ESTIMATING SEEMINGLY UNRELATED REGRESSION MODELS FROM INCOMPLETE CROSS-SECTION/TIME-SERIES DATA

BY
ERIK BIØRN

STATISTISK SENTRALBYRÅ
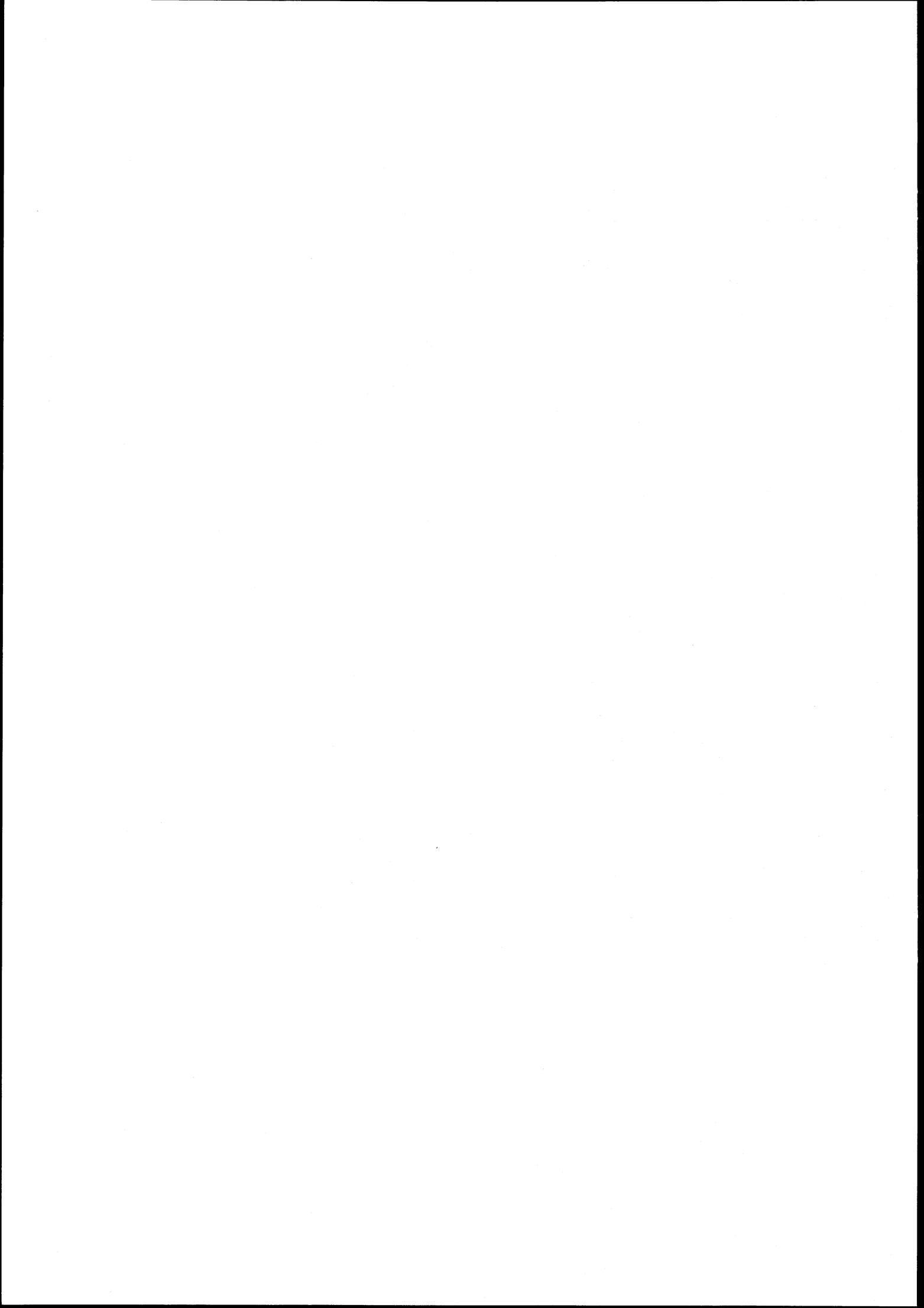CENTRAL BUREAU OF STATISTICS OF NORWAY

ESTIMATING SEEMINGLY UNRELATED REGRESSION MODELS FROM INCOMPLETE CROSS-SECTION/TIME-SERIES DATA

BY
ERIK BIØRN

PREFACE

Most of the theoretical contributions to models for handling combined cross-section/time-series data are based on the assumption that complete time-series of equal length exist for all observation units. This paper is concerned with the estimation of multi-equation models in situations where the observation units "rotate" over time. The data then become incomplete cross-section/time-series data. The situation with complete cross-section/time-series data emerges as a special case of this specification.

Econometric work with the Norwegian Surveys of Consumer Expenditure, which are based on rotating panels, has been a main motivation for exploring these problems. The conclusions, however, have for wider applicability.

A substantial part of the work reported in this paper was carried out during the author's visit to Institut National de la Statistique et des Etudes Economiques (INSEE), Paris, in the autumn 1980.

Central Bureau of Statistics, Oslo, 26 November 1981

Arne Øien

Mesteparten av de teoretiske bidrag som har vært gitt til modeller for behandling av kombinerte tverrsnitts-tidsseriedata, bygger på den forutsetning at komplette og like lange tidsserier foreligger for alle observasjonsenhetene. Denne rapporten behandler estimering av flerligningsmodeller i tilfelle hvor utvalget av observasjonsenheter "roterer" fra periode til periode. Dette gir et datamateriale i form av ufullstendige tverrsnitts-tidsseriedata. Situasjonen med komplette tverrsnitts-tidsseriedata framtrer som et spesialtilfelle av denne spesifikasjonen.

Økonometrisk arbeid med de norske forbruksundersøkelsene, som bygger på roterende utvalg, har gitt støtet til å utforske disse problemene noe nærmere. Konklusjonene har imidlertid mer generell gyldighet.

En vesentlig del av arbeidet som ligger til grunn for denne rapporten, ble utført under forfatterens studieopphold ved Institut National de la Statistique et des Études Économiques (INSEE), Paris, høsten 1980.

Statistisk Sentralbyrå, Oslo, 26. november 1981

Arne Øien

CONTENTS

ABSTRACT

The paper is concerned with the specification and estimation of multi-equation models in situations where the observation units "rotate" over time. The data then become incomplete cross-section/time-series data. The situation with complete cross-section/time-series data and several other practically interesting situations emerge as special cases of this specification. A procedure for iterative Full Information Maximum Likelihood estimation is outlined. Most attention is devoted to a disturbance components specification with individual components included and time specific components omitted.

1. INTRODUCTION

A conspicous feature of applied econometrics in recent years is the increasing trend in the utilization of combined cross-section/time-series (CS/TS) data. This reflects primarily the increasing availability of individual time series for empirical work, but also the growing theoretical insight in models and methods for handling such data. These include models with unobserved individual effects in general, and error components regression models in particular.

The well-known study of Balestra and Nerlove (1966), one of the first applications of error components models in econometrics, was concerned with the situation where N observation units are observed in T consecutive time periods. The stochastic disturbance in the regression equation considered was decomposed in two additive and independent parts, the first is specific to the observation unit - *the individual component* - while the second represents the combined effect of the observation unit and the time period - *the combined component*. An extended three components model which treats the time 'dimension' and the individual 'dimension' symmetrically - by including a *time (period) specific component* along with the individual and the combined ones - has been extensively discussed in several subsequent articles, e.g. Wallace and Hussain (1969), Maddala (1971), Mazodier (1971), Nerlove (1971b), and Mundlak (1978). In recent years, this three components, single equation regression model has been further extended to multi-equation models formulated as a system of seemingly unrelated regressions, in articles by Avery (1977) and Baltagi (1980).

A common feature of all the articles mentioned above (and a lot of others) is the assumption that the observation matrix is rectangular: the same N units are under observation in the same T successive time periods. The data set constitutes *complete cross-section/time-series data*. This specification is in fact so firmly established in the literature that the term *combined* CS/TS data has almost become synonymous with *complete* CS/TS data. Analytical and notational convenience may partly account for this. Another reason may be the fact that in many practical applications of CS/TS data, the observation units are some sorts of aggregates; in the Balestra-Nerlove (1966) study, for instance, the units were the different states in the USA. In such cases, complete time-series are readily available.

However, in the context of *sampling surveys data* from a population of genuine micro units, e.g. households or firms, this may be a rather extreme and unrealistic specification. Occasionally, the econometrician has access to socalled panel data (longitudional data), but owing to the serious problems of *non-response* when repetitive and time-consuming reporting is involved, this will in practice often be the exception rather than the rule. For this reason, models based on *incomplete CS/TS data* are well worth considering. The econometrician's philosophy should always be to construct his model according to the data situation actually at hand, rather than forcing his data into the straitjacket of preconceived simplistic model schemes. If he follows the latter strategy- e.g. by throwing away all individuals from which complete time series do not exist - he will usually waste a lot of information.

Some aspects of the problem of specification and estimation of *single equation* error components regression models from incomplete CS/TS data are discussed in Biørn (1981). The purpose of the present paper is to generalize this approach to a *multi-equation* model, specifically, a model written in the 'seemingly unrelated regressions' format. This is, formally, a multi-equation model in which all equations contain only one endogenous variable and with only predetermined variables occurring on the right hand side of each equation.

Rotating panels are applied in practice by several statistical agencies. The Central Bureau of Statistics of Norway, for instance, has followed this sort of sampling design for its household budget surveys since the year 1975. (About 25 per cent of the respondents in one year are asked to report again in the following year.) A main purpose is to reduce the degree of non-response. An econometric study based on these data is documented in Biørn and Jansen (1982). This work has, however, motivated a closer examination of the problems raised by incomplete CS/TS data in the context of multi-equation models with error components. It is the main results of this more general approach which are reported in the present paper. The conclusions are not confined to the field of micro consumer analysis, but

have, of course, far wider applicability.

The paper is organized as follows: Section 2 contains a brief general description of the model and the sampling design. Section 3 deals with the structure of the disturbance covariance matrix. We distinguish between the case where three disturbance components are included in all the equations and the case where time specific components are omitted. The full three components model is shown to contain all the models referred to above (and a lot of others) as special cases. Estimation procedures are discussed in section 4. We first state the Full Information Maximum Likelihood (FIML) estimation strategy in general terms for the full three components version of the model. Then we discuss more closely an iterative FIML estimation procedure for the simpler two components specification. Finally, in the concluding section 5, we attempt to place the work reported here in a wider context.

2. MODEL AND SAMPLING DESIGN

Consider a linear model with G equations, in the form of G 'seemingly unrelated regressions'. Formally, this is a linear simultaneous equations model in which each of the G endogenous variables is included in one and only one of the G equations, and with only predetermined variables on the right-hand side of each equation.[1] An example of this sort of model is a complete system of static consumer demand functions, where the volume of consumption of each commodity group is expressed as a function of income and prices (and possibly other exogenous variables). The observations on the variables are assumed to be combined CS/TS data. The i'th equation of the model thus has the form

$$(2.1) \qquad y_{iht} = x_{iht}\beta_i + e_{iht} \qquad (i=1, \ldots G),$$

Where β_i is the (column) vector of coefficients in the i'th equation, y_{iht} is the value of the i'th en-endogenous variable, x_{iht} is the (row) vector of exogenous (non-stochastic)[2] variables in the i'th equation, and e_{iht} is the corresponding disturbance. The subscripts h and t indicate the individual (micro unit) and time period, respectively. At this stage, we assume that h and t can be any positive integer, i.e., the model applies to all individuals in the population at all points of time.

We decompose e_{iht} into three additive components, an individual component u_{ih}, a time specific component v_{it} and a combined component (a remainder) w_{iht}:

$$(2.2) \qquad e_{iht} = u_{ih} + v_{it} + w_{iht} \qquad \text{for all h and t, and } i=1, \ldots, G.$$

All components have zero expectations:

$$(2.3) \qquad E(u_{ih}) = E(v_{it}) = E(w_{iht}) \qquad \text{for all h and t, and } i=1, \ldots, G,$$

and their second order moments satisfy

$$(2.4) \qquad E(u_{ih} u_{jk}) = \delta_{hk} \sigma_{ij}^u,$$

$$(2.5) \qquad E(v_{it} v_{js}) = \delta_{ts} \sigma_{ij}^v,$$

$$(2.6) \qquad E(w_{iht}w_{jks}) = \delta_{hk}\delta_{ts}\sigma_{ij}^w,$$

---

1) Cf. Zellner (1962) and Zellner and Huang (1962), for a general discussion of seemingly unrelated regression models.
2) We assume that no lagged endogenous variables are included. For a discussion of problems caused by the presence of such variables in single equation error components models, see Nerlove (1971 a).

(2.7) $E(u_{ih} \, v_{js}) = E(u_{ih} w_{jks}) = E(v_{it} \, w_{jks}) = 0$, for all h, k, t, s, and for i, j=1, ..., G.

where $\delta_{hk} = 1$ for k = h, and 0 for k $\neq$ h; $\delta_{ts} = 1$ for s = t, and 0 for s $\neq$ t. These assumptions have the following implications:

(i) Homoscedasticity of all disturbance components:

$$var \, (u_{ih}) = \sigma_{ii}{}^{u} \qquad \text{for all i and h,}$$

$$var \, (v_{it}) = \sigma_{ii}{}^{v} \qquad \text{for all i and t,}$$

$$var \, (w_{iht}) = \sigma_{ii}{}^{w} \qquad \text{for all i, h, and t.}$$

(ii) Constant covariances between disturbance components relating to different equations, but to the same individual or time period:

$$cov \, (u_{ih}, u_{jh}) = \sigma_{ij}{}^{u} \qquad \text{for all i, j, and h,}$$

$$cov \, (v_{it}, v_{jt}) = \sigma_{ij}{}^{v} \qquad \text{for all i, j, and t,}$$

$$cov \, (w_{iht}, w_{jht}) = \sigma_{ij}{}^{w} \qquad \text{for all i, j, h, and t.}$$

(iii) No correlation between disturbance components relating to different individuals and/or periods.

These assumptions are straightforward generalizations of those in the single equation model considered in Biørn (1981).

Assumptions (2.2) - (2.7) imply

(2.8) $E(e_{iht}) = 0$

$$(2.9) \qquad E(e_{iht} \, e_{jks}) = \begin{cases} \sigma_{ij}{}^{u} + \sigma_{ij}{}^{v} + \sigma_{ij}{}^{w} & \text{for k=h, s=t,} \\ \sigma_{ij}^{u} & \text{for k=h, s}\neq\text{t,} \\ \sigma_{ij}^{v} & \text{for k}\neq\text{h, s=t,} \\ 0 & \text{for k}\neq\text{h, s}\neq\text{t.} \end{cases}$$

The individuals are selected according to the following *sampling plan*: Let all individuals in the population be numbered consecutively, and let the sample in period 1 consist of individuals 1, 2, ..., N. In period 2, individuals 1, 2, ..., m ($0 \leq m \leq N$) are replaced by individuals N+1, N+2, ..., N+m. This procedure - dropping the first m individuals from the sample selected in the previous period and augmenting it by drawing m individuals from the population so that the sample size remains the same - continues in all the following periods until period T. In general, the sample in period t thus consists of the individuals with numbers (t-1)m+1, (t-1)m+2, ..., (t-1) m+N (t=1, 2, ..., T). The total number of individuals observed is

(2.10)   H = (T-1)m+N,

and the total number of observations is TN.

The special case with m=0, i.e. H=N, is the situation with *completely overlapping samples* studied by Avery (1977) and Baltagi (1980). At the other extreme, m = N, i.e. H = TN, corresponds to a sampling design with *non-overlapping samples*; all individuals are observed only once. If

0<m<N, we have a situation with *rotating (partly overlapping) samples*. In this intermediate situation, the data set will include some individuals observed only once, another subset observed twice, and - provided that 0<m<N/2 - still another subset observed three times, etc. If the 'rotation parameter' m exceeds half the size of each sample, $N/2 \leq m<N$, no individual is observed more than twice.

All sampling designs with 0<m<N thus have the common characteristic that the data set shows variation both along the 'time dimension' and the 'individual dimension', without giving complete time series of cross section data. To fix ideas, we have illustrated five typical cases in Figure 1 below for N = 1 000 and T = 5. We let the horizontal axis indicate the number of the individual and let the vertical axis represent the time period. Each asterisk represents N/4 = 250 observations. Case A illustrates the standard situation with completely overlapping samples, case E represents the other extreme situation with non-overlapping samples, whereas cases B, C, and D illustrate different designs with rotating samples.
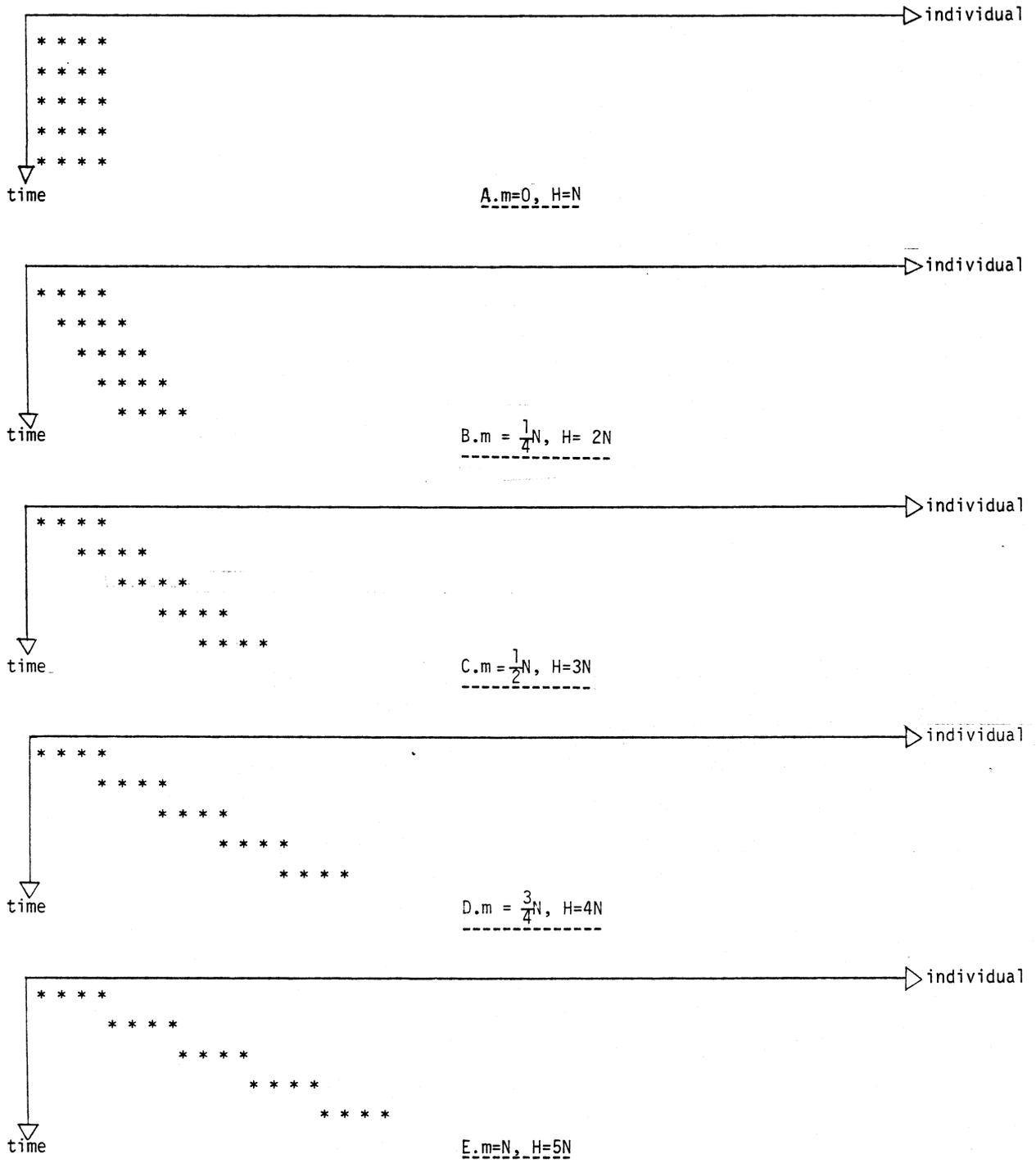
10

```
                                                    ▷individual
  │
  │ * * * *
  │ * * * *
  │ * * * *
  │ * * * *
  ▽ * * * *
time                        A.m=0, H=N


                                                    ▷individual
  │
  │ * * * *
  │   * * * *
  │     * * * *
  │       * * * *
  ▽         * * * *
time                        B.m = ¼N,  H= 2N


                                                    ▷individual
  │
  │ * * * *
  │   * * * *
  │     * * * *
  │       * * * *
  ▽         * * * *
time                        C.m = ½N,  H=3N


                                                    ▷individual
  │
  │ * * * *
  │   * * * *
  │     * * * *
  │       * * * *
  ▽         * * * *
time                        D.m = ¾N,  H=4N


                                                    ▷individual
  │
  │ * * * *
  │     * * * *
  │       * * * *
  │         * * * *
  ▽           * * * *
time                        E.m=N,  H=5N
```

Figure 1. Five typical sampling designs. N=1 000, T=5. Each asterisk (*) represents 250 observations.

3. THE DISTURBANCE COVARIANCE MATRIX

Before discussing estimation procedures, we need an analytical expression for the disturbance covariance matrix of the TN observations in our data set. This immediately raises the question how to arrange the GTN disturbances - TN disturbances from each of the G equations in the model - by equation number (i), individual (h), and time period (t). For obvious reasons, this is a more critical problem when working with incomplete cross-section/time-series data than it is in the more 'tidy' situation where all time series are complete. No general principle can be recommended for all situations. Which ordering of the disturbance vector will give the most tractable expression for the covariance matrix depends essentially on whether time specific disturbance components are included or not. If such components are present, the succession in time of the different observations (i.e. the subscript t) is, of course, crucial for the characterization of the covariance structure. If they are omitted, we are free to arrange the observations in a different order.

We shall therefore divide the following discussion into two parts. First, in section 3.1, we establish general formulae for the covariance matrix by employing a set of selection matrices to indicate which individuals in the population are included in the sample in the different periods. Then, in section 3.2, we examine more closely the case where time specific components are excluded. We show that, with a suitable reordering of the observations, the disturbance covariance matrix assumes a simple block-diagonal form, which will be convenient for the ensuing discussion of estimation problems.

3.1. The general case. Formulation by means of selection matrices

Let us start by considering the disturbance covariance matrix in the *hypothetic* case where individuals 1, 2, ..., $H[= (T - 1)m + N]$ are observed in *all* the T periods under consideration, i.e. the matrix which corresponds to complete CS/TS data with a panel consisting of H individuals. Arrange all disturbances from the i'th equation of the model in the H x T matrix

$$(3.1) \qquad \tilde{e}_i = \begin{pmatrix} e_{i11} \cdots\cdots\cdots\cdots e_{i1T} \\ \vdots \qquad\qquad\qquad \vdots \\ \vdots \qquad\qquad\qquad \vdots \\ e_{iH1} \cdots\cdots\cdots\cdots e_{iHT} \end{pmatrix} = (e_{i1} \cdots\cdots e_{iT}) \qquad (i = 1, ..., G),$$

its t'th column, e_{it}, containing the disturbances in period t from all the H individuals. According to the general specification of the variances and covariances, (2.9), we have

$$(3.2) \qquad E(e_{it}\, e_{js}') = \begin{cases} (\sigma_{ij}^u + \sigma_{ij}^w)I_H + \sigma_{ij}^v E_H & \text{for } s = t \quad \begin{pmatrix} i, j = 1, ..., G \\ t, s = 1, ..., T \end{pmatrix}, \\[2ex] \sigma_{ij}^u\, I_H & \text{for } s \neq t \end{cases}$$

where I_H is the H x H identity matrix and E_H is the H x H matrix with all elements equal to one.

Defining the TH x 1 vector[3]

$$(3.3) \qquad e_i = vec(\tilde{e}_i) = \begin{pmatrix} e_{i1} \\ \vdots \\ \vdots \\ e_{iT} \end{pmatrix} \qquad (i = 1, ..., G),$$

---

3) The operator 'vec' denotes vectorization, i.e. the stacking of all columns of a matrix into one long vector, and $\otimes$ is the Kronecker product operator. A list of main properties of Kronecker products, some of which will be utilized in the following, is given in Balestra (1975, section 2.1).

eq. (3.2) can alternatively be written as

(3.4) $E(e_i e_j') = \sigma_{ij}{}^W I_T \otimes I_H + \sigma_{ij}{}^V I_T \otimes E_H + \sigma_{ij}{}^U E_T \otimes I_H.$

The difference in notation apart, this equation is identical with eq. (26) in Avery (1977). (Recall that Avery, following the single equation models of Wallace and Hussain (1969) and Nerlove (1971b), has ordered the disturbances first by individual, second by period, whereas we have chosen the reverse ordering.)

The covariance matrix of the complete GTH x 1 disturbance vector that would result if all the H individuals were observed in all the T periods,

(3.5) $e = \begin{pmatrix} e_1 \\ \vdots \\ e_G \end{pmatrix} = \begin{pmatrix} vec(\tilde{e}_1) \\ \vdots \\ vec(\tilde{e}_G) \end{pmatrix}$,

can thus be written compactly as[4]

(3.6) $E(ee') = \Sigma_W \otimes I_T \otimes I_H + \Sigma_V \otimes I_T \otimes E_H + \Sigma_U \otimes E_T \otimes I_H$,

where

$$\Sigma_u = \begin{pmatrix} \sigma_{11}{}^u & \cdots\cdots & \sigma_{1G}{}^u \\ \vdots & & \vdots \\ \sigma_{G1}{}^u & \cdots\cdots & \sigma_{GG}{}^u \end{pmatrix},$$

$$\Sigma_v = \begin{pmatrix} \sigma_{11}{}^v & \cdots\cdots & \sigma_{1G}{}^v \\ \vdots & & \vdots \\ \sigma_{G1}{}^v & \cdots\cdots & \sigma_{GG}{}^v \end{pmatrix},$$

$$\Sigma_w = \begin{pmatrix} \sigma_{11}{}^w & \cdots\cdots & \sigma_{1G}{}^w \\ \vdots & & \vdots \\ \sigma_{G1}{}^w & \cdots\cdots & \sigma_{GG}{}^w \end{pmatrix}.$$

Now, a subset, consisting of N of the H individuals in this hypothetical panel, are *actually* selected. Define, in analogy with (3.1), the N x T matrix

(3.7) $\tilde{\epsilon}_i = \begin{pmatrix} \epsilon_{i11} & \cdots\cdots & \epsilon_{i1T} \\ \vdots & & \vdots \\ \epsilon_{iN1} & \cdots\cdots & \epsilon_{iNT} \end{pmatrix} = (\epsilon_{i1} \cdots \epsilon_{iT})$ $(i = 1, \ldots, G),$

where ϵ_{iht} is the disturbance in eq. no i *belonging to the h'th individual of those observed in the t'th period.* (Its population number is $(t - 1)m + h$.) The t'th column of this matrix, ϵ_{it}, contains the disturbances in period t of all the individuals actually observed in this period. Obviously,

_____

4) Cf. Avery (1977, eq. (27)) and Baltagi (1980, eq. (7)).

$$\epsilon_{iht} \equiv e_{i,(t-1)m+h,t} \qquad \left(\begin{array}{l} i = 1, \ldots, G \\ h = 1, \ldots, N \\ t = 1, \ldots, T \end{array} \right) .$$

This relationship can be stated compactly in vector notation as

(3.8) $$\qquad \epsilon_{it} = D_t e_{it} \qquad \left(\begin{array}{l} i = 1, \ldots, G \\ t = 1, \ldots, T \end{array} \right)$$

where D_t is a *sample design (selection) matrix* of dimension $N \times H$ ($= N \times [(T - 1) m + N]$) defined as follows:

(3.9) $$\qquad D_t = \left(0_{N, (t-1)m} \mid I_N \mid 0_{N, (T-t)m} \right) ,$$

$0_{N,n}$ denoting the $N \times n$ zero matrix. *Element (h, k) of D_t is 1 if the k'th individual in the population is the h'th individual in the sample in period t, and zero otherwise.* In completely overlapping samples ($m = 0$), we have in particular

(3.9a) $$\qquad D_t = I_N \qquad\qquad (t = 1, \ldots, T),$$

whereas non-overlapping samples ($m = N$) correspond to

(3.9b) $$\qquad D_t = (0_{1, t-1} \quad 1 \quad 0_{1, T-t}) \otimes I_N.$$

Since $D_t D_t' = I_N$ and $D_t E_H D_t' = E_N$ for all t, *irrespective of the sampling design*, it follows from (3.2) and (3.8) that

(3.10) $$E(\epsilon_{it} \, \epsilon_{js}') = D_t E(e_{it} \, e_{js}') \, D_s'$$

$$= \begin{cases} (\sigma_{ij}^u + \sigma_{ij}^w) \, I_N + \sigma_{ij}^v E_N & \text{for } s = t \\[2ex] \sigma_{ij}^u \, D_t D_s' & \text{for } s \neq t \end{cases}$$

$$(i, j, = 1, \ldots, G).$$

From this equation it is evident that σ_{ij}^u and σ_{ij}^w are not identifiable unless $D_t D_s' \neq 0_{N, N}$ for at least one $s \neq t$, which implies that at least two of the T samples overlap. If $D_t D_s' = 0_{N, N}$ for all t and $s \neq t$, i.e. all samples are non-overlapping, only σ_{ij}^v and $\sigma_{ij}^u + \sigma_{ij}^w$ can be identified from the data.

Defining the $TN \times 1$ vector

(3.11) $$\qquad \epsilon_i = \text{vec} \, (\tilde{\epsilon}_i) = \left(\begin{array}{c} \epsilon_{i1} \\ \vdots \\ \epsilon_{iT} \end{array} \right) \qquad\qquad (i = 1, \ldots, G),$$

and the stacked $TN \times H$ selection matrix

(3.12) $$\qquad D = \left(\begin{array}{c} D_1 \\ \vdots \\ D_T \end{array} \right) ,$$

eq. (3.10) can alternatively be written in the form

$$(3.13) \qquad E(\epsilon_i \epsilon_j') = \sigma_{ij}{}^W I_T \otimes I_N + \sigma_{ij}{}^V I_T \otimes E_N + \sigma_{ij}{}^u DD'$$

$$(i, j = 1, \ldots, G).$$

The covariance matrix of the stacked GTN x 1 vector

$$(3.14) \qquad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_G \end{pmatrix} = \begin{pmatrix} \text{vec } (\tilde{\epsilon}_1) \\ \vdots \\ \text{vec } (\tilde{\epsilon}_G) \end{pmatrix} \quad ,$$

can then be written compactly as

$$(3.15) \qquad E(\epsilon\epsilon') = \Omega = \Sigma_w \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N + \Sigma_u \otimes DD'.$$

This is the *general* expression for the disturbance covariance matrix in seemingly unrelated regressions models based on combined cross-section/time-series data.

In the special case with completely overlapping samples, it follows from (3.9a) and (3.12) that $D = 1_T \otimes I_N$, where 1_T denotes the T x 1 vector with all elements equal to one. Then $DD' = (1_T 1_T') \otimes I_N = E_T \otimes I_N$, and the covariance matrix becomes

$$(3.15a) \qquad \Omega = \Omega_C = \Sigma_w \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N + \Sigma_u \otimes E_T \otimes I_N,$$

the subscript C denoting 'complete'. Not surprisingly, we find that this expression is identical with (3.6) with H set equal to N. At the other extreme, non-overlapping samples are characterized by $D = I_T \otimes I_N$ (cf. (3.9b) and (3.12)), which implies the following covariance matrix:

$$(3.15b) \qquad \Omega = \Omega_N = (\Sigma_w + \Sigma_u) \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N,$$

the subscript N denoting 'non-overlapping'. This expression is equivalent to (3.6) with H set equal to N, Σ_w replaced by $\Sigma_w + \Sigma_u$, and with the last term omitted.

3.2. The two components case: Time specific components excluded

In the previous section, we gave a characterization of the covariance structure of the disturbances in the general case where both the individual (u_{ih}), the time specific (v_{it}), and the remainder disturbance component (w_{iht}) are included. This specification is theoretically attractive, not least because of its symmetry. However, for *practical* applications of disturbance components models in the context of genuine micro data, time specific effects are usually of far less importance than individual ones. This reflects, *inter alia*, the fact that the number of individuals by far exceeds the number of periods involved (e.g. T < 5 and N > 100). Then Σ_u and Σ_w dominate over Σ_v, and for practical purposes the latter may be ignored without essential loss.

In this section, we take a closer look at the covariance structure when $\Sigma_v = 0_{G,G}$. Of course, from (3.15) it then follows that

$$(3.16) \qquad E(\epsilon\epsilon') = \Omega = \Sigma_w \otimes I_T \otimes I_N + \Sigma_u \otimes DD'.$$

However, the ordering of the disturbances which leads to this expression for the covariance matrix does not exploit the fact that when time specific effects are ignored, the actual succession in time of the different observations is irrelevant. We can, without fear of ambiguity, drop the time subscript from

the model specification and rearrange the disturbance vector.[5] This is what we shall do in this section.

We arrange the individuals in groups according to the number of times each individual is under observation and identify each individual by its number in the group to which it belongs. Let, in general, $H(p)$ denote the number of individuals in the p'th group, i.e. the number of individuals observed p times. Let P denote the maximal number of replications. We arrange the $H(1)$ observations from the individuals observed only once in the first group, the $H(2)$ individuals observed twice and the corresponding $2H(2)$ observations in the second group, ..., the $H(P)$ individuals observed P times and the corresponding $PH(P)$ observations in the P'th group. Obviously, we have[6]

$$(3.17) \qquad \sum_{p=1}^{P} H(p) = H = \text{total number of individuals observed,}$$

and

$$(3.18) \qquad \sum_{p=1}^{P} p\,H(p) = TN = \text{total number of observations.}$$

To fix ideas, let us refer the values of P and $H(p)$ corresponding to the five sampling designs illustrated in Figure 1 above. We find

Case A: $P = 5$; $H(1) = H(2) = H(3) = H(4) = 0$, $H(5) = N = 1\ 000$.

Case B: $P = 4$; $H(1) = H(2) = H(3) = H(4) = \frac{N}{2} = 500$.

Case C: $P = 2$; $H(1) = N = 1\ 000$, $H(2) = 2N = 2\ 000$.

Case D: $P = 2$; $H(1) = 3N = 3\ 000$, $H(2) = N = 1\ 000$.

Case E: $P = 1$; $H(1) = 5N = 5\ 000$.

In the following, we use (h, p) as a shorthand notation for the h'th individual of those observed p times $(h = 1, \ldots, H(p); p = 1, \ldots, P)$ and define

$\epsilon_{i(h, p)q}$: disturbance in eq. no i relating to the q'th observation from individual (h, p)

$(i = 1, \ldots, G; q = 1, \ldots, p; h = 1, \ldots, H(p); p = 1, \ldots, P)$.

The $p \times 1$ vector containing the disturbances in eq. no i relating to individual (h, p) is

$$(3.19) \qquad \epsilon_{i(h, p)} = \begin{pmatrix} \epsilon_{i(h, p)1} \\ \vdots \\ \epsilon_{i(h, p)p} \end{pmatrix} \qquad \begin{array}{l} (i = 1, \ldots, G) \\ (h = 1, \ldots, H(p)) \\ (p = 1, \ldots, P). \end{array}$$

---

5) This may necessitate a redefinition of some of the structural variables in the model. When no lagged endogenous variables are included, (cf. footnote 2), this reformulation is trivial.

6) It is not essential for this reformulation of the model that the sample size is the same, and equal to N, in all the periods. The following formulae are equally valid in more general situations where the sample size changes from period to period; only the sequence, $H(1), \ldots, H(P)$ is of importance.

Since, when $\sigma_{ij}{}^v = 0$ we have

$$
E\left[\epsilon_{i(h,\,p)q}\;\epsilon_{j(k,\,n)r}\right] =
\begin{cases}
\sigma_{ij}{}^w + \sigma_{ij}{}^u & \text{for } (k,\,n) = (h,\,p) \text{ and } r = q \\[2mm]
\hline
\sigma_{ij}{}^u & \text{for } (k,\,n) = (h,\,p) \text{ and } r \neq q \\[4mm]
0 & \text{otherwise}
\end{cases}
$$

(cf. ((2.9)), it follows that

$$
(3.20) \qquad E\left[\epsilon_{i(h,\,p)}\;\epsilon'_{j(k,\,n)}\right] =
\begin{cases}
\sigma_{ij}{}^w I_p + \sigma^u{}_{ij} E_p & \text{for } (k,\,n) = (h,\,p) \\[3mm]
0_{p,\,n} & \text{for } (k,\,n) \neq (h,\,p)
\end{cases}
$$

$$(i = 1, \ldots, G;\; j = 1, \ldots, G)$$
$$(h = 1, \ldots, H(p);\; k = 1, \ldots, H(n))$$
$$(p = 1, \ldots, P;\; n = 1, \ldots, P).$$

It is convenient to arrange all disturbances from individual $(h,\,p)$ in the $p \times G$ matrix

$$
(3.21) \qquad \tilde{\epsilon}_{(h,\,p)} =
\begin{pmatrix}
\epsilon_{1(h,\,p)1} \cdots \epsilon_{G(h,\,p)1} \\
\vdots \qquad\qquad \vdots \\
\epsilon_{1(h,\,p)p} \cdots \epsilon_{G(h,\,p)p}
\end{pmatrix}
= \left(\epsilon_{1(h,\,p)} \cdots\cdots\cdots \epsilon_{G(h,\,p)}\right)
$$

$$(h = 1, \ldots, H(p))$$
$$(p = 1, \ldots, P),$$

or alternatively in the stacked $Gp \times 1$ vector

$$
(3.22) \qquad \epsilon_{(h,\,p)} = \text{vec}(\tilde{\epsilon}_{(h,\,p)}) =
\begin{pmatrix}
\epsilon_{1(h,\,p)} \\
\vdots \\
\epsilon_{G(h,\,p)}
\end{pmatrix}
$$

$$(h = 1, \ldots, H(p))$$
$$(p = 1, \ldots, P).$$

Eq. (3.20) can then be written as

$$
(3.23) \qquad E\left[\epsilon_{(h,\,p)}\;\epsilon'_{(k,\,n)}\right] =
\begin{cases}
\Omega_p & \text{for } (k,\,n) = (h,\,p) \\[3mm]
0_{Gp,\,Gn} & \text{for } (k,\,n) \neq (h,\,p)
\end{cases}
$$

$$(h = 1, \ldots, H(p);\; k = 1, \ldots, H(n))$$
$$(p = 1, \ldots, P;\; n = 1, \ldots, P),$$

where Ω_p is the following $Gp \times Gp$ matrix:

(3.24) $\qquad \Omega_p = \Sigma_w \otimes I_p + \Sigma_u \otimes E_p$ $\qquad\qquad\qquad$ (p = 1, ..., P).

$\qquad$ Let now $\epsilon_{(p)}$ denote the H(p)Gp x 1 vector containing all disturbances from the p'th group, i.e. from all individuals observed p times:
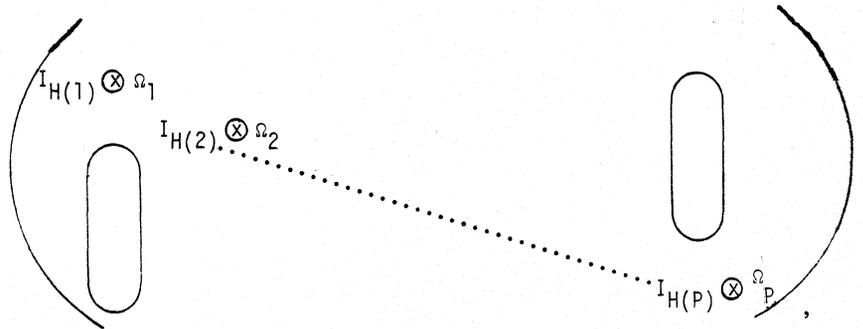
(3.25) $\qquad \epsilon_{(p)} = \begin{pmatrix} \epsilon_{(1,p)} \\ \vdots \\ \epsilon_{(H(p),p)} \end{pmatrix}$ $\qquad\qquad\qquad$ (p = 1, ..., P).

The covariance matrix of the complete disturbance vector

(3.26) $\qquad \epsilon_* = \begin{pmatrix} \epsilon_{(1)} \\ \vdots \\ \epsilon_{(P)} \end{pmatrix}$,

of dimension $\sum\limits_{p=1}^{P} H(p)Gp$ x 1, i.e. GTN x 1 (cf. (3.18)), can be written compactly as

(3.27) $\qquad E(\epsilon_* \epsilon_*') = \Omega_* =$

where Ω_p is defined as in eq. (3.24).

$\qquad$ Of course, eq. (3.27) expresses the same hypothesis regarding the second order moments of the disturbances as eq. (3.16). The difference lies in the different ordering of observations to form the vectors ϵ (cf. eqs. (3.7), (3.8), (3.11), and (3.14)) and ϵ_* (cf. eqs. (3.21), (3.22), (3.25), and (3.26)). The ordering which leads to ϵ_* implies that the corresponding covariance matrix, Ω_*, assumes a *block-diagonal* form: First the block $\Omega_1 = \Sigma_w + \Sigma_u$ is repeated H(1) times, then the block $\Omega_2 = \Sigma_w \otimes I_2 + \Sigma_u \otimes E_2$ is repeated H(2) times, the block $\Omega_3 = \Sigma_w \otimes I_3 + \Sigma_u \otimes E_3$ is repeated H(3) times, etc. Needless to say, this will be an obvious virtue when it comes to estimation.

4. ESTIMATION

$\qquad$ Following the organization of section 3, we shall also divide the discussion of estimation problems into two separate parts. First, in section 4.1 , we outline a procedure for Full Information Maximum Likelihood estimation on the basis of the general formulae for the disturbance vector ϵ and its covariance matrix Ω in section 3.1. Then, in section 4.2, we turn to the reordered disturbance vector ϵ_* , as given in section 3.2, and discuss in more detail a step-wise algorithm for estimating the two components model.

4.1. Outline of a procedure for FIML estimation in the general three components model

$\qquad$ Let y_i denote the TN x 1 vector of observations on the endogenous variable in the i'th of the G regression equations in (2.1), ordered in the same way as the disturbance vector ϵ_i (cf. (3.1), (3.8), and (3.11)), and let X_i be the corresponding matrix of observations on the exogenous variables. (We do not refer the formal definitions of y_i and X_i since they are not needed in the following: y_i is formed by an appropriate selection of the scalars y_{iht}, and X_i is constructed by an appropriate selection of the row vectors x_{iht}.) The i'th equation can then be written as

(4.1) $y_i = X_i \beta_i + \epsilon_i$ (i = 1, ..., G).

Combining all equations in one system, we get

(4.2) $y = X \beta + \epsilon$,

where ϵ is defined as in (3.14) and

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{pmatrix}, \quad X = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & X_G \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{pmatrix}.$$

Provided that the disturbances follow a multivariate normal distribution, the log-likelihood function of y is

(4.3) $L = L(y, X; \beta, \Omega) = -\dfrac{GTN}{2} \log (2\pi) - \dfrac{1}{2} \log |\Omega| - \dfrac{1}{2} \epsilon' \Omega^{-1} \epsilon$,

where Ω is defined as in (3.15), using ϵ as a shorthand notation for $y - X\beta$.

Formally, the problem of *Full Information Maximum Likelihood* (FIML) estimation of the complete model can now be expressed as follows: *Maximize L, or equivalently, minimize*

(4.4) $g = \log |\Omega| + \epsilon' \Omega^{-1} \epsilon$,

where

$$\Omega = \Sigma_w \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N + \Sigma_u \otimes DD'$$

with respect to β, Σ_u, Σ_v *and* Σ_w. In doing this, we pay regard to the definition of the stacked selection matrix D, which describes the sampling design (cf. eqs. (3.9) and (3.12)).

Direct minimization of g may, however, raise numerical problems and easily become prohibitive in terms of computer costs. If certain regularity conditions are satisfied, FIML estimates can in principle be attained - to a prescribed degree of approximation - by iteratively switching between the following two subproblems:[7]

(i) *Minimize* $Q = \epsilon' \Omega'^{-1} \epsilon$ *with respect to* β, *conditionally on* Ω.

(ii) *Minimize* $g = \log |\Omega| + \epsilon' \Omega^{-1} \epsilon$, *where* $\Omega = \Sigma_w \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N + \Sigma_u \otimes DD'$, *with respect to* Σ_u, Σ_v, *and* Σ_w, *conditionally on* β.

To start this algorithm, initial values have to be assigned to Σ_u, Σ_v and Σ_w. A convenient choice may be to set $\Sigma_u = \Sigma_v = 0$ and let Σ_w be a diagonal matrix with the i'th element equal to the estimated disturbance variance from a preliminary least squares regression on the i'th equation (i = 1, ..., G).

Solving *subproblem (i)* is equivalent to Aitken Generalized Least Squares (GLS) estimation of β. The resulting estimator is

(4.5) $\hat{\beta}_{GLS} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y$.

It is well-known that this estimator is more efficient than the one obtained by running Ordinary Least Squares (OLS) regression on each of the G equations (4.1) in all cases except the following two:[8]

---

7) If this zig-zag procedure converges towards a unique solution, it will give the FIML estimates. For a formal demonstration of this, see Oberhofer and Kmenta (1974).
8) Cf. Zellner (1962, p. 351) and Theil (1971, pp. 308 - 310). We implicitly assume that no cross-equational constraints have been imposed on the coefficient vectors β_1, ..., β_G. If such constraints apply, OLS will be inefficient also in cases (a) and (b).

(a) All G equations contain exactly the same explanatory variables - i.e. X can be written as $X = I_G \otimes \bar{X}$ - and Ω is a bloc-diagonal matrix of the form $\Omega = \Sigma \otimes I_{TN}$.

(b) No elements in the disturbance vector ϵ are correlated, i.e. Ω can be written as $\Omega = \text{diag} (\sigma_{11} \ldots \sigma_{GG}) \otimes I_{TN}$.

Neither of these specifications will be valid in a disturbance components model with either individual or time specific effects, or both, included. Thus, we can safely conclude that even the conditional GLS estimator which solves subproblem (i) is more efficient than OLS. This conclusion holds even if all equations contain exactly the same set of exogenous variable, i.e.

$$X_1 = \ldots = X_G = \underline{X} \ (\text{or } X = I_G \otimes \underline{X}).^{9)}$$

An expression for the inverse of the covariance matrix Ω is indispensible for solving subproblem (i) algebraically. Unfortunately, it is difficult to derive such an expression in the general case considered here - mainly due to the presence of the stacked selection matrix D in the expression for Ω.[10] Moreover, Ω will usually be a matrix of considerable dimension. If N = 100, T = 10, and G = 5 - to take a realistic example - then Ω will be a 5 000 x 5 000 matrix. All who are familiar with computer problems, know that the inversion of a matrix of this order is no trivial matter.

In two *special cases*, however, algebraic expressions for Ω^{-1} can be easily derived. These are the cases with *completely overlapping samples* ($m = 0$, i.e. $D = 1_T \otimes I_N$) and with *non-overlapping samples* ($m = N$, i.e. $D = I_T \otimes I_N$), cf. eqs. (3.15 a and b). By a slight reformulation, we find that the corresponding expressions for the covariance matrices can be written as

$$(4.6) \quad \Omega_C = \Sigma_w \otimes (I_T - \frac{E_T}{T}) \otimes (I_N - \frac{E_N}{N}) + (\Sigma_w + T\Sigma_u) \otimes \frac{E_T}{T} \otimes (I_N - \frac{E_N}{N})$$

$$+ (\Sigma_w + N\Sigma_v) \otimes (I_T - \frac{E_T}{T}) \otimes \frac{E_N}{N} + (\Sigma_w + T\Sigma_u + N\Sigma_v) \otimes \frac{E_T}{T} \otimes \frac{E_N}{N},$$

and

$$(4.7) \quad \Omega_N = (\Sigma_w + \Sigma_u) \otimes I_T \otimes (I_N - \frac{E_N}{N}) + (\Sigma_w + \Sigma_u + N\Sigma_v) \otimes I_T \otimes \frac{E_N}{N},$$

respectively. Now, it is easy to show that both $I_N - E_N/N$ and E_N/N are idempotent matrices and that

$$(I_N - \frac{E_N}{N}) \frac{E_N}{N} = \frac{E_N}{N} (I_N - \frac{E_N}{N}) = 0_{N,N},$$

regardless of the value of N. From this fact and familiar properties of Kronecker products[11] it then follows that[12]

$$(4.8) \quad \Omega_C^{-1} = \Sigma_w^{-1} \otimes (I_T - \frac{E_T}{T}) \otimes (I_N - \frac{E_N}{N}) + (\Sigma_w + T\Sigma_u)^{-1} \otimes \frac{E_T}{T} \otimes (I_N - \frac{E_N}{N})$$

$$+ (\Sigma_w + N\Sigma_v)^{-1} \otimes (I_T - \frac{E_T}{T}) \otimes \frac{E_N}{N} + (\Sigma_w + T\Sigma_u + N\Sigma_v)^{-1} \otimes \frac{E_T}{T} \otimes \frac{E_N}{N},$$

---

9) A similar observation has been made by Avery (1977), for the seemingly unrelated regressions model based on complete cross-section/time-series data.
10) In a recent paper, Wansbeek and Kapteyn (1981) have addressed a similar problem for a single equation model. Their formulae, however, turn out to be rather messy and do not invite attempts at generalization to multi-equation models.
11) See e.g. Theil (1971, pp. 303 - 306).
12) Cf. also Baltagi (1980), eqs. (7), (11), and Lemma. A lucid treatment of the corresponding situation in a single equation model is given in Mazodier (1971).

and

(4.9) $$\Omega_N^{-1} = (\Sigma_w + \Sigma_u)^{-1} \otimes I_T \otimes (I_N - \frac{E_N}{N}) + (\Sigma_w + \Sigma_u + N\Sigma_v)^{-1} \otimes I_T \otimes \frac{E_N}{N} \ .$$

The problem of calculating the inverse of the covariance matrix is then reduced to the problem of inverting the four GxG matrices

$$\Sigma_w,$$

$$\Sigma_w + T\Sigma_u,$$

$$\Sigma_w + N\Sigma_v,$$

and $$\Sigma_w + T\Sigma_u + N\Sigma_v,$$

in the case with completely overlapping samples, and to the problem of inverting the two GxG matrices

$$\Sigma_w + \Sigma_u$$

and $$\Sigma_w + \Sigma_u + N\Sigma_v,$$

in the case with non-overlapping samples. Inserting (4.8) and (4.9) into (4.5), we obtain the conditional GLS estimators of β which solve subproblem (i) in these two special cases.

The solution to *subproblem (ii)* is even more complicated since the minimand g contains not only the inverse of Ω, but also its determinant value, $|\Omega|$. Even in the special case with completely overlapping samples (m = 0, D = $1_T \otimes I_N$) it involves the solution of a highly non-linear equation system.[13]

By pointing out this unsolved problem, we end our discussion of estimation procedures for the general three components model. We now drop the time specific component; this will make things considerably easier, as we shall see in the next section.

4.2. Iterative FIML estimation in the two components model

Let $y_{i(h,p)}$ denote the px1 vector containing the observations on the i'th endogenous variable from individual (h,p), ordered in the same way as $\epsilon_{i(h,p)}$ (cf. eq. (3.19)), and let $X_{i(h,p)}$ be the corresponding matrix of observations on the exogenous variables. The i'th equation can then be written as

(4.10) $$y_{i(h,p)} = X_{i(h,p)} \beta_i + \epsilon_{i(h,p)}$$
$$(i = 1, \ldots, G)$$
$$(h = 1, \ldots, H(p))$$
$$(p = 1, \ldots, P).$$

Combining all G equations in one system, we have

---

[13] This problem has been thoroughly discussed by Amemiya (1971) for the corresponding single equation model. He shows, however, that simpler "analysis of variance estimates" exist, which are asymptotically equivalent to the Maximum Likelihood estimates. See also Magnus (1978) and Baltagi (1980).

(4.11) $\qquad y_{(h,p)} = X_{(h,p)} \beta + \epsilon_{(h,p)}$ $\qquad\qquad (h = 1, \ldots, H(p))$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (p = 1, \ldots, P),$

where $\epsilon_{(h,p)}$ is defined as in eq. (3.22) and

$$ y_{(h,p)} = \begin{pmatrix} y_{1(h,p)} \\ \vdots \\ y_{G(h,p)} \end{pmatrix} , \qquad X_{(h,p)} = \begin{pmatrix} X_{1(h,p)} \cdots\cdots\cdots 0 \\ \vdots \ddots \vdots \\ 0 \cdots\cdots\cdots\cdots X_{G(h,p)} \end{pmatrix} , \qquad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_G \end{pmatrix} $$

The log-likelihood function can be written in terms of ϵ_* and Ω_* (cf. (3.26) and (3.27)) as follows:

(4.12) $\qquad L = - \dfrac{GTN}{2} \log (2\pi) - \dfrac{1}{2} \log \left| \Omega_* \right| - \dfrac{1}{2} \epsilon_*{}' \Omega_*^{-1} \epsilon_*.$

From eqs. (3.23.) - (3.27) we find

(4.13) $\qquad \log \left| \Omega_* \right| = \displaystyle\sum_{p=1}^{P} H(p) \log \left| \Omega_p \right|$

$\qquad\qquad\qquad\qquad = \displaystyle\sum_{p=1}^{P} H(p) \log \left| \Sigma_w \otimes I_p + \Sigma_u \otimes E_p \right|$

and

(4.14) $\qquad \epsilon_*{}' \Omega_*^{-1} \epsilon_* = \displaystyle\sum_{p=1}^{P} Q_p,$

where

(4.15) $\qquad Q_p = \displaystyle\sum_{h=1}^{H(p)} \epsilon'_{(h,p)} \Omega_p^{-1} \epsilon_{(h,p)}$

$\qquad\qquad\qquad = \displaystyle\sum_{h=1}^{H(p)} \epsilon'_{(h,p)} [\Sigma_w \otimes I_p + \Sigma_u \otimes E_p]^{-1} \epsilon_{(h,p)}.$

Inserting (4.13) and (4.14) in (4.12) yields

(4.16) $\qquad L = - \dfrac{GTN}{2} \log (2\pi) - \dfrac{1}{2} \displaystyle\sum_{p=1}^{P} g_p,$

where

(4.17) $\qquad g_p = H(p) \log \left| \Omega_p \right| + Q_p = H(p) \log \left| \Sigma_w \otimes I_p + \Sigma_u \otimes E_p \right| + Q_p \qquad (p = 1, \ldots, P).$

In the expression for Q_p, cf. (4.15), we interpret $\epsilon_{(h,p)}$ as a shorthand for $y_{(h,p)} - X_{(h,p)}\beta$.

Thus, the log-likelihood function is, apart from an arbitrary constant, a sum of P terms, one for each of the P groups of individuals. The p'th term, $-\dfrac{1}{2} g_p$, represents those individuals which are observed p times. This separability is, of course, due to the block-diagonal structure of Ω_*, which reflects the omission of time specific disturbance components, and our particular way of ordering the elements in the disturbance vector ϵ_*.

In this simplified model, FIML estimation corresponds formally to minimizing $g = \sum_{p=1}^{P} g_p$ with respect to β, Σ_u and Σ_w, subject to (4.15) and (4.17). The iterative procedure for solving this problem consists of the following two subproblems:

(i) *Minimize* $Q = \sum_{p=1}^{P} Q_p$ *with respect to* β, *conditionally on* Ω_p ($p = 1, \ldots, P$).

(ii) *Minimize* $g = \sum_{p=1}^{P} g_p$ *with respect to* Σ_u *and* Σ_w, *conditionally on* β.

Let us consider these two subproblems in turn.

Subproblem (i): Minimization of Q with respect to β

The matrix Ω_p can be written as

$$\Omega_p = \Sigma_w \otimes (I_p - \frac{E_p}{p}) + (p\Sigma_u + \Sigma_w) \otimes \frac{E_p}{p} .$$

By utilizing the fact that $I_p - E_p/p$ and E_p/p are both idempotent for all positive integers p, we easily find that its inverse is[14]

(4.18)
$$\Omega_p^{-1} = \Sigma_w^{-1} \otimes (I_p - \frac{E_p}{p}) + (p\Sigma_u + \Sigma_w)^{-1} \otimes \frac{E_p}{p} \qquad (p = 1, \ldots, P).$$

Since $(p\Sigma_u + \Sigma_w)^{-1}$ is symmetric and non-singular, there exists a non-singular matrix S_p, of dimension $G \times G$, such that[15]

(4.19)
$$(p\Sigma_u + \Sigma_w)^{-1} = S_p{}' S_p \qquad (p = 0, 1, \ldots, P).$$

Combining (4.18) and (4.19), while again using the fact that $I_p - E_p/p$ and E_p/p are idempotent, we find that Ω_p^{-1} can be factorized as follows:

(4.20)
$$\Omega_p^{-1} = F_p{}' F_p \qquad (p = 1, \ldots, P),$$

where

(4.21)
$$F_p = S_0 \otimes (I_p - \frac{E_p}{p}) + S_p \otimes \frac{E_p}{p}$$

$$= [S_0 \otimes I_p] [I_G \otimes I_p - (I_G - S_0^{-1} S_p) \otimes \frac{E_p}{p}] \qquad (p = 1, \ldots, P).$$

Then the quadratic form relating to the p'th group of individuals, Q_p, can be written as a sum of squares

(4.22)
$$Q_p = \sum_{h=1}^{H(p)} \eta'{}_{(h,p)} \eta_{(h,p)} \qquad (p = 1, \ldots, P),$$

where $\eta_{(h,p)}$ is the $Gp \times 1$ vector of transformed disturbances

---

14) Confer the derivation of (4.8) - (4.9) from (4.6) - (4.7) above.
15) This factorization is not unique, but this is immaterial for the following application.

(4.23) $\qquad \eta_{(h,p)} = F_p \epsilon_{(h, p)}$ $\qquad$ (h = 1, .., H(p))

$\qquad$ (p = 1, ..., P).

The algorithm for solving subproblem (i) thus boils down to

(a) *Compute the matrices* $(p\Sigma_u + \Sigma_w)^{-1}$ *for p = 0, 1, ..., P, and perform the factorization (4.19)*

(b) *Compute F_p by means of (4.21) (p = 1, ..., P).*

(c) *Premultiply the observation matrices $y_{(h, p)}$ and $X_{(h, p)}$ from all individuals observed p times by the matrix F_p (p = 1, ..., P).*

(d) *Minimize the overall sum of squares of the transformed disturbances,* $Q = \sum_{p=1}^{P} Q_p = \sum_{p=1}^{P} \sum_{h=1}^{H(p)} \eta'_{(h, p)} \eta_{(h, p)}$ *with respect to* β.

Subproblem (ii): Minimization of g with respect to Σ_u and Σ_w

In order to solve the second subproblem, we need expressions for the first derivatives of $\log|\Omega_p|$ and Q_p with respect to Σ_u and Σ_w. We can obtain these rather easily by utilizing formulae for matrix derivatives developed in Balestra (1975, sections 5.2 and 5.3).[16]

From eq. (3.24) above and Balestra (1975, eq. (5.3.18)) it follows that

$$\frac{\partial \log|\Omega_p|}{\partial \Sigma_w} = [I_G \otimes (\text{vec } I_p)'] [\Omega_p^{-1} \otimes I_p] [I_G \otimes (\text{vec } I_p)],$$

$$\frac{\partial \log|\Omega_p|}{\partial \Sigma_u} = [I_G \otimes (\text{vec } I_p)'] [\Omega_p^{-1} \otimes E_p] [I_G \otimes (\text{vec } I_p)]$$

$$(p = 1, ..., P).$$

By inserting (4.18), while noting that

$$(\text{vec } I_p)' (I_p \otimes I_p) (\text{vec } I_p) = p,$$

$$(\text{vec } I_p)' (E_p \otimes I_p) (\text{vec } I_p) = (\text{vec } I_p)' (I_p \otimes E_p)(\text{vec } I_p) = p,$$

$$(\text{vec } I_p)' (E_p \otimes E_p)(\text{vec } I_p) = p^2,$$

these expressions can be simplified to

(4.24) $\qquad \dfrac{\partial \log|\Omega_p|}{\partial \Sigma_w} = (p - 1) \Sigma_w^{-1} + (p\Sigma_u + \Sigma_w)^{-1},$

(4.25) $\qquad \dfrac{\partial \log|\Omega_p|}{\partial \Sigma_u} = p (p\Sigma_u + \Sigma_w)^{-1}$

$$(p = 1, ..., P).$$

---

16) Confer also Chamberlain and Griliches (1975, appendix).

We have now two of the elements needed for calculating the first derivatives of g with respect to the covariance matrices. Let us turn to the quadratic form Q_p, as defined in (4.15). First, we note that the part of Q_p which relates to individual (h, p) can be reformulated as

$$\epsilon'_{(h, p)} \, \Omega_p^{-1} \, \epsilon_{(h, p)} = (\text{vec } \widetilde{\epsilon}_{(h, p)})' \left\{ \Sigma_w^{-1} \otimes (I_p - \tfrac{E_p}{p}) \right\} (\text{vec } \widetilde{\epsilon}_{(h, p)})$$

$$+ (\text{vec } \widetilde{\epsilon}_{(h, p)})' \left\{ (p\Sigma_u + \Sigma_w)^{-1} \otimes \tfrac{E_p}{p} \right\} (\text{vec } \widetilde{\epsilon}_{(h, p)}),$$

by utilizing (3.22) and (4.18). Since the trace of a quadratic matrix of the form $ABCD$ can be written as [17]

$$\text{tr } (ABCD) = (\text{vec } A')' \, (D' \otimes B) \, (\text{vec } C),$$

we can reformulate this equation as

$$\epsilon'_{(h, p)} \, \Omega_p^{-1} \, \epsilon_{(h, p)} = \text{tr } [\widetilde{\epsilon}'_{(h, p)} (I_p - \tfrac{E_p}{p}) \, \widetilde{\epsilon}_{(h, p)} \, \Sigma_w^{-1}]$$

$$+ \text{tr } [\widetilde{\epsilon}'_{(h, p)} \tfrac{E_p}{p} \, \widetilde{\epsilon}_{(h, p)} \, (p\Sigma_u + \Sigma_w)^{-1}].$$

Hence,

$$(4.26) \qquad Q_p = \text{tr}[(C_p - \overline{C}_p)\Sigma_w^{-1}] + \text{tr}[\overline{C}_p(p\Sigma_u + \Sigma_w)^{-1}]$$

$$(p = 1, \ldots, P),$$

where C_p and $\overline{C}_p$ are $G \times G$ matrices defined as

$$(4.27) \qquad C_p = \sum_{h=1}^{H(p)} \widetilde{\epsilon}'_{(h, p)} \, \widetilde{\epsilon}_{(h, p)},$$

$$(4.28) \qquad \overline{C}_p = \frac{1}{p} \sum_{h=1}^{H(p)} \widetilde{\epsilon}'_{(h, p)} \, E_p \, \widetilde{\epsilon}_{(h, p)} \qquad (p = 1, \ldots, P).$$

From eq. (4.26) above and Balestra (1975, eq. (5.2.19)) we derive

$$(4.29) \qquad \frac{\partial Q_p}{\partial \Sigma_w} = - \Sigma_w^{-1} (C_p - \overline{C}_p) \Sigma_w^{-1} - (p\Sigma_u + \Sigma_w)^{-1} \overline{C}_p (p\Sigma_u + \Sigma_w)^{-1},$$

$$(4.30) \qquad \frac{\partial Q_p}{\partial \Sigma_u} = - p(p\Sigma_u + \Sigma_w)^{-1} \overline{C}_p (p\Sigma_u + \Sigma_w)^{-1} \qquad (p = 1, \ldots, P).$$

The first order conditions for minimizing g with respect to the covariance matrices now follow by inserting (4.24), (4.25), (4.29), and (4.30) in

$$(4.31) \qquad \frac{\partial g}{\partial \Sigma_w} = \sum_{p=1}^{P} \frac{\partial g_p}{\partial \Sigma_w} = \sum_{p=1}^{P} \left\{ H(p) \frac{\partial \log |\Omega_p|}{\partial \Sigma_w} + \frac{\partial Q_p}{\partial \Sigma_w} \right\} = 0,$$

$$(4.32) \qquad \frac{\partial g}{\partial \Sigma_u} = \sum_{p=1}^{P} \frac{\partial g_p}{\partial \Sigma_u} = \sum_{p=1}^{P} \left\{ H(p) \frac{\partial \log |\Omega_p|}{\partial \Sigma_u} + \frac{\partial Q_p}{\partial \Sigma_u} \right\} = 0.$$

---

[17] See Balestra (1975, p. 20) or Magnus (1978, eq. (4)).

This gives a non-linear system of $G(G+1)$ equations in the unknown variances/covariances σ_{ij}^{w} and σ_{ij}^{u}, and defines our first-order conditions for solving subproblem (ii).

The numerical solution of (4.31) - (4.32) may be intricate, even for low values of P and/or G, and an iterative algorithm may be the only feasible approach. To obtain sensible *initial values* of Σ_w and Σ_u from which to start the iteration process, the following approach may be useful:

Assume, hypothetically, that our data consisted of observations from individuals observed p times only. The equation system for solving subproblem (ii) would then be

(4.33)
$$\frac{\partial g_p}{\partial \Sigma_w} = H(p) \left[(p - 1) \Sigma_w^{-1} + (p\Sigma_u + \Sigma_w)^{-1} \right]$$
$$- \Sigma_w^{-1} (C_p - \overline{C}_p) \Sigma_w^{-1} - (p\Sigma_u + \Sigma_w)^{-1} \overline{C}_p (p\Sigma_u + \Sigma_w)^{-1} = 0,$$

(4.34)
$$\frac{\partial g_p}{\partial \Sigma_u} = H(p) \, p \, (p\Sigma_u + \Sigma_w)^{-1} - p(p\Sigma_u + \Sigma_w)^{-1} \overline{C}_p (p\Sigma_u + \Sigma_w)^{-1} = 0.$$

This system can be easily solved. Let its solution be $\Sigma_w(p)$ and $\Sigma_u(p)$. From (4.34) it follows directly that

$$p\Sigma_u(p) + \Sigma_w(p) = \frac{\overline{C}_p}{H(p)}.$$

Inserting this into (4.33), we get

$$H(p) \, (p - 1) \, \Sigma_w(p)^{-1} = \Sigma_w(p)^{-1} (C_p - \overline{C}_p) \Sigma_w(p)^{-1}.$$

Hence,

(4.35)
$$\Sigma_w(p) = \frac{C_p - \overline{C}_p}{(p - 1) \, H(p)},$$

(4.36)
$$\Sigma_u(p) = \frac{p \, \overline{C}_p - C_p}{p \, (p - 1) \, H(p)}.$$

These are the estimates of Σ_w and Σ_u that we would obtain by utilizing data from the p'th group of individuals only. The corresponding estimator of the 'total' covariance matrix $\Sigma = \Sigma_u + \Sigma_w$ is

(4.37)
$$\Sigma(p) = \Sigma_u(p) + \Sigma_w(p) = \frac{C_p}{pH(p)}.$$

Usually, (4.35) and (4.36) will give different 'estimates' for different values of p. The estimates which solve subproblem (ii) may be considered as 'compromise values' of $\Sigma_w(p)$ and $\Sigma_u(p)$ for p = 1, ..., P. These considerations suggest that an appropriately weighted average of the P group specific estimates will be sensible initial values for an iterative solution of the complete equation system (4.31) - (4.32). We may, for instance, use the number of observations on which each estimate is based as weights. This would give the following initial values for Σ_u and Σ_w for starting the iteration process in subproblem (ii):

$$\Sigma_u^0 = \frac{\sum\limits_{p=1}^{P} \left\{ pH(p) \, \Sigma_u(p) \right\}}{\sum\limits_{p=1}^{P} pH(p)} = \frac{1}{TN} \sum\limits_{p=1}^{P} \left\{ \frac{p\overline{C}_p - C_p}{p - 1} \right\},$$

$$\Sigma_w^0 = \frac{1}{TN} \sum\limits_{p=1}^{P} \frac{p}{p - 1} (\overline{C}_p - C_p).$$

These values can be easily calculated once we have calculated the matrices C_p and $\overline{C}_p$ from the distur-
bances obtained by solving subproblem (i), cf. eqs. (3.21), (4.27) and (4.28).

A closer look at the single equation case

To get a closer understanding of the 'anatomy' of this iterative algorithm, let us see what it
implies in the case where the model contains one equation only. When $G = 1$, we have, with simplified
notation,

(4.38) $\Sigma_u = \sigma_u^2 = \rho\sigma^2$,

   $\Sigma_w = \sigma_w^2 = (1-\rho)\sigma^2$,

where $\sigma^2 = \sigma_u^2 + \sigma_w^2$ is the total disturbance variance and $\rho = \sigma_u^2/\sigma^2$ is the part of this total which
is due to individual variations. Alternatively, ρ can be interpreted as the coefficient of correlation
between two disturbances from the same individual.

Subproblem (i)

When $G = 1$, the S_p's, as defined in eq. (4.19), are scalars given by

$$S_0 = \frac{1}{\sigma_w} = \frac{1}{\sigma\sqrt{1-\rho}} \;,$$

$$S_p = \frac{1}{\sqrt{p\sigma_u^2 + \sigma_w^2}} = \frac{1}{\sigma\sqrt{1 + (p-1)\rho}} \qquad (p = 1, \ldots, P).$$

The expressions for the transformation matrices, (4.21), then become

(4.39) $F_p = \dfrac{1}{\sigma\sqrt{1-\rho}} \; [I_p - (1 - \alpha_p)\dfrac{E_p}{p})$ $(p = 1, \ldots, P)$,

where

(4.40) $\alpha_p = \left\{ \dfrac{1-\rho}{1 + (p-1)\rho} \right\}^{\frac{1}{2}}$ $(p = 1, \ldots, P)$.

All the matrices F_p contain $S_0 = \sigma^{-1}(1-\rho)^{-\frac{1}{2}}$ as a common scalar factor, which is irrelevant
for solving subproblem (i), so we may equally well use $F_p/S_0 = I_p - (1 - \alpha_p)E_p/p$ as transformation
matrices. Premultiplying the original disturbance vectors $\epsilon_{(h, p)}$ by these matrices to form the sum of
squares to be minimized then implies the following:[18]

> *Multiply the observations from individuals observed only once by $1 - (1 - \alpha_1) \doteq \alpha_1 \doteq (1 - \rho)^{\frac{1}{2}}$,
> subtract from the observations of all individuals observed p times ($p = 2, 3, \ldots, P$) a fraction
> $1 - \alpha_p \doteq 1 - [(1 - \rho)/(1 + (p - 1)\rho)]^{\frac{1}{2}}$ of the corresponding individual average,[19] and
> minimize the resulting sum of squares of disturbances.*

We note that $1-\alpha_p$, the fraction of the individual average to be subtracted, is an increasing function
of ρ, the individual share in the total variance. It is also an increasing function of p, the number
of replications. Or stated otherwise: The stronger the disturbances from the same individual are
correlated and the larger the number of times each individual is observed, the larger fraction of the
individual average should be subtracted to give the transformed disturbances. Table 1 reports the
value of $1 - \alpha_p$ for selected values of these two parameters.

---

18) The same algorithm is stated, without proof, in Biørn (1981, p. 229).
19) Recall that premultiplying a $p \times 1$ vector by E_p/p implies to replace all its elements by
their average value.

Table 1. The transformation parameter $1 - \alpha_p = 1 - \left\{ \dfrac{1 - \rho}{1 + (p - 1)\rho} \right\}^{\frac{1}{2}}$

p	ρ							
	0	0.01	0.1	0.2	0.5	0.8	0.9	0.999
1	0	0.0050	0.0513	0.1056	0.2929	0.5528	0.6838	0.9684
2	0	0.0100	0.0955	0.1835	0.4226	0.6667	0.7706	0.9776
3	0	0.0148	0.1340	0.2441	0.5000	0.7226	0.8110	0.9817
4	0	0.0196	0.1679	0.2929	0.5528	0.7575	0.8356	0.9842
5	0	0.0243	0.1982	0.3333	0.5918	0.7818	0.8526	0.9859
10	0	0.0470	0.3118	0.4655	0.6985	0.8438	0.8952	0.9900
20	0	0.0879	0.4429	0.5918	0.7818	0.8889	0.9257	0.9929
100	0	0.2947	0.7127	0.8039	0.9005	0.9501	0.9667	0.9968
∞	0	1	1	1	1	1	1	1

Subproblem (ii)

When $G = 1$, eqs. (4.24), (4.25), (4.29), and (4.30) can be simplified to

$$\frac{\partial \log |\Omega_p|}{\partial \sigma_w^2} = \frac{p-1}{\sigma_w^2} + \frac{1}{p\sigma_u^2 + \sigma_w^2},$$

$$\frac{\partial \log |\Omega_{\bar{p}}|}{\partial \sigma_u^2} = \frac{p}{p\sigma_u^2 + \sigma_w^2},$$

$$\frac{\partial Q_p}{\partial \sigma_w^2} = -\frac{1}{\sigma_w^4} (C_p - \bar{C}_p) - \frac{1}{(p\sigma_u^2 + \sigma_w^2)^2} \bar{C}_p,$$

$$\frac{\partial Q_p}{\partial \sigma_u^2} = -\frac{p}{(p\sigma_u^2 + \sigma_w^2)^2} \bar{C}_p.$$

Inserting these expressions into the first-order conditions for subproblem (ii), (4.31)-(4.32), while substituting (4.38), we obtain after some rearranging

$$(4.41) \quad \sum_{p=1}^{P} \{ \sigma^2 (1-\rho)^{-1} H(p)(p-1) - (1-\rho)^{-2}(C_p - \bar{C}_p) \}$$

$$+ \sum_{p=1}^{P} \{ \sigma^2 [1+(p-1)\rho]^{-1} H(p) - [1+(p-1)\rho]^{-2} \bar{C}_p \} = 0,$$

$$(4.42) \quad \sum_{p=1}^{P} \{ \sigma^2 [1+(p-1)\rho]^{-1} H(p)p - [1+(p-1)\rho]^{-2} p \bar{C}_p \} = 0.$$

A procedure for solving these two non-linear equations, and hence subproblem (ii), may for instance be: first, express σ^2 as a function of ρ by utilizing (4.42):

$$(4.43) \quad \sigma^2 = \frac{\sum\limits_{p=1}^{P} [1+(p-1)\rho]^{-2} p \, \overline{C}_p}{\sum\limits_{p=1}^{P} [1+(p-1)\rho]^{-1} pH(p)} \, ,$$

second, insert this expression in (4.41), and solve the resulting equation with respect to ρ by means of a grid search procedure, and third, insert the solution in (4.43). Given the complexity of our problem, this is a remarkably simple estimation procedure.

5. CONCLUSION

Conbined cross-section/time-series data constitute a very wide class of data structures, and both from a theoretical and, in particular, from a practical point of view the standard model which assumes complete time series to exist for all the observation units is a very restrictive one. In this paper, we have tried to widen the scope a bit by adopting a specification which allows for incomplete time series and which is applicable both in a multi-equation and a single equation context. Of course, the estimation problem becomes more cumbersome in this case than for the simpler models usually discussed in the literature. When the panel of individuals rotates, it is, in particular, the possible presence of time specific disturbance components which contributes to the complexity. This is the price we have to pay to make the model applicable to a wide class of practically important data structures.

If time specific components are omitted - which is realistic in many situations - the estimation problem may be simplified substantially, and an iterative algorithm for FIML estimation can be fairly easily implemented on a computer. This algorithm is particularly transparent in the single equation case - indeed it is surprisingly simple - but it is applicable also to linear multi-equation models, although it may involve a lot of programming work and a considerable amount of computer time. Some practical experiences should be gained before we can decide how well it will work.

The model specification studied in this paper includes several models discussed in the literature as special cases. To put our approach in perspective, let us give a few examples:

Three components specification: Σ_u, Σ_v, Σ_w unrestricted

    G > 1, m = 0: Avery (1977) (Only GLS), Baltagi (1980).

    G = 1, m = 0: Wallace-Hussain (1969), Nerlove (1971b), Mazodier (1971) (Only GLS), Amemiya (1971).

    G = 1, 0 < m < N: Wansbeek-Kapteyn (1981).

Two components specification: $\Sigma_v = 0$, Σ_u and Σ_w unrestricted

    G > 1, m = 0[20]: Chamberlain-Griliches (1975, appendix), Balestra (1975, Ch. 6).
    G = 1, m = 0[20]: Balestra-Nerlove (1966).
    G = 1, m = N/2[21]: Biørn (1981).

---

20) Or equivalenty: $G \geq 1$, $H(T) = N$, $H(p) = 0$ for $p \neq T$.
21) Or equivalenty: $G = 1$, $H(1) = N$, $H(2) = (T-1)N/2$, $H(p) = 0$ for $p > 2$.

One component specification: $\Sigma_v = \Sigma_u = 0$; Σ_w unrestricted

$G > 1$, $m \geq 0$: Conventional 'Seemingly Unrelated Regressions' model; Zellner (1962).

$G = 1$, $m \geq 0$: Standard linear regression model.

Our approach does not, however, include the models discussed in recent years in connection with the socalled "selectivity bias" (or "self-selection bias") problem.[22] We have implicitly disregarded the possibility that the panel may be subjected to a systematically changing degree of non-response by assuming that the sampling design is determined prior to and independent from the individual decisions represented by the structural equations of the model.

---

22) See e.g. Hausman and Wise (1977), Griliches, Hall, and Hausman (1978), and Maddala (1978).

REFERENCES

Amemiya, T. (1971): The Estimation of the Variances in a Variance - Components Model. *International Economic Review*, 12 (1971), pp. 1-13.

Avery, R.B. (1977): Error Components and Seemingly Unrelated Regressions. *Econometrica*, 45 (1977), pp. 199-209.

Balestra, P. (1975): *La Derivation Matricielle. Technique et Résultats pour Economistes*. (Paris: Sirey, 1975.)

Balestra P., and M. Nerlove (1966): Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. *Econometrica*, 34 (1966), pp. 585-612.

Baltagi, B.H. (1980): On Seemingly Unrelated Regressions with Error Components. *Econometrica*, 48 (1980), pp. 1547-1551.

Biørn, E. (1981): Estimating Economic Relations from Incomplete Cross-Section/Time-Series Data. *Journal of Econometrics*, 16 (1981), pp. 221-236.

Biørn, E., and E.S. Jansen (1982): *Econometrics of Incomplete Cross-Section/Time-Series Data: Consumer Demand in Norwegian Households 1975 - 1977*. Reports from the Central Bureau of Statistics of Norway, forthcoming. (Oslo: Central Bureau of Statistics, 1982.)

Chamberlain, G. and Z. Griliches (1975): Unobservables with a Variance-Components Structure: Ability, Schooling and the Economic Success of Brothers. *International Economic Review* 16 (1975), pp. 422-450.

Griliches, Z., B.H. Hall, and J.A. Hausman (1978): Missing Data and Self-Selection in Large Panels. *Annales de l'INSEE*, 30-31 (1978), pp. 137-176.

Hausman, J.A. and D.A. Wise (1977): Social Experimentation, Truncated Distributions, and Efficient Estimation. *Econometrica*, 45 (1977), pp. 919-938.

Maddala, G.S. (1971): The Use of Variance Components Models in Pooling Cross Section and Time Series Data. *Econometrica*, 39 (1971), pp. 341-358.

Maddala, G.S. (1978): Selectivity Problems in Longitudinal Data. *Annales de l'INSEE*, 30-31 (1978), pp. 423-450.

Magnus, J.R. (1978): Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix. *Journal of Econometrics*, 7 (1978), pp. 281-312.

Mazodier, P. (1971): L'Estimation des Modèles à Erreurs Composées. *Annales de l'INSEE*, 7 (mai-août 1971), pp.43-74.

Mundlak, Y. (1978): On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46 (1978), pp. 69-85.

Nerlove, M. (1971a): Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections. *Econometrica*, 39 (1971), pp. 359-382.

Nerlove, M. (1971b): A Note on Error Components Models. *Econometrica*, 39 (1971), pp. 383-396.

Oberhofer, W., and J. Kmenta (1974): A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models. *Econometrica*, 42 (1974), pp. 579-590.

Theil, H. (1971): *Principles of Econometrics*. (New York: John Wiley & Sons, Inc., 1971.)

Wallace, T.D., and A. Hussain (1969): The Use of Error Components Models in Combining Cross Section with Time Series Data. *Econometrica*, 37 (1969), pp. 55-72.

Wansbeek, T., and A. Kapteyn (1981): Maximum Likelihood Estimation in Incomplete Models. Unpublished paper, Central Bureau voor de Statistiek, Voorburg, the Netherlands, August 1981.

Zellner, A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57 (1962), pp. 348-368.

Zellner, A. and D.S. Huang (1962): Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations. *International Economic Review*, 3 (1962), pp. 300-313.

31

Printed 1981

Nr. 81/1 Erling J. Fløttum: National Accounts of Norway System and Methods of Estimation
Sidetall 101 Pris kr 15,00 ISBN 82-537-1555-2

- 81/2 Referansearkiv for naturressurs- og forurensningsdata 2. utgave Sidetall 424
Pris kr 20,00 ISBN 82-537-1233-2

- 81/3 Nils Håvard Lund: Byggekostnadsindeks for boliger Sidetall 127 Pris kr 15,00
ISBN 82-537-1232-4

- 81/4 Anne Lise Ellingsæter: Intervjuernes erfaringer fra arbeidskraftundersøkelsene
Rapport fra 99 intervjuere Field Work Experiences with the Labour Force Sample
Survey Reports from 99 Interviewers Sidetall 40 Pris kr 10,00 ISBN 82-537-1234-0

- 81/5 Bjørn Kjensli: Strukturundersøkelse for bygg og anlegg Vann- og kloakkanlegg
Sidetall 62 Pris kr 15,00 ISBN 82-537-1235-9

- 81/6 Erling Siring og Ib Thomsen: Metoder for estimering av tall for fylker ved hjelp av
utvalgsundersøkelser Sidetall 42 Pris kr 10,00 ISBN 82-537-1509-9

- 81/7 Arne Ljones og Hans Viggo Sæbø: Temperaturkorrigering av energiforbruket
Sidetall 43 Pris kr 10,00 ISBN 82-537-1507-2

- 81/8 Morten Reymert: En analyse av faktorinnsatsen i Norges utenrikshandel med utvik-
lingsland og industriland Sidetall 55 Pris kr 15,00 ISBN 82-537-1506-4

- 81/9 Petter Longva: A System of Natural Resource Accounts Eit rekneskapssystem for
naturressursar Sidetal 26 Pris kr 10,00 ISBN 82-537-1540-4

- 81/10 Stein Erland Brun: Tilgangen på arbeidskraft i fylkene for årene 1971 - 1979
Sidetall 72 Pris kr 15,00 ISBN 82-537-1514-5

- 81/11 Eva Ivås og Kjell Roland: MODIS IV Detaljerte virkningstabeller for 1979
Sidetall 264 Pris kr 20,00 ISBN 82-537-1515-3

- 81/12 Helge Brunborg, Jan Mønnesland og Randi Selmer: Framskriving av folkemengden etter
ekteskapelig status Sidetall 75 Pris kr 11,00 ISBN 82-537-1541-2

- 81/13 Adne Cappelen: Importinnhold i sluttleveringer Sidetall 20 Pris kr 10,00
ISBN 82-537-1545-5

- 81/14 MODIS IV Dokumentasjonsnotat nr. 16 Endringer i utgave 78-1 og 79-1 Sidetall 100
Pris kr 15,00 ISBN 82-537-1549-8

- 81/15 Skatter og overføringer til private Historisk oversikt over satser mv. Arene
1969 - 1981 Sidetall 74 Pris kr 15,00 ISBN 82-537-1554-4

- 81/16 Helgeturer 1978/79 Sidetall 23 Pris kr 10,00 ISBN 82-537-1560-9

- 81/17 Roy Østensen: Eie og bruk av personbil Foreløpige tall for 1979 og 1. kvartal 1980
Sidetall 42 Pris kr 10,00 ISBN 82-537-1566-8

- 81/18 Svein Homstvedt, Øyvind Lone og Tore Nesheim: Jordbruksareal ifølge jordregister og
utvalgstellinger. Metodiske forskjeller belyst med materiale fra Trøgstad kommune
Sidetall 62 Pris kr 15,00 ISBN 82-537-1598-6

- 81/19 Arne Faye: Holdninger til norsk utviklingshjelp 1980 Sidetall 62 Pris kr 15,00
ISBN 82-537-1562-5

- 81/20 Knut Fredrik Strøm: Konkurser i industri og varehandel Utvikling, hyppighet og
omfang Sidetall 31 Pris kr 10,00 ISBN 82-537-1569-2

- 81/21 Frank Foyn: Miljøverninvesteringer i industrien. Problemer ved kartlegging av
data Sidetall 34 Pris kr 10,00 ISBN 82-537-1591-9

Issued in the Series Reports from the Central Bureau of Statistics (REP) - ISSN 0332-8422
(cont.)

Printed 1981

Nr. 81/22   Petter R. Koren: Etterspørsel etter energi i norsk industri Sidetall 27
            Pris kr 10,00 ISBN 82-537-1592-7

-   81/23   Harald Bergland og Adne Cappelen: Produktivitet og sysselsetting i industrien
            Sidetall 75 Pris kr 15,00 ISBN 82-537-1600-1

-   81/24   Levekårsundersøkelsen 1980 Dokumentasjon Del I Sidetall 67 Pris kr 15,00
            ISBN 82-537-1612-5

-   81/25   Tor Haldorsen: Norske ferieformer Sidetall 112 Pris kr 15,00 ISBN 82-537-1611-7

-   81/26   Aktuelle skattetall 1981   Current Tax Data   Sidetall 46 Pris kr 10,00
            ISBN 82-537-1610-9

-   81/27   Tiril Vogt: Planregnskap Ressursregnskap for fysisk planlegging Sidetall 70
            Pris kr 15,00 ISBN 82-537-1614-1

-   81/28   Figurer i publikasjoner Sidetall 115 Pris kr 15,00 ISBN 82-537-1628-1

-   81/29   Tor Skoglund: Utprøving av modellen REGION mot fylkesfordelte nasjonalregnskapsdata
            for perioden 1973 - 1976 Sidetall 42 Pris kr 10,00 ISBN 82-537-1626-5

-   81/30   Adne Cappelen, Erik Garaas og Svein Longva: MODAG En modell for makroøkonomiske
            analyser Sidetall 70 Pris kr 15,00 ISBN 82-537-1630-3

-   81/31   Torstein Bye og Tor Eivind Høyland: Inntektsbegreper - Inntektsfordeling
            Sidetall 43 Pris kr 10,00 ISBN 82-537-1639-7

-   81/32   Totalregnskap for fiske- og fangstnæringen 1977 - 1980 Sidetall 39 Pris kr 10,00
            ISBN 82-537-1634-6

-   81/33   Erik Biørn: Estimating Seemingly Unrelated Regression Models from Incomplete
            Cross-Section/Time-Series Data Sidetall 32 Pris kr 10,00 ISBN 82-537-1635-4

-   81/34   Eva Ivås og Kjell Roland: MODIS IV Detaljerte virkningstabeller for 1980
            Sidetall 272 Pris kr 20,00 ISBN 82-537-1636-2

-   81/35   Sigurd Tveitereid og Jan Lædre: Markedsindikatorer for norsk eksport Sidetall 47
            Pris kr 15,00 ISBN 82-537-1645-1

Printed 1982

Nr. 82/1   Tor Haldorsen og Gunvor Iversen: Praktiske eksempler på måling av latente variable:
           Sammenhengen mellom subjektive og objektive indikatorer på arbeidsforhold
           ISBN 82-537-1649-4   ISSN 0332-8422

-   82/5   Naturressurser 1981 Sidetall 29 Pris kr 10,00 ISBN 82-537-1651-6

-   82/7   Øystein Engebretsen: Arealbruk i norske byer og tettsteder Sidetall 183
           Pris kr 15,00 ISBN 82-537-1720-2

-   82/9   Rolf Aaberge: Om måling av ulikskap Sidetall 73 Pris kr 15,00 ISBN 82-537-1721-0

Pris kr 10,00

Publikasjonen utgis i kommisjon hos H. Aschehoug & Co. og
Universitetsforlaget, Oslo, og er til salgs hos alle bokhandlere.

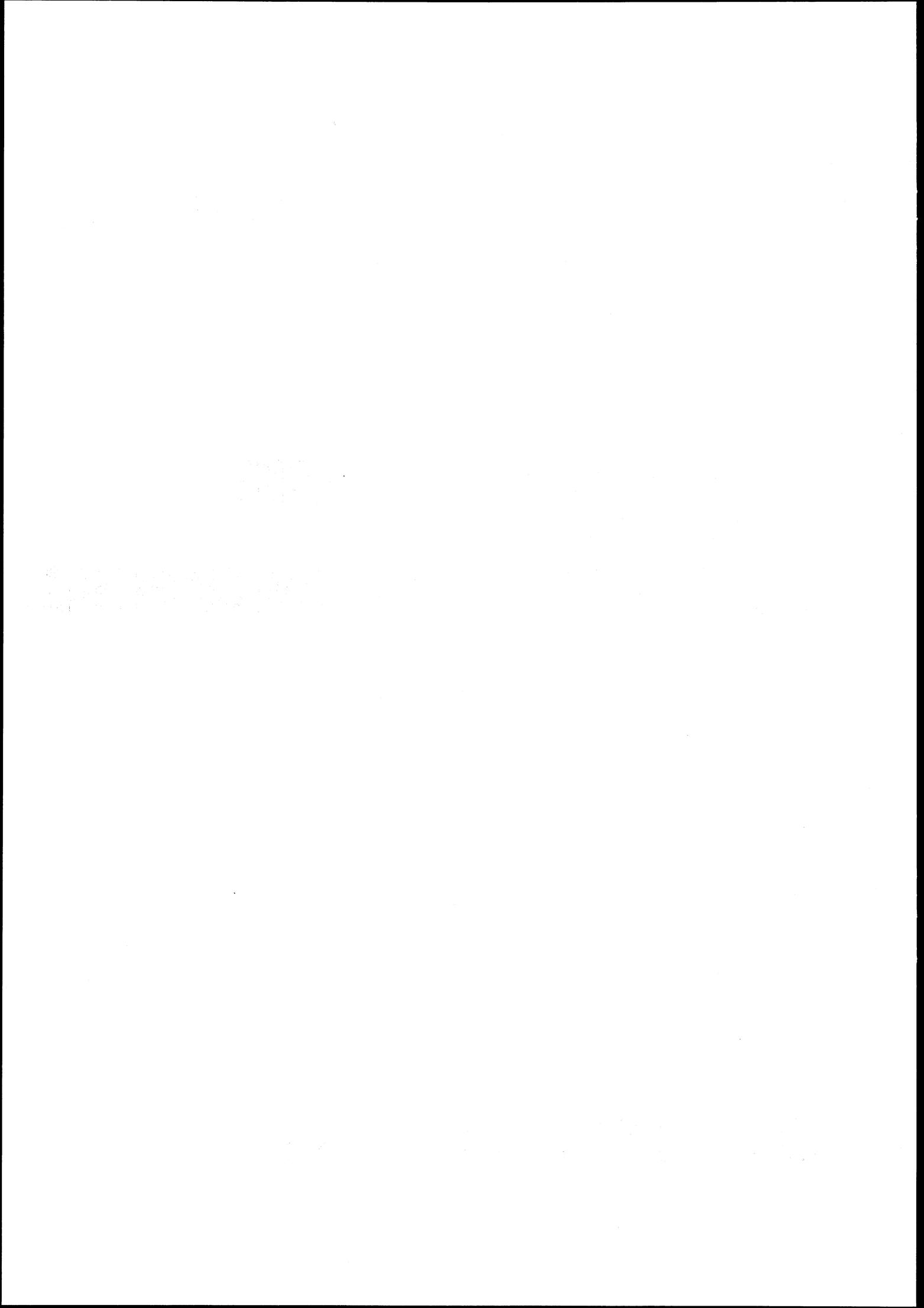# ESTIMATING SEEMINGLY UNRELATED REGRESSION MODELS FROM INCOMPLETE CROSS-SECTION/TIME-SERIES DATA

BY
ERIK BIØRN

STATISTISK SENTRALBYRÅ
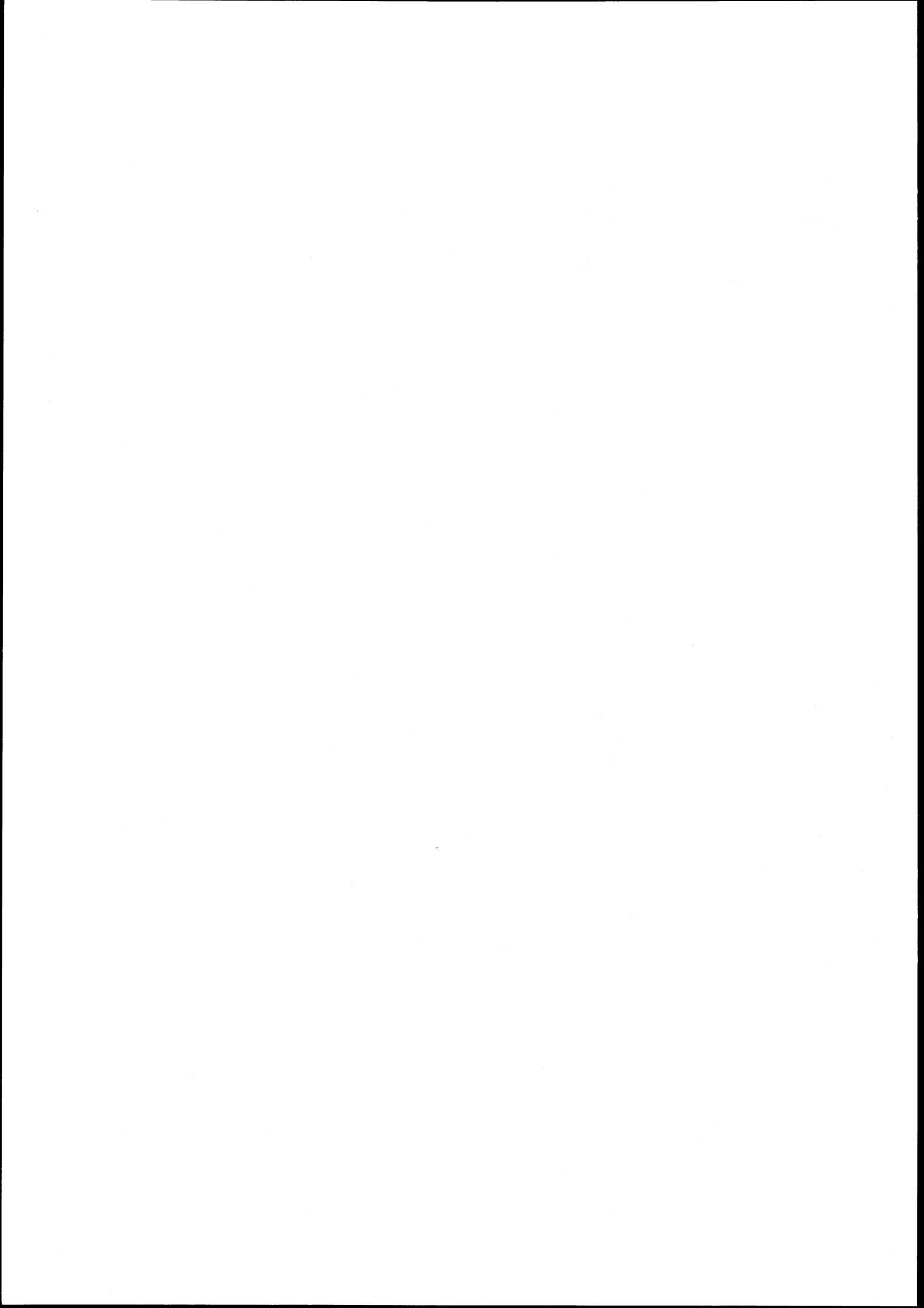CENTRAL BUREAU OF STATISTICS OF NORWAY

ESTIMATING SEEMINGLY UNRELATED REGRESSION MODELS FROM INCOMPLETE CROSS-SECTION/TIME-SERIES DATA

BY
ERIK BIØRN

PREFACE

Most of the theoretical contributions to models for handling combined cross-section/time-series data are based on the assumption that complete time-series of equal length exist for all observation units. This paper is concerned with the estimation of multi-equation models in situations where the observation units "rotate" over time. The data then become incomplete cross-section/time-series data. The situation with complete cross-section/time-series data emerges as a special case of this specification.

Econometric work with the Norwegian Surveys of Consumer Expenditure, which are based on rotating panels, has been a main motivation for exploring these problems. The conclusions, however, have for wider applicability.

A substantial part of the work reported in this paper was carried out during the author's visit to Institut National de la Statistique et des Études Économiques (INSEE), Paris, in the autumn 1980.

Central Bureau of Statistics, Oslo, 26 November 1981

Arne Øien

FORORD

Mesteparten av de teoretiske bidrag som har vært gitt til modeller for behandling av kombinerte tverrsnitts-tidsseriedata, bygger på den forutsetning at komplette og like lange tidsserier foreligger for alle observasjonsenhetene. Denne rapporten behandler estimering av flerligningsmodeller i tilfelle hvor utvalget av observasjonsenheter "roterer" fra periode til periode. Dette gir et datamateriale i form av ufullstendige tverrsnitts-tidsseriedata. Situasjonen med komplette tverrsnitts-tidsseriedata framtrer som et spesialtilfelle av denne spesifikasjonen.

Økonometrisk arbeid med de norske forbruksundersøkelsene, som bygger på roterende utvalg, har gitt støtet til å utforske disse problemene noe nærmere. Konklusjonene har imidlertid mer generell gyldighet.

En vesentlig del av arbeidet som ligger til grunn for denne rapporten, ble utført under forfatterens studieopphold ved Institut National de la Statistique et des Études Économiques (INSEE), Paris, høsten 1980.

Statistisk Sentralbyrå, Oslo, 26. november 1981

Arne Øien

CONTENTS

ABSTRACT

        The paper is concerned with the specification and estimation of multi-equation models in situations where the observation units "rotate" over time. The data then become incomplete cross-section/time-series data. The situation with complete cross-section/time-series data and several other practically interesting situations emerge as special cases of this specification. A procedure for iterative Full Information Maximum Likelihood estimation is outlined. Most attention is devoted to a disturbance components specification with individual components included and time specific components omitted.

1. INTRODUCTION

A conspicous feature of applied econometrics in recent years is the increasing trend in the utilization of combined cross-section/time-series (CS/TS) data. This reflects primarily the increasing availability of individual time series for empirical work, but also the growing theoretical insight in models and methods for handling such data. These include models with unobserved individual effects in general, and error components regression models in particular.

The well-known study of Balestra and Nerlove (1966), one of the first applications of error components models in econometrics, was concerned with the situation where N observation units are observed in T consecutive time periods. The stochastic disturbance in the regression equation considered was decomposed in two additive and independent parts, the first is specific to the observation unit - *the individual component* - while the second represents the combined effect of the observation unit and the time period - *the combined component*. An extended three components model which treats the time 'dimension' and the individual 'dimension' symmetrically - by including a *time (period) specific component* along with the individual and the combined ones - has been extensively discussed in several subsequent articles, e.g. Wallace and Hussain (1969), Maddala (1971), Mazodier (1971), Nerlove (1971b), and Mundlak (1978). In recent years, this three components, single equation regression model has been further extended to multi-equation models formulated as a system of seemingly unrelated regressions, in articles by Avery (1977) and Baltagi (1980).

A common feature of all the articles mentioned above (and a lot of others) is the assumption that the observation matrix is rectangular: the same N units are under observation in the same T successive time periods. The data set constitutes *complete cross-section/time-series data*. This specification is in fact so firmly established in the literature that the term *combined* CS/TS data has almost become synonymous with *complete* CS/TS data. Analytical and notational convenience may partly account for this. Another reason may be the fact that in many practical applications of CS/TS data, the observation units are some sorts of aggregates; in the Balestra-Nerlove (1966) study, for instance, the units were the different states in the USA. In such cases, complete time-series are readily available.

However, in the context of *sampling surveys data* from a population of genuine micro units, e.g. households or firms, this may be a rather extreme and unrealistic specification. Occasionally, the econometrician has access to socalled panel data (longitudional data), but owing to the serious problems of *non-response* when repetitive and time-consuming reporting is involved, this will in practice often be the exception rather than the rule. For this reason, models based on *incomplete CS/TS data* are well worth considering. The econometrician's philosophy should always be to construct his model according to the data situation actually at hand, rather than forcing his data into the strait-jacket of preconceived simplistic model schemes. If he follows the latter strategy- e.g. by throwing away all individuals from which complete time series do not exist - he will usually waste a lot of information.

Some aspects of the problem of specification and estimation of *single equation* error components regression models from incomplete CS/TS data are discussed in Biørn (1981). The purpose of the present paper is to generalize this approach to a *multi-equation* model, specifically, a model written in the *'seemingly unrelated regressions'* format. This is, formally, a multi-equation model in which all equations contain only one endogenous variable and with only predetermined variables occurring on the right hand side of each equation.

Rotating panels are applied in practice by several statistical agencies. The Central Bureau of Statistics of Norway, for instance, has followed this sort of sampling design for its household budget surveys since the year 1975. (About 25 per cent of the respondents in one year are asked to report again in the following year.) A main purpose is to reduce the degree of non-response. An econometric study based on these data is documented in Biørn and Jansen (1982). This work has, however, motivated a closer examination of the problems raised by incomplete CS/TS data in the context of multi-equation models with error components. It is the main results of this more general approach which are reported in the present paper. The conclusions are not confined to the field of micro consumer analysis, but

have, of course, far wider applicability.

The paper is organized as follows: Section 2 contains a brief general description of the model and the sampling design. Section 3 deals with the structure of the disturbance covariance matrix. We distinguish between the case where three disturbance components are included in all the equations and the case where time specific components are omitted. The full three components model is shown to contain all the models referred to above (and a lot of others) as special cases. Estimation procedures are discussed in section 4. We first state the Full Information Maximum Likelihood (FIML) estimation strategy in general terms for the full three components version of the model. Then we discuss more closely an iterative FIML estimation procedure for the simpler two components specification. Finally, in the concluding section 5, we attempt to place the work reported here in a wider context.

2. MODEL AND SAMPLING DESIGN

Consider a linear model with G equations, in the form of G 'seemingly unrelated regressions'. Formally, this is a linear simultaneous equations model in which each of the G endogenous variables is included in one and only one of the G equations, and with only predetermined variables on the right-hand side of each equation.[1] An example of this sort of model is a complete system of static consumer demand functions, where the volume of consumption of each commodity group is expressed as a function of income and prices (and possibly other exogenous variables). The observations on the variables are assumed to be combined CS/TS data. The i'th equation of the model thus has the form

$$(2.1) \qquad y_{iht} = x_{iht}\beta_i + e_{iht} \qquad (i=1, \ldots G),$$

Where β_i is the (column) vector of coefficients in the i'th equation, y_{iht} is the value of the i'th endogenous variable, x_{iht} is the (row) vector of exogenous (non-stochastic)[2] variables in the i'th equation, and e_{iht} is the corresponding disturbance. The subscripts h and t indicate the individual (micro unit) and time period, respectively. At this stage, we assume that h and t can be any positive integer, i.e., the model applies to all individuals in the population at all points of time.

We decompose e_{iht} into three additive components, an individual component u_{ih}, a time specific component v_{it} and a combined component (a remainder) w_{iht}:

$$(2.2) \qquad e_{iht} = u_{ih} + v_{it} + w_{iht} \qquad \text{for all h and t, and } i=1, \ldots, G.$$

All components have zero expectations:

$$(2.3) \qquad E(u_{ih}) = E(v_{it}) = E(w_{iht}) \qquad \text{for all h and t, and } i=1, \ldots, G,$$

and their second order moments satisfy

$$(2.4) \qquad E(u_{ih}\, u_{jk}) = \delta_{hk}\, \sigma_{ij}^{u},$$

$$(2.5) \qquad E(v_{it}\, v_{js}) = \delta_{ts}\, \sigma_{ij}^{v},$$

$$(2.6) \qquad E(w_{iht} w_{jks}) = \delta_{hk}\delta_{ts}\sigma_{ij}^{w},$$

---

1) Cf. Zellner (1962) and Zellner and Huang (1962), for a general discussion of seemingly unrelated regression models.
2) We assume that no lagged endogenous variables are included. For a discussion of problems caused by the presence of such variables in single equation error components models, see Nerlove (1971 a).

8

(2.7) $\quad E(u_{ih} v_{js}) = E(u_{ih}w_{jks}) = E(v_{it} w_{jks}) = 0$, for all h, k, t, s, and for i, j=1, ..., G.

where δ_{hk} = 1 for k = h, and 0 for k ≠ h; δ_{ts} = 1 for s = t, and 0 for s ≠ t. These assumptions have the following implications:

(i) Homoscedasticity of all disturbance components:

$$var (u_{ih}) = \sigma_{ii}^{u} \qquad \text{for all i and h,}$$

$$var (v_{it}) = \sigma_{ii}^{v} \qquad \text{for all i and t,}$$

$$var (w_{iht}) = \sigma_{ii}^{w} \qquad \text{for all i, h, and t.}$$

(ii) Constant covariances between disturbance components relating to different equations, but to the same individual or time period:

$$cov (u_{ih}, u_{jh}) = \sigma_{ij}^{u} \qquad \text{for all i, j, and h,}$$

$$cov (v_{it}, v_{jt}) = \sigma_{ij}^{v} \qquad \text{for all i, j, and t,}$$

$$cov (w_{iht}, w_{jht}) = \sigma_{ij}^{w} \qquad \text{for all i, j, h, and t.}$$

(iii) No correlation between disturbance components relating to different individuals and/or periods.

These assumptions are straightforward generalizations of those in the single equation model considered in Biørn (1981).

Assumptions (2.2) - (2.7) imply

(2.8) $\quad E(e_{iht}) = 0$

(2.9) $\quad E(e_{iht} e_{jks}) = \begin{cases} \sigma_{ij}^{u} + \sigma_{ij}^{v} + \sigma_{ij}^{w} & \text{for k=h, s=t,} \\ \sigma_{ij}^{u} & \text{for k=h, s≠t,} \\ \sigma_{ij}^{v} & \text{for k≠h, s=t,} \\ 0 & \text{for k≠h, s≠t.} \end{cases}$

The individuals are selected according to the following *sampling plan*: Let all individuals in the population be numbered consecutively, and let the sample in period 1 consist of individuals 1, 2, ..., N. In period 2, individuals 1, 2, ..., m ($0 \leq m \leq N$) are replaced by individuals N+1, N+2, ..., N+m. This procedure - dropping the first m individuals from the sample selected in the previous period and augmenting it by drawing m individuals from the population so that the sample size remains the same - continues in all the following periods until period T. In general, the sample in period t thus consists of the individuals with numbers (t-1)m+1, (t-1)m+2, ..., (t-1) m+N (t=1, 2, ..., T). The total number of individuals observed is

(2.10) $\quad H = (T-1)m+N,$

and the total number of observations is TN.

The special case with m=0, i.e. H=N, is the situation with *completely overlapping samples* studied by Avery (1977) and Baltagi (1980). At the other extreme, m = N, i.e. H = TN, corresponds to a sampling design with *non-overlapping samples*; all individuals are observed only once. If

0<m<N, we have a situation with *rotating (partly overlapping) samples*. In this intermediate situation, the data set will include some individuals observed only once, another subset observed twice, and - provided that 0<m<N/2 - still another subset observed three times, etc. If the 'rotation parameter' m exceeds half the size of each sample, N/2≤m<N, no individual is observed more than twice.

All sampling designs with 0<m<N thus have the common characteristic that the data set shows variation both along the 'time dimension' and the 'individual dimension', without giving complete time series of cross section data. To fix ideas, we have illustrated five typical cases in Figure 1 below for N = 1 000 and T = 5. We let the horizontal axis indicate the number of the individual and let the vertical axis represent the time period. Each asterisk represents N/4 = 250 observations. Case A illustrates the standard situation with completely overlapping samples, case E represents the other extreme situation with non-overlapping samples, whereas cases B, C, and D illustrate different designs with rotating samples.
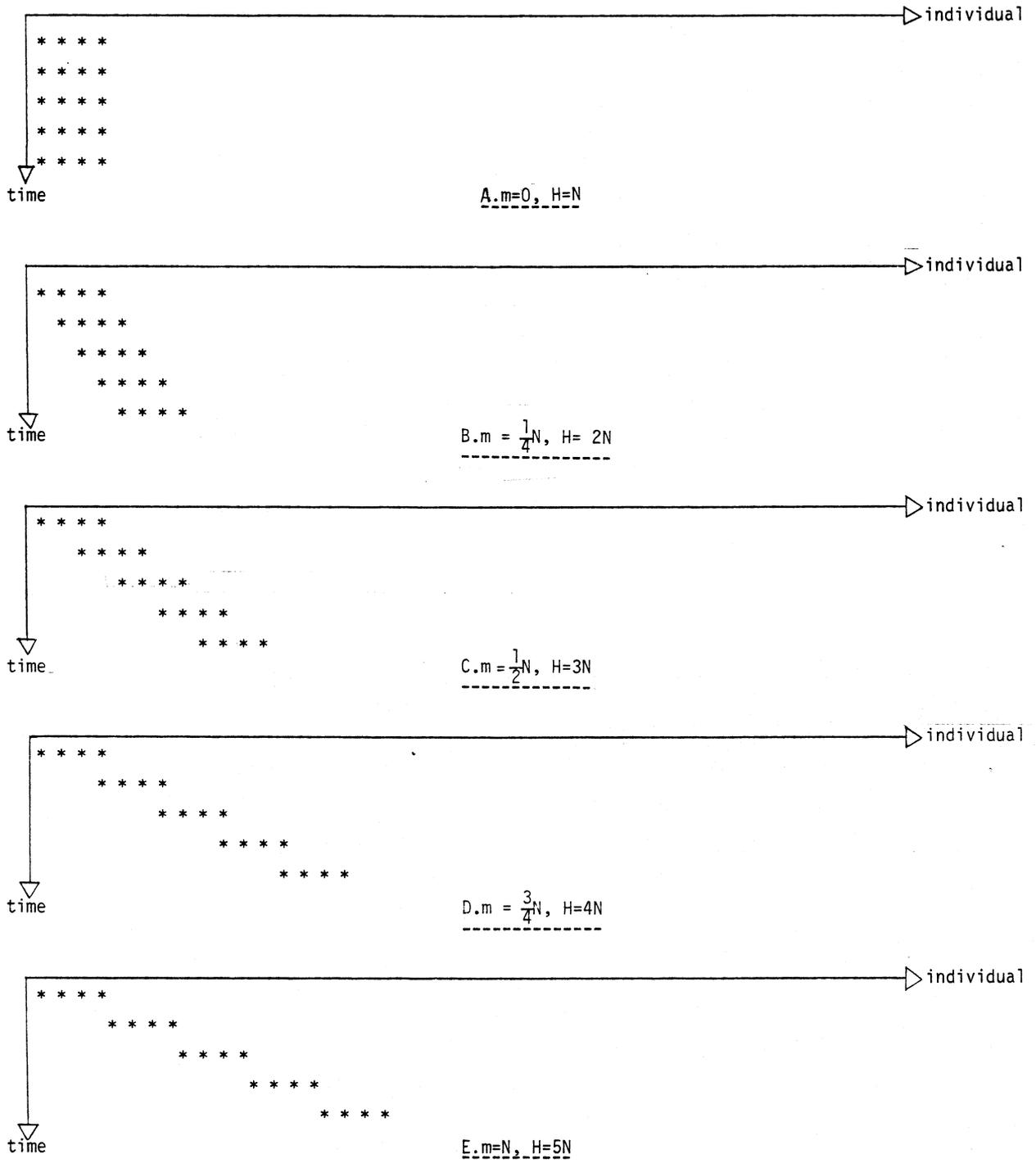
Figure 1. Five typical sampling designs. N=1 000, T=5. Each asterisk (*) represents 250 observations.

3. THE DISTURBANCE COVARIANCE MATRIX

Before discussing estimation procedures, we need an analytical expression for the disturbance covariance matrix of the TN observations in our data set. This immediately raises the question how to arrange the GTN disturbances - TN disturbances from each of the G equations in the model - by equation number (i), individual (h), and time period (t). For obvious reasons, this is a more critical problem when working with incomplete cross-section/time-series data than it is in the more 'tidy' situation where all time series are complete. No general principle can be recommended for all situations. Which ordering of the disturbance vector will give the most tractable expression for the covariance matrix depends essentially on whether time specific disturbance components are included or not. If such components are present, the succession in time of the different observations (i.e. the subscript t) is, of course, crucial for the characterization of the covariance structure. If they are omitted, we are free to arrange the observations in a different order.

We shall therefore divide the following discussion into two parts. First, in section 3.1, we establish general formulae for the covariance matrix by employing a set of selection matrices to indicate which individuals in the population are included in the sample in the different periods. Then, in section 3.2, we examine more closely the case where time specific components are excluded. We show that, with a suitable reordering of the observations, the disturbance covariance matrix assumes a simple block-diagonal form, which will be convenient for the ensuing discussion of estimation problems.

3.1. The general case. Formulation by means of selection matrices

Let us start by considering the disturbance covariance matrix in the *hypothetic* case where individuals 1, 2, ..., $H[= (T - 1)m + N]$ are observed in *all* the T periods under consideration, i.e. the matrix which corresponds to complete CS/TS data with a panel consisting of H individuals. Arrange all disturbances from the i'th equation of the model in the H x T matrix

$$(3.1) \qquad \tilde{e}_i = \begin{pmatrix} e_{i11} & \cdots\cdots\cdots\cdots & e_{i1T} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ e_{iH1} & \cdots\cdots\cdots\cdots & e_{iHT} \end{pmatrix} = (e_{i1} \cdots\cdots e_{iT}) \\ \qquad\qquad (i = 1, \ldots, G),$$

its t'th column, e_{it}, containing the disturbances in period t from all the H individuals. According to the general specification of the variances and covariances, (2.9), we have

$$(3.2) \qquad E(e_{it}\, e_{js}') = \begin{cases} (\sigma_{ij}^u + \sigma_{ij}^w)I_H + \sigma_{ij}^v E_H & \text{for } s = t \quad \begin{pmatrix} i, j = 1, \ldots, G \\ t, s = 1, \ldots, T \end{pmatrix}, \\[2ex] \sigma_{ij}^u\, I_H & \text{for } s \neq t \end{cases}$$

where I_H is the H x H identity matrix and E_H is the H x H matrix with all elements equal to one.

Defining the TH x 1 vector[3)]

$$(3.3) \qquad e_i = \text{vec}(\tilde{e}_i) = \begin{pmatrix} e_{i1} \\ \vdots \\ e_{iT} \end{pmatrix} \qquad\qquad (i = 1, \ldots, G),$$

___
3) The operator 'vec' denotes vectorization, i.e. the stacking of all columns of a matrix into one long vector, and $\otimes$ is the Kronecker product operator. A list of main properties of Kronecker products, some of which will be utilized in the following, is given in Balestra (1975, section 2.1).

eq. (3.2) can alternatively be written as

(3.4)
$$E(e_i e_j') = \sigma_{ij}{}^W I_T \otimes I_H + \sigma_{ij}{}^V I_T \otimes E_H + \sigma_{ij}{}^U E_T \otimes I_H.$$

The difference in notation apart, this equation is identical with eq. (26) in Avery (1977). (Recall that Avery, following the single equation models of Wallace and Hussain (1969) and Nerlove (1971b), has ordered the disturbances first by individual, second by period, whereas we have chosen the reverse ordering.)

The covariance matrix of the complete GTH x 1 disturbance vector that would result if all the H individuals were observed in all the T periods,

(3.5)
$$e = \begin{pmatrix} e_1 \\ \vdots \\ e_G \end{pmatrix} = \begin{pmatrix} \text{vec}(\tilde{e}_1) \\ \vdots \\ \text{vec}(\tilde{e}_G) \end{pmatrix},$$

can thus be written compactly as[4]

(3.6)
$$E(ee') = \Sigma_W \otimes I_T \otimes I_H + \Sigma_V \otimes I_T \otimes E_H + \Sigma_U \otimes E_T \otimes I_H,$$

where

$$\Sigma_u = \begin{pmatrix} \sigma_{11}{}^u \cdots\cdots\cdots \sigma_{1G}{}^u \\ \vdots \qquad\qquad \vdots \\ \sigma_{G1}{}^u \cdots\cdots\cdots \sigma_{GG}{}^u \end{pmatrix},$$

$$\Sigma_v = \begin{pmatrix} \sigma_{11}{}^v \cdots\cdots\cdots \sigma_{1G}{}^v \\ \vdots \qquad\qquad \vdots \\ \sigma_{G1}{}^v \cdots\cdots\cdots \sigma_{GG}{}^v \end{pmatrix},$$

$$\Sigma_w = \begin{pmatrix} \sigma_{11}{}^w \cdots\cdots\cdots \sigma_{1G}{}^w \\ \vdots \qquad\qquad \vdots \\ \sigma_{G1}{}^w \cdots\cdots\cdots \sigma_{GG}{}^w \end{pmatrix}.$$

Now, a subset, consisting of N of the H individuals in this hypothetical panel, are *actually* selected. Define, in analogy with (3.1), the N x T matrix

(3.7)
$$\tilde{\epsilon}_i = \begin{pmatrix} \epsilon_{i11} \cdots\cdots\cdots \epsilon_{i1T} \\ \vdots \qquad\qquad \vdots \\ \epsilon_{iN1} \cdots\cdots\cdots \epsilon_{iNT} \end{pmatrix} = (\epsilon_{i1} \cdots \epsilon_{iT}) \qquad (i = 1, \ldots, G),$$

where ϵ_{iht} is the disturbance in eq. no i *belonging to the h'th individual of those observed in the t'th period.* (Its population number is (t - 1)m + h.) The t'th column of this matrix, ϵ_{it}, contains the disturbances in period t of all the individuals actually observed in this period. Obviously,

---

4) Cf. Avery (1977, eq. (27)) and Baltagi (1980, eq. (7)).

$$\epsilon_{iht} \equiv e_{i,(t-1)m+h,t} \qquad \begin{pmatrix} i = 1, \ldots, G \\ h = 1, \ldots, N \\ t = 1, \ldots, T \end{pmatrix} \ .$$

This relationship can be stated compactly in vector notation as

$$(3.8) \qquad \epsilon_{it} = D_t e_{it} \qquad \begin{pmatrix} i = 1, \ldots, G \\ t = 1, \ldots, T \end{pmatrix}$$

where D_t is a *sample design (selection) matrix* of dimension $N \times H$ ($= N \times [(T-1)\,m + N\,]$) defined as follows:

$$(3.9) \qquad D_t = \left(O_{N,\ (t-1)m} \ \vdots \ I_N \ \vdots \ O_{N,\ (T-t)m} \right) \ ,$$

$O_{N,n}$ denoting the $N \times n$ zero matrix. *Element (h, k) of D_t is 1 if the k'th individual in the population is the h'th individual in the sample in period t, and zero otherwise.* In completely overlapping samples $(m = 0)$, we have in particular

$$(3.9a) \qquad D_t = I_N \qquad\qquad (t = 1, \ldots, T),$$

whereas non-overlapping samples $(m = N)$ correspond to

$$(3.9b) \qquad D_t = (O_{1,\ t-1} \quad 1 \quad O_{1,\ T-t}) \otimes I_N.$$

Since $D_t D_t' = I_N$ and $D_t E_H D_t' = E_N$ for all t, *irrespective of the sampling design,* it follows from (3.2) and (3.8) that

$$(3.10) \qquad E(\epsilon_{it}\,\epsilon_{js}') = D_t E(e_{it}\,e_{js}')\,D_s'$$

$$= \begin{cases} (\sigma_{ij}{}^u + \sigma_{ij}{}^w)\ I_N + \sigma_{ij}{}^v\ E_N & \text{for } s = t \\[2ex] \sigma_{ij}{}^u\ D_t D_s' & \text{for } s \neq t \end{cases}$$
$$(i, j, = 1, \ldots, G).$$

From this equation it is evident that $\sigma_{ij}{}^u$ and $\sigma_{ij}{}^w$ are not identifiable unless $D_t D_s' \neq O_{N,\ N}$ for at least one $s \neq t$, which implies that at least two of the T samples overlap. If $D_t D_s' = O_{N,\ N}$ for all t and $s \neq t$, i.e. all samples are non-overlapping, only $\sigma_{ij}{}^v$ and $\sigma_{ij}{}^u + \sigma_{ij}{}^w$ can be identified from the data.

Defining the $TN \times 1$ vector

$$(3.11) \qquad \epsilon_i = \text{vec}\,(\tilde{\epsilon}_i) = \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{iT} \end{pmatrix} \qquad\qquad (i = 1, \ldots, G),$$

and the stacked $TN \times H$ selection matrix

$$(3.12) \qquad D = \begin{pmatrix} D_1 \\ \vdots \\ D_T \end{pmatrix} \ ,$$

eq. (3.10) can alternatively be written in the form

(3.13) $E(\epsilon_i \epsilon_j') = \sigma_{ij}^{W} I_T \otimes I_N + \sigma_{ij}^{V} I_T \otimes E_N + \sigma_{ij}^{u} DD'$

                                                        (i, j = 1, ..., G).

The covariance matrix of the stacked GTN x 1 vector

(3.14) $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_G \end{pmatrix} = \begin{pmatrix} \text{vec } (\tilde{\epsilon}_1) \\ \vdots \\ \text{vec } (\tilde{\epsilon}_G) \end{pmatrix}$,

can then be written compactly as

(3.15) $E(\epsilon\epsilon') = \Omega = \Sigma_w \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N + \Sigma_u \otimes DD'$.

This is the *general* expression for the disturbance covariance matrix in seemingly unrelated regressions models based on combined cross-section/time-series data.

In the special case with completely overlapping samples, it follows from (3.9a) and (3.12) that $D = 1_T \otimes I_N$, where 1_T denotes the T x 1 vector with all elements equal to one. Then $DD' = (1_T 1_T') \otimes I_N = E_T \otimes I_N$, and the covariance matrix becomes

(3.15a) $\Omega = \Omega_C = \Sigma_w \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N + \Sigma_u \otimes E_T \otimes I_N$,

the subscript C denoting 'complete'. Not surprisingly, we find that this expression is identical with (3.6) with H set equal to N. At the other extreme, non-overlapping samples are characterized by $D = I_T \otimes I_N$ (cf. (3.9b) and (3.12)), which implies the following covariance matrix:

(3.15b) $\Omega = \Omega_N = (\Sigma_w + \Sigma_u) \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N$,

the subscript N denoting 'non-overlapping'. This expression is equivalent to (3.6) with H set equal to N, Σ_w replaced by $\Sigma_w + \Sigma_u$, and with the last term omitted.

3.2. The two components case: Time specific components excluded

In the previous section, we gave a characterization of the covariance structure of the disturbances in the general case where both the individual (u_{ih}), the time specific (v_{it}), and the remainder disturbance component (w_{iht}) are included. This specification is theoretically attractive, not least because of its symmetry. However, for *practical* applications of disturbance components models in the context of genuine micro data, time specific effects are usually of far less importance than individual ones. This reflects, *inter alia*, the fact that the number of individuals by far exceeds the number of periods involved (e.g. T < 5 and N > 100). Then Σ_u and Σ_w dominate over Σ_v, and for practical purposes the latter may be ignored without essential loss.

In this section, we take a closer look at the covariance structure when $\Sigma_v = 0_{G,G}$. Of course, from (3.15) it then follows that

(3.16) $E(\epsilon\epsilon') = \Omega = \Sigma_w \otimes I_T \otimes I_N + \Sigma_u \otimes DD'$.

However, the ordering of the disturbances which leads to this expression for the covariance matrix does not exploit the fact that when time specific effects are ignored, the actual succession in time of the different observations is irrelevant. We can, without fear of ambiguity, drop the time subscript from

the model specification and rearrange the disturbance vector.[5] This is what we shall do in this section.

We arrange the individuals in groups according to the number of times each individual is under observation and identify each individual by its number in the group to which it belongs. Let, in general, $H(p)$ denote the number of individuals in the p'th group, i.e. the number of individuals observed p times. Let P denote the maximal number of replications. We arrange the $H(1)$ observations from the individuals observed only once in the first group, the $H(2)$ individuals observed twice and the corresponding $2H(2)$ observations in the second group, ..., the $H(P)$ individuals observed P times and the corresponding $PH(P)$ observations in the P'th group. Obviously, we have[6]

$$(3.17) \qquad \sum_{p=1}^{P} H(p) = H = \text{total number of individuals observed,}$$

and

$$(3.18) \qquad \sum_{p=1}^{P} p\, H(p) = TN = \text{total number of observations.}$$

To fix ideas, let us refer the values of P and $H(p)$ corresponding to the five sampling designs illustrated in Figure 1 above. We find

Case A: $P = 5$; $H(1) = H(2) = H(3) = H(4) = 0$, $H(5) = N = 1\ 000$.

Case B: $P = 4$; $H(1) = H(2) = H(3) = H(4) = \frac{N}{2} = 500$.

Case C: $P = 2$; $H(1) = N = 1\ 000$, $H(2) = 2N = 2\ 000$.

Case D: $P = 2$; $H(1) = 3N = 3\ 000$, $H(2) = N = 1\ 000$.

Case E: $P = 1$; $H(1) = 5N = 5\ 000$.

In the following, we use (h, p) as a shorthand notation for the h'th individual of those observed p times $(h = 1, ..., H(p); p = 1, ..., P)$ and define

$\epsilon_{i(h, p)q}$: disturbance in eq. no i relating to the q'th observation from individual (h, p)

$(i = 1, ..., G; q = 1, ..., p; h = 1, ..., H(p); p = 1, ..., P)$.

The p x 1 vector containing the disturbances in eq. no i relating to individual (h, p) is

$$(3.19) \qquad \epsilon_{i(h, p)} = \begin{pmatrix} \epsilon_{i(h, p)1} \\ \vdots \\ \epsilon_{i(h, p)p} \end{pmatrix} \qquad \begin{matrix} (i = 1, ..., G) \\ (h = 1, ..., H(p)) \\ (p = 1, ..., P). \end{matrix}$$

---

5) This may necessitate a redefinition of some of the structural variables in the model. When no lagged endogenous variables are included, (cf. footnote 2), this reformulation is trivial.

6) It is not essential for this reformulation of the model that the sample size is the same, and equal to N, in all the periods. The following formulae are equally valid in more general situations where the sample size changes from period to period; only the sequence, $H(1), ..., H(P)$ is of importance.

Since, when $\sigma_{ij}^{\ v} = 0$ we have

$$
E\,[\epsilon_{i(h,\ p)q}\ \epsilon_{j(k,\ n)r}] = \begin{cases} \sigma_{ij}^{\ w} + \sigma_{ij}^{\ u} & \text{for } (k,\ n) = (h,\ p) \text{ and } r = q \\[2mm] \sigma_{ij}^{\ u} & \text{for } (k,\ n) = (h,\ p) \text{ and } r \neq q \\[2mm] 0 & \text{otherwise} \end{cases}
$$

(cf. ((2.9)), it follows that

$$
(3.20) \qquad E[\epsilon_{i(h,\ p)}\ \epsilon'_{j\,(k,\ n)}] = \begin{cases} \sigma_{ij}^{\ w}\ I_p + \sigma_{ij}^{\ u}\ E_p & \text{for } (k,\ n) = (h,\ p) \\[2mm] 0_{p,\ n} & \text{for } (k,\ n) \neq (h,\ p) \end{cases}
$$

$$
(i = 1,\ \ldots,\ G;\ j = 1,\ \ldots,\ G)
$$
$$
(h = 1,\ \ldots,\ H(p);\ k = 1,\ \ldots,\ H(n))
$$
$$
(p = 1,\ \ldots,\ P;\ n = 1,\ \ldots,\ P).
$$

It is convenient to arrange all disturbances from individual $(h,\ p)$ in the $p \times G$ matrix

$$
(3.21) \qquad \tilde{\epsilon}_{(h,\ p)} = \begin{pmatrix} \epsilon_{1(h,\ p)1} & \cdots & \epsilon_{G(h,\ p)1} \\ \vdots & & \vdots \\ \epsilon_{1(h,\ p)p} & \cdots & \epsilon_{G(h,\ p)p} \end{pmatrix} = (\epsilon_{1(h,\ p)}\ \cdots\cdots\ \epsilon_{G(h,\ p)})
$$

$$
(h = 1,\ \ldots,\ H(p))
$$
$$
(p = 1,\ \ldots,\ P),
$$

or alternatively in the stacked $Gp \times 1$ vector

$$
(3.22) \qquad \epsilon_{(h,\ p)} = \text{vec}(\tilde{\epsilon}_{(h,\ p)}) = \begin{pmatrix} \epsilon_{1(h,\ p)} \\ \vdots \\ \epsilon_{G(h,\ p)} \end{pmatrix}
$$

$$
(h = 1,\ \ldots,\ H(p))
$$
$$
(p = 1,\ \ldots,\ P).
$$

Eq. (3.20) can then be written as

$$
(3.23) \qquad E\,[\epsilon_{(h,\ p)}\ \epsilon'_{(k,\ n)}] = \begin{cases} \Omega_p & \text{for } (k,\ n) = (h,\ p) \\[2mm] 0_{Gp,\ Gn} & \text{for } (k,\ n) \neq (h,\ p) \end{cases}
$$

$$
(h = 1,\ \ldots,\ H(p);\ k = 1,\ \ldots,\ H(n))
$$
$$
(p = 1,\ \ldots,\ P;\ n = 1,\ \ldots,\ P),
$$

where Ω_p is the following $Gp \times Gp$ matrix:

(3.24) $\qquad \Omega_p = \Sigma_w \otimes I_p + \Sigma_u \otimes E_p$ $\hspace{3cm}$ $(p = 1, \ldots, P)$.

$\qquad$ Let now $\epsilon_{(p)}$ denote the $H(p)Gp \times 1$ vector containing all disturbances from the p'th group, i.e. from all individuals observed p times:

(3.25) $\qquad \epsilon_{(p)} = \begin{pmatrix} \epsilon_{(1,p)} \\ \vdots \\ \epsilon_{(H(p),p)} \end{pmatrix}$ $\hspace{3cm}$ $(p = 1, \ldots, P)$.

The covariance matrix of the complete disturbance vector

(3.26) $\qquad \epsilon_* = \begin{pmatrix} \epsilon_{(1)} \\ \vdots \\ \epsilon_{(P)} \end{pmatrix}$,

of dimension $\sum\limits_{p=1}^{P} H(p)Gp \times 1$, i.e. GTN $\times$ 1 (cf. (3.18)), can be written compactly as

(3.27) $\qquad E(\epsilon_* \epsilon_*') = \Omega_* = \begin{pmatrix} I_{H(1)} \otimes \Omega_1 \\ & I_{H(2)} \otimes \Omega_2 \\ & & \ddots \\ & & & I_{H(P)} \otimes \Omega_P \end{pmatrix}$,

where Ω_p is defined as in eq. (3.24).

$\qquad$ Of course, eq, (3.27) expresses the same hypothesis regarding the second order moments of the disturbances as eq. (3.16). The difference lies in the different ordering of observations to form the vectors ϵ (cf. eqs. (3.7), (3.8), (3.11), and (3.14)) and ϵ_* (cf. eqs. (3.21), (3.22), (3.25), and (3.26)). The ordering which leads to ϵ_* implies that the corresponding covariance matrix, Ω_*, assumes a *block-diagonal* form: First the block $\Omega_1 = \Sigma_w + \Sigma_u$ is repeated H(1) times, then the block $\Omega_2 = \Sigma_w \otimes I_2 + \Sigma_u \otimes E_2$ is repeated H(2) times, the block $\Omega_3 = \Sigma_w \otimes I_3 + \Sigma_u \otimes E_3$ is repeated H(3) times, etc. Needless to say, this will be an obvious virtue when it comes to estimation.

4. ESTIMATION

$\qquad$ Following the organization of section 3, we shall also divide the discussion of estimation problems into two separate parts. First, in section 4.1 , we outline a procedure for Full Information Maximum Likelihood estimation on the basis of the general formulae for the disturbance vector ϵ and its covariance matrix Ω in section 3.1. Then, in section 4.2, we turn to the reordered disturbance vector ϵ_* , as given in section 3.2, and discuss in more detail a step-wise algorithm for estimating the two components model.

4.1. Outline of a procedure for FIML estimation in the general three components model

$\qquad$ Let y_i denote the TN $\times$ 1 vector of observations on the endogenous variable in the i'th of the G regression equations in (2.1), ordered in the same way as the disturbance vector ϵ_i (cf. (3.1), (3.8), and (3.11)), and let X_i be the corresponding matrix of observations on the exogenous variables. (We do not refer the formal definitions of y_i and X_i since they are not needed in the following: y_i is formed by an appropriate selection of the scalars y_{iht}, and X_i is constructed by an appropriate selection of the row vectors x_{iht}.) The i'th equation can then be written as

(4.1) $y_i = X_i \beta_i + \epsilon_i$ $(i = 1, \ldots, G)$.

Combining all equations in one system, we get

(4.2) $y = X\beta + \epsilon$,

where ϵ is defined as in (3.14) and

$$
y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{pmatrix}, \quad
X = \begin{pmatrix} X_1 & 0 & \cdots\cdots & 0 \\ 0 & X_2 & \cdots\cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots\cdots & X_G \end{pmatrix}, \quad
\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{pmatrix}.
$$

        Provided that the disturbances follow a multivariate normal distribution, the log-likelihood function of y is

(4.3) $L = L(y, X; \beta, \Omega) = -\dfrac{GTN}{2} \log (2\pi) - \dfrac{1}{2} \log |\Omega| - \dfrac{1}{2} \epsilon'\Omega^{-1}\epsilon$,

where Ω is defined as in (3.15), using ϵ as a shorthand notation for $y - X\beta$.

        Formally, the problem of *Full Information Maximum Likelihood* (FIML) estimation of the complete model can now be expressed as follows: *Maximize L, or equivalently, minimize*

(4.4) $g = \log |\Omega| + \epsilon'\Omega^{-1}\epsilon$,

where

$$\Omega = \Sigma_w \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N + \Sigma_u \otimes DD'$$

with respect to β, Σ_u, Σ_v *and* Σ_w. In doing this, we pay regard to the definition of the stacked selection matrix D, which describes the sampling design (cf. eqs. (3.9) and (3.12)).

        Direct minimization of g may, however, raise numerical problems and easily become prohibitive in terms of computer costs. If certain regularity conditions are satisfied, FIML estimates can in principle be attained - to a prescribed degree of approximation - by iteratively switching between the following two subproblems:[7]

    (i) *Minimize* $Q = \epsilon'\Omega'^{-1}\epsilon$ *with respect to* β, *conditionally on* Ω.

    (ii) *Minimize* $g = \log |\Omega| + \epsilon'\Omega^{-1}\epsilon$, *where* $\Omega = \Sigma_w \otimes I_T \otimes I_N + \Sigma_v \otimes I_T \otimes E_N + \Sigma_u \otimes DD'$, *with respect to* Σ_u, Σ_v, *and* Σ_w, *conditionally on* β.

To start this algorithm, initial values have to be assigned to Σ_u, Σ_v and Σ_w. A convenient choice may be to set $\Sigma_u = \Sigma_v = 0$ and let Σ_w be a diagonal matrix with the i'th element equal to the estimated disturbance variance from a preliminary least squares regression on the i'th equation $(i = 1, \ldots, G)$.

        Solving *subproblem (i)* is equivalent to Aitken Generalized Least Squares (GLS) estimation of β. The resulting estimator is

(4.5) $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$.

It is well-known that this estimator is more efficient than the one obtained by running Ordinary Least Squares (OLS) regression on each of the G equations (4.1) in all cases except the following two:[8]

7) If this zig-zag procedure converges towards a unique solution, it will give the FIML estimates. For a formal demonstration of this, see Oberhofer and Kmenta (1974).
8) Cf. Zellner (1962, p. 351) and Theil (1971, pp. 308 - 310). We implicitly assume that no cross-equational constraints have been imposed on the coefficient vectors $\beta_1, \ldots, \beta_G$. If such constraints apply, OLS will be inefficient also in cases (a) and (b).

(a) All G equations contain exactly the same explanatory variables - i.e. X can be written as $X = I_G \otimes \bar{X}$ - and Ω is a bloc-diagonal matrix of the form $\Omega = \Sigma \otimes I_{TN}$.

(b) No elements in the disturbance vector ϵ are correlated, i.e. Ω can be written as $\Omega = \text{diag}(\sigma_{11} \ldots \sigma_{GG}) \otimes I_{TN}$.

Neither of these specifications will be valid in a disturbance components model with either individual or time specific effects, or both, included. Thus, we can safely conclude that even the conditional GLS estimator which solves subproblem (i) is more efficient than OLS. This conclusion holds even if all equations contain exactly the same set of exogenous variable, i.e.

$$X_1 = \ldots = X_G = \underline{X} \ (\text{or } X = I_G \otimes \underline{X}).^{9)}$$

An expression for the inverse of the covariance matrix Ω is indispensible for solving subproblem (i) algebraically. Unfortunately, it is difficult to derive such an expression in the general case considered here - mainly due to the presence of the stacked selection matrix D in the expression for Ω.[10] Moreover, Ω will usually be a matrix of considerable dimension. If N = 100, T = 10, and G = 5 - to take a realistic example - then Ω will be a 5 000 x 5 000 matrix. All who are familiar with computer problems, know that the inversion of a matrix of this order is no trivial matter.

In two *special cases*, however, algebraic expressions for Ω^{-1} can be easily derived. These are the cases with *completely overlapping samples* (m = 0, i.e. $D = 1_T \otimes I_N$) and with *non-overlapping samples* (m = N, i.e. $D = I_T \otimes I_N$), cf. eqs. (3.15 a and b). By a slight reformulation, we find that the corresponding expressions for the covariance matrices can be written as

$$(4.6) \quad \Omega_C = \Sigma_w \otimes (I_T - \frac{E_T}{T}) \otimes (I_N - \frac{E_N}{N}) + (\Sigma_w + T\Sigma_u) \otimes \frac{E_T}{T} \otimes (I_N - \frac{E_N}{N})$$
$$+ (\Sigma_w + N\Sigma_v) \otimes (I_T - \frac{E_T}{T}) \otimes \frac{E_N}{N} + (\Sigma_w + T\Sigma_u + N\Sigma_v) \otimes \frac{E_T}{T} \otimes \frac{E_N}{N},$$

and

$$(4.7) \quad \Omega_N = (\Sigma_w + \Sigma_u) \otimes I_T \otimes (I_N - \frac{E_N}{N}) + (\Sigma_w + \Sigma_u + N\Sigma_v) \otimes I_T \otimes \frac{E_N}{N},$$

respectively. Now, it is easy to show that both $I_N - E_N/N$ and E_N/N are idempotent matrices and that

$$(I_N - \frac{E_N}{N}) \frac{E_N}{N} = \frac{E_N}{N} (I_N - \frac{E_N}{N}) = 0_{N,N},$$

regardless of the value of N. From this fact and familiar properties of Kronecker products[11] it then follows that[12]

$$(4.8) \quad \Omega_C^{-1} = \Sigma_w^{-1} \otimes (I_T - \frac{E_T}{T}) \otimes (I_N - \frac{E_N}{N}) + (\Sigma_w + T\Sigma_u)^{-1} \otimes \frac{E_T}{T} \otimes (I_N - \frac{E_N}{N})$$
$$+ (\Sigma_w + N\Sigma_v)^{-1} \otimes (I_T - \frac{E_T}{T}) \otimes \frac{E_N}{N} + (\Sigma_w + T\Sigma_u + N\Sigma_v)^{-1} \otimes \frac{E_T}{T} \otimes \frac{E_N}{N},$$

---

9) A similar observation has been made by Avery (1977), for the seemingly unrelated regressions model based on complete cross-section/time-series data.
10) In a recent paper, Wansbeek and Kapteyn (1981) have addressed a similar problem for a single equation model. Their formulae, however, turn out to be rather messy and do not invite attempts at generalization to multi-equation models.
11) See e.g. Theil (1971, pp. 303 - 306).
12) Cf. also Baltagi (1980), eqs. (7), (11), and Lemma. A lucid treatment of the corresponding situation in a single equation model is given in Mazodier (1971).

and

$$(4.9) \qquad \Omega_N^{-1} = (\Sigma_w + \Sigma_u)^{-1} \otimes I_T \otimes (I_N - \frac{E_N}{N}) + (\Sigma_w + \Sigma_u + N\Sigma_v)^{-1} \otimes I_T \otimes \frac{E_N}{N} \ .$$

The problem of calculating the inverse of the covariance matrix is then reduced to the problem of inverting the four GxG matrices

$$\Sigma_w,$$

$$\Sigma_w + T\Sigma_u,$$

$$\Sigma_w + N\Sigma_v,$$

and $\quad \Sigma_w + T\Sigma_u + N\Sigma_v,$

in the case with completely overlapping samples, and to the problem of inverting the two GxG matrices

$$\Sigma_w + \Sigma_u$$

and $\quad \Sigma_w + \Sigma_u + N\Sigma_v,$

in the case with non-overlapping samples. Inserting (4.8) and (4.9) into (4.5), we obtain the conditional GLS estimators of β which solve subproblem (i) in these two special cases.

The solution to *subproblem (ii)* is even more complicated since the minimand g contains not only the inverse of Ω, but also its determinant value, $|\Omega|$. Even in the special case with completely overlapping samples (m = 0, D = $1_T \otimes I_N$) it involves the solution of a highly non-linear equation system.[13]

By pointing out this unsolved problem, we end our discussion of estimation procedures for the general three components model. We now drop the time specific component; this will make things considerably easier, as we shall see in the next section.

4.2. Iterative FIML estimation in the two components model

Let $y_{i(h,p)}$ denote the pxl vector containing the observations on the i'th endogenous variable from individual (h,p), ordered in the same way as $\epsilon_{i(h,p)}$ (cf. eq. (3.19)), and let $X_{i(h,p)}$ be the corresponding matrix of observations on the exogenous variables. The i'th equation can then be written as

$$(4.10) \qquad y_{i(h,p)} = X_{i(h,p)} \beta_i + \epsilon_{i(h,p)} \qquad\qquad \begin{array}{l} (i = 1, \ldots, G) \\ (h = 1, \ldots, H(p)) \\ (p = 1, \ldots, P). \end{array}$$

Combining all G equations in one system, we have

---

13) This problem has been thoroughly discussed by Amemiya (1971) for the corresponding single equation model. He shows, however, that simpler "analysis of variance estimates" exist, which are asymptotically equivalent to the Maximum Likelihood estimates. See also Magnus (1978) and Baltagi (1980).

(4.11) $\qquad y_{(h,p)} = X_{(h,p)}\,\beta + \epsilon_{(h,p)}$ $\qquad\qquad$ $(h = 1, \ldots, H(p))$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $(p = 1, \ldots, P),$

where $\epsilon_{(h,p)}$ is defined as in eq. (3.22) and

$$y_{(h,p)} = \begin{pmatrix} y_{1(h,p)} \\ \vdots \\ y_{G(h,p)} \end{pmatrix} \quad , \quad X_{(h,p)} = \begin{pmatrix} X_{1(h,p)} \cdots\cdots\cdots 0 \\ \ddots \\ 0 \cdots\cdots\cdots\cdots X_{G(h,p)} \end{pmatrix} \quad , \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_G \end{pmatrix}$$

$\qquad$ The log-likelihood function can be written in terms of ϵ_* and Ω_* (cf. (3.26) and (3.27)) as follows:

(4.12) $\qquad L = -\frac{GTN}{2} \log (2\pi) - \frac{1}{2} \log \left|\Omega_*\right| - \frac{1}{2} \epsilon_*{}' \Omega_*^{-1} \epsilon_*.$

From eqs. (3.23.) - (3.27) we find

(4.13) $\qquad \log \left|\Omega_*\right| = \sum_{p=1}^{P} H(p) \log \left|\Omega_p\right|$

$\qquad\qquad\qquad\qquad = \sum_{p=1}^{P} H(p) \log \left|\Sigma_w \otimes I_p + \Sigma_u \otimes E_p\right|$

and

(4.14) $\qquad \epsilon_*{}' \Omega_*^{-1} \epsilon_* = \sum_{p=1}^{P} Q_p,$

where

(4.15) $\qquad Q_p = \sum_{h=1}^{H(p)} \epsilon'_{(h,p)}\, \Omega_p^{-1}\, \epsilon_{(h,p)}$

$\qquad\qquad\qquad = \sum_{h=1}^{H(p)} \epsilon'_{(h,p)}\, [\Sigma_w \otimes I_p + \Sigma_u \otimes E_p]^{-1}\, \epsilon_{(h,p)}.$

Inserting (4.13) and (4.14) in (4.12) yields

(4.16) $\qquad L = -\frac{GTN}{2} \log (2\pi) - \frac{1}{2} \sum_{p=1}^{P} g_p,$

where

(4.17) $\qquad g_p = H(p) \log \left|\Omega_p\right| + Q_p = H(p) \log \left|\Sigma_w \otimes I_p + \Sigma_u \otimes E_p\right| + Q_p$ $\qquad$ $(p = 1, \ldots, P).$

In the expression for Q_p, cf. (4.15), we interpret $\epsilon_{(h,p)}$ as a shorthand for $y_{(h,p)} - X_{(h,p)}\beta$.
$\qquad$ Thus, the log-likelihood function is, apart from an arbitrary constant, a sum of P terms, one for each of the P groups of individuals. The p'th term, $-\frac{1}{2} g_p$, represents those individuals which are observed p times. This separability is, of course, due to the block-diagonal structure of Ω_*, which reflects the omission of time specific disturbance components, and our particular way of ordering the elements in the disturbance vector ϵ_*.

In this simplified model, FIML estimation corresponds formally to minimizing $g = \sum_{p=1}^{P} g_p$ with respect to β, Σ_u and Σ_w, subject to (4.15) and (4.17). The iterative procedure for solving this problem consists of the following two subproblems:

(i) *Minimize* $Q = \sum_{p=1}^{P} Q_p$ *with respect to* β, *conditionally on* Ω_p $(p = 1, \ldots, P)$.

(ii) *Minimize* $g = \sum_{p=1}^{P} g_p$ *with respect to* Σ_u *and* Σ_w, *conditionally on* β.

Let us consider these two subproblems in turn.

Subproblem (i): Minimization of Q with respect to β

The matrix Ω_p can be written as

$$\Omega_p = \Sigma_w \otimes (I_p - \frac{E_p}{p}) + (p\Sigma_u + \Sigma_w) \otimes \frac{E_p}{p} .$$

By utilizing the fact that $I_p - E_p/p$ and E_p/p are both idempotent for all positive integers p, we easily find that its inverse is[14]

(4.18) $$\Omega_p^{-1} = \Sigma_w^{-1} \otimes (I_p - \frac{E_p}{p}) + (p\Sigma_u + \Sigma_w)^{-1} \otimes \frac{E_p}{p}$$ $(p = 1, \ldots, P)$.

Since $(p\Sigma_u + \Sigma_w)^{-1}$ is symmetric and non-singular, there exists a non-singular matrix S_p, of dimension $G \times G$, such that[15]

(4.19) $$(p\Sigma_u + \Sigma_w)^{-1} = S_p' S_p$$ $(p = 0, 1, \ldots, P)$.

Combining (4.18) and (4.19), while again using the fact that $I_p - E_p/p$ and E_p/p are idempotent, we find that Ω_p^{-1} can be factorized as follows:

(4.20) $$\Omega_p^{-1} = F_p' F_p$$ $(p = 1, \ldots, P)$,

where

(4.21) $$F_p = S_0 \otimes (I_p - \frac{E_p}{p}) + S_p \otimes \frac{E_p}{p}$$

$$= [S_0 \otimes I_p] [I_G \otimes I_p - (I_G - S_0^{-1} S_p) \otimes \frac{E_p}{p}]$$ $(p = 1, \ldots, P)$.

Then the quadratic form relating to the p'th group of individuals, Q_p, can be written as a sum of squares

(4.22) $$Q_p = \sum_{h=1}^{H(p)} \eta'_{(h,p)} \eta_{(h,p)}$$ $(p = 1, \ldots, P)$,

where $\eta_{(h,p)}$ is the $Gp \times 1$ vector of transformed disturbances

---

14) Confer the derivation of (4.8) - (4.9) from (4.6) - (4.7) above.
15) This factorization is not unique, but this is immaterial for the following application.

(4.23) $\qquad \eta_{(h,p)} = F_p \epsilon_{(h, p)}$ $\qquad\qquad$ $(h = 1, .., H(p))$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $(p = 1, ..., P)$.

The algorithm for solving subproblem (i) thus boils down to

(a) *Compute the matrices* $(p\Sigma_u + \Sigma_w)^{-1}$ *for* $p = 0, 1, ..., P$, *and perform the factorization* (4.19)

(b) *Compute* F_p *by means of* (4.21) $(p = 1, ..., P)$.

(c) *Premultiply the observation matrices* $y_{(h, p)}$ *and* $X_{(h, p)}$ *from all individuals observed* p *times by the matrix* F_p $(p = 1, ..., P)$.

(d) *Minimize the overall sum of squares of the transformed disturbances,* $Q = \sum_{p=1}^{P} Q_p = \sum_{p=1}^{P} \sum_{h=1}^{H(p)} \eta'_{(h, p)} \eta_{(h, p)}$ *with respect to* β.

Subproblem (ii): Minimization of g with respect to Σ_u and Σ_w

In order to solve the second subproblem, we need expressions for the first derivatives of $\log|\Omega_p|$ and Q_p with respect to Σ_u and Σ_w. We can obtain these rather easily by utilizing formulae for matrix derivatives developed in Balestra (1975, sections 5.2 and 5.3).[16]

From eq. (3.24) above and Balestra (1975, eq. (5.3.18)) it follows that

$$\frac{\partial \log|\Omega_p|}{\partial \Sigma_w} = [I_G \otimes (\text{vec } I_p)'] \, [\Omega_p^{-1} \otimes I_p] \, [I_G \otimes (\text{vec } I_p)],$$

$$\frac{\partial \log|\Omega_p|}{\partial \Sigma_u} = [I_G \otimes (\text{vec } I_p)'] \, [\Omega_p^{-1} \otimes E_p] \, [I_G \otimes (\text{vec } I_p)]$$

$$(p = 1, ..., P).$$

By inserting (4.18), while noting that

$$(\text{vec } I_p)' \, (I_p \otimes I_p) \, (\text{vec } I_p) = p,$$

$$(\text{vec } I_p)' \, (E_p \otimes I_p) \, (\text{vec } I_p) = (\text{vec } I_p)' \, (I_p \otimes E_p)(\text{vec } I_p) = p,$$

$$(\text{vec } I_p)' \, (E_p \otimes E_p)(\text{vec } I_p) = p^2,$$

these expressions can be simplified to

(4.24) $\qquad \dfrac{\partial \log|\Omega_p|}{\partial \Sigma_w} = (p - 1) \, \Sigma_w^{-1} + (p\Sigma_u + \Sigma_w)^{-1},$

(4.25) $\qquad \dfrac{\partial \log|\Omega_p|}{\partial \Sigma_u} = p \, (p\Sigma_u + \Sigma_w)^{-1}$

$$(p = 1, ..., P).$$

---

16) Confer also Chamberlain and Griliches (1975, appendix).

We have now two of the elements needed for calculating the first derivatives of g with respect to the covariance matrices. Let us turn to the quadratic form Q_p, as defined in (4.15). First, we note that the part of Q_p which relates to individual (h, p) can be reformulated as

$$\epsilon'_{(h, p)} \, \Omega_p^{-1} \, \epsilon_{(h, p)} = (\text{vec } \widetilde{\epsilon}_{(h, p)})' \left\{ \Sigma_w^{-1} \otimes (I_p - \frac{E_p}{p}) \right\} (\text{vec } \widetilde{\epsilon}_{(h, p)})$$

$$+ (\text{vec } \widetilde{\epsilon}_{(h, p)})' \left\{ (p\Sigma_u + \Sigma_w)^{-1} \otimes \frac{E_p}{p} \right\} (\text{vec } \widetilde{\epsilon}_{(h, p)}),$$

by utilizing (3.22) and (4.18). Since the trace of a quadratic matrix of the form ABCD can be written as[17]

$$\text{tr} (ABCD) = (\text{vec } A')' \, (D' \otimes B) \, (\text{vec } C),$$

we can reformulate this equation as

$$\epsilon'_{(h, p)} \, \Omega_p^{-1} \, \epsilon_{(h, p)} = \text{tr} \, [\widetilde{\epsilon}'_{(h, p)} \, (I_p - \frac{E_p}{p}) \, \widetilde{\epsilon}_{(h, p)} \, \Sigma_w^{-1}]$$

$$+ \text{tr} \, [\widetilde{\epsilon}'_{(h, p)} \, \frac{E_p}{p} \, \widetilde{\epsilon}_{(h, p)} \, (p\Sigma_u + \Sigma_w)^{-1}].$$

Hence,

(4.26) $$Q_p = \text{tr}[(C_p - \overline{C}_p)\Sigma_w^{-1}] + \text{tr}[\overline{C}_p(p\Sigma_u + \Sigma_w)^{-1}]$$

$$(p = 1, \ldots, P),$$

where C_p and $\overline{C}_p$ are G x G matrices defined as

(4.27) $$C_p = \sum_{h=1}^{H(p)} \widetilde{\epsilon}'_{(h, p)} \, \widetilde{\epsilon}_{(h, p)},$$

(4.28) $$\overline{C}_p = \frac{1}{p} \sum_{h=1}^{H(p)} \widetilde{\epsilon}'_{(h, p)} \, E_p \, \widetilde{\epsilon}_{(h, p)} \qquad (p = 1, \ldots, P).$$

From eq. (4.26) above and Balestra (1975, eq. (5.2.19)) we derive

(4.29) $$\frac{\partial Q_p}{\partial \Sigma_w} = - \Sigma_w^{-1} \, (C_p - \overline{C}_p) \, \Sigma_w^{-1} - (p\Sigma_u + \Sigma_w)^{-1} \, \overline{C}_p \, (p\Sigma_u + \Sigma_w)^{-1},$$

(4.30) $$\frac{\partial Q_p}{\partial \Sigma_u} = - p(p\Sigma_u + \Sigma_w)^{-1} \, \overline{C}_p \, (p\Sigma_u + \Sigma_w)^{-1} \qquad (p = 1, \ldots, P).$$

The first order conditions for minimizing g with respect to the covariance matrices now follow by inserting (4.24), (4.25), (4.29), and (4.30) in

(4.31) $$\frac{\partial g}{\partial \Sigma_w} = \sum_{p=1}^{P} \frac{\partial g_p}{\partial \Sigma_w} = \sum_{p=1}^{P} \left\{ H(p) \frac{\partial \log |\Omega_p|}{\partial \Sigma_w} + \frac{\partial Q_p}{\partial \Sigma_w} \right\} = 0,$$

(4.32) $$\frac{\partial g}{\partial \Sigma_u} = \sum_{p=1}^{P} \frac{\partial g_p}{\partial \Sigma_u} = \sum_{p=1}^{P} \left\{ H(p) \frac{\partial \log |\Omega_p|}{\partial \Sigma_u} + \frac{\partial Q_p}{\partial \Sigma_u} \right\} = 0.$$

---

[17] See Balestra (1975, p. 20) or Magnus (1978, eq. (4)).

This gives a non-linear system of $G(G+1)$ equations in the unknown variances/covariances σ_{ij}^{w} and σ_{ij}^{u}, and defines our first-order conditions for solving subproblem (ii).

The numerical solution of (4.31) - (4.32) may be intricate, even for low values of P and/or G, and an iterative algorithm may be the only feasible approach. To obtain sensible *initial values* of Σ_w and Σ_u from which to start the iteration process, the following approach may be useful:

Assume, hypothetically, that our data consisted of observations from individuals observed p times only. The equation system for solving subproblem (ii) would then be

(4.33)
$$\frac{\partial g_p}{\partial \Sigma_w} = H(p) \, [(p - 1) \, \Sigma_w^{-1} + (p\Sigma_u + \Sigma_w)^{-1}]$$
$$- \Sigma_w^{-1} \, (C_p - \overline{C}_p) \, \Sigma_w^{-1} - (p\Sigma_u + \Sigma_w)^{-1} \, \overline{C}_p \, (p\Sigma_u + \Sigma_w)^{-1} = 0,$$

(4.34)
$$\frac{\partial g_p}{\partial \Sigma_u} = H(p) \, p \, (p\Sigma_u + \Sigma_w)^{-1} - p(p\Sigma_u + \Sigma_w)^{-1} \, \overline{C}_p \, (p\Sigma_u + \Sigma_w)^{-1} = 0.$$

This system can be easily solved. Let its solution be $\Sigma_w(p)$ and $\Sigma_u(p)$. From (4.34) it follows directly that

$$p\Sigma_u(p) + \Sigma_w(p) = \frac{\overline{C}_p}{H(p)}.$$

Inserting this into (4.33), we get

$$H(p) \, (p - 1) \, \Sigma_w(p)^{-1} = \Sigma_w(p)^{-1} \, (C_p - \overline{C}_p) \, \Sigma_w(p)^{-1}.$$

Hence,

(4.35)
$$\Sigma_w(p) = \frac{C_p - \overline{C}_p}{(p - 1) \, H(p)},$$

(4.36)
$$\Sigma_u(p) = \frac{p \, \overline{C}_p - C_p}{p \, (p - 1) \, H(p)}.$$

These are the estimates of Σ_w and Σ_u that we would obtain by utilizing data from the p'th group of individuals only. The corresponding estimator of the 'total' covariance matrix $\Sigma = \Sigma_u + \Sigma_w$ is

(4.37)
$$\Sigma(p) = \Sigma_u(p) + \Sigma_w(p) = \frac{C_p}{pH(p)} .$$

Usually, (4.35) and (4.36) will give different 'estimates' for different values of p. The estimates which solve subproblem (ii) may be considered as 'compromise values' of $\Sigma_w(p)$ and $\Sigma_u(p)$ for $p = 1, \ldots, P$. These considerations suggest that an appropriately weighted average of the P group specific estimates will be sensible initial values for an iterative solution of the complete equation system (4.31) - (4.32). We may, for instance, use the number of observations on which each estimate is based as weights. This would give the following initial values for Σ_u and Σ_w for starting the iteration process in subproblem (ii):

$$\Sigma_u^{0} = \frac{\sum\limits_{p=1}^{P} \left\{ pH(p) \, \Sigma_u(p) \right\}}{\sum\limits_{p=1}^{P} pH(p)} = \frac{1}{TN} \sum\limits_{p=1}^{P} \left\{ \frac{p\overline{C}_p - C_p}{p - 1} \right\} ,$$

$$\Sigma_w^{0} = \frac{1}{TN} \sum\limits_{p=1}^{P} \frac{p}{p - 1} \, (\overline{C}_p - C_p).$$

These values can be easily calculated once we have calculated the matrices C_p and $\overline{C}_p$ from the distur-
bances obtained by solving subproblem (i), cf. eqs. (3.21), (4.27) and (4.28).

A closer look at the single equation case

To get a closer understanding of the 'anatomy' of this iterative algorithm, let us see what it
implies in the case where the model contains one equation only. When $G = 1$, we have, with simplified
notation,

$$(4.38) \qquad \Sigma_u = \sigma_u^2 = \rho\sigma^2,$$

$$\Sigma_w = \sigma_w^2 = (1-\rho)\sigma^2,$$

where $\sigma^2 = \sigma_u^2 + \sigma_w^2$ is the total disturbance variance and $\rho = \sigma_u^2/\sigma^2$ is the part of this total which
is due to individual variations. Alternatively, ρ can be interpreted as the coefficient of correlation
between two disturbances from the same individual.

Subproblem (i)

When $G = 1$, the S_p's, as defined in eq. (4.19), are scalars given by

$$S_0 = \frac{1}{\sigma_w} = \frac{1}{\sigma\sqrt{1-\rho}} \quad,$$

$$S_p = \frac{1}{\sqrt{\rho\sigma_u^2 + \sigma_w^2}} = \frac{1}{\sigma\sqrt{1 + (p-1)\rho}} \qquad (p = 1, \ldots, P).$$

The expressions for the transformation matrices, (4.21), then become

$$(4.39) \qquad F_p = \frac{1}{\sigma\sqrt{1-\rho}} \; [I_p - (1 - \alpha_p)\frac{E_p}{p}] \qquad (p = 1, \ldots, P),$$

where

$$(4.40) \qquad \alpha_p = \left\{ \frac{1-\rho}{1 + (p-1)\rho} \right\}^{\frac{1}{2}} \qquad (p = 1, \ldots, P).$$

All the matrices F_p contain $S_0 = \sigma^{-1}(1-\rho)^{-\frac{1}{2}}$ as a common scalar factor, which is irrelevant
for solving subproblem (i), so we may equally well use $F_p/S_0 = I_p - (1 - \alpha_p)E_p/p$ as transformation
matrices. Premultiplying the original disturbance vectors $\epsilon_{(h, p)}$ by these matrices to form the sum of
squares to be minimized then implies the following:[18]

> Multiply the observations from individuals observed only once by $1 - (1 - \alpha_1) \doteq \alpha_1 \doteq (1 - \rho)^{\frac{1}{2}}$,
> subtract from the observations of all individuals observed p times (p = 2, 3, ... P) a fraction
> $1 - \alpha_p \doteq 1 - [(1 - \rho)/(1 + (p - 1)\rho)]^{\frac{1}{2}}$ of the corresponding individual average,[19] and
> minimize the resulting sum of squares of disturbances.

We note that $1 - \alpha_p$, the fraction of the individual average to be subtracted, is an increasing function
of ρ, the individual share in the total variance. It is also an increasing function of p, the number
of replications. Or stated otherwise: The stronger the disturbances from the same individual are
correlated and the larger the number of times each individual is observed, the larger fraction of the
individual average should be subtracted to give the transformed disturbances. Table 1 reports the
value of $1 - \alpha_p$ for selected values of these two parameters.

---

18) The same algorithm is stated, without proof, in Biørn (1981, p. 229).
19) Recall that premultiplying a p x 1 vector by E_p/p implies to replace all its elements by
their average value.

Table 1. The transformation parameter $1 - \alpha_p = 1 - \left\{ \dfrac{1 - \rho}{1 + (p - 1)\rho} \right\}^{\frac{1}{2}}$

p	ρ							
	0	0.01	0.1	0.2	0.5	0.8	0.9	0.999
1	0	0.0050	0.0513	0.1056	0.2929	0.5528	0.6838	0.9684
2	0	0.0100	0.0955	0.1835	0.4226	0.6667	0.7706	0.9776
3	0	0.0148	0.1340	0.2441	0.5000	0.7226	0.8110	0.9817
4	0	0.0196	0.1679	0.2929	0.5528	0.7575	0.8356	0.9842
5	0	0.0243	0.1982	0.3333	0.5918	0.7818	0.8526	0.9859
10	0	0.0470	0.3118	0.4655	0.6985	0.8438	0.8952	0.9900
20	0	0.0879	0.4429	0.5918	0.7818	0.8889	0.9257	0.9929
100	0	0.2947	0.7127	0.8039	0.9005	0.9501	0.9667	0.9968
∞	0	1	1	1	1	1	1	1

Subproblem (ii)

When $G = 1$, eqs. (4.24), (4.25), (4.29), and (4.30) can be simplified to

$$\frac{\partial \log |\Omega_p|}{\partial \sigma_w^2} = \frac{p-1}{\sigma_w^2} + \frac{1}{p\sigma_u^2 + \sigma_w^2},$$

$$\frac{\partial \log |\Omega_{\bar{p}}|}{\partial \sigma_u^2} = \frac{p}{p\sigma_u^2 + \sigma_w^2},$$

$$\frac{\partial Q_p}{\partial \sigma_w^2} = -\frac{1}{\sigma_w^4}(C_p - \bar{C}_p) - \frac{1}{(p\sigma_u^2 + \sigma_w^2)^2}\bar{C}_p,$$

$$\frac{\partial Q_p}{\partial \sigma_u^2} = -\frac{p}{(p\sigma_u^2 + \sigma_w^2)^2}\bar{C}_p.$$

Inserting these expressions into the first-order conditions for subproblem (ii), (4.31)-(4.32), while substituting (4.38), we obtain after some rearranging

$$(4.41) \quad \sum_{p=1}^{P}\{\sigma^2(1-\rho)^{-1} H(p)(p-1) - (1-\rho)^{-2}(C_p - \bar{C}_p)\}$$

$$+ \sum_{p=1}^{P}\{\sigma^2[1+(p-1)\rho]^{-1} H(p) - [1+(p-1)\rho]^{-2}\bar{C}_p\} = 0,$$

$$(4.42) \quad \sum_{p=1}^{P}\{\sigma^2[1+(p-1)\rho]^{-1} H(p)p - [1+(p-1)\rho]^{-2}p\,\bar{C}_p\} = 0.$$

A procedure for solving these two non-linear equations, and hence subproblem (ii), may for instance be: first, express σ^2 as a function of ρ by utilizing (4.42):

$$(4.43) \quad \sigma^2 = \frac{\sum\limits_{p=1}^{P} [1+(p-1)\rho]^{-2} \, p \, \overline{C}_p}{\sum\limits_{p=1}^{P} [1+(p-1)\rho]^{-1} pH(p)} \,,$$

second, insert this expression in (4.41), and solve the resulting equation with respect to ρ by means of a grid search procedure, and third, insert the solution in (4.43). Given the complexity of our problem, this is a remarkably simple estimation procedure.

5. CONCLUSION

Conbined cross-section/time-series data constitute a very wide class of data structures, and both from a theoretical and, in particular, from a practical point of view the standard model which assumes complete time series to exist for all the observation units is a very restrictive one. In this paper, we have tried to widen the scope a bit by adopting a specification which allows for incomplete time series and which is applicable both in a multi-equation and a single equation context. Of course, the estimation problem becomes more cumbersome in this case than for the simpler models usually discussed in the literature. When the panel of individuals rotates, it is, in particular, the possible presence of time specific disturbance components which contributes to the complexity. This is the price we have to pay to make the model applicable to a wide class of practically important data structures.

If time specific components are omitted - which is realistic in many situations - the estimation problem may be simplified substantially, and an iterative algorithm for FIML estimation can be fairly easily implemented on a computer. This algorithm is particularly transparent in the single equation case - indeed it is surprisingly simple - but it is applicable also to linear multi-equation models, although it may involve a lot of programming work and a considerable amount of computer time. Some practical experiences should be gained before we can decide how well it will work.

The model specification studied in this paper includes several models discussed in the literature as special cases. To put our approach in perspective, let us give a few examples:

Three components specification: Σ_u, Σ_v, Σ_w unrestricted

    G > 1, m = 0: Avery (1977) (Only GLS), Baltagi (1980).

    G = 1, m = 0: Wallace-Hussain (1969), Nerlove (1971b), Mazodier (1971) (Only GLS), Amemiya (1971).

    G = 1, 0 < m < N: Wansbeek-Kapteyn (1981).

Two components specification: $\Sigma_v = 0$, Σ_u and Σ_w unrestricted

    G > 1, m = 0[20]: Chamberlain-Griliches (1975, appendix), Balestra (1975, Ch. 6).
    G = 1, m = 0[20]: Balestra-Nerlove (1966).
    G = 1, m = N/2[21]: Biørn (1981).

---

20) Or equivalenty: $G \geqq 1$, $H(T) = N$, $H(p) = 0$ for $p \neq T$.
21) Or equivalenty: $G = 1$, $H(1) = N$, $H(2) = (T-1)N/2$, $H(p) = 0$ for $p > 2$.

One component specification: $\Sigma_v = \Sigma_u = 0$; Σ_w unrestricted

        $G > 1$, $m \geqq 0$:   Conventional 'Seemingly Unrelated Regressions' model; Zellner (1962).

        $G = 1$, $m \geqq 0$:   Standard linear regression model.

        Our approach does not, however, include the models discussed in recent years in connection with the socalled "selectivity bias" (or "self-selection bias") problem.[22]   We have implicitly disregarded the possibility that the panel may be subjected to a systematically changing degree of non-response by assuming that the sampling design is determined prior to and independent from the individual decisions represented by the structural equations of the model.

---

22) See e.g. Hausman and Wise (1977), Griliches, Hall, and Hausman (1978), and Maddala (1978).

REFERENCES

Amemiya, T. (1971): The Estimation of the Variances in a Variance - Components Model. *International Economic Review,* 12 (1971), pp. 1-13.

Avery, R.B. (1977): Error Components and Seemingly Unrelated Regressions. *Econometrica,* 45 (1977), pp. 199-209.

Balestra, P. (1975): *La Derivation Matricielle. Technique et Résultats pour Economistes.* (Paris: Sirey, 1975.)

Balestra P., and M. Nerlove (1966): Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. *Econometrica,* 34 (1966), pp. 585-612.

Baltagi, B.H. (1980): On Seemingly Unrelated Regressions with Error Components. *Econometrica,* 48 (1980), pp. 1547-1551.

Biørn, E. (1981): Estimating Economic Relations from Incomplete Cross-Section/Time-Series Data. *Journal of Econometrics,* 16 (1981), pp. 221-236.

Biørn, E., and E.S. Jansen (1982): *Econometrics of Incomplete Cross-Section/Time-Series Data: Consumer Demand in Norwegian Households 1975 - 1977.* Reports from the Central Bureau of Statistics of Norway, forthcoming. (Oslo: Central Bureau of Statistics, 1982.)

Chamberlain, G. and Z. Griliches (1975): Unobservables with a Variance-Components Structure: Ability, Schooling and the Economic Success of Brothers. *International Economic Review* 16 (1975), pp. 422-450.

Griliches, Z., B.H. Hall, and J.A. Hausman (1978): Missing Data and Self-Selection in Large Panels. *Annales de l'INSEE,* 30-31 (1978), pp. 137-176.

Hausman, J.A. and D.A. Wise (1977): Social Experimentation, Truncated Distributions, and Efficient Estimation. *Econometrica,* 45 (1977), pp. 919-938.

Maddala, G.S. (1971): The Use of Variance Components Models in Pooling Cross Section and Time Series Data. *Econometrica,* 39 (1971), pp. 341-358.

Maddala, G.S. (1978): Selectivity Problems in Longitudinal Data. *Annales de l'INSEE,* 30-31 (1978), pp. 423-450.

Magnus, J.R. (1978): Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix. *Journal of Econometrics,* 7 (1978), pp. 281-312.

Mazodier, P. (1971): L'Estimation des Modèles à Erreurs Composées. *Annales de l'INSEE,* 7 (mai-août 1971), pp.43-74.

Mundlak, Y. (1978): On the Pooling of Time Series and Cross Section Data. *Econometrica,* 46 (1978), pp. 69-85.

Nerlove, M. (1971a): Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections. *Econometrica,* 39 (1971), pp. 359-382.

Nerlove, M. (1971b): A Note on Error Components Models. *Econometrica,* 39 (1971), pp. 383-396.

Oberhofer, W., and J. Kmenta (1974): A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models. *Econometrica,* 42 (1974), pp. 579-590.

Theil, H. (1971): *Principles of Econometrics.* (New York: John Wiley & Sons, Inc., 1971.)

Wallace, T.D., and A. Hussain (1969): The Use of Error Components Models in Combining Cross Section with Time Series Data. *Econometrica,* 37 (1969), pp. 55-72.

Wansbeek, T., and A. Kapteyn (1981): Maximum Likelihood Estimation in Incomplete Models. Unpublished paper, Central Bureau voor de Statistiek, Voorburg, the Netherlands, August 1981.

Zellner, A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association,* 57 (1962), pp. 348-368.

Zellner, A. and D.S. Huang (1962): Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations. *International Economic Review,* 3 (1962), pp. 300-313.

Issued in the series Reports from the Central Bureau of Statistics (REP) - ISSN 0332-8422

Printed 1981

Nr. 81/1 Erling J. Fløttum: National Accounts of Norway System and Methods of Estimation
Sidetall 101 Pris kr 15,00 ISBN 82-537-1555-2

- 81/2 Referansearkiv for naturressurs- og forurensningsdata 2. utgave Sidetall 424
Pris kr 20,00 ISBN 82-537-1233-2

- 81/3 Nils Håvard Lund: Byggekostnadsindeks for boliger Sidetall 127 Pris kr 15,00
ISBN 82-537-1232-4

- 81/4 Anne Lise Ellingsæter: Intervjuernes erfaringer fra arbeidskraftundersøkelsene
Rapport fra 99 intervjuere Field Work Experiences with the Labour Force Sample
Survey Reports from 99 Interviewers Sidetall 40 Pris kr 10,00 ISBN 82-537-1234-0

- 81/5 Bjørn Kjensli: Strukturundersøkelse for bygg og anlegg Vann- og kloakkanlegg
Sidetall 62 Pris kr 15,00 ISBN 82-537-1235-9

- 81/6 Erling Siring og Ib Thomsen: Metoder for estimering av tall for fylker ved hjelp av
utvalgsundersøkelser Sidetall 42 Pris kr 10,00 ISBN 82-537-1509-9

- 81/7 Arne Ljones og Hans Viggo Sæbø: Temperaturkorrigering av energiforbruket
Sidetall 43 Pris kr 10,00 ISBN 82-537-1507-2

- 81/8 Morten Reymert: En analyse av faktorinnsatsen i Norges utenrikshandel med utvik-
lingsland og industriland Sidetall 55 Pris kr 15,00 ISBN 82-537-1506-4

- 81/9 Petter Longva: A System of Natural Resource Accounts Eit rekneskapssystem for
naturressursar Sidetal 26 Pris kr 10,00 ISBN 82-537-1540-4

- 81/10 Stein Erland Brun: Tilgangen på arbeidskraft i fylkene for årene 1971 - 1979
Sidetall 72 Pris kr 15,00 ISBN 82-537-1514-5

- 81/11 Eva Ivås og Kjell Roland: MODIS IV Detaljerte virkningstabeller for 1979
Sidetall 264 Pris kr 20,00 ISBN 82-537-1515-3

- 81/12 Helge Brunborg, Jan Mønnesland og Randi Selmer: Framskriving av folkemengden etter
ekteskapelig status Sidetall 75 Pris kr 11,00 ISBN 82-537-1541-2

- 81/13 Adne Cappelen: Importinnhold i sluttleveringer Sidetall 20 Pris kr 10,00
ISBN 82-537-1545-5

- 81/14 MODIS IV Dokumentasjonsnotat nr. 16 Endringer i utgave 78-1 og 79-1 Sidetall 100
Pris kr 15,00 ISBN 82-537-1549-8

- 81/15 Skatter og overføringer til private Historisk oversikt over satser mv. Årene
1969 - 1981 Sidetall 74 Pris kr 15,00 ISBN 82-537-1554-4

- 81/16 Helgeturer 1978/79 Sidetall 23 Pris kr 10,00 ISBN 82-537-1560-9

- 81/17 Roy Østensen: Eie og bruk av personbil Foreløpige tall for 1979 og 1. kvartal 1980
Sidetall 42 Pris kr 10,00 ISBN 82-537-1566-8

- 81/18 Svein Homstvedt, Øyvind Lone og Tore Nesheim: Jordbruksareal ifølge jordregister og
utvalgstellinger. Metodiske forskjeller belyst med materiale fra Trøgstad kommune
Sidetall 62 Pris kr 15,00 ISBN 82-537-1598-6

- 81/19 Arne Faye: Holdninger til norsk utviklingshjelp 1980 Sidetall 62 Pris kr 15,00
ISBN 82-537-1562-5

- 81/20 Knut Fredrik Strøm: Konkurser i industri og varehandel Utvikling, hyppighet og
omfang Sidetall 31 Pris kr 10,00 ISBN 82-537-1569-2

- 81/21 Frank Foyn: Miljøverninvesteringer i industrien. Problemer ved kartlegging av
data Sidetall 34 Pris kr 10,00 ISBN 82-537-1591-9

Issued in the Series Reports from the Central Bureau of Statistics (REP) - ISSN 0332-8422
(cont.)

Printed 1981

Nr. 81/22   Petter R. Koren: Etterspørsel etter energi i norsk industri Sidetall 27
            Pris kr 10,00 ISBN 82-537-1592-7

-   81/23   Harald Bergland og Adne Cappelen: Produktivitet og sysselsetting i industrien
            Sidetall 75 Pris kr 15,00 ISBN 82-537-1600-1

-   81/24   Levekårsundersøkelsen 1980 Dokumentasjon Del I Sidetall 67 Pris kr 15,00
            ISBN 82-537-1612-5

-   81/25   Tor Haldorsen: Norske ferieformer Sidetall 112 Pris kr 15,00 ISBN 82-537-1611-7

-   81/26   Aktuelle skattetall 1981   Current Tax Data   Sidetall 46 Pris kr 10,00
            ISBN 82-537-1610-9

-   81/27   Tiril Vogt: Planregnskap Ressursregnskap for fysisk planlegging Sidetall 70
            Pris kr 15,00 ISBN 82-537-1614-1

-   81/28   Figurer i publikasjoner Sidetall 115 Pris kr 15,00 ISBN 82-537-1628-1

-   81/29   Tor Skoglund: Utprøving av modellen REGION mot fylkesfordelte nasjonalregnskapsdata
            for perioden 1973 - 1976 Sidetall 42 Pris kr 10,00 ISBN 82-537-1626-5

-   81/30   Adne Cappelen, Erik Garaas og Svein Longva: MODAG En modell for makroøkonomiske
            analyser Sidetall 70 Pris kr 15,00 ISBN 82-537-1630-3

-   81/31   Torstein Bye og Tor Eivind Høyland: Inntektsbegreper - Inntektsfordeling
            Sidetall 43 Pris kr 10,00 ISBN 82-537-1639-7

-   81/32   Totalregnskap for fiske- og fangstnæringen 1977 - 1980 Sidetall 39 Pris kr 10,00
            ISBN 82-537-1634-6

-   81/33   Erik Biørn: Estimating Seemingly Unrelated Regression Models from Incomplete
            Cross-Section/Time-Series Data Sidetall 32 Pris kr 10,00 ISBN 82-537-1635-4

-   81/34   Eva Ivås og Kjell Roland: MODIS IV Detaljerte virkningstabeller for 1980
            Sidetall 272 Pris kr 20,00 ISBN 82-537-1636-2

-   81/35   Sigurd Tveitereid og Jan Lædre: Markedsindikatorer for norsk eksport Sidetall 47
            Pris kr 15,00 ISBN 82-537-1645-1

Printed 1982

Nr. 82/1   Tor Haldorsen og Gunvor Iversen: Praktiske eksempler på måling av latente variable:
           Sammenhengen mellom subjektive og objektive indikatorer på arbeidsforhold
           ISBN 82-537-1649-4   ISSN 0332-8422

-   82/5   Naturressurser 1981 Sidetall 29 Pris kr 10,00 ISBN 82-537-1651-6

-   82/7   Øystein Engebretsen: Arealbruk i norske byer og tettsteder Sidetall 183
           Pris kr 15,00 ISBN 82-537-1720-2

-   82/9   Rolf Aaberge: Om måling av ulikskap Sidetall 73 Pris kr 15,00 ISBN 82-537-1721-0

Pris kr 10,00

Publikasjonen utgis i kommisjon hos H. Aschehoug & Co. og
Universitetsforlaget, Oslo, og er til salgs hos alle bokhandlere.