derived without the benefit of matrix algebra they can more simply be derived using the least squares formulation (3.3) and (3.4).

Let S^2 be the common sampling variance

m be the number of matched units

$u = n - m$ be the number of unmatched units

ρ be the correlation between responses from the same individual at times $(t-1)$ and t .

Since we are dealing with only two times $t-1$ and t without loss of generality, let $t = 2$. Then

$$\hat{\tilde{Y}}_G = \begin{pmatrix} \hat{\bar{Y}}(1)_u \\ \hat{\bar{Y}}(1)_m \\ \hat{\bar{Y}}(2)_u \\ \hat{\bar{Y}}(2)_m \end{pmatrix} \quad (3.7)$$

where $\hat{\bar{Y}}(1)_u$ is mean of y for unmatched units at time 1
 $\hat{\bar{Y}}(1)_m$ is mean of y for matched units at time 1
 $\hat{\bar{Y}}(2)_u$ is mean of y for unmatched units at time 2
 $\hat{\bar{Y}}(2)_m$ is mean of y for matched units at time 2

and these are equivalent to $\hat{Y}_j(t)$, $t = 1, 2$, $j = 1, 2$, respectively, in the usual notation for elementary estimates. i.e.

$$\begin{pmatrix} \hat{\bar{Y}}(1)_u \\ \hat{\bar{Y}}(1)_m \\ \hat{\bar{Y}}(2)_u \\ \hat{\bar{Y}}(2)_m \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tau(1) \\ \tau(2) \end{pmatrix} + \begin{pmatrix} e(1)_u \\ e(1)_m \\ e(2)_u \\ e(2)_m \end{pmatrix} \quad (3.8)$$

from equation (3.3) where $e(1)_u$, $e(1)_m$, $e(2)_u$, $e(2)_m$ are the sampling errors corresponding to $\hat{\bar{Y}}(1)_u$, $\hat{\bar{Y}}(1)_m$, $\hat{\bar{Y}}(2)_u$, $\hat{\bar{Y}}(2)_m$ respectively.

Now

$$\text{var}(\hat{\tilde{Y}}_G) = \Sigma = S^2 \begin{bmatrix} 1/u & 0 & 0 & 0 \\ 0 & 1/m & 0 & \rho/m \\ 0 & 0 & 1/u & 0 \\ 0 & \rho/m & 0 & 1/m \end{bmatrix} \quad (3.9)$$

Thus

$$\Sigma^{-1} = \frac{1}{S^2} \begin{bmatrix} u & 0 & 0 & 0 \\ 0 & \frac{m}{1-\rho^2} & 0 & -\frac{m\rho}{1-\rho^2} \\ 0 & 0 & u & 0 \\ 0 & -\frac{m\rho}{1-\rho^2} & 0 & \frac{m}{1-\rho^2} \end{bmatrix} \quad (3.10)$$

and the improved sample survey estimates via equation (3.4) are

$$\hat{\tilde{\tau}}_0 = \begin{bmatrix} \hat{\tau}_0^{(2)}(1) \\ \hat{\tau}_0^{(2)}(2) \end{bmatrix} \quad (3.11)$$

where

$$\hat{\tau}_0^{(2)}(1) = \frac{1}{n^2 - u^2\rho^2} \left[u(n - u\rho^2) \hat{\bar{Y}}(1)_u + mn \hat{\bar{Y}}(1)_m + m\rho \hat{\bar{Y}}(2)_u - um\rho \hat{\bar{Y}}(2)_m \right] \quad (3.12)$$

$$\hat{\tau}_0^{(2)}(2) = \frac{1}{n^2 - u^2\rho^2} \left[m\rho \hat{\bar{Y}}(1)_u - m\rho \hat{\bar{Y}}(1)_m + u(n - u\rho^2) \hat{\bar{Y}}(2)_u + nm \hat{\bar{Y}}(2)_m \right] \quad (3.13)$$

Equation (3.13) is Jessen's estimator. Further, from least squares,

$$\text{var}(\hat{\tau}_0) = \frac{S^2}{n^2 - u^2\rho^2} \begin{pmatrix} n-uv^2 & mp \\ mp & n-uv^2 \end{pmatrix} \quad (3.14)$$

Thus

$$\text{var}(\hat{\tau}_0(2)) = \frac{S^2}{n^2 - u^2\rho^2} (n - uv^2) \quad (3.15)$$

and minimising this variance by differentiation (overlooking the fact that n and u must be integers) yields Jessen's result that the optimal overlap between surveys, π_{opt} , is

$$\pi_{\text{opt}} = \frac{m}{n} \Big|_{\text{opt}} = \frac{\sqrt{1 - \rho^2}}{1 + \sqrt{1 - \rho^2}} \quad (3.16)$$

giving

$$\text{var}(\hat{\tau}_0(2))_{\text{opt}} = \frac{S^2}{2n} [1 + \sqrt{1 - \rho^2}] \quad (3.17)$$

Note that ρ is assumed known, whereas in fact it must be estimated from the two samples via the matched units only. This requires the number of matched units to be relatively large, else the sample correlation coefficient is unlikely to be a good estimator of the population correlation coefficient. The problem of optimal allocation of numbers of matched and unmatched units, given that ρ must be estimated, is considered by Kulldorff (1963) and Robinson (1985). Note also that Jessen's method makes use of two independent estimates of $Y(2)$, namely $\hat{\bar{Y}}(2)_u$ and the double sample regression estimate $\hat{\bar{Y}}(2)_m + \rho(\hat{\bar{Y}}(1) - \hat{\bar{Y}}(1)_m)$ where $\hat{\bar{Y}}(1) = \frac{1}{n} (u \hat{\bar{Y}}(1)_u + m \hat{\bar{Y}}(1)_m)$ the traditional or conventional mean for $t = 1$. Weighting these two estimates inversely as their variances, which

are S^2/u and $S^2(1 - \rho^2)/m + \rho^2 S^2/n$ respectively, again gives equation (3.13). Other estimators than the double sample regression estimator are, however, possible, [e.g. the difference estimate $(\hat{Y}(2)_m - \hat{Y}(1)_m) + \hat{Y}(1)$ or the ratio estimate $(\hat{Y}(2)_m / \hat{Y}(1)_m) \cdot \hat{Y}(1)$ (see, for example, Hansen, Hurwitz and Madow (1953))] but their usage has been limited by the work of Patterson (1950) who showed that the regression estimate gives rise to best linear unbiased estimates (b.l.u.e.'s) under certain assumptions, including that ρ is known.

Patterson (1950) extended Jessen's (1942) results to sampling on h occasions with partial overlap of sampling units. His results too can be expressed as results in ordinary least squares applied to estimates of level and change in population means. We extend our previous definition and let $\underline{y}^{(h)} = (Y(1) \dots Y(h))^T = \underline{\tau}_{(h)} = (\tau(1) \dots \tau(h))^T$ be the population means of interest, and denote $\underline{y}^{(h)}$ by $\underline{y}$ and $\underline{\tau}_{(h)}$ by $\underline{\tau}$ where there is no ambiguity. The principal result is then an updating formula giving the b.l.u.e. of $Y(h) = \tau(h)$, after h surveys, $\hat{\tau}_0(h)$, in terms of the b.l.u.e. of $Y(h-1) = \tau(h-1)$ after $(h-1)$ surveys, $\hat{\tau}_0(h-1)$, together with suitable linear combinations of means of matched and unmatched units from time periods $(h-1)$ and h .

Patterson's (1950) results were originally derived using covariance properties of the estimates $\hat{\tau}_0(h-1)$ and $\hat{\tau}_0(h)$, but without the benefit of the matrix formulation used in the derivation below. Although not mathematically complicated in this alternative formulation, Patterson's (1950) results have been central in the repeated survey sampling literature and this alternative derivation is given here for completeness. As in Jessen's derivations, simple random sampling, exponential correlation structure, infinite population size, exchangeability of responses, equal sampling variance at each time period, are again assumed, although Patterson

himself extended his results to allow for unequal numbers, variances, or fractions of units replaced on different occasions. We note that the infinite population size implies the probability of reselection of a sampling unit, once it has dropped out of the sample, is zero.

Let $\hat{\tau}_0^{(t')}(t)$ be the b.l.u.e. of $\tau(t)$ given t' periods of data, noting that $\hat{\tau}_0^{(t)}(t)$ will usually be denoted simply as $\hat{\tau}_0(t)$. Let $\hat{\tau}_0^{(t')}(1,t) = (\hat{\tau}_0^{(t')}(1), \dots, \hat{\tau}_0^{(t')}(t))^T$, and $\hat{\tau}_0^{(t)}(1,t) = \hat{\tau}_0(t)$. Let $\underline{y}(t)$ be a vector of the $y_i(t)$ for $i \in s(t)$, the sample at time t , and $\underline{y}_{(h)} = (\underline{y}(1)^T, \underline{y}(2)^T, \dots, \underline{y}(h)^T)^T$. (Note then, that $\underline{y}_{(h)}$ is a vector of survey responses while $\underline{\tau}^{(h)} = \underline{\tau}_{(h)}$ is the vector of population means to be estimated). Also in a similar sense to previous notation, let

$\underline{y}_{(h)}_m$ = vector of $y_i(h)$'s "matched" with $y_i(h-1)$'s

i.e. $i \in s(t) \cap s(t-1)$

$\underline{y}_{(h-1)}_m$ = vector of $y_i(h-1)$'s "matched" with $y_i(h)$'s

i.e. $i \in s(t) \cap s(t-1)$

$\underline{y}_{(h)}_u$ = vector of $y_i(h)$'s "unmatched" with $y_i(h-1)$'s

i.e. $i \in s(t) \cap \overline{s(t-1)}$

Note that

(i) $\underline{y}_{(h)}_u$ is uncorrelated with $\underline{y}_{(h-1)}$, $\underline{y}_{(h)}_m$, $\underline{y}_{(h-1)}_m$.

(ii) $\underline{y}_{(h)}_m - \rho \underline{y}_{(h-1)}_m$ is uncorrelated with $\underline{y}_{(h-1)}$ and $\underline{y}_{(h)}_m$ since $\underline{y}_{(h)}_m = \rho \underline{y}_{(h-1)}_m + \text{error vector}$, by assumption of an exponential correlation structure.

(iii) The components of $\underline{y}_{(h)}_m$, $\underline{y}_{(h-1)}_m$ are, conventionally, random variables under the present model through the randomisation imposed by the sampling scheme. (They are then, implicitly

assumed to be exchangeable (see Chapter 3), and may therefore be viewed as having superpopulation properties.)

An appropriate linear model is

$$\tilde{y}_{(h)} = \begin{pmatrix} \tilde{y}_{(h-1)} \\ \tilde{y}_{(h)}_m - \rho \tilde{y}_{(h-1)}_m \\ \tilde{y}_{(h)}_u \end{pmatrix} = \begin{pmatrix} 1 & & & \\ & 1 & & 0 \\ & & 1 & \\ & 0 & -\rho 1 & 1 \\ & & & 1 \end{pmatrix} \begin{pmatrix} \tau(1) \\ \tau(2) \\ \vdots \\ \tau(h) \end{pmatrix} + \tilde{\epsilon}_{(h)} = X_{(h)} \tau_{(h)} + \tilde{\epsilon}_{(h)} \quad (3.18)$$

where $\text{cov}(\tilde{\epsilon}_{(h)}) = s^2 \begin{pmatrix} \frac{1}{s^2} \Sigma_{(h-1)} & 0 \\ \vdots & (1-\rho^2) I & 0 \\ 0 & 0 & I \end{pmatrix} = \Sigma_{(h)}$, say

and

$$X_{(h)} = \begin{pmatrix} 1 & & & \\ & 1 & & 0 \\ & & 1 & \\ & 0 & & 1 \\ & & -\rho 1 & 1 \\ & 0 & & 1 \\ & & 0 & 1 \end{pmatrix} = \begin{pmatrix} X_{(h-1)} & \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \\ 0 & \begin{pmatrix} -\rho 1 \\ 0 \\ 0 \end{pmatrix} \end{pmatrix}$$

Thus

$$(X_{(h)}^T \Sigma_{(h)}^{-1} X_{(h)}) \hat{\tau}_{(h)} = X_{(h)}^T \Sigma_{(h)}^{-1} \tilde{y}_{(h)} \quad (3.19)$$

are the normal equations (cf. equation (3.4)). $\hat{\tau}_{(h)}$ is the vector of

b.l.u.e.'s of $\tau(1)$ to $\tau(h)$ given data from time periods $1, 2, \dots, h$.

Dropping the subscript or superscript (h) except where necessary to maintain clarity, we have

$$X^T \Sigma^{-1} X = \frac{1}{S^2} \begin{pmatrix} & & 0 \\ X_{(h-1)}^T & -\rho \tilde{1}^T & \tilde{0}^T \\ \tilde{0}^T & \tilde{1}^T & \tilde{1}^T \end{pmatrix} \begin{pmatrix} \Sigma_{(h-1)}^{-1} & & 0 \\ 0 & \frac{1}{1-\rho^2} I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} X_{(h-1)} & \tilde{0} \\ -\rho \tilde{1} & \tilde{1} \\ \tilde{0} & \tilde{1} \end{pmatrix} \quad (3.20)$$

$$= \frac{1}{S^2} \begin{pmatrix} X_{(h-1)}^T \Sigma_{(h-1)}^{-1} X_{(h-1)} + \begin{pmatrix} 0 & \tilde{0} \\ \tilde{0}^T & \frac{m\rho}{1-\rho^2} \end{pmatrix} & \tilde{0} \\ \tilde{0}^T & -\frac{m\rho}{1-\rho^2} \end{pmatrix}$$

$$\begin{pmatrix} -\frac{m\rho}{1-\rho^2} & \frac{m}{1-\rho^2} + u \end{pmatrix}$$

which is an $(h \times h)$ matrix partitioned as

$$\begin{pmatrix} (h-1) \times (h-1) & (h-1) \times 1 \\ 1 \times (h-1) & 1 \times 1 \end{pmatrix},$$

and m and u are the number of matched and unmatched units at each time, respectively. Similarly,

$$X^T \Sigma^{-1} \tilde{y} = \frac{1}{S^2} \begin{pmatrix} X_{(h-1)}^T \Sigma_{(h-1)}^{-1} \tilde{y}_{(h-1)} + \begin{pmatrix} \tilde{0} \\ -\frac{m\rho}{1-\rho^2} (\hat{\tilde{y}}(h)_m - \rho \hat{\tilde{y}}(h-1)_m) \end{pmatrix} \\ \frac{m}{1-\rho^2} (\hat{\tilde{y}}(h)_m - \rho \hat{\tilde{y}}(h-1)_m) + u \hat{\tilde{y}}(h)_u \end{pmatrix} \quad (3.21)$$

where $\hat{\bar{Y}}(h)_m$ = mean of matched $Y_i(h)$'s
 $\hat{\bar{Y}}(h-1)_m$ = mean of matched $Y_i(h-1)$'s
 $\hat{\bar{Y}}(h)_u$ = mean of unmatched $Y_i(h)$'s

The normal equations (3.19) are thus

$$\begin{aligned} X_{(h-1)}^T \Sigma_{(h-1)}^{-1} X_{(h-1)} \hat{\tau}_0^{(h)}(1, h-1) + \frac{m\rho}{1-\rho^2} \hat{\tau}_0^{(h)}(h-1) \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \frac{m\rho}{1-\rho^2} \hat{\tau}_0^{(h)} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ = X_{(h-1)}^T \Sigma_{(h-1)}^{-1} \tilde{Y}_{(h-1)} - \frac{m\rho}{1-\rho^2} (\hat{\bar{Y}}(h)_m - \rho \hat{\bar{Y}}(h-1)_m) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{aligned} \quad (3.22)$$

where $\tilde{0}$ is $(h-2) \times 1$, and

$$-\frac{m\rho}{1-\rho^2} \hat{\tau}_0^{(h)}(h-1) + \frac{n-u\rho^2}{1-\rho^2} \hat{\tau}_0^{(h)} = \frac{m}{1-\rho^2} (\hat{\bar{Y}}(h)_m - \rho \hat{\bar{Y}}(h-1)_m) + u \hat{\bar{Y}}(h)_u \quad (3.23)$$

where n is the total of matched and unmatched units at each time period.

Multiplying equation (3.22) by $(X_{(h-1)}^T \Sigma_{(h-1)}^{-1} X_{(h-1)})^{-1}$, noting that

$$(X_{(h-1)}^T \Sigma_{(h-1)}^{-1} X_{(h-1)})^{-1} X_{(h-1)}^T \Sigma_{(h-1)}^{-1} \tilde{Y}_{(h-1)} = \hat{\tau}_0^{(h-1)}$$

the vector of b.l.u.e.'s of $\tau(1)$ to $\tau(h-1)$ given $(h-1)$ periods only, and denoting

$$\tilde{b} = \text{last column of } (X_{(h-1)}^T \Sigma_{(h-1)}^{-1} X_{(h-1)})^{-1} / S^2$$

yields

$$\hat{\tau}_0^{(h)}(1, h-1) = \frac{m\rho}{1-\rho^2} ((\hat{\tau}_0^{(h)}(h) - \rho \hat{\tau}_0^{(h)}(h-1)) - (\hat{\bar{Y}}(h)_m - \rho \hat{\bar{Y}}(h-1)_m)) \tilde{b} + \hat{\tau}_0^{(h-1)} \quad (3.24)$$

We take the last equation of the set of equations (3.24), namely

$$\hat{\tau}_0^{(h)}(h-1) = \frac{m\rho}{1-\rho^2} ((\hat{\tau}_0^{(h)}(h) - \rho\hat{\tau}_0^{(h)}(h-1)) - (\hat{Y}(h)_m - \rho\hat{Y}(h-1)_m))b_{h-1} + \hat{\tau}_0^{(h)}(h-1) \quad (3.25)$$

where b_{h-1} is the $(h-1)$ th element of $\underline{b}$, $((h-1) \times 1)$ (i.e. b_{h-1} is the $((h-1), (h-1))$ th element of $(X_{(h-1)}^T \Sigma_{(h-1)}^{-1} X_{(h-1)})^{-1} = \frac{1}{S^2}$ variance $(\hat{\tau}_0^{(h)}(h-1))$).

This equation and equation (3.23) can be solved simultaneously yielding

$$\hat{\tau}_0^{(h)}(h) = (1 - \varphi_h) [\hat{Y}(h)_m + \rho(\hat{\tau}_0^{(h)}(h-1) - \hat{Y}(h-1)_m)] + \varphi_h \hat{Y}(h)_u \quad (3.26)$$

which is equation (3.8) of Patterson's paper (1950), where

$$\varphi_h = \frac{u(1 - \rho^2(1 - mb_{h-1}))}{n - u\rho^2(1 - mb_{h-1})} = \frac{u(1 - \rho^2(1 - mb_{h-1}))}{m + u(1 - \rho^2(1 - mb_{h-1}))} \quad (3.27)$$

This result also appears unproven in Yates (1949). More generally, if numbers of matched and unmatched units and sample sizes differ from time period to time period

$$\varphi_h = \frac{u_h(1 - \rho^2(1 - m_h b_{h-1}))}{n_h - u_h \rho^2(1 - m_h b_{h-1})} \quad (3.28)$$

where u_h , m_h and n_h are, for time h , the number of matched, unmatched and total sampled units respectively.

The fact that

$$\text{var}(\hat{\tau}_0^{(h)}(h)) = \text{cov}(\hat{\tau}_0^{(h)}(h), \hat{Y}(h)_u) = \varphi_h \frac{S^2}{u_h} \quad (3.29)$$

follows from the following lemma, which (given the results of Duncan and Horn (1972) where it is shown that stochastic and non-stochastic least squares may be reparameterised to have the same solution) extends a result first proven by Fisher (1925).

Lemma 3.1

If $\hat{\tau}_0$ is a b.l.u.e. of τ given the linear model

$$y = X\tau + \epsilon$$

with $\text{var } \epsilon = \Sigma$, then if $\tilde{\tau}$ is any linear unbiased estimator of τ

$$\text{var}(\hat{\tau}_0) = \text{cov}(\hat{\tau}_0, \tilde{\tau}).$$

Proof:

Now $\hat{\tau}_0 = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y = L y$ say, with $\text{var } \hat{\tau}_0 = (X^T \Sigma^{-1} X)^{-1}$ and $\tilde{\tau}$ is a linear unbiased estimator of τ . Hence

$$\tilde{\tau} = K y \quad \text{for some matrix } K \text{ (linearity)}$$

$$\text{and} \quad E[K y] = \tau \quad \text{(unbiasedness)}$$

$$\text{i.e.} \quad K X \tau = \tau \quad \text{since } E[y] = X \tau$$

$$\text{Hence} \quad K X = I.$$

Thus

$$\begin{aligned}
 \text{cov}(\hat{\tau}_0, \hat{\tau}) &= \text{cov}(LY, KY) \\
 &= \text{cov}(KY, LY) \\
 &= K \Sigma L^T \\
 &= K \Sigma \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} \\
 &= K X (X \Sigma^{-1} X)^{-1} \\
 &= (X \Sigma^{-1} X)^{-1} \\
 &= \text{var}(\hat{\tau}_0) \quad \text{as required.}
 \end{aligned}$$

Note that the proof applies regardless of the form of X , τ and Σ (i.e. for any general linear model) and that τ may, in general, be stochastic or nonstochastic, providing that the expectation operator is defined accordingly, and the model defined as for Duncan and Horn (1972). In later sections a b.l.u.e. of τ will be denoted simply $\hat{\tau}$ when τ is stochastic.

Returning to the particular situation and considering the h^{th} component of the vectors $\hat{\tau}_0$ and $\hat{\tau}$, it follows as a corollary that when $\hat{\tau}$ has h^{th} component $\hat{Y}(h)_u$ the first part of equation (3.29) holds. Also, by the above lemma

$$\text{cov}(\hat{Y}(h)_m, \hat{\tau}_0(h)) = \text{var}(\hat{\tau}_0(h)) = \text{cov}(\hat{Y}(h)_u, \hat{\tau}_0(h)) \quad (3.30)$$

Thus from equations (3.26) and (3.29)

$$(1 - \varphi_h) \left(\frac{S^2}{m_h} (1 - \rho^2) + \rho^2 \text{var}(\hat{\tau}_0(h-1)) \right) = \varphi_h \frac{S^2}{u_h} \quad (3.31)$$

Solving equations (3.23) and (3.24) simultaneously, to eliminate $\hat{\tau}_0(h)$, also leads to

$$\hat{\tau}_0^{(h)}(h-1) = \hat{\tau}_0(h-1) - w_h(\hat{\tau}_0(h) - \hat{\bar{Y}}(h)_u) \quad (3.32)$$

where

$$w_h = \frac{\rho \varphi_{h-1} u_h}{u_{h-1}} = \rho b_{h-1} u_h$$

via equation (3.29) and the definition of b_{h-1} in equation (3.25). This formula, first derived by Patterson, is useful for updating the b.l.u.e. of $\tau(h-1)$ when given the extra data from the h^{th} period.

Also

$$\begin{aligned} \text{var}(\hat{\tau}_0^{(h)}(h-1)) &= \text{cov}(\hat{\tau}_0^{(h)}(h-1), \hat{\bar{Y}}(h-1)_u) \quad \text{by lemma (3.1)} \\ &= \text{cov}(\hat{\tau}_0(h-1), \hat{\bar{Y}}(h-1)_u) \\ &\quad - u_h \rho b_{h-1} (1 - \varphi_h) \text{cov}(\{\hat{\bar{Y}}(h)_u - \hat{\bar{Y}}(h)_m + \rho \hat{\bar{Y}}(h-1)_m \\ &\quad + \rho \hat{\tau}_0(h-1)\}, \hat{\bar{Y}}(h-1)_u) \quad \text{via equations (3.32) and (3.26)} \\ &= (1 - u_h \rho^2 b_{h-1} (1 - \varphi_h)) \text{cov}(\hat{\tau}_0(h-1), \hat{\bar{Y}}(h-1)_u) \\ &\quad \text{since } \text{cov}(\hat{\bar{Y}}(h)_u, \hat{\bar{Y}}(h-1)_u) = \text{cov}(\hat{\bar{Y}}(h)_m, \hat{\bar{Y}}(h-1)_u) \\ &\quad = \text{cov}(\hat{\bar{Y}}(h-1)_m, \hat{\bar{Y}}(h-1)_u) = 0 \\ &= (1 - u_h \rho b_{h-1} (1 - \varphi_h)) \text{var } \hat{\tau}_0(h-1) \quad \text{by lemma (3.1)} \\ &= (1 - u_h \rho b_{h-1} (1 - \varphi_h)) b_{h-1} s^2 \end{aligned}$$

via definition of b_{h-1} following equation (3.25). i.e.

$$\text{var}(\tau_0^{(h)}(h-1)) = \frac{s^2 \varphi_{h-1}}{u_{h-1}} (1 - w_h \rho (1 - \varphi_h)) \quad (3.33)$$

via the definition of w_h following equation (3.32).

In estimating change between two periods, i.e. estimating $[Y(h) - Y(h-1)] = [\tau(h) - \tau(h-1)]$, we can distinguish two possible estimates $[\hat{\tau}_0(h) - \hat{\tau}_0^{(h)}(h-1)]$ and $[\hat{\tau}_0(h) - \hat{\tau}_0(h-1)]$. The difference between these estimators is given via equation (3.32). Their corresponding variances are

$$\begin{aligned} \text{var}(\hat{\tau}_0(h) - \hat{\tau}_0^{(h)}(h-1)) &= \frac{\varphi_h}{u_h} s^2 + \frac{\varphi_{h-1} s^2}{u_{h-1}} \{1 - w_h \rho (1 - \varphi_h)\} - 2(1 - \varphi_h) \frac{\rho \varphi_{h-1} s^2}{u_{h-1}} \\ &= s^2 \left[\frac{\varphi_h}{u_h} + \frac{\varphi_{h-1}}{u_{h-1}} \{1 - w_h \rho (1 - \varphi_h) - 2\rho (1 - \varphi_h)\} \right] \end{aligned} \quad (3.34)$$

via equation (3.33), and

$$\text{var}(\hat{\tau}_0(h) - \hat{\tau}_0(h-1)) = \frac{\varphi_h s^2}{u_h} + \frac{\varphi_{h-1} s^2}{u_{h-1}} \{1 - 2\rho (1 - \varphi_h)\} \quad (3.35)$$

The difference between these two variances is

$$\frac{\varphi_{h-1} s^2}{u_{h-1}} w_h \rho (1 - \varphi_h) = \frac{s^2}{n} \left[g_{h-1}^2 \rho^2 \frac{u_h}{n} \left(1 - \frac{u_h g_h}{n} \right) \right] \quad (3.36)$$

where

$$g_h = \frac{\phi_h u_h}{n}$$

This difference in variance is clearly small comparative to terms of order $g_h \frac{s^2}{n}$, as is the difference between the two estimators, namely $\tau_0^{(h-1)} - \tau_0^{(h)}(h-1)$, comparative to either of them.

Patterson's (1950) estimators are dependent on there being some unmatched units at each time period (i.e. units unmatched with the previous time period). This is the compromise strategy we would need to adopt if level and change were to be estimated using the same repeated survey. If, however, estimates of change between two time periods were the only statistic of interest, the optimum strategy would be a panel study (i.e. to retain all sampling units from survey to survey) since the correlation between the same individual at two consecutive time periods, namely ρ , is nearly always positive. Thus estimated change for an individual has variance $s_t^2 + s_{t-1}^2 - 2\rho s_t s_{t-1}$, while that for two different individuals has variance $s_t^2 + s_{t-1}^2$ (again assuming infinite population size), where s_{t-1}^2, s_t^2 are the possibly different sampling variances at time $t-1$ and t respectively. If complete overlap is used the variance of the estimate of change is $2s^2(1-\rho)/n$ for fixed sample size n , and constant sampling variance s^2 .

Cochran (1963) investigated the optimum overlap for Patterson's (1950) estimator when all samples are of equal size, and showed that in the limit optimum overlap for estimating level would reselect half the sampling units on each occasion irrespective of the value of the correlation ρ .

The case when proportions of matched and unmatched units remain constant, rather than being allowed to change on each occasion, was also

considered by Patterson (1950) and Cochran (1963). We again get an equation of the form of equation (3.26) for the b.l.u.e. of $\tau(h)$ with variance given by equation (3.29) with $u_h = u$ and φ_h given by equation (3.27). Under these conditions φ_h tends quickly to a limiting value φ given by

$$\varphi = \frac{\sqrt{1 - \rho^2} [\sqrt{1 - (1 - 4mu/n^2)\rho} - \sqrt{1 - \rho^2}]}{2\rho^2 m/n} \quad (3.37)$$

Computations by Cochran have shown that there is little loss in the efficiency in estimating $\tau(h)$, if a constant weight φ and a fixed proportion of overlap, say $m/n = \frac{1}{2}$, is used. If a number of different variables are being estimated using the same sample survey there is also little loss in using an average value of φ for the different correlation coefficients being considered.

If we are considering estimates of change between two periods in addition to estimates of level, Cochran has shown numerically (without optimising for level and change simultaneously), that there is a considerable gain in a matching of $3/4$ of the sampled units between two time periods as against $1/2$. Such a procedure while only slightly increasing the variance of the estimate of level, markedly reduces the variance of the estimate of change.

Narain (1954) has derived formulae for b.l.u. estimation when the autocorrelation between successive survey responses follows an autoregressive process of order ≥ 2 .

Eckler's (1955) paper extended the results of Patterson (1950) for the first order autocorrelation structure by considering what he has

FIG. 1.3.1 ONE LEVEL ROTATION SCHEME

FIG. 1.3.2 TWO LEVEL ROTATION SCHEME

FIG. 1.3.3 THREE LEVEL ROTATION SCHEME

called one-level, two-level and three-level rotation sampling. A one-level rotation sample requires current information only from respondents at each time period, a two-level rotation sample requires of respondents their information for the current and the previous period, and in a three-level rotation sample respondents provide information for the current and two previous periods.

More precisely, a one-level rotation pattern is assumed to have n sample values in it at each time period $t = 1, 2, \dots, h$. m of the units are retained at t from the sample drawn at time $t-1$. The remaining $n-m = u$ units are replaced with u new ones which have never before been selected. A one-level rotation sample thus has the form illustrated in Fig. 1.3.1, where the x 's represent individual respondents (i.e. units sampled) in time periods 1 through h , and both matched and unmatched units sampled at any given time period are asked for information relevant to that time period only. (This last fact is indicated for time period h by the solid vertical and horizontal lines in Fig. 1.3.1.) The one-level rotation sample constitutes a particular type of sampling scheme for which Patterson's (1950) results apply. Note that Patterson's (1950) results do not necessarily require that a particular unit be dropped from the sample a given number of time periods after being introduced to it, as does the one-level rotation sample (e.g. in Fig. 1.3.1 above, three time periods elapse between introduction and departure of a unit from the sample).

A two-level rotation pattern has the form illustrated in Fig. 1.3.2. Each respondent at time period t is a new respondent and is to supply information for time periods t and $t-1$. This is indicated by the vertical and horizontal lines in Fig. 1.3.2. Thus information of the form $Y_i(t-1)$ (i.e. a given horizontal line of Fig. 1.3.2) is collected at two time periods, t and $t-1$. This leads to the equations

$$\hat{\tau}_0(h) = \bar{Y}(h)_1 - a_h \bar{Y}(h-1)_1 + a_h \hat{\tau}_0(h-1) \quad (3.38)$$

where $a_h = \frac{\rho}{2 - a_{h-1}\rho}$, $a_1 = 0$

with

$$\text{var}(\hat{\tau}_0(h)) = \frac{S^2}{n} (1 - a_h \rho) \quad (3.39)$$

The three-level rotation pattern has the form given in Fig. 1.3.3. All respondents at a given time, t , are new respondents and each is asked to supply information on time periods t , $t-1$ and $t-2$. Thus information of the form $Y_i(t-2)$ is collected at three time periods, namely $t-2$, $t-1$ and t . Relevant formulae for the three-level rotation pattern can be found in Eckler (1955). Eckler also provides a criterion, in the form of a graph, for determining which of one-level, two-level and three-level rotation sampling is the best procedure.

Eckler's results assume infinite population size, constant sampling variance S^2 independent of time, and an exponential correlation structure, together with the particular assumptions of one, two or three-level rotation sampling. His work has been extended to other rotation schemes by Manousakis (1977).

Other particular directions in which Patterson's (1950) work have been extended include finite population size adjustments, allowance for non-simple random sampling schemes, for auxiliary variables (Tikkiwal (1960), Sen (1972)), a multivariate extension to allow for simultaneous estimation of means for a number of variables in the same survey (Tikkiwal (1967)), and studies on the effect of rotation group bias (Bailar (1975)). The more important of these papers are discussed below.

We begin with finite population size adjustments. Rao and Graham (1964) developed a finite population theory for a composite estimator of the form

$$\tilde{\tau}_0(h) = (1 - \varphi)(\tilde{\tau}_0(h-1) + \hat{\bar{Y}}(h)_m - \hat{\bar{Y}}(h-1)_m) + \varphi \hat{\bar{Y}}(h)_u \quad (3.40)$$

where $\tilde{\tau}_0(h)$ is the estimator of $\tau(h)$ given h periods data (cf. equation (3.26) with $\rho = 1$, $\varphi_h = \varphi$). The variance of $\tilde{\tau}_0(h)$ is given in an earlier paper by Hansen, Hurwitz, Nisselson and Steinberg (1955), assuming infinite population size and an exponential correlation pattern. Rao and Graham show that the effect of finite population is to reduce this variance by a factor of S^2/N where N is the population size. They also consider the effect of an arithmetic rather than an exponential correlation pattern, i.e. the correlation structure of a moving average process. Equation (3.40) is not, however, a best estimate in the sense of minimum variance for either correlation structure. It generally, however, allows for considerable improvement in estimating current mean and change over the traditional estimates. Its importance lay in its use for the monthly Current Population Survey of the U.S. Bureau of the Census where the sample design used a rotation pattern in which a unit stayed in the sample for four occasions, temporarily left the sample for eight occasions, was reintroduced for a further four, then dropped out forever. The 75% month to month overlap and the 50% year to year overlap allowed for considerable improvement in estimation of both month to month and year to year change. As an aside, note that there is no inherent conflict with Patterson's requirement that once a unit is dropped from the sample its probability of reselection is zero, provided we allow that a unit is theoretically in the sample but dormant when it temporarily leaves the sample for eight occasions. Further details on the U.S. Bureau of the Census use of

rotating samples can be found in Woodruff (1963). Such composite estimators are also discussed in Wolter (1979).

Three central papers which have considered non-simple, random sampling techniques are Raj (1964, 1965) and Singh (1968). Raj (1964, 1965) considered sampling on two occasions only, where current estimates are of chief interest, but where sample units to be replaced are clusters selected with probability proportional to size (pps). He mentions two surveys where this sampling scheme operated, namely the U.S. Current Population Survey and the National Sample Survey of Urban India.

Raj weights two estimators $\hat{\bar{Y}}(2)_m$ and $\hat{\bar{Y}}(2)_u$ inversely as their variances where

$$\hat{\bar{Y}}(2)_m = \frac{1}{m} \sum_{a \in s(2)_m} \left(\frac{\bar{Y}_a(2) - \bar{Y}_a(1)}{p_a} \right) + \frac{1}{n} \sum_{a \in s(1)} \frac{\bar{Y}_a(1)}{p_a} \quad (3.41)$$

where a denotes cluster selected with pps, with means $\bar{Y}_a(1)$, $\bar{Y}_a(2)$ at times 1 and 2 respectively, $s(1)$ is the sample at time 1, $s(2)_m$ is that part of the sample at time 2 which is matched (i.e. the units also occur in the sample at time 1), and m clusters of n are matched, and

$$\hat{\bar{Y}}(2)_u = \frac{1}{u} \sum_{a \in s(2)_u} \frac{\bar{Y}_a(2)}{p_a} \quad (3.42)$$

where $s(2)_u$ is that part of the sample at time 2 which is unmatched, and u is the number of sampled but unmatched clusters. The resulting estimator has the property of minimum variance.

Singh (1968) considers the effect of multi-stage sample designs on Patterson's (1950) results. For a single stage stratified random sample

repeated on h occasions he shows that the b.l.u.e. of $\tau(h)$ is the appropriate linear combination of b.l.u.e.'s for the simple random samples within strata. Singh extends these results to two stage sample designs by assuming that the correlation between responses on different occasions can be expressed in two parts: namely the correlation between secondary sampling units (s.s.u.'s) within primary sampling units (p.s.u.'s) and the correlation between p.s.u. means, on the different occasions.

Singh gives formulae for the b.l.u.e. and its corresponding variance for the most recent occasion, and the b.l.u.e. together with its variance for a parameter average over all occasions, for three distinct rotation schemes involving a maximum of three time periods. Generalisation of the work of Raj (1964, 1965) and Singh (1968) to data from h time periods has been considered by Arnab (1979). Other developments for non-simple random sampling schemes include work on the design of rotation schemes for multi-stage samples (Jain (1978), (1981)).

The central result, however, for improvement of sample survey estimators from repeated surveys where parameters are viewed as non-stochastic (i.e. composite estimation) remains that of Gurney and Daly (1965). Although their result giving b.l.u.e.'s is essentially a well known result in least squares, the pioneer work of Jessen (1942) and of Patterson (1950) can never the less be seen as two special cases. The multivariate extension of Tikkiwal (1967) involves no major complications in principle. It is worth noting however that estimates are strictly b.l.u.e. iff sample survey error covariance matrices for elementary estimates are assumed known. Further comments on this assumption can be found at the end of the next section.

4. THE TIME SERIES APPROACH, WITH $\underline{Y}$ STOCHASTIC

Both the traditional sample design and its modification outlined in Section 3 treat the population means or totals $Y(t)$, as fixed constants. For traditional estimators this involves no contradiction, since if, for example, we have a simple random sample of size n from a population of size N , we estimate Y by

$$\hat{Y} = \frac{1}{n} \sum_{i \in s} Y_i$$

where s is the sample, and where the inherent randomisation in $\hat{Y}$ and Y_i only occurs via choosing each Y_i from a fixed set $\{Y_i : i = 1, 2, \dots, N\}$, i.e. sampling procedure randomisation only. Thus we take

$$Y = \frac{1}{N} \sum_{i=1}^N Y_i$$

as a fixed quantity rather than a random variable. This argument applies at each time t . For time series estimators ($\underline{Y} = (Y(1), Y(2), \dots, Y(h))^T$ stochastic) the argument goes a little deeper. Letting t denote time, the $\{Y_i(t) : i \in s(t) \cap s(t-1)\}$ are again assumed to have some known correlation structure, for example,

$$Y_i(t) - Y(t) = \rho(Y_i(t-1) - Y(t-1)) + \eta_i(t), \quad i \in s(t-1) \cap s(t).$$

If we consider, as an example, simple random sampling on two occasions from a population of size N , then since the selection of sample elements is random, any population element may be in the sample. The previous equation therefore applies for each and every population element, i.e.

$$Y_i(t) - Y(t) = \rho(Y_i(t-1) - Y(t-1)) + \eta_i(t), \quad i = 1, 2, \dots, N.$$

Irrespective of the stochastic status of Y , for the $\{Y_i(t) : i = 1, 2, \dots, N\}$ to be stochastic quantities in this equation, they must have stochastic properties not with respect to the sampling scheme but with respect to some superpopulation for which $\{y_i : i = 1, 2, \dots, N\}$ is a particular realisation. Thus (unless $\sum_{i=1}^N e_i(t)$ is fixed)

$$Y(t) = \frac{1}{N} \sum_{i=1}^N Y_i(t)$$

is strictly speaking no longer a fixed quantity but itself a random variable. The estimation methods of Section 3 have assumed it to be a fixed quantity.

Time series estimators for $\underline{Y}$ stochastic, by allowing $Y(t)$, $t = 1, 2, \dots, h$ to be random variables, permit standard time series methodology to be applied. Allowing $\underline{Y}$ to be stochastic is equivalent to a particular form of infinite superpopulation model. Analogous arguments apply to other sampling schemes, and when the parameter of interest is a total or a difference between two periods.

There are other arguments in favour of time series methods for improved estimation of means, totals and differences, including the use of repeated sample surveys in economic prediction where time series structure is often assumed, and the requirement for improved estimation when surveys do not overlap but a relationship between $Y(t)$ and $Y(t')$, $t = 1, 2, \dots, h$, $t' = 1, 2, \dots, t-1$, obviously exists.

As for $\underline{Y}$ non-stochastic, we consider the model of equation (1.2), namely $\hat{Y}(t) = Y(t) + e(t)$, where $e(t)$ is the sample survey error at time t , with joint superpopulation design expectation zero and variance S^2 , where $\hat{Y}(t)$ is the conventional sample survey estimator at time t , but where $Y(t)$ now has stochastic properties.

Blight and Scott (1973) were first to allow that the population means $Y(t)$ may exhibit such stochastic structure with respect to a superpopulation. The time series structure they imposed on the $Y(t)$ was that of a linear Markov process, namely

$$Y(t) - \mu = \lambda(Y(t-1) - \mu) + \alpha(t) \quad (4.1)$$

for $t = 2, 3, \dots$, where $\{\alpha(t) : t = 2, 3, \dots\}$ is an uncorrelated process with respect to the superpopulation having zero mean and variance σ_α^2 (i.e. $E(\alpha(t)^2) = \sigma_\alpha^2$, $E(\alpha(t), \alpha(t')) = 0$ for $t' = t-1, t-2, \dots$), and the mean $E(Y(t)) = \theta(t) = \mu$ at all times t . Blight and Scott (1973) assume μ is known and set $\mu = 0$ without loss of generality. The random variables representing the individual responses associated with the sampled population elements at time t , namely $\{Y_i(t) : i \in s(t)\}$, are assumed to have an exponential correlation pattern and the population elements are assumed exchangeable, i.e.

$$Y_i(t) - Y(t) = \rho(Y_i(t-1) - Y(t-1)) + \eta_i(t) \quad (4.2)$$

where $\{\eta_i(t) : t = 2, 3, \dots, i \in s(t)\}$ is a white noise process, with expectation defined either jointly with respect to the sample design and the superpopulation, or with respect to the sample design only, having zero mean and variance $(1-\rho^2)S^2$, with S^2 the population variance assumed constant over time, and where $\{\eta_i(t)\}$ is assumed uncorrelated with the $\{\alpha(t)\}$.

Although Blight and Scott (1973) suppose that the processes $\{\alpha(t)\}$ and $\{\eta_i(t)\}$ are normally distributed, and use a Bayesian framework, the estimators they obtain are minimum mean square, or best, linear unbiased estimators whatever the distribution of errors. The model of equation

(4.1), while recognised by Blight and Scott as likely to be an oversimplification, often succeeds in capturing a substantial amount of the serial correlation in a series. Equation (4.2) is simply a repetition of the assumption of much of Section 3, that there is an exponential correlation pattern in time for responses from a given sampling unit.

Retaining the notation of the previous section we let

$\underline{y}^{(h)} = (Y(1), Y(2), \dots, Y(h))^T = (\tau(1), \tau(2), \dots, \tau(h))^T = \underline{\tau}_{(h)}$. The b.l.u.e. of $\tau(t)$ given t' periods of data, is denoted $\hat{\tau}^{(t')}(t)$, with $\hat{\tau}^{(t)}(t)$ denoted $\hat{\tau}(t)$. (cf. Section 2 where $\tau(t)$ is non-stochastic and its b.l.u.e. given t' data periods is denoted $\hat{\tau}_0^{(t')}(t)$). $(\hat{\tau}^{(t')}(1), \hat{\tau}^{(t')}(2), \dots, \hat{\tau}^{(t')}(t))^T$ is denoted by $\hat{\underline{\tau}}^{(t')}(1, t)$, $(\hat{\tau}^{(t)}(t-1), \hat{\tau}^{(t)}(t))^T = (\hat{\tau}^{(t)}(t-1), \hat{\tau}(t))^T$ by $\hat{\underline{\tau}}_t$, and $(\tau(t-1), \tau(t))^T$ by $\underline{\tau}_t$.

The results of Blight and Scott (1973) can also be derived in a Kalman filter framework (see, for example, Anderson and Moore (1979)) for $t = 2, 3, \dots$ with parameter transition equation

$$\begin{pmatrix} \tau(t) \\ \tau(t-1) \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \tau(t-1) \\ \tau(t-2) \end{pmatrix} + \begin{pmatrix} \alpha(t) \\ 0 \end{pmatrix} \quad (4.3)$$

i.e. $\underline{\tau}_t = T_\lambda \underline{\tau}_{t-1} + \underline{\alpha}_t$, say

with $\text{var } \underline{\alpha}_t = \begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & 0 \end{pmatrix} = Q$, say,

and parameter regression structure

$$\begin{pmatrix} \hat{\bar{Y}}(t)_u \\ \hat{\bar{Y}}(t)_m - \rho \hat{\bar{Y}}(t-1)_m \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & -\rho \end{pmatrix} \begin{pmatrix} \tau(t) \\ \tau(t-1) \end{pmatrix} + \begin{pmatrix} \bar{\eta}(t)_u \\ \bar{\eta}(t)_m \end{pmatrix} \quad (4.4)$$

i.e.
$$\tilde{y}_t = R \tilde{\tau}_t + \tilde{\eta}_t$$

with
$$\text{var } \tilde{\eta}_t = s^2 \begin{pmatrix} 1/u(t) & 0 \\ 0 & (1-\rho^2)/m(t) \end{pmatrix} = \begin{pmatrix} w(t)_u & 0 \\ 0 & w(t)_m \end{pmatrix} = W_t$$

where $m(t)$ and $u(t)$ are respectively the number of matched and unmatched units at time t , var denotes variance with respect to the superpopulation, and var denotes variance with respect to the sampling scheme.

We let $V_t = \text{var}(\tilde{\tau}_t | \hat{\bar{Y}}(1), \tilde{y}_2, \tilde{y}_3, \dots, \tilde{y}_t)$ which is also equal to the joint sample design, superpopulation variance $\epsilon E(\hat{\tau}_t - \tau_t)^2$ (Duncan and Horn (1972:819)), and let $v(t) = \epsilon E(\hat{\tau}(t) - \tau(t))^2$. Using the usual Kalman filter updating equations (Anderson and Moore (1979), Duncan and Horn (1972)) which are, in this case, so constructed (via the definition of $\tilde{\tau}_t$) as to give b.l.u.e. of an augmented model suitable for fixed interval smoothing (Anderson and Moore (1979:187)), we have, for $t = 2, 3, \dots$

$$\hat{\tau}_t = T_\lambda \hat{\tau}_{t-1} + V_{t|t-1} R^T D_t^{-1} (y_t - R T_\lambda \hat{\tau}_{t-1}) \quad (4.5)$$

and

$$V_t = V_{t|t-1} - V_{t|t-1} R^T D_t^{-1} R V_{t|t-1} \quad (4.6)$$

where

$$V_{t|t-1} = T_\lambda V_{t-1} T_\lambda^T + Q$$

$$D_t = W_t + R V_{t|t-1} R^T$$

and where, following Blight and Scott (1973), we set

$$\hat{\tau}(1) = \hat{\bar{Y}}(1), \quad v(1) = S^2/n(1)$$

with $n(1)$ the sample size at time 1. (The other elements of $\hat{\tau}_1$ and V_1 may be chosen arbitrarily without affecting the estimates of $\hat{\tau}_t$ or V_t for $t = 2, 3, \dots$ since the parameter transition equation is Markov.)

Note that

$$\begin{aligned} V_{t|t-1} &= v(t-1) \begin{pmatrix} \lambda^2 & \lambda \\ \lambda & 1 \end{pmatrix} + \begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \lambda^2 v(t-1) + \sigma_\alpha^2 & \lambda v(t-1) \\ \lambda v(t-1) & v(t-1) \end{pmatrix} \end{aligned}$$

and that an alternative form of the Kalman filter updating equations are given in Duncan and Horn (1972:819) as

$$\hat{\tau}_t = V_t [V_{t|t-1}^{-1} T_\lambda \hat{\tau}_{t-1} + R^T W_t^{-1} y_t] \quad (4.7)$$

$$V_t = [V_{t|t-1}^{-1} + R W_t^{-1} R]^ {-1} \quad (4.8)$$

Here

$$V_{t|t-1}^{-1} + R^T W_t^{-1} R = \begin{bmatrix} \frac{1}{w(t)_u} + \frac{1}{w(t)_m} + \frac{1}{\sigma_\alpha^2} & -\left(\frac{\rho}{w(t)_m} + \frac{\lambda}{\sigma_\alpha^2}\right) \\ -\left(\frac{\rho}{w(t)_m} + \frac{\lambda}{\sigma_\alpha^2}\right) & \frac{\rho^2}{w(t)_m} + \frac{1}{v(t-1)} + \frac{\lambda^2}{\sigma_\alpha^2} \end{bmatrix}$$

and

$$V_t^{-1} |_{t-1}^T \hat{\tau}_{t-1} + R^T W_t^{-1} y_t = \left(\frac{\hat{Y}(t)_u}{w(t)_u} + \frac{\hat{Y}(t)_m - \rho \hat{Y}(t-1)_m}{w(t)_m}, \frac{\hat{\tau}(t-1)}{v(t-1)} - \rho \frac{Y(t)_m - \rho Y(t-1)_m}{w(t)_m} \right)^T$$

and the explicit formulae for $\hat{\tau}_t$ and V_t , given by Blight and Scott (1973), follow directly:

$$V_t = \frac{1}{|\Delta_t|} \Delta_t \quad (4.9)$$

where $|\Delta_t|$ is the determinant of Δ_t , and

$$\Delta_t = \begin{pmatrix} \frac{\rho^2}{w(t)_m} + \frac{1}{v(t-1)} + \frac{\lambda^2}{\sigma_\alpha^2} & \frac{\rho}{w(t)_m} + \frac{\lambda}{\sigma_\alpha^2} \\ \frac{\rho}{w(t)_m} + \frac{\lambda}{\sigma_\alpha^2} & \frac{1}{w(t)_m} + \frac{1}{w(t)_u} + \frac{1}{\sigma_\alpha^2} \end{pmatrix}$$

and

$$\hat{\tau}_t = V_t \left(\frac{\hat{Y}(t)_u}{w(t)_u} + \frac{\hat{Y}(t)_m - \rho \hat{Y}(t-1)_m}{w(t)_m}, \frac{\hat{\tau}(t-1)}{v(t-1)} - \rho \frac{\hat{Y}(t)_m - \rho \hat{Y}(t-1)_m}{w(t)_m} \right)^T \quad (4.10)$$

The b.l.u.e. of change between periods $t-1$ and t , given data to time t is given by $(1, -1) \hat{\tau}_t = \hat{\tau}(t) - \hat{\tau}(t-1)$. Blight and Scott (1973) also derive, in recursive form, the sample overlap which minimises $v(t)$ at each time t . Note that as $\sigma_\alpha^2 \rightarrow \infty$ in equations (4.9) and (4.10) these reduce to the results of Patterson (1950) given in equations (3.26),

(3.32) and (3.29), (3.33).

Scott and Smith (1974) used a more general approach than that allowed by Blight and Scott (1973). A stochastic model for the population means $\{Y(t)\}$ is assumed, although not defined explicitly. Standard time series techniques, under the assumption that $\{Y(t)\}$ have zero mean and are stationary or can be reduced to stationarity by differencing or subtraction of a deterministic component, are then applied to provide estimates of the $\{Y(t)\}$ for equally spaced times, t .

More explicitly, if $\hat{Y}(t)$ is the standard sample survey estimator at time t , with estimates to time h , and equation (1.2) holds with $Y(t)$ stochastic and stationary, then the problem of finding a minimum mean square linear unbiased estimator of $Y(t + \ell)$, namely $\hat{\tau}^{(t)}(t+\ell)$, given the sample survey estimates to time t , is the problem of extracting a signal $Y(t)$ with known zero mean in the presence of noise $e(t)$. Summarised below in equations (4.11) - (4.13) is the relevant theory of signal extraction given the semi-infinite past (i.e. $-\infty < t \leq h$). Further details are contained in Whittle (1963). The papers of Scott and Smith (1974), Jones (1975), and Scott, Smith and Jones (1977) all utilise this theory of signal extraction given the semi-infinite past.

Noting that ℓ may be positive, negative, or zero, we wish to determine $\hat{\tau}^{(t)}(t+\ell)$ as a linear combination of $\hat{Y}(t)$, $\hat{Y}(t-1)$, We assume that the spectral distributions of $\hat{Y}(t)$ and $\hat{\tau}^{(t)}(t+\ell)$ are absolutely continuous and contain no singular or discrete components. We then express $\hat{Y}(t)$, and $\hat{\tau}^{(t)}(t+\ell)$ each as a zero mean moving average process with respect to the same innovation process, $\{\epsilon(t)\}$, with variance σ_{ϵ}^2 , i.e.

$$\hat{Y}(t) = \sum_{j=0}^{\infty} b_j \epsilon(t-j)$$

$$\hat{\tau}^{(t)}(t+l) = \sum_0^{\infty} \phi_j \epsilon(t-j)$$

Then providing $B(z)$ defined below is analytic in $|z| < 1$, and

$$\lim_{r \uparrow 1} \int_{-\pi}^{\pi} |B(re^{-i\lambda})|^2 d\lambda < \infty \quad (\text{see Koopmans (1974:235,253)}), \text{ we may invert}$$

the moving average process for $\hat{Y}(t)$ and express $\hat{Y}(t)$ in autoregressive form.

$$\epsilon(t) = \frac{1}{B(B)} \hat{Y}(t) \quad \text{where } B(B) = \sum_{j=0}^{\infty} b_j B^j$$

$$\text{i.e.} \quad A(B) \hat{Y}(t) = \epsilon(t) \quad \text{where } A(B) = \sum_{j=0}^{\infty} a_j B^j = \frac{1}{B(B)}$$

If we now define

$$c_j = \frac{1}{\sigma_{\epsilon}} C(\tau(t), \epsilon(t-j)) \quad (4.11)$$

where C denotes covariance jointly with respect to the superpopulation and the sample design, we have

$$\begin{aligned} E(\hat{\tau}^{(t)}(t+l) - \tau(t+l))^2 &= \text{var}[\tau(t+l)] - 2\sigma_{\epsilon} \sum_0^{\infty} \phi_j c_{j+l} + \sigma_{\epsilon}^2 \sum_0^{\infty} \phi_j^2 \\ &= \text{var}[\tau(t+l)] - \sum_0^{\infty} c_{j+l}^2 + \sum_0^{\infty} (\sigma_{\epsilon} \phi_j - c_{j+l})^2 \\ &\geq \text{var}[\tau(t+l)] - \sum_l^{\infty} c_j^2 \\ &= \text{var}(\tau(t)) - \sum_{j=l}^{\infty} c_j^2 \end{aligned}$$

which is a minimum for $\sigma_{\epsilon} \phi_j = c_{j+l}$, $j = 0, 1, 2, \dots$

Then

$$\hat{\tau}^{(t)}(t+l) = \sum_0^{\infty} \phi_j B^j \epsilon(t)$$

$$= \left[\frac{g_{\tau\epsilon}(B)}{\sigma_{\epsilon}^2 B^{\ell}} \right]_+ \epsilon(t)$$

$$\text{where } g_{\tau\epsilon}(B) = \sum_{-\infty}^{\infty} C(\tau(t), \epsilon(t-s)) B^s$$

$$= \sum_{-\infty}^{\infty} C_{\tau\epsilon}(s) B^s = \sum_{-\infty}^{\infty} C_j B^j$$

$$= \left[\frac{g_{\tau\tau}(B)}{\sigma_{\epsilon}^2 B^{\ell} B(B^{-1})} \right]_+ \epsilon(t)$$

$$\text{where } g_{\tau\tau}(B) = \sum_{-\infty}^{\infty} C_{\tau\tau}(s) B^s = g_{\tau\hat{Y}}(B)$$

i.e.

$$\hat{\tau}^{(t)}(t+l) = \frac{1}{\sigma_{\epsilon}^2 B(B)} \left[\frac{g_{\tau\tau}(B)}{B^{\ell} B(B^{-1})} \right]_+ \hat{Y}(t) \quad (4.12)$$

and where $\left[\right]_+$ denotes that only nonnegative powers of B are to be considered in the expansion.

The variance of $[\hat{\tau}^{(t)}(t+l) - \tau(t+l)]$ is then given by

$$E(\hat{\tau}^{(t)}(t+l) - \tau(t+l))^2 = \text{var } \tau(t) - \sum_{j=\ell}^{\infty} C_j^2 \quad (4.13)$$

where C_j is defined by equation (4.11). $\hat{\tau}^{(t)}(t+l)$ is also the b.l.u.e. of $\tau(t+l) - \epsilon\tau(t+l) = Y(t+l) - \theta(t+l)$ given data to time t . Note that $C_{\tau\tau}(\ell) = C_{\hat{Y}\hat{Y}}(\ell) - C_{\epsilon\epsilon}(\ell)$ so that, with minor adjustments, $\hat{\tau}^{(t)}(t+l)$ can be written in terms of the autocovariance functions for the sample survey errors and the sample survey estimates. These results can be extended to include processes $\{\tau(t)\}$ which are not stationary but have stationary d^{th}

differences (see Whittle (1963), Chapter 8), although such procedures often produce difficulties in practice as they induce additional structure into the differenced survey estimates $\{\hat{Y}(t)\}$. If the d^{th} differenced survey estimates have an ARIMA structure with roots on the unit circle, as will be the case if $\tau(t)$ (i.e. $Y(t)$) requires higher order differencing than $\hat{Y}(t)$ to induce stationarity, then the d^{th} differenced survey estimates may not be representable in autoregressive form, and equations (4.12) and (4.13) do not apply. However, in comparison with formulations with $\underline{Y}$ non-stochastic, that of Scott and Smith (1974) by allowing $Y(t)$ to have stochastic structure has the advantage that it allows improvement of traditional sample survey estimates even when the surveys are non-overlapping. In such cases, since the surveys are non-overlapping, we have, from equation (1.2), that

$$\hat{Y}(t) = Y(t) + e(t),$$

where the $e(t)$ are uncorrelated with, for example, $\text{var } e(t) = S^2$. Given data to time h we have one estimate of $Y(h)$, namely the traditional sample survey estimator at time h , $\hat{Y}(h)$.

From time series analysis we have a second uncorrelated estimate of $Y(t)$, namely $\hat{\tau}^{(h-1)}(h)$, the time series predictor of $Y(h)$ given the data up to and including time period $h-1$. Let us suppose

$$\hat{\tau}^{(h-1)}(h) = \tau(h) + f(h).$$

Defining $\sigma_{h|h-1}^2 = \epsilon E(\hat{Y}(h) - \hat{\tau}^{(h-1)}(h))^2$ we have

$$\epsilon E(f(h))^2 = \sigma_{h|h-1}^2 - S^2$$

A b.l.u.e. of $Y(h)$, $\hat{\tau}(h)$, is formed from a weighted combination of $\hat{\tau}^{(h-1)}(h)$ and $\hat{Y}(h)$, i.e.

$$\hat{\tau}(h) = w\hat{Y}(h) + (1-w)\hat{\tau}^{(h-1)}(h)$$

where w remains to be determined. (For a justification that $\hat{\tau}(h)$ can be written in this form see Duncan and Horn (1972).) Then

$$E(\hat{\tau}(h) - \tau(h))^2 = (1-w)^2 \sigma_{h|h-1}^2 + (2w-1)S^2$$

and minimising with respect to w yields

$$w_0 = 1 - \frac{S^2}{\sigma_{h|h-1}^2}$$

and

$$E(\hat{\tau}(h) - \tau(h))^2 = w_0 S^2$$

with the proportional decrease in variance, by using $\hat{\tau}(h)$ rather than $\hat{Y}(h)$ to estimate $Y(h)$, being $S^2/\sigma_{h|h-1}^2$.

Scott and Smith (1974) also consider the special case of a first order autoregressive process for $Y(t)$ when the surveys do not overlap, and discuss the applicability of ARIMA (autoregressive integrated moving average) representations for non-overlapping surveys. Jones (1975) and Scott, Smith and Jones (1977) fit such a model to various data series from the U.K. Medical Data Index. For further details on ARIMA models see Box and Jenkins (1970).

For overlapping surveys, the theory remains that of signal extraction as outlined in equations (4.11) - (4.13). $\hat{\tau}(h)$ can thus be determined, on noting that we are predicting zero steps ahead. The theory itself takes care of the fact that the survey errors $\{e(t)\}$ are correlated. These results are in theory extendable to ARIMA (p,d,q) processes, although in practice the time series error process itself must be stationary on p^{th} differencing. Scott and Smith consider the particular case of the stationary ARIMA $(1,0,1)$ process for survey errors. The corresponding correlation structure is then of the form $c\rho^{|t-t'|}$ where c is a constant, with t and t' two time periods. The population means $\{Y(t)\}$ are then modelled with an AR(1) process, and explicit formulae for $\hat{\tau}(t)$ and its mean square error derived. Scott and Smith (1974) note that they consider an ARIMA $(1,0,1)$ (i.e. an ARMA $(1,1)$) a common structure for the sample survey error process $\{e(t)\}$.

In 1975 Jones produced an M.Sc. dissertation on the topic "Time Series Analysis of Repeated Surveys". It contained a summary of work to that date, some additional details on prediction and updating formulae provided the sample survey error process could be considered to be ARIMA, a discussion of methods of determining the time series structure of the sample survey error process, and several examples. Noting that only overlapping surveys have non-white noise sample survey error processes, the prediction and updating formulae and the correlation structure of the sample survey error are discussed here. The examples can be found in Jones (1975), and also in Scott, Smith and Jones (1977).

Jones (1975) notes that, if the sample survey error $\{e(t)\}$ follows a stationary ARMA (p,q) process,

$$\delta(B)e(t) = \beta(B)b(t) \quad (4.14)$$

i.e.

$$(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_p B^p) e(t) = (1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_q B^q) b(t)$$

where B is the backward shift operator, $\{b(t)\}$ is a zero mean i.i.d. process with variance σ_b^2 , $\beta(z)$ and $\delta(z)$ are analytic for $|z| \leq 1$, then equation (4.14) is invertible (see Priestley (1981:686)). In these circumstances, for $\ell > q$, there is a recursive algorithm for reducing computational burden at each time t , namely

$$\hat{\tau}^{(t)}(t+\ell) = \hat{Y}^{(t)}(t+\ell) - \sum_{i=1}^p \delta_i (\hat{\tau}^{(t)}(t+\ell-i) - \hat{Y}^{(t)}(t+\ell-i)) \quad (4.15)$$

where $\hat{Y}^{(t)}(t')$ is the time series estimate of $\hat{Y}(t')$ based on data to time t only. While continuing to assume sample survey errors have ARMA structure, Jones then develops formulae for the best linear unbiased estimate of $\tau(t)$ given data to $t-1$, t , $t+1$, Given short runs of data and the fact that Jones's development requires the semi-infinite past, these results are largely of theoretical interest. Some necessary conditions for such results are not however given in Jones (1975) and are given here for completeness.

Supposing again that $\{e(t)\}$ is an ARMA (p,q) process as defined by equation (4.14), we again assume that the traditional sample survey estimates $\{\hat{Y}(t)\}$ can be written in autoregressive form

$$A(B)\hat{Y}(t) = \epsilon(t) \quad \text{with} \quad A(B) = \sum_{j=0}^{\infty} a_j B^j$$

Then under certain conditions Jones forms $\hat{\tau}^{(t)}(t-1)$ the b.l.u.e. of $\tau(t-1)$ given data to time t , by the formula

$$\begin{aligned} \hat{\tau}^{(t)}(t-1) = \hat{\tau}^{(t-1)} + \frac{1}{\sigma_{\epsilon}^2} \sum_{m=1}^p \sum_{k=0}^q a_k \left[c_{e_m}^{(k-1)} - \frac{c_{e_m}^{(q)}}{\rho_m^{q-|k-1|}} \right] \epsilon(t) \\ + \frac{c_{e_m}^{(q)}}{\rho_m^{q+1} \sigma_{\epsilon}^2} \left[A(\rho_m) - (1 - \rho_m^2) \right] \epsilon(t) \end{aligned} \quad (4.16)$$

where equation (4.16) involves the sum of p independent ARMA $(1, q)$ processes, each further decomposable into the sum of two independent processes, an AR(1) and an MA($q-1$). Jones (1975) limits his discussion of the recursion of equation (4.15) to ARMA (p, q) processes of this restricted type. Equation (4.16) is related to the original $\{e(t)\}$ process by $e(t) = \sum_{m=1}^p e_m(t)$ where

$\{e_m(t)\} : m = 1, 2, \dots, p$ are the correlated random variables, for each of the p independent ARMA $(1, q)$ processes,

$\{\rho_m, m = 1, 2, \dots, p\}$ are the roots of $1 - \delta_1 z^{-1} - \dots - \delta_p z^{-p} = 0$,

$c_{e_m}^{(\ell)}$ is the covariance of e_m at lag ℓ .

Under the same conditions, Jones (1975:45) gives for $0 \leq \ell \leq q$,

$$\begin{aligned} \hat{\tau}^{(t)}(t+\ell) = \hat{Y}^{(t)}(t+\ell) - \frac{1}{\sigma_{\epsilon}^2} \sum_{m=1}^p \sum_{j=0}^{q-\ell} \left[\sum_{k=j}^{q-\ell} a_{k-j} \{c_{e_m}^{(\ell+k)} - \rho_m^{k-(q-\ell)} c_{e_m}^{(q)}\} \right] B^j \epsilon(t) \\ - \frac{1}{\sigma_{\epsilon}^2} \sum_{m=1}^p \left[c_{e_m}^{(q)} A(\rho_m) / \{\rho_m^{q-\ell} (1 - \rho_m^2)\} \right] \epsilon(t) \end{aligned} \quad (4.17)$$

and under the rather more general conditions pertaining to signal extraction as described for equations (4.11) - (4.13) for $\ell > q$,

$$\hat{\tau}^{(t)}(t+\ell) = \hat{Y}^{(t)}(t+\ell) - \frac{1}{\sigma_e^2} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_k c_e(\ell+j+k) \epsilon(t-j) \quad (4.18)$$

For equations (4.16) and (4.17) at least the following conditions must hold:

1. The elements of $\{\rho_m : m = 1, 2, \dots, p\}$ are real and distinct. This allows the ARMA (p, q) to be expressed as the sum of p independent ARMA $(1, q)$ processes.

$$2. \sum_{k=-q+1}^{q-1} \left[c_{e_m}(k) - \frac{c_{e_m}(q)}{\rho_m^q} \rho_m^{|k|} \right] e^{i\lambda k} \geq 0, \text{ for } m = 1, 2, \dots, p \text{ and}$$

$-\pi < \lambda \leq \pi$, i.e. the spectral densities of each of the p ARMA $(0, q-1)$ processes must be non-negative, as a condition for their existence.

3. $c_{e_m}(q)/\rho_m^q > 0$ for $m = 1, 2, \dots, p$, i.e. the spectral density of each of the p ARMA $(1, 0)$ processes exist by being strictly positive for at least some part of the interval $(-\pi, \pi]$. These spectral densities are of the form

$$f_m(\lambda) = \frac{\sigma_{b_m}^2}{2\pi} \frac{c_{e_m}(q)}{\rho_m^q} \frac{1 - \rho_m^2}{|1 - \rho_m e^{i\lambda}|^2}$$

for $m = 1, 2, \dots, p$, $-\pi < \lambda \leq \pi$, where $\sigma_{b_m}^2$ is the variance of b_m and

$$e_m(t) - \rho_m e_m(t-1) = b_m(t) - \beta_{1m} b_m(t-1) - \dots - \beta_{qm} b_m(t-q) \text{ say.}$$

Further details on conditions for invertibility and decomposition of ARMA processes can be found in Terasvirta (1977). Note that for an ARMA $(1, 1)$ process to be written as the sum of an ARMA $(1, 0)$ process and an independent ARMA $(0, 0)$ process (and thus for equation (4.16) to hold for this case), a necessary condition, from condition 2 above, is that

$\text{sgn}(\beta_1) = \text{sgn}(\rho)$, $|\rho| < 1$, $|\beta_1| < 1$, $|\beta_1| > |\rho|$, where sgn denotes sign and where the ARMA (1,1) process is

$$e(t) - \rho e(t-1) = b(t) - \beta_1 b(t-1)$$

Both Jones (1975) and Scott, Smith and Jones (1977) point out that in order to apply these results on signal extraction, and decompose $\{\hat{Y}(t)\}$ into signal $\{Y(t)\}$ and noise $\{e(t)\}$, it is necessary to estimate the correlation structure of the noise process $\{e(t)\}$, and the parameters $A(z)$ of the autoregressive process for $\hat{Y}(t)$. They estimate the autoregressive parameters by the standard techniques for fitting ARIMA models to a time series, a full description of which is given by Box and Jenkins (1970). Estimation of the correlation structure they consider to be one of two types, either a primary or a secondary analysis. A primary analysis makes direct use of the respondent information from each time period, $\{y_i(t) : i \in s(t), t \leq h\}$, to calculate directly the autocovariance function

$$g_{ee}(z) = \sum_{-\infty}^{\infty} c_{ee}(\ell) z^{\ell}.$$

A secondary analysis is usually undertaken when the individual respondent data is unavailable and the autocovariance function of $\{e(t)\}$ must be calculated using only the sample survey estimates $\{\hat{Y}(t) : t \leq h\}$, the noise or sample survey variance S^2 , and a knowledge of the survey design.

For a primary analysis both Jones (1975) and Scott, Smith and Jones (1977) show that even for single stage overlapping surveys, the sample survey error autocovariance function $g_{ee}(z)$ is not that of a simple

stochastic process, and that for two stage sampling schemes $g_{ee}(z)$ corresponds to a very complex stochastic process indeed. They note, however, that under certain circumstances considerable simplifications are possible. If there is no overlap for lags greater than q then $\{e(t)\}$ can be modelled by a moving average of order q regardless of how many stages there are in the sampling scheme. Master sampling schemes commonly used in market research also allow simplifications. For such schemes primary sampling units (psu's) remain unchanged over time and within psu's independent samples are taken at each time. Even if there were a rotation scheme for psu's with constant overlap, π , the correlation at lag ℓ would be of the form $\pi\rho(\ell)$, and if $\rho(\ell) = \rho^\ell$, the sample survey errors are then an ARMA (1,1) process.

Jones (1975) also considers overlapping surveys where individual responses are not known, so that while the form of the error process may be known via the sample design and rotation pattern, its parameters are not. He considers modelling the population means and the sample survey errors by independent ARMA processes, the parameters of which he recovers from the estimated parameters of the consequent ARMA process for the $\{\hat{Y}(t)\}$. The dimensions (i.e. the number of moving average and autoregressive terms) of the ARMA process for the population means $\{Y(t)\} = \{\tau(t)\}$ must be specified in advance, and specification difficulties often occur in practice. Terasvirta (1977:Theorem 2) gives conditions under which the sum of a number of seemingly unrelated stationary ARMA processes is itself an ARMA process. In signal plus noise models his condition reduces to the requirement that the component moving average polynomials in the backward shift operator, for signal (population means) and for noise (sample survey errors) have no common factor of modulus one. Misspecification of the ARMA models, while clearly possible, may not

be a major problem. Monte Carlo studies undertaken by Jones (1975) generally suggest that error structure misspecification, provided it is consistent with the data, leaves estimates of level and mean square error largely unchanged.

The efficiency of the various best linear unbiased estimators so far discussed in Sections 3 and 4 was made by Jones (1979) who compared the efficiencies of the estimators of Patterson (1950), Blight and Scott (1973), and Scott and Smith (1974), by computing the mean square error of the estimators of the current mean, and of the change in mean on the last two occasions under standardising conditions. Given t periods of data, we denote Patterson's (1950) estimator by $\hat{\tau}(t)^{(P)}$, Blight and Scott's (1973) estimator by $\hat{\tau}(t)^{(B,S)}$ and that of Scott and Smith (1974) by $\hat{\tau}(t)^{(S,S)}$.

Constant sample size, n , was assumed on each of h equally spaced occasions, with proportions π of sampled units matched at each time period. Units were assumed to be replaced using Eckler's (1955) one-level rotation pattern, and $\varepsilon(\tau(t)) = \theta(t) = \mu$ for all times t . Survey responses were assumed to follow an AR(1) process, and for the Blight and Scott (1973) and the Scott and Smith (1974) estimators, the population means were also assumed to follow an independent AR(1).

Values of standardised mean square error $SMSE[\hat{\tau}(h)^{(P)}]$, $SMSE[\hat{\tau}(h)^{(B,S)}]$ and $SMSE[\hat{\tau}(h)^{(S,S)}]$ were then calculated for the range of parameter values

$$\rho = 0.25, 0.5, 0.75, 0.9, 0.95$$

$$\lambda = 0.0, 0.5, 0.8$$

$$0 \leq \pi \leq 1$$

$$VR = 0.2, 1.0, 5.0$$

where ρ and λ are the AR parameters for the survey response process and the

population means respectively, π is the sample overlap, and VR is the signal to noise ratio defined by $C_{\tau\tau}(0)/(S^2/n)$.

Jones's (1979) results can be summarised as follows. Use of the time series relationship for population means leads to considerable gains in efficiency especially for small VR. As $VR \rightarrow \infty$, $\hat{\tau}(h)^{(B,S)} \rightarrow \hat{\tau}(h)^{(P)}$, and $\hat{\tau}(h)^{(S,S)} \rightarrow \hat{Y}(h)$. For low ρ values the proportion of overlap π has little effect of MSE of the estimators and $SMSE[\hat{\tau}(h)^{(B,S)}] \approx SMSE[\hat{\tau}(h)^{(S,S)}]$. For higher values of ρ , and especially if VR is large $SMSE[\hat{\tau}(h)^{(S,S)}] - SMSE[\hat{\tau}(h)^{(B,S)}]$ becomes more pronounced. Variation of λ , the autoregressive parameter of the signal process, most affects the SMSE estimates when the proportion of overlap, π , is low, since it is then the major determinant of the correlation structure of the $\{\hat{Y}(t)\}$ series. Similar values for $SMSE[\hat{\tau}(h)^{(S,S)}]$ result when more complex ARMA processes replace the AR(1) process, provided correlations are positive. The VR ratio is the most important factor in determining $SMSE[\hat{\tau}(h)^{(B,S)}]$ and $SMSE[\hat{\tau}(h)^{(S,S)}]$. For small VR, $\hat{\tau}(h)^{(S,S)}$ is nearly optimal. With increasing VR the separation of overlapping and non-overlapping is increasingly more effective, particularly for high values of ρ , and $\hat{\tau}(h)^{(P)}$ and $\hat{\tau}(h)^{(B,S)}$ thus perform better than $\hat{\tau}(h)^{(S,S)}$ over a large part of the range $0 \leq \pi \leq 1$.

Jones (1979) also considers estimates of change between two time periods, $\hat{\tau}(h) - \hat{\tau}^{(h)}(h-1)$, together with their mean square errors. For Scott and Smith's (1974) estimator it is assumed that to estimate change $\tau(h)$ and $\hat{\tau}(h)$ are replaced by their first differences before signal extraction is undertaken, which may not be an entirely satisfactory procedure if $\{\tau(t)\}$ is itself a stationary process.

Denoting $\hat{\tau}(h)^{(x)} - \hat{\tau}^{(h)}(h-1)^{(x)}$ by $\nabla \hat{\tau}(h)^{(x)}$ where $(x) = (P), (B,S)$, or (S,S) , Jones's (1979) results can be summarised as follows: For low ρ there

is little gain from considering overlapping and non-overlapping units separately so that $SMSE(\hat{V}\hat{\tau}(h)^{(P)})$ is close to one, and $SMSE(\hat{V}\hat{\tau}(h)^{(B,S)})$ is only slightly less than $SMSE(\hat{V}\hat{\tau}(h)^{(S,S)})$. As ρ increases the efficiency of Patterson's estimate of difference, $\hat{V}\hat{\tau}(h)^{(P)}$, improves considerably, approaching the efficiency of $\hat{V}\hat{\tau}(h)^{(B,S)}$ and in many cases improving on that of $\hat{V}\hat{\tau}(h)^{(S,S)}$, provided there is partial overlap. When the signal to noise ratio, VR , is small, time series estimators give considerable gain over $\hat{Y}(h) - \hat{Y}(h-1)$ or Patterson's estimator. As $VR \rightarrow \infty$ $\hat{V}\hat{\tau}(h)^{(B,S)} \rightarrow \hat{V}\hat{\tau}(h)^{(P)}$ and $\hat{V}\hat{\tau}(h)^{(S,S)} \rightarrow \hat{Y}(h) - \hat{Y}(h-1)$. The autoregressive parameter λ has considerable effect on the mean square error of the full time series estimators (i.e. that of Blight and Scott, and of Scott and Smith) via $C_{\tau\tau}(\ell)$, $\ell = 0, 1, 2, \dots$, particularly when $C_{ee}(\ell)$, $\ell = 0, 1, \dots$, is relatively weak.

Jones (1979) further comments that he believes the particular advantage of Scott and Smith's (1974) estimator is that it can be used quite generally, provided the noise process $\{e(t)\}$ is stationary. In practice this perceived advantage is not clear since parameters of the ARMA processes for signal and noise still require estimation. In comparison, Jones (1979) states that Patterson's (1950) estimator requires sample survey responses to be AR(1) and for Blight and Scott's (1973) estimator both sample responses and population means must be AR(1). These two estimators can be modified to allow for any sampling procedure but lose some simplicity in the process (see Chapter 5 for example). The results indicate that substantial gains in efficiency can be made, using full time series estimators comparative to the original sample survey estimators or Patterson's estimator, when $var(\hat{Y}(h))$ is high relative to $var(\tau(h))$, or when the correlation between sample units, ρ , is low.

However the major difficulty with estimates of population means using general time series techniques as in Scott, Smith and Jones (1977), remained the reliance on the theory of signal extractions given the semi-infinite past. Data series for repeated sample surveys tend to be short, and the assumption of a semi-infinite past is not in these circumstances a good one. Blight and Scott (1973) had avoided this difficulty, but only at the cost of imposing a particular simple time series structure, an AR(1), on both survey responses and population means. Jones (1980) provided an alternative to signal extraction given the semi-infinite past by formulating the time series estimation as a problem in a stochastic least squares. The Kalman filter used previously to derive the estimator of Blight and Scott (1973) is a particular case of stochastic least squares. More complicated ARMA signal plus ARMA noise models can also be fitted using Kalman filters. Jones (1980), however, outlined the more general and computationally intensive stochastic least squares formulation where the problem of signal extraction is framed in terms of all the available data between times 1 and h. If data is available as elementary estimates the stochastic model can be written in the form

$$\hat{\underline{y}}_G = X\underline{\tau} + \underline{e} \quad (4.19)$$

which is equation (3.3) but now with $\underline{\tau}$ stochastic. Assume $\epsilon(\underline{\tau}) = \underline{\mu}$ is known so that we may set $\underline{\mu} = 0$ without loss of generality, and let $\text{var}(\underline{\tau}) = T$. Extending the notation of equation (3.3), $\epsilon E(\underline{e}) = 0$ with covariance matrix Σ . Note that $\hat{\underline{y}}_G$ is an $(r \times 1)$ vector of elementary estimates from the h time periods, X is an $(r \times h)$ matrix of zeros and ones, $r \geq h$, $\underline{\tau} = (\tau(1), \dots, \tau(h))^T$, and r is the total number of elementary estimates over the h time periods. Then the b.l.u.e. of $\underline{\tau}$ is $\hat{\underline{\tau}}$ where

$$\hat{\tau} = (X^T \Sigma^{-1} X + T^{-1})^{-1} X^T \Sigma^{-1} Y_G \quad (4.20)$$

with
$$\hat{\tau} = (\hat{\tau}^{(h)}(1), \hat{\tau}^{(h)}(2), \dots, \tau(h))^T$$

and
$$\epsilon E(\hat{\tau} - \tau)(\hat{\tau} - \tau)^T = (X^T \Sigma^{-1} X + T^{-1})^{-1}$$

(cf. equation (3.4)). The b.l.u.e. of any linear combination $\ell^T \hat{\tau}$ is $\ell^T \hat{\tau}$.

The same result can also be derived by applying the theory of stochastic least squares to the least squares estimates of Gurney and Daly (1965).

The estimator of the form of equation (4.20) derived by Blight and Scott (1973) used model assumptions that equated the non-stochastic least squares estimator with Patterson's (1950) estimator, and as previously noted, assumed $\{Y(t)\}$ to follow a first order autoregressive model.

The reformulation of the Scott and Smith (1974) estimator may be derived as follows. We assume

$$\hat{y}_{(h)} = \tau + e \quad (4.21)$$

where $\hat{y}_{(h)}$ is the $(h \times 1)$ vector of sample survey estimates $(\hat{y}(1), \hat{y}(2), \dots, \hat{y}(h))^T$ (i.e. a special case of $\hat{y}_G$ with one elementary estimate per time period), $\tau = (\tau(1), \tau(2), \dots, \tau(h))^T$, e is now an $(h \times 1)$ vector of zero mean sample survey errors again with variance Σ estimable from the sample survey data, (cf. equation (4.19) with $X = I$, the identity matrix). Stochastic least squares then yields

$$\hat{\tau} = (\Sigma^{-1} + T^{-1})^{-1} \Sigma^{-1} \hat{y}_{(h)} \quad (4.22)$$

with

$$E(\hat{\tau} - \tau)(\hat{\tau} - \tau)^T = (\Sigma^{-1} + T^{-1})^{-1}$$

Jones (1980) points out the assumptions for least squares require only that $\hat{\tau}_{(h)}$ be any unbiased sampling estimate of τ provided that the associated errors e are independent of τ . In particular taking $\hat{\tau}_{(h)} = \hat{\tau}_0$ where $\hat{\tau}_0$ is derived by non-stochastic least squares, equates $\hat{\tau}_{(h)}$ to the Gurney and Daly (1965) estimator, which is (as has been noted previously) the generalisation of Patterson's (1950) estimator. Such a choice Jones notes, would, however, require either a short series of surveys or a recursive form of the estimator to be practicable in large scale surveys. (Appropriate recursive formulae are derived in Chapter 5.) A general choice of $\hat{\tau}_{(h)}$ however, allows that it may be possible, in a particular case, to pick an estimator $\hat{\tau}_{(h)}$ which is simple to calculate, allows for easy estimation of its error covariance structure and which, via equation (4.22) provides an estimator achieving most of the gains of equation (4.22) applied to Gurney and Daly's (1965) estimator. Note, however, as elaborated in Chapter 3, that such gains can also be achieved directly by applying equation (4.20) to the elementary estimates derived from matched and unmatched units assuming these were to be used to produce the Gurney and Daly estimator.

Equation (4.22) is not, however, in the most practical form since it depends on the covariance structure of τ , via T . Dagsvik (1978) and Helgeland (1982) were able to estimate T for Norwegian labour force data, but this was fortuitous in that the survey was designed to replace a repeated census, which provided the necessary estimates. However, from Rao (1973:33) with $\Lambda = E(\hat{\tau}_{(h)} - \tau)(\hat{\tau}_{(h)} - \tau)^T$, the variance of $\hat{\tau}_{(h)}$, it follows that

$$\begin{aligned}
 (\Sigma^{-1} + T^{-1})^{-1} &= \Sigma - \Sigma(\Sigma + T)^{-1}\Sigma \\
 &= \Sigma - \Sigma\Lambda^{-1}\Sigma
 \end{aligned}$$

so that equation (4.22) may be written as

$$\hat{\tau} = \hat{\tilde{y}}_{(h)} - \Sigma \Lambda^{-1} \hat{\tilde{y}}_{(h)} = \hat{\tilde{y}}_{(h)} - \hat{\tilde{e}} \quad (4.23)$$

where $\hat{\tilde{e}}$ is the b.l.u.e. of $\tilde{e}$ based on $\hat{\tilde{y}}_{(h)}$, and

$$E(\hat{\tau} - \tau)(\hat{\tau} - \tau)^T = \Sigma - \Sigma\Lambda^{-1}\Sigma$$

This formulation shows clearly the effect of allowing τ to be a random variable. The unbiased estimate $\hat{\tilde{y}}_{(h)}$ is adjusted by a factor, namely the b.l.u.e. of the sampling errors given $\hat{\tilde{y}}_{(h)}$.

Jones (1980) suggests that if $\{\tau(t)\} = \{Y(t)\}$ is stationary, and hence $\{\hat{Y}(t)\}$ is stationary, about a zero mean, the autocovariance for $\hat{\tilde{y}}_{(h)}$ at various lags can be estimated by

$$\hat{C}_{\hat{Y}}(\ell) = \frac{1}{h} \sum_{t=1}^{h-\ell} \hat{Y}(t) \hat{Y}(t+\ell) \quad (4.24)$$

for $\ell = 0, 1, \dots, h-1$, leading to an estimate for Λ^{-1} . Scott and Smith's (1974) signal extraction approach goes further in fitting an ARMA model to the observations which corresponds to the $\{\hat{C}_{\hat{Y}}(\ell), \ell = 0, 1, 2, \dots\}$ and using this model to determine the associated semi-infinite covariance structure. Jones suggests that the first of these alternatives is the better except perhaps for long runs of data $\{\hat{Y}(t) : t = 1, 2, \dots, h\}$. An even better alternative may be to smooth covariance estimation by fitting an ARMA model to the observations and using this model to determine $\hat{C}_{\hat{Y}}(\ell), \ell = 0, 1, 2, \dots, h-1$,

and hence the elements of the inverse matrix Λ^{-1} . Jones (1983) develops an alternative method which uses a stepwise regression procedure and utilises only those elements of $\{\hat{\tilde{C}}_Y(\ell) : \ell = 0, 1, 2, \dots, h-1\}$ which are known most accurately, namely those with small lags, ℓ .

The results for stochastic least squares equations (4.22) and (4.23) are derived assuming $\varepsilon(\tau) = 0$. In practice this is an unlikely assumption. If $\varepsilon(\tau)$ is unknown but τ stochastic then the b.l.u.e. of τ reduces to the b.l.u.e. obtained via non-stochastic least squares (Rao (1973)). The mean must therefore be known in order for the time series approach to be useful. Two approaches are discussed by Jones (1980). The first assumes τ can be differenced to stationarity, and the mean of the differenced series estimated and subtracted out. The second assumes a polynomial trend with constant coefficients, which is again estimated and subtracted out.

We first consider differencing to stationarity. We assume $\{\tau(t)\}$ and hence $\{\hat{Y}(t)\}$ can be differenced to stationarity. Dagsvik (1978) has shown that in this case $\tilde{e}$ depends on $\hat{\tilde{Y}}_{(h)}$ only through the differenced series $\Delta \hat{\tilde{Y}}_{(h)}$ where Δ is the appropriate differencing operator expressed in matrix form. Assuming $\Delta \tau$ is then also stationary and applying stochastic least squares to

$$\Delta \hat{\tilde{Y}}_{(h)} = \Delta \tau + \Delta \tilde{e} \quad (4.25)$$

yields

$$\hat{\tilde{e}} = \Sigma \Delta^T \Lambda^{-1} \Delta \hat{\tilde{Y}}_{(h)}$$

via

$$\begin{aligned} \Delta \hat{\tilde{e}} &= \Sigma_{\Delta \tilde{e}} \Lambda^{-1} \Delta \hat{\tilde{Y}}_{(h)} \\ &= \Delta \Sigma \Delta^T \Lambda^{-1} \Delta \hat{\tilde{Y}}_{(h)} \end{aligned}$$

$$\begin{aligned}
 \text{since } \Sigma_{\Delta \underline{e}} &= \epsilon E (\Delta \underline{e} - \overline{\Delta \underline{e}}) (\Delta \underline{e} - \overline{\Delta \underline{e}})^T \\
 &= \Delta \epsilon E (\underline{e} - \overline{\underline{e}}) (\underline{e} - \overline{\underline{e}}) \Delta^T \\
 &= \Delta \Sigma \Delta^T
 \end{aligned}$$

and thus

$$\hat{\underline{\tau}} = \hat{\underline{y}}_{(h)} - \Sigma \Delta^T \Lambda^{-1} \Delta \hat{\underline{y}}_{(h)} \quad (4.26)$$

with

$$\epsilon E (\hat{\underline{\tau}} - \underline{\tau}) (\hat{\underline{\tau}} - \underline{\tau})^T = \Sigma - \Sigma \Delta^T \Lambda^{-1} \Delta \Sigma \quad (4.27)$$

The dispersion matrix Λ can be estimated and inverted as in the stationary case, either directly or via signal extraction methods.

If, however, the mean is given by a polynomial trend with constant coefficients so that

$$\underline{\phi} = \underline{\tau} - H \underline{\alpha} \quad (4.28)$$

is stationary where $\underline{\alpha}^T = (\alpha_0, \alpha_1, \dots, \alpha_{p-1})^T$ is a vector of unknown polynomial coefficients either with respect to monomials or orthogonal polynomials in time, and H is an $(h \times p)$ matrix of specified constants determined as in Chapter 6, then we may write

$$\begin{aligned}
 \hat{\underline{y}}_{(h)} &= H \underline{\alpha} + \underline{\phi} + \underline{e} \\
 &= H \underline{\alpha} + [\underline{I}_h \vdots \underline{I}_h] \begin{bmatrix} \underline{\phi} \\ \dots \\ \underline{e} \end{bmatrix}
 \end{aligned} \quad (4.29)$$

and applying the results of Harville (1976) or Pfeiffermann (1984)

$$\hat{\alpha} = (H^T \Lambda^{-1} H)^{-1} H^T \Lambda^{-1} \hat{y}_{(h)}$$

$$\phi = \text{var}(\phi) \Lambda^{-1} (\hat{y}_{(h)} - H\hat{\alpha})$$

$$= (\hat{y}_{(h)} - H\hat{\alpha}) - \Sigma \Lambda^{-1} (\hat{y}_{(h)} - H\hat{\alpha})$$

and hence

$$\hat{z} = \hat{y}_{(h)} - \Sigma \Lambda^{-1} (\hat{y}_{(h)} - H\hat{\alpha}) \quad (4.30)$$

Jones (1980) notes that if $H\hat{\alpha}$ can be differenced to zero so that $\Delta H = 0$, or transposing $H^T \Delta^T = 0$ then via Rao (1973:77)

$$\Lambda^{-1} = \Delta^T (\Delta \Lambda \Delta^T)^{-1} \Delta + \Lambda^{-1} H (H^T \Lambda H)^{-1} H^T \Lambda^{-1} \quad (4.31)$$

and substitution of equation (4.31) into equation (4.30) yields equation (4.26). Jones (1980) states that differencing generally allows more flexibility than constant coefficient polynomials, and since if $H^T \Delta^T = 0$ the trend $H\hat{\alpha}$ can be estimated independently of $\hat{z}$, the differenced form of $\hat{z}$ is preferable in these circumstances.

A note is necessary however, as before, and places a rider on any advantage of differencing to stationarity. If at any stage d' ($d' < d$) of the d stages of differencing, the error series is an ARMA process, then after d stages of differencing, $(d-d')$ roots of the associated MA polynomial will be roots on the unit circle. Such a process will present considerable parameter estimation problems in practice, especially with short data series. In these circumstances a polynomial trend should be removed at or

before the d^{th} stage, in a simple step via equation (4.28).

The need to difference or remove a polynomial trend involves two of "an unlimited number of ways in which a process can be nonstationary" (Box and Jenkins (1970:87)). The use of nonlinear rather than linear transformations on sample survey estimates to find time series estimates, is not generally feasible however Jones (1980) notes, because it invalidates the basic signal plus noise model. It is the explicit estimability of the untransformed noise (i.e. sample survey error) process which allows the signal to be identified, and unless a primary analysis is possible on transformed individual data nonlinear transformations will produce an unidentifiable signal process. Nonlinear transformation and averaging are seldom commutative.

Jones (1983) provides a brief summary of repeated sample survey literature for both stochastic and nonstochastic population means. He emphasises that for the stochastic case it is important that the means $\{\varepsilon(Y(t))\}$, i.e. $\{\theta(t)\}$, be known and subtracted out before stochastic least squares is applied, else least squares estimates reduce to those for the nonstochastic case. Trend removal is thus a prerequisite to application of stochastic least squares. Jones also uses the stepwise regression procedure mentioned previously to avoid instability of unsmoothed lag covariance estimates for the sample survey error process, and extends this idea to differenced series. He thus also avoids fitting an ARMA process to the possibly differenced sample survey errors, a strategy which may have the cost of estimating a greater number of parameters. This brings up an interesting and important point. The work of Jessen (1942), Patterson (1950), Gurney and Daly (1965) and Blight and Scott (1973) all assume model structure and population correlations known.

Scott and Smith (1974), Scott, Smith and Jones (1977) and Jones (1980), while allowing more general time series models, again treat parameter estimates as parameters for mean square error determination. In none of these cases therefore are the estimates derived strictly best linear unbiased. One corollary is that given the relative simplicity of their updating equations (cf. Chapter 5), Kalman filters remain a neglected possibility for time series improvement of sample survey estimators, especially since ARMA processes can be represented in Kalman filter form. Cf. "there is the difficulty that the (Kalman filter) state transition equation will, in general, be unknown" (Jones (1980:225)). Some preliminary work on the fitting of ARMA signal plus ARMA noise models to series with more than one datum per time period, using Kalman filters, can be found in Haslett (1985) (c).

There are two further lines of enquiry that have been associated with repeated sample surveys in the literature.

The first seeks to extend into the time domain, various optimal sample survey strategies and estimators for one time period only. Optimality is often assessed, even for the one time period case, as a mean square error in a joint sample design and superpopulation sense, subject to design unbiasedness, although convergence and consistency properties of estimators are also important. For example, for a single time period, Cassel, Sarndal, and Wretman (1977:Chapter 4) consider design unbiased estimation under a superpopulation model, and Robinson (1982(a)) discusses convergence properties of the Horvitz-Thompson estimator. Nearly all the estimators proposed by Cassel, Sarndal and Wretman (1977) are for one time period only and are of the linear non-homogeneous estimator form discussed in detail in Chapter 2. Robinson (1982(b)) has extended the one period Horvitz-Thompson and multiple

regression estimators (see Model G_{MR} of Cassel, Sarndal and Wretman (1977:84)) under design unbiasedness constraints to incorporate time series information on individual respondents. Robinson's models and also those of Jessen (1942), Patterson (1950), Gurney and Daly (1965), Blight and Scott (1973), Scott and Smith (1974), and Jones (1980) can be seen as special cases of the linear non-homogeneous sample survey estimator extended into the time domain. This general estimator is the topic of Chapter 3.

The second but tangential line of enquiry has been the comprehensive study, primarily in the econometric literature, of time series/cross sectional data. A fairly complete survey is given in Judge, Griffith, Hill and Lee (1980:Chapter 8). The econometric model differs from the usual sample survey estimation model in two major respects, and is at the same time both more general and more particular. Its generality allows that the parameters to be estimated are not simply mean values over time but may be regression coefficients and/or covariances across individuals or time. Efficiency properties of estimators are paramount, and the mathematical setting is that of the General Linear Model. It is more particular in that almost invariably the data are assumed to be a panel study where all sampled units are sampled at all time periods, and each unit is assumed selected by simple random sampling. General linear model parameters may be either random or fixed. An example is given in Chapter 6, where the heart rate of five hospital patients are each monitored for the same number of time periods under three different treatment regimes, using a model due to Swamy (1971). A recent econometric development which may also be applicable to non-simple, random sample surveys has been to assume individual responses follow a random coefficient autoregressive model (e.g. Robinson (1978, 1986)).

Those papers which do address the sample design problem inherent in rotating and/or non-simple random samples fall into two groups, those that concern themselves with simple regression models and complex survey designs (e.g. Fuller (1975), DeMets and Halperin (1977), Nathan and Holt (1980), Holt, Smith and Winter (1980), Scott and Holt (1982)), and those few that have the complex regression models common to econometrics but relatively uncomplicated sample designs (e.g. Biorn (1981), Lenz and Robinson (1983)). The paper of Porter (1973) while seeming to fill some intermediate ground, has been shown by Pfeiffermann (1984) to involve stochastic rather than fixed design matrices, and therefore not yield b.l.u.e.'s.

There are subtle mathematical difficulties in using raw survey data (i.e. individual respondent information) from non-simple, random samples repeated over time in General Linear Models as the difficulties with Porter's (1973) paper attest. Considerable care, and perhaps a more general framework is necessary in such cases. Chapters 2 and 3 to follow, move some way to providing such a framework even for the econometric problem. The outline in these chapters is restricted to estimation of population means however, and suitable matrix augmentations would need to be used to extend the results to estimation of regression or other parameters.

While providing a framework for estimation of means the very generality of the methods of Chapters 2 and 3 would, in particular situations, often preclude computationally efficient estimation procedures. In Chapter 3 a condition under which individual information may be amalgamated without affecting estimate efficiency is derived for mean estimation. This allows the unbiased elementary estimates so formed to be used in some General Linear Model formulation without creating the potential bias problems often

associated with individual data. Efficiency is not, however, the only consideration. One apparent disadvantage of General Linear Models with either stochastic or non-stochastic coefficients is that general recursive prediction and updating formulae have not previously been available where new or potential new data to be added are known to be correlated with data already included in the estimates. This lack has been an inherent difficulty in those general stochastic formulations of Jones (1980) which do not have a Kalman filter form, and is commented on by Jones (1983:36). Chapters 4 and 5 provide a general solution to this problem, a solution which is not restricted to the sample survey/time series context.

CHAPTER 2

THE LINEAR NON-HOMOGENEOUS SAMPLE SURVEY ESTIMATOR

SUMMARY

For members of the class of linear non-homogeneous sample survey estimators, $\hat{Y} = W_{0(s)} + \sum_{i=1}^N W_{i(s)} Y_i$, of a mean or total, where the $\{W_{i(s)}\}$ are random variables associated with the survey design and the $\{Y_i\}$ random variables associated with the N population units, formulae for determination and estimation of variance are often presented as a series of particular results. In this chapter non-informative survey designs are considered, and general formulae for the expected value of $\hat{Y}$, and general formulae both for determination of bias, variance and mean square error and estimation of variance are derived, firstly in terms of the sampling distribution or design, $p(\cdot)$, secondly in terms of the joint distribution, ξ , of $Y_1, Y_2, \dots, Y_N$, and thirdly in terms of the joint $p\xi$ -distribution. The method subsumes all non-informative "traditional" designs with linear estimators, where a traditional design is one for which $Y_i = y_i$, a constant, for any sample in which unit i is included. These are the almost exclusive topic of many standard sampling texts where appropriate formulae for traditional sample survey estimators are derived without the benefit of general results. This chapter provides the foundation for the extension of these ideas to repeated sample surveys; this extension is the topic of Chapter 3.

* * * * *

1. INTRODUCTION

The variance, estimated variance and form of sample survey estimators of means or totals are often presented as a series of miscellaneous results. In this paper the class of linear non-homogeneous estimators of a mean or total

$$\hat{Y} = W_0(s) + \sum_{i=1}^N W_i(s)Y_i \quad (1.1)$$

is considered for non-informative sample designs (Cassell, Sarndal and Wretman (1977:12) hereafter denoted CSW). It is to be noted that "Informative designs are of rather limited interest from a practical point of view" (CSW (1977:13); see also CSW (1977:133)). When a mean is being estimated, the estimator of equation (1.1) will be denoted $\hat{\bar{Y}}$. The estimator, $\hat{Y}$, has also been called the generalised linear estimator (Basu(1971)). Y_i is a random variable associated with the i th of N population units; if unit i is not in a particular sample, s , the weight $W_i(s)$, in the formula for $\hat{Y}$, equals zero; if unit i is in the sample, $W_i(s)$ may depend on the sample drawn. All non-informative traditional sample survey estimators, linear in the Y_i , are therefore included as special cases. By a traditional sample survey estimator is meant one for which $Y_i = y_i$ whenever unit i is drawn in the sample, s . Such traditional sample survey designs are the topic of many standard sample survey texts, in which appropriate formulae are derived separately for each design and estimator (e.g., Cochran (1977), Raj (1968)).

One restriction is placed on the linear nonhomogeneous estimators for non-informative designs considered here. Estimators incorporating selection order of sampled units are not considered, since application of the Rao-Blackwell theorem to any estimator for which order of selection of the sample units is important, yields an estimator with smaller mean

square error, and for which order of selection is disregarded (CSW (1977:40)). Although the same reduction of mean square error is possible for estimators involving multiplicity of units, estimators of this type are considered, because they include the more commonly used estimators for with replacement sample designs.

General formulae for expected value, bias, variance, mean square error, and estimated variance of $\hat{Y}$ are given, firstly with respect to the sampling distribution inherent in the survey design, $p(\cdot)$, secondly with respect to the joint distribution, ξ , of $Y_1, Y_2, \dots, Y_N$, and thirdly with respect to the joint $p\xi$ distribution inherent in sample design under a superpopulation model. The superpopulation structure may in fact be degenerate (i.e., $Y_i = y_i$ for every sample in which i belongs to s , for $i=1, 2, \dots, N$). This chapter does not attempt to determine optimal designs, in the sense, for example, of minimum mean square error under particular model assumptions. The superpopulation model is discussed in CSW (1977:Chapter 4).

2. TERMINOLOGY

Let $\mathcal{U} = \{u_i : i=1, 2, \dots, N\}$ denote the N units of the finite population, where a random variable Y_i is associated with each unit i . Let ξ denote the joint distribution of $Y_1, Y_2, \dots, Y_N$. For the superpopulation, denote model or ξ -expectation by E . Let $\sigma_{ij} = E(Y_i Y_j) - E(Y_i)E(Y_j)$, and $\sigma_i^2 = E(Y_i^2) - E(Y_i)^2$. Let $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$ and $\bar{\theta} = E(\bar{Y})$ denote the parameters of interest and $\hat{Y}$ be the linear nonhomogeneous estimator of $\bar{Y}$ or $\bar{\theta}$. The extension of results to totals is straightforward. Let s denote a subset of units of the population, and let $\mathcal{S}$ be the set of all subsets, or samples, s . Let $p(s)$ denote the probability of selecting the sample s from $\mathcal{S}$.

A non-informative sample design is defined in CSW (1977:12) to be a sample design, $p(\cdot)$, which is not a function of the Y_i 's associated with the units i belonging to the sample, s (i.e., $i \in s$).

While informativeness is more a property of the sampling scheme (CSW (1977:15)) than of the design $p(\cdot)$ which the sampling scheme implements, this distinction is not critical here. A design may remain non-informative while depending on some auxiliary variable associated with each population unit i .

$\hat{Y}$ has been defined so that it is an estimator of Y or θ constructed as a linear combination of Y_i for $i \in s$, plus an additional constant which may be zero. Now define p - or design expectation of $g(\hat{Y})$ where g is a function of Y by

$$E(g(\hat{Y})) = \sum_{s \in \mathcal{S}} g(\hat{Y})p(s) = \sum_S g(\hat{Y})p(s).$$
 When $W_0(s) = 0$ for all $s \in \mathcal{S}$ the estimator $\hat{Y}_0 = \sum_{i=1}^N W_{i(s)}Y_i$ will be called linear homogeneous.

Let $S_i = \{s : i \in s\}$ and $S_{ij} = \{s : i \in s, j \in s\}$ for $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, N$, with $S_0 = \{s : W_0(s) \neq 0\}$ and $S_{i0} = \{s : i \in s \text{ and } W_0(s) \neq 0\}$ for $i = 1, 2, \dots, N$. In what follows, $i = 1, 2, \dots, N$ will be denoted "for all i " with a parallel notation for j . When zero is included in the range of i or j this will be stated explicitly.

Denote p -expectation of $g(\hat{Y})$ by $E[g(\hat{Y})]$. Note that when $p(\cdot)$ is non-informative, $E[g(\hat{Y})] = \epsilon E[g(\hat{Y})]$ (CSW (1977:109)). Note that where the two expectation operations, E and ϵ are applied sequentially, the operator applied first is a conditional expectation.

Define the p - or design bias of $\hat{Y}$ by $B(\hat{Y}) = E(\hat{Y}) - \bar{Y}$, the p - or design variance of $\hat{Y}$ by $V(\hat{Y}) = E[(\hat{Y} - E(\hat{Y}))^2]$, and the p - or design mean square error of $\hat{Y}$ by $E[(\hat{Y} - \bar{Y})^2]$. Analogously we define for $\hat{Y}$, the ξ - or model bias by $B(\hat{Y}) = E(\hat{Y}) - \bar{\theta}$, ξ - or model variance by $V(\hat{Y}) = E[(\hat{Y} - E(\hat{Y}))^2]$, ξ - or model mean square error by $E[(\hat{Y} - \bar{\theta})^2]$, $p\xi$ - bias by $E E(\hat{Y}) - \bar{\theta}$, $p\xi$ - variance by $E E[(\hat{Y} - E E(\hat{Y}))^2]$ and two possibly different forms of $p\xi$ - mean square error by $E E[(\hat{Y} - \bar{\theta})^2]$ and $E E[(\hat{Y} - \bar{Y})^2]$. $\hat{Y}$ is said to be p -, ξ - or $p\xi$ - unbiased when the corresponding bias is zero, for all possible populations, $\mathcal{U}$.

3. EXPECTATION, BIAS, VARIANCE, AND MEAN SQUARE ERROR OF $\hat{Y}$

The estimator $\hat{Y} = W_0(s) + \sum_{i=1}^N W_i(s) Y_i$ has properties both with respect to the distribution ξ and to the sample design $p(\cdot)$. The model expectation E is defined with respect to the distribution ξ , the joint distribution of $Y_1, Y_2, \dots, Y_N$, which is independent of the sampling distribution of $W_i(s)$ and of $W_i(s)W_j(s)$, for all i and j , since the sample design $p(\cdot)$ is non-informative. The expectation of the $W_i(s)$ and $W_i(s)W_j(s)$ are defined with respect to the p -expectation, E .

For p -, ξ - and $p\xi$ -, we derive below formulae for the expected value, the bias, the variance and the mean square error of $\hat{Y}$.

For all i and j , including i and j equal to zero, we define

$$W_i(s) = \begin{cases} 0 & \text{with probability } \sum_{s \notin S_i} p(s) \\ \beta_i(s) & \text{with probability } p(s) \text{ when } s \in S_i \end{cases} \quad (3.1)$$

$$W_i(s)W_j(s) = \begin{cases} 0 & \text{with probability } \sum_{s \notin S_{ij}} p(s) \\ \beta_i(s)\beta_j(s) & \text{with probability } p(s) \text{ when } s \in S_{ij} \end{cases} \quad (3.2)$$

Although much theoretical work has been done using the properties of $W_i(s)$ and $W_j(s)$ given above (e.g. CSW (1977)) the necessity of enumerating each possible sample $s \in \mathcal{S}$ makes the above definitions unwieldy for deriving p- and pξ- properties (e.g. expectation, bias, variance, or mean square error for $\hat{Y}$) given a particular sample design, $p(\cdot)$.

We therefore define for each i and j including zero

$$\Pi_i(q) = P(W_i(s) = w_i(q)) \quad (3.3)$$

and

$$\begin{aligned} \Pi_{ij}(q)r &= P(W_i(s) = w_i(q) \text{ and } W_j(s) = w_j(q)) \quad (3.4) \\ \text{for } q &= 0, 1, 2, \dots, k_i, \quad r = 0, 1, 2, \dots, k_j, \end{aligned}$$

where $P(A)$ denotes the probability of the event A .

Note $W_i(0) = w_j(0) = 0$

and $\{w_i(q)\} =$ the set of distinct values of $\beta_i(s)$ for samples, $s \in \mathcal{S}$, for which $p(s) > 0$.

The $\Pi_i(q)$ and $\Pi_{ij}(q)r$ are determined in the obvious way.

Note that $q \geq 1$ implies $i \in s$; that $r \geq 1$ implies $j \in s$; and that $\Pi_{ij}(q)r = \Pi_i(q)$ if $q = r$, and zero otherwise. The inclusion probability for unit i , Π_i , (i.e. the probability that $i \in s$) and the joint inclusion probability for units i and j , Π_{ij} , are given by:

$$\Pi_i = \sum_{s \in S_i} p(s) = \sum_{q=1}^{k_i} \Pi_i(q) \quad (3.5)$$

and

$$\Pi_{ij} = \sum_{s \in S_{ij}} p(s) = \sum_{q=1}^{k_i} \sum_{r=1}^{k_j} \Pi_{ij}(q)r \quad (3.6)$$

Hence

$$E(W_i(s)) = \sum_{q=1}^{k_i} w_i(q) \Pi_i(q) \quad (3.7)$$

for all i including zero.

Similarly

$$E(W_i(s)W_j(s)) = \sum_{q=1}^{k_i} \sum_{r=1}^{k_j} w_i(q)w_j(r) \Pi_i(q)\Pi_j(r) \quad (3.8)$$

and thus

$$\text{cov}(W_i(s)W_j(s)) = \sum_{q=1}^{k_i} \sum_{r=1}^{k_j} w_i(q)w_j(r) (\Pi_i(q)\Pi_j(r) - \Pi_i(q)\Pi_j(r)) \quad (3.9)$$

for all i and j including zero

We may define $\text{cov}(W_i(s)W_j(s))$ without ambiguity, and making no reference to the expectation operator E , since the design, $p(\cdot)$, is non-informative. If $\hat{Y}$ is homogeneous then $\text{var}(W_0(s)) = 0$ and $\text{cov}(W_0(s)W_i(s)) = 0$ for all i .

Many estimators, $\hat{Y}$, while of the form of equation (1.1), allow $W_i(s)$ to take only two values, 0 and $w_{i(1)}$ (= w_i say), for all $s \in \mathcal{S}$ with $p(s) > 0$. This class of estimator which we will call bivalued in $\{W_i(s)\}$ includes the expansion estimator for all without replacement sample designs, $p(\cdot)$, for which $\{W_i(s)\}$ depends only on whether $i \in s$. Many of these estimators are also homogeneous, e.g., the expansion estimator for simple random sampling without replacement, cluster sampling without replacement of clusters, stratified random sampling with units selected without replacement, and two stage cluster sampling with units selected without replacement at both stages. For bivalued $W_i(s)$, therefore $\Pi_{i(1)} = \Pi_i$ and $\Pi_{i(1)j(1)} = \Pi_{ij}$ for all i and j where Π_i and Π_{ij} , are given by equations (3.5) and (3.6).

Thus for $W_{i(s)}$ bivalued, and for all i and j

$$E(W_{i(s)}) = w_i \Pi_i \quad (3.10)$$

$$E(W_{i(s)} W_{j(s)}) = w_i w_j \Pi_{ij} \quad (3.11)$$

and
$$\text{cov}(W_{i(s)} W_{j(s)}) = w_i w_j (\Pi_{ij} - \Pi_i \Pi_j) \quad (3.12)$$

with $\text{var}(W_{0(s)})$ and $\text{cov}(W_{0(s)} W_{i(s)})$ equal to zero for homogeneous $\hat{Y}$, and given via equations (3.9) otherwise.

The formulae given in either equations (3.7) to (3.9), or (3.10) to (3.12), may be substituted as appropriate in the formulae which follow.

Theorem 3.1

The estimator $\hat{\bar{Y}} = W_{0(s)} + \sum_{i=1}^N W_{i(s)} Y_i$ of $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$, with $W_{i(s)} = 0$ if $i \notin s$ for all i , has p- or design expectation

$$E(\hat{\bar{Y}}) = E(W_{0(s)}) + \sum_{i=1}^N E(W_{i(s)}) Y_i \quad (3.13)$$

p- or design bias

$$B(\hat{\bar{Y}}) = E(W_{0(s)}) + \sum_{i=1}^N (E(W_{i(s)}) - \frac{1}{N}) Y_i \quad (3.14)$$

p- or design variance

$$\begin{aligned} V(\hat{\bar{Y}}) = \text{var}(W_{0(s)}) + 2 \sum_{i=1}^N Y_i \text{cov}(W_{0(s)}, W_{i(s)}) \\ + \sum_{i=1}^N \sum_{j=1}^N Y_i Y_j \text{cov}(W_{i(s)}, W_{j(s)}) \end{aligned} \quad (3.15)$$

and p- or design mean square error

$$E[(\hat{\bar{Y}} - \bar{Y})^2] = V(\hat{\bar{Y}}) + B(\hat{\bar{Y}})^2 \quad (3.16)$$

Proof

The results follow directly since the survey design, $p(\cdot)$ is non-informative. Setting $W_0(s) = 0$ for all $s \in \mathcal{S}$ in Theorem 2.1 yields the parallel results for the linear homogeneous estimator $\hat{Y}_0$.

The quantities $E(\hat{Y})$, $B(\hat{Y})$, $V(\hat{Y})$ and $E[(\hat{Y} - \bar{Y})^2]$, being functions of the random variables Y_i , $i=1,2,\dots,N$ with joint distribution ξ , are themselves random variables with respect to ξ . Determination of a realisation of each of $E(\hat{Y})$, $B(\hat{Y})$, $V(\hat{Y})$ and $E[(\hat{Y} - \bar{Y})^2]$ requires knowledge of one realisation of each of Y_i for each i .

Theorem 3.2

The estimator $\hat{Y} = W_0(s) + \sum_{i=1}^N W_i(s) Y_i$ of $\bar{\theta} = \xi(\bar{Y}) = \frac{1}{N} \sum_{i=1}^N \xi(Y_i)$, with $W_i(s) = 0$ if $i \notin s$, has model expectation

$$\xi(\hat{Y}) = W_0(s) + \sum_{i=1}^N W_i(s) \xi(Y_i) \quad (3.17)$$

model bias

$$\mathcal{B}(\hat{Y}) = W_0(s) + \sum_{i=1}^N (W_i(s) - \frac{1}{N}) \xi(Y_i) \quad (3.18)$$

model variance

$$\mathcal{V}(\hat{Y}) = \sum_{i=1}^N \sum_{j=1}^N W_i(s) W_j(s) \sigma_{ij} \quad (3.19)$$

and model mean square error

$$\xi[(\hat{Y} - \bar{\theta})^2] = \mathcal{V}(\hat{Y}) + \mathcal{B}(\hat{Y})^2 \quad (3.20)$$

Proof

Follows directly from the definitions of model expectation, bias, variance and mean square error, since the sample design is non-informative.

The quantities $\mathcal{E}(\hat{\bar{Y}})$, $\mathcal{B}(\hat{\bar{Y}})$, $\mathcal{V}(\hat{\bar{Y}})$ and $\mathcal{E}[(\hat{\bar{Y}} - \bar{\theta})^2]$ are random variables, like $E(\hat{\bar{Y}})$, $B(\hat{\bar{Y}})$, $V(\hat{\bar{Y}})$ and $E[(\hat{\bar{Y}} - \bar{Y})^2]$, but with respect to the sampling distribution $p(s)$ for $s \in \mathcal{S}$ (i.e. the sample design, $p(\cdot)$), not with respect to ξ the joint distribution of $Y_1, Y_2, \dots, Y_N$.

Theorem 3.3

The estimator $\hat{\bar{Y}} = W_0(s) + \sum_{i=1}^N W_i(s) Y_i$ of $\bar{\theta} = \frac{1}{N} \sum_{i=1}^N \mathcal{E}(Y_i)$, with $W_i(s)$ for $i \in s$, has $p\xi$ - expectation

$$E\mathcal{E}(\hat{\bar{Y}}) = E(W_0(s)) + \sum_{i=1}^N E(W_i(s)) \mathcal{E}(Y_i) \quad (3.21)$$

$p\xi$ - bias

$$E\mathcal{E}(\hat{\bar{Y}}) - \bar{\theta} = E(W_0(s)) + \sum_{i=1}^N (E(W_i(s)) - \frac{1}{N}) \mathcal{E}(Y_i) \quad (3.22)$$

$p\xi$ - variance

$$E\mathcal{E}[(\hat{\bar{Y}} - E\mathcal{E}(\hat{\bar{Y}}))^2] = \text{var } W_0(s) \quad (3.23)$$

$$\begin{aligned} &+ 2 \sum_{i=1}^N \mathcal{E}(Y_i) \text{cov}(W_0(s), W_i(s)) \\ &+ \sum_{i=1}^N \sum_{j=1}^N [\mathcal{E}(Y_i Y_j) E(W_i(s) W_j(s)) \\ &\quad - \mathcal{E}(Y_i) \mathcal{E}(Y_j) E(W_i(s)) E(W_j(s))] \end{aligned}$$

and $p\xi$ - mean square error

$$E\mathcal{E}[(\hat{\bar{Y}} - \bar{\theta})^2] = E\mathcal{E}[(\hat{\bar{Y}} - E\mathcal{E}(\hat{\bar{Y}}))^2] + (E\mathcal{E}(\hat{\bar{Y}}) - \bar{\theta})^2 \quad (3.24)$$

or

$$\begin{aligned} E\mathcal{E}[(\hat{\bar{Y}} - \bar{Y})^2] &= E\mathcal{E}[(\hat{\bar{Y}} - E\mathcal{E}(\hat{\bar{Y}}))^2] \\ &\quad + (E\mathcal{E}(\hat{\bar{Y}}) - \bar{\theta})^2 + \mathcal{E}(\hat{\bar{Y}} - \bar{\theta})^2 \end{aligned} \quad (3.25)$$

Proof

$E\hat{\xi}(\hat{Y}) = E(W_0(s)) + \sum_{i=1}^N E(W_i(s)) \xi(Y_i)$ since the sampling distribution of the $W_i(s)$, $i=0,1,\dots,N$, is independent of the joint distribution of $Y, Y, \dots, Y_N$. The formulae for $p\xi$ - bias follows directly.

To determine $p\xi$ - variance, note that, in the sequence of expectation operators $E\xi$, ξ is the superpopulation expectation conditional on the sample, s . It follows that

$$E\xi[\hat{Y} - E\xi(\hat{Y})]^2 = \xi V(\hat{Y}) + \nu E(\hat{Y}) \quad (3.26)$$

see Mood, Graybill and Boes (1974:159), C.R. Rao (1973:97)

Now defining $Y_0 = 1$,

$$\xi V(\hat{Y}) = \sum_{i=0}^N \sum_{j=0}^N \xi(Y_i Y_j) \text{cov}(W_i(s), W_j(s))$$

and

$$\begin{aligned} \nu E(\hat{Y}) &= \nu [E(W_0(s)) + \sum_{i=1}^N E(W_i(s)) Y_i] \\ &= \sum_{i=1}^N \sum_{j=1}^N E(W_i(s)) E(W_j(s)) [\xi(Y_i Y_j) - \xi(Y_i) \xi(Y_j)]. \end{aligned}$$

Equation (3.23) follows directly from equation (3.26).

The first formulae for $p\xi$ - mean square error follows from the expansion of $E\xi[(\hat{Y} - E\xi(\hat{Y})) + (E\xi(\hat{Y}) - \bar{\theta})]^2$ on noting that the cross-product term is zero. The second formulae, for the second, and distinct, form of $p\xi$ - mean square error, follows similarly.

Corollary 1

The two forms of the $p\xi$ - mean square error of $\hat{Y}$ are not equivalent. Call $E\xi[(\hat{Y} - \bar{Y})^2]$ the ξ -expectation of the design mean square error, and $E\xi[(\hat{Y} - \bar{\theta})^2]$ the p -expectation of the model mean

square error. Note that the E-expectation of the model mean square error equals the first form of the $p\xi$ -mean square error.

Proof

$$\begin{aligned} E\epsilon[(\hat{\bar{Y}} - \bar{\theta})^2] &= \epsilon E[(\hat{\bar{Y}} - \bar{\theta})^2] \text{ since } p(.) \text{ is non-informative} \\ &= \epsilon E[(\hat{\bar{Y}} - \bar{Y})^2] + \epsilon[(\hat{\bar{Y}} - \bar{\theta})^2] \\ &\neq \epsilon E[(\bar{Y} - \bar{Y})^2] \text{ in general.} \end{aligned}$$

Now for a non-informative design $\epsilon E[(\hat{\bar{Y}} - \bar{Y})^2] = E\epsilon[(\hat{\bar{Y}} - \bar{Y})^2]$. Thus $E\epsilon[(\hat{\bar{Y}} - \bar{\theta})^2] \neq E\epsilon[(\bar{Y} - \bar{Y})^2]$ in general. Further $E\epsilon[(\hat{\bar{Y}} - \bar{\theta})^2] = E[\epsilon[(\hat{\bar{Y}} - \bar{\theta})^2]]$ as required.

The condition for equality of the two forms of mean square error is that $\epsilon[(\hat{\bar{Y}} - \bar{\theta})^2] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij} = 0$.

CSW (1977:92) note that minimisation of the ξ expectation of the p -mean square error is the most fully explored of the possible criteria for judging the uncertainty of strategies, and their Chapters 4-6 are based on this criterion.

Conditions for p - and $p\xi$ - unbiasedness of $\hat{\bar{Y}}$ are given in CSW p. 27, and p. 92 respectively. Model of ξ - unbiasedness holds if $W_0(s) + \sum_{i=1}^N (W_i(s) - \frac{1}{N})\epsilon(Y_i) = 0$ for all possible realisations of $W_i(s)$. p - or ξ - unbiasedness is sufficient to imply $p\xi$ - unbiasedness (CSW (1977:92)).

4. UNBIASED ESTIMATION OF VARIANCE OF $\hat{\bar{Y}}$

For estimating p- and p ξ - variance of $\hat{\bar{Y}}$, we are interested in estimation given complete knowledge of the survey design, p(.), including $E(W_{i(s)})$ and $E(W_{i(s)}W_{j(s)})$ for all i and j including zero. Thus, for example, given a particular sample, s, we may know Y_i only for $i \in s$, but we will know $W_{i(s)}$ for all i including zero (with $W_{i(s)} = 0$ if $i \notin s$). We assume that $E(W_{i(s)})$ and $E(W_{i(s)}W_{j(s)})$ for all i and j including zero are also known, since these are determined solely by the survey design, p(.).

For the estimated ξ -variance of $\hat{\bar{Y}}$, knowledge of the survey design, p(.), yields no information. For estimating the ξ -variance of $\hat{\bar{Y}}$ we require estimators of the σ_{ij} , where σ_{ij} is the ξ -covariance between Y_i and Y_j .

It is possible to derive formulae for estimation of bias and mean square error, but these require knowledge of $\hat{\bar{Y}}$ or of $\bar{\theta}$. This is a serious restriction where information on $\hat{\bar{Y}}$ or $\bar{\theta}$ is restricted to their estimated values from a single sample, and estimation of bias and mean square error are therefore not considered here. We return to the question of variance estimation.

We define

$$X_i(s) = \begin{cases} 0 & \text{if } W_{i(s)} = 0 \\ 1 & \text{if } W_{i(s)} \neq 0 \end{cases} \quad \text{for all i including zero} \quad (4.1)$$

Further, to avoid indeterminate forms, we define

$$\frac{X_i(s)X_j(s)}{\Pi_{ij}} = 1 \quad \text{if } \Pi_{ij} = 0 \quad (4.2)$$

and

$$\frac{W_i(s)W_j(s)}{E(W_i(s)W_j(s))} = 1 \quad \text{if } E(W_i(s)W_j(s)) = 0 \quad (4.3)$$

Theorem 4.1

For the estimator $\hat{\bar{Y}} = W_0(s) + \sum_{i=1}^N W_i(s)Y_i$ of $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$, with $W_i(s) = 0$ if $i \notin s$, two p-unbiased estimators are given by

$$\begin{aligned} \widetilde{V(\hat{\bar{Y}})} = \text{var}(W_0(s)) + 2 \sum_{i \in s} X_0(s)Y_i \frac{\text{cov}(W_0(s), W_i(s))}{\Pi_{0i}} \\ + \sum_{i \in s} \sum_{j \in s} Y_i Y_j \frac{\text{cov}(W_i(s), W_j(s))}{\Pi_{ij}} \end{aligned} \quad (4.4)$$

and

$$\begin{aligned} \widehat{V(\hat{\bar{Y}})} = \text{var}(W_0(s)) + 2 \sum_{i \in s} W_0(s)W_i(s)Y_i \frac{\text{cov}(W_0(s), W_i(s))}{E(W_0(s)W_i(s))} \\ + \sum_{i \in s} \sum_{j \in s} W_i(s)W_j(s)Y_i Y_j \frac{\text{cov}(W_i(s), W_j(s))}{E(W_i(s)W_j(s))} \end{aligned} \quad (4.5)$$

provisional on the definitions (4.2) and (4.3) respectively when $\Pi_{ij} = 0$ or $E(W_i(s)W_j(s)) = 0$.

Proof

Equations (4.4) and (4.5) follow as a consequence of definitions (4.2) and (4.3), and noting that $E(X_i(s)X_j(s)) = \Pi_{ij}$ for all i and j including zero.

For any estimator for which $W_i(s)$ is bivalued for $i=1,2,\dots,N$ and for which either $W_0(s) = 0$ for all $s \in \mathcal{L}$ or $W_0(s)$ is itself bivalued, $\widehat{V(\hat{\bar{Y}})} = \widetilde{V(\hat{\bar{Y}})}$.

Theorem 4.2

For the estimator $\hat{\bar{Y}} = W_0(s) + \sum_{i=1}^N W_i(s)Y_i$ of $\bar{\theta} = \xi(\bar{Y}) = \frac{1}{N} \sum_{i=1}^N \epsilon(Y_i)$,

two ξ -unbiased estimators of ξ -variance are given by

$$\widehat{V(\hat{\bar{Y}})} = \sum_{i=1}^N \sum_{j=1}^N W_i(s)W_j(s) (Y_i - \epsilon(Y_i))(Y_j - \epsilon(Y_j)) \quad (4.6)$$

and

$$\widetilde{V(\hat{\bar{Y}})} = \sum_{i=1}^N \sum_{j=1}^N W_i(s)W_j(s) (Y_i Y_j - \epsilon(Y_i) \epsilon(Y_j)) \quad (4.7)$$

Proof

Taking expectations of $\widehat{V(\hat{\bar{Y}})}$ and $\widetilde{V(\hat{\bar{Y}})}$ with respect to ξ yields $\xi[\widehat{V(\hat{\bar{Y}})}] = \widehat{V(\hat{\bar{Y}})} = \xi[\widetilde{V(\hat{\bar{Y}})}]$ as required.

Define $Y_0 = \xi(Y_0) = 1$, and, if $\hat{\bar{Y}}$ is nonhomogeneous for the particular sample s , let $i \in s, j \in s$ include i and j equal to zero in Theorem 4.3 to follow.

Define

$$z_{ij}^{(1)} = Y_i Y_j E(W_i(s)W_j(s)) - \epsilon(Y_i)\epsilon(Y_j)E(W_i(s))E(W_j(s)) \quad (4.8)$$

$$z_{ij}^{(2)} = Y_i Y_j \text{cov}(W_i(s), W_j(s)) + (Y_i - \epsilon(Y_i))(Y_j - \epsilon(Y_j))E(W_i(s))E(W_j(s)) \quad (4.9)$$

$$z_{ij}^{(3)} = Y_i Y_j \text{cov}(W_i(s), W_j(s)) + \sigma_{ij} E(W_i(s))E(W_j(s)) \quad (4.10)$$

noting that $\sigma_{0j} = \sigma_{i0} = 0$

We then have the following theorem:

Theorem 4.3

The estimator $\hat{\bar{Y}} = W_0(s) + \sum_{i=1}^N W_i(s)Y_i$ of $\bar{\theta} = \xi(\hat{\bar{Y}}) = \frac{1}{N} \sum_{i=1}^N \epsilon(Y_i)$, with $W_i(s) = 0$ for $i \notin s$ has a number of $p\xi$ -unbiased estimators

of $p\xi$ -variance given by

$$\hat{R}_k = \sum_{i \in s} \sum_{j \in s} (z_{ij}^{(k)} / \pi_{ij})$$

and

$$\hat{R}_{k+3} = \sum_{i \in s} \sum_{j \in s} [W_i(s) W_j(s) z_{ij}^{(k)} / E(W_i(s) W_j(s))]$$

for $k=1,2,3$

where $z_{ij}^{(k)}$ for $k=1,2,3$ are defined by equations (4.8) to (4.10).

Proof

Follows directly from taking the $p\xi$ -expectation of R_1 to R_6 , and making use of equations (4.1) to (4.3) as appropriate.

5. THE TRUE AND ESTIMATED VARIANCE OF TRADITIONAL LINEAR

SAMPLE ESTIMATORS

The formulae appropriate to traditional survey design (i.e., survey design in the absence of a superpopulation model) can be deduced directly from Theorem 3.1 or Theorem 3.3 and Theorem 4.1 or 4.3 and their corollaries by setting $Y_i = y_i$ whenever unit i is drawn in a sample $s \in \mathcal{S}$. y_i is then the only possible value of Y_i for the fixed population, $\mathcal{U}$. Note however that while the usual formulae for true variance can be derived in this way, not all possible formulae for estimated variance can be deduced. This is discussed in more detail in Examples 5.2 and 5.4 below.

Let $\hat{\bar{y}}$ be $\hat{\bar{Y}}$ given the sample, s (and hence $W_i(s)$ for $i=0,1,\dots,N$, and y_i for $i \in s$). We consider estimation of the mean, $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$.

Example 5.1

Simple Random Sampling Without Replacement

$\hat{Y}$ is homogeneous with bivalued $w_i(s)$ for $i=1,2,\dots,N$

$$w_i(s) = \begin{cases} w_i = \frac{1}{n} & \text{if } i \in s \\ 0 & \text{if } i \notin s \end{cases}$$

$$\pi_i = \frac{n}{N}$$

$$\pi_{ij} = \begin{cases} \frac{n(n-1)}{N(N-1)} & , i \neq j \\ \frac{n}{N} & , i = j \end{cases}$$

Thus

$$\hat{Y} = \sum_{i=1}^N w_i(s) Y_i = \sum_{i \in s} w_i Y_i = \frac{1}{n} \sum_{i \in s} Y_i$$

$$\begin{aligned} V(\hat{Y}) &= \sum_{i=1}^N \sum_{j=1}^N w_i w_j (\pi_{ij} - \pi_i \pi_j) Y_i Y_j \\ &= \sum_{i=1}^N \frac{1}{n^2} \frac{n(N-n)}{N^2} Y_i^2 - \sum_{i=1}^N \sum_{j=1}^N \frac{1}{n^2} \frac{n(N-n)}{N(N-1)} Y_i Y_j \\ &= \frac{1}{n} \left(\frac{N-n}{N} \right) \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2. \end{aligned}$$

and from Theorem 4.1 or 4.3

$$\begin{aligned} \widehat{V(\hat{Y})} &= \widetilde{V(\hat{Y})} = \sum_{i \in s} \sum_{j \in s} w_i w_j \left(1 - \frac{\pi_i \pi_j}{\pi_{ij}} \right) Y_i Y_j \\ &= \frac{1}{n^2} \left(\frac{N-n}{N} \right) \left[\sum_{i \in s} Y_i^2 - \frac{1}{n-1} \left(\sum_{i \in s} \sum_{\substack{j \in s \\ i \neq j}} Y_i Y_j \right) \right] \\ &= \frac{1}{n} \left(\frac{N-n}{N} \right) \left[\frac{1}{n-1} \sum_{i \in s} (Y_i - \hat{Y})^2 \right] \end{aligned}$$

Example 5.2

Horwitz-Thompson Estimator

$\hat{Y}$ is homogeneous with bivalued $W_{i(s)}$ for $i=1,2,\dots,N$

$$W_{i(s)} = \begin{cases} w_i = \frac{1}{N\pi_i} & \text{for } i \in s \\ 0 & \text{for } i \notin s \end{cases}$$

Thus

$$\hat{Y} = \sum_{i=1}^N W_{i(s)} Y_i = \sum_{i \in s} \frac{Y_i}{N\pi_i},$$

$$\begin{aligned} V(\hat{Y}) &= \sum_{i=1}^N \sum_{j=1}^N w_i w_j (\pi_{ij} - \pi_i \pi_j) Y_i Y_j \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} Y_i Y_j \end{aligned}$$

and

$$\begin{aligned} \widehat{V(\hat{Y})} &= \widetilde{V(\hat{Y})} = \sum_{i \in s} \sum_{j \in s} \frac{1}{\pi_i \pi_j} \left(1 - \frac{\pi_i \pi_j}{\pi_{ij}}\right) Y_i Y_j \\ &= \sum_{i \in s} \frac{1 - \pi_i}{\pi_i^2} Y_i^2 + \sum_{i \in s} \sum_{\substack{j \in s \\ i \neq j}} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j \pi_{ij}} Y_i Y_j \end{aligned}$$

which is equation (9A.43) of Cochran (1977).

Note however that this estimate of variance is not unique c.f.

The Yates and Grundy (1953) estimator.

$$\sum_{i \in s} \sum_{\substack{j \in s \\ j > i}} \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j} \right).$$

Estimates of variance other than $\widehat{V(\hat{Y})}$ or $\widetilde{V(\hat{Y})}$, make use of algebraic identities particular to the form of $\hat{Y}$ for a given sampling scheme. Occasionally approximations are used (e.g., ratio estimation). For the Horwitz-Thompson estimator the identity used

to show the Yates and Grundy (1953) estimator to be p-unbiased for

$$V(\hat{\bar{y}}), \text{ is } \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N (\pi_{ij} - \pi_i \pi_j) = -\pi_i (1 - \pi_i).$$

$$\begin{aligned} \text{Thus } V(\hat{\bar{y}}) &= \sum_i \frac{1 - \pi_i}{\pi_i} y_i^2 + \sum_{i \neq j} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} y_i y_j \\ &= \sum_{i=1}^N \sum_{j>i}^N (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \end{aligned}$$

and the Yates and Grundy (1953) estimator is thus obviously a p-unbiased estimator of $V(\hat{\bar{y}})$. No generalisation to sample designs not confirming to the Howitz-Thompson pattern (e.g., with replacement sample designs) is possible, however.

Example 5.3

Two Stage Cluster Sampling : Sampling Without Replacement At Both Stages

Suppose we choose m clusters from M , and n_i units from N_i in selected cluster c .

Then

$$W_i(s) = \begin{cases} \frac{M}{N} \frac{N_c}{mn_c} & \text{for } i \in s, i \in \text{cluster } c \\ 0 & \text{for } i \notin s \end{cases}$$

$$\pi_i = \frac{mn_c}{MN_c} \quad \text{where } i \in \text{cluster } c$$

$$\pi_{ij} = \begin{cases} \frac{mn_c}{MN_c} & \text{where } i=j \in \text{cluster } c \\ \frac{m}{M} \frac{n_c(n_c-1)}{N_c(N_c-1)} & \text{where } i, j \in \text{cluster } c, i \neq j \\ \frac{m(m-1)}{M(M-1)} \frac{n_c}{N_c} \frac{n_{c'}}{N_{c'}} & \text{where } i \in \text{cluster } c, j \in \text{cluster } c', c \neq c' \end{cases}$$

$$\text{Let } \bar{y}_c = \frac{1}{N_c} \sum_{i \in c} y_i; \quad \text{Thus } \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{M} \sum_{c=1}^M N_c \bar{y}_c.$$

$$\text{Let } S_b^2 = \frac{1}{M-1} \sum_{c=1}^M (N_c \bar{y}_c - \bar{y})^2$$

$$\text{and } S_{wc}^2 = \frac{1}{N_c-1} \sum_{\substack{i=1 \\ i \in c}}^{N_c} (y_i - \bar{y}_c)^2 \text{ for } c=1, 2, \dots, M.$$

Thus

$$\begin{aligned} V(\hat{\bar{y}}) &= \sum_{i=1}^N \sum_{j=1}^N w_i w_j (\Pi_{ij} - \Pi_i \Pi_j) y_i y_j \\ &= \frac{1}{N^2} \frac{M}{m} \left\{ \frac{M-m}{M} \left(- \frac{1}{M-1} \sum_{\substack{i \in c \\ c \neq c'}} \sum_{\substack{j \in c' \\ c \neq c'}} y_i y_j \right) \right. \\ &\quad + \sum_{\substack{i \neq j \\ i, j \in c}} \sum_{l=1}^N \left(\frac{M-m}{M} - \frac{N_c - n_c}{n_c(N_c-1)} \right) y_i y_j \\ &\quad \left. + \sum_{l=1}^N \left(\frac{M-m}{M} + \frac{N_c - n_c}{n_c} \right) y_l^2 \right\} \\ &= \frac{1}{N^2} \frac{M}{m} \left\{ \frac{M-m}{M} \left[\sum_{c=1}^M N_c^2 \bar{y}_c^2 - \frac{1}{M-1} \left(\sum_{c=1}^M \sum_{\substack{c'=1 \\ c \neq c'}}^M N_c \bar{y}_c N_{c'} \bar{y}_{c'} \right) \right] \right. \\ &\quad \left. + \sum_{c=1}^M \frac{N_c - n_c}{N_c} \left[\left(\sum_{i \in c} y_i^2 \right) - \frac{1}{N_c-1} \sum_{\substack{i \in c \\ i \neq j}} \sum_{j \in c} y_i y_j \right] \right\} \\ &= \frac{1}{N^2} M^2 \left(\frac{1}{m} - \frac{1}{M} \right) S_b^2 + \frac{1}{N^2} \frac{M}{m} \sum_{c=1}^M N_c^2 \left(\frac{1}{n_c} - \frac{1}{N_c} \right) S_{wc}^2 \end{aligned}$$

where

$$\hat{\bar{y}} = \frac{1}{N} \frac{M}{m} \sum_{c \in S} \left[\frac{N_c}{n_c} \left(\sum_{\substack{i \in S \\ i \in c}} y_i \right) \right]$$

To derive the usual formula for estimated variance from Theorem 4.1

or 4.3, define

$$s_b^2 = \frac{1}{m-1} \sum_{c \in S} (N_c \bar{y}_c - \frac{1}{m} \sum_{c \in S} N_c \hat{\bar{y}}_c)^2$$

where $\hat{\bar{y}}_c = \frac{1}{n_c} \sum_{i \in c} y_i$ for the m clusters c which belong to s ($c \in s$)

and $s_{wc}^2 = \frac{1}{n_c - 1} \sum_{i \in c} (y_i - \hat{\bar{y}}_c)^2$ for each of the m clusters, $c \in s$.

Then

$$\begin{aligned}
 \widehat{V(\bar{y})} &= \widehat{V(\bar{y})} = \sum_{i \in s} \sum_{j \in s} w_i w_j \left(1 - \frac{\pi_i \pi_j}{\pi_{ij}} \right) y_i y_j \\
 &= \frac{1}{N^2} \frac{M}{m} \left\{ \frac{M-m}{m} \left[-\frac{1}{m-1} \sum_{\substack{i \in s, j \in s \\ i \in c, j \in c' \\ c \neq c'}} \frac{N_c}{n_c} \frac{N_{c'}}{n_{c'}} y_i y_j \right] \right. \\
 &\quad + \sum_{\substack{i \in s, j \in s \\ i, j \in c}} \left(\frac{M-m}{m} \frac{N_c^2}{n_c^2} - \frac{N_c(N_c - n_c)}{n_c^2(n_c - 1)} \right) y_i y_j \\
 &\quad \left. + \sum_{\substack{i \in s \\ i \in c}} \left(\frac{M-m}{m} \frac{N_c^2}{n_c^2} + \frac{N_c(N_c - n_c)}{n_c^2} \right) y_i^2 \right\} \\
 &= \frac{1}{N^2} \frac{M}{m} \left\{ \frac{M-m}{m} \left[\sum_{c \in s} \left(N_c^2 \hat{\bar{y}}_c^2 - \frac{1}{m-1} \sum_{\substack{c \in s, c' \in s \\ c \neq c'}} N_c N_{c'} \hat{\bar{y}}_c \hat{\bar{y}}_{c'} \right) \right] \right. \\
 &\quad \left. + \sum_{c \in s} \frac{N_c(N_c - n_c)}{n_c^2} \left[\sum_{i \in c} y_i^2 - \frac{1}{n_c - 1} \sum_{\substack{i, j \in c \\ i \neq j}} y_i y_j \right] \right\} \\
 &= \frac{1}{N^2} \frac{M}{m} \left\{ \frac{M-m}{m-1} \sum_{c \in s} \left[(N_c \hat{\bar{y}}_c)^2 - \frac{1}{m} \left(\sum_{c \in s} N_c \bar{y}_c \right)^2 \right] \right. \\
 &\quad \left. + \sum_{c \in s} \frac{N_c(N_c - n_c)}{n_c(n_c - 1)} \left[\sum_{i \in c} \left(y_i^2 - \frac{1}{n_c} \left(\sum_{i \in c} y_i \right)^2 \right) \right] \right\} \\
 &= \frac{M^2}{N^2} \left(\frac{1}{m} - \frac{1}{M} \right) s_b^2 + \frac{M}{N^2 m} \sum_{c \in s} N_c^2 \left(\frac{1}{n_c} - \frac{1}{N_c} \right) s_{wc}^2.
 \end{aligned}$$

Example 5.4

Varying Probability Sampling With Replacement

$\hat{\bar{y}}$ is homogeneous, but W_i is not bivalued for $i=1,2,\dots,N$. Let p_i be the probability of selection of unit i at any of n draws.

$$W_i(s) = \begin{cases} 0 & \text{if } i \notin s \\ \frac{q}{Nnp_i} & \text{if } i \in s, \text{ } q \text{ times} \end{cases}$$

i.e., $W_i(s)$ has $(n+1)$ potential values, including zero.

$$\Pi_i(q) = \frac{n!}{(n-q)!q!} p_i^q (1-p_i)^{n-q}$$

$$\Pi_i(q)j(r) = \frac{n!}{q!r!(n-q-r)!} p_i^q p_j^r (1-p_i-p_j)^{n-q-r}$$

Thus $\hat{\bar{y}} = \frac{1}{Nn} \sum_{d=1}^n \frac{y_d}{p_d}$ (where d denotes the unit drawn at the d th draw) which can be shown to be p -unbiased for $\bar{y}$ via the binomial theorem. Also $\text{cov}(W_i(s), W_j(s)) = -\frac{1}{nN^2}$ for $i \neq j$ via the binomial and multinomial theorems, and $\text{var}(W_i(s)) = \text{cov}(W_i(s), W_i(s)) = \frac{1}{N^2n} \left(\frac{1-p_i}{p_i} \right)$ via the binomial theorem, with $E(W_i(s)W_j(s)) = \frac{n-1}{nN^2}$ for $i \neq j$, and $E(W_i^2(s)) = \frac{1+p_i(n-1)}{p_i N^2 n}$.

Hence

$$\begin{aligned} V(\hat{\bar{y}}) &= \sum_{i=1}^N \sum_{j=1}^N \text{cov}(W_i(s), W_j(s)) y_i y_j \\ &= \frac{1}{nN^2} \left[\sum_{i=1}^N \frac{y_i^2}{p_i} - \left(\sum_{i \neq j}^N \sum_{j=1}^N y_i y_j + \sum_{i=1}^N y_i^2 \right) \right] \\ &= \frac{1}{nN^2} \left[\sum_{i=1}^N \left(\frac{y_i}{p_i} - N\bar{y} \right)^2 p_i \right]. \end{aligned}$$

Note: $\hat{\bar{y}}$ does not depend on the order of the draws, $d=1,2,\dots,n$.

For estimated variance we have from Theorem 4.1 or 4.3

$$\widehat{V(\hat{y})} = \frac{1}{N^2 n(n-1)} \left[\sum_{i \in s'} \left(\frac{a_i^2}{1+(n-1)p_i} \right) \left(\frac{y_i}{p_i} \right)^2 - \frac{1}{N} \left(\sum_{d=1}^n \frac{y_d}{p_d} \right)^2 \right]$$

where a_i = number of times unit i appears in the sample, s

$$s' = \{\text{distinct elements of } s\}$$

and d denotes the unit drawn at the d th drawn of n . As a preliminary to considering $\widetilde{V(y)}$ we note that:

For $i = j$

$$\Pi_{ij} = \Pi_i = 1 - (1-p_i)^n$$

For $i \neq j$

$$\Pi_{ij} = 1 - (1-p_i)^n - (1-p_j)^n + (1-p_i-p_j)^n.$$

Thus

$$\begin{aligned} \widetilde{V(\hat{y})} = \frac{1}{nN^2} & \left[\sum_{i \in s'} \left(\frac{1-p_i}{p_i} \right) \frac{(a_i y_i)^2}{(1-(1-p_i)^n)} \right. \\ & \left. - \sum_{\substack{i \in s' \\ i \neq j}} \sum_{j \in s'} \frac{(a_i y_i)(a_j y_j)}{(1-(1-p_i)^n - (1-p_j)^n + (1-p_i-p_j)^n)} \right] \end{aligned}$$

Unlike our other examples therefore neither $\widehat{V(\hat{y})}$ nor $\widetilde{V(\hat{y})}$ is the usual estimator of $V(\hat{y})$ given in the sample survey texts.

6. CONCLUSION

While we have derived very general formulae for statistical properties of the linear non-homogeneous estimator, given a survey design and a superpopulation structure, it has also been shown how, in the more restrictive traditional survey design context, these general formulae in many cases, give the formulae derived individually in standard sample survey texts. This chapter is therefore useful on four grounds: firstly, in deriving the appropriate formulae under a superpopulation model for use in optimisation contexts; secondly, in providing a more coherent framework for the teaching of traditional sample survey design; thirdly, in suggesting unbiased estimation formulae for variance of an estimator for particular sample surveys; and lastly, for suggesting how the very general formulae derived in this chapter can be extended to repeated sample surveys. This last is the subject of the next chapter.

CHAPTER 3

THE LINEAR NON-HOMOGENEOUS SAMPLE SURVEY ESTIMATOR IN A TIME SERIES CONTEXT

SUMMARY

The idea and framework of the linear non-homogeneous estimator is here extended from a single survey to repeated surveys. In addition, since many of the results which have been derived for estimators for repeated surveys fall under the heading of General Linear Models, the effect on parameter estimates in such models of data aggregation, such as occurs for elementary estimates, is discussed. These two threads are brought together to review previous published results, and some general conclusions on repeated survey estimators are drawn.

* * * * *

1. INTRODUCTION

The results of the previous chapter for the linear non-homogeneous sample survey estimator can be extended into the time domain. Many previously published models (e.g. Patterson (1950), Gurney and Daly (1965), Blight and Scott (1973), Scott and Smith (1974), Jones (1980)) can be seen as problems set in this context. The linear non-homogeneous sample survey/time series estimator (denoted hereafter the lnhssts estimator) because of its generality does not provide the most elegant solution to many of the problems that can be posed within its framework. Since many of the published models seek to minimise some form of variance or mean square error subject to certain constraints (e.g. unbiasedness), these problems are more elegantly solved by the techniques of least squares, generalised least squares, stochastic least squares or Kalman Filters, as appropriate.

The lnhssts estimator does, however, provide a method of systematising known results in the time domain, just as the linear homogeneous sample survey estimator systematises conventional sample survey theory, and this motivates extension of the linear non-homogeneous sample survey estimator into the time domain.

2. NOTATION

To simplify the notation we consider only the situation where the population consists of the same N units at each time $t = 1, 2, \dots, h$ at which the survey is repeated. The extension to situations where there are births and deaths in the population is, in principle, relatively straightforward, requiring only suitable augmentation of the appropriate vectors and matrices.

Let $\mathcal{U} = \{u_i : i = 1, 2, \dots, N\}$ denote the N units of the finite population as previously and let the random variable $Y_i(t)$ be associated with unit i at time $t = 1, 2, \dots, h$. Let $\underline{Y}_i = (Y_i(1), Y_i(2), \dots, Y_i(h))^T$. Let ξ denote the joint distribution of $\underline{Y}_1, \underline{Y}_2, \dots, \underline{Y}_N$ and denote ξ -expectation by ε . For $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, N$ let $\Sigma_{ij} = \varepsilon(\underline{Y}_i \underline{Y}_j^T) - \varepsilon(\underline{Y}_i) \varepsilon(\underline{Y}_j)^T$. Let $\bar{\underline{Y}} = \frac{1}{N} \sum_{i=1}^N \underline{Y}_i$ and $\bar{\theta} = \varepsilon(\bar{\underline{Y}})$ be the parameters of interest and let

$$\hat{\underline{Y}} = \underline{W}_0(\underline{s}) + \sum_{i=1}^N W_i(\underline{s}) \underline{Y}_i \quad (2.1)$$

where $\hat{\underline{Y}} = (\hat{Y}(1), \hat{Y}(2), \dots, \hat{Y}(h))^T$

$$\underline{W}_0(\underline{s}) = (W_0(\underline{s})(1), W_0(\underline{s})(2), \dots, W_0(\underline{s})(h))^T$$

$$W_i(\underline{s}) = (\underline{W}_i(\underline{s})(1), \underline{W}_i(\underline{s})(2), \dots, \underline{W}_i(\underline{s})(h)) \quad \text{for } i = 1, 2, \dots, N$$

$$\underline{Y}_i = (Y_i(1), Y_i(2), \dots, Y_i(h))^T \quad \text{for } i = 1, 2, \dots, N$$

$$\underline{s} = (s(1), s(2), \dots, s(h))^T$$

$W_i(\underline{s})$ is $h \times h$ for each $i = 1, 2, \dots, N$. $\hat{\underline{Y}}(t)$ and $\underline{Y}_i(t)$ for $t = 1, 2, \dots, h$ parallel $\bar{\underline{Y}}$ and $\underline{Y}_i$ of equation (1.1) of Chapter 2. For each $i = 1, 2, \dots, N$, the elements of the matrix of weights, $W_i(\underline{s})$, differ from the scalar $W_i(s)$ of the same equation only in that they may multiply the response of unit i at any time $t = 1, 2, \dots, h$ and are dependent on all the samples taken between times $t = 1$ and $t = h$, i.e. on $\underline{s} = (s(1), s(2), \dots, s(h))$. The fact that the weights may depend on all possible samples between times $t = 1$ and $t = h$ allows the weights to be influenced by the rotation pattern of the sample over time (i.e. the pattern in time with which individual units, u_i , arrive and leave

the sample). The symbol $\underline{s}$ then denotes a collection of h subsets of the population $\mathcal{U}$. For each such $\underline{s}$ these subsets are not necessarily mutually exclusive. Let S denote the set of all subsets $\underline{s}$ and $p(\underline{s})$ denote the probability of selecting the sample $\underline{s}$ from S . All survey schemes are again assumed non-informative. Define p - or design expectation of $g(\hat{\underline{Y}})$ where g is a function of $\hat{\underline{Y}}$ by

$$E(g(\hat{\underline{Y}})) = \sum_{\underline{s} \in S} g(\hat{\underline{Y}}) p(\underline{s}) = \sum_{\underline{s}} g(\hat{\underline{Y}}) p(\underline{s})$$

When $W_0(\underline{s})$ of equation 2.1 equals zero for all $\underline{s} \in S$ the estimator $\hat{\underline{Y}}$ will be called linear homogeneous.

Define p - or design bias of $\hat{\underline{Y}}$ by $B(\hat{\underline{Y}}) = E(\hat{\underline{Y}}) - \bar{\underline{Y}}$, the p - or design variance of $\hat{\underline{Y}}$ by $V(\hat{\underline{Y}}) = E[(\hat{\underline{Y}} - E(\hat{\underline{Y}}))(\hat{\underline{Y}} - E(\hat{\underline{Y}}))^T]$ and the p - or design mean square error of $\hat{\underline{Y}}$ by $E[(\hat{\underline{Y}} - \bar{\underline{Y}})(\hat{\underline{Y}} - \bar{\underline{Y}})^T]$. Analogously for $\hat{\underline{Y}}$ define ξ - or model bias by $\beta(\hat{\underline{Y}}) = \epsilon(\hat{\underline{Y}}) - \bar{\underline{\theta}}$, ξ - or model variance by $\mathcal{V}(\hat{\underline{Y}}) = E[(\hat{\underline{Y}} - \epsilon(\hat{\underline{Y}}))(\hat{\underline{Y}} - \epsilon(\hat{\underline{Y}}))^T]$, ξ - or model mean square error by $E[(\hat{\underline{Y}} - \bar{\underline{\theta}})(\hat{\underline{Y}} - \bar{\underline{\theta}})^T]$, $p\xi$ - bias by $E\epsilon(\hat{\underline{Y}}) - \bar{\underline{\theta}}$, $p\xi$ -variance by $E\epsilon[(\hat{\underline{Y}} - E\epsilon(\hat{\underline{Y}}))(\hat{\underline{Y}} - E\epsilon(\hat{\underline{Y}}))^T]$ and two possibly distinct forms of $p\xi$ -mean square error by $E\epsilon\hat{\underline{Y}} - \bar{\underline{\theta}}^T$ and $E\epsilon[(\hat{\underline{Y}} - \bar{\underline{Y}})(\hat{\underline{Y}} - \bar{\underline{Y}})^T]$.

Note that since the survey design is non-informative $E[g(\hat{\underline{Y}})] = E[g(\bar{\underline{Y}})]$. Define $\text{var}(W_{i(\underline{s})})$ and $\text{cov}(W_{i(\underline{s})}, W_{j(\underline{s})})$ for $i = 0, 1, 2, \dots, N$; $j = 0, 1, \dots, N$, in the usual way with respect to the design expectation, E .

Note also that $\hat{\underline{Y}}$ is a time series improved, rather than a traditional estimator of $\bar{\underline{Y}}$, unless the set $\{W_{i(\underline{s})} : \underline{s} \in S, i = 0, 1, 2, \dots, N\}$ contains only diagonal matrices when the estimator will be denoted $\hat{\underline{Y}}$. Such estimators $\hat{\underline{Y}}$ use data from more than one time period, and under certain conditions are equivalent to $\hat{\underline{I}}_0$ or $\hat{\underline{I}}$ of Chapter 1 (see Section 5).

3. EXPECTATION, BIAS, VARIANCE AND MEAN SQUARE ERROR

We thus have the following three theorems which are analogues of Theorems 3.1 to 3.3 of Chapter 2. For notational simplicity we adopt the convention that $\tilde{W}_0(\underline{s}) = \tilde{W}_0(\underline{s})\tilde{Y}_0$ for some suitably defined matrix $\tilde{W}_0(\underline{s})$ and vector $\tilde{Y}_0$ (e.g. $\tilde{Y}_0 = \underline{1}$ where $\underline{1}$ is a vector of 1's, and $\text{diag } \tilde{W}_0(\underline{s}) = \tilde{W}_0(\underline{s})$ with all off diagonal elements of $\tilde{W}_0(\underline{s})$ equal to zero).

Theorem 3.1

The lnhsts estimator of $\tilde{Y} = \frac{1}{N} \sum_{i=1}^N \tilde{Y}_i$, given by equation (2.1) with $\tilde{W}_i(\underline{s})(t) = 0$ if $i \notin s(t)$ for $i = 1, 2, \dots, N$ has p- or design expectation

$$E(\tilde{Y}) = E(\tilde{W}_0(\underline{s})) + \sum_{i=1}^N E(\tilde{W}_i(\underline{s}))\tilde{Y}_i \quad (3.1)$$

p- or design bias

$$B(\tilde{Y}) = E(\tilde{W}_0(\underline{s})) + \sum_{i=1}^N (E(\tilde{W}_i(\underline{s})) - \frac{1}{N} I)\tilde{Y}_i \quad (3.2)$$

p- or design variance

$$V(\tilde{Y}) = \sum_{i=0}^N \sum_{j=0}^N [E(\tilde{W}_i(\underline{s})\tilde{Y}_i\tilde{Y}_j^T\tilde{W}_j^T(\underline{s})) - E(\tilde{W}_i(\underline{s}))\tilde{Y}_i\tilde{Y}_j^TE(\tilde{W}_j^T(\underline{s}))] \quad (3.3)$$

p- or design mean square error

$$E[(\tilde{Y} - \tilde{Y})(\tilde{Y} - \tilde{Y})^T] = V(\tilde{Y}) + B(\tilde{Y})B(\tilde{Y})^T \quad (3.4)$$

where I is the $(h \times h)$ identity matrix and $\tilde{W}_0(\underline{s})\tilde{Y}_0 = \tilde{W}_0(\underline{s})$ by definition.

Proof:

The results follow directly since the design is non-informative.

Note that the expectation operator E in $E(W_i(\underline{s}) Y_i Y_j^T W_j^T(\underline{s}))$ treats $Y_i Y_j^T$ as fixed.

Theorem 3.2

The lnhssts estimator of $\bar{\theta} = \varepsilon(\bar{Y})$, given by equation (2.1), with $W_i(\underline{s})(t) = 0$ if $i \notin s(t)$ for $i = 1, 2, \dots, N$ has ξ - or model expectation

$$\varepsilon(\hat{\bar{Y}}) = W_0(\underline{s}) + \sum_{i=1}^N W_i(\underline{s}) \varepsilon(Y_i) \quad (3.5)$$

model bias

$$\mathcal{B}(\hat{\bar{Y}}) = W_0(\underline{s}) + \sum_{i=1}^N (W_i(\underline{s}) - \frac{1}{N} I) \varepsilon(Y_i) \quad (3.6)$$

model variance

$$v(\hat{\bar{Y}}) = \sum_{i=1}^N \sum_{j=1}^N W_i(\underline{s}) \Sigma_{ij} W_j^T(\underline{s}) \quad (3.7)$$

and model mean square error

$$\varepsilon[(\hat{\bar{Y}} - \bar{\theta})(\hat{\bar{Y}} - \bar{\theta})^T] = v(\hat{\bar{Y}}) + \mathcal{B}(\hat{\bar{Y}}) \mathcal{B}(\hat{\bar{Y}})^T \quad (3.8)$$

Proof:

The results follow directly since the design is non-informative.

Note that $\hat{\bar{Y}}$ is model unbiased for $\bar{\theta}$ iff $\mathcal{B}(\hat{\bar{Y}}) = 0$ for all possible joint realisations of $W_0(\underline{s})$ and $\{W_i(\underline{s}) : i = 1, 2, \dots, N\}$. Given a particular grouping of samples, $\underline{s} = (s(1), s(2), \dots, s(h))$,

we may thus determine the quantities $\epsilon(\hat{\tilde{Y}}|\tilde{s})$ and $\nu(\hat{\tilde{Y}}|\tilde{s})$. Given $\bar{\theta}$ or that $\epsilon(\tilde{Y}_i) = \bar{\theta}$ we may in addition determine $\beta(\hat{\tilde{Y}}|\tilde{s})$ and $\epsilon(\hat{\tilde{Y}} - \bar{\theta})(\hat{\tilde{Y}} - \bar{\theta})^T$.

Theorem 3.3

The lnhss estimator of $\bar{\theta}$, given by equation (2.1) with $\tilde{W}_i(\tilde{s})(t) = 0$ if $i \notin s(t)$ for $i = 1, 2, \dots, N$ has p ξ -expectation

$$E\epsilon(\hat{\tilde{Y}}) = E(W_0(\tilde{s})) + \sum_{i=1}^N E(W_i(\tilde{s}))\epsilon(\tilde{Y}_i) \quad (3.9)$$

p ξ -bias

$$E\epsilon(\hat{\tilde{Y}}) - \bar{\theta} = E(W_0(\tilde{s})) + \sum_{i=1}^N (E(W_i(\tilde{s})) - \frac{1}{N} I)\epsilon(\tilde{Y}_i) \quad (3.10)$$

p ξ -variance

$$\begin{aligned} E\epsilon[(\hat{\tilde{Y}} - E\epsilon(\hat{\tilde{Y}}))(\hat{\tilde{Y}} - E\epsilon(\hat{\tilde{Y}}))^T] \\ = \sum_{i=0}^N \sum_{j=0}^N [E\{W_i(\tilde{s})\epsilon(\tilde{Y}_i\tilde{Y}_j^T)W_j^T(\tilde{s})\} - E(W_i(\tilde{s}))\epsilon(\tilde{Y}_i)\epsilon(\tilde{Y}_j^T)E(W_j^T(\tilde{s}))\}] \end{aligned} \quad (3.11)$$

where $\epsilon(\tilde{Y}_0) = \tilde{Y}_0$ by definition, and two possibly different forms of p ξ -mean square error, namely

$$E\epsilon(\hat{\tilde{Y}} - \bar{\theta})(\hat{\tilde{Y}} - \bar{\theta})^T = E\epsilon[(\hat{\tilde{Y}} - E\epsilon(\hat{\tilde{Y}}))(\hat{\tilde{Y}} - E\epsilon(\hat{\tilde{Y}}))^T] + (E\epsilon(\hat{\tilde{Y}}) - \bar{\theta})(E\epsilon(\hat{\tilde{Y}}) - \bar{\theta})^T \quad (3.12)$$

and

$$\begin{aligned}
 E\epsilon \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} - \begin{pmatrix} \bar{Y} \\ \sim \end{pmatrix} \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} - \begin{pmatrix} \bar{Y} \\ \sim \end{pmatrix}^T &= E\epsilon \left[\begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} - E\epsilon \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} \right] \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} - E\epsilon \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix}^T \\
 &+ \left(E\epsilon \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} - \begin{pmatrix} \bar{Y} \\ \sim \end{pmatrix} \right) \left(E\epsilon \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} - \begin{pmatrix} \bar{Y} \\ \sim \end{pmatrix} \right)^T \\
 &- \epsilon \left[\begin{pmatrix} \bar{Y} \\ \sim \end{pmatrix} - \begin{pmatrix} \bar{\theta} \\ \sim \end{pmatrix} \right] \begin{pmatrix} \bar{Y} \\ \sim \end{pmatrix} - \begin{pmatrix} \bar{\theta} \\ \sim \end{pmatrix}^T \quad (3.13)
 \end{aligned}$$

Proof:

The proof parallels that for Theorem 3.3 of Chapter 2 given that the single design is non-informative on noting that using the multivariate extension of Rao (1973:97),

$$E\epsilon \left[\begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} - E\epsilon \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} \right] \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} - E\epsilon \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix}^T = \epsilon V \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix} + \tau E \begin{pmatrix} \hat{Y} \\ \sim \end{pmatrix}.$$

4. THE EFFECT OF DATA AGGREGATION ON PARAMETER ESTIMATES FROM GENERAL LINEAR MODELS

Because general linear model formulations are such an integral part of time series improvement of sample survey estimators, we digress to consider conditions under which stochastic and non-stochastic least squares parameter estimates using mean values as data are the same as those using individual respondent information. Initially, simple random sampling is considered, but other sample selection schemes are discussed in a later section. The results hold in general provided the design matrix X has the form given in equation (4.1) below.

To avoid possible confusion in the notation we again specify the parameters of interest, namely the sample survey means

$\underline{Y} = (Y(1), Y(2), \dots, Y(h))^T$ by $\underline{\tau} = (\tau(1), \tau(2), \dots, \tau(h))^T$, the best linear unbiased estimator of $\underline{Y}$ by $\hat{\underline{\tau}}$, and the traditional sample survey estimators of $\underline{Y}$ by $\hat{\underline{Y}} = (\hat{Y}(1), \hat{Y}(2), \dots, \hat{Y}(h))^T$. C.f. $\hat{\underline{Y}}$.

For individual sampled respondent information $\{y_{ij}(t) :$
 $j = 1, 2, \dots, n(t); t = 1, 2, \dots, h\}$, where $y_{ij}(t)$ denotes the response
of the j^{th} sampled individual of $n(t)$ at time t , and $y_{ij}(1)$ and $y_{ij}(2)$
do not necessarily denote responses from the same individual at times
1 and 2, the appropriate linear model is:

$$\begin{bmatrix} y_{i_1}(1) \\ y_{i_2}(1) \\ \vdots \\ y_{i_{n(1)}}(1) \\ \dots\dots\dots \\ y_{i_1}(2) \\ \vdots \\ y_{i_{n(2)}}(2) \\ \dots\dots\dots \\ \vdots \\ y_{i_{n(h)}}(h) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots\dots & 0 \\ 1 & 0 & & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & \\ \dots\dots & & & \\ 0 & 1 & & \\ & 1 & & \\ \vdots & \vdots & & \\ \vdots & \vdots & & \\ \vdots & 1 & & \\ \vdots & \dots\dots & & \\ \vdots & 0 & & \\ \vdots & \vdots & & 1 \\ 0 & 0 & \dots\dots & 1 \end{bmatrix} \begin{bmatrix} \tau(1) \\ \tau(2) \\ \vdots \\ \tau(h) \end{bmatrix} + \begin{bmatrix} e_{i_1}(1) \\ e_{i_2}(2) \\ \vdots \\ e_{i_{n(1)}}(1) \\ \dots\dots\dots \\ e_{i_1}(2) \\ \vdots \\ e_{i_{n(2)}}(2) \\ \dots\dots\dots \\ \vdots \\ e_{i_{n(h)}}(h) \end{bmatrix} \quad (4.1)$$

i.e. $\underline{y}_{(h)} = X \underline{\tau} + \underline{e}$

where variance of $\underline{e}$ is

$$\begin{bmatrix} \Sigma_{11} & \dots\dots & \Sigma_{1h} \\ \Sigma_{21} & & \\ \vdots & & \\ \Sigma_{h1} & & \Sigma_{hh} \end{bmatrix} = \Sigma, \text{ say.}$$

Stochastic least squares yields

$$\hat{\tau}_{\sim} = (T^{-1} + X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y_{\sim(h)} \quad (4.2)$$

or

$$\hat{\tau}_{\sim} = TX^T(\Sigma + XTX^T)^{-1} Y_{\sim(h)} \quad (4.3)$$

where $var(\tau_{\sim}) = T$

and ordinary least squares yields

$$\hat{\tau}_{\sim 0} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y_{\sim(h)} \quad (4.4)$$

For mean values at each time

$$\begin{bmatrix} \hat{Y}(1) \\ \hat{Y}(2) \\ \vdots \\ \hat{Y}(h) \end{bmatrix} = \begin{bmatrix} \tau(1) \\ \tau(2) \\ \vdots \\ \tau(h) \end{bmatrix} + \begin{bmatrix} \overline{e(1)} \\ \overline{e(2)} \\ \vdots \\ \overline{e(h)} \end{bmatrix} \quad (4.5)$$

i.e. $\hat{Y}_{\sim} = \tau_{\sim} + \bar{e}_{\sim}$

with the variance of $\bar{e}_{\sim}$ given by

$$(X^T X)^{-1} (X^T \Sigma X) (X^T X)^{-1}$$

Note:

$$X^T X = \begin{bmatrix} n_1 & & 0 \\ & n_2 & \\ & & \ddots \\ 0 & & & n_h \end{bmatrix} = nI \text{ where sample sizes are equal at each time } t \quad (4.6)$$

$$\hat{Y}_{\sim} = (X^T X)^{-1} X^T Y_{\sim(h)} = \frac{1}{n} X^T Y_{\sim(h)} \quad \text{with equal sample sizes at each time } t \quad (4.7)$$

Stochastic least squares yields

$$\tilde{\tau} = (T^{-1} + (X^T X)(X^T \Sigma X)^{-1} X^T X)^{-1} X^T X (X^T \Sigma X)^{-1} X^T X \hat{y}_{\sim} \quad (4.8)$$

or

$$\tilde{\tau} = T((X^T X)^{-1} (X^T \Sigma X)(X^T X)^{-1} + T)^{-1} \hat{y}_{\sim} \quad (4.9)$$

Ordinary least squares gives

$$\begin{aligned} \tilde{\tau}_0 &= [(X^T X)^{-1} X^T \Sigma X (X^T X)^{-1}]^{-1} (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} \hat{y}_{\sim} \\ &= \hat{y}_{\sim} = (X^T X)^{-1} X^T y_{\sim(h)} \end{aligned} \quad (4.10)$$

Thus for fixed τ , $\tilde{\tau}_0 = \hat{\tau}_0$ with $\text{var}(\tilde{\tau}_0) = \text{var}(\hat{\tau}_0) = (X^T X)^{-1} \sigma^2$
iff $\Sigma = \sigma^2 I$.

For stochastic τ the situation is a little more complicated. From equations (4.6) - (4.8),

$$\begin{aligned} \tilde{\tau} &= \{T^{-1} + (X^T X)(X^T \Sigma X)^{-1} X^T X\}^{-1} X^T X (X^T \Sigma X)^{-1} X^T X (X^T X)^{-1} X^T y_{\sim(h)} \\ &= \{X^T \Sigma X (X^T X)^{-1} T^{-1} + X^T X\}^{-1} X^T y_{\sim(h)} \\ &= T[X^T X \{(X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} + T\}]^{-1} X^T y_{\sim(h)} \end{aligned}$$

Thus $\tilde{\tau} = \hat{\tau}$ from equation (4.3) iff

$$X^T (\Sigma + X T X^T)^{-1} = [X^T X \{(X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} + T\}]^{-1} X^T$$

i.e.

$$X^T \{X(X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} X^T + X T X^T\} (\Sigma + X T X^T)^{-1} = X^T$$

i.e.

$$X^T \{ X(X^T X)^{-1} X^T \Sigma X(X^T X)^{-1} X^T \} = X^T \Sigma \quad (4.11)$$

To interpret equation (4.11) note that the variance of $\tilde{e}$ is Σ , and variance of $\tilde{e}$ is $(X^T X)^{-1} X^T \Sigma X(X^T X)^{-1}$. Also

$$X(X^T X)^{-1} X^T = \begin{bmatrix} 1/n_1 & \dots & 1/n_1 & & & \\ \vdots & & & & & 0 \\ 1/n_1 & \dots & 1/n_1 & & & \\ \hline & & & 1/n_2 & \dots & 1/n_2 \\ & & & \vdots & & \vdots \\ & & & 1/n_2 & \dots & 1/n_2 \\ & & & & \ddots & \\ & & & & & 1/n_h & \dots & 1/n_h \\ & & & & & \vdots & & \\ & & & & & 1/n_h & \dots & 1/n_h \end{bmatrix}$$

Thus $X(X^T X)^{-1} X^T \Sigma X(X^T X)^{-1} X^T$ has as its t - t' th block, a matrix with all elements equal to $1/(n_t n_{t'})$ times the sum of the elements of the sub-matrix $\Sigma_{tt'}$, where t, t' are two times between 1 and h . Thus $X^T \{ X(X^T X)^{-1} X^T \Sigma X(X^T X)^{-1} X^T \}$ has as its t - k th element, $1/n_t$, times the sum of all the elements of $\Sigma_{tt'}$, where k belongs to the t' th block, while $X^T \Sigma$ has as its t - k th element the sum of the m th column of $\Sigma_{tt'}$, where $m = k - \sum_{\ell=1}^{t'-1} n(\ell)$. Thus the condition for the equality of $\tilde{\tau}$ and $\hat{\tau}$ for stochastic $\tilde{\tau}$ is that each of the column sums of $\Sigma_{tt'}$, for given t and t' are equal. Under these conditions $\tilde{\tau}$ and $\hat{\tau}$ have equal mean square errors. Since $\Sigma_{tt'} = \Sigma_{t't}^T$ we have proved the following lemma:

Lemma 4.1

For simple random sampling, the condition for equality of the stochastic least squares estimator using individual respondent information, to the stochastic least squares estimator using mean values, and for the equality of their mean square errors, is that the matrix of covariances for individual respondent errors has the property that the sums of the columns in the tt' block are equal, and the sums of the rows in the tt' block are equal for $t = 1, 2, \dots, h$ and $t' = 1, 2, \dots, h$.

* * * * *

Sufficient conditions for Lemma 4.1 to hold are that for each pair of time periods t and t' there is equality of covariances between different individuals, and equality of variances for (re)sampled individuals. Lemma 4.1 is applied to a number of particular estimators in the following section.

The result can easily be extended to a partial condensation of individual respondent information, e.g. to a model involving matched and unmatched means in which case the equality conditions must be met in the appropriate blocks of the error covariance matrix for individual respondent information.

5. PREVIOUS PUBLISHED RESULTS AND THE lnhsts ESTIMATOR

Patterson's (1950) model can now be seen in the framework of the lnhsts estimator. Patterson considers the linear homogeneous estimator

$$\hat{Y}(t) = \sum_{i=1}^N \sum_{t'=1}^h W_{it}(t') Y_i(t') \quad \text{for } t = 1, 2, \dots, h, \quad (5.1)$$

where the $\{W_{it}(t')\}$ are random variables, the realised value of $W_{it}(t')$

for $i \notin s$ is zero. Letting $n(t')$ be the sample size at time t' , for each t we let the $n(t')$ non zero realisations of $W_{it}(t')$ at time t' be $\{w_{it}(t') : i = 1, 2, \dots, n(t')\}$. The variance of $\hat{\bar{Y}}(t)$ for $t = 1, 2, \dots, h$ is then minimised subject to the unbiasedness conditions given by Patterson as

$$\sum_{i \in s(t')} w_{it}(t') = \begin{cases} 1 & \text{for } t' = t, \\ 0 & \text{for } t' \neq t, \end{cases} \quad t' = 1, 2, \dots, h, \quad (5.2)$$

and the constraint that for $i \in s(t) \cap s(t-1)$ and hence $i = 1, 2, \dots, N$ and for $t = 2, 3, \dots, h$,

$$Y_i(t) - \bar{Y}(t) = \rho(Y_i(t-1) - \bar{Y}(t-1)) + \eta_i(t), \quad (5.3)$$

where $\cap$ denotes intersection. Nonstochastic least squares provides the most elegant method of solution. Note from Section 3 that, strictly speaking, each $Y_i(t)$, $i = 1, 2, \dots, N$ has stochastic properties only with respect to the superpopulation (cf. the less precise notation of Chapter 1) and that consequently equation (5.3) involves an implicit superpopulation model, unless i is redefined as the i th sampled unit. From a comparison of the unbiasedness conditions with Theorem 3.2 after stacking the $\hat{\bar{Y}}(t)$ for $t = 1, 2, \dots, h$ as a vector, unbiasedness can be viewed as a model unbiasedness with respect to a superpopulation model for which $\varepsilon(Y_i(t)) = \varepsilon(\bar{Y}(t))$ for all i and for $t = 1, 2, \dots, h$. Infinite population size is no real problem since the weights $\tilde{w}_{it}(t)$ are non-zero only if $i \in s(t)$. It is further assumed that $\varepsilon[Y_i(t) - \varepsilon\{Y_i(t)\}]^2 = S^2$ for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, h$. Thus $\varepsilon\{\eta_i(t)\} = 0$ and $\varepsilon\{\eta_i(t)^2\} = (1 - \rho^2)S^2$ for all i , and $t = 1, 2, \dots, h$. Assuming constant population variance over time, S^2 is traditionally defined as $1/(N-1)\sum(Y_i(t) - \bar{Y}(t))^2$ at each time t where the sum is from 1 to N , and $\bar{Y}(t)$ is non-random. In a superpopulation context, Patterson's prescription also has the property that $\bar{Y}(t) = \varepsilon(\bar{Y}(t)) = \bar{\theta}(t)$. One interpretation of Patterson's (1950) model then, involves optimisation with respect to a superpopulation, albeit one where the population mean $\bar{Y}(t)$ for $t = 1, 2, \dots, h$ has no stochastic properties.

The apparent anomaly of Patterson's (1950) model, that each $Y_i(t)$ has stochastic properties but their mean at time t , $\bar{Y}(t)$, does not, was pointed out in Blight and Scott (1973). The Blight and Scott (1973) model can also be seen in the context of the lnhssts estimator. Take

$$\hat{\bar{Y}}(t) = \sum_{i=1}^N \sum_{t'=1}^h w_{it}(t') Y_i(t') \quad \text{for } t = 1, 2, \dots, h \quad (5.4)$$

as previously. The unbiasedness conditions

$$\sum_{i \in S(t)} w_{it}(t') = \begin{cases} 1 & \text{for } t' = t \\ 0 & \text{for } t' \neq t \end{cases} \quad t = 1, 2, \dots, h \quad (5.5)$$

again apply, and since it is assumed that $\varepsilon(Y_i(t)) = \varepsilon(\bar{Y}(t))$ for all i and for $t = 1, 2, \dots, h$, the unbiasedness condition can be taken to be ξ unbiasedness. (Note that ξ unbiasedness is sufficient to imply $p\xi$ -unbiasedness.) Here however we have an additional constraint to

$$Y_i(t) - \bar{Y}(t) = \rho(Y_i(t-1) - \bar{Y}(t-1)) + \eta_i(t) \quad (5.6)$$

for all i , and $t = 2, 3, \dots, h$, namely that

$$(\bar{Y}(t) - \bar{\theta}) = \lambda(\bar{Y}(t-1) - \bar{\theta}) + \alpha(t) \quad (5.7)$$

for $t = 2, 3, \dots, h$, where $\bar{\theta} = \varepsilon(\bar{Y}(t))$ for all $t = 1, 2, \dots, h$.

Blight and Scott (1973) further assume that $\{\eta_i(t) : t = 2, 3, \dots, h; i = 1, 2, \dots, n(t)\}$ are uncorrelated, that they are also uncorrelated with the $\{\alpha(t) : t = 2, 3, \dots, h\}$ and that each $\eta_i(t)$ has mean zero and variance $(1 - \rho^2)S^2$ where S^2 is the population variance again expressible here as $\varepsilon(Y_i(t) - \varepsilon Y_i(t))^2$ for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, h$. The variance

of $\alpha(t)$ is defined to be σ_α^2 for $t = 2, 3, \dots, h$ as in Chapter 1.

Note that all the expectations discussed in the previous paragraph may, without ambiguity, be taken to be model or ξ -expectations because of the exchangeability (Cassell, Sarndal and Wretman (1977 : 72)) of the responses, $Y_i(t)$, for the units i at each time $t = 1, 2, \dots, h$.

With reference to Theorem 3.3 Blight and Scott (1973) minimise the ξ -expectation of the p-mean square error of $\hat{\bar{Y}}(t)$ for $t = 1, 2, \dots, h$, given by equation (3.13), subject to the restrictions of equations (5.5)-(5.7). The most elegant method of solution involves the Kalman Filter.

The exchangeability of the $\{Y_i(t)\}$ given t , where $t = 1, 2, \dots, h$, is not sufficient to guarantee that each of the estimates based on minimum mean square error, subject to the restrictions of equations (5.5) to (5.7) is a symmetric function of $\{Y_i(t) : i = 1, 2, \dots, n(t)\}$ for each $t = 1, 2, \dots, h$. (The exchangeability of the $\{Y_i(t)\}$ over all times t and $\bar{Y}(t)$ non-stochastic for $t = 1, 2, \dots, h$, would be sufficient conditions but would imply $\bar{Y}(t) = \bar{Y}(t')$ for $t = 1, 2, \dots, h$; $t' = 1, 2, \dots, h$. See particularly Lemma 3.6 Cassell, Sarndal and Wretman (1977).)

Blight and Scott (1973) consider the minimisation in terms of statistics sufficient for $\{\bar{Y}(t) : t = 2, 3, \dots, h\}$ under multivariate normality assumptions, namely $\{\hat{\bar{Y}}(t)', \hat{\bar{Y}}(t)", \hat{\bar{Y}}(t)''' : t = 2, 3, \dots, h\}$, where $\hat{\bar{Y}}(t)'$ is the mean of units matched with observations in the previous sample, $\hat{\bar{Y}}(t)''$ is the mean of units unmatched with the previous sample, and $\hat{\bar{Y}}(t)'''$ is the mean of units matched with the following sample. However derivation of the Blight and Scott (1973) estimator requires no explicit additional restrictions that the weights $\{W_{it}(t)\}$ are equal for units which contribute to the same mean (i.e. $\hat{\bar{Y}}(t)', \hat{\bar{Y}}(t)''$ or $\hat{\bar{Y}}(t)'''$ for

$t = 2, 3, \dots, h$). That these weights will be equal follows from Lemma 4.1 since covariance between different individuals is taken to be zero, and for the same individual is $\rho^{|t-t'|} s^2$, where t and t' are the two time periods for which individual correlations are being considered, and s^2 the sampling variance at each time $t = 1, 2, \dots, h$.

There are three further points of interest. The first is that Lemma 4.1 is satisfied for the model of Blight and Scott (1973) when the model uses as data appropriate matched and unmatched means (since Blight and Scott's (1973) model implies only one non-zero element in each row or column of a given block and equality of these non-zero elements within each block). The second is that Lemma 4.1 will be approximately satisfied when the population is large and the sample design is so constructed as to make diagonal elements of Σ_{tt} , approximately equal. The third is that even where the necessary condition for Lemma 4.1 to hold is not satisfied, but the signal to noise variance ratio is high (as indicated by the relative "magnitudes" of XTX^T and Σ as measured by their respective traces), equation (4.11) will be approximately satisfied. Jones (1983:45) has noted that under the condition that $\{\bar{Y}(t)\}$ is an autoregressive process of order one, using stochastic least squares on sample data "will only be worthwhile when the signal to noise variance ratio is small". He has also indicated that his results may be generalisable beyond first order autoregressive models (Jones (1983:53)). This would suggest that for stochastic parameters τ for which Lemma 4.1 is not satisfied exactly, there may well be considerable gains in using individual respondent information rather than mean values at least for samples selected by simple random sampling in construction of an appropriate linear model. Stochastic least squares does not generally apply, without some suitable bias adjustment, for samples selected by random sampling schemes other than

simple random sampling. These are discussed later in this section.

Cassell, Sarndal and Wretman (1977:92) note that minimisation of the ξ expectation of the p-mean square error is the most fully explored of the possible criteria for judging the uncertainty of sampling strategies when only one time period is involved, and their comment is also true for sampling strategies for repeated surveys. The paper of Scott and Smith (1974) as well as Jones (1980) are also based on this criterion. These papers are discussed below.

We discuss Scott and Smith (1974) and Jones (1980) concurrently since as noted by Jones (1983:15) they differ primarily in the method of estimation of the covariance matrix, for the data used in formulating the linear model. Both methods generally deal with a minimisation problem involving h mean estimates, one from each time $t = 1, 2, \dots, h$, although as noted by Jones (1983:15) this "ignores the potential gains to be made from distinguishing matched and unmatched sample units". To be particular, if the sampling scheme is simple random sampling and if we use the least squares framework, Lemma 4.1 will be satisfied under quite general conditions when the distinction between matched and unmatched units is maintained, as for Blight and Scott (1973), but if the matching distinction is dropped then Lemma 4.1 will not be satisfied if there is sample overlap between time periods. For sampling schemes other than simple random sampling the situation is less clear, principally because schemes other than simple random sampling are often chosen as more efficient because sampling units do not have the property that $\epsilon(Y_i(t)) = \theta(t)$. The ξ -unbiasedness condition given by equation (5.2) or (5.5) is not then appropriate; exchangeability does not hold.

We consider the simple random sampling case in detail first. In this case the lnhssts estimator is given by equation (2.1), which we note is of the form $(a + \underline{\ell}^T \underline{y}_{(h)})$ for each time t , where $\underline{y}_{(h)}$ consists of all sampled responses to time h . We wish to choose a and $\underline{\ell}$ so that for each time t ,

$$E[\bar{Y}(t) - (a + \underline{\ell}^T \underline{y}_{(h)}) - E\{\bar{Y}(t) - (a + \underline{\ell}^T \underline{y}_{(h)})\}]^2$$

is minimised subject to an unbiasedness condition given by $E\{\bar{Y}(t) - (a + \underline{\ell}^T \underline{y}_{(h)})\} = 0$ or, since responses are exchangeable, by $E\{\bar{Y}(t) - (a + \underline{\ell}^T \underline{y}_{(h)})\} = 0$ as in equation (5.2) or (5.5). This is equivalent to minimisation of the ξ expectation of the p-mean square error of $\bar{Y}(t) - (a + \underline{\ell}^T \underline{y}_{(h)})$ subject to design unbiasedness. The mean vector $\underline{y} = \bar{\underline{y}} = (\bar{Y}(1), \bar{Y}(2), \dots, \bar{Y}(h))^T$ is here stochastic. With each component of $\underline{y}_{(h)}$ being a design unbiased estimator of its corresponding mean $\bar{Y}(t)$, a comparison with Rao (1973:234) shows that the problem as formulated is a problem in stochastic least squares. Using stochastic least squares instead of signal extraction given the semi-infinite past, is in fact Jones' (1980) modification of the technique of Scott and Smith (1974). As noted previously the particular formulation of Jones (1980) which uses only the h sample means $(1/n(t)) \sum_i Y_i(t)$ where the sum is over $i \in s(t)$ only, is exactly equivalent to the above stochastic least squares problem only if Lemma 4.1

The best method of fitting linear models to survey data selected by random sampling schemes other than simple random sampling has been a debated topic even in the context of a single survey (see, for example, Porter (1973), Du Mouchel and Duncan (1983)). The essential difficulty is that if individual respondent information is used and selection probabilities ignored then parameter estimates (e.g. means) will be biased when the selection probabilities are not independent of the model parameters, or, more precisely, when $\epsilon E(Y_i(t)) \neq \epsilon(\bar{Y}(t))$. Various ways around this problem while continuing to use individual information have been proposed. References have been given in Chapter 1. All these methods involve approximations to the "true" least squares estimate. The alternative procedure, again approximate, is to replace individual respondent information in the linear model with appropriate mean values (e.g. $\{\hat{Y}(t) : t = 1, 2, \dots, h\}$ or $\{\hat{Y}(t)', \hat{Y}(t)'', \hat{Y}(t)''' : i = 1, 2, 3, \dots, h\}$) which are p-unbiased estimates of the $\{\bar{Y}(t) : t = 1, 2, \dots, h\}$. Using such mean values in the linear model is equivalent to a minimisation of the lnhssts estimator given by equation 2.1, subject to p ξ -unbiasedness (see equation (3.10)) and certain specified dependencies between the weights $\{w_{it}(t')\}$ at each time t . Fixed relationships between the weights for all individuals sampled at time t' , for given t , gives the Scott and Smith (1974) type estimator, provided $w_{it}(t')/w_{it''}(t')$ is also constrained to be independent of i . Lemma 4.1 with the individual respondent information replaced by $\{\hat{Y}(t)', \hat{Y}(t)'', \hat{Y}(t)''' : t = 1, 2, \dots, h\}$ shows that using these matched and unmatched means will again be more efficient than using $\{\hat{Y}(t) : t = 1, 2, \dots, h\}$ as data, unless there is no matching or all the lag-correlations between matched means are zero, when efficiencies will be equal.

Although taking means of data is less efficient than using individual information, especially where signal to noise variance is small, this may be a necessary procedure to avoid bias problems for non-simple, random sample selection schemes. Using mean information is equivalent to using constrained weights for individual information within each such means when minimising mean square error. The Kalman Filter (and other general linear model formulations) require ad hoc approximations when non-simple, random sampling is used, and these present a greater hurdle than the inefficiency of general linear models using mean data. Despite efficiency losses, general linear models based on mean data remain more efficient than traditional sample survey estimators irrespective of the signal to noise variance ratio, and only small efficiency losses are likely where each block Σ_{tt} , of Σ of equation (4.1) is diagonally dominant with approximately equal diagonal elements within, but not necessarily between, blocks (i.e. times).

CHAPTER 4

RECURSIVE ESTIMATION OF THE GENERAL LINEAR MODEL FOR A PARAMETER VECTOR OF FIXED LENGTH

SUMMARY

The central part played by general linear models in repeated sample surveys is clear from Chapters 1 and 3. With an increasing number of time periods and more than one elementary estimate at each time period, updating and prediction formulae for general linear model parameter estimates provide computational advantages over re-estimation from scratch at each time period. Appropriate updating formulae have not been previously available however, for correlated data added more than one datum at a time, as would be the case for >1 elementary estimate added at each time point. The updating formulae for a parameter vector of fixed length, as for polynomial models in time, are developed in this chapter, and utilised in Chapter 6. The more general case where parameter vectors may increase in length with addition of data from new time points is a topic of Chapter 5. In the present chapter computational formulae for prediction, and recursive updating, of parameter estimates from general linear models with fixed length parameter vectors are derived, together with a sequence of univariate recursive residuals for testing the constancy of the regression relation over time. The results are applicable to any such linear model, not only to repeated sample survey data, and are discussed in this more general framework.

* * * * *

1. INTRODUCTION

For the linear regression model, recursive estimation is a technique for updating parameter estimates and the residual sum of squares given additional data. The technique was introduced by Plackett (1950), and has been shown by Brown et al (1975) to produce a set of independent residuals to enable testing for constancy of the regression model given the additional data.

The assumption of these two papers is that errors from the regression model are independent. An extension to correlated errors and one additional datum is due to McGilchrist and Sandland (1979).

In this chapter the results of McGilchrist and Sandland (1979) are extended to the case where more than one correlated datum are added simultaneously. These additional data may, for example, be observations on different individuals at the same time point or observations at different time points which become available simultaneously. While the requirement for a known error variance-covariance matrix is a restriction, in repeated applications to the same type of data good estimates are often available.

2. DEFINITIONS AND NOTATION

In the interest of clarity, the notation, as far as possible, parallels that of McGilchrist and Sandland (1979). Dimensions of matrices and vectors are given for the same reason.

Suppose that observations are said to belong to a particular grouping, t , when they are added simultaneously to the linear model. The symbol t may represent time. Let $n(t)$ be the sample size in grouping t . Assume the linear regression of $\underline{y}(t)$ the data in grouping t $((n(t) \times 1))$ on an $(n(t) \times p)$ matrix of independent variables, $\underline{x}_t$, is

$$\begin{array}{ccccccc} \underline{y}(t) & = & \underline{x}_t & \underline{\beta} & + & \underline{e}(t) \\ (n(t) \times 1) & & (n(t) \times p) & (p \times 1) & & (n(t) \times 1) \end{array}$$

where $\underline{\beta}$ is a parameter vector independent of t , and $\underline{e}(t)$ an error term. Suppose we have data up to grouping h .

(For repeated sample surveys, where elementary estimates are used as data, the 'sample size' is strictly speaking, in our previous notation, equal to K_t , the number of elementary estimates at time t . Within this chapter, we will continue to use $n(t)$ as the number of data in the linear model from time t , however.)

Then

$$\begin{array}{ccccccc} \underline{Y}_h & = & X_h & \underline{\beta} & + & \underline{e}_h \\ (N_h \times 1) & & (N_h \times p) & (p \times 1) & & (N_h \times 1) \end{array}$$

where

$$\underline{Y}_h = (\underline{Y}(1)^T, \underline{Y}(2)^T, \dots, \underline{Y}(h)^T)^T$$

$$X_h = (x_1^T \ x_2^T \ x_3^T \ \dots, \ x_h^T)^T$$

$$\underline{e}_h = (\underline{e}(1)^T, \underline{e}(2)^T, \dots, \underline{e}(h)^T)^T$$

with the variance of $\underline{e}_h$ equal to Σ_h and

$$N_h = \sum_{t=1}^h n(t)$$

Let

$$\Sigma_{h+1} = \begin{pmatrix} \Sigma_h & \Sigma_{h;h+1} \\ \Sigma_{h;h+1}^T & \Sigma_{h+1,h+1} \end{pmatrix}$$

$$X_{h+1} = \begin{pmatrix} X_h \\ x_{h+1} \end{pmatrix}$$

$$\underline{Y}_{h+1} = \begin{pmatrix} \underline{Y}_h \\ \underline{Y}_{(h+1)} \end{pmatrix}$$

Also define

$$\begin{aligned}
 (a) \quad G_{1h} &= \Sigma_h^{-1} \Sigma_{h;h+1} & (b) \quad Z_h &= x_{h+1}^T - X_h^T G_{1h} \\
 (N_h \times n(h+1)) & & (p \times n(h+1)) & \\
 (c) \quad G_{2h} &= \left(X_h^T \Sigma_h^{-1} X_h \right)^{-1} Z_h & (d) \quad C_{1h} &= \Sigma_{h+1,h+1} - \Sigma_{h;h+1}^T G_{1h} \\
 (p \times n(h+1)) & & (n(h+1) \times n(h+1)) & \\
 (e) \quad C_{2h} &= G_{2h}^T Z_h & (f) \quad R_{1h} &= \underline{y}^{(h+1)} - G_{1h}^T \underline{y}_h \\
 (n(h+1) \times n(h+1)) & & (n(h+1) \times 1) & \\
 (g) \quad R_{2h} &= G_{2h}^T X_h^T \Sigma_h^{-1} \underline{y}_h & (h) \quad W_h &= (C_{1h} + C_{2h})^{-\frac{1}{2}} (R_{1h} - R_{2h}) \\
 (n(h+1) \times 1) & & (n(h+1) \times 1) &
 \end{aligned}$$

Definitions (c), (e), (g); and (h) apply only for $\text{rank}(X_h) \geq p$.

The expectation operator used in this chapter is denoted E . For repeated sample surveys, the question whether such expectation is a design, a superpopulation, or a joint design-superpopulation expectation depends on the particular General Linear Model (see, for example, Chapters 1 and 3).

3. UPDATING FORMULAE AND RECURSIVE RESIDUALS

Theorem 1.

The best linear unbiased mean square error predictor of $\tilde{Y}_{(h+1)}^T$, for arbitrary $\tilde{Y}_{(h)}$, based on $\tilde{Y}_{(h)}$ is

$$\tilde{Y}_{(h+1)}^{\hat{(h)}} = \tilde{B}_{*}^T \tilde{Y}_{(h)}$$

$$\text{where } B_{*} = G_{1h} + \Sigma_h^{-1} X_h G_{2h} \\ (N_h \times n(h+1))$$

Proof:

Letting $\hat{Y}_{(h+1)}^{(h)} = B_{*}^T \tilde{Y}_{(h)}$ and choosing B to minimise

$$\tilde{E}[(Y_{(h+1)} - \hat{Y}_{(h+1)}^{(h)}) (Y_{(h+1)} - \hat{Y}_{(h+1)}^{(h)})^T] \quad \text{subject to}$$

$$\tilde{E}[Y_{(h+1)} - \hat{Y}_{(h+1)}^{(h)}] = 0 \quad \text{where } E \text{ denotes the expectation operator,}$$

yields the required result.

Note that

$$(G_{1h} + \Sigma_h^{-1} X_h G_{2h})^T \tilde{Y}_{(h)} = \Sigma_{h,h+1}^T \{ \Sigma_h^{-1} - \Sigma_h^{-1} X_h (X_h^T \Sigma_h^{-1} X_h)^{-1} X_h^T \Sigma_h^{-1} \} \tilde{Y}_{(h)} \\ + X_{h+1}^T (X_h^T \Sigma_h^{-1} X_h)^{-1} X_h^T \Sigma_h^{-1} \tilde{Y}_{(h)}$$

where the first term is the prediction of $\underline{e}(h+1)$, the errors for grouping $(h+1)$, based on past errors for groupings 1 through h and the known covariance matrix Σ_{h+1} , and the second term is $x_{h+1} \hat{\beta}_h$ where $\hat{\beta}_h$ is the best linear unbiased parameter estimate based on the combined data for groupings 1 to h .

i.e. $\hat{\beta}_h = (X_h^T \Sigma_h^{-1} X_h)^{-1} X_h^T \Sigma_h^{-1} Y_h$. The matrix

$\Sigma_h^{-1} - \Sigma_h^{-1} X_h (X_h^T \Sigma_h^{-1} X_h)^{-1} X_h^T \Sigma_h^{-1}$ is idempotent and a

projection operator.

Thus $\underline{y}^{(h)}(h+1) = x_{h+1} \hat{\beta}_h + \Sigma_{h;h+1}^T \Sigma_h^{-1} (\underline{y}_h - X_h \hat{\beta}_h)$

Corollary 1.1

$$(i) \quad E[(\underline{\hat{y}}^{(h)}(h+1) - \underline{y}(h+1))(\underline{\hat{y}}^{(h)}(h+1) - \underline{y}(h+1))^T] = C_{1h} + C_{2h}$$

$$(ii) \quad \underline{y}(h+1) - \underline{\hat{y}}^{(h)}(h+1) = \underline{R}_{1h} - \underline{R}_{2h} = (C_{1h} + C_{2h})^{\frac{1}{2}} \underline{w}_h$$

Proof: (i) follows on expansion, after some algebra, on

noting that

$$E[(\underline{\hat{y}}^{(h)}(h+1) - \underline{y}(h+1))(\underline{\hat{y}}^{(h)}(h+1) - \underline{y}(h+1))^T] = (B_{*}^T \begin{smallmatrix} \vdots \\ -I \end{smallmatrix}) E[\underline{y}_{h+1} \underline{y}_{h+1}^T] \begin{pmatrix} B_{*} \\ \vdots \\ -I \end{pmatrix}$$

and that

$$(B_{*}^T \begin{smallmatrix} \vdots \\ -I \end{smallmatrix}) E(\underline{y}_{h+1}) = 0$$

where 0 is a matrix of zero elements.

(ii) follows directly on noting that

$$\underline{y}_{(h+1)} - \underline{y}_{(h)}^{(h)} = - (B_*^T : -I) \underline{y}_{h+1}$$

See also Bibby & Toutenburg (1977: 84-88)

Theorem 2.

$$(i) \quad (X_{h+1}^T \Sigma_{h+1}^{-1} X_{h+1})^{-1} = (X_h^T \Sigma_h^{-1} X_h)^{-1} - G_{2h} (C_{1h} + C_{2h})^{-1} G_{2h}^T$$

$$(ii) \quad \Sigma_{h+1}^{-1} X_{h+1} = \begin{pmatrix} \Sigma_h^{-1} X_h \\ 0 \end{pmatrix} - \begin{pmatrix} G_{1h} \\ -I \end{pmatrix} C_{1h}^{-1} Z_h^T$$

$$(iii) \quad \underline{y}_{h+1}^T \Sigma_{h+1}^{-1} X_{h+1} = \underline{y}_h^T \Sigma_h^{-1} X_h + \underline{z}_{1h}^T C_{1h}^{-1} Z_h^T$$

Proof. (i) The result follows on noting that

$$\Sigma_{h+1}^{-1} = \begin{pmatrix} \Sigma_h^{-1} + G_{1h} C_{1h}^{-1} G_{1h}^T & -G_{1h} C_{1h}^{-1} \\ -C_{1h}^{-1} G_{1h}^T & C_{1h}^{-1} \end{pmatrix}$$

(see, for example Rao (1973:33), Example (2.7))

$$= \begin{pmatrix} \Sigma_h^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} G_{1h} \\ -I \end{pmatrix} C_{1h}^{-1} (G_{1h}^T : -I)$$

and that for A and D non-singular matrices of orders m and n and B m×n,

$$(A+BDB^T)^{-1} = A^{-1} - A^{-1} B (B^T A^{-1} B + D)^{-1} B^T A^{-1}$$

(see Rao (1973:33), Example (2.9)).

$$\text{Here } A = X_h^T \Sigma_h^{-1} X_h, B = Z_h, D = C_{1h}^{-1}$$

(ii) and (iii) follow directly.

Lemma 1.

For any two symmetric invertible matrices A and B of order k.

$$(i) \quad A^{-1} - (A+B)^{-1}BA^{-1} = (A+B)^{-1}$$

$$(ii) \quad A^{-1} - A^{-1}BA^{-1} + A^{-1}B(A+B)^{-1}BA^{-1} = (A+B)^{-1}$$

Proof: Follow from multiplying the left hand side of each of (i) and (ii) both on the left and on the right by (A+B) and showing that this is the identity matrix of order k in all cases.

Note that (i) implies $A^{-1} - A^{-1}B(A+B)^{-1} = (A+B)^{-1}$

Theorem 3.

Let $\hat{\beta}_{\sim h} = (X_h^T \Sigma_h^{-1} X_h)^{-1} X_h^T \Sigma_h^{-1} y_{\sim h}$ be the generalised least squares best linear unbiased estimator obtained from data from groupings 1 through h. A recursive estimator given data for grouping h+1, for $\hat{\beta}_{\sim h+1}$ is

$$\begin{aligned} \hat{\beta}_{\sim h+1} &= \hat{\beta}_{\sim h} + G_{2h} (C_{1h} + C_{2h})^{-1} (R_{1h} - R_{2h}) \\ &= \hat{\beta}_{\sim h} + G_{2h} (C_{1h} + C_{2h})^{-1} W_{\sim h} \end{aligned}$$

Proof: The proof follows from the use of Theorem 2 (i) and (iii) and Lemma 1 (i) in its transposed form.

Lemma 2.

If two matrices F and G are positive definite and of order k , and E is any conformable matrix then

- (i) $F+G$ is positive definite
- (ii) $E^T F E$ is positive definite
- (iii) F^{-1} is positive definite

Proof:

$$(i) \quad \underline{x}^T (F+G) \underline{x} = \underline{x}^T F \underline{x} + \underline{x}^T G \underline{x} > 0$$

$$(ii) \quad \underline{y}^T (E^T F E) \underline{y} = (E \underline{y})^T F (E \underline{y}) = \underline{x}^T F \underline{x} > 0$$

$$\text{for } \underline{x} = E \underline{y}$$

$$(iii) \quad \text{Set } E = E^T = F^{-1} \text{ in (ii) above.}$$

Theorem 4.

$$\text{If } S_h = \tilde{Y}_h^T \{ \Sigma_h^{-1} - \Sigma_h^{-1} X_h (X_h^T \Sigma_h^{-1} X_h)^{-1} X_h^T \Sigma_h^{-1} \} \tilde{Y}_h$$

is the residual sum of squares then

$$\begin{aligned} \text{(i)} \quad S_{h+1} &= S_h + \tilde{Y}_{h+1}^T F_{h+1} \tilde{Y}_{h+1} \\ &= S_h + (\tilde{R}_{1h} - \tilde{R}_{2h})^T (C_{1h} + C_{2h})^{-1} (\tilde{R}_{1h} - \tilde{R}_{2h}) \end{aligned}$$

where

$$F_{h+1} = \begin{pmatrix} B_* \\ -I \end{pmatrix} (C_{1h} + C_{2h})^{-1} (B_*^T, -I)$$

and B_* is defined in Theorem 1.

- (ii) if $\tilde{e}_h$ is distributed as $N(0, \Sigma_h)$ then $\tilde{Y}_t^T F_{t \sim t} \tilde{Y}_t$ are each independent chi-square random variables on $n(t)$ degrees of freedom for $t = K, K+1, \dots, h$ where K is the least value of t for which the (row) rank of $X_{K-1} \geq p$. The usual value of K is 2.
- (iii) if $\tilde{e}_h$ is distributed as $N(0, \Sigma_h)$ then the $n(t)$ components of $F_{t \sim t}^{1/2} \tilde{Y}_t$ are each distributed as i.i.d. $N(0, 1)$ random variables for $t = K, K+1, \dots, h$ where the independence of the components extends to independence of components belonging to different groupings, t , and where K is the least value of t for which the rank of $X_{K-1} \geq p$. Again the usual value of K is 2.

Proof:

(i) It can be shown using Theorem 2(i) and (iii) that

$$\tilde{y}_{h+1}^T \Sigma_{h+1}^{-1} \tilde{y}_{h+1} = \tilde{y}_h^T \Sigma_h^{-1} \tilde{y}_h + \tilde{R}_{1h} C_{1h}^{-1} \tilde{R}_{1h}$$

and that

$$\begin{aligned} & \tilde{y}_{h+1}^T \Sigma_{h+1}^{-1} X_{h+1} (X_{h+1}^T \Sigma_{h+1}^{-1} X_{h+1})^{-1} X_{h+1}^T \Sigma_{h+1}^{-1} \tilde{y}_{h+1} \\ &= \tilde{y}_h^T \Sigma_h^{-1} X_h (X_h^T \Sigma_h^{-1} X_h)^{-1} X_h^T \Sigma_h^{-1} \tilde{y}_h \\ &+ \tilde{R}_{1h}^T \{C_{1h}^{-1} C_{2h} C_{1h}^{-1} - C_{1h}^{-1} C_{2h} (C_{1h} + C_{2h})^{-1} C_{2h} C_{1h}^{-1}\} \tilde{R}_{1h} \\ &+ \tilde{R}_{1h}^T \{C_{1h}^{-1} - C_{1h}^{-1} C_{2h} (C_{1h} + C_{2h})^{-1}\} \tilde{R}_{2h} \\ &+ \tilde{R}_{2h}^T \{C_{1h}^{-1} - (C_{1h} + C_{2h})^{-1} C_{2h} C_{1h}^{-1}\} \tilde{R}_{1h} - \tilde{R}_{2h}^T (C_{1h} + C_{2h})^{-1} \tilde{R}_{2h} \end{aligned}$$

Thus using Lemma 1 (i) and (ii)

$$\begin{aligned} S_{h+1} &= S_h + (\tilde{R}_{1h} - \tilde{R}_{2h})^T (C_{1h} + C_{2h})^{-1} (\tilde{R}_{1h} - \tilde{R}_{2h}) \\ &= S_h + \tilde{y}_{h+1}^T F_{h+1} \tilde{y}_{h+1} \end{aligned}$$

(ii) Given Σ_{h+1} and hence Σ_h and $\Sigma_{h+1,h+1}$ are positive definite, C_{1h} and C_{2h} are positive definite, C_{1h} because it is the conditional covariance of $\tilde{Y}_{h+1}$ given $\tilde{Y}_h$, and C_{2h} by

Lemma 2 (i), (ii), and (iii). Therefore
by Lemma 2 (i), (ii) and (iii)

$$F_{h+1} = \begin{pmatrix} B_*^* \\ -I \end{pmatrix} (C_{1h} + C_{2h})^{-1} (B_*^T, -I)$$

is positive definite. Thus by the extended
Fisher-Cochran Theorem (Graybill 1961 Theorem 4.22),
 $Y_{h+1}^T F_{h+1} Y_{h+1}$ is distributed as χ^2 on $n(h+1)$ degrees
of freedom when $e_{h+1} \sim N(0, \Sigma_{h+1})$, since S_h and S_{h+1}
are distributed as χ^2 on N_h and N_{h+1} degrees of
freedom and $n(h+1) = N_{h+1} - N_h$.

(iii) Noting that

$$\begin{aligned} \text{var}(\tilde{R}_{1h} - \tilde{R}_{2h}) &= \text{var}[(B_*^T, -I) \tilde{Y}_{h+1}] \\ &= E[(\hat{\tilde{Y}}^{(h)}(h+1) - \tilde{Y}(h+1))(\hat{\tilde{Y}}^{(h)}(h+1) - \tilde{Y}(h+1))^T] \\ &\quad \text{by Theorem 1(i)} \\ &= C_{1h} + C_{2h} \quad \text{by Theorem 1(i)} \end{aligned}$$

it follows that

$$\text{var}[(C_{1h} + C_{2h})^{-1/2}(\tilde{R}_{1h} - \tilde{R}_{2h})] = I$$

and the $n(h+1)$ components of $(C_{1h} + C_{2h})^{-1/2}(\tilde{R}_{1h} - \tilde{R}_{2h})$
are thus i.i.d. If K is the least value of t for
which $\text{rank}(X_{K-1}) \geq p$ then these components are
also mutually independent for different groupings t ,

since the $(h-K+1)$ vectors constructed from these components for each grouping, namely $\tilde{W}_t$, for $t = K, K+1, \dots, h$ have the property that $\tilde{W}_t^T \tilde{W}_t$ have mutually independent chi-square distributions.

Note also that

$$\begin{aligned} & \tilde{Y}_{h+1}^T F_{h+1} \tilde{Y}_{h+1} \\ &= \tilde{Y}_{h+1}^T \begin{pmatrix} B_{*}^* \\ -I \end{pmatrix} \left[(B_{*}^T, -I) \begin{pmatrix} \Sigma_h & \Sigma_{h;h+1} \\ \Sigma_{h;h+1}^T & \Sigma_{h+1;h+1} \end{pmatrix} \begin{pmatrix} B_{*}^* \\ -I \end{pmatrix} \right]^{-1} (B_{*}^T, -I) \tilde{Y}_{h+1} \end{aligned}$$

The maximum possible number of such components can be generated by regrouping the data in grouping $K-1$ into two groups, the division being made following the p th linearly independent row vector of X_h .

4. CUSUM PROCEDURES

The components of $\tilde{W}_t$ for $t = K, K+1, \dots$ where K is the least value of t for which $\text{rank}(X_{K-1}) \geq p$, are independent both within and between different values of $t \geq K$, and provide the analogues of residuals transformed to independence used for cusum procedures in Brown et al (1975) and discussed McGilchrist and Sandland (1979). Such cusum procedures allow testing for departures from the model $\tilde{Y} = X\tilde{\beta} + \tilde{e}$ caused by $\tilde{\beta}$ changing between groupings or periods, t . Let W_j for $j = N_{K-1}+1, N_{K-1}+2, \dots, N_h$ denote the components of $\tilde{W}_t$ for $t = K, K+1, \dots, h$. Under the null hypothesis of constant $\tilde{\beta}$, Brown et al (1975) used a Wiener process to approximate the cumulative sums $\sum W_j$, summing over j from p to $N_h = n$ when adding only one datum at each time t . The probability that a Wiener process should go outside a family of straight line pairs was then used to test the null hypothesis. In addition, they used a cusum of squares test. McGilchrist and Sandland (1979) have noted that the results of Sections 2.3 and 2.4 of the paper of Brown et al (1975) apply without alteration using residuals modified in an analogous manner to the $\tilde{W}_t$ defined in Section 1 of this paper, provided the $\tilde{W}_t$ contain only one component.

For the present paper the cusum and the cusum of squares tests of Brown et al (1975) can also be used without modification except to note that the $n(t)$ such residuals derived simultaneously from grouping t are the components of $\tilde{W}_t$, not necessarily univariate. The range of t is again $t = K, K+1, \dots$ where K is the least value of t for which $\text{rank}(X_{K-1}) \geq p$.

Different sets of recursive residuals $\{\tilde{W}_t: t=K, K+1, \dots, h\}$ can be generated from the same data depending on the choice of $(C_{1t} + C_{2t})^{-1/2}$ of equation (h). $(C_{1t} + C_{2t})^{-1/2}$ can be any matrix A_t satisfying $A_t^T A_t = (C_{1t} + C_{2t})^{-1}$. Using the Choleski decomposition of $(C_{1t} + C_{2t})^{-1}$ yields the same recursive residuals as would be derived by adding the additional data one datum at a time and using the formulae of McGilchrist and Sandland (1979).

CHAPTER 5

UPDATING GENERAL LINEAR MODELS TO INCLUDE ADDITIONAL DATA AND/OR REGRESSORS

SUMMARY

Like Chapter 4, this chapter derives prediction and updating formulae for General Linear Models which are applicable beyond the sample survey context. These formulae further extend the results of Chapter 4. Specifically, updating formulae for minimum mean square linear parameter estimates and their mean square error, applicable where a general linear model is updated by the inclusion of additional regressors and/or data, and where the variance matrices for the model error and the parameter estimates are known and non-singular, are outlined. When variance matrices are estimated rather than known the equations derived give estimated generalised least squares estimators. The formulae are derived for both fixed and random parameter vectors. For models possible when additional data and regressors are added simultaneously, it is shown that a generalisation of the Kalman Filter, which allows for arbitrary forms of model error autocorrelation given non-singular error variance, is possible.

* * * * *

1. INTRODUCTION

Given the general linear model

$$\underline{y}_1^* = \underline{x}_{11}^* \underline{\beta}_1 + \underline{e}_1^* \quad \text{with} \quad \text{var}(\underline{e}_1^*) = \Sigma_{11}^* \quad (1.1)$$

where $\underline{y}_1^*$ is the data to which an initial model is being fitted by least squares, $\underline{x}_{11}^*$ is a matrix of independent variables, $\underline{\beta}_1$ is a fixed parameter vector to be estimated, and $\underline{e}_1^*$ an error term with dispersion matrix Σ_{11}^* , then the minimum variance or best, linear unbiased estimator of $\underline{\beta}_1$ is

$$\underline{\beta}_1^* = \left(\underline{x}_{11}^{*T} \Sigma_{11}^{*-1} \underline{x}_{11}^* \right)^{-1} \underline{x}_{11}^{*T} \Sigma_{11}^{*-1} \underline{y}_1^* \quad (1.2)$$

provided Σ_{11}^{*-1} and $\left(\underline{x}_{11}^{*T} \Sigma_{11}^{*-1} \underline{x}_{11}^* \right)^{-1}$ exist.

Suppose instead $\underline{\beta}_1$ is stochastic with $\epsilon(\underline{\beta}_1) = \underline{\mu}_1$ known, and the stochastic structure of $\underline{\beta}_1$

$$\underline{\beta}_1 = \underline{\mu}_1 + \underline{u}_1 \quad \text{with} \quad \text{var}(\underline{u}_1) = \Sigma_{\underline{\beta}_1}, \quad (1.3)$$

$\Sigma_{\underline{\beta}_1}$ non-singular, is included within a condensed model

$$\begin{bmatrix} \underline{\mu} \\ \underline{y}_1^* \end{bmatrix} = \begin{bmatrix} \underline{I} \\ \underline{x}_{11}^* \end{bmatrix} \underline{\beta}_1 + \begin{bmatrix} -\underline{u} \\ \underline{e}_1^* \end{bmatrix} \quad (1.4)$$

$$\text{with} \quad \text{var} \begin{bmatrix} -\underline{u} \\ \underline{e}_1^* \end{bmatrix} = \begin{bmatrix} \Sigma_{\underline{\beta}_1} & 0 \\ 0 & \Sigma_{11}^* \end{bmatrix}$$

i.e.,

$$\underline{\tilde{y}}_1 = \underline{\tilde{x}}_{11} \underline{\beta}_1 + \underline{\tilde{e}}_1 \quad \text{with} \quad \text{var}(\underline{\tilde{e}}_1) = \underline{\tilde{\Sigma}}_{11} \text{ say} \quad (1.5)$$

(For repeated sample surveys, ϵ denotes superpopulation expectation, but var now denotes joint superpopulation-design variance.)

Then Duncan and Horn (1972 : Theorem 2.3) have shown that the usual minimum mean square linear estimator (mmsle) of β_1

$$\beta_1^+ = \left(\Sigma_{\beta_1}^{-1} + X_{11}^{*T} \Sigma_{11}^{*-1} X_{11}^* \right)^{-1} \left(\Sigma_{\beta_1}^{-1} \mu + X_{11}^{*T} \Sigma_{11}^{*-1} Y_1^* \right) \quad (1.6)$$

is given by

$$\tilde{\beta}_1 = \left(\tilde{X}_{11}^T \tilde{\Sigma}_{11}^{-1} \tilde{X}_{11} \right)^{-1} \tilde{X}_{11}^T \tilde{\Sigma}_{11}^{-1} \tilde{Y}_1 \quad (1.7)$$

For stochastic β_1 , and unknown μ , the mmsle of β_1 is again given by equation (1.2) (Rao 1973 : 234).

The similarity of equations (1.1) and (1.5), and of equations (1.2) and (1.7) is immediately apparent and allow a statement of the problem which incorporates both fixed and random parameter vectors, β_1 . Both β_1^* and $\tilde{\beta}_1$ of equations (1.2) and (1.7) give mmsle's even for β_1 of equation (1.1) fixed.

In what follows below we discuss the estimator $\hat{\beta}_1$ of β_1 and its consequent updating formulae, where $\hat{\beta}_1 = \beta_1^*$ when β_1 is fixed, or random with unknown mean, and $\hat{\beta}_1 = \tilde{\beta}_1$ when β_1 is random with known mean. The model is then

$$Y_1 = X_{11}\beta_1 + e_1 \quad \text{with} \quad \text{var}(e_1) = \Sigma_{11} \quad (1.8)$$

as given by equation (1.1) and (1.5) as appropriate with mmsle,

$$\hat{\beta}_1 = \left(X_{11}^T \Sigma_{11}^{-1} X_{11} \right)^{-1} X_{11}^T \Sigma_{11}^{-1} Y_1 \quad (1.9)$$

given respectively by equation (1.2) or (1.7). A parallel translation process must be applied to the updating formulae when given additional data, regressors, parameters or their covariances.

We note further that $\text{var}(\hat{\beta}_1 - \beta_1)$ equals the mean square error of $(\hat{\beta}_1 - \beta_1)$ with

$$\text{var}(\hat{\beta}_1 - \beta_1) = \left(X_{11}^T \Sigma_{11}^{-1} X_{11} \right)^{-1} \quad (1.10)$$

and

$$\begin{aligned} \text{RSS}_1 &= \underline{Y}_1^T \left(\Sigma_{11}^{-1} - \Sigma_{11}^{-1} X_{11} \left(X_{11}^T \Sigma_{11}^{-1} X_{11} \right)^{-1} X_{11}^T \Sigma_{11}^{-1} \right) \underline{Y}_1 \quad (1.11) \\ &= \underline{Y}_1^T R_{11} \underline{Y}_1, \quad \text{say} \end{aligned}$$

where $\text{var}(\hat{\beta}_1 - \beta_1) = \text{var} \hat{\beta}_1$ in the fixed parameter case, and RSS_1 denotes the residual sum of squares for the mmsle solution to equation (1.8). The proofs in the random parameter case are given in Duncan and Horn (1972 : Lemma 2.4 and 2.6).

Given these preliminaries we may distinguish three cases, given additional regressors and/or data

Model 1

$$\underline{Y}_1 = (X_{11} \ X_{12}) \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} + \underline{e}_2 \quad (1.12)$$

$$\text{with } \text{var}(\underline{e}_1) = \text{var}(\underline{e}_2) = \Sigma_{11}$$

Model 2

$$\begin{pmatrix} \underline{Y}_1 \\ \underline{Y}_2 \end{pmatrix} = \begin{pmatrix} X_{11} \\ X_{12} \end{pmatrix} \beta_1 + \begin{pmatrix} \underline{e}_1 \\ \underline{e}_2 \end{pmatrix} \quad (1.13)$$

$$\text{with } \text{var} \begin{pmatrix} \underline{e}_1 \\ \underline{e}_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Model 3

$$\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} + \begin{pmatrix} \tilde{e}_1 \\ \tilde{e}_2 \end{pmatrix} \quad (1.14)$$

$$\text{with } \text{var} \begin{pmatrix} \tilde{e}_1 \\ \tilde{e}_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

(a) $X_{12} \neq 0$

(b) $X_{12} = 0$

Note that while $\Sigma_{12} = \Sigma_{21}^T$, $X_{12} \neq X_{21}^T$ in general. In subsequent sections we will denote the mmsle of $\tilde{\beta} = \tilde{\beta}_1$ (Model 2) or

$\tilde{\beta} = (\tilde{\beta}_1^T \tilde{\beta}_2^T)^T$ (Models 1 and 3) of equations (1.12) to (1.14) by

$\tilde{\beta} = \tilde{\beta}_1$ or $\tilde{\beta} = (\tilde{\beta}_1^T \tilde{\beta}_2^T)^T$ as appropriate.

2. MODEL 1 : ADDITIONAL REGRESSORS

The solution to the problem of incorporating additional regressors into the model of equation (1.8) to give the model

$$\tilde{y}_1 = (X_{11} \ X_{12}) \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} + \tilde{e}_2 \quad (2.1)$$

with $\text{var}(\tilde{e}_1) = \text{var}(\tilde{e}_2) = \Sigma_{11}$, was given by Cochran (1938) for

$\Sigma_{11} = \sigma^2 I$ and $\tilde{\beta}_2 = \beta_2$, a scalar. Quenouille (1950) extended

the argument to the case $\Sigma_{11} = \sigma^2 I$ and $\tilde{\beta}_2$ a vector.

Quenouille's (1950) results are easily modified to allow for more general error covariances, simply by premultiplying equation (2.1)

by $\Sigma_{11}^{-1/2}$ in which case the transformed error covariance structure

becomes I the identity matrix, and redefinition of

$\Sigma_{11}^{-1/2} \underline{y}_1$, $\Sigma_{11}^{-1/2} (X_{11}; X_{12})$ and $\Sigma_{11}^{-1/2} \underline{e}_2$ returns us to

Quenouille's (1950) model. The fact that both Cochran (1938), and Quenouille's (1950) results were derived for (β_1^T, β_2^T) fixed, presents no bar to their application to random (β_1^T, β_2^T) , with the provisos outlined in Section 1.

Theorem 2.1

Given the model of equation (1.8) with mmsle $\hat{\beta}$ given by equation (1.9), the model of equation (2.1) has mmsle

$\tilde{\beta} = (\tilde{\beta}_1^T, \tilde{\beta}_2^T)^T$ given by

$$\tilde{\beta}_1 = \hat{\beta}_1 - (X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1} X_{11}^T \Sigma_{11}^{-1} X_{12} \tilde{\beta}_2 \quad (2.2)$$

$$\tilde{\beta}_2 = (X_{12}^T R_{11} X_{12})^{-1} X_{12}^T R_{11} \underline{y}_1$$

residual sum of squares RSS_2 given by

$$RSS_2 = RSS_1 - \tilde{\beta}_2^T X_{12}^T R_{11} \underline{y}_1 \quad (2.3)$$

and minimum mean square error

$$\text{var} \begin{pmatrix} \tilde{\beta}_1 - \beta_1 \\ \tilde{\beta}_2 - \beta_2 \end{pmatrix} = \begin{bmatrix} (X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1} + L_{21}^T M_{22}^{-1} L_{21} & -L_{21}^T M_{22}^{-1} \\ -M_{22}^{-1} L_{21} & M_{22}^{-1} \end{bmatrix} \quad (2.4)$$

where

$$\begin{aligned} R_{11} &= \Sigma_{11}^{-1} - \Sigma_{11}^{-1} X_{11} (X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1} X_{11}^T \Sigma_{11}^{-1} \\ L_{21} &= X_{12}^T \Sigma_{11}^{-1} X_{11} (X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1} \\ M_{22} &= X_{12}^T R_{11} X_{12} \end{aligned} \quad (2.5)$$

Proof:

The proof for the fixed parameter case is contained in Quenouille (1950). The extension to random parameter vectors involves no complication. See also Seber (1977 : 65-67).

Note that since Σ_{11}^{-1} , $(X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1}$ and R_{11} are known from the initial model fit, determination of $\tilde{\beta}$, RSS_2 and $\text{var}(\tilde{\beta} - \beta)$ involve matrix inversion of maximum order p_2 where β_2 is $(p_2 \times 1)$. Further,

$$\text{var} \begin{pmatrix} \tilde{\beta}_1 - \beta_1 \\ \tilde{\beta}_2 - \beta_2 \end{pmatrix} = \begin{pmatrix} (X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1} & 0 \\ 0 & 0 \end{pmatrix} \quad (2.6)$$

$$+ \begin{pmatrix} L_{21}^T \\ -I \end{pmatrix} M_{22}^{-1} (L_{21} \quad -I)$$

and

$$\begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ 0 \end{pmatrix} - \begin{pmatrix} L_{21} \\ -I \end{pmatrix} M_{22}^{-1} X_{12}^T R_{11} \tilde{y}_1$$

i.e.,

$$\begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ 0 \end{pmatrix} + \begin{pmatrix} L_{21} \\ -I \end{pmatrix} H_{21} \tilde{y}_1 \quad (2.7)$$

where $H_{21} = -M_{22}^{-1} X_{12}^T R_{11}$

3. MODEL 2 : ADDITIONAL DATA

When additional data are incorporated into the linear model of equation (1.8), then with the exception of premultiplication of the linear model for the combined data by the Choleski decomposition of the inverse of their covariance matrix, no transformation exists which will transform the correlated error case into the uncorrelated error case. The uncorrelated error case is discussed in Seber (1977 : 336-338). The correlated error case for a single additional datum without use of the Choleski decomposition is outlined in McGilchrist and Sandland (1979). The correlated error case for more than one additional datum added simultaneously is contained in Haslett (1985)(b) or in Chapter 4.

We summarise the results for the correlated error case and more than one additional data. The additional data give the updated linear model:

$$\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix} = \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} \tilde{\beta}_1 + \begin{pmatrix} \tilde{e}_1 \\ \tilde{e}_2 \end{pmatrix} \quad (3.1)$$

with

$$\text{var} \begin{pmatrix} \tilde{e}_1 \\ \tilde{e}_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Theorem 3.1

Given the model of equation (1.8) with mmsle given by equation (1.9), the model of equation (3.1) has mmsle

$$\tilde{\beta}_1 = \hat{\beta}_1 + G_B(C_A + C_B)^{-1}(\tilde{R}_A - \tilde{R}_B) \quad (3.2)$$

residual sum of squares RSS_2 given by

$$RSS_2 = RSS_1 + (\tilde{R}_A - \tilde{R}_B)^T (C_A + C_B)^{-1} (\tilde{R}_A - \tilde{R}_B) \quad (3.3)$$

and minimum mean square error

$$\text{var}(\tilde{\beta}_1 - \beta_1) = \text{var}(\hat{\beta}_1 - \beta_1) - G_B (C_A + C_B)^{-1} G_B^T \quad (3.4)$$

where

$$\begin{aligned} (a) \quad \tilde{R}_A &= Y_2 - G_A^T Y_1 \\ (b) \quad \tilde{R}_B &= G_B^T X_{11}^T \Sigma_{11}^{-1} Y_1 \\ (c) \quad C_A &= \Sigma_{22} - \Sigma_{21} G_A \\ (d) \quad C_B &= G_B^T Z_A \\ (e) \quad G_A &= \Sigma_{11}^{-1} \Sigma_{12} \\ (f) \quad G_B &= (X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1} Z_A \\ (g) \quad Z_A &= X_{21}^T - X_{11}^T G_A \end{aligned} \quad (3.5)$$

Proof:

See Chapter 4, Theorem 2(i), 3 and 4(i).

.

Using the model (3.1) it is possible to generate a series of uncorrelated recursive residuals given by the components of

$$\tilde{w}_2 = (C_A + C_B)^{-1/2} (\tilde{R}_A - \tilde{R}_B). \text{ These can be used to test whether}$$

the realisation of $\tilde{\beta}_1$ is in fact fixed (i.e., there has been only one realisation of $\tilde{\beta}_1$), via the use of cusum and cusum of squares tests. (Brown, Durbin and Evans (1975)).

The properties of the recursive residuals are discussed in Chapter 4 and the extension to random β_1 is without difficulty provided the test statistics derived from the recursive residuals are used to test for a single realisation of β_1 .

When $(\tilde{y}_1^T \tilde{y}_2^T)$ is $1 \times n$, and β_1 random with given mean μ , it is possible to generate n such recursive residuals. When β_1 ($p \times 1$) is fixed, or random with unknown mean, a maximum possible number recursive residuals is $n - p$.

4. MODEL 3(a) : ADDITIONAL DATA AND REGRESSORS, $X_{12} \neq 0$

Starting with the initial model given by equation (1.8) the addition of extra data and regressors for the correlated error case gives

$$\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \quad (4.1)$$

$$\text{with var} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

We assume in this section that $X_{12} \neq 0$. More particularly, except in the final paragraph of this section, we consider the case where $X_{12}^T \Sigma_{11}^{-1} X_{12}$ is of full rank. The case of $X_{12} = 0$ is discussed in Section 5. For notational convenience equation (4.1) may be written

$$\tilde{y} = X\beta + e \quad \text{with} \quad \text{var}(e) = \Sigma \quad (4.2)$$

We aim to determine $\tilde{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \underline{y}$, the residual sum of squares associated with the least squares solution $\tilde{\beta}$, and the mean square error of $\tilde{\beta}$, each in terms the corresponding estimates or parameters of the model of equation (1.8) plus some additional terms. Sandland (1980) considers this model with $\underline{y}_2 = Y_2$, $\underline{\beta}_2 = \beta_2$ but adds the additional regressor and datum in separate steps.

Define

$$\begin{aligned}
 (a) \quad G_A &= \Sigma_{11}^{-1} \Sigma_{12} & (n) \quad H_C &= C_P^{-1} (G_P^T X_{11}^T - X_{12}^T) \Sigma_{11}^{-1} \\
 (b) \quad C_A &= \Sigma_{22} - \Sigma_{21} G_A & (o) \quad \underline{R}_A &= \underline{y}_2 - G_A^T \underline{y}_1 \\
 (c) \quad Z_A &= X_{21}^T - X_{11}^T G_A & (p) \quad \underline{R}_D &= (H_A^T X_{11}^T - H_B^T X_{12}^T) \Sigma_{11}^{-1} \underline{y}_1 \\
 (d) \quad Z_B &= X_{22}^T - X_{12}^T G_A \\
 (e) \quad P_{11} &= X_{11}^T \Sigma_{11}^{-1} X_{11} \\
 (f) \quad P_{12} &= X_{11}^T \Sigma_{11}^{-1} X_{12} & (4.3) \\
 (g) \quad P_{21} &= X_{12}^T \Sigma_{11}^{-1} X_{11} \\
 (h) \quad P_{22} &= X_{12}^T \Sigma_{11}^{-1} X_{12} \\
 (i) \quad C_D &= \begin{pmatrix} Z_A^T & Z_B^T \end{pmatrix} \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}^{-1} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix} \\
 (j) \quad G_P &= P_{11}^{-1} P_{12} \\
 (k) \quad C_P &= P_{22} - P_{21} P_{11}^{-1} P_{12} \\
 (l) \quad H_A &= (P_{11}^{-1} + G_P C_P^{-1} G_P^T) Z_A - G_P C_P^{-1} Z_B \\
 (m) \quad H_B &= C_P^{-1} (G_P^T Z_A - Z_B)
 \end{aligned}$$

Note particularly that

$$C_D = \begin{pmatrix} Z_A^T & Z_B^T \end{pmatrix} \begin{pmatrix} P_{11}^{-1} + G_P C_P^{-1} G_P^T & -G_P C_P^{-1} \\ -C_P^{-1} G_P^T & C_P^{-1} \end{pmatrix} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix}$$

via Rao (1973 : 33) Example 2.7

$$= H_A^T Z_A - H_B^T Z_B.$$

Further, given that Σ_{11}^{-1} , $(X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1}$, $X_{11}^T \Sigma_{11}^{-1}$ and

$X_{11}^T \Sigma_{11}^{-1} \tilde{y}_1$ are retained from the least squares solution of

equation (1.8), no matrix inversions of order greater than

n_2 (where $\tilde{y}_2$ is $(n_2 \times 1)$) are required in forming the terms

of equations (4.3) (a) - (p).

Theorem 4.1

$$\begin{aligned} \text{(i)} \quad (X^T \Sigma^{-1} X)^{-1} &= \begin{pmatrix} (X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1} & 0 \\ 0 & 0 \end{pmatrix} \\ &+ \begin{pmatrix} G_P \\ -I \end{pmatrix} C_P^{-1} (G_P^T \mid -I) \\ &- \begin{pmatrix} H_A \\ -H_B \end{pmatrix} (C_A + C_D)^{-1} (H_A \mid -H_B) \end{aligned}$$

$$\text{(ii)} \quad \Sigma^{-1} X = \begin{pmatrix} \Sigma_{11}^{-1} X_{11} & \Sigma_{11}^{-1} X_{12} \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} G_A \\ -I \end{pmatrix} C_A^{-1} (Z_A^T \mid Z_B^T)$$

$$\text{(iii)} \quad \tilde{y}^T \Sigma^{-1} X = \tilde{y}_1^T \Sigma_{11}^{-1} (X_{11} \mid X_{12}) + \tilde{y}_A^T C_A^{-1} (Z_A^T \mid Z_B^T)$$

Proof:

(i) The result follows on noting that

$$\Sigma^{-1} = \left(\begin{array}{c|c} \Sigma_{11}^{-1} + G_A C_A^{-1} G_A^T & -G_A C_A^{-1} \\ \hline -C_A^{-1} G_A^T & C_A^{-1} \end{array} \right) \quad (4.4)$$

see for example Rao (1973 : 33) Example 2.7

$$= \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} G_A \\ -I \end{pmatrix} C_A^{-1} (G_A^T \quad -I)$$

and thus

$$X^T \Sigma^{-1} X = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} + \begin{pmatrix} Z_A \\ Z_B \end{pmatrix} C_A^{-1} (Z_A^T \quad Z_B^T) \quad (4.5)$$

Further for A and D nonsingular matrices of orders m and n say, and B m × n

$$(A + BDB^T)^{-1} = A^{-1} - A^{-1} B (B^T A^{-1} B + D^{-1})^{-1} B^T A^{-1} \quad (4.6)$$

(see for example Rao (1973 : 33) Example 2.9)

$$\text{Here } A = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \quad B = \begin{pmatrix} Z_A \\ Z_B \end{pmatrix}, \quad D = C_A^{-1}$$

The result follows on noting the definitions of

C_D , H_A and H_B , and expanding A^{-1} in terms of

Rao (1973 : 33) Example 2.7 as illustrated above for Σ^{-1}

(ii) and (iii) follow directly using the above formula for Σ^{-1} and suitably partitioning $\underline{y}$ and $\underline{x}$.

Corollary 4.1

$$\text{var}(\tilde{\underline{\beta}} - \underline{\beta}) = \text{var} \begin{pmatrix} \tilde{\underline{\beta}}_1 - \underline{\beta}_1 \\ \tilde{\underline{\beta}}_2 - \underline{\beta}_2 \end{pmatrix} \quad (4.7)$$

$$= \begin{pmatrix} \text{var}(\hat{\underline{\beta}}_1 - \underline{\beta}_1) & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \underline{G}_P \\ -\underline{I} \end{pmatrix} \underline{C}_P^{-1} (\underline{G}_P^T \underline{I} - \underline{I})$$

$$- \begin{pmatrix} \underline{H}_A \\ -\underline{H}_B \end{pmatrix} (\underline{C}_A + \underline{C}_D)^{-1} (\underline{H}_A^T \underline{I} - \underline{H}_B^T)$$

Proof:

The result follows directly from equation (1.10) on noting that

$$\text{var}(\hat{\underline{\beta}} - \underline{\beta}) = (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1}.$$

A crucial point in the proof of Theorem 4.1(i) is that $\underline{X}_{12} \neq 0$.

For if $\underline{X}_{12} = 0$ then $\underline{P}_{12} = 0$, $\underline{P}_{21} = 0$ and $\underline{P}_{22} = 0$ and thus

$$\underline{A} = \begin{pmatrix} \underline{P}_{11} & \underline{P}_{12} \\ \underline{P}_{21} & \underline{P}_{22} \end{pmatrix} \quad \text{is singular.}$$

Theorem 4.2

The mmsle of β is given by

$$\begin{aligned} \tilde{\beta} &= \begin{pmatrix} \hat{\beta}_1 \\ 0 \end{pmatrix} + \begin{pmatrix} G_P \\ -I \end{pmatrix} H_C \tilde{y}_1 \\ &+ \begin{pmatrix} H_A \\ -H_B \end{pmatrix} (C_A + C_D)^{-1} (\tilde{R}_A - \tilde{R}_B) \end{aligned} \quad (4.8)$$

Proof:

The result follows by substitution from Theorem 4.1 (i) and (ii) on noting that

$$\begin{aligned} \tilde{\beta} &= (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \tilde{y} \\ &= \begin{pmatrix} P_A^{-1} Z_A C_A^{-1} \tilde{R}_A \\ 0 \end{pmatrix} + \begin{pmatrix} G_P C_P^{-1} (G_P^T Z_A - Z_B) C_A^{-1} \tilde{R}_A \\ -C_P^{-1} (G_P^T Z_A - Z_B) C_A^{-1} \tilde{R}_A \end{pmatrix} = \begin{pmatrix} H_A C_A^{-1} \tilde{R}_A \\ -H_B C_A^{-1} \tilde{R}_A \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} & - \begin{pmatrix} H_A \\ -H_B \end{pmatrix} (C_A + C_D)^{-1} (H_A^T X_{11}^T - H_B^T X_{12}^T) \Sigma_{11}^{-1} \tilde{y}_1 \\ & + \begin{pmatrix} H_A \\ -H_B \end{pmatrix} \left[I - (C_A + C_D)^{-1} (H_A^T Z_A - H_B^T Z_B) \right] C_A^{-1} \tilde{R}_A \\ & = \begin{pmatrix} H_A \\ -H_B \end{pmatrix} (C_A + C_D)^{-1} (\tilde{R}_A - \tilde{R}_B) \end{aligned}$$

An updating formula for the residual sum of squares for the least squares fit to equation (4.1), RSS, in terms of the residual sum of squares derivative from equation (1.8) is also possible, and is given in Theorem 4.3 both for completeness, and to exhibit its structure. Computationally however the better procedure would calculate RSS directly from the terms in Theorem 4.1 on noting that

$$RSS = \tilde{Y}^T (\Sigma^{-1} - \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}) \tilde{Y} \quad (4.9)$$

$$= \tilde{Y}^T \Sigma^{-1} \tilde{Y} - \tilde{Y}^T \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \tilde{Y}$$

and that updating formulae for Σ^{-1} , $\tilde{Y}^T \Sigma^{-1} \tilde{Y}$, $(X^T \Sigma^{-1} X)^{-1}$ are given by Theorem 4.1.

For notational simplicity we define some further terms additional to those of equation (4.3).

$$\begin{aligned} (a) \quad H_E &= C_A^{-1} - C_A^{-1} Z_A^T P_{11}^{-1} Z_A C_A^{-1} \\ (b) \quad H_D &= (G_P^T X_{11}^T - X_{12}^T) \Sigma_{11}^{-1} \\ (c) \quad H_P &= P_{11}^{-1} X_{11}^T \Sigma_{11}^{-1} \\ (d) \quad C_C &= (H_A X_{11}^T - H_B X_{12}^T) \Sigma_{11}^{-1} \\ (e) \quad C_H &= H_P^T Z_A + H_D^T H_B \quad (4.10) \\ (f) \quad C_I &= H_B C_A^{-1} \\ (g) \quad C_F &= C_D C_A^{-1} \\ (h) \quad C_L &= C_P^{-1} H_D + C_I G_A^T \\ (i) \quad C_M &= C_C + C_F G_A^T \end{aligned}$$

Note that C_F is symmetric.

We then have the following theorem:

Theorem 4.3

The residual sum of squares RSS, for the least squares fit to equation (4.1) is given in terms of the residual sum of squares RSS₁, of the initial model of equation (1.8) as

$$RSS = RSS_1 \quad (4.11)$$

$$\begin{aligned} & + (\tilde{Y}_1^T \tilde{Y}_2^T) \left[\begin{pmatrix} G_A \\ -I \end{pmatrix} H_E (G_A^T \mid -I) - \begin{pmatrix} C_L^T \\ -C_E^T \end{pmatrix} C_P (C_L \mid -C_E) \right. \\ & + \begin{pmatrix} C_M^T \\ -C_F \end{pmatrix} (C_A + C_D)^{-1} (C_M \mid -C_F) \\ & \left. + \begin{pmatrix} C_D \\ 0 \end{pmatrix} C_A^{-1} (G_A^T \mid -I) + \begin{pmatrix} G_A \\ -I \end{pmatrix} C_A^{-1} (C_D^T \mid 0) \right] \begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{pmatrix} \end{aligned}$$

Proof:

The proof proceeds from equation (4.9) and the results of Theorem (4.1), by suitable collection of terms. In this connection, note that

$$\begin{aligned} (a) \quad & \begin{pmatrix} G_A \\ -I \end{pmatrix} C_A^{-1} (G_A^T \mid -I) - \begin{pmatrix} G_A \\ -I \end{pmatrix} C_A^{-1} Z_A^T P_{11}^{-1} Z_A C_A^{-1} (G_A^T \mid -I) \\ & = \begin{pmatrix} G_A \\ -I \end{pmatrix} H_C (G_A^T \mid -I) \end{aligned}$$

$$\begin{aligned} (b) \quad & \begin{pmatrix} H_D^T \\ 0 \end{pmatrix} C_P^{-1} (H_D \mid 0) + \begin{pmatrix} G_A \\ -I \end{pmatrix} C_A^{-1} H_B^T C_P H_B C_A^{-1} (G_A^T \mid -I) \\ & = \begin{pmatrix} C_L^T \\ -C_I^T \end{pmatrix} C_P (C_L \mid -C_I) \end{aligned}$$

$$\begin{aligned}
 (c) \quad & \begin{pmatrix} C_A^T \\ C_C \\ 0 \end{pmatrix} (C_A + C_D)^{-1} (C_C \vdots 0) - \begin{pmatrix} C_A^T \\ C_C \\ 0 \end{pmatrix} (C_A + C_D)^{-1} (G_A^T \vdots \\
 & - \begin{pmatrix} G_A \\ -I \end{pmatrix} C_A^{-1} C_D (C_A + C_D)^{-1} (C_C \vdots 0) + \begin{pmatrix} G_A \\ -I \end{pmatrix} C_A^{-1} C_D (C_A + \\
 & = \begin{pmatrix} C_M^T \\ -C_F \end{pmatrix} (C_A + C_D)^{-1} (C_M \vdots -C_F).
 \end{aligned}$$

$$(d) \quad \begin{pmatrix} H_P^T Z_A \\ 0 \end{pmatrix} C_A^{-1} (G_A^T \vdots -I) + \begin{pmatrix} H_D^T \\ 0 \end{pmatrix} H_B C_A^{-1} (G_A^T \vdots -I) = \begin{pmatrix} C_H \\ 0 \end{pmatrix}$$

There may be situations in which $X_{12}^T \Sigma_{11}^{-1} X_{12}$ is not of rank, but for which $X_{12} \neq 0$. Since Σ_{11}^{-1} is positive definite, this can only occur when at least one of the columns of X_{12} is a combination of the others. It is then possible to temporarily redefine β_2 (together with X_{12} and X_{22}) so that each such column of X_{12} is the zero vector. Determining the mmsle to include additional data can then proceed in two steps. Firstly, the parameters for which corresponding columns of X_{12} are non-zero are fitted to the initial data only, via the methods of Section 2. Secondly this augmented model is taken as the initial model and itself redefined in terms of equation (1.8). X_{12} thus redefined is the zero matrix in the model of equation (4.1). The mmsle of equation (4.2) can then be determined by the method of the previous section since $X_{12} = 0$, and the original β recovered from the temporary redefinition.

5. MODEL 3(b) : ADDITIONAL DATA AND REGRESSORS, $X_{12} = 0$

As in Section 4 we start with the initial model given by equation (1.8) and the addition of extra data and regressors for the correlated case gives:

$$\begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} + \begin{pmatrix} \tilde{e}_1 \\ \tilde{e}_2 \end{pmatrix} \quad (5.1)$$

$$\text{with } \text{var} \begin{pmatrix} \tilde{e}_1 \\ \tilde{e}_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

or

$$\tilde{Y} = X\tilde{\beta} + \tilde{e} \quad \text{with } \text{var}(\tilde{e}) = \Sigma. \quad (5.2)$$

Equations (5.1) and (5.2) differ from equations (4.1) and (4.2) respectively only in that $X_{12} = 0$ here. Again we aim to determine $\hat{\tilde{\beta}} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \tilde{Y}$, and the mean square error of $\tilde{\beta}$ each in terms of the corresponding term from the model of equation (1.8) plus some additional terms. The updating formula for the residual sum of squares is not given here explicitly. Its general form is given in Theorem 4.3, equation (4.11) and appropriate replacement of terms from equations (4.3) and (4.10) with the parallel terms for $X_{12} = 0$ (see particularly equation (5.3) below) will yield the correct formula, if required.

Define G_A , C_A , Z_A , Z_B , P_{11} , R_A as in equation (4.3) noting that since $X_{12} = 0$, $Z_B = X_{22}^T$.

Define:

$$\begin{aligned}
 (a) \quad Q_{11} &= Z_A^T C_A^{-1} Z_A \\
 (b) \quad Q_{12} &= Z_A^T C_A^{-1} Z_B^T \\
 (c) \quad Q_{21} &= Z_B^T C_A^{-1} Z_A^T \\
 (d) \quad Q_{22} &= Z_B^T C_A^{-1} Z_B^T \\
 (e) \quad G_Q &= P_{11}^{-1} Q_{12} \\
 (f) \quad C_Q &= Q_{22} - Q_{21} G_Q \quad (5.3) \\
 (g) \quad C_E &= (Z_A^T \quad Z_B^T) \begin{pmatrix} P_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}^{-1} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix} \\
 (h) \quad K_A &= (P_{11}^{-1} + G_Q C_Q^{-1} G_Q^T) Z_A \\
 (i) \quad K_B &= C_Q^{-1} G_Q^T Z_A \\
 (j) \quad K_C &= C_Q^{-1} G_Q^T X_{11}^T \Sigma_{11}^{-1} \\
 (k) \quad R_E &= K_A^T X_{11}^T \Sigma_{11}^{-1} Y_1
 \end{aligned}$$

Note that

$$\begin{aligned}
 C_E &= (Z_A^T \quad Z_B^T) \left(\begin{array}{c|c} P_{11}^{-1} + G_Q C_Q^{-1} G_Q^T & -G_Q C_Q^{-1} \\ \hline -C_Q^{-1} G_Q^T & C_Q^{-1} \end{array} \right) \begin{pmatrix} Z_A \\ Z_B \end{pmatrix} \\
 &= K_A^T Z_A - K_B^T Z_B.
 \end{aligned}$$

Further, given Σ_{11}^{-1} , $(X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1}$, $X_{11}^T \Sigma_{11}^{-1} \tilde{y}_1$ and $X_{11}^T \Sigma_{11}^{-1}$

no matrix inversions of order greater than n_2 (where $\tilde{y}_2$ is

$(n_2 \times 1)$) are required in forming the terms of equation (5.3)(a) - (k).

Theorem 5.1

$$(i) \quad (X^T \Sigma^{-1} X)^{-1} = \left(\begin{array}{c|c} (X_{11}^T \Sigma_{11}^{-1} X_{11})^{-1} & 0 \\ \hline 0 & 0 \end{array} \right) + \left(\begin{array}{c} G_Q \\ 0 \\ -I \end{array} \right) C_Q^{-1} (G_Q^T \vdots -I)$$

$$- \left(\begin{array}{c} K_A \\ -K_B \end{array} \right) (C_A + C_E)^{-1} (K_A^T \vdots -K_B^T)$$

$$(ii) \quad \Sigma^{-1} X = \left(\begin{array}{c|c} \Sigma_{11}^{-1} X_{11} & 0 \\ \hline 0 & 0 \end{array} \right) - \left(\begin{array}{c} G_A \\ -I \end{array} \right) C_A^{-1} (Z_A^T \vdots Z_B^T)$$

$$(iii) \quad \tilde{y}^T \Sigma^{-1} X = \tilde{y}_1^T \Sigma_{11}^{-1} (X_{11} \vdots X_{12}) + \tilde{y}_A^T C_A^{-1} (Z_A^T \vdots Z_B^T)$$

Proof:

(i) The proof in principle closely follows that of Theorem 4.1.

Again Σ^{-1} is given by equation (4.4) and

$$X^T \Sigma^{-1} X = \left(\begin{array}{cc} P_{11} & 0 \\ 0 & 0 \end{array} \right) + \left(\begin{array}{c} Z_A \\ Z_B \end{array} \right) C_A^{-1} (Z_A^T \vdots Z_B^T) \quad (5.4)$$

$$= \left(\begin{array}{cc} P_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array} \right) + \left(\begin{array}{c} Z_A \\ 0 \end{array} \right) C_A^{-1} (Z_A^T \vdots 0)$$

c.f. equation (4.5).

We again use equation (4.6) (Rao (1973 : 33)), namely

$$(A + BDB^T)^{-1} = A^{-1} - A^{-1} B(B^T A^{-1} B + D^{-1})^{-1} B^T A^{-1} \quad \text{where here}$$

$$A = \begin{pmatrix} P_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \quad B = \begin{pmatrix} Z_A \\ 0 \end{pmatrix}, \quad D = C_A^{-1}.$$

A^{-1} exists provided X_{22} is of full column rank. If X_{22} is not of full column rank, then β_2 contains a parameter redundancy which can be removed, making X_{22} full column rank in the modified model. The result follows from the definitions of C_E , K_A and K_B and the expansion of A^{-1} .

(ii) and (iii) follow directly from Theorem 4(ii) and (iii) on setting $X_{12} = 0$.

Corollary 5.1

$$\begin{aligned} \text{var}(\tilde{\beta} - \beta) &= \text{var} \begin{pmatrix} \tilde{\beta}_1 - \beta_1 \\ \tilde{\beta}_2 - \beta_2 \end{pmatrix} \\ &= \begin{pmatrix} \text{var}(\hat{\beta}_1 - \beta_1) & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} G_Q \\ -I \end{pmatrix} C_Q^{-1} (G_Q^T \parallel -I) \\ &\quad - \begin{pmatrix} K_A \\ -K_B \end{pmatrix} (C_A + C_E)^{-1} (K_A^T \parallel -K_B^T) \end{aligned} \quad (5.5)$$

Proof:

Directly from equation (1.10) and Theorem 5.1 since

$$\text{var}(\tilde{\beta} - \beta) = (X^T \Sigma^{-1} X)^{-1}$$

Theorem 5.2

The mmsle of β is given by

$$\begin{aligned} \tilde{\beta} = & \begin{pmatrix} \hat{\beta}_1 \\ \tilde{0} \end{pmatrix} + \begin{pmatrix} G_Q \\ -I \end{pmatrix} K_C \tilde{y}_1 \\ & + \begin{pmatrix} K_A \\ -K_B \end{pmatrix} (C_A + C_E)^{-1} (\tilde{R}_A - \tilde{R}_E) \end{aligned} \quad (5.6)$$

Proof:

The proof parallels that of Theorem 4.2 with $G_Q, C_Q, C_E, K_A, K_B, K_C, \tilde{R}_E$ replacing $G_P, C_P, C_D, H_A, H_B, H_C$ and $\tilde{R}_D$ respectively.

The residual sum of squares of the augmented model is best calculated via equation (4.9) and the results of Theorem (5.1).

6. STRUCTURAL PARALLELS IN THE UPDATING EQUATIONS

The very strong structural parallels between the formulae of Sections 4 and 5 are immediately apparent. This is not unexpected since both sections involve the addition of extra data and regression simultaneously and the methods of derivation given here are strongly parallel. We restrict comparisons therefore to the formulae of Section 2 for additional regressors with those of Section 4, and to the formulae of Section 3 for additional data with those of Section 4.

Equation (4.8) gives the updating formula for the estimate of the parameter β of equation (4.2). Comparison of equation (4.8) with equations (2.7) and (3.2) shows that a term paralleling the second term of equation (4.8) occurs as the second term of equation (2.7) and that the third term of equation (4.8) is in part paralleled by the second of equation (3.2). Similar parallels are also apparent on comparison of equations (2.6), (3.4) and (4.7) which give the mean square error of the updated estimates. Comparison of the formulae for updated residual sums of squares (equations (2.3), (3.3), and (4.11)) also reveals some parallel terms. (The term equivalent to $\tilde{\beta}_2^T X_{12}^T R_{11} \tilde{y}_1$ of equation (2.3) has been absorbed into the term $\tilde{y}_1^T C_L^T C_P C_L \tilde{y}_1$ of equation (4.11)).

7. UPDATING ALGORITHMS AND A GENERALISED KALMAN FILTER

The Kalman Filter (Anderson and Moore (1979)) can be viewed as the solution of a regression model (Duncan and Horn (1972)). A typical form of the Kalman Filter for a k variate time series $y_1, y_2, \dots, y_n$ is given by the recursive equations

$$\alpha_t = T_t \alpha_{t-1} + u_t \quad (7.1)$$

$$y_t = X_t \alpha_t + \varepsilon_t \quad (7.2)$$

where

$$\text{var} \begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix} = \begin{pmatrix} Q_t & 0 \\ 0 & R_t \end{pmatrix} \quad \text{and} \quad \varepsilon \begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (7.3)$$

for $t = 1, 2, \dots, h$.

The starting equation for α_t may be written

$$\alpha_1 = \mu_1 + u_1 \quad (7.4)$$

with μ_1 assumed known.

While Anderson and Moore (1977) discuss Kalman Filters for which contemporaneous correlation between u_t and ε_t is non-zero, all Kalman Filters assume that the error vectors

$\{(u_t^T, \varepsilon_t^T) : t = 1, 2, \dots, h\}$ are non-autocorrelated.

Duncan and Horn (1972) illustrate the relationship between the Kalman Filter of equations (7.1) and (7.4) and the regression model given by equation (7.5) below, and show that both give the same estimator of α_t : namely its mmsle, when $\text{var}(u_t^T, \varepsilon_t^T)^T$ is non-singular for $t' = 1, 2, \dots, t$.

$$\begin{bmatrix} \mu_1 \\ y_1 \\ 0 \\ y_2 \\ \vdots \\ 0 \\ y_{t-1} \\ 0 \\ y_t \end{bmatrix} = \begin{bmatrix} I & 0 & & & 0 \\ X_1 & 0 & & & 0 \\ -T_2 & I & 0 & & 0 \\ 0 & X_2 & 0 & & 0 \\ \vdots & & & & \\ \cdot & & -T_{t-1} & I & 0 \\ \cdot & 0 & & X_{t-1} & 0 \\ \cdot & & 0 & -T_t & I \\ 0 & & & 0 & X_t \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{t-1} \\ \alpha_t \end{bmatrix} + \begin{bmatrix} -u_1 \\ \varepsilon_1 \\ -u_2 \\ \varepsilon_2 \\ \vdots \\ -u_{t-1} \\ \varepsilon_{t-1} \\ -u_t \\ \varepsilon_t \end{bmatrix} \quad (7.5)$$

i.e.,

$$\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix} = \begin{pmatrix} X_{11} & 0 \\ X_{21} & X_{22} \end{pmatrix} \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \quad (7.6)$$

say, where the partitioning is given to illustrate the relationship with the formulae of Section 5. Noting that $\tilde{\alpha}_t$ of equation (7.5) is $\tilde{\beta}_2$ of equation (7.6) and (5.1), it is clear that the mmsle of $\tilde{\beta}_2$:

$$\tilde{\beta}_2 = -K_C \tilde{y}_1 - K_B (C_A + C_E)^{-1} (\tilde{R}_A - \tilde{R}_E)$$

from equations (5.3) and (5.6), is the mmsle of $\tilde{\alpha}_t$ given by the Kalman Filter.

Further for the Kalman Filter the additional data $(0^T : \tilde{y}_t^T)^T$ is usually added in two steps, the first giving the mmsle of $\tilde{\alpha}_t$ given the data $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_{t-1}$ to time $t-1$, denoted $\tilde{\alpha}_t|_{t-1}$. The formulae of Sections 4 and 5 can be used to achieve the same objective, the formulae of Section 5 being used to determine $\tilde{\alpha}_t|_{t-1}$, and those of Section 4 to determine $\tilde{\alpha}_t$ the mmsle of $\tilde{\alpha}_t$ given data $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_t$ to time t .

The formulae of Sections 4 and 5 are however more general, and allow generalisation of the Kalman Filter to allow autocorrelation between error terms $\{(u_t^T : \tilde{e}_t^T) : t = 1, 2, \dots, h\}$ provided the associated error covariances are non-singular. Parameter vectors with moving average, autoregressive and autoregressive-moving average structure can thus be incorporated into random coefficient regression models without using state space Markovian forms having singular covariance matrices.

The usual state space forms are given for example, in Anderson and Moore (1979 : 17).

Again this generalisation may proceed for each time t in either one or two steps. Either $\begin{pmatrix} 0^T \\ \tilde{\alpha}_t \end{pmatrix} \begin{pmatrix} y_t^T \end{pmatrix}$ is incorporated in one step using the formulae of Section 5 to give $\tilde{\alpha}_t$ directly, or the incorporation proceeds in two stages, the first incorporating 0 (i.e., $\varepsilon(\alpha_t)$) as data to give $\tilde{\alpha}_t|_{t-1}$ using the formulae of Section 5, the second incorporating y_t as data to yield $\tilde{\alpha}_t$ via the formulae of Section 4.

CHAPTER 6

POLYNOMIAL MODELS FOR REPEATED SAMPLE SURVEY DATA

SUMMARY

The existence of projection and updating formulae for general linear models where additional data are added but the length of the parameter vector is fixed, as outlined in Chapter 4, make the use of polynomial models in time to model underlying trends, an attractive proposition. When additional data are added, polynomial models require no supplementation of the parameter vector containing the polynomial coefficients, provided the degree of the polynomial is unchanged. Multiple observations, or elementary estimates in the sample survey context, at each time, also provide no theoretical problems. In this chapter polynomial models are discussed, together with their reparameterisation in terms of orthogonal polynomials, the results of a simulation study to assess the ability of fitted models to adequately model polynomial generated data are outlined, and medical data on the effect of the drug Propranolol on heart rate are analysed using fixed and random coefficient polynomial models.

* * * * *

1. INTRODUCTION

The use of polynomial models in time for modelling trends in short sequences of time series data has been the topic of considerable literature (e.g. Graybill (1976), Seber (1977), Box and Jenkins (1970), Fuller (1976), Priestley (1981)) and is a part of a more general topic of general linear models which are polynomial in their independent variables while being linear in their coefficients or parameters. Polynomials in time have often been applied to growth studies, for example in forestry and medicine, and this also forms a part of the more general study of growth curves (e.g. Ratkowsky (1983)), a topic not restricted to general linear model formulations. Multiple observations at each time produce no major theoretical complications, provided each such observation is an unbiased estimate of the appropriate underlying population mean.

One advantage of polynomial models in time is that they provide predictors of future values of the dependent variable for both uncorrelated and correlated model errors (see, for example, Chapter 4, Theorem 1, and Bibby and Toutenberg (1977:87)). Further, updating formulae for estimates of polynomial coefficients given additional data, together with cusum and cusum of squares tests for testing the constancy of the regression over time, are available both for uncorrelated errors (Brown et al. (1975)), and for correlated errors with additional data added one datum at a time (McGilchrist and Sandland (1979)) or as groups of data added simultaneously (see Chapter 4, Theorem 3).

Godolphin and Stone (1980) have shown that polynomial projecting predictors based on Kalman filter equations can be expressed in the same form which describes many classical polynomial predictors, and that such Kalman filter predictors, for which forecasts ℓ steps ahead based on data to time t follow a polynomial of order $p-1$, will coincide with the

corresponding least-squares predictors for an integrated moving average (IMA) $(0,p,r)$ process with $r \leq p$. There is thus an equivalence, at least in a predictive sense, between polynomial models with uncorrelated error structure and a subclass of ARIMA models. Given the class of time series data problems to which ARIMA models are applicable (see Chapter 1, however, for a note of caution), the utility of polynomial models in time for modelling certain types of time series data becomes clearer. Polynomial models with correlated error structure further extend the utility of the class of polynomial models. Polynomial plus autocorrelated error models are not, however, a subset of ARIMA models or vice versa.

In the context of repeated sample surveys then, polynomial models provide not only a method of smoothing (possibly multiple) mean estimates at each time for which data is already available, but also of testing the adequacy of such smoothing, together with formulae for predicting future mean values. Jessen (1942) proposed improvement of sample survey estimates from simple random samples at two periods only, by assuming that units matched between periods have correlation ρ between periods and that units appearing only at one time period are uncorrelated with all other responses. Jessen's results can be seen in the context of polynomial estimators as a polynomial model of order one fitted to two time periods; no other polynomial model is possible of course, since there are only two time periods. However, Jessen's interest was improvement of estimates for the two time periods for which he had data, rather than prediction of future values, and the extension of Jessen's (1942) results by Patterson (1950) while permitting a general linear model formulation as a polynomial model, was structured by Patterson to permit no prediction of future values.

The degree of the polynomial model equivalent to Patterson's model is $(h-1)$, where h is the number of time periods, so that the polynomial formulation imposes no additional structure on the underlying mean estimates and, since there are still h parameters to be estimated, is simply a reparameterisation of Patterson's original problem.

2. THE MODEL

To state the general polynomial model with fixed polynomial coefficients explicitly, we are considering a model linear in its parameters β the polynomial coefficients, which can be written given data to time h , as:

$$\underset{\sim}{y}_h = X_h \underset{\sim}{\beta} + \underset{\sim}{e}_h \quad (2.1)$$

with $\text{var}(\underset{\sim}{e}_h) = \Sigma_h$, $E(\underset{\sim}{e}_h) = \underset{\sim}{0}$, $\underset{\sim}{\beta}$ ($p \times 1$) containing the polynomial coefficients, $\underset{\sim}{y}_h$ the data from time periods 1 to h , and the rows of X_h having the form $\underset{\sim}{t}^T = (1, t, t^2, \dots, t^{p-1})$ for $t = 1, 2, \dots, h$. $\underset{\sim}{y}_h$ is not $(h \times 1)$ when there is more than one datum at any time period. The parameters of real interest are the h underlying population means at each time t , namely $\underset{\sim}{Y} = (Y(1), \dots, Y(h))^T$ the components of which are given by $\underset{\sim}{t}^T \underset{\sim}{\beta}$ for $t = 1, 2, \dots, h$. Here E denotes design expectation and var design variance, as discussed in Chapters 2 and 3.

The data $\underset{\sim}{y}_h$ used in the linear model may in the case of simple random sampling consist of individual responses from the times $t = 1, 2, \dots, h$, but more generally will have elementary estimates or the h traditional estimates $\hat{Y}(1), \hat{Y}(2), \dots, \hat{Y}(h)$ as its component elements. Elementary estimates as discussed in Chapter 1, consist of unbiased sub-means formed as (weighted) averages of individual responses from the same time period, where those individuals in an elementary grouping arrive and leave the

set of repeated surveys, together as a group. Normality assumptions for such elementary estimates at a particular time are not unreasonable, via a central limit theorem approximation, given the large sample sizes of many repeated sample surveys (Cochran (1977)).

Generally Σ_h is not known. However, in the sample survey context good estimates are often possible where individual information from surveys has been preserved. Providing removal of the polynomial trend is assumed to produce a stationary time series, estimates of Σ_{h+1} are often possible even without the extra data for time $h+1$. Covariance estimation using individual responses from several time periods has been called primary analysis by Scott, Smith and Jones (1977).

Given the model of equation (2.1) and the data to time h , β is estimated in principle by the generalised least squares formula

$$\hat{\beta}_{\sim} = (X_h^T \Sigma_h^{-1} X_h)^{-1} X_h^T \Sigma_h^{-1} y_{\sim h}$$

The matrix $(X_h^T \Sigma_h^{-1} X_h)$ is however generally ill-conditioned, i.e. small changes in the components of $(X_h^T \Sigma_h^{-1} X_h)$ lead to relatively large changes in $(X_h^T \Sigma_h^{-1} X_h)^{-1}$. The special case where $\Sigma_h = I$ is discussed for example in Froberg (1969). The ill-conditioning problem can be circumvented by the use of orthogonal polynomials (Phillips and Taylor (1973)), in which case we may reparamaterise our estimation problem with data to time h as:

$$y_{\sim h} = P_{h\sim} \alpha + e_{\sim h} \quad (2.2)$$

with

$$P_{h\sim} \alpha = X_{h\sim} \beta \quad (2.3)$$

where $\underline{\alpha}$ is a vector of coefficients to the orthogonal polynomials $\psi_0(t), \psi_1(t), \dots, \psi_{p-1}(t)$; P_h is a matrix with the elements of each row determined by the evaluation of the orthogonal polynomials at the corresponding value of t ; and P_h incorporates the orthogonality property so that $P_h^T P_h = I$, the identity matrix. Further details may be found in the MACRO INITPOL section of Appendix 1.

Having determined $\hat{\underline{\beta}}$ via $\hat{\underline{\alpha}}$, for the data to time h , we may use these parameter estimates to predict future population mean estimates. Such projections would usually be restricted to time $(h+1)$, i.e. one step ahead predictors, to avoid the cumulative effect of possible model misspecification errors. For the one step ahead predictor, the underlying mean estimate is:

$$\hat{Y}^{(h)}(h+1) = \underline{x}_{h+1}^T \hat{\underline{\beta}} = \underline{p}_{h+1}^T \hat{\underline{\alpha}} \quad (2.4)$$

with $\underline{x}_{h+1}^T = (1, (h+1)^2, \dots, (h+1)^{p-1})$

$$\underline{p}_{h+1}^T = (\psi_0(h+1), \psi_1(h+1), \dots, \psi_{p-1}(h+1))$$

Note that no information on the covariance structure for data at time $(h+1)$ is required. Theorem 1 of Chapter 4 also provides predictions for the elementary estimates, one step ahead, but does require knowledge or estimates of the correlation structure of the errors at time $(h+1)$ with errors from previous time periods. A possible method of estimating such correlations in the absence of data at time $(h+1)$ has been mentioned previously. Further estimation details for polynomial projections may be found in the MACRO POLYPROJ section of Appendix 1.

Given additional data at time $(h+1)$ the formulae of Chapter 4 also provide an updating procedure whereby the estimate of $\hat{\underline{\beta}}$ (or $\hat{\underline{\alpha}}$) may be updated, and cusum or cusum of squares tests carried out to test for

the constancy of the regression. These updating equations and test procedures are incorporated in MACRO POLYUPDT, further details of which are to be found in Appendix 1. Alternative covariance estimates for Σ_h are available which assume a particular type of covariance structure and do not require a primary analysis if elementary estimate information is used as data in the polynomial model. Such covariance estimation procedures are called secondary analysis methods in the terminology of Scott, Smith and Jones (1977). For example, polynomial models with random polynomial coefficients are possible, at least for certain survey design rotation patterns, and for such models the random components may be seen as a modification to the covariance structure in a redefined model which incorporates superpopulation structure. We consider an example.

Within the general class of random coefficient models, random coefficient models of the Swamy type (Swamy (1971), Judge et al. (1980)) allow that different individuals or elementary estimates follow different polynomials over time, that these polynomial coefficients are random and each have the same mean and variance. For the i^{th} individual or elementary grouping:

$$\hat{y}_{\sim i} = x_i (\bar{\beta} + \mu_{\sim i}) + e_{\sim i}, \quad i = 1, 2, \dots, K \quad (2.5)$$

with $\beta_{\sim i} = \bar{\beta} + \mu_{\sim i}$

$$E(\mu_{\sim i}) = 0; \quad E(\mu_{\sim i} \mu_{\sim i}^T) = \Delta; \quad E(\mu_{\sim i} \mu_{\sim j}^T) = 0 \text{ for } i \neq j$$

Additionally, it is assumed that $E(e_{\sim i} e_{\sim i}^T | i) = \sigma_{ii} I$, that $E(e_{\sim i} e_{\sim j}^T) = 0$ for $i \neq j$, and in the sample survey context that all elementary groupings remain in the sample for the same number of time periods; this last assumption does not necessarily imply a panel study since the total number

of time periods, denoted h , may exceed the number of time periods an elementary grouping remains in the sample. The matrix X_i for polynomial models again consists of rows having the form $\tilde{t}^T = (1, t, t^2, \dots, t^{p-1})$ with $1 \leq t \leq h$. Here E and ϵ are design and superpopulation expectations as for Chapters 2 and 3.

In principle, the covariance matrix of the composite disturbances:

$$\begin{bmatrix} X_1 \\ \\ X_2 \\ \\ \\ \\ X_k \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_k \end{bmatrix}$$

is then block diagonal with i^{th} diagonal block

$$\phi_{ii} = X_i \Delta X_i^T + \sigma_{ii} I \quad (2.6)$$

If we let

$$\tilde{b}_i = (X_i^T X_i)^{-1} X_i^T \hat{Y}_i \quad (2.7)$$

Swamy (1971) has shown that we may estimate σ_{ii} , Δ and hence ϕ_{ii} by

$$\hat{\sigma}_{ii} = \tilde{e}_i^T \tilde{e}_i / (h-p) \quad (2.8)$$

$$\hat{\Delta} = S_b / (K-1) \quad (2.9)$$

where

$$\tilde{e}_i = \hat{Y}_i - X_i \tilde{b}_i$$

and

$$S_b = \sum_{i=1}^K b_i b_i^T - \frac{1}{N} \left(\sum_{i=1}^K b_i \right) \left(\sum_{i=1}^K b_i \right)^T$$

In practice, the model would be reparameterised with respect to the appropriate orthogonal polynomials. Further details may be found in the MACRO ESTCOV section of Appendix 1 where a slightly different notation is used, in Judge et al. (1980) or in Swamy (1971).

Repeated measures type experimental designs may be considered as a special case of a sample survey, namely a panel survey selected by simple random sampling where elementary estimates are individual responses. Given h periods of data (or equivalently h repeated measures) with all individual responses being i.i.d., the usual ANOVA procedure treats time as a randomisable crossed effect (i.e. all individuals are observed at each time). When responses on any individual are correlated over time however, or there is heteroscedasticity of responses between individuals, the i.i.d. assumption is violated, and the analysis of variance is invalidated. An alternative procedure is then to treat responses from particular individuals as polynomials in time. This matter is discussed further in Section 4, where an example, which can be considered either as a sample survey or as an experimental design, is analysed.

3. A SIMULATION STUDY

One measure of the efficacy of polynomial models is their ability to closely approximate the true values of the underlying means $Y(1), Y(2), \dots, Y(h)$. If these underlying means are non-stochastic and themselves follow a polynomial model, estimates of the polynomial coefficients would be expected to closely approximate the true coefficients. Errors of interpolation and extrapolation should then be small. This would suggest a simulation study in which polynomials of a particular

degree were generated with known coefficients, and fitted polynomials were judged against the true polynomials on the basis both of equality of polynomial order and similarity of polynomial coefficients. This criterion is not entirely satisfactory however, since for example a cubic may be well approximated in a part of its range by a quadratic. The predictions obtained from a quadratic fitted to simulated cubic data within this range may differ little from those for a fitted cubic despite the order being incorrect for the fitted quadratic; the fitted quadratic coefficients may also differ markedly from the true cubic coefficients despite the fit being judged as adequate on the basis of the deviations from the true cubic curve.

It is on this basis of deviation of the fitted from the true polynomial that the primary criterion of adequate fit is based. For a particular polynomial fit, interpolation error may be estimated by the mean square error of the fit, MSE, defined by:

$$MSE = \frac{1}{h} \sum_{t=1}^h (F_t - A_t)^2 \quad (3.1)$$

where F_t and A_t are the values of the fitted and actual polynomials respectively at time t , where $t = 1, 2, \dots, h$.

Extrapolation error may be estimated by an alternative criterion, one step ahead mean square prediction error. The square of the one step ahead prediction error (MSEP) is also calculated in the simulation study but is inherently less accurate than MSE since only a single one step ahead prediction error based on data to a fixed time h is available from each polynomial simulation.

The parameters used in the simulation study are outlined below. The total number of time periods is denoted h ; the number of elementary estimates per time period by K ; the correlation at lag ℓ between elementary estimates from the same elementary group about the true polynomial value by $\rho_\ell = \rho^\ell$; the variance of simulated elementary estimates about the true polynomial at time t by σ^2 ; the degree of the true polynomial by $p-1$; the minimum and maximum of $\underline{Y} = (Y(1), Y(2), \dots, Y(h))^T$ the values of the true polynomial at times $t = 1, 2, \dots, h$, by $\min(\underline{Y})$ and $\max(\underline{Y})$ respectively; N is the total population size at each time t ; and s the number of simulations for fixed h, K, ρ, σ^2 and p . The correlation between different elementary means both at the same time and at different times is assumed to be zero (since for elementary estimates from the same time t it is approximately $-\sigma^2/(N-1)$, see for example Chapter 2).

Parameter values for simulation study were:

- (a) $h = 10$
- (b) $K = 5$
- (c) $\rho_\ell = \rho^\ell$ with $\rho = 0.5, 0.7, 0.9$
- (d) $\sigma^2 = 1, 9, 25$
- (e) $\min(\underline{Y}) = 15; \max(\underline{Y}) = 45$
- (f) $p-1 = 2, 3, 4$
- (g) $s = 5$

giving $3 \times 3 \times 3 \times 5 = 135$ simulations in all.

Note that for each of the $K = 5$ data points (i.e. elementary estimates), assumed normally distributed at each time t , σ^2 is the corresponding variance. Given the sample size in each elementary estimate is $n/5$ where n is the total sample size at each time period, the population variance for the population sampled is $n\sigma^2/5$. If, for example, at

each time $t = 1, 2, \dots, h$, population values were to be normally distributed, and no more than 1 in 200 were to be negative, the maximum sample size per elementary estimate would be 25 for $\sigma^2 = 1$, ≈ 3 for $\sigma^2 = 9$ and 1 for $\sigma^2 = 25$, substantially smaller sample sizes than customary for repeated sample surveys. The simulation therefore provides a stringent test for polynomial models with correlated errors applied to aggregated sample survey data.

For the simulation, 45 different sets of random seeds were used to generate independent normally distributed variables for the nine combinations $\rho = 0.5, 0.7, 0.9$ and $\sigma^2 = 1, 4, 9$. These realisations were transformed using a Choleski decomposition of the appropriate inverted Toeplitz matrix to give transformed elementary estimate errors which were then added to an appropriate polynomial of order 2, 3 and 4. The same transformed errors were thus added to three polynomials of different orders. The fifteen underlying polynomials themselves were generated randomly, by specifying p equally spaced points on the interval $[0, 10]$, where $(p-1)$ is the order of the polynomial; at each of these p times a uniform random number was then generated on the interval $[0, 1]$ and the p coordinate pairs used to specify a unique polynomial of order $(p-1)$. This polynomial was then scaled so that $\min(\tilde{Y}) = 15$ and $\max(\tilde{Y}) = 45$; polynomials so generated were then chosen so that the absolute value of the leading coefficient was > 0.7 for quadratics, > 0.25 for cubics and > 0.1 for quartics, and duplicates of very similar polynomials were excluded. These values of the leading coefficient were sufficiently large that generally $(p-2)$ turning points appeared in the interval $[0, 10]$ for each polynomial of order $(p-1)$.

For each simulated polynomial plus error, the polynomial degree for the "best" fit was chosen using the F statistic defined below:

$$F = \frac{RSS_{p'} - RSS_{p'-1}}{ESS_{min} / DF_{ESS_{min}}} \quad (3.2)$$

where RSS denotes regression sum of squares, with p' and $p'-1$ the degree of the fitted polynomials

ESS_{min} is the minimum over all polynomials fitted to the data, of the error sum of squares, and

$DF_{ESS_{min}}$ its corresponding degrees of freedom.

This F statistic thus makes an ad hoc adjustment to the more usual form of the F statistic, which tends to be unstable when, for the same data, a different error covariance matrix is estimated for each polynomial (as with the random coefficient regression model discussed in Section 4). In the present instance the F test was carried out at the $(1-\alpha) = 1\%$ level. At this 1% level the correct degree for the fitted polynomial was chosen in all 135 simulations; 3 of 135 simulations showed a significant F value at the 5% level. This may, in part however, be a consequence of the method of constructing polynomials (see the final sentence of the previous paragraph).

Independent recursive residuals calculated as detailed in Chapter 4 were analysed both using cusum tests and a chisquare statistic based on $hK-p'$ degrees of freedom (where $p'-1$ is the degree of the fitted polynomial); the chisquare test is based on the fact that, for the correct fit, recursive residuals are i.i.d. $N(0,1)$ random variables. At the 5% level, 3/135 simulations failed the cusum test, and 10/135 failed the chisquare test; 2/135 failed the chisquare test at the 1% level. Failure of these two tests was related to the particular set of errors chosen, rather than the degree of the true polynomial, since the same set of correlated errors fitted to polynomials of different degrees gave very similar values of $(1-\alpha)$ for these tests. Cusum tests were dependent

on which arbitrarily chosen "individual" supplied the p' points for the initial polynomial fit of order $p'-1$ and particularly on the relative level and polynomial "shape" of these points; for chisquare tests a similar difficulty occurred for high ρ values when the 1st "individual" fitted had a level in the first p' periods markedly different from the average over individuals (assuming a polynomial of degree $p'-1$ is again being fitted) since then the recursive residuals had a markedly non-zero mean, and the chisquare test statistic (which tests whether recursive residuals are all $N(0,1)$ random variables) was consequently large.

Normality tests were significant at the 5% level in 12/135 cases, and at the 1% level in 3/135 cases, and like the chisquare and cusum tests were dependent on the particular set of simulated correlated errors rather than the degree of polynomial to which they were fitted. The highest correlation (i.e. $\rho = 0.9$) accounted for 10/135 test failures at the 5% level, and all failures at the 1% level; for high correlations, the samples of size 5 at time period 1 for each simulation tend to have similar distributions at consequent time periods and normality testing of correlated residuals therefore gives a higher percentage of failures than would be expected from i.i.d. data - normality tests based on recursive residuals rather than model residuals may give more accurate results, but would be susceptible to model error variance-covariance misspecification.

Average values for root mean square error ($\overline{STD}$) and for root mean square prediction error ($\overline{STDP}$) for each combination of values for σ^2 , ρ , and $(p-1)$ are given in Table 6.3.1. $\overline{STD}$ is based on 50 data points and $\overline{STDP}$ on 5. The most marked effect is the increase of $\overline{STD}$ and $\overline{STDP}$ with increasing σ^2 . Other effects were also considered. Analysis of variance techniques applied to $\overline{STD}$ and $\overline{STDP}$ as dependent variables with

TABLE 6.3.1

Average root mean square error ($\overline{\text{STD}}$) and average root mean square error of prediction ($\overline{\text{STDP}}$) for simulated polynomials with correlated error

error variance σ^2	correlation ρ	degree of polynomial $p-1$	$\overline{\text{STD}}$ *	$\overline{\text{STDP}}$ **
1	0.5	2	0.231	0.298
1	0.5	3	0.324	0.837
1	0.5	4	0.335	1.239
1	0.7	2	0.282	0.215
1	0.7	3	0.315	0.486
1	0.7	4	0.323	0.511
1	0.9	2	0.450	0.562
1	0.9	3	0.454	0.609
1	0.9	4	0.453	0.672
9	0.5	2	1.054	2.420
9	0.5	3	1.075	2.329
9	0.5	4	0.888	2.345
9	0.7	2	0.803	1.650
9	0.7	3	0.886	1.736
9	0.7	4	0.949	1.393
9	0.9	2	0.914	1.323
9	0.9	3	0.988	1.463
9	0.9	4	1.152	1.665
25	0.5	2	2.062	2.562
25	0.5	3	2.106	3.235
25	0.5	4	2.227	6.370
25	0.7	2	1.762	3.229
25	0.7	3	2.029	3.834
25	0.7	4	2.112	4.796
25	0.9	2	2.260	3.824
25	0.9	3	2.266	3.560
25	0.9	4	2.291	4.076

* based on 5 simulated polynomials, and 10 data per polynomial

** based on 5 simulated polynomials, and 1 datum per polynomial

the design defined by particular polynomials (A,B,C,D,E) nested within degree ($p-1 = 2,3,4$) and crossed with correlation ($\rho = 0.5, 0.7, 0.9$), analysed separately for each variance ($\sigma^2 = 1,9,25$), yielded the following results.

For $\overline{\text{STD}}$, none of the effects, "correlation", "degree" or their interaction were significant for any of the separate analyses for each value of σ^2 (the minimum value of $(1-\alpha)$ assuming normal residuals was 0.36 - one residual set of three was however significantly non-normal at a $(1-\alpha)$ level of 0.01). For $\overline{\text{STDP}}$, only "correlation" was significant at the 5% level for $\sigma^2 = 1$, with the same result for $\sigma^2 = 9$; no effects were significant for $\sigma^2 = 25$; all ANOVA's for $\overline{\text{STDP}}$ passed the normality tests on residuals with $(1-\alpha) > 0.03$. Looking at Table 6.3.1, the effect due to "correlation" has been to increase $\overline{\text{STDP}}$ at the lowest value of ρ for $\sigma^2 = 1$, and $\sigma^2 = 9$; from a further analysis of the data used to produce Table 6.3.1, this effect is marked where there is a turning point in the actual polynomial in the range $t = 9$ to $t = 10$ (particularly where this is not adequately tracked in the fitted polynomial due to the pattern of individual variation at $t = 10$), or where the actual polynomial has steep gradient in the interval $t = 10$ to $t = 11$ (prediction error was calculated at $t = 11$).

The values of the actual polynomial in this simulation study were always in the range 15 to 45. For $\sigma^2 = 1$ the average prediction error of 0.79 for $\rho = 0.5$, 0.40 for $\rho = 0.7$ and 0.61 for $\rho = 0.9$, given the range of the true polynomial, is relatively small. For $\sigma^2 = 9$ (i.e. $\sigma = 3$) average prediction error (i.e. average $\overline{\text{STDP}}$) is 2.36 for $\rho = 0.5$, 1.59 for $\rho = 0.7$ and 1.48 for $\rho = 0.9$; the corresponding figure for $\sigma^2 = 25$ ($\sigma = 5$) is 3.94 ("correlation" is not a significant effect for $\overline{\text{STDP}}$ for $\sigma^2 = 25$).

For $\sigma^2 = 25$, therefore, relative prediction errors can be reasonably substantial, at least when the range (15-45) of the values of $Y(1), Y(2), \dots, Y(10)$ is equal to their approximate mean, i.e. 30. Average errors of fit (i.e. average $\overline{\text{STD}}$) however, appear to remain small comparative to $\overline{\text{STD}}$, 0.35 for $\sigma^2 = 1$, 0.97 for $\sigma^2 = 9$, 2.12 for $\sigma^2 = 25$, suggesting that true polynomials can be approximated, at least within their range, by fitted polynomials, even for large variances σ^2 , a small number of time periods ($h = 10$) and a considerable relative variation (i.e. range/average) in values of the true polynomial. It is worth noting that for $\sigma^2 = 25$ the $\pm 2\sigma$ confidence interval, corresponding approximately to 95% confidence limits at each time $t = 1, 2, \dots, 10$, is ± 10 ($\sigma^2 = 9$ gives ± 6 ; $\sigma^2 = 1$ gives ± 2), with standard error of the mean at each time if the data were non-autocorrelated being 2.24 for $\sigma^2 = 25$ (1.34 for $\sigma^2 = 9$, and 0.45 for $\sigma^2 = 1$). Average $\overline{\text{STD}}$ values are therefore very reasonable, and even the prediction error for $\sigma^2 = 25$ is not surprising when viewed in this light.

As noted previously, relative variance is very much greater in this simulation than would be usual for elementary estimates based on means for repeated sample survey data. It is therefore very encouraging that the fitted and actual polynomials are as similar as could be expected for all values of σ^2 , given 10 time points, and that even one step ahead prediction errors are generally of reasonable size. The exceptions to adequate prediction occur where σ^2 is very high, and the last time point for which data is available is close to a turning point in the true underlying polynomial. Even this exception could be controlled in part by increasing the number of data at each time point beyond the five data used in this simulation, even if the cost were to increase σ^2 . It is also encouraging that in a separate set of simulations with correctly specified correlation but with scale factor σ^2 underestimated

by a factor of 9 ($\sigma^2 = 81$, $\hat{\sigma}^2 = 9$), polynomials of the correct order were chosen in 45/45 cases.

4. AN EXAMPLE : THE EFFECT OF PROPRANOLOL ON HEARTRATE

Maling and Dollery (1980) studied the effects of the drug propranolol on heart rates and blood pressure of five patients hospitalised for high blood pressure and associated cardiac problems. Each patient was given three treatments on separate occasions, a placebo, an acute dose (one dose at the beginning of the timed period), and a chronic dose (two doses per day for 4 days prior to, plus one dose at the beginning of the timed period). Heart rate and blood pressure (both diastolic and systolic), were measured at 15 minute intervals for the first hour, and at hourly intervals to the eleventh hour. Lunch was eaten by all patients at the fourth or fifth hour and dinner at the ninth or tenth. The data was analysed by Maling and Dollery (1980) as an analysis of variance with time and treatment as independent categorical variables. Providing that all model errors are i.i.d. the design is a repeated measures design, (Time x Treatment x Subject) where x denotes "crossing" of effects. Fitting of a "meal" effect was not possible because lunch and dinner were not given to all patients at the same time. Here we consider the heart rate data only, and as an alternative analysis fit a polynomial in time for each of the three treatments. The five different individuals can either be taken to follow different random polynomials in time, (i.e. a random coefficient model) or, as a less efficient alternative, the mean estimates at each time for a given treatment may be modelled as a polynomial in time without replication under the assumption that resultant model errors are i.i.d.

In the discussion that follows there are three different sets of model errors to be distinguished. For each treatment, the ANOVA model

of Maling and Dollery (1980) has the model errors which are for each individual at each time, and measure individual deviations from the usual mean value averaged across individuals at that time for that treatment. For the random coefficient (r.c.) polynomial the model errors for each individual at each time for each treatment, measure individual deviations from the polynomial regression fitted to time series data for that individual only, for that treatment. Finally, for the polynomial model for mean data, model errors measure departures of the mean values taken over all individuals at each time for each treatment, from the polynomial fitted to the mean data (rather than to the individual responses) for that treatment.

One disadvantage of the small sample in this study is there is no a priori reason that the model errors should be normally distributed when a random coefficient model (or an ANOVA model) is used to analyse this data, invalidating the usual F tests together with critical regions for cusum and cusum of squares tests. Parameter estimates remain best linear unbiased however. This likely non-normality then, would also have presented problems for the original analysis performed by Maling and Dollery (1980); the problem is in part circumvented by the use of data averaged over individuals, but even with such an analysis normality assumptions could remain in some doubt, since such averages are taken over only five individuals.

The data is thus not ideal for an analysis of this kind, and differs from the situation for many large scale repeated sample surveys in that the sample size is small, individual information is used in the general linear model rather than the more aggregated elementary estimates, and the sampling scheme here is simple random sampling. Elementary estimates which fulfil the unbiasedness requirement for use in linear models have not however been commonly collected for large scale sample

surveys for a sufficient period of time, at least in New Zealand, and the data analysed here are used, despite their shortcomings, to provide some indication of the utility of polynomial methods in more pertinent situations.

The medical questions relating to heart rate which this study was designed to assess can be summarised as follows:

1. Does propranolol lower heart rate?
2. If so, to what extent and how rapidly, and is the effect sustained for the chronic dosage?
3. Given that ingestion of food is known to increase heart rate for 1-2 hours following a meal, does propranolol dampen this response?

A number of further comments on the statistical techniques used, and the statistics derived from them, are necessary before attempting to answer these questions. A number of relevant statistics from the random coefficient model analysis and the analysis of means are contained in Tables 6.4.1 and 6.4.2 respectively. For the r.c. polynomial model, the white noise test is based on the portmanteaux statistic of Ljung and Box (1978) tested for each individual at maximum lag of six,

$$x^2 = h(h+2) \sum_{\ell=1}^6 (r_{\ell}/(h-\ell))$$

where

$$r_{\ell} = \sum_{t=1}^{h-\ell} (\hat{e}_t \hat{e}_{t+\ell}) / \sum_{t=1}^h \hat{e}_t^2$$

h is the number of time periods,

$\hat{e}_t$ is the estimated residual at time t , and

the subscript denoting individual i has been suppressed for clarity.

A similar but untabulated white noise test is used for the polynomial based on mean data. For r.c. polynomial fits, the normality test uses the Kolmogorov-Smirnov statistic, $\hat{\sigma}_{ii}$ is defined by equation (2.8), and the cusum and cusum of squares tests are not performed because effects due to individuals are very pronounced and there is no natural ordering of the data.

For the polynomial model based on mean data, cusum and cusum of squares tests are performed; the normality test for residuals is the Shapiro-Wilks (1965).

The tabulated F statistic is again of the modified form defined in equation (3.2) and tests whether a polynomial of degree p' is a significantly better fit than one of degree $(p'-1)$.

Tabulated probabilities for the white noise test, the normality test, the cusum of squares and the F test are the probabilities of test statistics at least as large under the respective null hypotheses of no autocorrelation, normal residuals and an adequate fit for a polynomial of degree $(p'-1)$, i.e. $(1-\alpha)$. For the cusum test A denotes acceptance of the null hypothesis of no change in $(p'-1)^{\text{st}}$ degree polynomial coefficients with the increased number of time periods beyond p' at the 95% level, and R denotes rejection of this null hypothesis at the same level.

The individual data for placebo, acute dose and chronic dose, together with the mean values of this data for each dose at each time are plotted in Fig. 6.4.1. Also plotted are the r.c. polynomial and the polynomial based on mean data for each dose of degrees 4, 5 and 5,

TABLE 6.4.1
Summary Statistics for the Random Coefficient Polynomial Model

Treatment	degree of polynomial	White noise test (1- α)					$\hat{\sigma}_{ii}$					monomial coefficient by degree					Kolmogorov-Smirnov normality test	statistic	df	nominal p	
		1	2	3	4	5	1	2	3	4	5	0	1	2	3	4	5				
Placebo	1	0.103	0.002	0.879	0.567	0.318	5.83	9.24	16.57	10.35	22.75	62.61	0.3881	-	-	-	-	< 0.01	-	-	
	2	0.032	0.161	0.734	0.442	0.385	5.50	7.42	17.07	10.34	24.16	61.50	1.2950	-0.088115	-	-	-	< 0.01	5.34	1,72	0.024
	3	0.034	0.165	0.584	0.389	0.420	5.96	8.04	18.25	11.03	26.05	61.72	0.9072	0.008199	-0.05945	-	-	< 0.01	5.40	1,71	0.023
	4	0.009	0.201	0.714	0.217	0.583	6.09	2.60	18.44	10.99	15.92	63.87	-5.3662	2.8534	-0.4206	0.0189	-	< 0.01	2.53	1,70	0.116
Acute Propranolol	1	0.070	0.000	0.035	0.290	0.205	25.95	6.16	15.36	12.95	45.53	56.60	-0.3522	-	-	-	-	> 0.15	-	-	
	2	0.014	0.008	0.040	0.162	0.046	22.32	5.08	16.29	13.15	32.30	58.10	-1.6405	0.1272	-	-	-	> 0.15	2.82	1,72	0.10
	3	0.440	0.685	0.040	0.036	0.007	13.87	1.85	17.65	10.45	26.52	60.38	-5.6938	1.1257	-0.06125	-	-	> 0.15	4.01	1,71	0.05
	4	0.260	0.650	0.038	0.199	0.015	6.29	1.94	19.23	7.98	18.74	62.50	-11.270	3.5745	-0.4126	0.01586	-	> 0.15	4.63	1,70	0.035
	5	0.289	0.284	0.003	0.025	0.565	6.87	1.50	19.00	2.45	7.81	63.61	-15.400	6.5101	-1.1577	0.09338	-0.002823	> 0.15	4.62	1,69	0.035
Chronic Propranolol	1	0.081	0.366	0.279	0.065	0.889	11.13	3.74	3.60	9.40	6.29	50.49	-0.02980	-	-	-	-	> 0.15	-	-	
	2	0.213	0.056	0.316	0.048	0.403	8.59	3.42	3.85	9.35	4.88	51.44	-0.8321	0.07890	-	-	-	> 0.15	3.39	1,72	0.070
	3	0.071	0.195	0.430	0.041	0.215	5.79	3.28	3.61	10.11	4.66	52.48	-2.5991	0.5127	-0.02657	-	-	> 0.15	4.96	1,71	0.030
	4	0.048	0.154	0.046	0.008	0.011	6.17	1.19	3.26	9.77	4.23	53.79	-6.3221	2.1883	-0.2703	0.01109	-	0.058	4.88	1,70	0.030
	5	0.055	0.081	0.074	0.006	0.007	6.79	1.25	3.43	10.00	3.31	54.14	-7.9915	3.4459	-0.5979	0.04573	-0.001276	0.108	5.34	1,69	0.024

TABLE 6.4.2
Summary Statistics for the Polynomial Model for Mean Data

Treatment	degree of polynomial	monomial coefficient by degree					Shapiro-Wilks normality test of residuals (1- α)	Cusum test	Cusum of squares test (1- α)	statistic	F df	nominal p
Placebo	1	62.72	0.3697	-	-	-	0.495	R	> 0.20	-	-	-
	2	61.67	1.2512	-0.08626	-	-	0.725	R	> 0.20	5.77	1,12	0.04
	3	61.96	0.7353	0.04181	-0.007906	-	0.778	R	> 0.20	0.37	1,11	0.56
	4	64.15	-5.4802	2.8506	-0.4172	0.01867	0.972	A	0.02	16.44	1,10	< 0.01
	5	64.25	-5.8654	3.1325	-0.4898	0.02628	0.945	A	> 0.20	0.03	1,9	0.87
Acute Propranolol	1	56.75	-0.3845	-	-	-	0.958	R	> 0.20	-	-	-
	2	58.52	-1.8602	0.1444	-	-	0.592	R	> 0.20	36.72	1,12	< 0.001
	3	60.79	-5.9123	1.1503	-0.06210	-	0.594	A	0.10	51.50	1,11	< 0.001
	4	62.67	-11.2274	3.5523	-0.4121	0.01596	0.282	A	> 0.20	27.30	1,10	< 0.001
	5	63.63	-15.1387	6.4139	-1.1487	0.09324	0.737	A	0.10	7.05	1,9	0.025
Chronic Propranolol	1	50.48	-0.02732	-	-	-	0.348	R	0.02	-	-	-
	2	51.51	-0.8921	0.08463	-	-	0.360	R	0.10	14.93	1,12	0.003
	3	52.48	-2.6033	0.5094	-0.02622	-	0.796	A	> 0.20	10.87	1,11	0.008
	4	53.66	-5.9590	2.0259	-0.2472	0.01008	0.884	A	> 0.20	12.89	1,10	0.005
	5	54.18	-8.0963	3.5896	-0.6497	0.05231	0.690	A	0.10	2.49	1,9	0.145

FIG. 6.4.1 EFFECT OF PROPRANOLOL ON HEART RATE

and the mean data polynomials all of degree 5 for placebo, acute and chronic doses of propranolol. Looking at the mean data it is clear that polynomials of order less than three are inadequate. Increasing polynomial order fits the mean data for each dose more and more closely. Beyond polynomial order five it does so without altering the substantive conclusions to the medical questions at issue. Given the inefficiencies of the SAS (1982) Matrix procedure, with the consequent cost of program execution, computer analyses for polynomials of order greater than five have not been run.

The answers to the medical questions are that:

1. Propranolol lowers heartrate in both acute and chronic doses.
2. The effect of the acute dose is to lower the average unmedicated heartrate by approximately 12 beats per minute from 63 to 51 in the first two hours, an effect which is sustained beyond 11 hours; chronic doses produce a sustained but slightly variable lowering effect of around 14 beats per minute over the study period.
3. Propranolol in acute and chronic doses appears to dampen the usual increase in heartrate which accompanies ingestion of food, although a larger study with more statistical power would be desirable to substantiate this claim.

The analysis also raises a number of interesting statistical issues. The original ANOVA type examination of the data used by Maling and Dollery (1980) makes no allowance for model errors for individuals which at each dosage level are clearly autocorrelated and not identically distributed. It is pleasing then that the medical conclusions

of the original analysis are unaltered when using the more sophisticated polynomial models.

The r.c. polynomial model requires that the errors for each individual at each dose, about the random polynomial for that individual, be non-autocorrelated. While the nominal white noise test probabilities in Table 6.4.1 may leave this assumption in some doubt for some individuals, it is clear that the ability of this type of polynomial to make allowance for heteroscedasticity (see $\hat{\sigma}_{ii}$ of Table 6.4.1) gives it considerable advantage over ANOVA models for this data. The variation across individuals at each time point for a given treatment is not markedly dissimilar (cf. $\hat{\sigma}_{ii}$ of Table 6.4.1) and the white noise model residuals hypothesis for the polynomials based on mean data yields nominal chi square statistics of 3.01, 9.55, 7.30 respectively for placebo, acute and chronic doses, each based on lag 6 (and hence 6 degrees of freedom), leading to acceptance of these white noise hypotheses each at nominal type 1 error level of greater than 0.10. Given the significance of normality tests tabulated in Table 6.4.2, the assumptions necessary for mean data polynomial fits thus appear well satisfied. The possible violation of the white noise requirement for the r.c. polynomials would then appear not to be of critical importance, given the obvious similarity of the two types of polynomials at each dosage level, when plotted in Fig. 6.4.1.

It is difficult to draw general conclusions from analysis of one set of data, as the variation about overall trends is much larger in this case than if normally distributed elementary estimates from large scale repeated sample surveys were used as raw data. However, the polynomial models have advantages over the use of ANOVA in this case, and given normality assumptions and the recoverability of simulated polynomials even in the presence of high individual variability, similar

advantages seem more than plausible when comparing polynomial model estimates with conventional sample survey estimates for repeated surveys.

APPENDIX 1

THE POLYMACR SAS MACROS

GENERAL COMMENTS

POLYMACR SAS is a series of macros written in the macro language of the 1982 version of the Statistical Analysis System (SAS) package, for use on IBM systems under CMS (Conversational Monitoring System).¹ It is intended for the determination and fitting of an appropriate polynomial regression to time series data, possibly for a number of individuals, where the error correlation structure is known (MACRO INITPOL), for providing polynomial projections both including and not including model error estimates (MACRO POLPROJ), and for updating the polynomial regression coefficients given additional data (MACRO POLUPDT). A number of statistical tests are undertaken to aid model fitting (e.g. MACRO POLSTAT) and to test for constancy of the regression relationship over time. Optionally a further macro (MACRO ESTRCOV) allows fitting and testing of a random coefficient polynomial model where data is available on more than one individual, together with tests for the appropriateness of this model. MACRO ESTRCOV can be used where appropriate to provide an estimated error covariance matrix for the fitting of the initial appropriate polynomial regression.

Macros are invoked by using a percentage sign followed by the macro name and, in brackets, the value of the macro

¹. By appropriate changes to FILE DEFINITIONS the macros may also be run under other IBM operating systems.

parameters e.g. %POLSTAT(STLIB). No semicolon is necessary. If macro invocations are to be stored in a separate SAS file they should include the following statements previous to macro invocation:

```
CMS FI MACROIN DISK POLYMACR SAS A;
```

```
%INCLUDE MACROIN;
```

Further details on SAS macros and the SAS82 macro language are contained in the 1982 edition of the SAS manual (Basics).

Data set names, or library names, passed as parameters to macros must be SAS data sets and libraries respectively. Variables within such data sets can be user-defined, although the macros will use Y for the dependent variable and T for time, regardless of these user given variable names. Values of Y, T and correlations between the corresponding model error must be given to the macros, where required, in three separate SAS data sets, with the SAS data sets containing Y and T having one variable per observation. The SAS data set containing the correlations or covariances for MACRO INITPOL will, for each observation in turn, contain correlations (or covariances) with all other observations, and will therefore contain the same number of variables as observations. For MACRO POLPROJ and MACRO POLUPDT the SAS data set of correlations will contain the same number of observations as new data points, and as many variables as total data points, old plus new.

Further details are given below on the methods and use of each of these macros, including the necessary and optional parameters for their use.

The order of positional parameters for macros INITPOL, POLSTAT, POLPROJ, POLUPDT and ESTRCOV are given below. Note that all MACRO parameters are positional.

```
%MACRO    INITPOL (MAXDEGRE, TDAT, YDAT, INDDAT, SIGMDAT,  
                STATLIBN, LIBNAME, ODATSAVE, PLOTTYPE,  
                DATAPR) ;
```

```
%MACRO    POLSTAT (STATLIB2)
```

```
%MACRO    POLPROJ (LIBNAM3, PROJT, PROJIND, PROJSIGM, PLOTTYP3,  
                TDAT3, YDAT3, INDDAT3);
```

```
%MACRO    POLUPDT (NEWT, NEWY, NEWIND, NEWSIGM, LIBNAM4, UDSTATUS,  
                PLOTTYP4, TDAT4, YDAT4, SIGMDAT4, INDDAT4);
```

```
%MACRO    ESTRCOV (MAXDEG, TDTA, YDTA, INDDTA, OUTCOV, RESLIB,  
                SAMEX, REORD);
```

A note of caution should perhaps be sounded here. These macros make use of PROC MATRIX and the maximum number of data points which can be processed is therefore set at 181 (SAS 1982 (Statistics) p552). The amount of memory required and the time taken for execution increase significantly when large numbers of data points are processed.

Polynomial regression is therefore best used with short data series, both for this reason and because, for longer series, more sophisticated time series techniques are often available.

1 MACRO INITPOL

MACRO INITPOL is designed to allow the choice of an appropriate polynomial in time to fit initial data values in the dependent variable, given time t , possibly for a number of different individuals, assuming a known covariance or correlation structure in the errors to the polynomial model. While estimated covariances or correlations may be used, the chosen model will not then necessarily be best linear unbiased. To be more specific suppose that

$$\tilde{Y} = X\beta + e \quad (1)$$

where $\tilde{Y}$ is the vector of dependent variables ($L \times 1$)

β is a vector of coefficients to the monomials

$1, t, t^2, \dots, t^{p-1}$ and is ($p \times 1$)

X has ℓ -th row $(1 \quad t_\ell \quad t_\ell^2 \quad t_\ell^3 \quad \dots \quad t_\ell^{p-1})$ and is ($L \times p$)

e is a vector of model errors ($L \times 1$) with variance-

covariance matrix, or correlation matrix $\Sigma, (L \times L)$

Note that L may equal NT if data is collected from each of N individuals at T distinct time points, but that this is not a requirement for use of this macro. The ordinary least squares solution to the model of equation (1) involves inversion of the matrix $X^T X$ which is known to be ill conditioned (Froberg (1969)) i.e. small changes in the components of $X^T X$ lead to relatively large changes in $(X^T X)^{-1}$ and hence

in the least squares estimate of β . This problem remains for the generalised least squares solution even when $\Sigma \neq \sigma^2 I$, where I is the identity matrix and σ^2 an appropriate scale factor. Further discussion on this and related problems is to be found in Seber (1977:Chapter 8).

The solution to the problem of ill-conditioning adopted in this macro is to reparameterise the model (1) with respect to orthogonal polynomials (Phillips and Taylor (1973)).

$$\tilde{Y} = P\alpha + e \quad (2)$$

where $P\alpha = X\beta$

α is the vector of coefficients to the orthogonal polynomials, denoted $\psi_0(t), \psi_1(t), \psi_2(t), \dots, \psi_{p-1}(t)$ where $\psi_0(t)$ is some constant.

P has ℓ -th row $(\psi_0(t_\ell), \psi_1(t_\ell), \psi_2(t_\ell), \dots, \psi_{p-1}(t_\ell))$

Note that $\psi_0(t)$ is a constant for all t , that $\psi_j(t)$ is of degree j in t for $j = 0, 1, 2, \dots, p-1$, and that the $\{\psi_j(t)\}$ have the property that

$$\sum_{\ell=1}^L \psi_j(t_\ell) \psi_{j'}(t_\ell) = \begin{matrix} 0 & \text{if } j \neq j' \\ 1 & \text{if } j = j' \end{matrix}$$

i.e. $P^T P = I$, the identity matrix.

$P^T P$ is thus not ill-conditioned since $(P^T P)^{-1} = I$.

The generalised least squares estimate of α is then

$$\hat{\alpha} = (P^T \Sigma^{-1} P)^{-1} P^T \Sigma^{-1} \tilde{Y}$$

The macro defines a matrix CONV ($p \times p$) by

$$\hat{\beta} = \text{CONV } \hat{\alpha}$$

letting CONV = C for notational convenience. We note that

$$\begin{aligned} \tilde{Y} &= X\beta + e \\ &= X(CC^{-1})\beta + e \\ &= XC\alpha + e \end{aligned}$$

Thus from equation (2) $P = XC$ and therefore

$$C = (P^T X)^{-1} P^T P = (P^T X)^{-1}$$

It also follows that $\hat{\beta} = C\hat{\alpha}$, where

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \tilde{Y}$$

is the generalised least squares solution to equation (1).

The value of $(p-1)$ i.e. the degree of the polynomial fitted, is a required parameter for MACRO INITPOL, denoted &MAXDEGRE. A sequence of values, including the polynomial of largest plausible degree should be fitted using INITPOL. One invocation of INITPOL is required for each value of &MAXDEGRE. The choice between these models is aided by MACRO POLSTAT, discussed below, and by cusum and cusum of squares tests which test the constancy of the polynomial regression as data points beyond the p -th are included in the model. These are also discussed in more detail below.

MACRO INITPOL determines condition numbers of inverted matrices with respect to the infinity norm (Froberg 1969:93).

This condition number, defined for a square matrix G ($q \times q$) as the product of

$$\max_i \left(\sum_{j=1}^q |g_{ij}| \right) \quad \text{and} \quad \max_i \left(\sum_{j=1}^q |g^{(ij)}| \right)$$

where the ij -th element of G is g_{ij} and the ij -th element of G^{-1} is $g^{(ij)}$, gives an indication of ill-conditioning. Large condition numbers indicate ill-conditioning.

The macro also determines total sum of squares, error sum of squares and regression sum of squares for each value of &MAXDEGR, defined respectively as

$$\tilde{\mathbf{Y}}^T \Sigma^{-1} \tilde{\mathbf{Y}}, \tilde{\mathbf{Y}}^T (\Sigma^{-1} - \Sigma^{-1} \mathbf{P} (\mathbf{P}^T \Sigma^{-1} \mathbf{P})^{-1} \mathbf{P}^T \Sigma^{-1}) \tilde{\mathbf{Y}}$$

and

$$\tilde{\mathbf{Y}}^T (\Sigma^{-1} \mathbf{P} (\mathbf{P}^T \Sigma^{-1} \mathbf{P})^{-1} \mathbf{P}^T \Sigma^{-1}) \tilde{\mathbf{Y}}$$

together with their associated degrees of freedom.

Various plots are produced: a plot of the dependent variable against time with the fitted polynomial superimposed, of the model residuals against time, of the estimated versus the actual values of the dependent variable, and the models residuals against time. Each of these are plotted, noting the "individual" from which a particular datum is derived. The plot of the dependent variable against time with the superimposed polynomial may optionally be plotted using PROC GPLOT, to give a more elegant picture.

MACRO INITPOL also calculates a sequence of independent standardised residuals, distinct from the model residuals, for $\ell = p'+1, p'+2, \dots, L$. Details of the mathematics of this matrix calculation are given in Haslett (1985)(b). Under the null hypothesis of no change in β from the p' -th observation onward these $(L-p')$ residuals are distributed as $N(0,1)$, and can therefore be used for the cusum and cusum of squares tests of Brown et al (1975). The cusum (cumulative sum) test allows a statistical test for any

systematic drift in these residuals indicating a slowly changing polynomial regression. A cusum plot is drawn together with 95% upper and lower confidence intervals for the null hypothesis of no systematic drift.

Data points outside these bounds indicate that the null hypothesis of no systematic drift should be rejected at the 95% level. Where a cusum test is desired beginning at some point other than the $(p+1)$ st, this can be obtained by moving a cardboard replica of the wedge shape between the confidence intervals to that other point, while retaining its horizontal orientation. Points outside of the wedge again imply rejection of the null hypothesis at the 95% level. A cusum of squares (cumulative sum of squares test) is also undertaken (Brown et al (1975)) which allows testing of random drift from the fitted polynomial regression curve. CPLUS and CMINUS produced by MACRO INITPOL can be compared with tabulated values (Table.1. Durbin (1969)) (see also Table.1. below) for significance level α by comparing them with the tabulated value, C_0 , for significance level $\alpha/2$, and degrees of freedom n (or the linear interpolation between n_l and n_u as appropriate). Where either CPLUS or CMINUS exceeds C_0 the null hypothesis of no significant random fluctuation about the polynomial regression line is rejected, at the significance level α .

Cusum and cusum of squares tests should be disregarded if data from different individuals is used and the underlying polynomial signal is known to follow different polynomials for different individuals (e.g. different individual levels, random coefficient models with different coefficients for different individuals). To use cusum and cusum of squares tests in these cases MACRO INITPOL should be involved using as data the average values at each time point, with appropriate modifications to the model error variance-covariance matrix.

Table 1. Significance values for $c+$ and $c-$;

$$\text{pr}(c+ > \text{tabled value}) = \text{pr}(c- > \text{tabled value}) \approx 2\alpha$$

n	α					n	α				
	0.10	0.05	0.025	0.01	0.005		0.10	0.05	0.025	0.01	0.005
1	0.40000	0.45000	0.47500	0.49000	0.49500	41	0.14916	0.17213	0.19254	0.21667	0.23310
2	0.35044	0.44306	0.50855	0.56607	0.59596	42	0.14761	0.17034	0.19050	0.21436	0.23081
3	0.35477	0.41811	0.46702	0.53456	0.57990	43	0.14611	0.16858	0.18852	0.21212	0.22839
4	0.33435	0.39075	0.44641	0.50495	0.54210	44	0.14466	0.16688	0.18661	0.20995	0.22605
5	0.31556	0.37359	0.42174	0.47692	0.51576	45	0.14325	0.16524	0.18475	0.20785	0.22377
6	0.30244	0.35522	0.40045	0.45440	0.48988	46	0.14188	0.16364	0.18295	0.20581	0.22157
7	0.28991	0.33905	0.38294	0.43337	0.46761	47	0.14055	0.16208	0.18120	0.20383	0.21943
8	0.27828	0.32538	0.36697	0.41522	0.44819	48	0.13926	0.16058	0.17950	0.20190	0.21735
9	0.26794	0.31325	0.35277	0.39922	0.43071	49	0.13800	0.15911	0.17785	0.20003	0.21534
10	0.25884	0.30221	0.34022	0.38481	0.41517	50	0.13678	0.15769	0.17624	0.19822	0.21337
11	0.25071	0.29227	0.32894	0.37187	0.40122	51	0.13559	0.15630	0.17468	0.19645	0.21146
12	0.24325	0.28330	0.31869	0.36019	0.38856	52	0.13442	0.15495	0.17316	0.19473	0.20961
13	0.23639	0.27515	0.30935	0.34954	0.37703	53	0.13330	0.15363	0.17168	0.19305	0.20780
14	0.23010	0.26767	0.30081	0.33980	0.36649	54	0.13221	0.15235	0.17024	0.19142	0.20604
15	0.22430	0.26077	0.29296	0.33083	0.35679	55	0.13113	0.15110	0.16884	0.18982	0.20432
16	0.21895	0.25439	0.28570	0.32256	0.34784	56	0.13009	0.14939	0.16746	0.18828	0.20265
17	0.21397	0.24847	0.27897	0.31489	0.33953	57	0.12907	0.14870	0.16613	0.18677	0.20101
18	0.20933	0.24296	0.27270	0.30775	0.33181	58	0.12807	0.14754	0.16482	0.18529	0.19942
19	0.20498	0.23781	0.26685	0.30108	0.32439	59	0.12710	0.14641	0.16355	0.18385	0.19786
20	0.20089	0.23298	0.26137	0.29484	0.31784	60	0.12615	0.14530	0.16230	0.18245	0.19635
21	0.19705	0.22844	0.25622	0.28898	0.31149	62	0.12431	0.14316	0.15990	0.17973	0.19341
22	0.19343	0.22416	0.25136	0.28346	0.30552	64	0.12255	0.14112	0.15760	0.17713	0.19061
23	0.19001	0.22012	0.24679	0.27825	0.29989	66	0.12087	0.13916	0.15540	0.17464	0.18792
24	0.18677	0.21630	0.24245	0.27333	0.29456	68	0.11926	0.13728	0.15329	0.17226	0.18535
25	0.18370	0.21268	0.23835	0.26866	0.28951	70	0.11771	0.13548	0.15127	0.16997	0.18288
26	0.18077	0.20924	0.23445	0.26423	0.28472	72	0.11622	0.13375	0.14932	0.16777	0.18051
27	0.17799	0.20596	0.23074	0.26001	0.28016	74	0.11479	0.13208	0.14745	0.16566	0.17823
28	0.17533	0.20283	0.22721	0.25600	0.27582	76	0.11341	0.13048	0.14565	0.16363	0.17604
29	0.17280	0.19985	0.22383	0.25217	0.27168	78	0.11208	0.12894	0.14392	0.16167	0.17392
30	0.17037	0.19709	0.22061	0.24851	0.26772	80	0.11079	0.12745	0.14224	0.15978	0.17188
31	0.16805	0.19427	0.21752	0.24501	0.26393	82	0.10955	0.12601	0.14063	0.15795	0.16992
32	0.16582	0.19166	0.21457	0.24165	0.26030	84	0.10835	0.12462	0.13907	0.15619	0.16802
33	0.16368	0.18915	0.21173	0.23843	0.25683	86	0.10719	0.12327	0.13756	0.15449	0.16618
34	0.16162	0.18674	0.20901	0.23534	0.25348	88	0.10607	0.12197	0.13610	0.15284	0.16440
35	0.15964	0.18442	0.20639	0.23237	0.25027	90	0.10493	0.12071	0.13468	0.15124	0.16268
36	0.15774	0.18218	0.20387	0.22951	0.24718	92	0.10393	0.11949	0.13331	0.14970	0.16101
37	0.15590	0.18003	0.20144	0.22676	0.24421	94	0.10291	0.11831	0.13198	0.14820	0.15940
38	0.15413	0.17796	0.19910	0.22410	0.24134	96	0.10192	0.11716	0.13070	0.14674	0.15783
39	0.15242	0.17595	0.19684	0.22154	0.23857	98	0.10096	0.11604	0.12944	0.14533	0.15631
40	0.15076	0.17402	0.19465	0.21906	0.23589	100	0.10002	0.11496	0.12823	0.14396	0.15483

From BIOMETRIKA (1969) 56 1-15. "Tests for serial correlation in regression analysis based on the periodogram of least squares residuals". J. Durbin.

Once a polynomial model of appropriate degree has been fitted, the data sets produced by MACRO INITPOL should be saved using a permanent library name determining possible model projections (MACRO POLPROJ) or for later model updating (MACRO POLUPDT).

The macro parameters to be specified for MACRO INITPOL are as follows. Mandatory parameters are underlined.

&MAXDEGRE

specifies the maximum degree polynomial to be fitted in a particular macro run. i.e. (p-1) in equations (1) and (2) e.g. &MAXDEGRE set to one will fit a straight line; set to two it will fit a quadratic, etc.

&TDAT

specifies a SAS data set, possibly with a two level name, which contains the time data for each Y value i.e. value of the dependent variable. The first p time values in &TDAT must be distinct, where &MAXDEGRE = p-1.

The SAS data set containing the time data must contain no other variable. The naming of this variable is arbitrary but MACRO INITPOL will rename it T.

&YDAT

specifies the SAS data set (possibly with a two level name) which contains the dependent variable. This data set must be ordered in exactly the same order as the corresponding data in &TDAT. The SAS data set containing the dependent variable must contain no others. The naming of this variable is arbitrary but MACRO INITPOL will rename it Y.

&INDDAT

specifies the "individual" or grouping to which specific datum belong. It must be numeric and ordered in the same order as &YDAT, &TDAT (and &SIGMDAT if specified). Its use within MACRO INITPOL is entirely for plotting purposes. However where there is more than one individual, special attention should be paid to the appropriateness of cusum and cusum of squares tests as discussed previously. The naming of this variable is arbitrary although MACRO INITPOL will rename it IND. Values of IND greater than nine will cause IND not be used for plotting purposes. Values less than one are not allowed.

&SIGMDAT

contains the variance-covariance matrix of the polynomial model errors, or its estimate. Alternatively the model error correlation matrix or its estimate may be given. The covariances or correlations if specified must be given in the same order as &TDAT, &YDAT (and &INDDAT if specified). &SIGMDAT must contain as many variables as observations. The naming of these variables is arbitrary. If &SIGMDAT is not specified it is assumed that $\text{corr}(\underline{e}\underline{e}^T) = I$ where $\underline{e}$ is defined in equation (1) or (2) and I is the identity matrix of appropriate order.

&STATLIBN

specifies a library name in which sums of squares and associated degrees of freedom are accumulated for later use by MACRO POLSTAT, which tests for an appropriate degree polynomial via an F test. For a given set of data and a range of values of &MAXDEGRE, the value of &STATLIBN should remain fixed. The default value of &STATLIBN is WORK, and if the default value is used MACRO POLSTAT must be invoked during the same program run to

avoid loss of the information in the SAS data set containing the sums of squares. Usually the default setting for &STATLIBN is used only after the appropriate polynomial model has been fitted, and the invocation of MACRO INITPOL is being used purely to provide a series of SAS data sets for later projection or updating, in following program runs, using MACRO POLPROJ or POLUPDT. Setting &STATLIBN to a value other than WORK creates one observation in one SAS data set for each invocation of INITPOL. Do not use the same value of &STATLIBN (other than WORK in different program runs) for different raw data.

&LIBNAME

specifies a permanent library name to which data sets for later projection or updating by other macros. The default is WORK, and this should be used while testing data for its appropriate polynomial, unless ample disk space for storage of permanent SAS data sets is available. See also &ODATSAVE. The data set &LIBNAME..INITPLOT may be later used, if desired, to produce a PROC GPLOT type plot of the fitted polynomial and the data from which it was derived.

&ODATSAVE

if set to NOSAVE does not save the original data (&YDAT, &TDAT, &INDDAT, &SIGMDAT) within the SAS library &LIBNAME. This is particularly useful where such data sets (particularly &SIGMDAT) are large and possible duplication viewed as undesirable. It does however require explicit specification of the data sets containing Y, T, the individual indicator variable, and the correlations or covariances if MACRO POLPROJ or POLUPDT are to be invoked later. The default value causes &TDAT, &YDAT, &INDDAT and &SIGMDAT to be saved within the SAS library &LIBNAME, unless &LIBNAME = WORK.

&PLOTTYPE

if set to GPLOT will produce a SASGRAPH plot, using PROC GPLOT and your installation plotter, of the data points with the fitted polynomial superimposed in green. At some installations only one SASGRAPH plot per SAS program is possible. The default value gives an ordinary PROC PLOT type plot within the programs output listing.

&DATAPR

if set to PRINT provides a more extensive printout of interim calculations and includes a printout of the original data in &TDAT, &YDAT, &INDDAT and &SIGMDAT. The default setting is to suppress printing of this extra information.

2. MACRO POLSTAT

MACRO POLSTAT provides a statistical procedure for testing for the appropriate order of the fitted polynomial (Seber 1977:217). It should only be invoked after several invocations of MACRO INITPOL. One such invocation of MACRO INITPOL should fit a polynomial of degree sufficiently large that the actual polynomial is of lesser degree. Under the assumption of normally distributed model errors MACRO POLSTAT carries out a series of F tests to test whether the addition of extra terms to the polynomial (i.e. increasing values of &MAXDEGRE) provide a statistically significant improvement to the polynomial model fit. Under the normality assumption for model errors the test is exact when the variance-covariance error matrix is known, and is otherwise approximate. A series of F statistics are calculated as

$$F_{j-j', L-m} = \left(\frac{ESS_j - ESS_{j'}}{j' - j} \right) / \left(\frac{ESS_{m+1}}{L-m-1} \right)$$

where ESS denotes error sum of squares

$j > j'$

m is the maximum value of &MAXDEGRE used in
invoking MACRO INITPOL

This provides a statistical test of whether the transition from a polynomial of degree $(j'-1)$ provides a significantly better fit than a polynomial of degree $(j-1)$.

The probability of such an F statistic is printed, together with the test statistic in descending order the values of &MAXDEGRE used in invoking MACRO INITPOL. An FPROB value is printed beside each value of DF_RSS i.e. the degrees of freedom for residual sum of squares i.e. &MAXDEGRE +1. FPROB is the probability under the null hypothesis that given the polynomial is of degree &MAXDEGR=j that an F statistic of this magnitude would be generated by fitting a polynomial of degree j' rather than j . Note that the value of FPROB for the transition from a polynomial of degree j to one of degree j' ($j' > j$) appears beside the value of DF_RSS= $j-1$ not beside the value of DF_ESS= $j'-1$.

Particular care should be taken in using these F statistics and their associated values of FPROB where the model error variance-covariance matrix (or correlation matrix) is re-estimated for different values of &MAXDEGRE for use in MACRO INITPOL. Negative F values are possible under these circumstances, because a change of error covariance or correlation matrix alters the underlying

metric of the space used to determine total, regression and error sums of squares. Here the total sum of squares used for calculating F statistics is always that from the polynomial of degree m defined above.

MACRO POLSTAT should be used in conjunction with model residual plots, cusum tests and cusum of squares tests to determine a polynomial of degree appropriate to the data. The parameter associated with MACRO POLSTAT is given below:

&STATLIB2

is the name of the SAS data library, containing the data set of sums of squares and associated degrees of freedom, to be analysed by the F statistic procedure.

The value of &STATLIB2 will be identical to the value of &STATLIBN used in invoking MACRO INITPOL. The default value of &STATLIB2 is WORK, when the macro may be invoked either by %POLSTAT or &POLSTAT() or %POLSTAT(WORK).

3 MACRO POLPROJ

MACRO POLPROJ is designed for determining polynomial projections, possibly including model error projections, for a polynomial model determined via MACRO INITPOL (and MACRO POLSTAT). Two types of projection are undertaken, the first being pure polynomial projection and the second being polynomial projection incorporating a model error estimate based on known covariance or correlation properties of future observations. Where the correlation between past and future data values is zero there is no distinction between these two forms of projection.

The pure polynomial projection at times $\tilde{t}(h+1)$ are given by

$$\tilde{Q}(h+1) = x_{h+1} \hat{\beta}_h \quad (3)$$

where $x_{h+1} = \begin{pmatrix} 1 & \tilde{t}(h+1) & \tilde{t}^2(h+1) & \dots & \tilde{t}^{p-1}(h+1) \end{pmatrix}$

and $\tilde{t}^j(h+1)$ = the elementwise j -th power of the elements of $\tilde{t}(h+1)$

$\hat{\beta}_h = (X_h^T \Sigma^{-1} X_h)^{-1} X_h^T \Sigma^{-1} \tilde{Y}_h = (X_h^T \Sigma_h^{-1} X_h)^{-1} X_h^T \Sigma_h^{-1} \tilde{Y}_h$ say, is the best

linear estimate of β from equation (1) and MACRO INITPOL, given the data to time h .

$\tilde{Q}(h+1)$ is named YPOLPROJ in the MACRO POLPROJ output.

The polynomial projection $\hat{\tilde{Y}}^{(h)}(h+1)$, incorporating an estimated error term, is given by

$$\hat{\tilde{Y}}^{(h)}(h+1) = \tilde{x}_{h+1} \hat{\beta}_h + \Sigma_{h,h+1}^T \Sigma_h^{-1} (\tilde{y}_h - \tilde{x}_h \hat{\beta}_h) \quad (4)$$

where the model of equation (1) is rewritten as

$$\tilde{y}_h = \tilde{x}_h \beta + e_h$$

with $\text{var-cov}(e_h) = \Sigma_h$ with best linear unbiased estimate of β for the data $\tilde{y}_h$, given by $\hat{\beta}_h$ and where

$$\text{var-cov}(e_{h+1}) = \Sigma_{h+1} = \begin{pmatrix} \Sigma_h & \Sigma_{h,h+1} \\ \Sigma_{h,h+1}^T & \Sigma_{h+1,h+1} \end{pmatrix}$$

incorporates the known correlations of future observations into the model error covariance matrix.

The first term in equation (4) is $\tilde{Q}(h+1)$ the pure polynomial projection and the second is an estimate of the model error term at times $t(h+1)$. Further details and the derivation of these results are given in Bibby and Toutenberg (1977) and in Haslett (1985)(b).

As was the case with MACRO INITPOL, condition numbers of inverted matrices are calculated with respect to the infinity norm.

MACRO POLPROJ produces a plot of the fitted polynomial and its projection up to the maximum value of T specified, with known data values plotted together with any projections including error terms calculated via equation (4). Projected values are either plotted using "individual" numbers in a different colour if a SASGRAPH plot is requested, or using the corresponding alphabet letter for PROC PLOT type plots (e.g. A corresponds to individual 1, B to individual 2, etc).

Parameters which must or may be specified for MACRO POLYPROJ are given below. Those that are mandatory are underlined.

&LIBNAM3

specifies the SAS library name used when MACRO INITPOL was invoked with appropriate parameters (including a statistically appropriate value of &MAXDEGRE). The default value of &LIBNAM3, namely WORK, requires that the appropriate invocation of MACRO INITPOL mentioned above has been performed using &LIBNAME=WORK in the same program run.

&PROJT

specifies the SAS data set containing the vector of times at which projections are required. It may be a two level name. This SAS data set must contain only one variable per observation. The variable may be given any name but MACRO POLPROJ will rename the variable T.

&PROJIND

specifies the "individual" or grouping to which projected values will be deemed to belong. It must be numeric and sorted in the same order as &PROJT (and &PROJSIGM if specified). Values greater than nine will result in data plots without indicator variables. Values less than one are not allowed. Only one variable is allowed in the data set containing the indicator variable, which may be arbitrarily named. MACRO POLPROJ will rename the variable IND or PIND. The variable is used solely for plotting purposes.

&PROJSIGM

specifies the SAS data set, which may have a two level name, and which contains the matrix of covariances or correlations of model errors for the required projections, both with model

errors of past observations and with model errors for the projections themselves. If &SIGMDATA of MACRO INITPOL contained correlations so must &PROJSIGM, with a parallel comment applying for covariances. The observations in &PROJSIGM must be sorted in the same order as those for &PROJ and the variables in the same order as the data in &TDAT (or &YDAT) merged with the data in &TPROJ i.e. the data in &PROJSIGM viewed as a matrix contains the submatrix

$$\begin{pmatrix} \Sigma_{h,h+1}^T & \Sigma_{h+1,h+1} \end{pmatrix}$$

of Σ_{h+1} of equation (4). If &PROJSIGM is not specified for a particular invocation of MACRO POLPROJ, pure polynomial projection only is performed.

&PLOTTP3

if set to GPLOT produces a PROC GPLOT type plot of the projected polynomial together with known and projected data values. Projected data values are printed by their value of IND (see &PROJIND) or if IND is not specified as diamonds of a different colour to known values, except where these values correspond to pure polynomial projection. The

default value gives a PROC PLOT type plot within the program output listing with projected values coded as letters (A=1, B=2, etc.) to distinguish them from known values. At some installations only one PROC GPLOT type plot is possible per program run.

&TDAT3

should be specified only if &ODATSAVE=NOSAVE was specified when MACRO INITPOL was invoked. In this case only, &TDAT3 should be set to the value of &TDAT used in MACRO INITPOL, and this is optional where MACRO INITPOL was invoked in the same program run.

&YDAT3

should be specified only if &ODATSAVE=NOSAVE was specified when MACRO INITPOL was invoked. In this case only, &YDAT3 should be set to the value of &YDAT used in MACRO INITPOL, and this is optional where MACRO INITPOL was invoked in the same program run.

&INDDAT3

should be specified only if &ODATSAVE=NOSAVE was specified together with a value for &INDDAT when MACRO INITPOL was invoked. In this case only, &INDDAT3 should be set to the value of &INDDAT used in MACRO INITPOL, and this is

optional where MACRO INITPOL was invoked
in the same program run.

4 MACRO POLUPDT

MACRO POLUPDT allows initial generalised least square polynomial coefficients estimates, $\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$ in the notation of equation (1), to be updated to incorporate additional data with known covariance or correlation properties. Again, as for MACRO INITPOL numerical calculations are carried out with respect to orthogonal polynomials, and estimated covariances or correlations may be used. Estimated correlations do not however produce optimal solutions in the sense of best linear unbiasedness (see for example Judge et al (1980)). The constancy of $\hat{\beta}$ given $\hat{\beta}_h$ and its updated version $\hat{\beta}_{h+1}$ are tested via χ^2 statistic, cusum and cusum of squares tests. The Chi-square test is based on the result that, for normally distributed model errors and $n(h+1)$ additional data points, the difference of the error sums of squares of the original and updated models has a chi-square distribution on $n(h+1)$ degrees of freedom, under the null hypothesis of constant $\hat{\beta}$ (Haslett (1985)(b)). The cusum and cusum of squares tests parallel those for MACRO INITPOL and the interested reader is referred to the comments made there, and to Brown et al (1975).

The updating formulae used by MACRO POLUPDT make explicit use of Theorems 2, 3 and 4 of Haslett (1985)(b). The interested reader is referred to that paper, a modification made here is to reparameterise the polynomial model with respect to the polynomials orthogonal with respect to the original data for calculation purposes, and to later update these orthogonal polynomials to make them orthogonal with respect to the entire set of data. This allows MACRO POLUPDT to be used for further updating of parameter estimates at a later invocation.

Parameters which must or may be given for MACRO POLYUPDT are given below. Optional parameters are not underlined.

&NEWT

specifies the SAS data set containing the vector of times at which data is available for model parameter updating. It may be a two level name. This data set must contain only one variable per observation. The variable may be given any name but MACRO POLUPDT will rename the variable T. If notice is to be taken of cusum and cusum of squares tests it is generally best if &NEWT is sorted by T.

&NEWY

specifies the SAS data set (possibly with a two level name) which contains the additional values of the dependent variable. This SAS data set must contain only one variable per observation which may be given any name but which will be renamed Y by MACRO POLUPDT. &NEWT, &NEWY, (together with &NEWIND and &NEWSIGM) must be sorted in the same order at the time of invocation of MACRO POLUPDT.

&NEWIND

specifies the "individual" to which projected values have been deemed to belong. It must be numeric and sorted in the same order as &NEWT. Values greater than nine will result in data plots without indicator variables. Values less than one are prohibited. Only one variable is allowed in the data set &NEWIND and this variable may be arbitrarily named. MACRO POLUPDT will rename the variable IND and use it solely for plotting purposes.

&NEWSIGM

specifies the SAS data set, which may have a two level name, and which contains the matrix of covariances or correlations of the model errors for the new data, both with the model errors of

past observations and the model errors for the new data themselves. If &SIGMDATA of MACRO INITPOL contained correlations then so must &NEWSIGM, with a parallel comment applying for covariances. The default value of &NEWSIGM (i.e. if &NEWSIGM is not specified) is to set the variance of model errors for observations to one, and covariances between different model errors to zero.

It we let the correlation or covariance of the "old" data (i.e. the data already incorporated into the polynomial parameter estimates) be Σ_h and that for all the data, "old" and "new" to Σ_{h+1} , then

$$\Sigma_{h+1} = \begin{pmatrix} \Sigma_h & \Sigma_{h;h+1} \\ \Sigma_{h;h+1}^T & \Sigma_{h+1,h+1} \end{pmatrix} \quad \text{say}$$

and the data in &NEWSIGM when viewed as a matrix should contain $(\Sigma_{h;h+1}^T : \Sigma_{h+1,h+1})$. When &NEWSIGM is not specified it is assumed that the model error covariances $(\Sigma_{h;h+1}^T : \Sigma_{h+1,h+1})$ are given by $(0:I)$.

`&LIBNAM4`

specifies the SAS library name used when MACRO INITPOL was invoked with appropriate parameters, including a statistically appropriate value of `&MAXDEGRE`. The default value of `&LIBNAM4`, namely WORK, requires that the appropriate invocation of MACRO INITPOL mentioned above has occurred using `&LIBNAME=WORK` in the same program run.

`&UDSTATUS`

if set to UPDATE updates the whole of the SAS library `&LIBNAM4` to include modifications made necessary by the new data. If not set to UPDATE almost no change is made to `&LIBNAM4` (see below). The default is to not update the SAS library data sets. Regardless of whether `&UDSTATUS` is set to update the data set `&LIBNAM4..UPDTPLT` is produced or updated. This contains the information necessary to generate a SASGRAPH type plot of the updated polynomial and the data both "old" and "new". (Note that if `&UDSTATUS` is set to UPDATE then `&LIBNAM4..INITPLOT` will also be updated to include the additional data.)

`&PLOTYP4`

if set to GPLOT produces a PROC GPLOT type plot of the updated polynomial together with the data both "old" and "new". Values are plotted by their value of IND (see `&INDDAT4`). The default gives a PROC PLOT type plot within the program output listing. At some installations only one PROC GPLOT type plot is possible per program run. One method of surmounting this difficulty, if desired, is outlined under `&UDSTATUS` above.

`&TDAT4`

should be specified only if `&ODATSAVE=NOSAVE` was specified when MACRO INITPOL was invoked. In this case only, `&TDAT4` should be set to the value of `&TDAT` used in MACRO INITPOL, and this is optional where MACRO INITPOL is invoked in the same program run if `&UDSTATUS` is set to UPDATE, `&TDAT4` will be updated regardless of the data library or data set to which it belongs, and the variable within it renamed T.

`&YDAT4`

should be specified only if `&ODATSAVE=NOSAVE` was specified when MACRO INITPOL was invoked. In this case only, `&YDAT4` should be set to the value of `&YDAT` used in MACRO INITPOL, and this is optional where MACRO INITPOL is invoked in the same program run. If `&UDSTATUS` is set to UPDATE, `&YDAT4` will

be updated regardless of the data library or data set to which it belongs, and the variable within it renamed Y.

&SIGMDAT4

should be specified only if &ODATSAVE=NOSAVE was specified when MACRO INITPOL was invoked.

In this case only, &SIGMDAT used in MACRO INITPOL and this is optional whose MACRO INITPOL is invoked in the same program run.

If &UDSTATUS is set to UPDATE, &SIGMDAT4 will be updated regardless of the data library or data set to which it belongs and the variables within it renamed COL1, COL2, ..., COLk where k is the total number of data points "old" plus "new".

&INDDAT4

should be specified only if &ODATSAVE=NOSAVE was specified together with a value for &INDDAT when MACRO INITPOL was invoked. In this case only, &INDDT4 should be set to the value of &INDDAT used in MACRO INITPOL and this is optional where MACRO INITPOL is invoked in the same program run. If &UDSTATUS is set to UPDATE, &INDDAT4 will

be updated regardless of the SAS library or
date set to which it belongs, and the variable
within it renamed IND.

5 MACRO ESTRCOV

MACRO ESTRCOV estimates the error variance-covariance matrix for a random coefficient model of the Swamy type. (Judge et al 1980:347.)

$$\text{i.e. } y_{it} = \sum_{k=1}^p \beta_{ki} x_{kit} + e_{it} \quad \begin{array}{l} i = 1, 2, \dots, N' \\ t = 1, 2, \dots, T \end{array}$$

where the $\{\beta_{ki}\}$ are random variables. Note that the model requires a value of y for every individual $i = 1, 2, \dots, N'$ at every time $t = 1, 2, \dots, T$. These times t may be different for each i , but must be T in number for each i . Further, in the macro, a polynomial model of degree $p-1$ fitted to the data implies that $x_{kit} = t_i^{k-1}$. It is assumed that $t_i = t$ for all i if the option &SAMEX=SAMEXMAT is specified. The macro options are further explained below, and within the macro code.

For the i -th individual the associated $(p \times 1)$ response vector β_i (i.e. the polynomial coefficients for individual i) is regarded as a random vector drawn from a probability distribution, with mean $\bar{\beta}$ and covariance matrix Δ . For the i -th individual the model can thus be written as

$$\hat{\tilde{Y}}_i = X_i(\bar{\beta} + \tilde{\mu}_i) + \tilde{e}_i \quad i = 1, 2, \dots, N'$$

where $\hat{\tilde{Y}}_i$ and X_i contain, respectively, the observations on the dependent variable Y , and on the explanatory variables namely the monomials in t (i.e. $1, t, t^2, \dots, t^{p-1}$), and $\bar{\beta} + \mu_i = \beta_i$. Then the model may be written in matrix notation as

$$\tilde{Y} = X\bar{\beta} + Z\mu + e$$

where $\tilde{Y} = (\tilde{Y}_1^T, \tilde{Y}_2^T, \dots, \tilde{Y}_{N'}^T)^T \quad (NT \times 1)$

$$X = (X_1^T, X_2^T, \dots, X_{N'}^T)^T \quad (NT \times p)$$

$$Z = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_{N'} \end{bmatrix} \quad (NT \times Np) \text{ block diagonal}$$

$$e = (e_1^T, e_2^T, \dots, e_{N'}^T)^T \quad (NT \times 1)$$

$$\mu = (\mu_1^T, \mu_2^T, \dots, \mu_{N'}^T)^T \quad (Np \times 1)$$

and $\bar{\beta}$ is as defined previously $(p \times 1)$

If the data in &YDTA, &TDTA and &INDDTA had been previously sorted by individual and subsorted

by time, the covariance matrix of the composite disturbance, $\underline{z}_i^u + \underline{e}_i$, is $\emptyset$ where $\emptyset$ block diagonal with i -th diagonal block given by

$$\emptyset_{ii} = X_i \Delta X_i^T + \sigma_{ii} I$$

The variance-covariance matrix of the data in the order actually in &YDTA, &TDTA and &INDDTA given, is called Σ and is the appropriate reordering of $\emptyset$. Note that $E(\underline{e}_i \underline{e}_i^T) = \sigma_{ii} I$, and that $E(\underline{e}_i \underline{e}_j^T) = 0$ for $i \neq j$, is assumed. This assumption is checked, for data from evenly spaced time points for each individual, by the PROC ARIMA section of the macro.

Let

$$\underline{b}_i = (X_i^T X_i)^{-1} X_i^T \underline{y}_i$$

i.e. $\underline{b}_i$ is the best linear unbiased estimator of $\underline{\beta}_i$ given the data for individual i only. Then Swamy (1971) shows that we may estimate σ_{ii} , and Δ by

$$\hat{\sigma}_{ii} = \frac{\tilde{\underline{e}}_i^T \tilde{\underline{e}}_i}{T-p}$$

$$\hat{\Delta} = \frac{S_b}{N-1}$$

where

$$\tilde{\underline{e}}_i = \hat{\underline{y}}_i - X_i \underline{b}_i$$

and

$$S_b = \sum_{i=1}^N \underline{b}_i \underline{b}_i^T - \frac{1}{N} \sum_{i=1}^N \underline{b}_i \sum_{i=1}^N \underline{b}_i^T$$

$\hat{\sigma}_{ii}$ is unbiased, but $\hat{\Delta}$ trades the unbiasedness property to ensure the more critical property of non-negative definiteness.

All least squares calculations rather than being calculated with respect to X and Z are undertaken with respect to the corresponding orthogonal polynomials (Seber (1977:214-238)) to avoid numerical computation problems. Condition numbers of inverted matrices with respect to the infinity norm (Froberg 1969:93) are calculated to give an indication of any numerical instabilities in the least squares parameter fitting calculations.

The macro parameters to be specified for the macro ESTRCOV are as follows:

&MAXDEG

specifies the maximum degree polynomial to be fitted. This should be the same as the value of &MAXDEGRE used in the corresponding invocation of MACRO INITPOL.

&TDTA

specifies the SAS data set containing the values of the dependent variable, and consists of NT observations on this variable.

&YDTA

specifies the SAS data set containing the values of the dependent variable, and consists of NT observations on this variable.

&INDDTA

specifies the SAS data set containing the indicator variable for NT observations. This indicator variable specifies to which individual the corresponding observation for time and the dependent variable belongs. N the maximum value of this indicator variable must be at least two.

NOTE: &TDTA, &YDTA and &INDDTA must each be sorted in the same order.

&OUTCOV

specifies the SAS data set, which may be a two level name, to which the estimate of the matrix Σ is output during execution of the macro. This is usually used as the input covariance matrix data-set &SIGMDATA for MACRO INITPOL.

&RESLIB

specifies the SAS library name to which the residuals $\tilde{e}_i$ are output for each i. Such data sets will have the name &RESLIB..RESID&I e.g. for &RESLIB = RES and &I = 3 the data set will be RES.RESID3. The

default value for &RESLIB is WORK. A permanent library name allows for further possible checking of the assumptions $E(\tilde{e}_i \tilde{e}_i^T) = \sigma_{ii} I$ and $E(\tilde{e}_i \tilde{e}_j^T) = 0$ for $i \neq j$.

&SAMEX

when &SAMEX is set to SAMEXMAT the assumption is that all individuals $i = 1, 2, \dots, N$ each have T observations at the same times $t = 1, 2, \dots, T$. This implies $X_i = W$ say, for all $i = 1, 2, \dots, N$ with corresponding modifications to the formulae given previously. Significant reductions in execution time can be expected by specifying SAMEXMAT, when appropriate.

&REORD

when &REORD = NOREORD the matrix $\emptyset$ is set to Σ without reordering of rows and columns i.e. the assumption is that the data in &YDTA, &TDTA and &INDDTA are sorted by individual and subsorted by time. Significant reductions in execution time are possible when &REORD = NOREORD is appropriate and specified.

TEST PROGRAM FOR POLYMACR MACROS

```
OPTIONS NOSOURCE2 MPRINT ;

CMS FILEDEF POLYMAC DISK POLNMAC SAS A;

%INCLUDE POLYMAC;

DATA TDATA ;
  INPUT T          /* VECTOR OF TIMES */ ;
  CARDS;
          8
          10
          12
          16
          20
          ;
DATA YDATA ;
  INPUT Y          /* VECTOR OF OBSERVATIONS */ ;
  CARDS;
          0.88
          1.22
          1.64
          2.72
          3.96
          ;
          * MACRO INVOCATIONS FOLLOW ;

%INITPOL(2,TDATA,YDATA)

%INITPOL(3,TDATA,YDATA)

%POLSTAT
```

```

DATA PRJ_T;
  INPUT T /* TIME AT WHICH PROJECTION REQUIRED */ ;
  CARDS;
    60
  ;
DATA PRJ_SIGM;
  INPUT COL1 COL2 COL3 COL4 COL5 COL6
        /* CORRELATION SUBMATRIX */ ;
  CARDS;
    0    0    0    0    0    1
  ;
        * MACRO INVOCATION FOLLOWS ;

```

%POLPROJ(,PRJ_T,,PRJ_SIGM)

```

DATA NEW_T;
  INPUT T /* VECTOR OF NEW TIMES */ ;
  CARDS;
    30
    40
    60
    100
  ;
DATA NEW_Y;
  INPUT Y /* VECTOR OF NEW OBSERVATIONS */ ;
  CARDS;
    7.66
    11.96
    21.56
    43.16
  ;
DATA NEW_SIGM;
  INPUT COL1 COL2 COL3 COL4 COL5 COL6 COL7 COL8 COL9
        /* CORRELATION SUBMATRIX */ ;
  CARDS;
    0    0    0    0    0    1    0    0    0
    0    0    0    0    0    0    1    0    0
    0    0    0    0    0    0    0    1    0
    0    0    0    0    0    0    0    0    1
  ;
        * MACRO INVOCATION FOLLOWS ;

```

%POLUPDT(NEW_T,NEW_Y,,NEW_SIGM,,UPDATE)

* THE POLYMACR MACROS

WRITTEN BY STEPHEN HASLETT
INSTITUTE OF STATISTICS AND OPERATIONS RESEARCH
VICTORIA UNIVERSITY OF WELLINGTON
PRIVATE BAG
WELLINGTON
NEW ZEALAND

OCTOBER 1983 ;

```
%GLOBAL TDATA YDATA SIGMDATA INDDATA GLIBNAME GSTATLIB;  
  
%MACRO INITPOL(MAXDEGRE,TDAT,YDAT,INDDAT,SIGMDAT,  
STATLIBN,LIBNAME,ODATSAVE,PLOTTYPE,DATAPR);
```

```
%LET TDATA=&TDAT;;  
%LET YDATA=&YDAT;;  
%LET SIGMDATA=&SIGMDAT;;
```

* THE ABOVE THREE LET STATEMENTS PRODUCE GLOBAL
-MACRO- VARIABLES FOR USE IN THE -MACRO-POLPROJ-
AND -MACRO-POLUPDT- ;

```
%IF &STATLIBN=%STR() %THEN %LET STATLIBN=WORK;;  
%LET GSTATLIB=&STATLIBN;;
```

* THIS STATEMENT ALLOWS &STATLIBN TO BE THE DEFAULT
LIBRARY NAME FOR THE LATER -MACRO-POLSTAT-.
SEE ALSO THE GLOBAL STATEMENT PRECEEDING THIS
-MACRO-;
* &STATLIBN IS THE NAME OF THE SAS LIBRARY, GIVEN BY
THE USER, AND INTENDED TO CONTAIN SUMMARY STATISTICS
FROM CONTINUED CALLS OF THIS -MACRO- . TO ADD FURTHER
INFORMATION FROM A PARTICULAR CALL, &STATLIBN MUST
OF COURSE REMAIN UNCHANGED FROM THE LIBRARY NAME
GIVEN IN PREVIOUS CALLS;

```
%IF &LIBNAME =%STR() %THEN %LET LIBNAME =WORK;;  
%LET GLIBNAME=&LIBNAME;;
```

* THE DEFAULT SAS LIBRARY NAME FOR DATA SETS
CREATED BY THIS -MACRO- IS WORK . IF THE
DATA IS TO BE LATER UPDATED, A PERMANENT
DATA LIBRARY NAME WILL BE REQUIRED, BUT
THIS WILL ONLY BE NECESSARY AFTER THE ORDER

OF THE POLYNOMIAL, APPROPRIATE TO THE DATA USED
FOR POLYNOMIAL FITTING, HAS BEEN DECIDED ;

%IF &ODATSAVE =%STR() %THEN %LET ODATSAVE =SAVE;;

* &ODATSAVE FOR ORIGINAL DATA SAVE.

- 1) FOR &LIBNAME=WORK (THE DEFAULT LIBRARY NAME)
&ODATSAVE IS INOPERATIVE.
- 2) IF SET TO SAVE, THE DEFAULT SETTING FOR
&LIBNAME^=WORK, IT SAVES &TDAT,&YDAT, AND &SIGMDAT
WITHIN &LIBNAME..TDATA, &LIBNAME..YDATA, AND
&LIBNAME..SIGMDATA RESPECTIVELY.
- 3) IF SET TO NOSAVE WHEN &LIBNAME^=WORK, IT IS
NECESSARY TO EXPLICITLY SPECIFY THE DATA SETS
CONTAINING T-DATA, Y-DATA, AND SIGMA-DATA IN
THE -MACRO-POLPROJ- AND -MACRO-POLUPDT- .
THE SETTING NOSAVE REDUCES REQUIRED DISK
STORAGE SPACE FOR SAS LIBRARY NAME, &LIBNAME,
AT THE COST OF REQUIRING EXPLICIT SPECIFICATION
OF THE DATA SETS CONTAINING T-DATA, Y-DATA AND
SIGMA-DATA IN THESE LATER -MACROS- ;

* THE DEFAULT FOR &SIGMDAT, IF NOT SPECIFIED
EXPLICITLY, IS THE IDENTITY MATRIX OF APPROPRIATE
DIMENSION;

* &PLOTTYPE=GLOT GIVES A SASGRAPH PLOT,
FOR THE PLOT OF FITTED POLYNOMIAL AND THE
DATA, Y, AGAINST TIME, T.
THE DEFAULT IS A PROC PLOT TYPE PLOT;

* &DATAPR=PRINT PROVIDES A MORE EXTENSIVE PRINTOUT
OF BOTH DATA AND INTERIM CALCULATIONS ;

```
PROC MATRIX;  
  FETCH Y DATA=&YDAT;;  
  OUTPUT Y DATA=YDT2(RENAME=(COL1=Y));
```

```
PROC SUMMARY DATA=YDT2;  
  VAR Y;  
  OUTPUT OUT=NROWYDT N=NROWY;
```

```
DATA YTEM;  
  SET NROWYDT;  
  CALL SYMPUT('NRWY',NROWY);  
RUN;
```

PROC MATRIX FUZZ;

FETCH T DATA=&TDAT;;

OUTPUT T DATA=TDATTEMP(RENAME=(COL1=T));

* T IS THE VECTOR OF TIMES;

FETCH Y DATA=&YDAT;;

OUTPUT Y DATA=YDATTEMP(RENAME=(COL1=Y));

* Y IS THE VECTOR OF THE DEPENDENT VARIABLE;

%IF &SIGMDAT=%STR() %THEN %DO;

%STR(SIGMA=I(NROW(T)));

OUTPUT SIGMA DATA=SIGMDATA;

%LET SIGMDAT=SIGMDATA;

%LET SIGMDATA=&SIGMDAT;

%END;

%ELSE

%STR(FETCH SIGMA DATA=&SIGMDAT;;);

* SIGMA IS THE CORRELATON MATRIX (OR COV)
MATRIX FOR ERRORS IN THE LINEAR MODEL;

%IF &INDDAT=%STR() %THEN %DO;

IND = J(NROW(T),1,1);

OUTPUT IND DATA=INDDTEMP(RENAME=(COL1=IND));

%LET INDDAT=INDDTEMP;;

%END;

%ELSE

%DO;

FETCH IND DATA=&INDDAT;;

OUTPUT IND DATA=INDDTEMP(RENAME=(COL1=IND));

%END;

INDMAX=MAX(IND);

OUTPUT INDMAX DATA=INDMAX1(RENAME=(COL1=INDMAX));

%LET INDDATA=&INDDAT;;

* IND IS A NUMERIC VARIABLE WHICH INDICATES
TO WHICH 'INDIVIDUAL' THE Y VALUES
BELONG;

MAXDEGR = &MAXDEGRE;;

* MAXDEGR IS THE MAXIMUM DEGREE ORTHOGONAL
POLYNOMIAL TO BE FITTED ;

NOTE THE LINEAR MODEL BEING FITTED IS

NOTE

NOTE $Y = X \cdot BETA + E$ WITH VAR-COV(E)=SIGMA

NOTE

NOTE THIS MODEL IS FITTED VIA A REPARAMETRISATION WITH RESPECT TO;
NOTE POLYNOMIALS, ORTHOGONAL WITH RESPECT TO THE COMPONENTS OF THE;

NOTE VECTOR , T. THE REPARAMETRISED MODEL IS

NOTE

NOTE $Y = P \cdot ALPHA + E$ WITH VAR-COV(E) = SIGMA;

NOTE

NOTE

```

NOTE          NOTE THAT    B=CONV*A      X*CONV=P      ;
NOTE          BETA=CONV*ALPHA      ;
NOTE          B = BETAHAT    A = ALHAHAT      ;
NOTE          ;
NOTE          THUS    Y = X      *      BETA + E      ;
NOTE          I.E.    Y = X*(CONV* INV(CONV))*BETA + E      ;
NOTE          I.E.    Y = (X*CONV)*(INV(CONV))*BETA) + E      ;
NOTE          I.E.    Y = P      *      ALPHA      + E      ;
NOTE          ;
NOTE          ;
NOTE          ;

```

```

NOTE    TO INDICATE THE EXTENT OF ILL-CONDITIONING OF INVERTED MATRICES,;
NOTE    THE CONDITION NUMBER OF EACH INVERTED MATRIX IS CALCULATED ;
NOTE    WITH RESPECT TO THE INFINITY NORM, ;
NOTE    REF : FROBERG,C.E. 'INTRODUCTION TO NUMERICAL ANALYSIS';
NOTE    ADDISON-WESLEY 2ND EDITION 1969 P93. ;
NOTE    ;
NOTE    ;

```

NOTE MAXDEGR IS THE ORDER OF THE MAXIMUM DEGREE POLYNOMIAL BEING FITTED;

```

PRINT MAXDEGR;
OUTPUT MAXDEGR DATA=&LIBNAME..MAXDDATA;

```

```
%IF &DATAPR=PRINT %THEN %DO;
```

```

NOTE    T IS THE VECTOR OF TIMES CORRESPONDING TO THE DATA, Y, WITH;
NOTE    ERROR VAR-COV MATRIX ,SIGMA ;

```

```

PRINT T;
PRINT Y;
PRINT SIGMA;

```

```
%END;
```

```

* X FORMED ITERATIVELY BELOW IS THE MATRIX OF MONOMIALS
(1, T, T**2 ETC.) CORRESPONDING TO P. X(I,J)=T(I)**(J-1) ;

```

```

X = J(NROW(T),1,1) ;
DO DEGR = 1 TO MAXDEGR BY 1;
X = X || T##DEGR ;
END;

```

```

NOTE    X IS THE MATRIX OF MONOMIALS, 1,T,T**2,...T*MAXDEGR;
NOTE    CORRESPONDING TO THE OBSERVATIONS IN T.;
NOTE    X(I,J)=T(I)**(J-1) ;

```

```

PRINT X;
OUTPUT X DATA=&LIBNAME..XDATA;

```

```

P = ORPOL(T,MAXDEGR);

%IF &DATAPR=PRINT %THEN %DO;

NOTE    P CONSISTS OF THE POLYNOMIALS ORTHOGONAL          ;
NOTE    WITH RESPECT TO T, EVALUATED AT EACH OF          ;
NOTE    THE COMPONENTS OF T. P HAS MAXDEGR+1 COLS.        ;
NOTE    P(I,J)=VALUE OF (J-1)TH ORTHOG POLY AT T(I) ;

                                PRINT P;

%END;                                OUTPUT P DATA=&LIBNAME..PDATA;

                                SIGMAINV = SOLVIT(SIGMA,I(NROW(SIGMA)));

%IF &DATAPR=PRINT %THEN %DO;

NOTE    SIGMAINV IS THE INVERSE OF THE CORRELATION OR COVARIANCE MATRIX,;
NOTE    SIGMA, OF THE ERRORS OF THE OBSERVATIONS, Y ;

                                PRINT SIGMAINV;

%END;                                OUTPUT SIGMAINV DATA=&LIBNAME..INVSIGDT;

NOTE    CNSIGMA IS THE CONDITION NUMBER OF SIGMA WITH RESPECT TO THE;
NOTE    INFINITY NORM ;

                                CNSIGMA = MAX(ABS(SIGMA)(,+)) * MAX(ABS(SIGMAINV)(,+));

                                PRINT CNSIGMA;

NOTE    A IS THE LEAST SQUARES SOLUTION WITH RESPECT TO THE ORTHOGONAL;
NOTE    POLYNOMIALS. IT IS A VECTOR OF THE ESTIMATED COEFFICIENTS OF;
NOTE    THE ORTHOGONAL POLYNOMIALS;

                                %STR(INVPTSP = INV(P%'*SIGMAINV*P)) ;
                                %STR(PTS = P%'*SIGMAINV ) ;
                                A = INVPTSP * PTS * Y ;

                                PRINT A;
                                OUTPUT A DATA=&LIBNAME..ADATA(RENAME=(COL1=A));

```


* THESE COEFFICIENTS, A, ARE NOW CONVERTED TO THE MONONOMIAL
COEFFICIENTS , DENOTED BY B ;

```
%STR(XTSP = X%'*SIGMAINV*P) ;
    INVXTSP = INV(XTSP) ;
```

NOTE CNXTSP IS THE CONDITION NUMBER OF (XT*SIGMAINV*P) WITH RESPECT;
NOTE TO THE INFINITY NORM ;

```
CNXTSP = MAX(ABS(XTSP)(,+))*MAX(ABS(INVXTSP)(,+));
    PRINT CNXTSP;
```

NOTE B=CONV*A X*CONV=P BETA=CONV*ALPHA;

```
%STR(INVCONV = INVXTSP*(X%'*SIGMAINV*X));
    CONV = INV(INVCONV);
    PRINT CONV;
    OUTPUT CONV DATA=&LIBNAME..CONVDATA;
```

* NOTE: X'*SIGMAINV*X IS ILL-CONDITIONED.
HENCE CONV IS NOT CALCULATED AS
CONV=INV(X'*SIGMAINV*X)*XTSP ;

NOTE CNCONV IS THE CONDITION NUMBER OF CONV WITH RESPECT;
NOTE TO THE INFINITY NORM ;

```
CNCONV = MAX(ABS(CONV)(,+))*MAX(ABS(INVCONV)(,+));
    PRINT CNCONV;
```

NOTE B IS THE VECTOR OF ESTIMATED COEFFICIENTS OF THE MONONOMIALS;
NOTE 1, T, T**2,...,T**MAXDEGR , WITH VARIANCE VARB;

```
B = CONV * A ;
```

```
    PRINT B;
    OUTPUT B DATA=&LIBNAME..BDATA(RENAME=(COL1=B));
```

```
%STR(VARB = CONV * INVXTSP * CONV%');
    PRINT VARB;
    OUTPUT VARB DATA=&LIBNAME..VARBDATA;
```

* THE ESTIMATED VALUES FOR Y ARE NOW CALCULATED;

```
YEST = P * A ;
RESID= Y-YEST ;
```

```
    OUTPUT YEEST DATA=YEESTDATA(RENAME=(COL1=YEEST));
    OUTPUT RESID DATA=RESIDATA(RENAME=(COL1=RESID));
```

```

NOTE    THE SUM OF SQUARES FOR THE POLYNOMIAL MODEL OF MAXIMUM    ;
NOTE    DEGREE, MAXDEGR, IS NOW CALCULATED .                      ;
NOTE    TSS=TOTAL SUM OF SQUARES (NO MEAN CORRECTION)             .;
NOTE    DF_TSS=DEGREES OF FREEDOM OF TOTAL SUM OF SQUARES        .;
NOTE    RSS=REGRESSION SUM OF SQUARES (INCLUDING MEAN CORRECTION);
NOTE    DF_RSS=DEGREES OF FREEDOM OF REGRESSION SUM OF SQUARES;
NOTE    ESS=ERROR SUM OF SQUARES                                  .;
NOTE    DF_ESS=DEGREES OF FREEDOM OF ERROR SUM OF SQUARES        .;

```

```

%STR(TSS = Y%' * SIGMAINV * Y) ;
DF_TSS = NROW(Y);
%STR(ESS = Y%' * (SIGMAINV - PTS%'*INVPTSP*PTS) * Y) ;
DF_ESS = NROW(Y)-MAXDEGR;
RSS = TSS - ESS ;
DF_RSS = MAXDEGR;

PRINT TSS DF_TSS RSS DF_RSS ESS DF_ESS;
RESULT = MAXDEGR || TSS || DF_TSS || RSS || DF_RSS || ESS || DF_ESS;
OUTPUT RESULT DATA=REGTEST (RENAME=(COL1=DEGREE
                                   COL2=TSS COL3=DF_TSS
                                   COL4=RSS COL5=DF_RSS
                                   COL6=ESS COL7=DF_ESS));

```

* VALUES OF THE APPROXIMATING POLYNOMIAL IN T ARE NOW CALCULATED
OVER THE RECORDED RANGE OF T ,FOR LATER USE IN PLOTTING
ROUTINES ;

```

D1T= 0:200 ;
%STR(D1 = D1T%') ;
D2 = J(201,1,1) ;
INCR = (MAX(T)-MIN(T))/200;
D3 = MIN(T)*D2 + INCR*D1;
    * D3 CONSISTS OF 201 ELEMENTS RANGING FROM
    MIN(T) TO MAX(T) BY EQUAL INCREMENTS OF INCR;

```

```

XTEST=J(201,1,1);
DO DEGREE=1 TO MAXDEGR BY 1;
XTEST = XTEST || D3##DEGREE;
END;

```

* EACH COLUMN OF XTEST CONSISTS OF ONE OF THE
MONONOMIALS 1,X,X**2....,X**MAXDEGR.
XTEST IS 201*(MAXDEGR+1);

```

YTEST = XTEST * B;
    * YTEST ARE THE VALUES OF THE APPROXIMATING
    POLYNOMIAL AT EACH ELEMENT OF D3;

```

```

OUTPUT YTEST DATA=YESTDAT2(RENAME=(COL1=YEST));
OUTPUT D3 DATA=D3DATA(RENAME=(COL1=T));

```

* THE FOLLOWING SECTION OF CODE DETERMINES A SET OF INDEPENDENT RESIDUALS (FOR J=MAXDEGR+2 TO NROW(Y)) WHICH ARE THE COMPONENTS OF THE VECTOR W_H;

```
%LET INDRESNO=0;
%LET ODDRESNO=0;
```

```
%IF %EVAL(&NRWY-&MAXDEGRE-1)>=1 %THEN %DO;
```

```
TEMPVAL2=MAXDEGR+2;
TEMPVEC2=TEMPVAL2:NCOL(SIGMA);
TEMPVAL1=MAXDEGR+1;
TEMPVEC1=1:TEMPVAL1;
SIGH1H1 = SIGMA(TEMPVEC2,TEMPVEC2);
SIGH_H1T = SIGMA(TEMPVEC2,TEMPVEC1);
SIGH_H1 = SIGMA(TEMPVEC1,TEMPVEC2);
SIGH_ = SIGMA(TEMPVEC1,TEMPVEC1);
INVSIGH = SOLVIT(SIGH,I(NROW(SIGH)));
NOTE CNSIGH IS THE CONDITION NUMBER OF SIGH WITH;
NOTE RESPECT TO THE INFINITY NORM ;
CNSIGH=MAX(ABS(SIGH)(,+))*MAX(ABS(INVSIGH)(,+));
PRINT CNSIGH;
```

```
YH = Y(TEMPVEC1,);
YH1 = Y(TEMPVEC2,);
PH = ORPOL(T(TEMPVEC1,1),MAXDEGR);
XH = X(TEMPVEC1,);
XH1 = X(TEMPVEC2,);
CONV2 = SOLVIT(XH,PH);
PH1 = XMULT(XH1,CONV2);
```

```
G1H_A = INVSIGH*SIGH_H1;
%STR(ZH_P = PH1%' - PH%'*G1H_A) ;
%STR(PHTSPH=XMULT(PH%',' ,XMULT(INVSIGH,PH)));
INPHTSPH =SOLVIT(PHTSPH,I(NROW(PHTSPH)));
G1H_BP = INPHTSPH * ZH_P;
```

```
C1H_A = SIGH1H1-SIGH_H1T*G1H_A;
%STR(C1H_B = G1H_BP%'*ZH_P);
%STR(R1H_A = YH1 - G1H_A%'*YH);
%STR(R1H_B = G1H_BP%'*PH%'*INVSIGH*YH);
COR_RES = R1H_A - R1H_B;
NOTE COR_RES = R1H_A-R1H_B I.E. THE CORRELATED RECURSIVE RESIDUALS;
PRINT COR_RES;
```

```
M1 =C1H_A+C1H_B ;
```

```
EIGEN EIGVAM1 EIGVEM1 M1 ;
NOTE EIGVAM1 ARE THE EIGENVALUES OF M1 WHERE M1=C1H_A+C1H_B ;
NOTE M1 IS THE COVARIANCE OR CORRELATION MATRIX OF COR_RES;
```

```
PRINT EIGVAM1;
EIGVAM2 =EIGVAM1 ##0.5;
EIGVEM2 =EIGVEM1 ;
```

* M2 HAS THE PROPERTY THAT $M2 \cdot M2 = M1$. THIS IS IMPLIES AND IS IMPLIED JOINTLY BY $EIGVAM2=EIGVAM1##0.5$ AND $EIGVEM2=EIGVEM1$;

```
%STR(M2 = XMULT(EIGVEM2,XMULT(DIAG(EIGVAM2),EIGVEM2%')));
INVM2 = SOLVIT(M2,I(NROW(M2)));
```

NOTE THE COMPONENTS OF W_H ARE MUTUALLY INDEPENDENT $N(0,1)$ RANDOM;
 NOTE VARIABLES PROVIDED THE LINEAR MODEL ERRORS ARE MULTIVARIATE ;
 NOTE NORMAL, NOT NECESSARILY INDEPENDENT. ;

```
W_H = INVM2 * (R1H_A-R1H_B);
```

```
PRINT W_H;
OUTPUT W_H DATA=&LIBNAME..INDRESID(RENAME=(COL1=W));
```

NOTE NRW_H IS THE NUMBER OF RECURSIVE RESIDUALS;

```
NRW_H = NROW(W_H);
PRINT NRW_H;
```

NOTE VARW_H IS AN ESTIMATED VARIANCE FOR W_H . ;
 NOTE UNDER THE NULL HYPOTHESIS THAT THE DATA FOLLOWS;
 NOTE A POLYNOMIAL OF THE SPECIFIED ORDER, AND ASSUMING THAT;
 NOTE SIGMA IS KNOWN, THE TRUE VARIANCE OF W_H EQUALS ONE. ;
 NOTE THE W_H ARE THEN INDEPENDENT $N(0,1)$, PROVIDED THE W_H ARE;
 NOTE NORMALLY DISTRIBUTED, AND $CHISTAT=(VARW_H \cdot NRW_H)$ IS THEN ;
 NOTE CHI-SQUARE ON NRW_H DEGREES OF FREEDOM. ;
 NOTE THE PROBABILITY OF AN OBSERVED STATISTIC AT LEAST AS ;
 NOTE LARGE AS CHISTAT UNDER THE NULL HYPOTHESIS GIVEN THE;
 NOTE NORMALITY ASSUMPTION IS DENOTED BY CHIPROB;

```
VARW_H = %STR(W_H% ' * W_H)/NRW_H;
PRINT VARW_H;
```

```
CHISTAT =VARW_H#NRW_H;
CHIPROB = 1-PROBCHI(CHISTAT,NRW_H);
PRINT CHISTAT CHIPROB NRW_H;
```

```
INDRESN=ROUND(NROW(X)-(MAXDEGR+1));
ODDRESN=ROUND(MOD(INDRESN,2));
OUTPUT INDRESN DATA=INDRESND(RENAME=(COL1=INDRESN));
OUTPUT ODDRESN DATA=ODDRESND(RENAME=(COL1=ODDRESN));
```

```
PNUM=J(NROW(W_H),1,TEMPVAL1);
OUTPUT PNUM DATA=&LIBNAME..PNUM(RENAME=(COL1=PNUM));
TNUM=J(NROW(W_H),1,NROW(X));
```

```
OUTPUT TNUM DATA=&LIBNAME..TNUM(RENAME=(COL1=TNUM));

%END;

RUN;

%IF %EVAL(&NRWY-&MAXDEGRE-1)>=1 %THEN %DO;

    DATA INDRESN2;
        SET INDRESND;
        CALL SYMPUT('INDRESNO',INDRESN);
    RUN;

    DATA ODDRESN2;
        SET ODDRESND;
        CALL SYMPUT('ODDRESNO',ODDRESN);
    RUN;

%END;

* THE PROC APPEND BELOW APPENDS TEST STATISTICS FOR THE CURRENT
  VALUE OF &MAXDEGRE, TO TEST STATISTICS FOR OTHER CALLS OF
  -MACRO-INITPOL- IN THE CURRENT RUN, FOR POSSIBLE
  LATER CALCULATION OF F STATISTICS TO DETERMINE THE APPROPRIATE
  ORDER OF THE FITTED POLYNOMIAL.
  F STATISTICS ARE CALCULATED AND TABULATED BY -MACRO-POLSTAT-;

PROC APPEND BASE=&STATLIBN..TESTSTAT DATA=REGTEST;

%* IF &LIBNAME^=WORK AND &ODATSAVE^=NOSAVE THE FOLLOWING DATA
  STEPS CREATE COPIES OF Y , T AND SIGMA ;

%IF %EVAL(&LIBNAME^=WORK AND &ODATSAVE^=NOSAVE)=1 %THEN %DO;

    DATA &LIBNAME..TDATA;
        SET &TDAT;

    DATA &LIBNAME..YDATA;
        SET &YDAT;

    DATA &LIBNAME..SIGMDATA;
        SET &SIGMDAT;

    DATA &LIBNAME..INDDATA;
        SET &INDDAT;

%END;
```

```
%IF &LIBNAME=WORK %THEN %DO;

    DATA TDATA;
        SET &TDAT;

    DATA YDATA;
        SET &YDAT;

    DATA SIGMDATA;
        SET &SIGMDAT;

    DATA INDDATA;
        SET &INDDAT;

                                %END;

* THE FOLLOWING DATA STEPS AND PROCEDURES SET UP AND PRODUCE PLOTS
  OF THE DATA, RESIDUALS TO THE MODEL AND THE APPROXIMATING
  POLYNOMIAL;

DATA &LIBNAME..SUMMDATA;
    MERGE TDATA YDATA INDDATA YESDATA RESIDATA;
    DROP ROW;

PROC PRINT DATA=&LIBNAME..SUMMDATA;

PROC PLOT DATA=&LIBNAME..SUMMDATA;
    PLOT Y*T=IND
        YEST*T=IND
        RESID*T=IND
        Y*YEST=IND
        RESID*Y=IND;

PROC UNIVARIATE DATA=&LIBNAME..SUMMDATA PLOT FREQ NORMAL;
    VAR RESID;

DATA PLOTDAT2;
    SET &LIBNAME..SUMMDATA;
    KEEP Y T IND;

DATA PLOTDAT3 ;
    MERGE YESDAT2 D3DATA;
    Y=YEST;
    IND=.;
    KEEP Y T IND;

DATA &LIBNAME..INITPLOT;
    SET PLOTDAT2 PLOTDAT3;
```

```
%IF &PLOTTYPE=GLOT %THEN %DO;
```

```
PROC SORT DATA=&LIBNAME..INITPLOT;  
  BY IND;
```

```
DATA INDMAX2;  
  SET INDMAX1;  
  CALL SYMPUT('MAXIND',INDMAX);  
  RUN;
```

```
OPTIONS NOTERMIAL NOSYMBOL DEVICE=&SYSDEVIC;;  
TITLE FITTED POLYNOMIAL PLOTTED AGAINST TIME;  
SYMBOL1 C=GREEN V=NONE I=JOIN;  
SYMBOL2 C=BLACK V=DIAMOND I=NONE;  
%LET LIM1=%EVAL(&MAXIND+1);  
%IF &MAXIND>1 %THEN %DO COUNT=3 %TO &LIM1;  
  %LET COUNT2=%EVAL(&COUNT-1);  
  %IF &MAXIND<10 %THEN  
    %STR(SYMBOL&COUNT C=BLACK V=&COUNT2 I=NONE;);  
  %ELSE  
    %STR(SYMBOL&COUNT C=BLACK V=DIAMOND I=NONE;);  
%END;  
FOOTNOTE1 IF DATA CONSISTS OF MORE THAN ONE INDIVIDUAL, THE FIRST IS;  
FOOTNOTE2 CODED AS A DIAMOND AND THE REMAINDER BY THE CORRESPONDING ;  
FOOTNOTE3 NUMBER, IN BLACK. THE FITTED POLYNOMIAL IS THE LINE;  
FOOTNOTE4 PLOTTED IN GREEN. ;
```

```
PROC GLOT DATA=&LIBNAME..INITPLOT;;  
  %IF &MAXIND<10 %THEN  
    %STR(PLOT Y*T=IND / OVERLAY );;  
  %ELSE  
    %STR(PLOT Y*T=IND / OVERLAY NOLEGEND;);  
  RUN;  
  %END;  
  %ELSE %DO;
```

```
PROC SORT DATA=&LIBNAME..INITPLOT;;  
  BY DESCENDING IND;
```

```
PROC PLOT DATA=&LIBNAME..INITPLOT;  
  PLOT Y*T=IND / OVERLAY;  
  RUN;  
  %END;
```

* CUSUM AND CUSUM OF SQUARES TESTS ARE UNDERTAKEN BELOW;

```
%IF &INDRESNO^=0 %THEN %DO;
```

```
DATA CUSUMDT1;
```

```
MERGE &LIBNAME..INDRESID
      &LIBNAME..PNUM
      &LIBNAME..TNUM;
WTOT + W;
R = _N_+PNUM;
CUSUM95U = (0.948*(2*R + TNUM - 3*PNUM)) / SQRT(TNUM-PNUM);
CUSUM95L = -CUSUM95U;

PROC PRINT DATA=CUSUMDT1;
  RUN;

TITLE3 CUSUM PLOT OF WTOT WITH 95 PERCENT CONFIDENCE INTERVALS;
TITLE4      VERSUS OBSERVATION NUMBER, R ;

PROC PLOT DATA=CUSUMDT1;
  PLOT WTOT*R='*'
        CUSUM95U*R='.'
        CUSUM95L*R='.' /HZERO OVERLAY;

RUN;

TITLE3; TITLE4;

%END;

%IF &INDRESNO>=2 %THEN %DO;

  DATA CUSUMDT2;
    MERGE &LIBNAME..INDRESID
          &LIBNAME..PNUM
          &LIBNAME..TNUM;
    R = _N_+PNUM;
    WSQTOT + W**2;
    M = (TNUM-PNUM)/2;

  PROC SORT DATA=CUSUMDT2;
    BY DESCENDING R;

  DATA CUSUMDT3;
    SET CUSUMDT2;
    IF _N_=1 THEN WSQTMAX=WSQTOT;
    RETAIN WSQTMAX;
    SR = WSQTOT/WSQTMAX;
    IF _N_=1 THEN DELETE;

  PROC SORT DATA=CUSUMDT3;
    BY R;

  %IF &ODDRESNO=0 %THEN %DO;

    DATA CUSUMDT4;
      SET CUSUMDT3;
```



```
SRADJ = SR - (_N_/(TNUM-PNUM));  
N = M-1;  
KEEP SRADJ SR R N;
```

```
PROC MATRIX;  
  FETCH RAW DATA=CUSUMDT4;  
  SRADJ = RAW(,3);  
  SR = RAW(,2);  
  R = RAW(,1);  
  N = RAW(1,4);  
  CPLUS=MAX(SRADJ);  
  CMINUS=MAX(-SRADJ);
```

```
NOTE  CPLUS AND CMINUS ARE THE STATISTICS FOR THE CUSUM OF ;  
NOTE  SQUARES TEST GIVEN IN  BROWN ET AL, JRSS(B) 37 NO2 1975  ;  
NOTE  P155.  FOR THE SIGNIFICANCE LEVEL ALPHA, TABULATED VALUES;  
NOTE  IN  DURBIN, BIOMETRIKA 56 1969 P4 TABLE 1, GIVEN C0, THE  ;  
NOTE  VALUE IN THE TABLE AT N  AND ALPHA/2.                      ;  
NOTE  IF EITHER CPLUS OR CMINUS EXCEEDS C0 THEN THE HYPOTHESIS;  
NOTE  OF CONSTANT REGRESSION COEFFICIENTS IS REJECTED AT THE  ;  
NOTE  SIGNIFICANCE LEVEL ALPHA.  AN EYEBALL TEST IS ALSO        ;  
NOTE  RECOMMENDED BY  BROWN ET AL 1975  TO GIVE AN INDICATION  ;  
NOTE  OF HOW CLOSE EACH OF THE COMPONENTS OF SRADJ IS TO C0  ;
```

```
PRINT  CPLUS CMINUS N SRADJ SR R;
```

```
%END;
```

```
%IF &ODDRESNO=1 %THEN %DO;
```

```
DATA CUSUMDT4;  
  SET CUSUMDT3;  
  SRADJ = SR - (_N_/(TNUM-PNUM));  
  N_U = M-0.5;  
  N_L = M-1.5;  
  KEEP SRADJ SR R N_L N_U;
```

```
PROC MATRIX;  
  FETCH RAW DATA=CUSUMDT4;  
  SRADJ = RAW(,3);  
  SR = RAW(,2);  
  R = RAW(,1);  
  N_L = RAW(1,5);  
  N_U = RAW(1,4);  
  CPLUS=MAX(SRADJ);  
  CMINUS=MAX(-SRADJ);
```

```
NOTE  CPLUS AND CMINUS ARE THE STATISTICS FOR THE CUSUM OF ;  
NOTE  SQUARES TEST GIVEN IN  BROWN ET AL, JRSS(B) 37 NO2 1975  ;  
NOTE  P155.  FOR THE SIGNIFICANCE LEVEL ALPHA, TABULATED VALUES;  
NOTE  IN  DURBIN, BIOMETRIKA 56 1969 P4 TABLE 1, GIVEN C0, THE  ;
```

NOTE VALUE LINEARLY INTERPOLATED FROM THE TABLE FOR ALPHA/2, ;
 NOTE BETWEEN N_L AND N_U. ;
 NOTE IF EITHER CPLUS OR CMINUS EXCEEDS C0 THEN THE HYPOTHESIS;
 NOTE OF CONSTANT REGRESSION COEFFICIENTS IS REJECTED AT THE ;
 NOTE SIGNIFICANCE LEVEL ALPHA. AN EYEBALL TEST IS ALSO ;
 NOTE RECOMMENDED BY BROWN ET AL 1975 TO GIVE AN INDICATION ;
 NOTE OF HOW CLOSE EACH OF THE COMPONENTS OF SRADJ IS TO C0 ;

PRINT CPLUS CMINUS N_L N_U SRADJ SR R;

%END;

%END;

%MEND INITPOL;

%MACRO POLSTAT(STATLIB2);

%IF &GSTATLIB=%STR() %THEN %LET GSTATLIB=WORK;;
 %IF &STATLIB2=%STR() %THEN %LET STATLIB2=&GSTATLIB;;

* THE DEFAULT VALUE OF &STATLIB2 IS THE VALUE OF &STATLIBN WHEN
 -MACRO-INITPOL- WAS LAST INVOKED DURING THE CURRENT PROGRAM
 RUN. IF -MACRO-INITPOL- HAS NOT BEEN SO INVOKED, THE DEFAULT
 VALUE OF &STATLIB2 IS WORK ;

* THE DATA STEP, AND PROC PRINT BELOW PRODUCE AND
 PRINT TEST STATISTICS (F TESTS) TO TEST THE SIGNIFICANCE OF
 THE ADDITION OF EXTRA TERMS TO THE ORDER OF THE FITTED
 POLYNOMIAL. THE ERROR SUM OF SQUARES AND TOTAL SUM OF SQUARES
 FOR THE HIGHEST ORDER POLYNOMIAL FITTED TO THIS STAGE IS USED
 TO CALCULATE THE F TEST. NEGATIVE F STATISTICS CAN SOMETIMES
 RESULT WHEN DIFFERENT ESTIMATED ERROR COVARIANCE MATRICES ARE USED
 FOR DIFFERENT VALUES OF DEGREE, AS THIS ALTERS THE METRIC OF THE
 UNDERLYING SPACE AND ALTERS SUM OF SQUARES CALCULATIONS ACCORDINGLY.
 THE TSS FOR THE MAXIMUM VALUE OF DEGREE IS USED AS THE TSS FOR ALL
 OTHER VALUES OF DEGREE IN CALCULATING F STATISTICS.

$$F = ((RSSADJ - LAGRSSAJ) / (DF_RSS - LAGDFRSS)) / (ESSMIN / DFESSMIN)$$

 DFNUMER=DF_RSS-LAGDFRSS DFDENOM=DFESSMIN;

PROC SORT DATA=&STATLIB2..TESTSTAT;
 BY DESCENDING DEGREE;

```
DATA &STATLIB2..SUMMSTAT;
  SET &STATLIB2..TESTSTAT;

  IF _N_=1 THEN DO;
    ESSMIN=ESS; DFESSMIN=DF_ESS; TSSMAXD=TSS;
    F=.; FPROB=.;
    DFNUMER=.; DFDENOM=.;
    OUTPUT;
  END;

  RETAIN ESSMIN DFESSMIN TSSMAXD;
    RSSADJ=TSSMAXD-ESS;
    LAGRSSAJ=LAG(RSSADJ);
    LAGDFRSS=LAG(DF_RSS);

  IF _N_>1 THEN DO;
    F=((RSSADJ-LAGRSSAJ)/(DF_RSS-LAGDFRSS))/(ESSMIN/DFESSMIN);
    DFNUMER=LAGDFRSS-DF_RSS; DFDENOM=DFESSMIN;
    IF F > 0 THEN FPROB=PROBF(F,DFNUMER,DFDENOM);
    ELSE FPROB=0;
    OUTPUT;
  END;

PROC PRINT DATA=&STATLIB2..SUMMSTAT;

RUN;

%MEND POLSTAT;

%MACRO POLPROJ(LIBNAM3,PROJT,PROJIND,PROJSIGM,
               PLOTTYP3,TDAT3,YDAT3,INDDAT3);;

* -MACRO-POLPROJ- VIA PROC MATRIX DETERMINES POLYNOMIAL
  PROJECTIONS BASED ON THE CHOSEN MODEL AT A GIVEN FUTURE TIME ,WHEN
  ERROR CORRELATION WITH THE DATA ALREADY COLLECTED IS ALREADY KNOWN.
  THE MODEL MUST HAVE PREVIOUSLY BEEN CHOSEN USING -MACRO-INITPOL-.
  TWO TYPES OF PROJECTION ARE UNDERTAKEN, THE FIRST BEING PURE
  PROJECTION OF THE POLYNOMIAL, THE SECOND INCORPORATING ,IN ADDITION,
  AN ESTIMATE OF THE APPROPRIATE ERROR IN THE LINEAR MODEL, GIVEN
  CORRELATIONS WITH PAST OBSERVATIONS ;

  * &PROJT IS THE SAS DATASET CONTAINING VALUES OF T FOR WHICH A
    PROJECTION IS REQUIRED;
  * &PROJSIGM IS THE SAS DATASET CONTAINING THE TRANSPOSED MATRIX
```

OF CORRELATIONS OF ERRORS AT TIMES TPROJ WITH ERRORS AT TIMES (VEC(T)',TPROJ');

- * &PROJIND IS THE SAS DATASET CONTAINING VALUES OF A NUMERIC INDICATOR VARIABLE RELATING THE 'INDIVIDUALS' TO THEIR PROJECTED Y VALUES. IT MAY BE LEFT MISSING ESPECIALLY FOR PURE POLYNOMIAL PROJECTION;

%IF &GLIBNAME=%STR() %THEN %LET GLIBNAME=WORK;;

%IF &LIBNAM3=%STR() %THEN %LET LIBNAM3=&GLIBNAME;;

- * &LIBNAME3 IS BY DEFAULT THE LIBRARY NAME USED WHEN -MACRO-INITPOL- WAS LAST INVOKED, DURING THIS PROGRAM RUN. THAT LIBRARY NAME MAY BE WORK .
IF &LIBNAM3 IS SPECIFIED IT IS ASSUMED -MACRO-INITPOL- HAS ALREADY BEEN INVOKED, POSSIBLY IN A PREVIOUS RUN WITH &LIBNAME^=WORK, AND THE RESULTING SAS DATASETS ARE AVAILABLE TO THIS MACRO. NOTE THAT &GLIBNAME IS SET TO &LIBNAME BY -MACRO-INITPOL-. THE INVOCATION OF -MACRO-INITPOL- SHOULD, OF COURSE, HAVE FITTED THE POLYNOMIAL MODEL OF THE APPROPRIATE MAXIMUM DEGREE;

- * &PLOTTP3=GLOT GIVES A SASGRAPH PLOT,
FOR THE PLOT OF FITTED POLYNOMIAL AND THE DATA, Y, AGAINST TIME, T.
THE DEFAULT IS A PROC PLOT TYPE PLOT;

- * SPECIFICATION OF THE SAS DATASET CONTAINING THE MATRIX, SIGMA (THE MATRIX OF COVARIANCES OR CORRELATIONS OF THE ERRORS OF THE LINEAR MODEL AT TIMES, T, AT WHICH Y IS KNOWN) IS UNNECESSARY FOR THIS MACRO. THE INVERSE OF SIGMA IS HOWEVER CONTAINED IN THE SAS LIBRARY &LIBNAM3 IN THE SAS DATASET &LIBNAM3..INVSIGDT AND IS USED BY THIS MACRO;

PROC MATRIX FUZZ;

FETCH TPROJ DATA=&PROJT;;

- * TPROJ IS THE VALUES OF T FOR WHICH A PROJECTION IS REQUIRED;

FETCH MAXDEGR DATA=&LIBNAM3..MAXDDATA;

- * MAXDEGR IS THE MAXIMUM DEGREE OF THE POLYNOMIAL PREVIOUSLY FITTED TO THE DATA;

FETCH CONV DATA=&LIBNAM3..CONVDATA;

- * B = CONV*A;

FETCH A DATA=&LIBNAM3..ADATA;

- * A IS THE VECTOR OF L-S COEFFICIENTS OF THE POLYNOMIALS, ORTHOGONAL WITH RESPECT TO TIME, BASED ON THE VECTOR, T ;

FETCH B DATA=&LIBNAM3..BDATA;

- * B IS THE VECTOR OF L-S COEFFICIENTS OF THE MONOMIALS, 1,T,T**2,...,T**MAXDEGR;

```
%IF &TDAT3=%STR() %THEN %LET TDAT3=&LIBNAM3..TDATA;

FETCH T DATA=&TDAT3;;
      OUTPUT T DATA=TDATTEMP(RENAME=(COL1=T));
      * T IS THE VECTOR OF TIMES AT WHICH THE
        DEPENDENT VARIABLE IS KNOWN;

%IF &YDAT3=%STR() %THEN %LET YDAT3=&LIBNAM3..YDATA;

FETCH Y DATA=&YDAT3;;
      OUTPUT Y DATA=YDATTEMP(RENAME=(COL1=Y));
      * Y IS THE VECTOR OF THE DEPENDENT VARIABLE,
        AND IS OF THE SAME DIMENSION AS T;

%IF &INDDAT3=%STR() %THEN %LET INDDAT3=&LIBNAM3..INDDATA;
      FETCH IND DATA=&INDDAT3;;
      OUTPUT IND DATA=INDDTEMP(RENAME=(COL1=IND));
      INDMAX=MAX(IND);
      OUTPUT INDMAX DATA=INDMAX1(RENAME=(COL1=INDMAX));
      * IND IS HERE THE VARIABLE WHICH INDICATES
        TO WHICH 'INDIVIDUAL' THE ORIGINAL
        Y VALUES BELONG;

%IF &PROJIND=%STR() %THEN %DO;
      IND = J(NROW(TPROJ),1,1);
      OUTPUT IND DATA=INDPROJ(RENAME=(COL1=IND));
      %END;
%ELSE
      %DO;
      FETCH IND DATA=&PROJIND;;
      OUTPUT IND DATA=INDPROJ(RENAME=(COL1=IND));
      %END;
      INDPMAX=MAX(IND);
      OUTPUT INDPMAX DATA=INDPMAX1(RENAME=(COL1=INDPMAX));
      * INDPROJ IS A TEMP SAS DATASET CONTAINING VALUES OF A NUMERIC
        INDICATOR VARIABLE, IND, RELATING THE 'INDIVIDUALS' TO
        THEIR PROJECTED Y VALUES;

NOTE      MAXDEGR IS THE MAXIMUM DEGREE OF THE      ;
NOTE      POLYNOMIAL PREVIOUSLY FITTED TO THE DATA;
NOTE      BY -MACRO-INITPOL- ;

      PRINT MAXDEGR;

NOTE      TPROJ IS THE VALUES OF T FOR WHICH A ;
NOTE      PROJECTION IS REQUIRED;
      PRINT TPROJ ;

* XH1 FORMED ITERATIVELY IS THE VECTOR OF MONONOMIALS
  (1, T, T**2 ETC.) FORMED FROM TPROJ.  XH1(J)=TPROJ**(J-1) ;
```

```
      XH1 = J(NROW(TPROJ),1,1) ;  
DO DEGR = 1 TO MAXDEGR BY 1;  
      XH1 = XH1 || TPROJ##DEGR ;  
END;
```

```
PH1 = XH1*CONV;
```

NOTE YPOLPROJ GIVES THE VALUES OF THE FITTED POLYNOMIAL, PREVIOUSLY;
NOTE DETERMINED VIA -MACRO-INITPOL-, AT THE TIMES GIVEN BY THE;
NOTE VECTOR TPROJ. I.E. PURE SIGNAL EXTRACTION ;

```
YPOLPROJ = PH1 * A;
```

```
PRINT YPOLPROJ ;
```

* VALUES OF THE APPROXIMATING POLYNOMIAL IN T ARE NOW CALCULATED
OVER THE RECORDED RANGE OF TTEST ,FOR LATER USE IN PLOTTING
ROUTINES ;

```
TTEST=T//TPROJ;  
D1T= 0:200 ;  
%STR(D1 = D1T%') ;  
D2 = J(201,1,1) ;  
INCR = (MAX(TTEST)-MIN(TTEST))#/200;  
D3 = MIN(TTEST)*D2 + INCR*D1;  
      * D3 CONSISTS OF 201 ELEMENTS RANGING FROM  
      MIN(TTEST) TO MAX(TTEST), WHERE TTEST IS THE  
      VECTOR T//TPROJ, BY EQUAL INCREMENTS OF INCR;
```

```
      XTEST=J(201,1,1);  
DO DEGREE=1 TO MAXDEGR BY 1;  
      XTEST = XTEST || D3##DEGREE;  
END;
```

* EACH COLUMN OF XTEST CONSISTS OF ONE OF THE
MONOMIALS 1,X,X**2....,X**MAXDEGR.
XTEST IS 201*(MAXDEGR+1);

SIGMA IS THE CORRELATION MATRIX (OR COV
MATRIX) FOR ERRORS IN THE LINEAR MODEL
FOR THE KNOWN VALUES, Y ;

FETCH P DATA=&LIBNAM3..PDATA;

* P IS THE MATRIX OF ORTHOGONAL POLYNOMIALS,
BASED ON THE VECTOR T, WITH NROW(Y) ROWS;

FETCH SIGMPROJ DATA=&PROJSIGM;;

* SIGMPROJ IS THE TRANSPOSED MATRIX OF
CORRELATIONS OF ERRORS AT TIMES TPROJ
WITH ERRORS AT TIMES (VEC(T)',TPROJ');

TEMPVAL1=NROW(T)+1;
TEMPVEC1=TEMPVAL1:NCOL(SIGMPROJ);
SIGH1H1 = SIGMPROJ(,TEMPVEC1);
TEMPVEC2=1:NROW(T);
SIGH_H1T = SIGMPROJ(,TEMPVEC2);
%STR(SIGH_H1 = SIGH_H1T%';)

G1H_A = SIGMAINV*SIGH_H1;
%STR(ZH = PH1%' - P%'*G1H_A) ;
%STR(PTSP=P%'*SIGMAINV*P) ;
INVPTSP=INV(PTSP);
NOTE CNPTSP IS THE CONDITION NUMBER OF PTSP WITH;
NOTE RESPECT TO THE INFINITY NORM ;
CNPTSP=MAX(ABS(PTSP)(,+))*MAX(ABS(INVPTSP)(,+));
PRINT CNPTSP;

G1H_B = INVPTSP * ZH;

C1H_A = SIGH1H1-SIGH_H1T*G1H_A;
%STR(C1H_B = G1H_B%'*ZH;)
%STR(R1H_B = G1H_B%'*P%'*SIGMAINV*Y;)

BSTAR = G1H_A + (SIGMAINV * P * G1H_B) ;

%STR(YPROJ = BSTAR%' * Y);

VARYPROJ = C1H_A + C1H_B;

NOTE YPROJ IS THE VECTOR OF PROJECTIONS ;
NOTE CORRESPONDING TO THE VECTOR TPROJ, THE ;
NOTE VECTOR OF TIMES AT WHICH A PROJECTION ;
NOTE INCLUDING ESTIMATED ERROR IS REQUIRED, ;
NOTE WITH THE ASSOCIATED ERROR CORRELATION ;
NOTE MATRIX, SIGMPROJ. SIGMPROJ IS THE ;
NOTE TRANSPOSED MATRIX OF CORRELATIONS OF THE ;
NOTE ERRORS AT TIMES TPROJ WITH THOSE AT ;
NOTE TIMES (VEC(T)',TPROJ') ;

```
PRINT YPROJ ;

      NOTE      THE PREDICTION VARIANCE OF YPROJ,      ;
      NOTE      ASSUMING THE CORRECT POLYNOMIAL MODEL HAS;
      NOTE      BEEN FITTED IS VARYPROJ;
PRINT VARYPROJ;

      NOTE      TPROJ IS THE VALUES OF T FOR WHICH A ;
      NOTE      PROJECTION IS REQUIRED;
PRINT TPROJ ;

      NOTE      SIGMPROJ IS THE TRANSPOSED MATRIX OF    ;
      NOTE      CORRELATIONS OF ERRORS AT TIMES  TPROJ ;
      NOTE      WITH ERRORS AT TIMES  (VEC(T)',TPROJ');
PRINT SIGMPROJ ;

      OUTPUT YPROJ DATA=YPROJ(RENAME=(COL1=YPROJ));

      %END;
```

* THE FOLLOWING DATA STEPS AND PROCEDURES SET UP AND PRODUCE A PLOT
OF THE DATA AND THE APPROXIMATING POLYNOMIAL;

```
DATA PLOTDAT1;
  MERGE YDATTEMP TDATTEMP INDDTEMP;
  KEEP Y T IND;
```

```
%IF &PROJSIGM^=%STR( ) %THEN %DO;
```

```
DATA PINDPROJ;
  SET INDPROJ;
  IF IND=1 THEN IND=11 ;
  IF IND=2 THEN IND=12 ;
  IF IND=3 THEN IND=13 ;
  IF IND=4 THEN IND=14 ;
  IF IND=5 THEN IND=15 ;
  IF IND=6 THEN IND=16 ;
  IF IND=7 THEN IND=17 ;
  IF IND=8 THEN IND=18 ;
  IF IND=9 THEN IND=19 ;
```

```
DATA PLOTDAT2;
  MERGE YPROJ TPROJ PINDPROJ;
  Y=YPROJ;
  KEEP Y T IND ;

      %END;
```



```
DATA PLOTDAT3 ;
MERGE YESTDAT2 D3DATA;
Y=YEST;
IND=.;
KEEP Y T IND;

DATA &LIBNAM3..PROJPLOT;
%IF &PROJSIGM^=%STR() %THEN %DO;
SET PLOTDAT3 PLOTDAT2 PLOTDAT1;
%END;
%ELSE %DO;
SET PLOTDAT3 PLOTDAT1;
%END;
IF IND=. THEN CHARIND='.';
IF IND=1 THEN CHARIND='1';
IF IND=2 THEN CHARIND='2';
IF IND=3 THEN CHARIND='3';
IF IND=4 THEN CHARIND='4';
IF IND=5 THEN CHARIND='5';
IF IND=6 THEN CHARIND='6';
IF IND=7 THEN CHARIND='7';
IF IND=8 THEN CHARIND='8';
IF IND=9 THEN CHARIND='9';
IF IND=11 THEN CHARIND='A';
IF IND=12 THEN CHARIND='B';
IF IND=13 THEN CHARIND='C';
IF IND=14 THEN CHARIND='D';
IF IND=15 THEN CHARIND='E';
IF IND=16 THEN CHARIND='F';
IF IND=17 THEN CHARIND='G';
IF IND=18 THEN CHARIND='H';
IF IND=19 THEN CHARIND='I';
DROP IND;

PROC SORT DATA=&LIBNAM3..PROJPLOT;
BY CHARIND;

%IF &PLOTTP3=GLOT %THEN %DO;

PROC SORT DATA=&LIBNAM3..PROJPLOT;
BY CHARIND;

DATA INDMAX2;
SET INDMAX1;
CALL SYMPUT('MAXIND',INDMAX);
RUN;

DATA INDPMAX2;
SET INDPMAX1;
CALL SYMPUT('MAXPIND',INDPMAX);
```

```

RUN;

GOPTIONS NOTERMAL NOSYMBOL DEVICE=&SYSDEVIC;;
TITLE FITTED POLYNOMIAL PLOTTED AGAINST TIME;
SYMBOL1 C=GREEN V=NONE I=JOIN;
SYMBOL2 C=RED V=DIAMOND I=NONE;
%LET LIM1=%EVAL(&MAXPIND+1);
%IF &MAXPIND>1 %THEN %DO COUNT=3 %TO &LIM1;
    %LET COUNT2=%EVAL(&COUNT-1);
    %IF %EVAL(&MAXIND<10 AND &MAXPIND<10)=1 %THEN
        %STR(SYMBOL&COUNT C=RED V=&COUNT2 I=NONE;);
    %ELSE
        %STR(SYMBOL&COUNT C=RED V=DIAMOND I=NONE;);
%END;
    %LET COUNT=%EVAL(&MAXPIND+2);
    SYMBOL&COUNT C=BLACK V=DIAMOND I=NONE;
%LET LIM2=%EVAL(&MAXPIND+3);
%LET LIM3=%EVAL(&MAXPIND+&MAXIND+1);
%IF &MAXIND>1 %THEN %DO COUNT=&LIM2 %TO &LIM3;
    %LET COUNT2=%EVAL(&COUNT-&MAXPIND-1);
    %IF %EVAL(&MAXIND<10 AND &MAXPIND<10)=1 %THEN
        %STR(SYMBOL&COUNT C=BLACK V=&COUNT2 I=NONE;);
    %ELSE
        %STR(SYMBOL&COUNT C=BLACK V=DIAMOND I=NONE;);
%END;
FOOTNOTE1 IF DATA CONSISTS OF MORE THAN ONE INDIVIDUAL, THE FIRST IS;
FOOTNOTE2 CODED AS A DIAMOND AND THE REMAINDER BY THE CORRESPONDING ;
FOOTNOTE3 NUMBER, IN BLACK. THE PROJECTED DATA IS CODED ;
FOOTNOTE4 SIMILARLY IN RED. THE FITTED POLYNOMIAL IS THE LINE ;
FOOTNOTE5 PLOTTED IN GREEN. ;
PROC GLOT DATA=&LIBNAM3..PROJPLOT;
    PLOT Y*T=CHARIND / OVERLAY NOLEGEND;
    RUN;
                                %END;

%IF &PLOTTP3^=GLOT %THEN %DO;

PROC SORT DATA=&LIBNAM3..PROJPLOT;
    BY DESCENDING CHARIND;

PROC PLOT DATA=&LIBNAM3..PROJPLOT;
    PLOT Y*T=CHARIND / OVERLAY;
    RUN;
                                %END;

%MEND POLPROJ;
```

```
%MACRO POLUPDT(NEWY,NEWY,NEWIND,NEWSIGM,LIBNAM4,UDSTATUS,  
                PLOTTYP4,TDAT4,YDAT4,SIGMDAT4,INDDAT4);;
```

* -MACRO-POLUPDT- UPDATES ESTIMATES OF THE POLYNOMIAL COEFFICIENTS WITH RESPECT TO THE MONOMIALS, GIVEN NEW DATA AND ITS ASSOCIATED CORRELATIONS WITH PREVIOUS DATA. IF &UDSTATUS IS SET TO UPDATE, THE DATA SETS IN THE SAS DATABASE &LIBNAM4 ARE UPDATED TO INCLUDE BOTH THE NEW DATA AND THE CONSEQUENT UPDATED ESTIMATES ;

* THIS MACRO IS DESIGNED TO ALLOW USE EVEN WHERE NO RECURSIVE RESIDUALS HAVE BEEN CALCULATED WHEN INITPOL WAS USED. I.E. INITPOL HAS BEEN INVOKED USING DATA FROM ONLY MAXDEGR+1 DISTINCT TIME POINTS;

```
%IF &GLIBNAME=%STR() %THEN %LET GLIBNAME=WORK;;  
%IF &LIBNAM4=%STR() %THEN %LET LIBNAM4=&GLIBNAME;;
```

* &LIBNAM4 IS BY DEFAULT THE LIBRARY NAME USED WHEN -MACRO-INITPOL- WAS LAST INVOKED, DURING THIS PROGRAM RUN. THAT LIBRARY NAME MAY BE WORK .
IF &LIBNAM4 IS SPECIFIED IT IS ASSUMED -MACRO-INITPOL- HAS ALREADY BEEN INVOKED, POSSIBLY IN A PREVIOUS RUN WITH &LIBNAM4=WORK, AND THE RESULTING SAS DATASETS ARE AVAILABLE TO THIS MACRO. THAT INVOCATION OF -MACRO-INITPOL- SHOULD, OF COURSE, HAVE FITTED THE POLYNOMIAL MODEL OF THE APPROPRIATE MAXIMUM DEGREE;

* &NEWIND IS THE SAS DATASET CONTAINING VALUES OF A NUMERIC INDICATOR VARIABLE RELATING THE 'INDIVIDUALS' TO THE NEW Y VALUES. IT MAY BE LEFT MISSING, IN WHICH CASE ALL Y VALUES ARE ASSUMED TO BE FROM THE SAME INDIVIDUAL DURING PLOTTING PROCEDURES;

* SETTING &UDSTATUS =UPDATE CAUSES AN UPDATE TO THE SAS DATASETS IN LIBRARY &LIBNAM4 , AND IN THE SAS DATASETS CONTAINING T-DATA Y-DATA AND SIGMA-DATA (IRRESPECTIVE OF THE LIBRARY NAME IN WHICH THESE ARE STORED). BY DEFAULT, NO UPDATE IS UNDERTAKEN;

* &PLOTTYP4=GLOT GIVES A SASGRAPH PLOT,
FOR THE PLOT OF FITTED POLYNOMIAL AND THE
UPDATED DATA, Y, AGAINST TIME, T.
THE DEFAULT IS A PROC PLOT TYPE PLOT;

```
%IF &TDAT4=%STR() %THEN %LET TDAT4=&LIBNAM4..TDATA;
```

```
%IF &YDAT4=%STR() %THEN %LET YDAT4=&LIBNAM4..YDATA;  
%IF &SIGMDAT4=%STR() %THEN %LET SIGMDAT4=&LIBNAM4..SIGMDATA;  
%IF &INDDAT4=%STR() %THEN %LET INDDAT4=&LIBNAM4..INDDATA;
```

PROC MATRIX FUZZ;

FETCH T DATA=&TDAT4;;

* T IS THE VECTOR OF TIMES FOR THE OLD DATA
I.E. THE DATA ALREADY IN &LIBNAM4;

FETCH Y DATA=&YDAT4;;

* Y IS THE VECTOR OF THE DEPENDENT VARIABLE
FOR THE OLD DATA;

FETCH SIGMA DATA=&SIGMDAT4;;

* SIGMA IS THE CORRELATON MATRIX (OR COV
MATRIX) FOR ERRORS IN THE LINEAR MODEL
FOR THE OLD DATA;

FETCH SIGMAINV DATA=&LIBNAM4..INVSIGDT;

* SIGMAINV IS INV(SIGMA) , WHERE
SIGMA IS THE CORRELATON MATRIX (OR COV
MATRIX) FOR ERRORS IN THE LINEAR MODEL
FOR THE OLD DATA;

FETCH IND DATA=&INDDAT4;

* IND IS A NUMERIC VARIABLE WHICH INDICATES
TO WHICH 'INDIVIDUAL' THE Y VALUES
BELONG;

FETCH P DATA=&LIBNAM4..PDATA;

* P IS THE MATRIX OF ORTHOGONAL POLYNOMIALS
BASED ON THE VECTOR T, WITH NROW(Y) ROWS;

FETCH X DATA=&LIBNAM4..XDATA;

* X IS THE MATRIX OF THE MONONOMIALS 1, T,
...T**MAXDEGR EVALUATED AT EACH OF THE
COMPONENTS OF THE VECTOR T;

FETCH B DATA=&LIBNAM4..BDATA;

* B IS THE VECTOR OF L-S COEFFICIENTS TO THE
MONONOMIALS FOR THE OLD DATA;

FETCH A DATA=&LIBNAM4..ADATA;

* A IS THE VECTOR OF L-S COEFFICIENTS TO THE
ORTHOGONAL POLYNOMIALS FOR THE OLD DATA;

FETCH CONV DATA=&LIBNAM4..CONVDATA;

* B = CONV*A;

FETCH MAXDEGR DATA=&LIBNAM4..MAXDDATA;

* MAXDEGR IS THE APPROPRIATE MAXIMUM DEGREE
POLYNOMIAL PREVIOUSLY FITTED TO THE OLD DATA;

FETCH TNEW DATA=&NEWT;;

* TNEW IS THE VALUES OF T FOR WHICH AN
UPDATE OF B IS REQUIRED;

FETCH YNEW DATA=&NEWY;;

* YNEW IS THE VALUES OF Y FOR WHICH AN
UPDATE OF B IS REQUIRED;

```

%IF &NEWSIGM^=%STR() %THEN %DO;
  FETCH SIGMNEW DATA=&NEWSIGM;;
  * SIGMNEW IS THE TRANSPOSED MATRIX OF
    CORRELATIONS OF ERRORS AT TIMES TNEW
    WITH ERRORS AT TIMES (VEC(T)',TNEW)';
  %END;

%ELSE %DO;
  TEMP1=J(NROW(TNEW),NROW(T),0);
  TEMP2=I(NROW(TNEW));
  SIGMNEW=TEMP1 || TEMP2;
  %END;

%IF &NEWIND=%STR() %THEN %DO;
  INDNEW = J(NROW(TNEW),1,1);
  OUTPUT INDNEW DATA=INDNEW(RENAME=(COL1=IND));
  %END;
%ELSE %DO;
  FETCH INDNEW DATA=&NEWIND;;
  %END;
  * INDNEW IS A NUMERIC INDICATOR VARIABLE RELATING
    'INDIVIDUALS' TO THEIR NEW Y VALUES;

  TEMPVAL1=NROW(T)+1;
  TEMPVEC1=TEMPVAL1:NCOL(SIGMNEW);
  SIGH1H1 = SIGMNEW(,TEMPVEC1);
  TEMPVEC2=1:NROW(T);
  SIGH_H1T = SIGMNEW(,TEMPVEC2);
  %STR(SIGH_H1 = SIGH_H1T%';)

  * XH1 FORMED ITERATIVELY IS THE VECTOR OF MONONOMIALS
    (1, T, T**2 ETC.) FORMED FROM TNEW. XH1(J)=TNEW**(J-1) ;

  XH1 = J(NROW(TNEW),1,1) ;
  DO DEGR = 1 TO MAXDEGR BY 1;
    XH1 = XH1 || TNEW##DEGR ;
  END;

  PH1 = XH1*CONV;

  G1H_A = SIGMAINV*SIGH_H1;
  %STR(ZH = PH1%' - P%'*G1H_A) ;
  %STR(PTSP=P%'*SIGMAINV*P) ;
  INVPTSP=INV(PTSP);
  NOTE CNPTSP IS THE CONDITION NUMBER OF PTSP WITH;
  NOTE RESPECT TO THE INFINITY NORM ;
  CNPTSP=MAX(ABS(PTSP)(,+) )*MAX(ABS(INVPTSP)(,+) );
  PRINT CNPTSP;

  G1H_B = INVPTSP * ZH;

  C1H_A = SIGH1H1-SIGH_H1T*G1H_A;
  %STR(C1H_B = G1H_B%'*ZH);

```

```
%STR(R1H_A = YNEW - G1H_A% '*Y);
%STR(R1H_B = G1H_B% '*P% '*SIGMAINV*Y);
```

```
M1=C1H_A+C1H_B;
INVM1=INV(M1);
NOTE CNM1 IS THE CONDITION NUMBER OF CNM1 WITH ;
NOTE RESPECT TO THE INFINITY NORM, WHERE ;
NOTE M1=C1H_A+C1H_A*C1H_B*INV(C1H_A);
CNM1=MAX(ABS(M1)(, +))*MAX(ABS(INVM1)(, +));
PRINT CNM1;
```

```
ANEW = A + G1H_B*INVM1*(R1H_A-R1H_B);
```

```
* ANEW IS THE VECTOR OF THE COEFFICIENTS OF THE POLYNOMIALS
ORTHOGONAL WITH RESPECT TO TIME BASED ON THE VECTOR T
ONLY. THE COEFFICIENTS THEMSELVES ARE HOWEVER UPDATED TO
INCLUDE THE NEW INFORMATION CONTAINED IN TNEW, YNEW AND
SIGMANEW. C.F. AUPD ;
```

```
BUPD = CONV * ANEW;
```

```
%STR(VARBUPD = CONV * (INVPTSP-G1H_B*INVM1*G1H_B%') * CONV%');
```

```
NOTE BUPD IS THE UPDATED ESTIMATE OF BETA GIVEN THE ;
NOTE ADDITIONAL DATA. BUPD IS THUS THE UPDATE OF B ;
NOTE WITH ESTIMATE OF VARIANCE VARBUPD;
PRINT BUPD;
PRINT VARBUPD;
```

```
DELTA_B = BUPD - B;
```

```
NOTE DELTA_B IS THE CHANGE IN THE ESTIMATED MONONOMIAL ;
NOTE COEFFICIENTS , AS A CONSEQUENCE OF ADDING THE ;
NOTE ADDITIONAL DATA;
PRINT DELTA_B;
```

```
TUPD = T // TNEW ;
OUTPUT TUPD DATA=TUPDDATA(RENAME=(COL1=T));
```

```
YUPD = Y // YNEW ;
OUTPUT YUPD DATA=YUPDDATA(RENAME=(COL1=Y));
```

```
INDUPD = IND // INDNEW ;
OUTPUT INDUPD DATA=INDUPDDT(RENAME=(COL1=IND));
INDUDMAX=MAX(INDUPD);
OUTPUT INDUDMAX DATA=INDMAX1(RENAME=(COL1=INDMAX));
```

SIGMAUPD = (SIGMA|SIGH_H1) // (SIGH_H1T|SIGH1H1) ;

INVSIGUD = INV(SIGMAUPD);

* X FORMED ITERATIVELY BELOW IS THE MATRIX OF MONOMIALS
(1, T, T**2 ETC.) CORRESPONDING TO P. X(I,J)=T(I)**(J-1) ;

X = J(NROW(T),1,1) ;
DO DEGR = 1 TO MAXDEGR BY 1;
X = X || T**DEGR ;
END;

XUPD = X // XH1;

PUPD = ORPOL(TUPD,MAXDEGR);

%STR(XTUSPU = XUPD%'*INVSIGUD*PUPD) ;
INXTUSPU = INV(XTUSPU) ;

NOTE CNXTUSPU IS THE CONDITION NUMBER OF (XUPDT*INVSIGUD*PUPD);
NOTE WITH RESPECT TO THE INFINITY NORM ;

CNXTUSPU = MAX(ABS(XTUSPU)(,+))*MAX(ABS(INXTUSPU)(,+));
PRINT CNXTUSPU;

NOTE BUPD=CONVUPD*AUPD XUPD*CONVUPD=PUPD ;
NOTE BETA=CONVUPD*ALPHAUPD ;
* ALPHAUPD ARE COEFFICIENTS WITH RESPECT TO PUPD
C.F. ALPHA WHICH ARE COEFFICIENTS WITH RESPECT TO P ;

%STR(INCONUPD = INXTUSPU*(XUPD%'*INVSIGUD*XUPD));
CONVUPD = INV(INCONUPD);
PRINT CONVUPD;

* NOTE: XUPD'*INVSIGUD*XUPD IS ILL-CONDITIONED.
HENCE CONVUPD IS NOT CALCULATED AS
CONVUPD=INV(XUPD'*INVSIGUD*X)*XTUSPU ;

NOTE CNCONUPD IS THE CONDITION NUMBER OF CONVUPD WITH RESPECT;
NOTE TO THE INFINITY NORM ;

CNCONUPD = MAX(ABS(CONVUPD)(,+))*MAX(ABS(INCONUPD)(,+));
PRINT CNCONUPD;

* AUPD IS THE VECTOR OF COEFFICIENTS OF THE POLYNOMIALS ORTHOGONAL
WITH RESPECT TO TIME BASED ON THE VECTOR TUPD ,AND CONTAINED IN
THE MATRIX PUPD ;

AUPD = INCONUPD * BUPD;

* THE NEXT SECTION OF CODE CALCULATES A STATISTIC DISTRIBUTED AS AN CHI-SQUARE STATISTIC UNDER THE NULL HYPOTHESIS OF NO PARAMETER CHANGE FOR BETA. THE ERRORS IN THE LINEAR MODEL ARE ASSUMED TO HAVE A NORMAL DISTRIBUTION FOR THIS TEST;

```
%STR(PTS = P%'*SIGMAINV);
%STR(ESS = Y%' * (SIGMAINV - PTS%'*INVPTSP*PTS) * Y );
      DF_ESS = NROW(Y)-MAXDEGR;
%STR(PTSUPD = PUPD%'*INVSIGUD);
%STR(PTSPUPD = PUPD%'*INVSIGUD*PUPD);
      INPTSPUD = INV(PTSPUPD);
NOTE  CNPTSPUD IS THE CONDITION NUMBER OF PTSPUPD WITH RESPECT;
NOTE  TO THE INFINITY NORM ;
      CNPTSPUD = MAX(ABS(PTSPUPD)(,+))*MAX(ABS(INPTSPUD)(,+));
      PRINT CNPTSPUD;
%STR(ESSUPD = YUPD%' * (INVSIGUD - PTSUPD%'*INPTSPUD*PTSUPD) * YUPD );
      DF_ESSUD = NROW(YUPD)-MAXDEGR;

      C_S = ESSUPD-ESS;
      DFC_S=DF_ESSUD-DF_ESS;
      CHIPROB=PROBCHI(C_S,DFC_S);

NOTE  THE FOLLOWING IS A TEST STATISTIC FOR TESTING THE CONSTANCY ;
NOTE  OF BETA ,GIVEN THE NEW DATA . THE STATISTIC IS DISTRIBUTED AS ;
NOTE  CHI-SQUARE UNDER THE NULL HYPOTHESIS OF NO PARAMETER CHANGE ;
NOTE  FOR BETA. THE ERRORS IN THE LINEAR MODEL ARE ASSUMED TO HAVE A;
NOTE  NORMAL DISTRIBUTION FOR THIS TEST;
      PRINT C_S ;
      PRINT DFC_S ;
      PRINT CHIPROB;
```

/* IF &UDSTATUS=UPDATE THEN THE NEXT SECTION OF CODE UPDATES
THE SAS DATABASE &LIBNAM4 TO INCLUDE THE NEW DATA;

%IF &UDSTATUS=UPDATE %THEN %DO;

```
      OUTPUT BUPD DATA=&LIBNAM4..BDATA(RENAME=(COL1=B));

      OUTPUT VARBUPD DATA=&LIBNAM4..VARBDATA;

      OUTPUT TUPD DATA=&TDAT4(RENAME=(COL1=T));

      OUTPUT YUPD DATA=&YDAT4(RENAME=(COL1=Y));

      OUTPUT INDUPD DATA=&INDDAT4(RENAME=(COL1=IND));
```



```
OUTPUT SIGMAUPD DATA=&SIGMDAT4;

OUTPUT INVSIGUD DATA=&LIBNAM4..INVSIGDT;

OUTPUT AUPD DATA=&LIBNAM4..ADATA(RENAME=(COL1=A));

OUTPUT PUPD DATA=&LIBNAM4..PDATA;

OUTPUT CONVUPD DATA=&LIBNAM4..CONVDATA;

%END;
```

* THE ESTIMATED VALUES FOR UPDATED Y ARE NOW CALCULATED;

```
YEST = PUPD * AUPD ;
RESID= YUPD-YEST ;

OUTPUT YEEST DATA=YEESTDATA(RENAME=(COL1=YEEST));
OUTPUT RESID DATA=RESIDATA(RENAME=(COL1=RESID));
```

* VALUES OF THE APPROXIMATING POLYNOMIAL IN T ARE NOW CALCULATED
OVER THE RECORDED RANGE OF TUPD ,FOR LATER USE IN PLOTTING
ROUTINES ;

```
D1T= 0:200 ;
%STR(D1 = D1T%') ;
D2 = J(201,1,1) ;
INCR = (MAX(TUPD)-MIN(TUPD))/200;
D3 = MIN(TUPD)*D2 + INCR*D1;
* D3 CONSISTS OF 201 ELEMENTS RANGING FROM
MIN(TUPD) TO MAX(TUPD) BY EQUAL INCREMENTS
OF INCR;
```

```
XTEST=J(201,1,1);
DO DEGREE=1 TO MAXDEGR BY 1;
XTEST = XTEST || D3##DEGREE;
END;
```

* EACH COLUMN OF XTEST CONSISTS OF ONE OF THE
MONONOMIALS 1,X,X**2....,X**MAXDEGR.
XTEST IS 201*(MAXDEGR+1);

```
YTEST = XTEST * BUPD;
```

* YTEST ARE THE VALUES OF THE APPROXIMATING
POLYNOMIAL AT EACH ELEMENT OF D3;

```
OUTPUT YTEST DATA=YEESTDAT2(RENAME=(COL1=YEEST));
OUTPUT D3 DATA=D3DATA(RENAME=(COL1=T));
```

```

* THE FOLLOWING SECTION OF CODE DETERMINES A SET OF INDEPENDENT
  RESIDUALS (FOR J=PNO TO NROW(YUPD)) WHICH ARE THE COMPONENTS OF
  THE VECTOR WH;
  * PNO IS THE NUMBER OF OBSERVATIONS IN THE LINEAR MODEL BEFORE
    THE FIRST RECURSIVE RESIDUAL IS CALCULATED;

  SIGH = SIGMA;
  YH   = Y;
  YH1  = YNEW;
  PH   = P;
  INVSIGH = SIGMAINV;

  G1H_A = INVSIGH*SIGH_H1;
  %STR(ZH_P = PH1%' - PH%'*G1H_A) ;
  %STR(PHTSPH=XMULT(PH%','XMULT(INVSIGH,PH)));
  INPHTSPH =SOLVIT(PHTSPH,I(NROW(PHTSPH)));
  NOTE CNPHTSPH IS THE CONDITION NUMBER OF PHTSPH WITH;
  NOTE RESPECT TO THE INFINITY NORM ;
  CNPHTSPH=MAX(ABS(PHTSPH)(,+))*MAX(ABS(INPHTSPH)(,+));
  PRINT CNPHTSPH;

  G1H_BP = INPHTSPH * ZH_P;

  C1H_A = SIGH1H1-SIGH_H1T*G1H_A;
  %STR(C1H_B = G1H_BP%'*ZH_P);
  %STR(R1H_A = YH1 - G1H_A%'*YH);
  %STR(R1H_B = G1H_BP%'*PH%'*INVSIGH*YH);

  COR_RES = R1H_A - R1H_B;
  NOTE COR_RES = R1H_A-R1H_B I.E. THE CORRELATED RECURSIVE RESIDUALS;
  PRINT COR_RES;

  M1 =C1H_A+C1H_B ;

  EIGEN EIGVAM1 EIGVEM1 M1 ;
  NOTE EIGVAM1 ARE THE EIGENVALUES OF M1 WHERE M1=C1H_A+C1H_B ;
  NOTE M1 IS THE COVARIANCE OR CORRELATION MATRIX OF COR_RES;
  PRINT EIGVAM1;
  EIGVAM2 =EIGVAM1 ##0.5;
  EIGVEM2 =EIGVEM1 ;

  * M2 HAS THE PROPERTY THAT M2*M2 =M1 . THIS IS IMPLIES AND IS IMPLIED
    JOINTLY BY EIGVAM2=EIGVAM1##0.5 AND EIGVEM2=EIGVEM1 ;

  %STR(M2 = XMULT(EIGVEM2,XMULT(DIAG(EIGVAM2),EIGVEM2%')));
  INVM2 = SOLVIT(M2,I(NROW(M2)));

  NOTE THE COMPONENTS OF WH ARE MUTUALLY INDEPENDENT N(0,1) RANDOM;
  NOTE VARIABLES PROVIDED THE LINEAR MODEL ERRORS ARE MULTIVARIATE ;
  NOTE NORMAL, NOT NECESSARILY INDEPENDENT. ;

```

```
W_H = INVM2 * (R1H_A-R1H_B);
```

```
PRINT W_H;
```

```
OUTPUT W_H DATA=INDRES2(RENAME=(COL1=W));
```

```
* THE FOLLOWING DATA STEPS AND PROCEDURES SET UP AND PRODUCE PLOTS  
  OF THE DATA, RESIDUALS TO THE MODEL AND THE APPROXIMATING  
  POLYNOMIAL;
```

```
DATA SUMMDATA;
```

```
  MERGE TUPDDATA YUPDDATA INDUPDDT YESTDATA RESIDATA ;  
  DROP ROW;
```

```
PROC PRINT DATA=SUMMDATA;
```

```
PROC PLOT DATA=SUMMDATA;
```

```
  PLOT Y*T=IND  
        YEST*T=IND  
        RESID*T=IND  
        Y*YEST=IND  
        RESID*Y=IND;
```

```
PROC UNIVARIATE DATA=SUMMDATA PLOT FREQ NORMAL;  
  VAR RESID;
```

```
DATA PLOTDAT2;  
  SET SUMMDATA;  
  KEEP Y T IND;
```

```
DATA PLOTDAT3 ;  
  MERGE YESTDAT2 D3DATA;  
  Y=YEST;  
  IND=. ;  
  KEEP Y T IND;
```

```
DATA &LIBNAM4..UPDTPLOT;  
  SET PLOTDAT3 PLOTDAT2;
```

```
%IF &PLOTTP4=GLOT %THEN %DO;
```

```
PROC SORT DATA=&LIBNAM4..UPDTPLOT;  
  BY IND;
```

```
DATA INDMAX2;  
  SET INDMAX1;
```

```
CALL SYMPUT('MAXIND',INDMAX);
RUN;

OPTIONS NOTERMAL NOSYMBOL DEVICE=&SYSDEVIC;;
TITLE FITTED POLYNOMIAL PLOTTED AGAINST TIME;
SYMBOL1 C=GREEN V=NONE I=JOIN;
SYMBOL2 C=BLACK V=DIAMOND I=NONE;
%LET LIM1=%EVAL(&MAXIND+1);
%IF &MAXIND>1 %THEN %DO COUNT=3 %TO &LIM1;
  %LET COUNT2=%EVAL(&COUNT-1);
  %IF &MAXIND<10 %THEN
    %STR(SYMBOL&COUNT C=BLACK V=&COUNT2 I=NONE;);
  %ELSE
    %STR(SYMBOL&COUNT C=BLACK V=DIAMOND I=NONE;);
%END;
FOOTNOTE1 IF DATA CONSISTS OF MORE THAN ONE INDIVIDUAL, THE FIRST IS;
FOOTNOTE2 CODED AS A DIAMOND AND THE REMAINDER BY THE CORRESPONDING ;
FOOTNOTE3 NUMBER, IN BLACK.      THE FITTED POLYNOMIAL IS THE LINE;
FOOTNOTE4 PLOTTED IN GREEN.      ;
FOOTNOTE5;
PROC GLOT DATA=&LIBNAM4..UPDTPLT;
  %IF &MAXIND<10 %THEN
    %STR(PLOT Y*T=IND / OVERLAY ;);
  %ELSE
    %STR(PLOT Y*T=IND / OVERLAY NOLEGEND;);
  RUN;
  %END;
  %ELSE %DO;

PROC SORT DATA=&LIBNAM4..UPDTPLT;
  BY DESCENDING IND;

PROC PLOT DATA=&LIBNAM4..UPDTPLT;
  PLOT Y*T=IND / OVERLAY;
  RUN;
  %END;
```

* CUSUM AND CUSUM OF SQUARES TESTS ARE UNDERTAKEN NEXT;

/* THIS SECTION OF CODE MAKES ALLOWANCE FOR THE POSSIBILITY THAT
NO RECURSIVE RESIDUAL DATA SET EXISTS. I.E.-MACRO-POLUPDT- HAS NOT
BEEN RUN PREVIOUSLY ON THIS TIME SERIES;

```
PROC APPEND BASE=&LIBNAM4..INDRESID DATA=INDRES2;

DATA INDRES2;
```

```
SET &LIBNAM4..INDRESID;

PROC SUMMARY DATA=&LIBNAM4..INDRESID;
  VAR W;
  OUTPUT OUT=NUM1 N=WN1;

DATA NUM1B;
  SET NUM1;
  CALL SYMPUT( 'NUMB1',WN1);

PROC SUMMARY DATA=INDRESID2;
  VAR W;
  OUTPUT OUT=NUM2 N=WN2;

DATA NUM2B;
  SET NUM2;
  CALL SYMPUT( 'NUMB2',WN2);
RUN;

%LET NUMDIF=%EVAL(&NUMB1-&NUMB2);

%IF &NUMDIF=0 %THEN %DO;
  PROC DELETE DATA=&LIBNAM4..INDRESID;
%END;

%ELSE %DO;
  DATA TEMPDT;
    SET &LIBNAM4..INDRESID;
    DIFF=&NUMDIF;
    IF _N_ > DIFF THEN DELETE;
    DROP DIFF;

    DATA &LIBNAM4..INDRESID;
    SET TEMPDT;
%END;

%LET INDRESNO=&NUMB1;

PROC MATRIX;

  FETCH YUPD DATA=YUPDDATA;
  INDRESN=&INDRESNO;
  ODDRESN=ROUND(MOD(INDRESN,2));
  OUTPUT ODDRESN DATA=ODDRESND(RENAME=(COL1=ODDRESN));

  * PNO IS THE NUMBER OF OBSERVATIONS IN THE LINEAR MODEL BEFORE
  THE FIRST RECURSIVE RESIDUAL IS CALCULATED;
  PNO = NROW(YUPD)-INDRESN;
```

```
PNUM=J(INDRESN,1,PNO);
  OUTPUT PNUM DATA=PNUM(RENAME=(COL1=PNUM));
TNUM=J(INDRESN,1,NROW(YUPD));
  OUTPUT TNUM DATA=TNUM(RENAME=(COL1=TNUM));

RUN;

DATA ODDRESN2;
  SET ODDRESND;
  CALL SYMPUT('ODDRESNO',ODDRESN);
RUN;

* THE CUSUM AND CUSUM OF SQUARES TESTS FOLLOW;

DATA CUSUMDT1;
  MERGE INDRES2
    PNUM
    TNUM;
  WTOT + W;
  R = _N_ +PNUM;
  * PNUM IS THE NUMBER OF OBSERVATIONS IN THE LINEAR MODEL BEFORE
    THE FIRST RECURSIVE RESIDUAL IS CALCULATED;
  CUSUM95U = (0.948*(2*R + TNUM - 3*PNUM)) / SQRT(TNUM-PNUM);
  CUSUM95L = -CUSUM95U;

PROC PRINT DATA=CUSUMDT1;
RUN;

TITLE3 CUSUM PLOT OF WTOT WITH 95 PERCENT CONFIDENCE INTERVALS;
TITLE4    VERSUS OBSERVATION NUMBER, R ;

PROC PLOT DATA=CUSUMDT1;
  PLOT WTOT*R='*'
    CUSUM95U*R='.'
    CUSUM95L*R='.' /HZERO OVERLAY;

RUN;

TITLE3; TITLE4;

%IF &INDRESNO>=2 %THEN %DO;

DATA CUSUMDT2;
  MERGE INDRES2
    PNUM
    TNUM;
  R = _N_ +PNUM;
  WSQTOT + W**2;
  M = (TNUM-PNUM)/2;

PROC SORT DATA=CUSUMDT2;
```

```
BY DESCENDING R;

DATA CUSUMDT3;
  SET CUSUMDT2;
  IF _N_=1 THEN WSQTMAX=WSQTOT;
  RETAIN WSQTMAX;
  SR = WSQTOT/WSQTMAX;
  IF _N_=1 THEN DELETE;

PROC SORT DATA=CUSUMDT3;
  BY R;

%IF &ODDRESNO=0 %THEN %DO;

  DATA CUSUMDT4;
    SET CUSUMDT3;
    SRADJ = SR - (_N_/(TNUM-PNUM));
    N = M-1;
    KEEP SRADJ SR R N;

  PROC MATRIX;
    FETCH RAW DATA=CUSUMDT4;
    SRADJ = RAW(,3);
    SR = RAW(,2);
    R = RAW(,1);
    N = RAW(1,4);
    CPLUS=MAX(SRADJ);
    CMINUS=MAX(-SRADJ);

    NOTE CPLUS AND CMINUS ARE THE STATISTICS FOR THE CUSUM OF ;
    NOTE SQUARES TEST GIVEN IN BROWN ET AL, JRSS(B) 37 NO2 1975 ;
    NOTE P155. FOR THE SIGNIFICANCE LEVEL ALPHA, TABULATED VALUES;
    NOTE IN DURBIN, BIOMETRIKA 56 1969 P4 TABLE 1, GIVEN C0, THE ;
    NOTE VALUE IN THE TABLE AT N AND ALPHA/2. ;
    NOTE IF EITHER CPLUS OR CMINUS EXCEEDS C0 THEN THE HYPOTHESIS;
    NOTE OF CONSTANT REGRESSION COEFFICIENTS IS REJECTED AT THE ;
    NOTE SIGNIFICANCE LEVEL ALPHA. AN EYEBALL TEST IS ALSO ;
    NOTE RECOMMENDED BY BROWN ET AL 1975 TO GIVE AN INDICATION ;
    NOTE OF HOW CLOSE EACH OF THE COMPONENTS OF SRADJ IS TO C0 ;

    PRINT CPLUS CMINUS N SRADJ SR R;

  %END;

%IF &ODDRESNO=1 %THEN %DO;

  DATA CUSUMDT4;
    SET CUSUMDT3;
    SRADJ = SR - (_N_/(TNUM-PNUM));
    N_U = M-0.5;
    N_L = M-1.5;
```

KEEP SRADJ SR R N_L N_U;

```
PROC MATRIX;
  FETCH RAW DATA=CUSUMDT4;
  SRADJ = RAW(,3);
  SR = RAW(,2);
  R = RAW(,1);
  N_L = RAW(1,5);
  N_U = RAW(1,4);
  CPLUS=MAX(SRADJ);
  CMINUS=MAX(-SRADJ);
```

```
NOTE  CPLUS AND CMINUS ARE THE STATISTICS FOR THE CUSUM OF ;
NOTE  SQUARES TEST GIVEN IN  BROWN ET AL, JRSS(B) 37 NO2 1975  ;
NOTE  P155.  FOR THE SIGNIFICANCE LEVEL ALPHA, TABULATED VALUES;
NOTE  IN  DURBIN, BIOMETRIKA 56 1969 P4 TABLE 1, GIVEN C0, THE  ;
NOTE  VALUE LINEARLY INTERPOLATED FROM THE TABLE FOR ALPHA/2,  ;
NOTE  BETWEEN N_L AND N_U.                                     ;
NOTE  IF EITHER CPLUS OR CMINUS EXCEEDS C0 THEN THE HYPOTHESIS;
NOTE  OF CONSTANT REGRESSION COEFFICIENTS IS REJECTED AT THE  ;
NOTE  SIGNIFICANCE LEVEL ALPHA.  AN EYEBALL TEST IS ALSO      ;
NOTE  RECOMMENDED BY  BROWN ET AL 1975  TO GIVE AN INDICATION ;
NOTE  OF HOW CLOSE EACH OF THE COMPONENTS OF SRADJ IS TO C0  ;
```

```
PRINT  CPLUS CMINUS N_L N_U SRADJ SR R;
```

```
%END;
```

```
%END;
```

```
%IF &UDSTATUS=UPDATE %THEN %DO;
```

```
DATA &LIBNAM4..SUMMDATA;
  SET SUMMDATA;
```

```
DATA &LIBNAM4..INITPLOT;
  SET &LIBNAM4..UPDTPLOT;
```

```
DATA &LIBNAM4..TNUM;
  SET TNUM;
```

```
PROC APPEND BASE=&LIBNAM4..INDRESID DATA=INDRESID2;
```

```
RUN;
```

```
%END;
```

```
%MEND POLUPDT;
```


%MACRO ESTRCOV(MAXDEG, TDTA, YDTA, INDDTA, OUTCOV, RESLIB, SAMEX, REORD);;

* -MACRO-ESTRCOV- ESTIMATES THE VARIANCE-COVARIANCE MATRIX OF THE ERRORS IN A RANDOM COEFFICIENT MODEL. THE 'DESIGN' MATRIX IN THIS CASE IS FORMED FROM THE TIMES IN &TDTA (I.E. THE VECTOR OF TIMES, T, TO WHICH THE DATA, Y, CORRESPONDS) BY VERTICAL CONCATENATION OF THE VALUES 1, T(I), T(I)**2, ..., T(I)**&MAXDEG, FOR EACH TIME T(I) IN &TDTA. THE DATA Y IS CONTAINED IN &YDTA. THE VAR-COV MATRIX SO FORMED CAN THEN BE USED AS THE MATRIX OF ERROR COVARIANCES TO BE USED IN -MACRO-INITPOL- ;

* RANDOM COEFFICIENT MODELS ARE DISCUSSED IN :

'THE THEORY AND PRACTICE OF ECONOMETRICS' JUDGE, G.G, GRIFFITHS, W.E, HILL, R.C, AND LEE, T-S. JOHN WILEY AND SONS, NEW YORK (1980) P347-351

'STATISTICAL INFERENCE IN RANDOM COEFFICIENT REGRESSION MODELS' SWAMY, P.A.V.B. SPRINGER-VERLAG, BERLIN, NEW YORK (1971) ;

* &INDDTA IS THE DATA SET CONTAINING DESIGNATIONS OF INDIVIDUALS, NUMBERED 1,2,3,...INDMAX, CORRESPONDING TO THE DATA IN BOTH &TDTA AND IN &YDTA. &INDDTA MUST BE SPECIFIED AND INDMAX MUST BE GREATER THAN OR EQUAL TO TWO;

* &OUTCOV IS THE SAS DATASET WHICH ON COMPLETION OF EXECUTION OF -MACRO-ESTRCOV- CONTAINS THE ERROR VAR-COV MATRIX;

* &RESLIB IS THE SAS LIBRARY NAME OF THE DATA SETS CONTAINING RESIDUALS TO THE INDIVIDUAL MODELS, ASSUMING ZERO CORRELATION BETWEEN OBSERVATIONS ON INDIVIDUALS AT DIFFERENT TIMES. THESE DATA SETS CAN BE USED TO CHECK THIS ASSUMPTION, FOR OBSERVATIONS EQUALLY SPACED IN TIME BY INDIVIDUAL. THE DEFAULT VALUE IS WORK;

* IF THE DATA CONSISTS OF N INDIVIDUALS EACH RECORDED AT THE SAME DISTINCT TIMES T AND EACH INDIVIDUAL HAS THE SAME X MATRIX IN THE MODEL, THEN SPECIFYING &SAMEX=SAMEXMAT WILL GREATLY REDUCE EXECUTION TIME FOR THIS MACRO;

* IF &REORD=NOREORD THEN DATA, Y, AND THE CORRESPONDING T AND IND, ARE ASSUMED TO BE ORDERED FIRSTLY BY INDIVIDUAL AND THEN BY TIME. THIS OPTION WILL REDUCE EXECUTION TIME FOR THIS MACRO;

```
%IF &RESLIB=%STR() %THEN %LET RESLIB=WORK;

PROC MATRIX FUZZ;
  FETCH IND DATA=&INDDTA;
  OUTPUT IND DATA=INDDATA(RENAME=(COL1=IND));
  INDMAX=MAX(IND);
  OUTPUT INDMAX DATA=INDMAXD(RENAME=(COL1=INDMAX));

  DATA INDTEMP;
  SET INDMAXD;
  CALL SYMPUT('MAXIND', INDMAX);
  RUN;

%DO I=1 %TO &MAXIND;

  DATA YDAT&I;
  MERGE &YDTA INDDATA;
  IF IND=&I THEN OUTPUT;
  ELSE RETURN;
  DROP IND ROW;

  DATA TDAT&I;
  MERGE &TDTA INDDATA;
  IF IND=&I THEN OUTPUT;
  ELSE RETURN;
  DROP IND ROW;

  DATA ORDDAT&I;
  SET INDDATA;
  ORD= N ;
  IF IND=&I THEN OUTPUT;
  ELSE RETURN;
  DROP IND ROW;

%END;

%IF &SAMEX^=SAMEXMAT %THEN %DO I=1 %TO &MAXIND;

  PROC MATRIX FUZZ;
  * T&I IS THE VECTOR OF TIMES CORRESPONDING TO THE DATA, Y&I ;
  FETCH Y&I DATA=YDAT&I;
  FETCH T&I DATA=TDAT&I;;

  NOTE MAXDG IS THE ORDER OF THE MAXIMUM DEGREE POLYNOMIAL;
  NOTE BEING FITTED;
  MAXDG = &MAXDEG;;
  PRINT MAXDG;

  * TNUM&I IS THE NUMBER OF TIMES (NOT NECESSARILY DISTINCT) AT WHICH
  Y&I (I.E. Y FOR INDIVIDUAL &I) HAS BEEN RECORDED;
  TNUM&I=NROW(T&I);
```

OUTPUT TNUM&I DATA=TNUMD&I;;

NOTE X&I FORMED ITERATIVELY BELOW IS THE MATRIX OF MONONOMIALS, ;
NOTE 1, T, T**2 ETC., CORRESPONDING TO P&I. X&I(I,J)=T&I(I)**(J-1);

```

      X&I = J(NROW(T&I),1,1) ;
DO   DEGR = 1 TO MAXDG BY 1;
      X&I = X&I || T&I##DEGR ;
END;
```

OUTPUT X&I DATA=XDATA&I;;
PRINT X&I;

* P&I CONSISTS OF THE POLYNOMIALS ORTHOGONAL
WITH RESPECT TO T&I, EVALUATED AT EACH OF
THE COMPONENTS OF T&I. P&I HAS MAXDG+1 COLS.
P&I(I,J)=VALUE OF (J-1)TH ORTHOG POLY AT T&I(I) ;

P&I = ORPOL(T&I,MAXDG);

* A&I IS THE LEAST SQUARES SOLUTION WITH RESPECT TO THE ORTHOGONAL
POLYNOMIALS. IT IS A VECTOR OF THE ESTIMATED COEFFICIENTS OF
THE ORTHOGONAL POLYNOMIALS;

```

      %STR(INVPTP&I = INV(P&I%'*P&I)) ;
      %STR(A&I = INVPTP&I * P&I%'* Y&I) ;
```

* THESE COEFFICIENTS, A&I, ARE NOW CONVERTED TO THE MONONOMIAL
COEFFICIENTS , DENOTED BY B&I ;

```

      %STR(XTP&I = X&I%'*P&I) ;
      INVXTP&I = INV(XTP&I) ;
```

NOTE CNXTP&I IS THE CONDITION NUMBER OF (XT&I*P&I) WITH RESPECT;
NOTE TO THE INFINITY NORM ;

```

      CNXTP&I = MAX(ABS(XTP&I)(,+))*MAX(ABS(INVXTP&I)(,+));
      PRINT CNXTP&I;
```

NOTE B&I=CONV&I*A&I X&I*CONV&I=P&I BETA&I=CONV&I*ALPHA&I;

```

      %STR(INCONV&I = INVXTP&I*(X&I%'*X&I));
      CONV&I = INV(INCONV&I);
```

* NOTE: X&I'*X&I IS ILL-CONDITIONED.
HENCE CONV&I IS NOT CALCULATED AS
CONV&I=INV(X&I'*X&I)*XTP&I ;

NOTE CNCONV&I IS THE CONDITION NUMBER OF CONV&I WITH RESPECT;
NOTE TO THE INFINITY NORM ;

```

      CNCONV&I = MAX(ABS(CONV&I)(,+))*MAX(ABS(INCONV&I)(,+));
      PRINT CNCONV&I;
```

NOTE B&I IS THE VECTOR OF ESTIMATED COEFFICIENTS OF THE MONONOMIALS;
 NOTE 1, T, T**2,...,T**MAXDG BASED ON THE DATA FOR ;
 NOTE INDIVIDUAL, &I, ASSUMING COVARIANCE BETWEEN THESE INDIVIDUALS;
 NOTE TO BE THE IDENTITY MATRIX;

B&I = CONV&I * A&I ;

PRINT B&I;
 OUTPUT B&I DATA=BDATA&I(RENAME=(COL1=B&I));

* THE ESTIMATED VALUES FOR Y ARE NOW CALCULATED;

YEST&I = P&I * A&I ;
 RESID&I= Y&I-YEST&I ;

OUTPUT RESID&I DATA=&RESLIB..RESID&I(RENAME=(COL1=RESID&I));

PROC ARIMA DATA=&RESLIB..RESID&I;
 IDENTIFY VAR=RESID&I NLAG=6;

%END;

%IF &SAMEX=SAMEXMAT %THEN %DO;

PROC MATRIX FUZZ;

* T IS THE VECTOR OF TIMES CORRESPONDING TO THE DATA, Y,
 FOR EACH INDIVIDUAL;
 FETCH T DATA=TDAT1;;

NOTE MAXDG IS THE ORDER OF THE MAXIMUM DEGREE POLYNOMIAL;
 NOTE BEING FITTED;
 MAXDG = &MAXDEG;;
 PRINT MAXDG;

* TNUM IS THE NUMBER OF TIMES (NOT NECESSARILY DISTINCT) AT WHICH
 Y FOR EACH INDIVIDUAL HAS BEEN RECORDED;
 TNUM=NROW(T);
 OUTPUT TNUM DATA=TNUMD;;

NOTE X FORMED ITERATIVELY BELOW IS THE MATRIX OF MONONOMIALS, ;
 NOTE 1, T, T**2 ETC., CORRESPONDING TO P. X(I,J)=T(I)**(J-1);

X = J(NROW(T),1,1) ;
 DO DEGR = 1 TO MAXDG BY 1;
 X = X || T##DEGR ;
 END;

OUTPUT X DATA=XDATA;;

PRINT X;

* P CONSISTS OF THE POLYNOMIALS ORTHOGONAL
WITH RESPECT TO T, EVALUATED AT EACH OF
THE COMPONENTS OF T. P HAS MAXDEGR+1 COLS.
P(I,J)=VALUE OF (J-1)TH ORTHOG POLY AT T(I) ;

P = ORPOL(T,MAXDG);
%STR(INVPTP = INV(P%'*P)) ;

%STR(XTP = X%'*P) ;
INVXTP = INV(XTP) ;

NOTE CNXTP IS THE CONDITION NUMBER OF (XT*P) WITH RESPECT;
NOTE TO THE INFINITY NORM ;
CNXTP = MAX(ABS(XTP)(,+))*MAX(ABS(INVXTP)(,+));
PRINT CNXTP;

%STR(INCONV = INVXTP*(X%'*X));
CONV = INV(INCONV);

* NOTE: X'*X IS ILL-CONDITIONED.
HENCE CONV IS NOT CALCULATED AS
CONV=INV(X'*X)*XTP ;

NOTE CNCONV IS THE CONDITION NUMBER OF CONV WITH RESPECT;
NOTE TO THE INFINITY NORM ;
CNCONV = MAX(ABS(CONV)(,+))*MAX(ABS(INCONV)(,+));
PRINT CNCONV;

%DO I=1 %TO &MAXIND;

FETCH Y&I DATA=YDAT&I;

* A&I IS THE LEAST SQUARES SOLUTION WITH RESPECT TO THE ORTHOGONAL
POLYNOMIALS. IT IS A VECTOR OF THE ESTIMATED COEFFICIENTS OF
THE ORTHOGONAL POLYNOMIALS;

%STR(A&I = INVPTP * P%'* Y&I) ;

* THESE COEFFICIENTS, A&I, ARE NOW CONVERTED TO THE MONONOMIAL
COEFFICIENTS , DENOTED BY B&I ;

NOTE B&I IS THE VECTOR OF ESTIMATED COEFFICIENTS OF THE MONONOMIALS;
NOTE 1, T, T**2,...,T**MAXDG BASED ON THE DATA FOR ;
NOTE INDIVIDUAL, &I, ASSUMING COVARIANCE BETWEEN THESE INDIVIDUALS;
NOTE TO BE THE IDENTITY MATRIX;

B&I = CONV * A&I ;

```
PRINT B&I;
OUTPUT B&I DATA=BDATA&I(RENAME=(COL1=B&I));

* NOTE  B&I=CONV*A&I    X*CONV=P    BETA&I=CONV*ALPHA&I;

* THE ESTIMATED VALUES FOR Y ARE NOW CALCULATED;

YEST&I = P * A&I ;
RESID&I= Y&I-YEST&I ;

OUTPUT RESID&I DATA=&RESLIB..RESID&I(RENAME=(COL1=RESID&I));

%END;

%DO I=1 %TO &MAXIND;

PROC ARIMA DATA=&RESLIB..RESID&I;;
IDENTIFY VAR=RESID&I NLAG=6;

%END;

%END;

%IF &REORD^=NOREORD %THEN %DO;

* ORDDATA CONTAINS INFORMATION WHICH CAN LATER BE USED TO REORDER
  THE VAR-COV MATRIX, PHI, TO FORM THE VAR-COV MATRIX, SIGMA. THE
  MATRIX SIGMA WILL CONTAIN THE COVARIANCES BETWEEN OBSERVATIONS
  IN THE SAME ORDER AS THE DATA IN &YDATA ;

DATA ORDDATA;
SET ORDDAT1;

%DO I=2 %TO &MAXIND;
  %* THE DATA STEP AND PROC DATASETS BELOW CAN BE REPLACED BY
    PROC APPEND BASE=ORDDATA DATA=ORDDAT&I SEMICOLON
    BUT ONLY IF THE BUG IN SAS82 TEST RELEASE PROC APPEND HAS
    BEEN FIXED. THIS BUG CAUSES EXECUTION TIMES OF 60SECS CPU
    FOR APPENDING 15 OBS WITH ONE VARIABLE;
  DATA ORDDAT;
    SET ORDDATA ORDDAT&I;

  PROC DATASETS DDNAME=WORK NOLIST;
    EXCHANGE ORDDAT=ORDDATA;
%END;

%END;
```

```
%IF &SAMEX^=SAMEXMAT %THEN %DO;
```

```
PROC MATRIX FUZZ;
```

```
%DO I=1 %TO &MAXIND;
```

```
    FETCH E&I DATA=&RESLIB..RESID&I;
```

```
    FETCH TNUM&I DATA=TNUMD&I;
```

```
    FETCH X&I DATA=XDATA&I;
```

```
    FETCH B&I DATA=BDATA&I;
```

```
    MAXDG = &MAXDEG;;
```

```
NOTE SIG&I&I BELOW IS EQUATION 8.5.13 OF JUDGE ET AL, 1980 ;
```

```
NOTE AND MEASURES THE VARIANCE ABOUT THE POLYNOMIAL FITTED ;
```

```
NOTE TO THE DATA FOR INDIVIDUAL &I, GIVEN ZERO CORRELATION ;
```

```
NOTE BETWEEN ERRORS FOR THAT INDIVIDUAL AT DIFFERENT TIMES. ;
```

```
%STR(SIG&I&I = (E&I%'*E&I)#/(TNUM&I-MAXDG));
```

```
PRINT SIG&I&I;
```

```
%STR(BBT&I = B&I*B&I%');
```

```
%END;
```

```
    DIMEN=MAXDG+1;
```

```
    SB=J(DIMEN,DIMEN,0);
```

```
    SB1=J(DIMEN,DIMEN,0);
```

```
    SB2=J(DIMEN,1,0);
```

```
%DO I=1 %TO &MAXIND;
```

```
    SB1=SB1+BBT&I;
```

```
%END;
```

```
%DO I=1 %TO &MAXIND;
```

```
    SB2=SB2+B&I;
```

```
%END;
```

```
* SB BELOW IS EQUATION(8.5.15) OF JUDGE ET AL (1980);
```

```
    INDMAX=&MAXIND;
```

```
%STR(SB = SB1 - ((SB2*SB2%')#/INDMAX));
```

```
NOTE COV_B BELOW IS DELTA OF EQUATION 8.5.14 OF JUDGE ET AL, 1980 ;
```

```
NOTE AND MEASURES THE ESTIMATED COVARIANCE BETWEEN THE COMPONENTS ;
```

```
NOTE OF B I.E. THE POLYNOMIAL COEFFICIENTS ESTIMATED OVER ALL ;
```

```
NOTE INDIVIDUALS, WHERE THE POLYNOMIAL HAS MAXIMUM DEGREE, &MAXDEG;;
```

```
COV_B = SB #/ (INDMAX-1);
```

```
PRINT COV_B;
```

```
NOTE PHI&I&I BELOW IS EQUATION 8.5.7 OF JUDGE ET AL, 1980 ;
```

```
%DO I=1 %TO &MAXIND;
    %STR(PHI&I&I = (X&I*COV_B*X&I%' ) + (I(NROW(X&I))#SIG&I&I));

                                PRINT  PHI&I&I;

%END;

* PHI BELOW IS THE BLOCK DIAGONAL MATRIX FORMED FROM PHI11 TO
  PHIMM WHERE M IS INDMAX, AND PHIJJ FOR J=1,2,...,M IS
  EQUATION(8.5.7) OF JUDGE ET AL (1980);

  PHI=BLOCK(PHI11,PHI22);
  %IF &MAXIND>2 %THEN %DO I=3 %TO &MAXIND;
    PHI=BLOCK(PHI,PHI&I&I);
  %END;

* PHI MAY NOW BE REORDERED TO FORM SIGMA AS EXPLAINED ABOVE;

%IF &REORD^=NOREORD %THEN %DO;

  FETCH ORDER DATA=ORDDATA;
    DIMEN2=NROW(PHI);
    TEMP=J(DIMEN2,DIMEN2,0);
    DO K=1 TO DIMEN2;
      ORD=ORDER(K,1);
      TEMP(ORD,)=PHI(K,);
    END;
    SIGMA=J(DIMEN2,DIMEN2,0);
    DO K=1 TO DIMEN2;
      ORD=ORDER(K,1);
      SIGMA(,ORD)=TEMP(,K);
    END;
  %END;

  %ELSE %STR(SIGMA=PHI,);

%END;

%IF &SAMEX=SAMEXMAT %THEN %DO;

  PROC MATRIX FUZZ;

    FETCH TNUM DATA=TNUMD;
    FETCH X DATA=XDATA;

    %DO I=1 %TO &MAXIND;

      FETCH E&I DATA=&RESLIB..RESID&I;
      FETCH B&I DATA=BDATA&I;
      MAXDG = &MAXDEG;;
```


NOTE SIG&I&I BELOW IS EQUATION 8.5.13 OF JUDGE ET AL, 1980 ;
 NOTE AND MEASURES THE VARIANCE ABOUT THE POLYNOMIAL FITTED ;
 NOTE TO THE DATA FOR INDIVIDUAL &I, GIVEN ZERO CORRELATION ;
 NOTE BETWEEN ERRORS FOR THAT INDIVIDUAL AT DIFFERENT TIMES. ;

```
%STR(SIG&I&I = (E&I% '*E&I) # / (TNUM-MAXDG));
                                PRINT SIG&I&I;
```

```
%STR(BBT&I = B&I*B&I% ');
```

```
%END;
```

```
DIMEN=MAXDG+1;
SB=J(DIMEN,DIMEN,0);
SB1=J(DIMEN,DIMEN,0);
SB2=J(DIMEN,1,0);
```

```
%DO I=1 %TO &MAXIND;
    SB1=SB1+BBT&I;
```

```
%END;
```

```
%DO I=1 %TO &MAXIND;
    SB2=SB2+B&I;
```

```
%END;
```

* SB BELOW IS EQUATION(8.5.15) OF JUDGE ET AL (1980);

```
INDMAX=&MAXIND;
%STR(SB = SB1 - ((SB2*SB2%') # / INDMAX));
```

NOTE COV_B BELOW IS DELTA OF EQUATION 8.5.14 OF JUDGE ET AL, 1980 ;
 NOTE AND MEASURES THE ESTIMATED COVARIANCE BETWEEN THE COMPONENTS ;
 NOTE OF B I.E. THE POLYNOMIAL COEFFICIENTS ESTIMATED OVER ALL ;
 NOTE INDIVIDUALS, WHERE THE POLYNOMIAL HAS MAXIMUM DEGREE, &MAXDEG;;

```
COV_B = SB # / (INDMAX-1);
```

```
PRINT COV_B;
```

NOTE PHI11, PHI22,.....ETC. BELOW IS EQUATION 8.5.7 ;
 NOTE OF JUDGE ET AL, 1980 ;

```
%STR(GAMMA = X*COV_B*X% ');
%DO I=1 %TO &MAXIND;
    PHI&I&I = GAMMA + (I(NROW(X)) # SIG&I&I);
```

```
PRINT PHI&I&I;
```

```
%END;
```

* PHI BELOW IS THE BLOCK DIAGONAL MATRIX FORMED FROM PHI11 TO

PHIMM WHERE M IS INDMAX, AND PHIJJ FOR J=1,2,...,M IS
EQUATION(8.5.7) OF JUDGE ET AL (1980);

```
PHI=BLOCK(PHI11,PHI22);
%IF &MAXIND>2 %THEN %DO I=3 %TO &MAXIND;
  PHI=BLOCK(PHI,PHI&I&I);
%END;
```

* PHI MAY NOW BE REORDERED TO FORM SIGMA AS EXPLAINED ABOVE;

```
%IF &REORD^=NOREORD %THEN %DO;
```

```
  FETCH ORDER DATA=ORDDATA;
  DIMEN2=NROW(PHI);
  TEMP=J(DIMEN2,DIMEN2,0);
  DO K=1 TO DIMEN2;
    ORD=ORDER(K,1);
    TEMP(ORD,)=PHI(K,);
  END;
  SIGMA=J(DIMEN2,DIMEN2,0);
  DO K=1 TO DIMEN2;
    ORD=ORDER(K,1);
    SIGMA(,ORD)=TEMP(,K);
  END;
```

```
%END;
```

```
%ELSE %STR(SIGMA=PHI;);
```

```
%END;
```

NOTE SIGMA IS THE VARIANCE COVARIANCE MATRIX OF THE ERRORS TO ;
NOTE THE LINEAR MODEL OF POLYNOMIALS, FORMULATED AS A SWAMY ;
NOTE RANDOM COEFFICIENT MODEL. IT IS ORDERED IN THE SAME WAY AS ;
NOTE THE GIVEN VECTOR OF OBSERVATIONS, TIMES, AND INDIVIDUALS. ;

PRINT SIGMA;

OUTPUT SIGMA DATA=&OUTCOV;

RUN;

%MEND ESTRCOV;

REFERENCES

1. Anderson, B.D.O. and Moore, J.B. (1979) *Optimal filtering*. Prentice-Hall, Englewood Cliffs, New Jersey.
2. Arnab, R. (1979) *On strategies of sampling finite populations on successive occasions with varying probabilities*. Sankhya C 41 (2) p141-155.
3. Bailar, B.A. (1975) *The effect of rotation group bias on estimates from panel surveys*. J. Amer. Statis. Assoc. 70 p23-30.
4. Barnett, V. (1974) *Elements of sampling theory*. English University Press, London.
5. Basu, D. (1971) *An essay on the logical foundations of survey sampling, part one*. in Foundations of Statistical Inference, editors V.P. Godambe and D.A. Sprott. Holt, Rinehart and Winston, Toronto p203-242.
6. Bibby, J. and Toutenburg, H. (1977) *Prediction and improved estimation in linear models*. Wiley, New York.
7. Biorn, E. (1981) *Estimating economic relations from incomplete cross section/ time series data*. J. of Econometrics 16 p221-236.

8. Blight, B.J.N. and Scott, A.J. (1973) *A statistical model for repeated surveys*. J. Roy. Statist. Soc. B 35 p61-66.
9. Box, G.E.P. and Cox, D.R. (1964) *An analysis of transformations*. J. Roy. Statist. Soc. B 26 p211-243.
10. Box, G.E.P. and Jenkins, G.M. (1970) *Time series analysis: forecasting and control*. Holden-Day, San Francisco.
11. Brown, R.L., Durbin, J. and Evans, J.M. (1975) *Techniques for testing the constancy of regression relationships over time*. J. Roy. Statist. Soc. B 37 p149-163.
12. Cassell, C-M, Sarndal, C-E and Wretman, J.H. (1977) *Foundations of inference in survey sampling*. Wiley, New York.
13. Cochran, W.G. (1938) *The omission or addition of an independent variate in multiple linear regression*. J. Roy. Stat. Soc. Suppl. 5 p171-176.
14. Cochran, W.G. (1963) *Sampling techniques*. Wiley, New York. 2nd edition.
15. Cochran, W.G. (1977) *Sampling techniques*. Wiley, New York. 3rd edition.
16. Dagsvik, J. (1978) *Time series estimators for repeated sample surveys;*
Part 1: Empirical results from the Norwegian labour force surveys,
Part 2: Calculation of prediction formulae using a matrix function approach.
Central Bureau of Statistics, Oslo.

17. DeMets, D. and Halperin, M. (1977) *Estimation of a simple regression coefficient in samples arising from a sub-sampling procedure.* Biometrics 33 p47-56.
18. DuMouchal, W.H. and Duncan, G.J. (1983) *Using sample survey weights in multiple regression analyses of stratified samples.* J. Amer. Statist. Assoc. 78 p535-543.
19. Duncan, D.B. and Horn, S.D. (1972) *Linear dynamic recursive estimation from the viewpoint on regression analysis.* J. Amer. Statist. Assoc. 67 p815-821.
20. Durbin, J. (1969) *Tests for serial correlation in regression analysis based on the periodogram of least squares residuals.* Biometrika 56 p1-15.
21. Eckler, A.R. (1955) *Rotation sampling.* Ann. Math. Statist. 26 p664-685.
22. Fisher, R.A. (1925) *Theory of statistical estimation.* Proc. Camb. Phil. Soc. 22 p700-725.
23. Froberg, C-E. (1969) *Introduction to numerical analysis.* Addison-Wesley, Massachusetts. 2nd edition.
24. Fuller, W. (1975) *Regression analysis for sample survey.* Sankhya C 87 (3) p117-132.
25. Fuller, W.A. (1976) *Introduction to statistical time series.* Wiley, New York.

26. Godolphin, E.J. and Stone, J.M. (1980) *On the structural representation for polynomial-projecting predictor models based on the Kalman filter.* J. Roy. Statist. Soc. B 42 (1) p35-45.
27. Graham, J.E. (1973) *Composite estimation in two cycle rotation sample designs.* Commun. in Statist. 1 (5) p419-431.
28. Granger, C.W.J. and Morris, M.J. (1976) *Time series modelling and interpretation.* J. Roy. Statist. Soc. A 139 (2) p246-257.
29. Graybill, F.A. (1961) *An introduction to linear statistical models, Vol 1.* McGraw-Hill, New York.
30. Graybill, F.A. (1976) *Theory and application of the linear model.* Duxbury Press, Massachusetts.
31. Gurney, M. and Daly, J.F. (1965) *A multivariate approach to estimation in periodic sample surveys.* Proc. of the Soc. Statist. Section. Amer. Statist. Assoc. p242-257.
32. Hannan, E.J. (1970) *Multiple time series.* Wiley, New York.
33. Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953) *Sample survey methods and theory.* Wiley, New York.

34. Hansen, M.H., Hurwitz, W.N., Nisselson, H. and Steinberg, J. (1955) *The redesign of the census current population survey*. J. Amer. Statist. Assoc. 50 p701-719.
35. Harville, D. (1976) *Extention of the Gauss-Markov theorem to include the estimation of random effects*. Ann. of Statist. 4 (2) p384-395.
36. Haslett, S.J. (1984) *Data projecting and parameter updating in linear models: five SAS macros with hints*. Proc of the SUNZ '83 Conference, Wellington, Nov 1983. SAS Institute (NZ). p169-183.
37. Haslett, S.J. (1985)(a) *The linear non-homogeneous estimator in sample surveys*. Sankhya B 47 (1) p101-117.
38. Haslett, S.J. (1985)(b) *Recursive estimation of the general linear model with dependent errors and multiple additional observations*. Aust. J. Statist. 27 (2) p183-188.
39. Haslett, S.J. (1985)(c) *A SAS macro for maximum likelihood estimation of ARMA processes using the EM algorithm*. Proc of the SUNZ '84 Conference, Wellington, Nov 1984. SAS Institute (NZ). p218-231.
40. Helgeland, J. (1982) *Time series models for repeated sample surveys*. in Time Series Analysis : Theory and Practice I, editor O.D. Anderson. North Holland, Amsterdam.

41. Holt, D., Smith, T.M.F. and Winter, P.D. (1980) *Regression analysis of data from complex surveys*. J. Roy. Statist. Soc. A 143 (4) p474-487.
42. Jain, A.K. (1978) *Estimation from a stratified random sample under changes in strata composition*. Proc. Section on Survey Research Methods. Amer. Statist. Assoc. p642-646.
43. Jain, A.K. (1981) *Design of a rotation scheme for a stratified multi-stage sample*. J. Statist. Planning and Inference. 5 p57-69.
44. Jessen, R.J. (1942) *Statistical investigation of a sample survey for obtaining farm facts*. Iowa Agric. Exp. Sta. Res. Bull. 304.
45. Jones, R.G. (1975) *Time series analysis of repeated surveys*. M.Sc. Dissertation, University of Southampton.
46. Jones, R.G. (1979) *The efficiency of time series estimators for repeated surveys*. Aust. J. of Statist. 21 (1) p45-56.
47. Jones, R.G. (1980) *Best linear unbiased estimators for repeated surveys*. J. Roy. Statist. Soc. B 42 (2) p221-226.
48. Jones, R.G. (1983) *Estimation procedures for repeated surveys*. Ph.D Thesis, Australian National University, Canberra

49. Judge, G.G., Griffiths, W.E., Hill, R.C. and Lee, T-C. (1980) *The theory and practice of econometrics*. Wiley, New York.
50. Koopmans, L.H. (1974) *The spectral analysis of time series*. Academic Press, New York.
51. Kulldorff, G. (1963) *Some problems in optimal allocation for sampling on two occasions*. Rev. Int. Statist. Inst. 31 p24-57.
52. Lenz, H.J. and Robinson, P.M. (1983) *Sampling of cross-sectional time series*. in Recent Trends in Statistics, editor S. Heiler. Vandenhoeck and Ruprecht, Gottingen. p56-68.
53. Ljung, G.M. and Box, G.E.P. (1978) *On a measure of lack of fit of time series models*. Biometrika 65 p297-303.
54. McGilchrist, C.A. and Sandland, R.L. (1979) *Recursive estimation of the general linear model with dependent errors*. J. Roy. Stat. Soc. B 41 p65-68.
55. Maling, T.J.B. and Dollery, C.T. (1980) *The effect of acute and chronic B-adrenoceptor blockade on sympathetic activity*. Australia New Zealand J. of Medicine 10 (2) p261.
56. Manousakis, E. (1977) *Repeated sampling with partial replacement of units*. Annals of Statist. 5 p795-802.

57. Mood, A.M., Graybill, F.A. and Boes, D.C. (1974) *Introduction to the theory of statistics*. McGraw-Hill International, Tokyo. 3rd edition.
58. Narain, R.D. (1954) *The general theory of sampling on successive occasions*. Bull. Internat. Statist. Instit. 34 p87-89.
59. Nathan, G. and Holt, D. (1980) *The effect of survey design on regression analysis*. J. Roy. Statist. Soc. B 42 (3) p377-386.
60. Patterson, H.D. (1950) *Sampling on successive occasions with partial replacement of units*. J. Roy. Statist. Soc. B 12 (2) p241-255.
61. Pfeiffermann, D. (1984) *On extensions of the Gauss-Markov theorem to the case of stochastic regression coefficients*. J. Roy. Statist. Soc. B 46 (1) p139-148.
62. Phillips, G.M. and Taylor, P.J. (1973) *Theory and applications of numerical analysis*. Academic Press, London.
63. Plackett, R.L. (1950) *Some theorems in least squares*. Biometrika 37 p149-157.
64. Porter, R.D. (1973) *On the use of survey sample weights in the linear model*. Annals of Economic Social Measurement 2 (2) p141-158.

65. Priestley, M.B. (1981) *Spectral analysis and time series*. Academic Press, London.
66. Quenouille, M.H. (1950) *An application of least squares to family diet surveys*. *Econometrica* 18 p27-44.
67. Raj, D. (1964) *On double sampling for pps estimation*. *Ann. Math. Statist.* 35 p900-902.
68. Raj, D. (1965) *On sampling over two occasions with pps*. *Ann. Math. Statist.* 36 p326-330.
69. Raj, D. (1968) *Sampling theory*. McGraw-Hill, New York.
70. Rao, C.R. (1973) *Linear statistical inference and its applications*. Wiley, New York. 2nd edition.
71. Rao, J.N.K. and Graham, J.E. (1964) *Rotation designs for sampling on repeated occasions*. *J. Amer. Statist. Assoc.* 59 p492-509.
72. Ratkowsky, D.A. (1983) *Nonlinear regression modelling: a unified approach*. Marcel Dekker, New York.
73. Robinson, P.M. (1978) *Statistical inference for a random coefficient autoregressive model*. *Scand. J. Statist.* 5 p163-168.

74. Robinson, P.M. (1982)(a) *On the convergence of the Horvitz-Thompson estimator.* Aust. J. of Statist. 24 (2) p234-238.
75. Robinson, P.M. (1982)(b) *Economic time series analysis and sample survey theory.* in Games, Economic Dynamics and Time Series Analysis, editors M. Deistler et al. Physica-Verlag, Wurzburg. p322-333.
76. Robinson, P.M. (1985) *Approximate optimal allocation in repeated surveys from a finite population.* J. of Statist. Planning and Inference 11 (2) p135-148.
77. Robinson, P.M. (1986) *On a model for a time series of cross sections.* in Essays in Time Series and Allied Processes, editors J. Gani and M.B. Priestley. J. Applied Prob. Special Vol. 23A p113-126.
78. Royall, R.M. and Herson, J. (1973)(a) *Robust estimation in finite populations I.* J. Amer. Statist. Assoc. 68 p880-889.
79. Royall, R.M. and Herson, J. (1973)(b) *Robust estimation in finite populations II; stratification on a size variable.* J. Amer. Statist. Assoc. 68 p890-893.
80. Sandland, R.L. (1980) *Stochastic growth processes.* Ph.D. Thesis, University of New South Wales.
81. SAS Institute Inc. (1982) *SAS/ETS Users Guide.* SAS Institute Inc., Cary, North Carolina. 1982 edition.

82. SAS Institute Inc. (1982) *SAS Users Guide: Basics*. SAS Institute Inc., Cary, North Carolina. 1982 edition.
83. SAS Institute Inc. (1982) *SAS Users Guide: Statistics*. SAS Institute Inc., Cary, North Carolina. 1982 edition.
84. Scott, A.J., Brewer, K.R.W. and Ho, E.W.H. (1978) *Finite population sampling and robust estimation*. J. Amer. Statist. Assoc. 73 p359-361.
85. Scott, A.J. and Holt, D. (1982). *The effect of two stage sampling on ordinary least squares methods*. J. Amer. Statist. Assoc. 77 p848-854.
86. Scott, A.J. and Smith, T.M.F. (1974) *Analysis of repeated surveys using time series techniques*. J. Amer. Statist. Assoc. 69 p674-678.
87. Scott, A.J., Smith, T.M.F. and Jones, R.G. (1977) *The application of time series methods to the analysis of repeated surveys*. Internat. Statist. Review 45 p13-28.
88. Seber, G.A.F. (1977) *Linear Regression Analysis*. Wiley, New York.
89. Sen, A.R. (1972) *Successive sampling with p ($p \geq 1$) auxilliary variables*. Annals of Math. Statist. 43 (6) p2031-2034.
90. Shapiro, S.S. and Wilk, M.B. (1965) *An analysis of variance test for normality (complete samples)*. Biometrika 52 p591-611.

91. Singh, D. (1968) *Estimates in successive sampling using a multistage design*. J. Amer. Statist. Assoc. 63 p99-112.
92. Swamy, P.A.V.B. (1971) *Statistical inference in random coefficient regression models*. Springer-Verlag, New York.
93. Terasvirta, T. (1977) *The invertibility of sums of discrete MA and ARMA processes*. Scand. J. of Statist. 4 p165-170.
94. Tikkiwal, B.D. (1960) *On the theory of classical regression and double sampling estimation*. J. Roy. Statist. Soc. B 22 p131-138.
95. Tikkiwal, B.D. (1967) *Theory of multiphase sampling from a finite or an infinite population on successive occasions 1,2*. Review of the Internat. Statist. Instit. 35 (3) p247-263.
96. Whittle, P. (1963) *Prediction and estimation by least squares methods*. English University Press, London.
97. Wolter, K.M. (1979) *Composite estimation in finite populations*. J. Amer. Statist. Assoc. 74 p604-613.
98. Woodruff, R.S. (1963) *The use of rotating samples in the Census Bureau's monthly samples*. J. Amer. Statist. Assoc. 58 p454-467.

99. Yates, F. (1949) *Sampling methods for censuses and surveys*. Griffin and Co. London.
100. Yates, F. (1960) *Sampling methods for censuses and surveys*. Griffin and Co. London 2nd edition.
101. Yates, F. and Grundy, P.M. (1953) *Selection without replacement from within strata with probability proportional to size*. J. Roy. Statist. Soc. 15 p253-261.