The Econometrics of Panel Data

Michel Mouchart [1]

Institut de statistique
Université catholique de Louvain (B)

3rd March 2004

[1]Preliminary version- Not to be quoted-Comments welcome.

Contents

Preface

Panel data: the subject matter
The subject matter of Panel data basically deals with statistical data made of repeated measurements on a same statistical unit. Often, but not always, the several measurements are operated at different times; thus, in the econometric literature, frequent references are made to "pooling cross-sections and time-series" but it should be emphasized that, in Panel data, the different cross-sections bear on the same statistical units. Often the statistical units are "individuals", as human being, households or companies, but Panel data may also deal with statistical units that are "aggregates" as industrial sectors, regions or countries. Thus, Panel data are often considered as a subject matter for microeconometrics rather than for macroeconometrics, but this is actually somewhat arbitrary, although possibly justified if one considers the volume of the relevant literature published in these two fields.

The subject matter of Panel Data is also closely related to :

models for spatio-temporal data

multilevel analysis

The **Main objective** of this textbook is to expose the basic tools for modelling Panel Data, rather than a survey of the state of the art at a given date. This text is accordingly a first textbook on Panel data. A particular emphasis is given to the acquisition of the analytical tools that will help the readers to further study forthcoming materials in that field, not only in the traditional sources of econometrics but also in biometrics and a bio-statistics; indeed, these fields, under the heading of "Repeated measurements", have produced many earlier original contributions, often as extensions of the literature on the analysis of variances that first introduced the distinction between so-called "fixed effects" and "random effects" models. Thus this text tends to favour cross-fertilisations between different fields of the application of statistical methods.

The **Main readership** is targeted towards graduate students in econometrics or in statistics.

The basic **Prerequisite** for this text is a basic course in econometrics. Complements are included, and clearly separated, at the end of the volume with a double objective . Firstly, this text is intended to be reasonably self-contained for an heterogenous audience. It is indeed the experience that graduate programs in econometrics, or in economics, are attended by students with largely varying backgrounds: this may be viewed as a positive feature, provided due allowance is made to this heterogeneity. Thus such materials will be superfluous for some readers, or students, although being niew for other ones. This separation may possibly alleviate some unnecessary frustration of the reader. Secondly, a particular effort has been made to focus the attention on what is properly genuine to the field of Panel data; for this purpose it seemed suitable to physically separate, in the text, on one side, those materials pertaining to general linear algebra or to general statistical methods, and, on the other side, the kernel subject matter of panel data. By so-doing, the reader, and the teacher, may organize more efficiently the learning of the very topic of this textbook.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Nature and Characteristics of Panel Data

1.1.1 Definition and examples

Introductory example

$$
\begin{array}{lll}
x_{it} & : & x : \text{income} \\
 & & i : \text{household} \quad i = 1, 2, \ldots, N \\
 & & t : \text{period} \quad\quad t = 1, 2, \ldots, T
\end{array}
$$

Basic Ideas

1. *Data pooling times series and cross section.* More specifically:
A time series of cross-sections on (more or less) the *same* individuals
A cross-section of time-series on (more or less) the *same* periods
Typically obtained by repeating, at different times, surveys on (more or les) the same individuals.
2. *Individuals* may be true individuals or households, companies, countries etc. They are actually "statistical units" or "units of observation".

3. *For the sample size*, the number of individuals (N) and of time periods (T) may be strikingly different : there may be 10 or 10,000, or more, individuals and 2 or 10,000 , or more (as is the case in financial high-frequency data) time periods. This is crucial for modelling!

4. Also :

- *repeated measurements* (mainly in : biometrics)

  i.e. emphasis on : repetition of measurements on a same individual

  example : "growth curve", X_{it} : weight of rat i at time t

- *functional data*

  in non parametric methods, emphasis on "a function of time" observed on several individuals

  $\rightarrow$ often in discrete times, sometines in (almost) continuous time

  examples

  electro cardiogram

  stock exchange data.

- *longitudinal data*

  particularly in demography, sociology

5. Moreover, many communalities with : spatio-temporal data, multilivel analysis

A remark on notation

$$Y_{it} = y \leftrightarrow \quad \begin{aligned} X_k &= (Y_k, Z_k, T_k) \\ x_k &= (y, i, t) \end{aligned}$$

with k : index of a unit of observation
Y_k : variable(s) of interest
Z_k : indicator (or, identifier) of "individual"
T_k : indicator (or, identifier) of time

Generally :

$$Y_{it} \sim N(...) \quad \leftrightarrow \quad (Y_k|Z_k = i,\ T_k = t) \sim N(...)$$

i.e. Z_k, T_k : are typically treated as "explanatory" variables (subject to different types of coding)

1.2 Types of Panel Data

1.2.1 Balanced versus non-balanced data

Balanced case	:	$i = 1, 2, \ldots, N$	$t = 1, 2, \ldots, T$
Unbalanced case	:	$i = 1, 2, \ldots, N$	$t = 1, 2, \ldots, T_i$
			or: $t = t_{j_1}, t_{j_2}, \ldots, t_{j_{T_i}}$

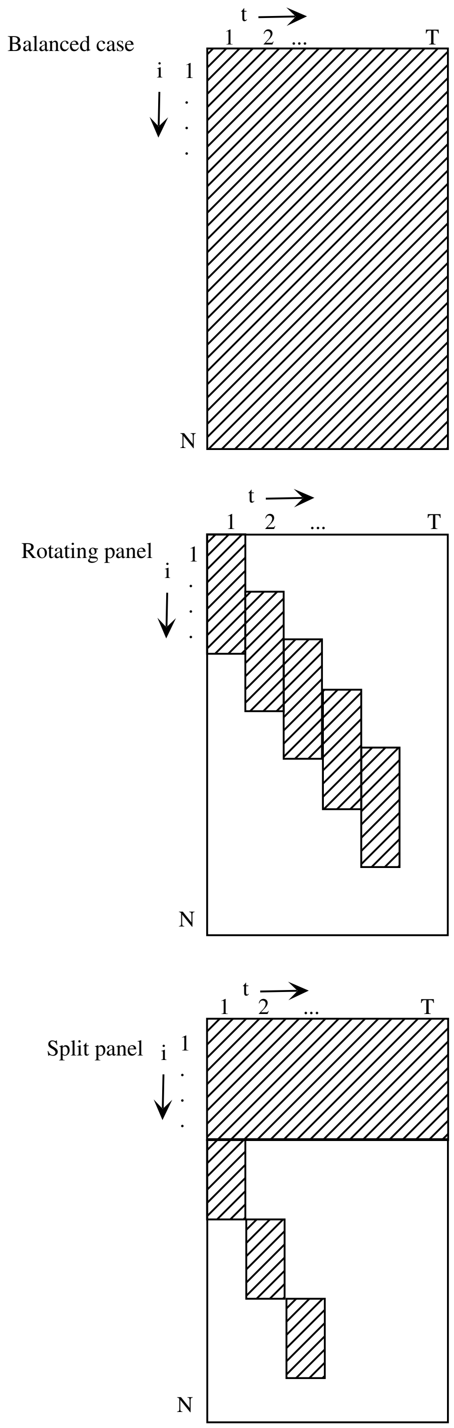
1.2.2 Limit cases : Time-series and Cross-section data

Two limit cases: a "pure" cross section with only one time period or a "pure" time-series with only one individual

These two limit cases act as frameworks of reference when modelling.

1.2.3 Rotating and Split Panel

Figure 1.1: *Rotating and Split Panel*



1.3 More examples

Panel study of income dynamics (PSID)

From : Survey Research Center (U. of Michigan - USA)
Begin : 1968 with 4802 families
Features :
annual interviews
± 5000 variables : socio economic characteristics
± 31000 individuals (also in derivative families)
oversampling of poor households

National longitudinal surveys of labour market experience (NLS)

From : Center of Human Resource Research (Ohio State University)
Begin : 1966
Features :

- Five distinct segments of The Labour Force

older men (45-49 in 66) :	5020
young men (14-24 in 66) :	5225
mature women (30-44 in 67) :	5083
young women (14-21 in 68) :	5159
youths (14-24 in 79) :	12686

- Oversampling of : blacks - hispanics -poor whites - military
- Thousands of variables (emphasis on labour supply)

Longitudinal retirement history supply

- men and non-married women 58-63 in 69
- 11 153 individuals
- variables : demographics, employment histories, debt, assets, incomes, health and living expenses.
- main interest: attitudes towards work and retirement

Social security adminstration's continuous work history sample

Labour Department's continuous wage and benefit history

Labour Department's continuous longitudinal manpower survey

Negative income tax experiments (several)

Current Population Survey (CPS)
basis for several panels,
Features (with respect to SPID) :

- fewer variables - shorter period
- but : larger sample - wider representativity

Consumer's panel(*e.g.* Nielsen)

In Europe :

- German Social Economic Panel (Germany)
- Household Market and Non market acivities (Sweden)
- Intomart Duch Panel of Households (Netherlands)

More generally
Data are available in many fields of application:

- econometrics: labopur market, fianncial market, marketing, *etc.*
- demography
- sociology
- political opinion surveys
- bio-medical fields (in particular, some clinical trials)

1.4 Some applications

- PSID and NLS :

  both designed

  - to study the nature and causes of poverty
  - to explain changes in economic well-being
  - to study the effects of economic and social programs

  600 published articles and monographs : wide range of topics including:

  - labor supply
  - family economic status
  - effects of transfer income programs
  - family composition changes
  - residential mobility

- Also, papers on:

  evaluation of training program for unemployed individuals

  financial markets

  demography

  poll of political opinions

  medical applications

  Efficiency analysis (production frontier)

1.5 Statistical modelling : benefits and limitations of panel data

1.5.1 Some characteristic features of P.D.

Object of this subsection : features to bear in mind when modelling P.D.

- Size : often

  N (♯ of individual(s)) is large

  T_i (size of individual time series) is small

  thus:$N >> T_i$ BUT this is not always the case

  ♯ of variables is large (often: multi-purpose survey)

•• Sampling : often

  individuals are selected randomly

  Time is not

  rotating panels / split panels } : individuals are partly renewed at each period

• • • non independent data

  among data relative to a same individual: because of unobservable characteristics of each individual

  among individuals : because of unobservable characteristics common to several individuals

  between time periods : because of dynamic behaviour

1.5.2 Some benefits from using P.D.

a) Controlling for individual heterogeneity

Example : state cigarette demand (Baltagi and Levin 1992)

- Unit : 46 american states
- Time period : 1963-1988
- endogenous variable : cigarette demand
- explanatory variables : lagged endogenous, price, income
- consider other explanatory variables :

Z_i :	time invariant
	religion ($\pm$ stable over time)
	education
	etc.
W_t	state invariant
	TV and radio advertising (national campaign)

Problem : many of these variables are not available

This is HETEROGENEITY (also known as "frailty")

(remember !) omitted variable $\Rightarrow$ bias (unless very specific hypotheses)

Solutions with P.D.

- dummies (specific to i and/or to t)

  WITHOUT "killing" the data

•• differences w.r.t. to i-averages
$i.e.$: $y_{it} \mapsto (y_{it} - \bar{y}_{i.})$

b) more information data sets

- larger sample size due to pooling $\left[\begin{array}{l}\text{individual}\\ \text{time}\end{array}\right.$ dimension

  In the balanced case: NT observations
  In the unbalanced case: $\sum_{1\leq i\leq N} T_i$ observations

•• more variability

  → less collinearity (as is often the case in time series)

  often : variation between units is much larger than variation within units

c) better to study the dynamics of adjustment

- distinguish

    repeated cross-sections : different individuals in different periods

    panel data : SAME individuals in different periods

•• cross-section : photograph at one period

  repeated cross-sections : different photographs at different periods

  only panel data to model HOW individuals ajust over time . This is crucial for:

    policy evaluation

    life-cycle models

    intergenerational models

d) Identification of parameters that would not be identified with pure cross-sections or pure time-series:

example 1 : does union membership increase wage ?
P.D. allows to model BOTH union membership and individual characteristics for the individuals who enter the union during the sample period.

example 2 : identifying the turn-over in the female participation to the labour market.
Notice: the female, or any other segment !

i.e. P.D. allows for more sophisticated behavioural models

e) • estimation of aggregation bias

•• often : **more precise measurements** at the micro level

1.5.3 Problems to be faced with P.D.

a) data collection problems

i.e. survey : often P.D. come from survey.

→ problems of coverage
of non response
of recall (*i.e.* incorrect remembering)
of frequency of interviewing
of interview spacing
of reference period
of bounding
of time-in-sample bias

b) measurement errors

i.e. faulty responses due to unclear questions
including : cultural differences language - translation, etc.
to memory errors
to deliberate distortion of responses
to inappropriate informante
to interviewer effects
to misrecording of responses
to coding inacurracy

c) selectivity problem

basic idea : the mere presence of a data in a sample might be informative
i.e. problem of missing data.

More specifically :

• self-selectivity

•• non response
item non response
unit non response

• • • attrition
i.e. non response in subsequent waves

d) short time-series dimension
often, BUT not always

Chapter 2

One-Way Component Regression Model

2.1 Introduction

2.1.1 The data

Let us consider data on an endogenous variable y and K exogenous variables z, with the following form:

$$y_{it},\ z_{itk} \quad i = 1, \cdots N,\ \ t = 1, \cdots T,\ \ k = 1, \cdots K,$$

where: N is the number of individuals, indexed by i, T is the number of time periods, indexed by t, and K is the number of explanatory variables, indexed by k. Thus, in the balanced case, we have NT observations on $(K+1)$ variables; the data may be re-arranged into:

$$x'_{it} = (y_{it},\ z'_{it})$$

where z_{it} is a $K-$dimensional vector.

It is convenient to first stack the data into individual data sets as follows:

$$\begin{aligned} y'_i &= (y_{i1}, y_{i2},\ \cdots,\ y_{it}, \cdots,\ y_{iT}) \\ Z_i &= [z_{i1},\ z_{i2},\ \cdots,\ z_{it}, \cdots,\ z_{iT}]' \qquad : T \times K \\ X_i &= [y_i \quad Z_i] \qquad : T \times (1+K) \end{aligned}$$

The data matrix is accordingly structured as follows:

$$X = [y \;\; Z] \quad : NT \times (1+K)$$

where:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_N \end{bmatrix} \qquad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \cdots \\ Z_N \end{bmatrix}$$

Note that we have arranged the y_{it}'s into an $NT-$column vector as follows:

$$y = (y_{11}, y_{12}, \cdots, y_{1T}, y_{21}, \cdots, y_{2T}, \cdots y_{it}, \cdots y_{NT})'$$

Thus, i, the individual index, is the "slow moving" one whereas t, the time index, is the "fast moving" one. In terms of the usual rule in matrix calculus, *viz.* first index for row and second index for column, the notation y_{it} may be interpreted as a row-vectorization (see Section 6.9) of an $N \times T$ matrix Y:

$$Y = [y_1, y_2, \ldots, y_N]' \quad : N \times T \qquad\qquad y = [Y']^v$$

2.1.2 A basic model

As a starting point, let us consider a linear regression model taking into account that each of the N individuals is endowed with a specific *non-observable*, or not made explicit, characteric, say μ_i. It is accordingly natural to specify a regression equation of the following type:

$$y_{it} = \alpha + z_{it}'\beta + \mu_i + \varepsilon_{it} \tag{2.1}$$

or, equivalently:

$$y = \iota_{NT}\alpha + Z\,\beta + Z_\mu \mu + \varepsilon \tag{2.2}$$

where ι_{NT} is an $NT-$ dimensional vector, the components of which are 1 , Z_μ is an $NT \times N$ matrix of N individual dummy variables:

$$\begin{aligned} Z_\mu &= I_{(N)} \otimes \iota_T \\ &= \begin{bmatrix} \iota_T & 0 & \cdots & 0 \\ 0 & \iota_T & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \iota_T \end{bmatrix} \end{aligned} \tag{2.3}$$

and μ is an N-dimensional vector : $\mu = (\mu_1, \mu_2, \cdots, \mu_N)'$ of individual effects. Equation (2.2) may also be written:

$$y = Z_* \delta + Z_\mu \mu + \varepsilon \tag{2.4}$$

where:

$$Z_* = [\iota_{NT} \;\; Z] \qquad \delta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Model (2.2), or (2.4), is referred to as a *one-way component* model because it makes explicit *one* factor characterizing the data, namely, in this case, the "individual" one. A similar model would be obtained by replacing the individual characteristic μ_i by a time-dependent characteristic λ_t. This chapter concentrates the attention on the individual characteristic, leaving as an exercise developping the formal analogy of a purely time-dependent characteristic.

An essential feature of the individual effect μ_i, incorporated in model (2.2), or (2.4), is to be "time-invariant"; this means that it represents the global effect, on the endogenous variable y, of all the characteristics of the individual that are not changing over the period of observation. Thus μ_i embodies the effect of truely time-invariant characteristics, such as gender, place of birth etc but also the impact of pre-sample behaviour or unchanged behaviour during the period of observation. This fact suggests that specification (2.2) or (2.4) is not complete: a particular care will be necessary to introduce specific hypotheses regarding the behaviour of this individual effect with respect to the explanatory variables Z and to the residual term ε. Note also that the randomness of Z_μ is basically linked to the selection process of the individuals pertaining to the panel, i.e. to the design of the experiment. These two features, time-invariance of the individual effects and selection of the individuals, are at the root of the many subtleties involved in the analysis of panel data. The next sections of this chapter propose a first run of contributions to the solution of those problems.

Exercise. Consider some panel data you have in mind , or the panel data briefly sketched in the first chapter, and discuss carefully the meaning of the individual effects and the particular attention one should pay for completing the modelling. ■

2.1.3 A Remark on the randomness of Z_μ: a finite population approach

Let us look more closely on the design of a survey producing an actual panel data and take the approach of a survey from a finite population.

We start from a real population of statistical units identified by a label, or identifier, l_i,

$$\mathcal{N} = \{l_i \; : i \in I\}$$

where I is a finite set. In most, but NOT in all cases, the identifiers contain no relevant information; in such a case, the population is as well represented as:

$$\mathcal{N} = \{i \; : i \in I\}$$

where I is now a finite subset of the natural numbers: $I \subset I\!N$). A random sample S , say of size N, may be viewed as a random subset of that population:

$$S = \{s_{l_1}, \; s_{l_2}, \; \cdots , \; s_{l_N}\} \subset \mathcal{N}$$

for a sampling without replacement , or as a point $S \in \mathcal{N}^N$ for a sampling with replacement. Again, once the identifiers are non-informative, the selected members of the sample may be arbitrarily re-labelled as:

$$S = \{1, 2, \cdots , N\}$$

Thus, when we speak of the randomness of Z_μ we have in mind the randomness of the sampling design: the "i" of the individual in the panel is an arbitrary label (*i.e.* "i " instead of "l_i ") for a randomly selected individual. The structure $Z_\mu = I_{(N)} \otimes \iota_T$ makes precisely explicit the fact that in a balanced panel the *same* individuals have been selected T times. As a consequence, we shall *never* consider that Z_μ is a matrix made of non-random elements; this makes meaningful the discussion about the stochastic independence, or the lack of, between Z and Z_μ, or between ε and Z_μ. This is indeed crucial when, later on, missing data, selection bias or attrition will be discussed.

Remark.
In some contexts, it may be useful to represent the sample result by a selection matrix rather than by a set of labels. More specifically, let N^* be the size

of the population: $|\,\mathcal{N}\,| = N^*$ and define an $N \times N^*$ selection matrix as a matrix the rows of which are rows of the identity matrix I_{N^*} , say e_j; thus

$$S = [e_{l_1}, e_{l_2}, \cdots, e_{l_N}]'$$

is a matrix representation of the sample $(s_{l_1},\ s_{l_2},\ \cdots,\ s_{l_N})$. In this approach, the structure of the Z_μ matrix suggests that the sample of the N individuals in the panel corresponds to a selection matrix $S = [I_{(N)}\ \ 0]$ where 0 is an $N \times (N^* - N)$ matrix and, therefore, that the selected individuals are the first N ones of an (arbitrarily labelled) finite population.

Let now $\xi = (\xi_{l_1}, \xi_{l_2}, \cdots, \xi_{l_{N^*}})'$ be an N^*-vector representing the values of a given characteristic for each member of the population and let $Y = (Y_{l_1}, Y_{l_2}, \cdots, Y_{l_N})$ be an N- random vector of the same characteristic measured for each member of the random sample S. A useful aspect of the selection matrix lies in the identity: $Y = S\,\xi$

Remark. It is crucial to distinguish the *selection* of the individuals, represented by the matrix Z_μ, and the *effect* of the selected individuals on the distribution of the endogenous variable y, represented by the vector μ.

2.2 The Fixed Effect Model

2.2.1 The model

The conceptually simplest approach for completing model (2.2), or (2.4), is to consider the unobservable individual effects μ_i as unknown parameters, to be estimated along with the usual hypotheses about the residuals: centered, spherical and uncorrelated with the explanatory variables. The hypotheses of the Fixed Effect Model, along with the parameter space, may therefore be written as follows:

$$\begin{aligned} y &= Z_*\,\delta + Z_\mu \mu + \varepsilon && (2.5)\\ \varepsilon &\sim (0,\ \sigma^2_\varepsilon\, I_{(NT)}) \qquad \varepsilon \perp (Z,\ Z_\mu) && (2.6)\\ \theta_{FE} &= (\alpha,\ \beta',\ \mu',\ \sigma^2_\varepsilon)' = (\delta',\ \mu',\ \sigma^2_\varepsilon)' && (2.7)\\ &\in \Theta_{FE} = \mathbb{R}^{1+K+N} \times \mathbb{R}_+ && (2.8) \end{aligned}$$

As such, this is a standard linear regression model, describing the first two moments of a process generating the distribution of $(y \mid Z,\, Z_\mu,\, \theta_{FE})$:

$$\begin{aligned} \mathbb{E}\,[y \mid Z,\, Z_\mu,\, \theta_{FE}) &= Z_*\,\delta + Z_\mu \mu && (2.9) \\ V(y \mid Z,\, Z_\mu,\, \theta_{FE}) &= \sigma^2_\varepsilon\, I_{(NT)} && (2.10) \end{aligned}$$

Three particular features should however call our attention.

Firstly, one should control the plausibility of the hypotheses. As far as the design of the panel is concerned, there is no requirement that the selected individuals should be perfectly representative of some underlying population: Z_μ , i.e. the labels, or identifiers, of the selected individuals, could be non independent of Z and could even represent some biased, or purposive, sampling *provided* the selection process operates independently of the "unexplained" component of the model, namely the individual disturbances ε_{it}, *i.e.* that $Z_*\,\delta + Z_\mu \mu$ truly represents the conditional expectation (2.9).

Secondly, the vector of parameters μ is often a nuisance parameter of large dimension. This makes the blind use of a standard regression packages inefficient and possibly impossible once N may be of the order of magnitude of 10,000, or more. Note also a problem a strict multicollinearity: the sum of the columns of Z_μ is exactly equal to ι_{NT}, the first column of Z_*. Fortunately enough, many specialised packages handle those problems in an efficient way; it is therefore fitting to understand how they operate.

Thirdly, even if μ is not a nuisance parameter, it is still an incidental parameter: when a new individual is added to the sample, a new parameter μ_i is added to θ_{FE}. This is evidently crucial for the asymptotic properties of the sampling model and makes still more useful to separate the estimation of the structural parameters $(\,\delta,\, \sigma^2_\varepsilon)$ from the estimation of the incidental parameter μ. This feature motivates the structure of the exposition in the next sections.

2.2.2 Estimation of the regression coefficients

The multicollinearity problem suggests to rewrite equation (2.4) as follows:

$$y \,=\, Z\,\beta \,+\, Z^*_\mu \begin{bmatrix} \alpha \\ \mu \end{bmatrix} + \,\varepsilon \qquad (2.11)$$

where:

$$Z^*_\mu = [\iota_{NT} \;\; Z_\mu]$$

Clearly, the column spaces generated by Z_μ and by Z^*_μ are the same. Let us introduce the projectors onto that space and onto its orthogonal complement:

$$P_\mu = Z_\mu(Z'_\mu Z_\mu)^{-1}Z'_\mu \qquad M_\mu = I_{(NT)} - P_\mu \tag{2.12}$$

Using the decomposition of multiple regressions (Appendix B) provides an easy way for constructing the OLS estimator of β, namely:

$$\hat{\beta}_{OLS} = (Z'M_\mu Z)^{-1}\, Z'M_\mu\, y \tag{2.13}$$

$$V(\hat{\beta}_{OLS} \mid Z,\, Z_\mu,\, \theta_{FE}) = \sigma^2_\varepsilon\, (Z'M_\mu Z)^{-1} \tag{2.14}$$

Thus, provided M_μ has been obtained, the evaluation of (2.13) requires the inversion of a $K \times K$ matrix only. Fortunately enough, the inversion of $Z'_\mu Z_\mu$, an $N \times N$ matrix, may be described analytically, and therefore interpreted, avoiding eventually the numerical inversion of a possibly large-dimensional matrix. Indeed, the following properties of Z_μ are easily checked (*i.e. Exercise* : Check them !).
Defining:

$$J_T = \iota_T\, \iota'_T \qquad \bar{J}_T = \frac{1}{T}\, J_T$$

one may check:

- $Z_\mu Z'_\mu = I_{(N)} \otimes J_T \quad : NT \times NT$
- $Z'_\mu Z_\mu = T.I_{(N)} \quad : N \times N$
- $P_\mu = I_{(N)} \otimes \bar{J}_T \quad : NT \times NT$
- $M_\mu = I_{(N)} \otimes (I_T - \bar{J}_T) = I_{(N)} \otimes E_T \quad : NT \times NT$

where:

$$E_T = I_T - \bar{J}_T$$

Furthermore:

$$r(J_T) = r(\bar{J}_T) = 1 \qquad r(E_T) = T - 1 \tag{2.15}$$

$$r(P_\mu) = N \qquad r(M_\mu) = N(T-1) \tag{2.16}$$

Notice the action of these projections:

$$P_\mu\, y = [\bar{J}_T\, y_i] = [\iota_T\, \bar{y}_{i.}] \tag{2.17}$$

$$M_\mu\, y = [E_T\, y_i] = [y_{it} - \bar{y}_{i.}] \tag{2.18}$$

where

$$\bar{y}_{i.} = \frac{1}{T} \sum_{1 \leq t \leq T} y_{it}$$

Thus, the projection P_μ transforms the individual data y_{it} into individual averages (for each time t, therefore, repeated T times within the NT-vector $P_\mu y$) whereas the projection M_μ transforms the individual data y_{it} into deviations from the individual averages; and similarly for each column of Z. Thus, the OLS estimator of β in (2.13) amounts to an ordinary regression of y on Z , after transforming all observations into deviations from individual means.

Exercise. Reconsider the estimator $\hat{\beta}_{OLS}$ in (2.13) and check that it may be viewed equivalently as:
- the part relative to the OLS estimation of β in the regression of y on Z, Z_μ;
- the OLS estimation of the transformed regression of $M_\mu y$ on $M_\mu (Z, Z_\mu)$, in which case there is one observation (more precisely, one degree of freedom) lost per individual; comment this feature and discuss the estimation (or, identification) of α and of μ in that transformed model;
- the GLS estimation of the transformed regression of $M_\mu y$ on $M_\mu (Z, Z_\mu)$, taking into account that the residuals of the transformed model have a covariance matrix equal to $\sigma_\varepsilon^2 M_\mu$. ■

Remark. The OLS estimator (2.13) is also known as the "*Least Squares Dummy Variable*" (LSDV) estimator in reference to the presence of the dummy variables Z_μ in the FE model (2.5) . It is also called the "*Within estimator*", eventually denoted as $\hat{\beta}_W$ because it is based, in view of 2.18, on the deviation from the individual averages, *i.e.* on the data transformed "within" the individuals.

2.2.3 Estimation of the individual effects

Remember that a (strict) multicollinearity problem is also an identification problem. Identifying restrictions aim accordingly not only to "solve" the multicollinearity problem but also to endow with interpretation otherwise meaningless parameters.

In the present case, model (2.2), or (2.4), introduces N parallel hyperplanes: one for each individual. Thus, N parameters are sufficient for identifying each of these hyperplanes, namely $\alpha + \mu_i$: one of the $N+1$ parameters in $(\alpha \;\; \mu')$ is redundant. The easiest form of identifying restrictions is a linear homogenous restriction:

$$c' \begin{bmatrix} \alpha \\ \mu \end{bmatrix} = 0 \qquad \text{with } c \text{ such that } c \in I\!R^{N+1} \qquad \text{and } r \begin{bmatrix} Z^*_\mu \\ c' \end{bmatrix} = N+1 \tag{2.19}$$

Thus, infinitely many specifications of c may be used, but 3 of them are practically useful.

- $\alpha = 0$
  In this case, equation (2.5) is simplified into :

$$y = Z\beta + Z_\mu \mu + \varepsilon \tag{2.20}$$

  and the μ_i's are the ordinates at the origin of each individual hyperplane and eventually measure the "individual effects". In this case, the μ_i's are estimated as :

$$\hat{\mu} = (Z'_\mu M_Z Z_\mu)^{-1} Z'_\mu M_Z y \tag{2.21}$$

$$\hat{\mu}_i = \bar{y}_{i.} - \bar{z}'_{i.} \hat{\beta}_{OLS} \qquad 1 \leq i \leq N \tag{2.22}$$

  where

$$M_Z = I_{(NT)} - Z Z^+ \qquad \bar{z}_{i.} = \frac{1}{T} \sum_{1 \leq t \leq T} z_{it} \quad : K \times 1$$

- $\sum_{1 \leq i \leq N} \mu_i = 0$

  Now, α is an "average" ordinate at the origin and the μ_i's are the differences between that "average" ordinate at the origin and the ordinates at the origin of each individual hyperplanes. Thus, α represents an "average" effect and the μ_i's measure eventually the "individual effects" in deviation form. The coefficients α and μ_i are accordingly estimated as follows:

$$\hat{\alpha} = \bar{y}_{..} - \bar{z}'_{..} \hat{\beta}_{OLS} \tag{2.23}$$

$$\hat{\mu}_i = \bar{y}_{i.} - \hat{\alpha} - \bar{z}'_{i.} \hat{\beta}_{OLS} \tag{2.24}$$

$$= (\bar{y}_{i.} - \bar{y}_{..}) - (\bar{z}_{i.} - \bar{z}_{..})' \hat{\beta}_{OLS} \qquad 1 \leq i \leq N \tag{2.25}$$

  where

$$\bar{z}_{..} = \frac{1}{NT} \sum_{1 \leq i \leq N} \sum_{1 \leq t \leq T} z_{it} \quad : K \times 1$$

- $\mu_N = 0$

  This restriction corresponds to eliminate the individual parameter corresponding to , say, the last one. Here, α is the ordinate at the origin of the hyperplane corresponding to the eliminated individual, used as a reference, and the μ_i 's are the differences between that reference ordinate at the origin and the ordinates at the origin of each remaining individual hyperplanes. Thus, α represents a "reference" effect and the μ_i 's measure eventually the "individual effects" in deviation from that reference. The coefficients α and μ_i are accordingly estimated as follows:

$$\hat{\alpha} = \bar{y}_{N.} - \bar{z}'_{N.} b_{OLS} \qquad (2.26)$$

$$\begin{aligned} \hat{\mu}_i &= \bar{y}_{i.} - \hat{\alpha} - \bar{z}'_{i.} \beta_{OLS} \\ &= (\bar{y}_{i.} - \bar{y}_{N.}) - (\bar{z}_{i.} - \bar{z}_{N.})' \beta_{OLS} \qquad 1 \leq i \leq N-1 \end{aligned} \qquad (2.27)$$

Each of these identifying restriction is potentially useful: context will suggest which one is the most appropriate, but the third one, namely $\mu_N = 0$, is particularly useful when introducing more than one fixed effect.

2.2.4 Testing for individual specificities

Once it is reckognized that individuals display some specific characteristics, it is natural to ask how far are the individuals different and how statistically significant are these differences. As in usual testing problems, this section makes use of an implicit hypothesis of normality.

Model (2.1), or (2.5), may be considered within a sequence of nested models, corresponding to different levels of specificity of the individuals. Using the identifying restriction $\sum_{1 \leq i \leq N} \mu_i = 0$, one may consider:

- M_1 : a completely unreduced model: each individual has different value for the regression parameter, namely:

$$y_{it} = \alpha + z'_{it} \beta_i + \mu_i + \varepsilon_{it} \qquad (2.28)$$

  Note that this model requires $T > K + 1$ for being meaningful.

- M_2 : FE model, namely: $\beta_i = \beta$ for all i

$$y_{it} = \alpha + z'_{it}\beta + \mu_i + \varepsilon_{it} \tag{2.29}$$

- M_3 : completely identical individuals : $\beta_i = \beta$ and $\mu_i = 0$ for all i (perfect poolability)

$$y_{it} = \alpha + z'_{it}\beta + \varepsilon_{it} \tag{2.30}$$

These nested models are obtained as linear restrictions in linear models. Using the procedure reminded in Appendix B, define the residual sums of squares, relatively to each null and alternative hypotheses, $S_i^2 \quad i = 0, 1$, and l_i as the corresponding numbers of degrees of freedom. Under the null hypothesis, the statistic:

$$F = \frac{(S_0^2 - S_1^2)/(l_0 - l_1)}{S_1^2/l_1}, \tag{2.31}$$

is distributed as an F-distribution with $(l_0 - l_1, l_1)$ degrees of freedom. Table 2.1 gives the relevant degrees of freedom for the different tests of interest.

Table 2.1: *Tests of Poolability in the One-Way FE Model*

Null Hypothesis	Alternative Hypothesis	d.f. under H_0 l_0	d.f. under H_1 l_1	$l_0 - l_1$
M_2	M_1	$N(T-1) - K$	$N(T-(K+1))$	$K(N-1)$
M_3	M_2	$NT - K - 1$	$N(T-1) - K$	$N - 1$
M_3	M_1	$NT - K - 1$	$N(T-(K+1))$	$(N-1)(K+1)$

2.2.5 Estimation of the residual variance: A Remark on the d.f. in the FE model

A natural way for evaluating the residual sum of squares in the FE model (2.1), or (2.5), is to run the "within" regression; this is the regression on the data transformed into deviations from individual averages, *i.e.* an OLS of $M_\mu y$ on $M_\mu Z$ (without a constant term because all data are taken in deviation form). When this transformation is done prior to enter the (transformed) data into a standard regression program, one should pay attention that the program will typically consider " $NT - K$ " degrees of freedom for the residual sum of squares (NT observations - K regression coefficients) whereas it should be "$N(T-1)-K$ " because the matrix of residual variances-covariances of the transformed model, namely M_μ, has rank $N(T-1)$ rather than NT. The unbiased estimator of σ^2_ε should be accordingly corrected.

2.2.6 Asymptotic Properties in the FE model

Let us distinguish the increasing behaviour of N and of T .

(i) If N is fixed and $T \rightarrow \infty$, the estimators of β and of $\alpha+\mu_i$ are consistent.

(ii) If T is fixed and $N \rightarrow \infty$, the estimator of β is still consistent but the estimators of $\alpha + \mu_i$ are not any more consistent because of the incidental nature of μ_i .

2.2.7 Time-invariant explanatory variables and modelling the individual effects

Introducing a different individual effect μ_i for each individual may be viewed as an extreme case where each individual is deemed to be characterised by a different constant term *and* where the source of that heterogeneity is not amenable to any explanation, one possible reason being that μ_i actually aggregates, into one term, the effect of too many time-invariant (or, individual specific) explanatory variables. A less extreme modelling strategy could consider to "explain" the vector μ by a set of individual-specific variables. Let us now have a closer look to such a strategy.

For the sake of illustration, let us specify that in equation (2.20), the matrix of explanatory variables Z is partitionned into two pieces: $Z = [Z_1, Z_2]$, (Z_i is now $NT \times K_i, K_1 + K_2 = K$) where Z_2 represents the observations of K_2 time-invariant variables. Thus Z_2 may be written as :

$$Z_2 = \bar{Z}_2 \otimes \iota_T \qquad \text{with } \bar{Z}_2 : N \times K_2$$

i.e. $\bar{Z}_2$ gives the individual specific values of K_2 explanatory variables of the N individuals. Such a specification might be interpreted as a first trial to introduce also time-invariant explanatory variables, along with the individual effect μ_i's, or as an exploration of the effect of introducing inadvertendly such variables (a not unrealistic issue when dealing with large-scale data bases!)

Exercise. Check that: $\bar{Z}_2 \otimes \iota_T = Z_\mu \bar{Z}_2$ ■

Thus, equation (2.20) now becomes:

$$y = Z_1 \beta_1 + Z_\mu \bar{Z}_2 \beta_2 + Z_\mu \mu + \varepsilon \tag{2.32}$$

Because $\mathcal{C}(Z_\mu \bar{Z}_2) \subset \mathcal{C}(Z_\mu)$, we note that:

$$r(Z_\mu \bar{Z}_2 \;\; Z_\mu) = r(Z_\mu) = N \qquad \text{and} \qquad \begin{bmatrix} \beta_2 \\ \mu \end{bmatrix} \in I\!R^{K_2+N}$$

Thus a blind OLS estimation of (2.32) would face a strict multicollinearity problem, eventually requiring K_2 identifying restrictions, typically in the form:

$$[R_1 \;\; R_2] \begin{bmatrix} \beta_2 \\ \mu \end{bmatrix} = 0 \quad R_1 : K_2 \times K_2 \;\; R_2 : K_2 \times N$$

such that:

$$r \begin{bmatrix} Z_\mu \bar{Z}_2 & Z_\mu \\ R_1 & R_2 \end{bmatrix} = N + K_2$$

Let us have a closer look on different possiblilities.

- $K_2 = N$
  In this case, $\bar{Z}_2$ is $N \times N$ and the restriction $\beta_2 = 0$ is equivalent to $\mu = 0$ and is just-identifying. Equivalently, these restrictions may take the form:

  $$\mu = \bar{Z}_2 \beta_2 \qquad \beta_2 = \bar{Z}_2^{-1} \mu$$

  that amounts to "explain" the individual effects μ by the explanatory variables $\bar{Z}_2$.

- $K_2 < N$
  In this case, the restriction $\mu = 0$ is overidentifying at the order $N-K_2$. Indeed, under that restriction, the "within estimator", obtained by an $\text{OLS}(M_\mu y : M_\mu Z_1)$ would give an unbiased but inefficient estimator of β_1 because the transformation M_μ "eats up" N degrees of freedom but eliminates only $K_2(< N)$ parameters, a loss of efficiency corresponding to a loss of $N - K_2$ degrees of freedom. In other terms, the restriction $\mu = 0$ leads to an $\text{OLS}(y : Z_1, Z_\mu \bar{Z}_2)$ and this regression may be viewed as restricting the regression (2.20) through

$$\mu = \bar{Z}_2\beta_2 \qquad \beta_2 = (\bar{Z}_2'\bar{Z}_2)^{-1}\bar{Z}_2'\mu$$

  i.e. $N-K_2$ restrictions on μ leaving K_2 degrees of freedom for μ. Note that the restriction $\beta_2 = 0$, in (2.32) , would be just-identifying but means giving up the idea of "explaining" the individual effects μ.

- $K_2 > N$
  In this case, the restriction $\mu = 0$ is not sufficient to identify because $r(Z_\mu\bar{Z}_2) \leq \min(N, K_2)$; $K_2 - N$ more restrictions are accordingly needed. Notice however that the "within estimator" for β_1 is still BLUE even though β_2 is not identified.

Exercise. Extend this analysis to the case where some variables are time-invariant for *some* individuals but not for all. ■

2.3 The Random Effect Model

2.3.1 Introduction

Suppose that $T = 2$ and $N = 10,000$. In such a situation, the $10,000$ μ_i are likely to be considered not any more as parameters of interest. Furthermore, the FE effect model would be a model with $10,000 + K + 1$ parameters with 20,000 observations: not a palatable situation from a statistical point of view. It is often more natural to consider the individual, but unobservable, μ_i 's as random drawings from a population of interest. Thus the idea of the Random Effect model is to consider the individual effects as latent random variables and to formally incorporate them into the residual term of a linear model. In the econometric literature, this is also known as a case in "non-observable heterogeneity".

2.3.2 The model

The basic model (2.1) is now written as:

$$y_{it} = \alpha + z'_{it}\beta + \upsilon_{it} \qquad \text{with } \upsilon_{it} = \mu_i + \varepsilon_{it} \tag{2.33}$$

or, equivalently:

$$y = \iota_{NT}\alpha + Z\beta + \upsilon \qquad \text{with } \upsilon = Z_\mu \mu + \varepsilon \tag{2.34}$$

Such a model will be made reasonably easy to estimate under the following assumptions.

$$\mu_i \sim Ind.\mathcal{N}(0,\ \sigma^2_\mu) \tag{2.35}$$

$$\varepsilon_{it} \sim Ind.\mathcal{N}(0,\ \sigma^2_\varepsilon) \tag{2.36}$$

$$Z \perp\!\!\!\perp Z_\mu \perp\!\!\!\perp \mu \perp\!\!\!\perp \varepsilon \tag{2.37}$$

Thus the parameters are now:

$$\theta_{RE} = (\alpha,\ \beta',\ \sigma^2_\mu,\ \sigma^2_\varepsilon)' \in \Theta_{RE} = \mathbb{R}^{1+K} \times \mathbb{R}^2_+ \tag{2.38}$$

Therefore the RE effect model is a linear regression model with non-spherical residuals, describing the first two moments of a process generating the distribution of $(y \mid Z,\ Z_\mu,\ \theta_{RE})$ as follows:

$$\mathbb{E}\,[y \mid Z,\ Z_\mu,\ \theta_{RE}) = \iota_{NT}\alpha + Z\beta \tag{2.39}$$

$$V(y \mid Z,\ Z_\mu,\ \theta_{RE}) = \Omega \tag{2.40}$$

where

$$\begin{aligned}\Omega &= \mathbb{E}\,(\upsilon\,\upsilon' \mid Z,\, Z_\mu,\, \theta_{RE}) \;=\; V(Z_\mu\,\mu + \varepsilon \mid Z,\, Z_\mu,\, \theta_{RE}) \\ &= \sigma_\mu^2\, Z_\mu\, Z_\mu' \;+\; \sigma_\varepsilon^2\, I_{(NT)} \;=\; I_{(N)} \,\otimes\, [\sigma_\mu^2\, J_T \,+\, \sigma_\varepsilon^2\, I_{(T)}] \end{aligned} \tag{2.41}$$

This covariance matrix displays a structure of a block-diagonal matrix where the blocks are characterized by homoscedastic and equi-correlated residuals:

$$\begin{aligned} cov(\upsilon_{it}, \upsilon_{js}) &= \sigma_\mu^2 + \sigma_\varepsilon^2 && i = j\;,\; t = s \\ &= \sigma_\mu^2 && i = j\;,\; t \neq s \\ &= 0 && i \neq j \end{aligned} \tag{2.42}$$

In view of (2.42), the RE model is also called a "Variance components model".

Exercise. Check that $\sigma_\mu^2 = 0$ does *not* imply that the conditional distribution of $(y \mid Z,\, Z_\mu)$ be a degenerate one (*i.e.* of zero variance). ■

2.3.3 Estimation of the regression coefficients

Similarly to (2.4), let us re-write (2.34) as follows:

$$y \;=\; Z_*\,\delta \;+\; \upsilon \tag{2.43}$$

where:

$$Z_* \;=\; [\iota_{NT}\;\; Z] \qquad \delta \;=\; \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Under non-spherical residuals, the natural- and BLU - estimator of the regression coefficients is obtained by the so-called Generalized Least Squares:

$$\hat{\delta}_{GLS} \;=\; (Z_*'\,\Omega^{-1}\,Z_*)^{-1} Z_*'\,\Omega^{-1}\,y \tag{2.44}$$

Given the size of Ω, namely $NT \times NT$, it is useful, both for computational purposes and for easing the interpretation, to make use of its spectral decomposition (*Hint*: review Section 7.5 of Complement B):

$$\Omega \;=\; \sum_{1 \le i \le p} \lambda_i Q_i$$

Under (2.41), we obtain that $p = 2$, and therefore $Q_1 \,+\, Q_2 \;=\; I_{(NT)}$. The spectral decomposition of the residual covariance matrix is characterised in Table 2.2.

Table 2.2: *Spectral decomposition for the One-way-RE model*

i	Q_i	λ_i	mult(λ_i)=$r(Q_i)$	$Q_i y$
1	$I_{(N)} \otimes E_T$	σ_ε^2	$N(T-1)$	$[y_{it} - \bar{y}_{i.}]$
2	$I_{(N)} \otimes \bar{J}_T$	$\sigma_\varepsilon^2 + T\sigma_\mu^2$	N	$[\bar{y}_{i.}] \otimes \iota_T$
Sum	$I_N \otimes I_T$		NT	

Exercise. Check and compare the results with (2.15) to (2.18). ■

Thus Q_1 and Q_2 in the RE model are respectively the same as M_μ and P_μ in the FE model.

Let us now use the results of Section 7.5 of Complement B (chap.7) and note first that

$$Q_1 \, \iota_{NT} = 0 \qquad Q_2 \, \iota_{NT} = \iota_{NT}$$

Thus, when, in view of (7.47), we combine the estimations of the two OLS of $Q_i \, y$ on $Q_i \, Z_*$ for i=1, 2 , the estimation of α is non-zero in the second one only which is the "between" regression among the individual averages:

$$\bar{y}_{i.} = \alpha + \bar{z}_{i.}' \, \beta + \bar{v}_{i.} \tag{2.45}$$

Therefore

$$\hat{\alpha}_{GLS} = \bar{y}_{..} - \bar{z}_{..}' \, \hat{\beta}_{GLS} \tag{2.46}$$

and the estimation of β in the model transformed by Q_2 is denoted $\hat{\beta}_B$ ("B" is for "between"). The OLS of $Q_1 \, y$ on $Q_1 \, Z$ is the regression "within" without a constant term, given that $Q_1 \, \iota_{NT} = 0$; this is the regression on the deviations from the individual averages, *i.e.* of $[y_{it} - \bar{y}_{i.}]$ on $[z_{it} - \bar{z}_{i.}]$. This estimation of β is denoted $\hat{\beta}_W$ ("W" is for "within"). As a conclusion, we complete (2.46) with:

$$\hat{\beta}_{GLS} = W_1 \, \hat{\beta}_W + W_2 \, \hat{\beta}_B \tag{2.47}$$

with:

$$W_1 = [\sigma_\varepsilon^{-2} Z'Q_1Z + \lambda_2^{-1} Z'Q_2Z]^{-1} \sigma_\varepsilon^{-2} Z'Q_1Z$$
$$W_2 = [\sigma_\varepsilon^{-2} Z'Q_1Z + \lambda_2^{-1} Z'Q_2Z]^{-1} \lambda_2^{-1} Z'Q_2Z$$

Note the slight difference with the particular case at the end of Section 6 in Complement B. There, we had $Q_2 = \bar{J}_n$ and $(p-1)$ terms in the summation, whereas here we have $Q_2 = I_{(N)} \otimes \bar{J}_T$ and $p = 2$ terms in the summation.

Both $\hat{\beta}_B$ and $\hat{\beta}_W$, and therefore $\hat{\beta}_{GLS}$, are unbiased and consistent estimator of β once $N \to \infty$.

Remark. Alternatively, one can use directly the relationship:

$$[V(\varepsilon)]^r = \Omega^r = \sum \lambda_i^r Q_i = (\sigma_\varepsilon^2)^r Q_1 + (\sigma_\varepsilon^2 + T\sigma_\mu^2)^r Q_2$$

and apply to (2.44) with $r = -1$. ■

2.3.4 Estimation of the Variance components

Introduction

In general, the estimator (2.47) is not feasible because it depends on the unknown characteristic values λ_i. A natural way to come around that problem is to plug in suitable estimators, say $\hat{\lambda}_i$. Moreover, the variances themselves, σ_ε^2 and σ_μ^2, may also be parameter of interest. For instance, σ_μ^2 is a natural measure of the heterogeneity among the individuals, and eventually a basis for testing the significance of that heterogeneity. One way of estimating the variances σ_ε^2 and σ_μ^2 is to first estimate the characteristic values λ_1 and λ_2 and thereafter deduce an estimator for the variances.

Moment estimation through the eigenvalues

The estimation of the characteristic values may follow the general procedure reminded in Section 7.5 (Complement B), taking advantage of the fact that $p = 2$. For the estimation of λ_1, we use $y'R_2\,y$, the sum of square residuals of the "Within" regression, without constant term because $M_2 = Q_1$ and $Q_1\,\iota_{NT} = 0$, *i.e.* :

$$R_2 = [I_{(N)} \otimes E_T] - [I_{(N)} \otimes E_T]Z(Z'[I_{(N)} \otimes E_T]Z)^{-1}Z'[I_{(N)} \otimes E_T] \quad (2.48)$$

Given that $r_1 = N(T-1)$, we obtain a natural unbiased estimator of λ_1 ,

$$\hat{\lambda}_1 = \frac{y' R_2\, y}{N(T-1) - K} \tag{2.49}$$

Similarly for λ_2 , we use $y' R_1\, y$, where, using (7.58) and reminding that $M_1 = Q_2 = P_\mu$,

$$R_1 = [I_{(N)} \otimes \bar{J}_T] - [I_{(N)} \otimes \bar{J}_T] Z_* (Z'_* [I_{(N)} \otimes \bar{J}_T] Z_*)^{-1} Z'_* [I_{(N)} \otimes \bar{J}_T] \tag{2.50}$$

Thus, $y' R_1\, y$ is the sum of square residuals of the "Between" regression with constant. Given that $r_2 = N$, we obtain an unbiased (actually, Best Quadratic Unbiased) estimator of λ_1 ,

$$\hat{\lambda}_2 = \frac{y' R_1\, y}{N - K - 1} \tag{2.51}$$

From these estimations of the characteristic values we may solve from Table 2.2 to obtain unbiased estimators of the variances:

$$\hat{\sigma}^2_\varepsilon = \hat{\lambda}_1 = \frac{y' R_2\, y}{N(T-1) - K} \tag{2.52}$$

$$\hat{\sigma}^2_\mu = \frac{\hat{\lambda}_2 - \hat{\lambda}_1}{T} \tag{2.53}$$

Note that for some samples, $\hat{\sigma}^2_\mu$ might be negative. A simple solution to this problem is to replace, in such cases, that estimation by zero, accepting eventually a biased estimator. It may be shown that the probability that (2.53) produces a negative estimate is substantial only when σ^2_μ is close to zero, in which case an estimate equal to zero is reasonable.

Remark. Formally, the regression on the model transformed by Q_2 is the "Between" regression (2.45) on the individual averages, *but* with each individual average observation repeated T times, indeed $Q_2\, y = [\iota_T\, \bar{y}_{i.}]$. Notice however that the "Between" regression (2.45), without repetition of the individual averages, may be obtained as follows. Define:

$$Q^*_2 = I_{(N)} \otimes \frac{1}{\sqrt{T}}\, \iota'_T \qquad : N \times NT$$

and notice that:

$$Q^*_2\, y = [I_{(N)} \otimes \frac{1}{\sqrt{T}}\, \iota'_T]\, y = \sqrt{T} \begin{bmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \\ \cdots \\ \bar{y}_{N.} \end{bmatrix} = \sqrt{T}\, [\bar{y}_{i.}]$$

$$Q_2^* Q_2^{*'} \; = \; I_N \qquad Q_2^{*'} Q_2^* \; = \; Q_2 \qquad Q_2^* \, Q_2 \, Q_2^{*'} \; = \; I_{(N)}$$
$$V(\, Q_2^* \, \varepsilon \,) \; = \; \lambda_2 \, I_{(N)}$$

Thus, the regression of $Q_2^* \, y$ on $Q_2^* \, Z$ boils down to the regression (2.45) with each observation multiplied by $\sqrt{T}$; this regression has spherical residuals with residual variance equal to λ_2. Therefore, $y' R_2 y$ is equal to the sum of square residuals of the "Between" regression (2.45), without repetition of the individual averages, multiplied by the factor T.

Direct moment estimation

One way to avoid a negative estimate for σ_μ^2 is to use a direct estimate as follows

$$\hat{\hat{\sigma}}_\mu^2 \; = \; \frac{1}{N-1} \sum_{1 \leq i \leq N} \hat{\mu}_i^2 \tag{2.54}$$

where $\hat{\mu}_i$ is the FE estimate of μ_i under the identifying restriction $\sum_{1 \leq i \leq N} \hat{\mu}_i \; = \; 0$.

2.3.5 Estimation in the RE model under a Normality Assumption

Up to now, the assumption of normality has not been used. The attention has been focused on moment estimators. Once μ and ε are normally and independently distributed, υ is also normally distributed:

$$\upsilon \; \sim \; \mathcal{N}(0, \, \Omega) \tag{2.55}$$

From (7.18) in Section 7.5.1 , we know that $V(Q_i \upsilon) \; = \; \lambda_i Q_i$. Therefore,

$$\frac{\upsilon' Q_i \upsilon}{\lambda_i} \; \sim \; \chi^2_{(r_i)} \tag{2.56}$$

Moreover, $V(\frac{\upsilon' Q_i \upsilon}{\lambda_i}) \; = \; 2 \, r_i$ and $V(\frac{\upsilon' Q_i \upsilon}{r_i}) \; = \; \frac{2\lambda_i^2}{r_i}$. Thus, once $N \; \rightarrow \; \infty$,:

$$\begin{pmatrix} \sqrt{NT} \, [\, \frac{\upsilon' Q_1 \upsilon}{r_1} \; - \; \lambda_1 \,] \\ \\ \sqrt{N} \, [\, \frac{\upsilon' Q_2 \upsilon}{r_2} \; - \; \lambda_2 \,] \end{pmatrix} \longrightarrow \mathcal{N} \left(0, \, \begin{bmatrix} 2 \, \lambda_1^2 & 0 \\ \\ 0 & 2 \, \lambda_2^2 \end{bmatrix} \right) \tag{2.57}$$

(approximating $r_1 \; = \; N(T-1)$ by NT).

Remark

Remember that :

$$Q_1 \upsilon = [\upsilon_{it} - \bar{\upsilon}_{i.}] \qquad Q_2 \upsilon = [\bar{\upsilon}_{i.}\ \iota_T]$$

Therefore:

$$\upsilon' Q_1 \upsilon = \sum_i \sum_j (\upsilon_{it} - \bar{\upsilon}_{i.})^2 \qquad \upsilon' Q_2 \upsilon = T \sum_i \bar{\upsilon}_i^2$$

$$\hat{\lambda}_1 = \frac{\sum_i \sum_j (\upsilon_{it} - \bar{\upsilon}_{i.})^2}{N(T-1)} \qquad \hat{\lambda}_2 = \frac{T \sum_i \bar{\upsilon}_i^2}{N} = T\, (\bar{\bar{\upsilon}}^2)$$

Let us now turn to the maximum likelihood estimation in the RE model under the normality assumption. The data density relative to (2.43) is

$$\begin{aligned} p(y \mid Z,\, Z_\mu, \theta_{RE}) \;=\; & (2\,\pi)^{-\frac{1}{2}NT} \mid \Omega \mid^{-\frac{1}{2}} \\ & \exp -\tfrac{1}{2}\, (y - Z_* \,\delta)'\, \Omega^{-1}\, (y - Z_*\, \delta) \end{aligned} \tag{2.58}$$

Two issues are of interest: the particular struture of Ω, and of its spectral decomposition, and the fact that in frequent cases N is very large (and much larger than T).

2.3.6 Testing problems

2.3.7 Predictions

Predicting future observations in the RE effect model is a case of prediction with non-spherical residuals, a case reminded in section 7.4. Thus we have a sample generated through the model described in (2.33) to (2.41) and we want to predict future observations, $y_{i,T+S} \quad i = 1, \cdots N; S \geq 1$, relatively to given values of the exogenous variables, $z_{i,\,T+t} \quad i = 1, \cdots N; S \geq 1$, with $z_{i,\,T+S} \in \mathbb{R}^k$, and generated under the same sampling conditions, in particular:

$$\upsilon_{i,\,T+S} = \mu_i + \varepsilon_{i,\,T+S}$$

Therefore

$$\begin{aligned} cov(\upsilon_{i,T+S},\, \upsilon_{i',t}) &= 0 \qquad i \neq i' \\ &= \sigma_\mu^2 \qquad i = i' \qquad (2.59) \\ cov(\upsilon_{i,T+S},\, \upsilon) &= \sigma_\mu^2 (e_{i,N} \otimes \iota_T) \qquad (2.60) \\ cov(\upsilon_{i,T+S}, u')V(u)^{-1} &= \sigma_\mu^2 (e'_{i,N} \otimes \iota'_T)[\frac{1}{\lambda_1} Q_1 \,+\, \frac{1}{\lambda_2} Q_2] \\ &= \frac{\sigma_\mu^2}{T\,\sigma_\mu^2 \,+\, \sigma_\varepsilon^2}\,(e'_{i,N} \otimes \iota'_T) \qquad (2.61) \end{aligned}$$

(where $e_{i,N}$ denotes the $i-$th column of the identity matrix I_N)

Exercise Check it!

Therefore:

$$\begin{aligned} \mathbb{E}\,(\upsilon_{i,T+(S)} \mid \upsilon \,=\, \hat{\upsilon}) &= \frac{\sigma_\mu^2}{T\,\sigma_\mu^2 \,+\, \sigma_\varepsilon^2}\,(e'_{i,N} \otimes \iota'_T)\hat{\upsilon} \\ &= \frac{T\,\sigma_\mu^2}{T\,\sigma_\mu^2 \,+\, \sigma_\varepsilon^2}\,\overline{\hat{\upsilon}}_{i.} \qquad (2.62) \end{aligned}$$

We then obtain the BLU predictor:

$$\hat{y}_{i,T+S} \,=\, \hat{\alpha} \,+\, z'_{i,T+S}\hat{\beta}_{GLS} \,+\, \frac{T\,\sigma_\mu^2}{T\,\sigma_\mu^2 \,+\, \sigma_\varepsilon^2}\,\overline{\hat{\upsilon}}_{i.} \qquad (2.63)$$

2.4 Comparing the Fixed Effect and the Random Effect Models

2.4.1 Comparing the hypotheses of the two Models

The RE model and the FE model *may* be viewed within a hierarchical specification of a unique encompassing model. From this point of view, the two models are not fundamentally different, they rather correspond to different levels of analysis within a unique hierarchical framework. More specifically, from a Bayesian point of view, where all the variables (latent or manifest) and parameters are jointly endowed with a (unique) probability measure, one may consider the complete specification of the law of $(y,\, \mu,\, \theta \;\mid\; Z,\, Z_\mu)$ as follows:

$$\begin{aligned}(y \mid \mu,\, \theta,\, Z,\, Z_\mu) &\sim \mathcal{N}(\, Z_* \,\delta + Z_\mu \mu,\, \sigma^2_\varepsilon\, I_{(NT)}) \qquad (2.64)\\ (\mu \mid \theta,\, Z,\, Z_\mu) &\sim \mathcal{N}(0,\, \sigma^2_\mu\, I_{(N)}) \qquad (2.65)\\ (\theta \mid Z,\, Z_\mu) &\sim Q \qquad (2.66)\end{aligned}$$

where Q is an arbitrary prior probability on $\theta = (\delta,\, \sigma^2_\varepsilon,\, \sigma^2_\mu)$. Parenthetically, note that this complete specification assumes:

$$y \perp\!\!\!\perp \sigma^2_\mu \mid \mu,\, \delta,\, \sigma^2_\varepsilon,\, Z,\, Z_\mu \qquad \mu \perp\!\!\!\perp (\theta,\, Z,\, Z_\mu) \mid \sigma^2_\mu$$

The above specification implies:

$$(y \mid \theta,\, Z,\, Z_\mu) \sim \mathcal{N}(\, Z_* \,\delta\, ,\, \sigma^2_\mu\, Z_\mu\, Z'_\mu \,+\, \sigma^2_\varepsilon\, I_{(NT)}) \qquad (2.67)$$

Thus the FE model, *i.e.* (2.64), considers the distribution of $(y \mid \mu,\, \theta,\, Z,\, Z_\mu)$ as the sampling distribution and the distributions of $(\mu \mid \theta,\, Z,\, Z_\mu)$ and $(\theta \mid Z,\, Z_\mu)$ as prior specification. The RE model, *i.e.* (2.67), considers the distribution of $(y \mid \theta,\, Z,\, Z_\mu)$ as the sampling distribution and the distribution of $(\theta \mid Z,\, Z_\mu)$ as prior specification. Said differently, in the RE model, μ is treated as a latent (*i.e.* not obervable) variable whereas in the FE model μ is treated as an incidental parameter. Moreover, the RE model is obtained from the FE model through a marginalization with respect to μ.

These remarks make clear that the FE model and the RE model should be expected to display different sampling properties. Also, the inference on μ is an estimation problem in the FE model whereas it is a prediction problem in the RE model: the difference between these two problems regards the difference in the relevant sampling properties, *i.e.* w.r.t. the distribution of $(y \mid \mu,\, \theta,\, Z,\, Z_\mu)$ or of $(y \mid \theta,\, Z,\, Z_\mu)$, and eventually of the relevant risk functions, *i.e.* the sampling expectation of a loss due to an error between an estimated value and a (fixed) parameter or between a predicted value and the realization of a (latent) random variable.

This fact does however not imply that both levels might be used indifferently. Indeed, from a sampling point of view:

(i) the dimensions of the parameter spaces are drastically different. In the FE model, when N, the number of individuals, increases, the μ_i 's being incidental parameters also increases in number: each new individual introduces a new parameter. This makes a crucial difference for the asymptotic properties of the two models.

(ii) the conditions of validity for the design of the panel are also drastically different. The (simple version of the) RE model requires Z, Z_μ, μ and ε to be mutually independent . This means that the individuals, coded in Z_μ , should have been selected by a representative sampling of a well defined population of reference, in such a way that the corresponding μ_i's are an IID sample from a $N(0,\ \sigma^2_\mu)$ population, independently of the exogenous variables contained in Z . The difficult issue is however to evaluate whether the (random) *latent* individual effects μ_i are actually independent of the *observable* exogenous variables z_{it} . In contrast to these requirements, the FE model, considering the sampling *conditional* to a fixed realization of μ , does *not* require the μ_i's to be "representative", not even to be uncorrelated with Z.

Choosing between the FE and the RE models should therefore be based on the following considerations:

(i) what is the parameter of interest: is it the particular realization of the individual values μ_i or their distribution, characterized by σ^2_μ ? In particular, when N is large, the vector $\mu = (\mu_1, \mu_2, \cdots, \mu_N)'$ is not likely to be of interest and the FE model suffers from a serious problem of degrees of reedom.

(ii) are the conditions of validity of these two models equally plausible? In particular, is it plausible that μ and Z are independent? Is the sampling truly representative of a population of μ_i's of interest? If not, the hypothesis underlying the RE model are not satisfied whereas the FE model only considers that the individuals are not identical.

(iii) what are the relevant asymptotic properties: $N \rightarrow \infty$ and/or $T \rightarrow \infty$?

2.4.2 "Testing" the two Models

As such, the RE models and the FE model are not nested models, *i.e.* one does not correspond to a restriction of the other one. Comparing the properties of the estimator of β , note that , in the FE model, the "Within" estimator

$$\hat{\beta}_W = (Z'M_\mu Z)^{-1} Z'M_\mu y = (Z' Q_1 Z)^{-1} Z' Q_1 y \qquad (2.68)$$

is BLUE and consistent, if either N or T tends to infinity . In the RE model, $\hat{\beta}_W$ is still unbiased and consistent; indeed even if μ and Z are not uncorrelated, the "Within" estimator wipes out the effect of μ in the estimation of β, because $M_\mu Z_\mu = 0$. However, the GLS estimator of the RE model, (2.47), is not anymore unbiased nor consistent once μ and Z are not uncorrelated.

Thus, in the spirit of Hausman (1978) specification test, a comparaison between $\hat{\beta}_W$ and $\hat{\beta}_{GLS}$ of (2.47) may be used for thesting the hypothesis of uncorrelatedness between μ and Z which is *one* of the condition of validity of (the simple version of) the RE model. More specifically, let us consider, as a maintained hypothesis, a more general version of the RE model, namely equation (2.33), or (2.34), along with the assumptions (2.35) to (2.37), *except* that the independence between μ and Z is not any more maintained and take, as the null hypothesis to be tested, that μ and Z are uncorrelated. Thus, under this maintained hypothesis, the parameter becomes:

$$\theta_{(1)} = (\alpha,\, \beta,\, \sigma^2_\varepsilon,\, \sigma^2_\mu,\, \rho) = (\theta_{RE},\, \rho) \qquad \text{where } \rho = cov(\mu,\, z) \in I\!R^K$$

and the null hypothesis is:

$$H_0 : \quad \rho = 0$$

Now, the Within estimator, $\hat{\beta}_W$ is unbiased and consistent under the null *and* under the alternative hypothesis. Indeed, in the RE model transformed by $M_\mu = Q_1$, the residual $Q_1\, \upsilon = Q_1(Z_\mu\mu + \varepsilon) = Q_1\, \varepsilon$ is uncorrelated with Z, even under the alternative hypothesis (*provided* that the independence between Z_μ and ε is maintained!). Note however that $\hat{\beta}_{GLS}$ is consistent under the null but not under the alternative hypothesis. In a specification test approach, it is accordingly natural to analyze, under the null hypothesis, the distribution of the difference:

$$d = \hat{\beta}_W - \hat{\beta}_{GLS} = [\,(Z' Q_1 Z)^{-1} Z' Q_1 - (Z' \Omega^{-1} Z)^{-1} Z' \Omega^{-1}]y \qquad (2.69)$$

where $(Z,\, y)$ are taken in deviation from the mean in β_{GLS}. Marking the moments under the null hypothesis by a subscript "0", we have:

$$
\begin{aligned}
V_0(\hat{\beta}_W) &= \sigma_\varepsilon^2 \,(Z'\,Q_1\,Z)^{-1} && (2.70)\\
V_0(\hat{\beta}_{GLS}) &= (Z'\,\Omega^{-1}\,Z)^{-1} && (2.71)\\
cov_0(\hat{\beta}_W,\, \beta_{GLS}) &= cov((Z'\,Q_1\,Z)^{-1}\,Z'\,Q_1\varepsilon, (Z'\,\Omega^{-1}\,Z)^{-1}\,Z'\,\Omega^{-1}[Z_\mu\mu\,+\,\varepsilon]) &&\\
&= (Z'\,\Omega^{-1}\,Z)^{-1} && (2.72)
\end{aligned}
$$

(taking into account that $Q_1\,\Omega^{-1} \,=\, \sigma_\varepsilon^{-2}Q_1$). Therefore:

$$V_0(d) \,=\, \sigma_\varepsilon^2\,(Z'\,Q_1\,Z)^{-1} \,-\, (Z'\,\Omega^{-1}\,Z)^{-1} \qquad (2.73)$$

A Hausman test statistic is accordingly given by:

$$H \,=\, d'\,[\hat{V}_0(d)]^{-1}d$$

where $\hat{V}_0(d)$ is obtained by replacing, in (2.73), σ_ε^2 and Ω by the value of consistent estimators, to be used also for obtaining, in d, a feasible version of $\hat{\beta}_{GLS}$. It may then be shown that, under the null hypothesis, H is asymptotically distributed as χ^2_K.

Chapter 3

Two-Way Component Regression Model

3.1 Introduction

With the same data as in Chap.2, let us now consider a linear regression model taking into account that not only each of the N individuals is endowed with a specific *non-observable* characteric, say μ_i, but also each time-period is endowed with a specific *non-observable* characteric, say λ_t. In other words, the μ_i's are individual-specific and time-invariant whereas the λ_t's are time-specific and individual-invariant.

Similarly to the individual effect μ_i, the time effect λ_t represents the global impact of all individual-invariant factors. For instance, λ_t might represent the effect of a strike at specific periods, or the general economic environment proper to each period (macro-economic conditions) etc, the understanding being that those time characteristics equally affect all the individuals of the panel.

It is accordingly natural to specify a regression equation of the following type:

$$y_{it} = \alpha + z_{it}'\beta + \mu_i + \lambda_t + \varepsilon_{it} \tag{3.1}$$

or, equivalently:

$$y = \iota_{NT}\alpha + Z\beta + Z_\mu \mu + Z_\lambda \lambda + \varepsilon \tag{3.2}$$

where Z_μ and μ have been defined in Chap.2, Z_λ is an $NT \times T$ matrix of T

time-specific dummy variables:

$$Z_\lambda = \iota_N \otimes I_{(T)},$$

and λ is a T-dimensional vector of time-specific (or, individual-invariant) effects: $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_T)'$.

Model (3.1), or (3.2), is referred to as a *two-way components* model because it makes explicit *two* factors characterizing the data, namely the "individual" one and the "time" one.

As in Chap. 2 , one may model both individual and time effects as fixed or as random. In some cases, it will be useful to consider "mixed" models, *i.e.* one effect being fixed, the other one being random.

3.2 The Fixed Effect Model

3.2.1 The model

Similarly to the one-way model, one may specify that the unobservable effects μ_i and λ_t are unknown parameters, that assume fixed values for the data under consideration, equivalently : that the model is conditional to a specific but unknown realization of $\mu = (\mu_1, \mu_2, \cdots, \mu_N)'$ and $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_T)'$. The hypotheses of the two-way Fixed Effect Model, along with the parameter space, may accordingly be written as follows:

$$y = \iota_{NT}\alpha + Z\beta + Z_\mu \mu + Z_\lambda \lambda + \varepsilon \quad (3.3)$$

$$\varepsilon \sim (0, \sigma^2_\varepsilon I_{(NT)}) \qquad \varepsilon \perp (Z, Z_\mu, Z_\lambda) \quad (3.4)$$

$$\theta_{FE} = (\alpha, \beta', \mu', \lambda', \sigma^2_\varepsilon)' \in \Theta_{FE} = I\!R^{1+K+N+T} \times I\!R_+ \quad (3.5)$$

As such, this is again a standard linear regression model, describing the first two moments of a process generating the distribution of $(y \mid Z, Z_\mu, Z_\lambda, \theta_{FE})$:

$$I\!E\,[y \mid Z, Z_\mu, Z_\lambda, \theta_{FE}) = \iota_{NT}\alpha + Z\beta + Z_\mu \mu + Z_\lambda \lambda \quad (3.6)$$

$$V(y \mid Z, Z_\mu, Z_\lambda, \theta_{FE}) = \sigma^2_\varepsilon I_{(NT)} \quad (3.7)$$

3.2.2 Estimation of the regression coefficients

Similarly to the one-way FE model, we also have a (strict) multicollinearity problem: the $NT \times (1 + N + T)$ matrix $[\iota_{NT}\ Z_\mu\ Z_\lambda]$ has rank equal to

$N + T - 1$ (*Exercise :* Check that $Z_\mu \iota_N = Z_\lambda \iota_T = \iota_{NT}$). We therefore need two identifying restrictions. For reasons of symmetry, the often used restrictions are:

$$\sum_{1 \le i \le N} \mu_i = \sum_{1 \le t \le T} \lambda_t = 0 \tag{3.8}$$

Equally useful restrictions are:

$$\mu_N = \lambda_T = 0 \tag{3.9}$$

As for the one-way FE model, the estimation of β will be made easier, both for computational purposes and for interpretation, if we first "eliminate" $(\alpha,\ \mu,\ \lambda)$ by projecting the data on the orthogonal complement to the space generated by the matrix $[\iota_{NT}\ Z_\mu\ Z_\lambda]$. Let us denote that projector as $M_{\mu,\lambda}$. Taking advantage of the basic properties of the matrices Z_μ and Z_λ (see Appendix of this chapter), one obtains:

$$\begin{aligned}
M_{\mu,\lambda} &= I_{(NT)} - [Z_\mu,\, Z_\lambda]\,[Z_\mu,\, Z_\lambda]^+ \qquad \text{by definition} \\
&= I_{(N)} \otimes I_{(T)} - \bar{J}_N \otimes I_{(T)} - I_{(N)} \otimes \bar{J}_T + \bar{J}_N \otimes \bar{J}_T \\
&= E_N \otimes E_T \qquad \text{with : } E_N = I_{(N)} - \bar{J}_N \qquad (3.10) \\
r(M_{\mu,\lambda}) &= tr(M_{\mu,\lambda}) \\
&= tr(E_N).tr(E_T) \\
&= (N-1)(T-1) \qquad (3.11)
\end{aligned}$$

Exercise : Check it! ■

.

Exercise : Check also:

$$\begin{array}{lll}
(\bar{J}_N \otimes I_{(T)})y = \iota_N \otimes [\bar{y}_{.t}] & & (I_{(N)} \otimes \bar{J}_T)y = [\bar{y}_{i.}] \otimes \iota_T \\
(\bar{J}_N \otimes \bar{J}_T)y = \bar{J}_{NT}y = \iota_{NT}\bar{y}_{..} & & M_{\mu,\lambda}Z_\lambda = 0 \\
M_{\mu,\lambda}Z_\mu = 0 & & M_{\mu,\lambda}\iota_{NT} = 0
\end{array}$$
■

The action of the transformation $M_{\mu,\lambda}$ is therefore to "sweep" from the data the individual-specific and the time-specific effects in the following sense:

$$\begin{aligned}
M_{\mu,\lambda}\, y &= [y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..}] \qquad (3.12) \\
&= M_{\mu,\lambda}\, Z\beta + M_{\mu,\lambda}\, \varepsilon \qquad (3.13)
\end{aligned}$$

said differently: $M_{\mu,\lambda}\, y$ produces the residuals from an ANOVA-2 (without interaction, because there is only one observation per cell (i,t)). Using the decomposition of multiple regressions (Appendix B) provides again an easy way for constructing the OLS estimator of β, namely:

$$\hat{\beta}_{OLS} = (Z' M_{\mu,\lambda} Z)^{-1} Z' M_{\mu,\lambda}\, y \tag{3.14}$$

$$V(\hat{\beta}_{OLS} \mid Z,\, Z_\mu,\, Z_\lambda,\, \theta_{FE}) = \sigma^2_\varepsilon\, (Z' M_{\mu,\lambda} Z)^{-1} \tag{3.15}$$

Notice that the meaning of (3.12) is that the estimator (3.14) corresponds to a regression of the vector y on K vectors of residuals from an ANOVA-2 (two fixed factors) on each of the columns of matrix Z.

3.2.3 Estimation of the specific effects

Under the restrictions (3.8), the parameters (α, μ, λ) are estimated as follows:

$$\hat{\alpha} = \bar{y}_{..} - \hat{\beta}'_{OLS}\bar{z}_{..} \tag{3.16}$$

$$\hat{\mu}_i = (\bar{y}_{i.} - \bar{y}_{..}) - \hat{\beta}'_{OLS}(\bar{z}_{i.} - \bar{z}_{..}) \tag{3.17}$$

$$\hat{\lambda}_t = (\bar{y}_{.t} - \bar{y}_{..}) - \hat{\beta}'_{OLS}(\bar{z}_{.t} - \bar{z}_{..}) \tag{3.18}$$

Exercise : Adjust these estimators for the case of restrictions (3.9). ■

3.2.4 Testing the specificities

The general method is the same as for the one-way fixed effect model (Chapter 2). The main difference regards the models, or hypotheses, of possible interest. The two-way model offers indeed a much wider bunch of interests. Using the identifying restrictions (3.8), one may consider the following models. A tempting starting point would be to consider:

M_1 : Completely unreduced model: each combination of individual and time period has different parameters, namely:

$$y_{it} = \alpha_{it} + z'_{it}\beta_{it} + \varepsilon_{it} \tag{3.19}$$

Note however that this model requires repeated observations for *each* pair (i,t). In the completely balance case, one would have R repeated observations for *each* pair (i,t), thus RNT observations. The residual Sum of Squares of the OLS regressions on (3.19) would have $NT(R-K-1)$ degrees of freedom, requiring accordingly that $R > K+1$ for the model to be meaningful. In practice, such a situation is not frequent (in econometrics!) and will not be pursued here.

A more natural starting point is to consider purely additive models, *i.e.* models without parameters of interaction. We shall accordingly focus the attention on the following models, which form a nested sequence selected from a larger number of possible simplifications of model M_1:

- M_2 : Unreduced additive model: each individual and time period has different specific parameters, namely:

$$y_{it} = \alpha + z'_{it}\beta_i + z'_{it}\gamma_t + \mu_i + \lambda_t + \varepsilon_{it} \tag{3.20}$$

  Model M_2 may also be written as:

$$y_{it} = \alpha + z'_{it}[\beta_i + \gamma_t] + \mu_i + \lambda_t + \varepsilon_{it}$$

  and corresponds therefore to the restriction $\beta_{it} = \beta_i + \gamma_t$ in Model M_1 .

- M_3 : No time-specific regression effect: $\gamma_t = 0$

$$y_{it} = \alpha + z'_{it}\beta_i + \mu_i + \lambda_t + \varepsilon_{it} \tag{3.21}$$

- M_4 : No time-effect at all: $\gamma_t = 0$ and $\lambda_t = 0$

$$y_{it} = \alpha + z'_{it}\beta_i + \mu_i + \varepsilon_{it} \tag{3.22}$$

- M_{4bis} : Standard two-way FE model, *i.e.* no time-specific nor individual-specific regression effect: $\beta_i = \beta$ and $\gamma_t = 0$

$$y_{it} = \alpha + z'_{it}\beta + \mu_i + \lambda_t + \varepsilon_{it} \tag{3.23}$$

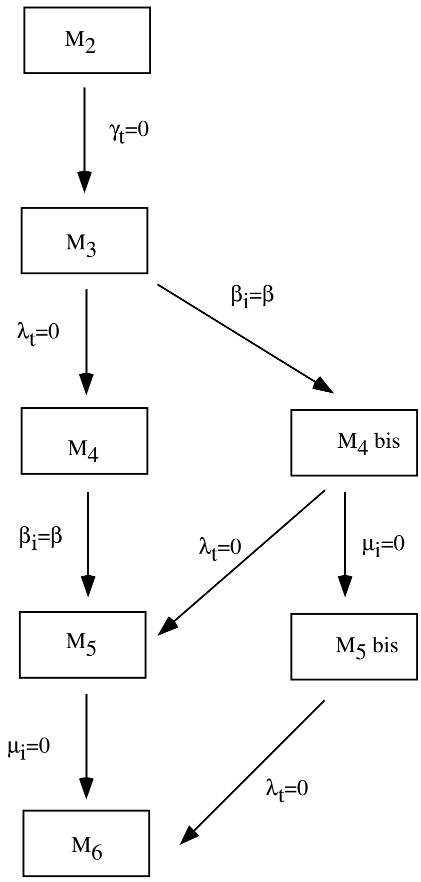
- M_5 : One-way individual-FE model, *i.e.* no time-effect and identical individual regression effect: $\beta_i = \beta$, $\gamma_t = 0$ and $\lambda_t = 0$:

$$y_{it} = \alpha + z'_{it}\beta + \mu_i + \varepsilon_{it} \tag{3.24}$$

- M_{5btis} One-way time-FE model, *i.e.* no individual- effect and identical time regression effect: $\beta_i = \beta$, $\gamma_t = 0$ and $\mu_i = 0$:

$$y_{it} = \alpha + z'_{it}\beta + \lambda_t + \varepsilon_{it} \tag{3.25}$$

Figure 3.1: *Nesting properties of Models in a Two-way FE framework*



- M_6 : No time nor individual effect : $\beta_i = \beta$, $\gamma_t = 0$ and $\mu_i = \lambda_t = 0$ for all i, t (perfect poolability)

$$y_{it} = \alpha + z_{it}'\beta + \varepsilon_{it} \tag{3.26}$$

Figure (3.1) illustrates the nesting properties of these models. Note that, for being meaningful, model M_2 requires $\min\{N, T\} > K+1$ and models M_3 and M_4 require $T > K + 1$.

The testing procedure is the same as in Chapter 2 and makes use of the same F-statistic (2.31) with the degrees of freedom evaluated in Table (3.1), which is a simple extension of Table 2.1.

Table 3.1: *Tests of Poolability in the Two-Way FE Model*

H_0	H_1	d.f. under H_0 l_0	d.f. under H_1 l_1	$l_0 - l_1$
M_3	M_2	$N(T-K-1)-T+1$	$NT-(K+1)(N+T)+1$	TK
M_4	M_3	$N(T-K-1)$	$N(T-K-1)-T+1$	$T-1$
M_5	M_4	$N(T-1)-K$	$N(T-K-1)$	$K(N-1)$
M_6	M_5	$NT-K-1$	$N(T-1)-K$	$N-1$
M_6	M_2	$NT-K-1$	$NT-(K+1)(N+T)+1$	$(N+T-1)(K+1)-1$
M_{4bis}	M_3	$(T-1)(N-1)-K$	$N(T-K-1)-T+1$	$K(N-1)$
M_5	M_{4bis}	$N(T-1)-K$	$(T-1)(N-1)-K$	$T-1$
M_6	M_{4bis}	$NT-K-1$	$(T-1)(N-1)-K$	$N+T-2$
M_{5bis}	M_{4bis}	$NT-K-T$	$(T-1)(N-1)-K$	$N-1$
M_6	M_{5bis}	$NT-K-1$	$NT-K-T$	$T-1$

3.2.5 Estimation of the residual variance: A Remark on the d.f. in the FE model

Similarly to the remark made in section 2.2.5, the "within regression", *i.e.* the OLS($M_{\mu,\lambda}\, y \,:\, M_{\mu,\lambda}\, Z$) (without constant term) gives the same residuals, and therefore the same sum of squares, as the OLS($y \,:\, \iota_{NT}$, Z, Z_μ, Z_λ). Remember, however, that when the data ($M_{\mu,\lambda}\, y$, $M_{\mu,\lambda}\, Z$) are introduced directly into a standard regression package, the program, when evaluating the unbiased estimator of the residual variance, assumes wrongly $NT - K$ degrees of freedom whereas it should be $(N - 1)(T - 1) - K$; in other terms, the value of the computed estimate should be corrected by a factor of $(NT - K)[(N - 1)(T - 1) - K]^{-1}$.

3.3 The Random Effect Model

3.3.1 Introduction

Here, the randomness of the time-effect should receive a particular attention, given that it should be independent of the exogenous variables included in Z. Thus, λ_t might represent the impact of climatic variability *provided* it is not associated with any variables included in Z. One should accordingly be very cautious before imputing to λ_t the impact of the general economic environment, unless *all* the variables included in Z where independent of that economic environment. This is a typical difficulty when dealing with non-observable heterogeneity.

3.3.2 The model

$$y_{it} = \alpha + z_{it}'\beta + \upsilon_{it} \qquad \upsilon_{it} = \mu_i + \lambda_t + \varepsilon_{it} \tag{3.27}$$

or, equivalently:

$$y = \iota_{NT}\alpha + Z\beta + \upsilon \qquad \upsilon = Z_\mu \mu + Z_\lambda \lambda + \varepsilon \tag{3.28}$$

$$\varepsilon \sim (0,\ \sigma_\varepsilon^2\ I_{(NT)}) \tag{3.29}$$

$$\mu \sim (0,\ \sigma_\mu^2\ I_{(N)}) \tag{3.30}$$

$$\lambda \sim (0,\ \sigma_\lambda^2\ I_{(T)}) \tag{3.31}$$

$$\mu \perp\!\!\!\perp \lambda \perp\!\!\!\perp \varepsilon \perp\!\!\!\perp Z \perp\!\!\!\perp (Z_\mu\,,\ Z_\lambda) \tag{3.32}$$

$$\theta_{RE} = (\alpha,\ \beta',\ \sigma_\varepsilon^2,\ \sigma_\mu^2,\ \sigma_\lambda^2\ ) \in \Theta_{RE} = I\!R^{1+K} \times I\!R_+^3 \tag{3.33}$$

Exactly as in the one-way RE model, this is a linear regression model with non-spherical residuals, describing the first two moments of a process generating the distribution of $(y \mid Z,\ Z_\lambda,\ Z_\mu,\ \theta_{RE})$ as follows:

$$I\!E\,[y \mid Z,\ Z_\mu,\ Z_\lambda,\ \theta_{RE}) = \iota_{NT}\alpha + Z\beta \tag{3.34}$$

$$V(y \mid Z,\ Z_\mu,\ Z_\lambda,\ \theta_{RE}) = \Omega \tag{3.35}$$

where

$$\begin{aligned}\Omega &= \mathbb{E}\,(\upsilon\,\upsilon' \mid Z,\, Z_\mu,\, Z_\lambda,\, \theta_{RE}) \\ &= \sigma^2_\mu\, Z_\mu\, Z'_\mu \;+\; \sigma^2_\lambda\, Z_\lambda\, Z'_\lambda \;+\; \sigma^2_\varepsilon\, I_{(NT)} \\ &= \sigma^2_\varepsilon\,(I_{(N)} \,\otimes\, I_{(T)}) \,+\, \sigma^2_\mu\,(I_{(N)} \,\otimes\, J_T) \,+\, \sigma^2_\lambda\,(J_N \,\otimes\, I_{(T)}) \qquad (3.36)\end{aligned}$$

Exercise
Check that the structure of the variances and covariances can also be described as follows:

$$\begin{aligned}cov(\upsilon_{it}, \upsilon_{js}) &= cov(\mu_i \,+\, \lambda_t \,+\varepsilon_{it},\; \mu_j \,+\, \lambda_s \,+\varepsilon_{js}) && (3.37)\\ &= \sigma^2_\mu \,+\sigma^2_\lambda \,+\, \sigma^2_\varepsilon \quad i=j\;,\; t=s && (3.38)\\ &= \sigma^2_\mu \quad i=j\;,\; t\neq s && (3.39)\\ &= \sigma^2_\lambda \quad i\neq j\;,\; t=s && (3.40)\\ &= 0 \quad i\neq j\;,\; t\neq s && (3.41)\end{aligned}$$

3.3.3 Estimation of the regression coefficients

Exactly as in the one-way RE model, let us re-write (3.34) as :

$$y \,=\, Z_*\,\delta \,+\, \upsilon \qquad (3.42)$$

where:

$$Z_* \,=\, [\iota_{NT}\;\; Z] \qquad \delta \,=\, \begin{bmatrix}\alpha\\ \beta\end{bmatrix}$$

The Generalized Least Squares estimator of the regression coefficients is again written as :

$$\hat\delta_{GLS} \,=\, (Z'_*\,\Omega^{-1}\,Z_*)^{-1} Z'_*\,\Omega^{-1}\,y \qquad (3.43)$$

Given the size of Ω, namely $NT \times NT$, it is again useful, both for computational purposes and for easing the interpretation, to make use of its spectral decomposition :

$$\Omega \,=\, \sum_{1\le i\le p} \lambda_i Q_i$$

Under (3.36), we obtain that $p = 4$. Table 3.2 gives the projectors Q_i, the corresponding eigenvalues λ_i, their respective multiplicities along with the transformations they operate on $\mathbb{R}^{NT}$.
As exposed in Section 7.5., the GLS estimator may be obtained by performing an OLS regression on the model (3.28) transformed successively by each projector Q_i of Table 3.2; more explicitly:

Table 3.2: *Spectral decomposition for the Two-way-RE model*

i	Q_i	λ_i	mult(λ_i)=$r(Q_i)$	$Q_i y$
1	$E_N \otimes E_T$	σ_ε^2	$(N-1)(T-1)$	$[y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..}]$
2	$E_N \otimes \bar{J}_T$	$\sigma_\varepsilon^2 + T\sigma_\mu^2$	$N-1$	$[\bar{y}_{i.} - \bar{y}_{..}] \otimes \iota_T$
3	$\bar{J}_N \otimes E_T$	$\sigma_\varepsilon^2 + N\sigma_\lambda^2$	$T-1$	$\iota_N \otimes [\bar{y}_{.t} - \bar{y}_{..}]$
4	$\bar{J}_N \otimes \bar{J}_T$	$\sigma_\varepsilon^2 + T\sigma_\mu^2 + N\sigma_\lambda^2$	1	$\bar{y}_{..}\iota_{NT}$
Sum	$I_N \otimes I_T$		NT	

$Q_1 y$ corresponds to a "within regression" (see FE: b_W)
$Q_2 y$ corresponds to a "between individuals" regression("b_{BI}")
$Q_3 y$ corresponds to a "between time periods" regression("b_{BT}")
$Q_4 y$ corresponds to an adjustment to the means (see below)

Let us indeed consider the fourth regression. Notice that:

$$Q_4\,\iota_{NT} = \iota_{NT} \qquad Q_4\,Z_\mu = 0 \quad Q_4\,Z_\lambda = 0$$

Therefore, the transformed model $Q_4 y = Q_4\,Z_*\,\delta + Q_4\,\upsilon$, has one degree of freedom and may be written as:

$$\iota_{NT}\,\bar{y}_{..} = \iota_{NT}\,\alpha + \iota_{NT}\,\bar{z}'_{..}\,\beta + \iota_{NT}\,\bar{\varepsilon}_{..}$$

Thus:

$$\hat{\alpha}_{(4)} = \bar{y}_{..} - \bar{z}'_{..}\,\hat{\beta}_{(4)}, \quad \hat{\beta}_{(4)} \text{ arbitrary}, \qquad \hat{\bar{\varepsilon}}_{..(4)} = 0$$

and the other three transformed models $Q_i\,y = Q_i\,Z\,\beta + Q_i\,\upsilon$ $(i = 1, 2, 3)$ have no constant terms (*i.e.* $\hat{\alpha}_{(i)} = 0$) because $Q_i\,\iota_{NT} = 0$ $(i = 1, 2, 3)$.

Therefore, using the results of Section 7.5., we obtain:

$$\hat{\alpha}_{GLS} = \hat{\alpha}_{(4)} = \bar{y}_{..} - \bar{z}'_{..}\hat{\beta}_{(GLS)} \tag{3.44}$$

$$\hat{\beta}_{GLS} = \sum_{1\leq i\leq 3} W_i^* b_i \tag{3.45}$$

where:

$$W_i^* = [\sum_{1\leq j\leq 3} \lambda_j^{-1} Z'Q_jZ]^{-1}\lambda_i^{-1}Z'Q_iZ \tag{3.46}$$

and where b_i the OLS estimator of the regression of Q_iy on Q_iZ:

$$\begin{aligned} b_i &= [Z'Q_iZ]^{-1}\,Z'Q_iy \qquad \text{when } r_i \geq K & (3.47)\\ &= \text{any solution of } [Z'Q_iZ]\; b_i = Z'Q_iy \quad \text{, in general} & (3.48) \end{aligned}$$

Therefore, (3.45) and (3.46) may also be written as:

$$\hat{\beta}_{GLS} = [\sum_{1\leq j\leq 3} \lambda_j^{-1}Z'Q_jZ]^{-1}][\sum_{1\leq i\leq 3} \lambda_i^{-1}Z'Q_iy] \tag{3.49}$$

Thus, if $T < K$, b_3 is arbitrary but $Z'Q_3Z\,b_3$ is well defined given that $Z'Q_3Z\,b_3 = Z'Q_3y$ and this is the only value entering the construction of $\hat{\beta}_{GLS}$.

Alternative derivation

From Table 3.2 one may write:

$$\begin{aligned} \Omega^{-\frac{1}{2}}\,y &= \sum_j \lambda_j^{-\frac{1}{2}}\,Q_j\,y\\ &= [\lambda_1^{-\frac{1}{2}}\,y_{it} + (\lambda_2^{-\frac{1}{2}} - \lambda_1^{-\frac{1}{2}})\bar{y}_{i.} + (\lambda_3^{-\frac{1}{2}} - \lambda_1^{-\frac{1}{2}})\bar{y}_{.t}\\ &\qquad + (\lambda_1^{-\frac{1}{2}} - \lambda_2^{-\frac{1}{2}} - \lambda_3^{-\frac{1}{2}} + \lambda_4^{-\frac{1}{2}})\bar{y}_{..}]\\ \sigma_\varepsilon\,\Omega^{-\frac{1}{2}}\,y &= [y_{it} - \theta_1\bar{y}_{i.} - \theta_2\bar{y}_{.t} + \theta_3\bar{y}_{..}] & (3.50) \end{aligned}$$

where:

$$\begin{aligned} \theta_1 &= 1 - \sigma_\varepsilon\,\lambda_2^{-\frac{1}{2}}\\ \theta_2 &= 1 - \sigma_\varepsilon\,\lambda_3^{-\frac{1}{2}}\\ \theta_3 &= \theta_1 + \theta_2 + \sigma_\varepsilon\,\lambda_4^{-\frac{1}{2}} - 1 & (3.51) \end{aligned}$$

We therefore may obtain the GLS estimator $\hat{\delta}_{GLS}$ in equation (3.43) through an OLS of $y^* = \sigma_\varepsilon\,\Omega^{-\frac{1}{2}}\,y$ on $Z_*^* = \sigma_\varepsilon\,\Omega^{-\frac{1}{2}}\,Z_*$.

Remarks

(i) If $\sigma_\mu^2 \to 0$ *and* $\sigma_\lambda^2 \to 0$, then, from Table 3.2, $\lambda_i \to \sigma_\varepsilon^2 \qquad i = 2, 3, 4$ and therefore $\hat{\delta}_{GLS} \to \hat{\delta}_{OLS}$.

(ii) If $N \to \infty$ *and* $T \to \infty$, then $\lambda_i^{-1} \to 0 \qquad i = 2, 3, 4$ and therefore, from (3.46), $\hat{\delta}_{GLS} \to \hat{\delta}_{OLS}$.

(iii) When $T\,\sigma_\mu^2\,\sigma_\varepsilon^{-2} \to \infty$, *i.e.* $T\,\sigma_\mu^2 >> \sigma_\varepsilon^2$, then $\beta_{GLS} \to \beta_{BI}$. Similarly, when $N\,\sigma_\lambda^2\,\sigma_\varepsilon^{-2} \to \infty$, *i.e.* $N\,\sigma_\lambda^2 >> \sigma_\varepsilon^2$, then $\beta_{GLS} \to \beta_{BT}$.

3.3.4 Estimation of the Variances

The estimators developped in Section 3.3.3 are feasible only if the variances, or the eigenvalues of the residual Variance matrix, are either known or replaced by suitably estimated values. Two strategies may be proposed: either estimate first the eigenvalues under the constraint that the 4 eigenvalues are functions of 3 variances only, or estimate directly the variances and deduce estimations of the eigenvalues.

Direct Moment estimation of the Variances

The estimation of the characteristic values may follow the general procedure reminded in Section 7.5 (Complement B), with $p = 4$.

Moment estimation through the eigenvalues

3.3.5 Testing problems

3.3.6 Predictions

3.4 Mixed Effect Models

3.4.1 Introduction

In not infrequent cases, particularly when N is large and T is small, it is natural to treat the time-effect as a fixed one, capturing effects possibly not taken into account by the time-dependent variables included in Z. This would accomodate for a possible association between Z and Z_λ. But the individual effect would be treated as a random one, if only because of a large number of individuals present in the panel. Such models are called "Mixed Models"; they are characterized by N latent variables, namely μ, and $T(+K)$ parameters, equivalenlty: conditional on T unknown realisations of λ_t.

3.4.2 The model

The hypotheses of the two-way Mixed Effects Model, along with the parameter space, may be written as follows:

$$y_{it} = \alpha + z'_{it}\beta + \lambda_t + \upsilon_{it} \qquad \upsilon_{it} = \mu_i + \varepsilon_{it} \tag{3.52}$$

or, equivalently:

$$y = \iota_{NT}\alpha + Z\beta + Z_\lambda\lambda + \upsilon \qquad \upsilon = Z_\mu\mu + \varepsilon \tag{3.53}$$

under the following assumptions:

$$\mu_i \sim Ind.N(0,\, \sigma^2_\mu) \tag{3.54}$$
$$\varepsilon_{it} \sim Ind.N(0,\, \sigma^2_\varepsilon) \tag{3.55}$$
$$(Z,\, Z_\lambda) \perp\!\!\!\perp Z_\mu \perp\!\!\!\perp \mu \perp\!\!\!\perp \varepsilon \tag{3.56}$$

Thus the parameters are now:

$$\theta_{ME} = (\alpha,\, \beta',\, \lambda',\, \sigma^2_\mu,\, \sigma^2_\varepsilon)' \in \Theta_{ME} = I\!R^{1+K+T} \times I\!R^2_+ \tag{3.57}$$

This is again a linear regression model with non-spherical residuals, describing the first two moments of a process generating the distribution of $(y \mid Z,\, Z_\lambda,\, Z_\mu,\, \theta_{ME})$ as follows:

$$I\!E\,[y \mid Z,\, Z_\mu,\, Z_\lambda,\, \theta_{ME}) = \iota_{NT}\alpha + Z\beta + Z_\lambda\lambda \tag{3.58}$$
$$V(y \mid Z,\, Z_\mu,\, Z_\lambda,\, \theta_{ME}) = \Omega \tag{3.59}$$

where

$$\begin{aligned}\Omega &= V(\, Z_\mu \mu \,+\, \varepsilon \mid Z,\, Z_\mu,\, Z_\lambda,\, \theta_{ME}) \\ &= \sigma^2_\mu\, Z_\mu\, Z'_\mu \,+\, \sigma^2_\varepsilon\, I_{(NT)} \\ &= \sigma^2_\varepsilon\,(I_{(N)} \otimes I_{(T)}) \,+\, \sigma^2_\mu\,(I_{(N)} \otimes J_T) \\ &= I_{(N)} \otimes [(\sigma^2_\varepsilon \,+\, T\,\sigma^2_\mu)\bar{J}_T \,+\, \sigma^2_\varepsilon E_T] \\ &= \sum_{1\le i\le 2} \lambda_i\, Q_i \qquad (3.60)\end{aligned}$$

Table 3.3: *Spectral decomposition for the Two-way Mixed Effects model*

i	Q_i	λ_i	mult(λ_i)=$r(Q_i)$ $= r_i$	$Q_i y$
1	$I_N \otimes E_T$	σ^2_ε	$N(T-1)$	$[y_{it} - \bar{y}_{i.}]$
2	$I_N \otimes \bar{J}_T$	$\sigma^2_\varepsilon + T\sigma^2_\mu$	N	$[\bar{y}_{i.}] \otimes \iota_T$
Sum	$I_N \otimes I_T$		NT	

$$tr\Omega = \sum_{1\le i\le 2} \lambda_i\, r_i \,=\, NT(\sigma^2_\varepsilon \,+\, \sigma^2_\mu) \qquad (3.61)$$

$$r(\Omega) = r_1 \,+\, r_2 \,=\, NT \qquad (3.62)$$

3.4.3 Estimation of the Regression coefficients

Notice that the structure of the residual term is the same as in the One-way RE model, as seen in Section 2.3. Therefore the spectral decomposition of the residual covariance matrix, given in Table 3.3 , is the same as that given in Table 2.2; in particular, $p \,=\, 2$, $Q_1 \,=\, M_\mu$ and $Q_2 \,=\, P_\mu$.

Two different strategies may be considered for organizing the estimation of the regression coefficients, depending on the priority one gives for the FE or for the RE and depending of the relative magnitudes of N and T.

From equation (3.53, we notice:

$$\begin{aligned} r[\iota_{(NT)}\, Z_\lambda] &= r[Z_\lambda] = T \quad \text{indeed} : Z_\lambda \iota_{(NT)} = \iota_{(T)} \qquad (3.63) \\ M_\lambda &= I_{(NT)} - Z_\lambda Z_\lambda^+ = E_N \otimes I_T \\ & \qquad\qquad r(M_\lambda) = (N-1)T \qquad (3.64) \\ M_\lambda\, y &= [y_{it} - \bar{y}_{.t}] = M_\lambda Z\beta + M_\lambda u \\ V(M_\lambda u) &= \Omega_\lambda \quad \text{, say} \\ &= M_\lambda\, \Omega\, M_\lambda \\ &= \sum_{1\leq i\leq 2} \lambda_i M_\lambda Q_i M_\lambda \qquad (3.65) \end{aligned}$$

The OLS estimator on the transformed model $M_\lambda\, y$ is not efficient; indeed: $M_\lambda\, \Omega\, M_\lambda [M_\lambda Z][M_\lambda Z]^+$ is not symetric (see Section 7.3). In order to make use of the GLS estimator, we first notice that (3.65) actually is the spectral decomposition of Ω_λ; indeed the terms $M_\lambda Q_i M_\lambda$ are projections:

$$\begin{aligned} M_\lambda Q_1 M_\lambda &= (E_N \otimes I_T)(I_N \otimes E_T)(E_N \otimes I_T) = E_N \otimes E_T \\ M_\lambda Q_2 M_\lambda &= (E_N \otimes I_T)(I_N \otimes \bar{J}_T)(E_N \otimes I_T) = E_N \otimes \bar{J}_T \end{aligned}$$

More precisely, Table 3.4 gives the complete spectral decomposition of Ω_λ. Therefore, Ω_λ is singular; more precisely:

$$r(\Omega_\lambda) = r_1 + r_2 = (N-1)\,T < NT$$

(equivalently: $r(\Omega_\lambda) = r(M_\lambda) = (N-1)\,T$) Note that $\bar{J}_N \otimes \bar{J}_T$, projecting an NT-vector into its overall average, actually projects on a space *included* in the invariant subspace corresponding to the third eigenvalue (and to the projection $\bar{J}_N \otimes I_T$).
As a consequence, the GLS estimator may be obtained as follows.

$$\begin{aligned} b_{GLS} &= \sum_{1\leq i\leq 2} W_i^*\, b_i \qquad \text{with :} \\ W_i^* &= [\sum_{1\leq j\leq 2} \lambda_j^{-1} Z' M_\lambda Q_j M_\lambda Z]^{-1} \lambda_i^{-1} Z' M_\lambda Q_i M_\lambda Z \\ b_i &: OLS(Q_i M_\lambda y \,:\, Q_i M_\lambda Z) \qquad \textit{without}\ \text{constant term}(M_\lambda \iota NT = 0) \end{aligned}$$

Table 3.4: *Spectral decomposition for the Transformed Model in the Mixed Effects model*

i	Q_i	λ_i	mult(λ_i)=$r(Q_i)$ $= r_i$	$Q_i y$
1	$E_N \otimes E_T$	σ_ε^2	$(N-1)(T-1)$	$[y_{it} - \bar{y}_{i.} - y_{.t} + y_{..}]$
2	$E_N \otimes \bar{J}_T$	$\sigma_\varepsilon^2 + T\sigma_\mu^2$	$N - 1$	$[\bar{y}_{i.} - y_{..}] \otimes \iota_T$
3	$\bar{J}_N \otimes I_T$	0	T	$\iota_N \otimes [y_{.t}]$
Sum	$I_N \otimes I_T$		NT	

More specifically:

$$\begin{array}{lcll} Q_1 M_\lambda & = & E_N \otimes E_T & Q_1 M_\lambda y = [y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..}] \\ & & & i.e.\ b_1 = \text{OLS}([y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..}] \\ & & & \quad : [z_{itk} - \bar{z}_{i.k} - \bar{z}_{.tk} + \bar{z}_{..k} : 1 \le k \le K]) \\ & & & \textit{Remark}\ r(Q_1 M_\lambda) = (N-1)(T-1); \\ Q_2 M_\lambda & = & E_N \otimes \bar{J}_T & Q_2 M_\lambda y = [\bar{y}_{i.} - \bar{y}_{..}] \otimes \iota_T \\ & & & \text{in practice: } b_2 = \text{OLS}([\bar{y}_{i.} - \bar{y}_{..}] \\ & & & \quad : [\bar{z}_{i.k} - \bar{z}_{..k} : 1 \le k \le K]) \\ & & & i.e.\ \text{"between individuals",} \\ & & & \quad \text{without repeating } T \text{ times the same regression.} \\ & & & \textit{Remark}\ r(Q_2 M_\lambda) = N - 1; \\ & & & \quad \text{thus } N \text{ observations but } N-1 \text{ d.f.} \\ & & & \quad \text{and } r(Q_2 M_\lambda Z) = N - 1 \text{ provided } K < N - 1 \end{array}$$

Alternative strategy

When T is small, one may proceed as in the One-Way RE with (Z, Z_λ) as regressors. Because of (3.63) we need the restriction $\alpha = 0$ and proceed, as in Section 2.3, but with a regression with T different constant terms.

More specifically, let us consider:

$$\begin{aligned}
y &= Z_*\delta + u \qquad Z_* = [Z,\, Z_\lambda] \\
& \qquad\qquad \delta = \begin{bmatrix} \beta \\ \lambda \end{bmatrix} \\
& \qquad\qquad u = Z_\mu + \varepsilon \\
d_{GLS} &= \sum_{1\le i\le 2} W_i^*\, b_i \qquad d_i : \mathrm{OLS}(Q_i y \,:\, Q_i Z_*) \qquad (3.66)
\end{aligned}$$

where:

$$\begin{aligned}
Q_1 Z_\lambda &= \iota_N \otimes E_T \qquad Q_1 Z_\lambda \lambda = \iota_N \otimes [\lambda_t - \lambda_.] \\
Q_2 Z_\lambda &= \iota_N \otimes \bar{J}_T \qquad Q_2 Z_\lambda \lambda = \iota_{NT}\, \bar{\lambda}_. \qquad (3.67)
\end{aligned}$$

3.4.4 Estimation of the variance components

3.4.5 Testing problems

3.5 Hierarchical Models

3.5.1 Introduction

Motivation

In Section 2.2.7, we have considered the possibility of "explaining" the individual effect μ_i by means of individual- specific variables. A particular case would be to consider a categorization of the individuals as an explanatory device.

Consider, for instance, a panel on milk production (average per cow) on farms j at time t. Suppose now that one wants to introduce the idea that the farm effect is expected to be different according to the region i. Thus, the data take the form x_{ijt} and one may considers that the regions operate a categorization of the farms in the panel. One natural question to be faced is "Is the region effect sufficient to explain the individual effects?", or equivalently "Is the individual effect still significant, after taking into account the region effect?"

Clearly, the two factors "individual" and "region" are ordered: individuals are "within" region and not the contrary; thus in x_{ijt}, i stands for a "principal" factor, say A, and j stands for a "secondary" one, say B. In other words, this means that for each level of factor A, there is a supplementary factor B. Other examples of such a situation include:

j for companies; i for industrial sector
j for individuals; i for profession
j for household; i for municipality
etc

For the ease of exposition, we shall use, in this section, the words "sector i" for the main factor and "individual j" for the secondary factor. Note that the econometric literature also uses the expression "nested effects" whereas the biometric literature rather uses "hierarchical models". As a matter of fact, this section may also be viewed as an introduction to the important topic of "multi-level analysis".

Different grounds for Hierarchical Models

Two different, although not mutually exclusive, grounds for hierarchical models should be distinguished:
(i)*Hierarchy by design.* This is a hierarchy introduced at the stage of designing a multistage sampling. This is the case, for instance, when sampling first the sectors i's and thereafter sampling, in each sampled sectors, the individuals j's.
(ii) *Hierarchy a posteriori.* This is a hierarchy at the stage of modelling and follows from the main motivation suggested above. As this hierarchy may be introduced *after* the sampling has been designed, and realized, it is qualified as " a posteriori"

Notice that these two grounds are indenpendent in the sense that neither one implies the other one. But this double aspect of hierarchisation is most important when evaluating the plausibility of the hypotheses used in the different models to be exposed below. The hierarchical aspect is rooted in the fact that (ij) represents the label of *one* individual: individual j in sector i.

The basic equation

Let us first organize the data $x_{ijt} = (y_{ijt}, z'_{ijt})'$ where y_{ijt} is an endogenous variable and z_{ijt} is a K-vector of exogenous variables. In the balanced case, we specify:

$$\begin{aligned} i &= 1, \cdots M && \text{slow moving} \\ j &= 1, \cdots N && \text{middle moving} \\ t &= 1, \cdots T && \text{fast moving} \end{aligned}$$

Thus, we have MNT observations stacked as follows:

$$\begin{aligned} y = (&y_{111}, \cdots, y_{11T}, y_{121}, \cdots, y_{12T}, \cdots, y_{1N1}, \cdots, y_{1NTt}, \\ &\cdots, y_{M11} \cdots, y_{M1T}, \cdots, y_{MN1}, \cdots, y_{MNT})' \end{aligned}$$

Assuming, for the sake of exposition, that there is no time effect, we start from the basic equation:

$$y_{ijt} = \alpha + z'_{ijt}\beta + \mu_i + \nu_{ij} + \varepsilon_{ijt} \tag{3.68}$$

or, equivalently:

$$y = \iota_{MNT}\alpha + Z\beta + Z_\mu\mu + Z_\nu\nu + \varepsilon \tag{3.69}$$

where

$$\begin{aligned} Z_\mu &= I_{(M)} \otimes \iota_{(NT)} = I_{(M)} \otimes \iota_{(N)} \otimes \iota_{(T)} \qquad : MNT \times M \\ &= \begin{bmatrix} \iota_{(NT)} & 0 & \cdots & 0 \\ 0 & \iota_{(NT)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \iota_{(NT)} \end{bmatrix} & (3.70) \\ Z_\nu &= I_{(MN)} \otimes \iota_T = I_{(M)} \otimes I_{(N)} \otimes \iota_T \qquad : MNT \times MN \\ &= \begin{bmatrix} \iota_T & 0 & \cdots & 0 \\ 0 & \iota_T & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \iota_T \end{bmatrix} & (3.71) \\ Z &: MNT \times K \qquad \mu \in I\!R^M \qquad \nu \in I\!R^{MN} & (3.72) \end{aligned}$$

3.5.2 Hierarchical Fixed Effects

The model

$$y = \iota_{MNT}\alpha + Z\beta + Z_\mu\mu + Z_\nu\nu + \varepsilon \tag{3.73}$$

$$\varepsilon \sim (0,\ \sigma_\varepsilon^2\, I_{(MNT)}) \qquad \varepsilon \perp (Z,\ Z_\mu,\ Z_\nu) \tag{3.74}$$

$$\theta_{HFE} = (\alpha,\ \beta',\ \mu',\ \nu',\ \sigma_\varepsilon^2)' \in \Theta_{HFE} = I\!R^{1+K+M+MN} \times I\!R_+ \tag{3.75}$$

As such, this is again a standard linear regression model, describing the first two moments of a process generating the distribution of $(y \mid Z,\ Z_\mu,\ Z_\nu,\ \theta_{HFE})$ as follows:

$$\mathbb{E}\,[y \mid Z,\ Z_\mu,\ Z_\nu,\ \theta_{HFE}) = \iota_{MNT}\alpha + Z\beta + Z_\mu\mu + Z_\nu\nu \tag{3.76}$$

$$V(y \mid Z,\ Z_\mu,\ Z_\nu,\ \theta_{HFE}) = \sigma_\varepsilon^2\, I_{(MNT)} \tag{3.77}$$

Estimation of the regression coefficients

Clearly, the space generated by $[\iota_{MNT},\ Z_\mu]$ is included in the space generated by $[Z_\nu]$.

Exercises.
(i) Find a matrix A such that $Z_\nu A = Z_\mu$. What is the order of A?
(ii) Equivalently, show that : $Z_\nu Z_\nu^+ Z_\mu = Z_\mu$
(iii) Repeat the same with ι_{MNT} instead of Z_μ. ■

The projection on the orthogonal complement of that space is characterised as follows:

$$M_\nu = I_{(MNT)} - Z_\nu Z_\nu^+ = I_{(MN)} \otimes E_T \tag{3.78}$$
$$r(M_\nu) = MNT - NT = NT(M - 1) \tag{3.79}$$
$$M_\nu y = [y_{ijt} - \bar{y}_{ij.}] \tag{3.80}$$

Furthermore:

$$M_\nu y = M_\nu Z \beta + M_\nu \varepsilon \tag{3.81}$$

Using the decomposition of multiple regressions (Section 7.2) provides an easy way for constructing the OLS estimator of β , namely

$$\hat{\beta}_{OLS} = (Z' M_\nu Z)^{-1} Z' M_\nu y \tag{3.82}$$
$$V(\hat{\beta}_{OLS} \mid Z, Z_\mu, Z_\nu, \theta_{FE}) = \sigma_\varepsilon^2 (Z' M_\nu Z)^{-1} \tag{3.83}$$

This is simply the OLS estimation of a regression, *without constant*, of the data (on y and Z) in deviation from the time- averages.

Remark. Again, if the data are entered directly in that form in a regression package, care should be taken for the account of degrees of freedom. Indeed, the program will correctly evaluate the total residual sum of squares but will assume $MNT - K$ degrees of freedom whereas it should $r(M_\nu) - K = MN(T-1) - K$. Thus, when the program gives $s^2 = (MNT - K)^{-1} RSS$ (RSS = Residual Sum of Squares), the unbiased estimator of σ_ε^2 is

$$s_c^2 = (MNT - K)[MN(T-1) - K]^{-1} s^2$$

■

Estimation of the individual effects

Note that the (partitionned) matrix $[\iota_{MNT}, Z_\mu, Z_\nu]$ has dimension $MNT \times (1 + M + MN)$ but rank equal to MN. We accordingly need to introduce

$M + 1$ identifying restrictions. Many packages use the following, or similar, ones:

$$\mu_M = 0 \quad \text{and} \quad \nu_{iN} = 0 \;\; i = 1, \cdots M$$

In such a case, α represents the constant term of individual MN. These restrictions have the advantage of offering more flexibility at the stage of model building, *i.e.* when the factors to be introduced in the model are still under discussions. Alternative restrictions, often used in a more standard framework of ANOVA, could be, in the balanced case,:

$$\sum_i \mu_i = 0 \qquad \sum_j \nu_{ij} = 0 \;\; i = 1, \cdots M$$

but will not be discussed here.

Exercise. Describe the estimators $\hat{\alpha}, \hat{\mu}, \hat{\nu}$

Testing problems

The general approach of Section 3.2.4 , along with Figure 3.1 and Table 3.1, is again applicable. The main issue is to make explicit different hypotheses of potential interest:

- $H_m \; : \; \mu_{ij} = \alpha + \mu_i + \nu_{ij}$
- $H_1 \; : \; \mu_i = 0 \quad i = 1, \cdots M$
- $H_2 \; : \; \nu_{ij} = 0 \quad i = 1, \cdots M; \; j = 1, \cdots N$
- $H_3 \; : \; \mu_i = 0, \; \nu_{ij} = 0 \quad i = 1, \cdots M; \; j = 1, \cdots N$ (perfect poolability)
- $H_4 \; : \; \nu_{ij} = 0 \quad i$ fixed; $\; j = 1, \cdots N$.

and to make precise the maintained hypotheses for each case of hypothesis testing.

Exercise. Check that the F-statistic for testing H_2 against H_m has $(M(N - 1), \; MN(T-1) - K)$ degrees of freedom. ■

3.5.3 Hierarchical Random Effects

The model

$$y_{ijt} = \alpha + z'_{ijt}\beta + \upsilon_{ijt} \qquad \upsilon_{ijt} = \mu_i + \nu_{ij} + \varepsilon_{ijt} \tag{3.84}$$

or, equivalently:

$$\begin{aligned} y &= \iota_{NT}\alpha + Z\,\beta + \upsilon = Z_*\,\delta + \upsilon \\ \upsilon &= Z_\mu\mu + Z_\nu\nu + \varepsilon, \end{aligned} \tag{3.85}$$

where $Z_* = [\iota_{NT} \;\; Z]$ and $\delta = [\alpha,\, \beta']'$, under the following assumptions:

$$\varepsilon \sim (0,\, \sigma^2_\varepsilon\, I_{(MNT)}) \tag{3.86}$$

$$\mu \sim (0,\, \sigma^2_\mu\, I_{(M)}) \tag{3.87}$$

$$\nu \sim (0,\, \sigma^2_\nu\, I_{(MN)}) \tag{3.88}$$

$$\mu \perp\!\!\!\perp \nu \perp\!\!\!\perp \varepsilon \perp\!\!\!\perp Z \perp\!\!\!\perp (Z_\mu\,,\, Z_\nu) \tag{3.89}$$

$$\theta_{HRE} = (\alpha,\, \beta',\, \sigma^2_\varepsilon,\, \sigma^2_\mu,\, \sigma^2_\nu\,) \in \Theta_{HRE} = I\!R^{1+K} \times I\!R^3_+ \tag{3.90}$$

This is a linear regression model with non-spherical residuals, describing the first two moments of a process generating the distribution of $(y \mid Z,\, Z_\nu,\, Z_\mu,\, \theta_{HRE})$ as follows:

$$\mathbb{E}\,[y \mid Z,\, Z_\mu,\, Z_\nu,\, \theta_{HRE}) = \iota_{NT}\alpha + Z\,\beta \tag{3.91}$$

$$V(y \mid Z,\, Z_\mu,\, Z_\nu,\, \theta_{HRE}) = \Omega \tag{3.92}$$

where

$$\begin{aligned} \Omega &= \mathbb{E}\,(\upsilon\,\upsilon' \mid Z,\, Z_\mu,\, Z_\nu,\, \theta_{HRE}) \\ &= \sigma^2_\mu\, Z_\mu\, Z'_\mu + \sigma^2_\nu\, Z_\nu\, Z'_\nu + \sigma^2_\varepsilon\, I_{(MNT)} \\ &= \sigma^2_\mu\, I_{(M)} \otimes J_{(N)} \otimes J_{(T)} + \sigma^2_\nu\, I_{(M)} \otimes I_{(N)} \otimes J_{(T)} \\ &\qquad + \sigma^2_\varepsilon\, I_{(M)} \otimes I_{(N)} \otimes I_{(T)} \end{aligned} \tag{3.93}$$

Exercises

(i) In order to verify (3.93), check that:$I_{MN} = I_M \otimes I_N$ and $J_{NT} = J_N \otimes J_T$.

(ii) check that the structure of the variances and covariances can also be described as follows:

$$\begin{aligned} cov(\upsilon_{ijt}, \upsilon_{i'j't'}) &= cov(\mu_i + \nu_{ij} + \varepsilon_{ijt},\ \mu_i' + \nu_{i'j'} + \varepsilon_{i'j't'}) \\ &= \sigma_\mu^2 + \sigma_\nu^2 + \sigma_\varepsilon^2 \quad i = i',\ j = j',\ t = t' \\ &= \sigma_\mu^2 + \sigma_\nu^2 \quad i = i',\ j = j',\ t \neq t' \\ &= \sigma_\mu^2 \quad i = i',\ j \neq j' \\ &= 0 \quad i \neq i' \end{aligned} \tag{3.94}$$

Estimation of the regression coefficients

Under (3.93), we obtain that $p = 3$. Table 3.5 gives the projectors Q_i, the corresponding eigenvalues λ_i, their respective multiplicities along with the transformations they operate in $I\!R^{MNT}$.

Table 3.5: *Spectral decomposition for the Two-way HRE model*

i	Q_i	λ_i	mult(λ_i)=$r(Q_i)$	$Q_i y$
1	$I_M \otimes I_N \otimes E_T$	σ_ε^2	$MN(T-1)$	$[y_{ijt} - \bar{y}_{ij.}]$
2	$I_M \otimes E_N \otimes \bar{J}_T$	$\sigma_\varepsilon^2 + T\sigma_\nu^2$	$M(N-1)$	$[\bar{y}_{ij.} - \bar{y}_{i..}] \otimes \iota_T$
3	$I_M \otimes \bar{J}_N \otimes \bar{J}_T$	$\sigma_\varepsilon^2 + T\sigma_\nu^2 + NT\sigma_\mu^2$	M	$[\bar{y}_{i..}] \otimes \iota_{NT}$
Sum	I_{NMT}		MNT	

When evaluating the GLS estimator of (3.85) by means of the spectral decomposition (7.25):

$$\hat{\delta}_{GLS} = [\sum_{1 \leq i \leq p} W_i]^{-1} [\sum_{1 \leq i \leq p} W_i d_i]$$

the 3 OLS estimators d_i are obtained by OLS estimations on the data successively transformed into deviations from the time-averages, $[y_{ijt} - \bar{y}_{ij.}]$, deviations between the time-averages and the time-and-sector averages, $[\bar{y}_{ij.} - \bar{y}_{i..}]$ and the time-and-sector averages, $[\bar{y}_{i..}]$. Note that the first two regressions are *without constant terms* and that $Q_1 = M_\nu$. Thus the GLS estimation of α is obtained from the third regression only:

$$\hat{\alpha}_{GLS} = \bar{y}_{...} - \bar{z}'_{...}\hat{\beta}_{GLS} \tag{3.95}$$

3.5.4 Hierarchical Mixed Effects

The model

$$y_{ijt} = \alpha + z'_{ijt}\beta + \mu_i + \upsilon_{ijt} \qquad \upsilon_{ijt} = \nu_{ij} + \varepsilon_{ijt} \tag{3.96}$$

or, equivalently:

$$y = \iota_{NT}\alpha + Z\,\beta + Z_\mu\,\mu + \upsilon \qquad \upsilon = Z_\nu\nu + \varepsilon \tag{3.97}$$

under the following assumptions:

$$\begin{aligned}
\varepsilon &\sim (0,\ \sigma^2_\varepsilon\, I_{(MNT)}) && (3.98)\\
\nu &\sim (0,\ \sigma^2_\nu\, I_{(MN)}) && (3.99)\\
\nu \perp\!\!\!\perp \varepsilon &\perp\!\!\!\perp (Z,\ Z_\mu) \perp\!\!\!\perp Z_\nu && (3.100)\\
\theta_{HME} = (\alpha,\ \beta',\ \mu,\ \sigma^2_\varepsilon,\ \sigma^2_\nu\,) &\in \Theta_{HME} = I\!R^{1+K+M} \times I\!R^2_+ && (3.101)\\
& && (3.102)
\end{aligned}$$

This is a linear regression model with non-spherical residuals, describing the first two moments of a process generating the distribution of $(y \mid Z,\ Z_\nu,\ Z_\mu,\ \theta_{HME})$ as follows:

$$\begin{aligned}
I\!E\,[y \mid Z,\ Z_\mu,\ Z_\nu,\ \theta_{HME}) &= \iota_{NT}\alpha + Z\,\beta + Z_\mu\,\mu && (3.103)\\
V(y \mid Z,\ Z_\mu,\ Z_\lambda,\ \theta_{HME}) &= \Omega && (3.104)
\end{aligned}$$

where

$$\begin{aligned}
\Omega &= I\!E\,(\upsilon\,\upsilon' \mid Z,\ Z_\mu,\ Z_\nu,\ \theta_{HME})\\
&= \sigma^2_\nu\, Z_\nu\, Z'_\nu + \sigma^2_\varepsilon\, I_{(MNT)}\\
&= \sigma^2_\varepsilon\, I_{(M)} \otimes I_{(N)} \otimes I_{(T)} + \sigma^2_\nu\, I_{(M)} \otimes I_{(N)} \otimes J_{(T)} && (3.105)
\end{aligned}$$

Thus, the structure of the variances and covariances is as follows:

$$\begin{aligned} cov(\upsilon_{ijt}, \upsilon_{i'j't'}) &= cov(\nu_{ij} + \varepsilon_{ijt}, \nu_{i'j'} + \varepsilon_{i'j't'}) \\ &= \sigma^2_\nu + \sigma^2_\varepsilon \quad i = i', j = j', t = t' \\ &= \sigma^2_\nu \quad i = i', j = j', t \neq t' \\ &= 0 \quad i \neq i' \text{ and/or } j \neq j' \end{aligned} \tag{3.106}$$

Estimation of the regression coefficients

As the residuals are not spherical, we again make use of the spectral decomosition of the residual covariance matrix. Under (3.105), we obtain that $p = 2$. Table 3.6 gives the projectors Q_i, the corresponding eigenvalues λ_i, their respective multiplicities along with the transformations they operate in $I\!R^{MNT}$.

Table 3.6: *Spectral decomposition for the Two-way HME model*

i	Q_i	λ_i	mult(λ_i)=$r(Q_i)$	$Q_i y$
1	$I_{MN} \otimes E_T$	σ^2_ε	$MN(T-1)$	$[y_{ijt} - \bar{y}_{ij.}]$
2	$I_{MN} \otimes \bar{J}_T$	$\sigma^2_\varepsilon + T\sigma^2_\nu$	MN	$[\bar{y}_{ij.}] \otimes \iota_T$
Sum	I_{MNT}		MNT	

3.6 Appendix

3.6.1 Summary of Results on Z_μ and Z_λ

$$
\begin{array}{lllllll}
Z_\mu & = & I_{(N)} \otimes \iota_T & : NT \times N & Z_\lambda & = & \iota_N \otimes I_{(T)} & : NT \times T \\
Z_\mu \iota_N & = & \iota_{NT} & & Z_\lambda \iota_T & = & \iota_{NT} & \\
Z_\mu Z_\mu' & = & I_{(N)} \otimes J_T & : NT \times NT & Z_\lambda Z_\lambda' & = & J_N \otimes I_{(T)} & : NT \times NT \\
Z_\mu' Z_\mu & = & T.I_{(N)} & : N \times N & Z_\lambda' Z_\lambda & = & N.I_{(T)} & : T \times T \\
Z_\mu Z_\mu^+ & = & I_{(N)} \otimes \bar{J}_T & : NT \times NT & Z_\lambda Z_\lambda^+ & = & \bar{J}_N \otimes I_{(T)} & : NT \times NT \\
Z_\mu' Z_\lambda & = & \iota_N \otimes \iota_T' & : N \times T & Z_\lambda' Z_\mu & = & \iota_N' \otimes \iota_T & : T \times N \\
 & = & \iota_N \iota_T' & & & = & \iota_T \iota_N' &
\end{array}
$$

with: $J_N = \iota_N \iota_N'$ and $\bar{J}_N = \frac{1}{N} \iota_N \iota_N'$. Furthermore:

$$
Z_\mu Z_\mu^+ y \quad = \quad [\bar{y}_{i.}] \otimes \iota_T \quad : NT \times 1 \qquad Z_\lambda Z_\lambda^+ y \quad = \quad \iota_N \otimes [\bar{y}_{.t}] \quad : NT \times 1
$$

Chapter 4

Incomplete Panels

4.1 Introduction

4.1.1 Different types of incompleteness

According to the object of missingness

Item non response
Unit non-response

According to the cause of missingness

By design of the panel

- Rotating panel: at each wave, part of the individuals are replaced by new ones.
- Split panel: two parts, a panel and a series of cross sections
- Complex sampling: for instance, in a hierarchical sampling, one may have different number of individuals in each sector and a different number of observations for each individuals; in such a case, one obtains: $i = 1, \cdots, M \;\; ; j = 1, \cdots, N_i \;\; ; t = 1, \cdots, T_{ij}$

By behaviour of the respondents
Unit non response
This is: individuals are selected but do not respond, from the first wave on or in later waves. The later case, of answering first waves but giving up at later

waves, is also called *attrition*. Causes of unit non response include absence at the time of the interview. These absences may be due, in turn, to death, home moving or removal from the frame of the sampling. Unit non response is a major problem in most actual panels and a basic motivation for rotating panel.

Item non-response
Behavioural causes of item non response include refusal of answering sensitive questions, misunderstanding of some questions, ignorance of the requested information .

4.1.2 Different problems raised by incompleteness

Analytical and computational issues

In the balanced case, the matrices related to the sampling design have a Kronecker product form, *e.g.* $Z_\mu = I_{(N)} \otimes \iota_T$ or $Z_\lambda = \iota_N \otimes I_{(T)}$. Many of the analytical developments take advantage of that structure. Once the panel data are not anymore balanced, this structure is lost: the formulae, and the computations, become more complicated and, in some cases, untractable.This implies difficulties both for the analytical treatment and for the interpretation of the results. Thus, this is an issue on operationality.

Sampling bias

Once some data are missing, the presence or absence of the data may be associated to the behaviour under analysis. For instance, in a survey on individual mobility made by telephone interviews, the higher probability of non response comes precisely from the more mobile individuals. We should then admit that in such a case, the missingness process should not be ignored, *i.e.* is *not ignorable*, or equivalently that the selection mechanism is *informative*.

In other words, one should recognize that the empirical information has two components: the fact of the data being available or not, as distinct from the value of the variables being measured. A proper building of the likelihood function should model these two aspects of the data; failure of recognizing this fact may imply inadequate likelihood function and eventually inconsistency in the inference. This is an issue on statistical inference and raises a preliminary question: how far is it possible to detect, or to test, whether a missingness process is ignorable or not.

It should be stressed that the question of ignorability should be considered seriously whether the missingness is attibuted to the design or the the behaviour of the respondent. For instance, many panels are made rotating in order to "replace" the nonrespondent individuals; in such a case, one should take care of checking: why have some individuals failed to repond and how the replacing individuals have been selected.

4.2 Randomness in Missing Data

4.2.1 Random missingness

Let
$\xi = (\eta, \zeta)$ denote the complete, or potential, data
S denote the selection mechanism
$x = S\xi = (S\eta, S\zeta) = (y, z)$ denote the available data

For instance, in the "pure" survey case- *i.e.* a one-wave panel- ξ could be an $N \times (k + 1)$ matrix where N represents the designed sample size, *i.e.* the sample size if there were no unit non response and S would be an $n \times N$ selection matrix where each row is a (different) row of $I_{(N)}$, *i.e.* premultiplying ξ by one row of S "selects" one row of ξ; n represents the actually available sample size and x is now an $n \times (k + 1)$ matrix of available data.
The main point is that the actual data, x, is a deterministic function of a latent vector, or matrix, ξ and of a random S. In this case, it may be meaningful to introduce an indicator of the "missing" part of the data, in the form of a selection matrix $\bar{S}$, with the idea that $\bar{S}\xi$ denotes the actually missing data. Note that $\bar{S}$ might be known to the statistician but not $\bar{S}\xi$ and that, when S is known, the knowledge of x and $\bar{S}\xi$ would be equivalent to the complete knowledge of ξ. Note that when S and $\bar{S}$ are represented by selection matrices of order $n \times N$ and $(Nn) \times N$ respectively, the matrix $[S' \; \bar{S}']$ is an $N \times N$ permutation matrix and knowing S is equivalent to know $\bar{S}$.

It is of importance to distinguish the cases where the available data (D) are $D = x$ or $D = (x, S)$.

4.2.2 Ignorable and non ignorable missingness

Once it is recognized that the selection mechanism is stochastic, by behaviour of the individuals and/or by design of the survey, one should check that the selection has not biased the representativity of the sampling. In other words, *the problem* is to find conditions that would allow one to model the available data "as if" they were complete. These conditions are called conditions of "ignorability". The idea is that when such conditions are not satisfied, the selection mechanism should also be modelled as a part of the Data Generating Process.

Missing completely at random (MCAR)

$$X \perp\!\!\!\perp S \qquad \text{or} \quad \xi \perp\!\!\!\perp S$$

Missing at random (MAR)

$$y \perp\!\!\!\perp S \mid Z \qquad \text{or} \quad \eta \perp\!\!\!\perp S \mid \zeta$$

4.3 Non Balanced Panels with random missingness

4.3.1 Introduction

The basic problem

When the selection mechanism is completely at random, that the panel data are non balanced is the only problem to take care of. In such a case, the design matrices, such as Z_μ , Z_λ or Z_ν are not anymore in the form of Kronecker product. This implies more cumbersome computations and less transparent interpretations of the projections used in the decomposition of the multiple regressions, for the FE models, or in the spectral decomposition of the residual covariance matrix, for the RE models.

In what follows, we shall sketch the main analytical problems arising from incomplete data, with a particular attention to the computational and the interpretational issues. It should however be pointed out that other attitudes have also developped for facing that problem of data incompleteness. One alternative approach starts from the remark that it would still be possible to make use of the standard balanced data methods, provided one may succeed in transforming the incomplete data into balanced ones. There are many such procedures but two are particularly popular, even though not always advisable:

- *Imputation methods* are particularly used in case of item nonresponse. These methods consist in "imputing" to a missing value a value evaluated from that given by "similar" respondents. For instance, it can be the answer given by the most similar respondent, taking into account a (reasonable) number of individual characteristics, or the estimated conditional expectation from a regression of the variable corresponding to the missing data on other (reasonably chosen - and available!) variables. The usual criticism against the use of such methods is to notice that it boils down to a multiple use of a same data and that it may not correspond to a proper specification of the actual Data Generating Process.

- *Weighting methods* are particularly used in case of unit nonreponse or when an individual is characterized by a large number of missing responses. These methods consist in discarding the non-respondent, or the too poorly respondent, individuals and to reweight the remaining ones so as to recover an adequate representativity of the sample.

Some useful notations

Consider an individual i with T_i data, out of T potential data, available at the times $t_i(1) < \cdots < t_i(T_i)$ and define:

$$\mathcal{T}_i = \{t_i(1), \cdots t_i(T_i)\} \subset \{1, \cdots, T\} \qquad \text{with} \quad \cup_{1 \leq i \leq N} \mathcal{T}_i = \{1, \cdots, T\}$$

$$S_i = \begin{bmatrix} e'_{t_i(1)} \\ \vdots \\ e'_{t_i(T_i)} \end{bmatrix} \qquad : T_i \times T \quad \text{(selection matrix)}$$

$$R_i = S_i' S_i = \underset{1 \leq t \leq T_i}{diag} \left(I\!I_{\{t \in \mathcal{T}_i\}} \right) \quad : T \times T \quad i.e. \;\; r_{i,st} = 1 \Leftrightarrow s = t \in \mathcal{T}_i$$

For the complete sample, we also use:

$$S = \underset{1 \leq i \leq N}{diag} (S_i) \qquad : \; T_. \times NT \quad (\text{where } T_. = \sum_{1 \leq i \leq N} T_i)$$

$$S_\mu = \underset{1 \leq i \leq N}{diag} (\iota_{T_i}) \qquad : \; T_. \times N$$

Exercise. Check that :

$$\begin{aligned} S_i^+ &= S_i' \\ R_i^+ &= R_i \\ S_i\,S_i' &= I_{(T_i)} \\ (S_i'\,S_i)^+ &= R_i^+ \,=\, R_i \\ S_\mu^+ &= \underset{1\leq i\leq N}{diag}\,(\frac{1}{T_i}\iota_{T_i}') \\ S_\mu\,S_\mu' &= \underset{1\leq i\leq N}{diag}\,(T_i\,\bar{J}_{T_i}) \end{aligned}$$

■

4.3.2 One-way models

Fixed Effect

$$y_{it} \,=\, \alpha \,+\, z_{it}'\beta \,+\, \mu_i \,+\, \varepsilon_{it} \qquad t \,=\, t_i(1),\cdots,t_i(T_i);\; i \,=\, 1,\cdots,N \tag{4.1}$$

or, equivalently:

$$y \,=\, \iota_{T.}\alpha \,+\, Z\,\beta \,+\, S_\mu\mu \,+\varepsilon \tag{4.2}$$

In the case of a FE model, the unbalanced case implies minor modifications of the balanced case only. In particular, $\iota_{T.}$ is again in the column space of S_μ; the projection on its orthogonal complement is:

$$M_{S,\mu} \,=\, I_{(T.)}S_\mu\,S_\mu^+ \,=\, \underset{1\leq i\leq N}{diag}\,(E_{T_i}) \tag{4.3}$$

$$M_{S,\mu}\,y \,=\, [E_{T_i}\,y_i] \,=\, [y_{it} \,-\, \bar{y}_{i.}] \tag{4.4}$$

where

$$\bar{y}_{i.} \,=\, \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} y_{it}$$

Using the same argument as in Section 2.2.2, we obtain:

$$\hat{\beta}_{OLS} \,=\, (Z'M_{S,\mu}Z)^{-1}\,Z'M_{S,\mu}\,y \tag{4.5}$$

$$V(\hat{\beta}_{OLS}\mid Z,\,S_\mu,\,\theta_{FE}) \,=\, \sigma_\varepsilon^2\,(Z'M_{S,\mu}Z)^{-1} \tag{4.6}$$

Thus, as compared with the balanced case, $\hat{\beta}_{OLS}$ is again in the form of a "Within estimator" and the only change is that the individual averages are computed with respect to different sizes T_i.

Random Effect

The model

$$y_{it} = \alpha + z_{it}'\beta + \upsilon_{it} \qquad \upsilon_{it} = \mu_i + \varepsilon_{it} \qquad t \in \Im_i;\ i = 1, \cdots, N \tag{4.7}$$

or, equivalently:

$$\begin{aligned} y &= \iota_{T.}\alpha + Z\beta + \upsilon = Z_* \delta + u \\ & \qquad \text{where } Z_* = [\iota_{T.} \ \ Z] \quad \delta = [\alpha \ \ \beta']' \qquad (4.8) \\ \upsilon &= S_\mu \mu + \varepsilon \qquad (4.9) \end{aligned}$$

under the same assumptions as in the balanced case, in particular:

$$Z \perp\!\!\!\perp S_\mu \perp\!\!\!\perp \mu \perp\!\!\!\perp \varepsilon \tag{4.10}$$

Note that the assumption of independence between S_μ and (Z, μ, ε) is actually an ignorability assumption.

GLS Estimation of the regression coefficients

It is easily checked that

$$\begin{aligned} V(\upsilon) &= \Omega = V(S_\mu \mu + \varepsilon) = \sigma_\mu^2 S_\mu S_\mu' + \sigma_\varepsilon^2 I_{(T.)} \\ &= \underset{1 \leq i \leq N}{diag} (\Omega_i) \qquad (4.11) \\ \Omega_i &= \sigma_\varepsilon^2 E_{T_i} + (\sigma_\varepsilon^2 + T_i \sigma_\mu^2) \bar{J}_{T_i} \quad r(\Omega_i) = T_i \\ &= \sigma_\varepsilon^2 E_{T_i} + \omega_i^2 \bar{J}_{T_i} \qquad \text{where: } \omega_i^2 = \sigma_\varepsilon^2 + T_i \sigma_\mu^2 \qquad (4.12) \end{aligned}$$

Therefore:

$$V(\upsilon) = \sigma_\varepsilon^2 \underset{1 \leq i \leq N}{diag} (E_{T_i}) + \sum_{1 \leq j \leq N} \omega_i^2 \underset{1 \leq i \leq N}{diag} (\delta_{ij} \bar{J}_{T_i}) \tag{4.13}$$

Thus, the eigenvalues of the residual covariance matrix are : σ_ε^2 and $\omega_i^2 = \sigma_\varepsilon^2 + T_i \sigma_\mu^2$ and there are as many different eigenvalues as different values of T_i plus one. Let be $(p-1)$ different values of T_i and order these values as follows: $T_{(1)} < \cdots T_{(j)} < \cdots < T_{(p-1)}$ and let $p_{(j)}$ be the number of individuals with a same number of observations $T_{(j)}$. Thus:

$$\sum_j p_{(j)} = N \quad \text{and} \quad \sum_j p_{(j)} T_{(j)} = T.$$

After ordering the individuals in increasing order of number of observations, the spectral decomposition of Ω takes the form:

$$\Omega \;=\; \sigma^2_\varepsilon \underset{1\leq i\leq N}{diag}\;(E_{T_i}) \;+ \sum_{1\leq j\leq p-1} \omega^2_{(j)} \underset{1\leq i\leq p-1}{diag}\;(\delta_{ij} I_{(p_{(j)})} \otimes \bar{J}_{T_{(j)}}) \tag{4.14}$$

Table 4.1: *Spectral decomposition for the One Way RE Unbalanced*

i	Q_i	λ_i	mult(λ_i) $=r(Q_i)$	$Q_i\, y$
1	$\underset{1\leq j\leq p-1}{diag}\;(I_{(p_{(j)})} \otimes E_{T_{(j)}})$	σ^2_ε	$T_. - N$	$[y_{ij} - \bar{y}_{i.}]$
$j \;=\; 1, \cdots p-1$	$\underset{1\leq k\leq p-1}{diag}\;(\delta_{jk} I_{(p_{(j)})} \otimes \bar{J}_{T_{(j)}})$	$\omega^2_{(j)} \;=\; \sigma^2_\varepsilon \;+\; T_{(j)}\sigma^2_\mu$	$p_{(j)}$	$([\bar{y}_{i.}]_{T_i=T_{(j)}} \otimes \delta_{jk}\, \iota_{T_{(j)}})$ $k \;=\; 1, \cdots p-1$
Sum	$I_{T_.}$		$T_.$	

Table 4.1 gives the projectors Q_i, the corresponding eigenvalues λ_i, their respective multiplicities along with the transformations they operate in $I\!R^{T_.}$. Note however that the p distinct eigenvalues $(\sigma^2_\varepsilon, \omega^2_{(j)} \;\; j \;=\; 1, \cdots, p-1)$ are subject to $p-2$ restrictions because they depend on only 2 variation-free parameters, namely σ^2_ε and σ^2_μ.

From the spectral decomposition of the residual covariance matrix, given in Table 4.1 , one may obtain the GLS estimator of $\delta \;=\; [\alpha \;\; \beta']'$ by the usual weighted average of the OLS estimators on the data transformed successively by the projectors Q_i. Taking advantage of the particular form of the last $p-1$ projectors, one may use the following shortcut. Note first that (4.11) and (4.12) imply:

$$\sigma_\varepsilon\, \Omega^{-\frac{1}{2}} \;=\; \underset{1\leq i\leq N}{diag}\;[\frac{\sigma_\varepsilon}{\omega_i}\bar{J}_{T_i} \;+\; E_{T_i}] \tag{4.15}$$

The GLS estimator is then obtained as:

$$\hat{\delta}_{GLS} = (Z'_{**}Z_{**})^{-1}Z'_{**}y_* \quad \text{where} \quad \begin{aligned} Z_{**} &= \sigma_\varepsilon \Omega^{-\frac{1}{2}} Z_* \\ y_* &= \sigma_\varepsilon \Omega^{-\frac{1}{2}} y = [y_{it}\theta_i \, \bar{y}_{i.}] \\ \theta_i &= 1\frac{\sigma_\varepsilon}{\omega_i} \in [0,\ 1] \end{aligned} \qquad (4.16)$$

Thus, $\hat{\delta}_{GLS}$ may be obtained by means of an OLS on the data transformed into weighted deviations from the individual averages: $y_{it}\theta_i \, \bar{y}_{i.}$, $z_{itk}\theta_i \, \bar{z}_{i.k} \quad k = 1, \cdots K+1$.

Estimation of the variances

The estimator $\hat{\delta}_{GLS}$ is feasible as soon as the two variances σ^2_ε and σ^2_μ are known or, at least, have been estimated. With this purpose in mind, let us consider the "Within" and the "Between" projections;

$$Q_1 = \underset{1\leq i\leq N}{diag}\,(E_{T_i}) \qquad r(Q_1) = \sum_{1\leq i\leq N}(T_i 1) = T_. N \qquad (4.17)$$

$$Q_2 = \underset{1\leq i\leq N}{diag}\,(\bar{J}_{T_i}) \qquad r(Q_2) = N \qquad (4.18)$$

Clearly:

$$Q_i = Q'_i = Q^2_i \quad i = 1, 2 \quad Q_1 Q_2 = 0 \quad Q_1 + Q_2 = I_{T.}$$

Let us now evaluate the expectation of the associated quadratic forms:

$$\mathbb{E}\,(\upsilon' Q_j \, \upsilon) = tr Q_j \, \Omega = tr Q_j \, [\underset{1\leq i\leq N}{diag}\,(\sigma^2_\varepsilon E_{T_i} + \omega^2_i \bar{J}_{T_i})] \qquad j = 1, 2. \quad (4.19)$$

Therefore

$$\begin{aligned} \mathbb{E}\,(\upsilon' Q_1 \, \upsilon) &= tr[\underset{1\leq i\leq N}{diag}\,(E_{T_i})][\underset{1\leq i\leq N}{diag}\,(\sigma^2_\varepsilon E_{T_i} + \omega^2_i \bar{J}_{T_i})] \\ &= (T_. N)\, \sigma^2_\varepsilon \qquad (4.20) \\ \mathbb{E}\,(\upsilon' Q_2 \, \upsilon) &= tr[\underset{1\leq i\leq N}{diag}\,(\bar{J}_{T_i})][\underset{1\leq i\leq N}{diag}\,(\sigma^2_\varepsilon E_{T_i} + \omega^2_i \bar{J}_{T_i}] \\ &= N\,\sigma^2_\varepsilon + T_.\,\sigma^2_\mu \qquad (4.21) \end{aligned}$$

Thus, IF the true residuals υ were observable, one could estimate the variances unbiasedly as follows:

$$\tilde{\sigma}^2_\varepsilon = \frac{\upsilon' Q_1 \, \upsilon}{T_. N} \qquad (4.22)$$

$$\tilde{\sigma}^2_\mu = \frac{\upsilon' Q_2 \, \upsilon \frac{N}{T_. N} \upsilon' Q_1 \, \upsilon}{T_.} \qquad (4.23)$$

As the true residuals υ are not observable, a simple idea is to replace them by the OLS residuals: $\hat{\upsilon} = M_* y = M_* \upsilon$ where $M_* = [I_{(T.)} Z_* Z_*^+]$. Taking into account that both M_* and Q_j are projectors, the expectation of these estimators are:

$$\begin{aligned}
\mathbb{E}\,(\hat{\upsilon}' Q_j\, \hat{\upsilon}) &= tr M_*\, Q_j\, M_*\, \Omega \;=\; tr M_*\, Q_j\, M_*\, [\underset{1 \le i \le N}{diag}\,(\sigma_\varepsilon^2 E_{T_i} \,+\, \omega_i^2 \bar{J}_{T_i})] \\
&= \sigma_\varepsilon^2\, tr Q_j\, M_* \,+\, \sigma_\mu^2\, tr M_*\, Q_j\, M_* S_\mu\, S_\mu' \\
& \qquad\qquad\qquad\qquad j \,=\, 1,\, 2. \qquad\qquad (4.24)
\end{aligned}$$

where we have made use of the identity: $S_\mu\, S_\mu' \,=\, \underset{1 \le i \le N}{diag}\,(T_i\, \bar{J}_{T_i})$.

These two expectations provides therefore in linear system of 2 equations in the two unknown variances $(\sigma_\varepsilon^2,\, \sigma_\mu^2)$. Denoting these estimators as : $q_j = \hat{\upsilon}' Q_j\, \hat{\upsilon} \quad j \,=\, 1,\, 2$, we obtain unbiased estimators of these variances by solving:

$$q_1 \;=\; a_{11}\sigma_\varepsilon^2 \,+\, a_{12}\sigma_\mu^2 \qquad (4.25)$$

$$q_2 \;=\; a_{21}\sigma_\varepsilon^2 \,+\, a_{22}\sigma_\mu^2 \qquad (4.26)$$

where, making use of the identity: $Q_2\, S_\mu\, S_\mu' \,=\, S_\mu\, S_\mu'\, Q_2 \,=\, S_\mu\, S_\mu'$

$$\begin{aligned}
a_{11} &= tr Q_1\, M_* \,=\, T_{\!.} N tr\, Q_1\, Z_*\, Z_*^+ \\
&= T_{\!.} N(K+1) tr\, Q_2\, Z_*\, Z_*^+ \\
a_{12} &= tr M_*\, Q_1\, M_* S_\mu\, S_\mu' \\
&= tr\, [Z_*\, Z_*^+ Z_*\, Z_*^+ Q_2\, Z_*\, Z_*^+] S_\mu\, S_\mu' \\
a_{21} &= tr Q_2\, M_* \,=\, N tr\, Q_2\, Z_*\, Z_*^+ \\
a_{22} &= tr M_*\, Q_2\, M_* S_\mu\, S_\mu' \\
&= tr\, S_\mu\, S_\mu' 2\, tr\, Z_*\, Z_*^+\, S_\mu\, S_\mu' \,+\, tr\, Z_*\, Z_*^+\, Q_2\, Z_*\, Z_*^+ S_\mu\, S_\mu'
\end{aligned}$$

Note that that plugging the OLS residuals into the estimators q_j is one among several alternative choices, such as, for instance, using the residuals of the "Within" regression, or the residuals of the "Between regression" or using the "Within resuiduals" for q_1 and the "Between" residuals for q_2. These modifications affect the matrix M_* above but leaves unchanged the general structure of the analytical manipulations.

4.3.3 Two-way models

Fixed Effects

Random Effects

4.4 Selection bias with non-random missingness

Chapter 5

Dynamic models with Panel Data

5.1 Introduction

5.2 Residual dynamics

5.3 Endogenous dynamics

Chapter 6

Complements of Linear Algebra (Appendix A)

6.1 Notations

r(A)	=	rank of matrix A
tr(A)	=	trace of (square) matrix A $(=\sum_{i=1}^{n} a_{ii})$
$\lvert A\rvert$	=	det (A) = determinant of matrix A
$I_{(n)}$	=	unit matrix of order n, also : identity operator on $I\!R^n$
e_i	=	i-th column of the unit matrix $I_{(n)}$
ι	=	$(1\,1\ldots1)'$
I	=	identity operator on an abstract vector space V
		$[I(x) = x \qquad \forall\, x \in V]$

Let

$$f:\ A \rightarrow B$$
$$A_0 \subseteq A \qquad B_0 \subseteq B$$

Then (definition):

$$f(A_0) \equiv \{b \in B \mid \exists\, a \in A_0 : f(a) = b\}$$
$$f^{-1}(B_0) \equiv \{a \in A \mid \exists\, b \in B_0 : f(a) = b\}$$

In particular:

$$Im(f) = f(A)$$
$$Ker(f) = f^{-1}(\{0\})$$

Remark
When f is a linear application represented by a matrix F, one may write

FA_0 instead of $f(A_0)$ and $F^{-1}A_0$ instead of $f^{-1}(A_0)$. In this later case, F^{-1} does not represent the inverse of matrix F. In particular, $F^{-1}A_0$ makes sense even if matrix F is singular. Context should avoid any possible ambiguity.

6.2 Partitionned Matrices and Sum of Quadratic Forms

Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$B \equiv A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where :

A and $B : n \times n$

A_{ij} and $B_{ij} : n_i \times n_j \qquad i,j = 1,2 \qquad n_1 + n_2 = n$

Remark. In what follows, $i = 1$ or $2, j = 1$ or 2 and $i \neq j$.

Theorem 6.2.1 *(Determinant of partitionned matrices)*

If :

$r\ (A_{ii}) = n_i$

then:

$$\begin{aligned} 1)\ |A| &= |A_{ii}|\,.\,|A_{jj} - A_{ji}A_{ii}^{-1}A_{ij}| \\ 2)\ r(A) &= n_i + r(A_{jj} - A_{ji}A_{ii}^{-1}A_{ij}) \end{aligned}$$

Corollary
Let $\quad C : n \times r$ and $D : r \times n$
then:
$|I_{(n)} + CD| = |I_{(r)} + DC|$

Theorem 6.2.2 *(Inverse of partitionned matrices)*
If :

$r(A_{ii}) = n_i$
$r(A_{jj} - A_{ji}A_{ii}^{-1}A_{ij}) = n_j$
$B = A^{-1}$

then:

$B_{jj} = [A_{jj} - A_{ji}A_{ii}^{-1}A_{ij}]^{-1}$
$B_{ii} = A_{ii}^{-1} + A_{ii}^{-1}A_{ij}B_{jj}A_{ji}A_{ii}^{-1}$
$B_{ij} = -A_{ii}^{-1}A_{ij}B_{jj}$
$B_{ji} = -B_{jj}A_{ji}A_{ii}^{-1}$

If, furthermore :

$r(A_{jj}) = n_j$

then :

$A_{jj}^{-1}A_{ji}B_{ii} = B_{jj}A_{ji}A_{ii}^{-1}$

Corollary

Let $A : n \times n$
$C : n \times p$
$D : p \times p$
$E : p \times n$

then,under evident rank conditions :

$$[A - CDE]^{-1} = A^{-1} + A^{-1}C[D^{-1} - EA^{-1}C]^{-1}EA^{-1}$$

Corollary

$$[I + ab']^{-1} = I - \frac{ab'}{1 + a'b}$$

Theorem 6.2.3 *(Decomposition of Quadratic Forms)*

Let $Q = x'A\,x$
$x' = (x_1' \;\; x_2') \qquad x_i : n_i \times 1 \qquad i = 1, 2.$
$r(A_{ii}) = n_i$
$r(B_{jj}) = n_j$
$x_{i|j} = -A_{ii}^{-1}A_{ij}x_j = B_{ij}B_{jj}^{-1}x_j$

then $Q = (x_i - x_{i|j})'\,A_{ii}(x_i - x_{i|j}) \; + \; x_jB_{jj}^{-1}x_j$

6.3 Invariant spaces- Characteristic polynomial- Eigenvectors - Eigenvalues.

Let V: vector space of finite dimension; thus, let us consider $V = I\!R^n$
A: $V \to V$, linear(thus, one may consider A represented by a an $n \times n$ matrix)
E: s.v.s. of V

Definition E is an **invariant sub-space** of A
$\Leftrightarrow\ x \in E \Rightarrow Ax \in E$
$\Leftrightarrow\ AE \subseteq E$

Invariant sub-spaces of a linear transformation (or, of a matrix) A, may be characterized by a basis, *i.e.* a family of non null vectors linearly independent and solution of the **characteristic equation** :

$$Ax = \lambda x$$

When (λ_i, x_i) is a solution of the characteristic equation, we say that x_i is an **eigenvector** (or characteristic vector) associated to the **eigenvalue** (or characteristic value) λ_i. Any eigenvalue is a root of the **characteristic polynomial** of A :

$$\varphi_A(\lambda) = |A - \lambda I|$$

As $\varphi_A(\lambda)$ is a polynomial in λ of degree n, the equation $\varphi_A(\lambda) = 0$ admits n solutions, therefore, there are n eigenvalues (real or complex, distinct or multiple).

Definition. The set of all eigenvalues , λ_i, of A , along with their multiplicities, r_i, is called the **spectrum** of A. It is denoted as :

$$S = S(A) = \{\lambda_1^{r_1}, \lambda_2^{r_2}, \cdots, \lambda_p^{r_p}\}$$

(with $\sum_{i=1}^{p} r_i = n$ when A is diagonalisable)

Theorem 6.3.1 *(Properties of Eigenvalues and of Eigenvectors)*

1. *Let* $A:\ n \times n$*(alternatively* : $A : I\!R^n \to I\!R^n$*, linear)*
   then

i) $r(A) \geq$ *number of non-zero eigenvalues. (with equality when A is diagonalizable).*

ii) $|A|$ = *product of eigenvalues.*

iii) $tr(A)$ = *sum of eigenvalues.*

iv) The set of all eigenvectors associated to a same eigenvalue, completed by the null vector, conform an invariant subspace of A.

2. *Let* $A : n \times m$ *and* $B : m \times n$ *(alternatively :* $A : \mathbb{R}^m \to \mathbb{R}^n$, *linear and* $B : \mathbb{R}^n \to \mathbb{R}^m$, *linear)*
then AB and BA have the same non-zero eigenvalues.

Theorem 6.3.2 *(Cayley - Hamilton)*

Let $A : n \times n$

$$\varphi_A(\lambda) = |A - \lambda I| = \sum_{i=o}^{n} \alpha_i \lambda^i$$

then $\sum_{i=0}^{n} \alpha_i \lambda^i = 0 \Rightarrow \sum_{i=0}^{n} \alpha_i A^i = 0$

Theorem 6.3.3 *(Real Eigenvalues) In each of the following three cases, all the eigenvalues are real:*

(i) $A = A'$ *(symmetric)*

(ii) $A = A^2$ *(projection)*

(iii) A is a product of two symmetric matrices, one of which is non-singular.

6.4 Orthogonal and Similar Matrices

Theorem 6.4.1 *(Alternative Definitions of Orthogonal Matrices)*

Let $A : n \times n$

then the following properties are equivalent and define an ***orthogonal matrix****:*

(i) $A\,A' = I_{(n)}$

(ii) $A'A = I_{(n)}$

(iii) $A' = A^{-1}$

(iv) each rows,(columns) have unit length and are mutually orthogonal.

Theorem 6.4.2 *(Properties of orthogonal matrices)*
If $A(n \times n)$ *is orthogonal*
then

1) $\sum_{i=1}^{n} a_{ij}^2 = \sum_{j=1}^{n} a_{ij}^2 = 1$

2) $|a_{ij}| \leq 1 \quad \forall (i,j)$

3) B *orthogonal* $\Rightarrow A\,B$ *orthogonal*

4) $|A| \in \{-1, +1\}$

Corollary
The orthogonal $(n \times n)$ matrices, along with their product, form a (non-commutative) group.

Definition
Two matrices $A(n \times n)$ and $B(n \times n)$ are **similar** if there exists a non-singular matrix $P(n \times n)$ such that $A = PBP^{-1}$

Lemma
The similarity among the $(n \times n)$ matrices is an equivalence relation .

Theorem 6.4.3 *(Properties of similar matrices)*
Let A and B be two similar matrices.
then

(i) $r(A) = r(B)$

(ii) $|A| = |B|$

(iii) $tr(A) = tr(B)$

(iv) $\varphi_A(\lambda) = \varphi_B(\lambda)$ *and therefore* $|A - \lambda I_{(n)}| = 0 \Leftrightarrow |B - \lambda I_{(n)}| = 0$

6.5 P.D.S. Matrices and Quadratic Forms

Theorem 6.5.1 *(Alternative Definitions of P.S.D.S.matrices)*

Let $A = A' : (n \times n)$
then the following properties are equivalent and define the matrix A to be ***Positive SemiDefinite Symmetric*** *(P.S.D.S.)*

(i) $x' A\, x \geq 0 \quad \forall x \in I\!R^n$

(ii) $\exists R : (n \times n)$ *such that* $A = RR'$

(iii) all principal minors of A are non-negative

(iv) all eigenvalues of A are non-negative

Theorem 6.5.2 *(Characteristic Properties of P.S.D.S. Matrices)*

$$A \ \ P.S.D.S. \Longleftrightarrow [\,\forall\, B \ P.S.D.S \ \Rightarrow \ tr(AB) \geq 0]$$

Theorem 6.5.3 *(Alternative Definitions of P.D.S. matrices.)*

Let $A = A' : (n \times n)$

then the following properties are equivalent and define the matrix A to be ***Positive Definite Symmetric*** *(P.D.S.)*

(i) *A is P.S.D.S. and is non singular*

(ii) $x' A\, x > 0 \quad \forall x \neq 0 \quad x \in I\!R^n$

(iii) $\exists R : (n \times n)$ *non singular such that* $A = RR'$

(iv) *all principal minors of A are strictly positive*

(v) *all eigenvalues of A are strictly positive*

Theorem 6.5.4 *(Characteristic Properties of P.D.S. Matrices)*

$$\begin{aligned} A \ \ P.D.S \quad &\Longleftrightarrow \quad [\forall B \ P.D.S. \ : \ tr(AB) > 0] \\ &\Longleftrightarrow A^{-1} \ \ P.D.S \end{aligned}$$

Theorem 6.5.5 *(Properties of P.D.S. and P.S.D.S. Matrices)*

Let $A : (n \times n)$ *P.D.S.*
$B : (n \times n)$ *P.S.D.S.* $\quad D : (n \times n)$ *P.S.D.S.*
$C : (n \times m) \quad r(C) = m \leq n$
$c > 0$

then

(i) *cA is P.D.S. and cB is P.S.D.S.*

(ii) $A + B$ *is P.D.S.*

(iii) $C'C$ *is P.D.S.* $(m \times m)$

(iv) CC' *is P.S.D.S.*$(n \times n)$

(v) $r(B + D) \geq max\{r(B), r(D)\}$

Remark Properties (i) and (ii) above show that the set of $(n \times n)$ P.S.D.S. matrices is a **cone** denoted as $\mathcal{C}_{(n)}$

Theorem 6.5.6 *(Alternative Definitions of inequality among P.S.D.S. matrices)*

Let A and B be two $(n \times n)$ *P.S.D.S. matrices then the following properties are equivalent and define a partial preorder on the cone* $\mathcal{C}_{(n)}$ *denoted as:*
$A \leq B$ *(in the P.D.S. sense)* $(resp. <)$

(i) $x'A\,x \leq x'B\,x \quad \forall x \in I\!R^n (resp. <)$

(ii) $|A - \lambda B| = 0 \Rightarrow \lambda \in [0, 1] (resp. \lambda \in]0, 1])$

(iii) $B - A$ *is P.S.D.S. (resp. P.D.S.)*

Theorem 6.5.7 *(Properties of the inequality among P.S.D.S. matrices)*

If $A \leq B$ *(in the P.S.D.S. sense)*

then

(i) $a_{ii} \leq b_{ii} \quad i = 1, ..., n$

(ii) $tr(A) \leq tr(B)$

(iii) $|A| \leq |B|$

Remark Therefore, A is P.D.S. (resp. P.S.D.S.) may be written as $A > 0$ (resp. $A \geq 0$)

6.6 Diagonalisation of symmetric matrices

In this section, we only consider symmetric matrices :$A = A'$

Theorem 6.6.1 *(One symmetric matrix)*

Let $A = A' : n \times n$
$\lambda_i (i = 1, ..., m)$ *the* m *distinct eigenvalues of* A
E_i $(i = 1, ..., m)$ *the subspaces generated by the eigenvectors associated to* λ_i

then *(i)* $i \neq j \Rightarrow E_i \perp E_j$
(ii) $dim E_i =$ *multiplicity of* $\lambda_i = n_i$ *and* $\sum_{i=1}^{m} n_i = n$
(iii) $\mathbb{R}^n = E_1 \oplus E_2 \oplus ... \oplus E_m$

Building an orthogonal matrix Q the columns q_i of which form an orthonormalized basis for each E_i and a diagonal matrix Λ with the eigenvalues (repeated according to their respective multiplicities) the following corollary is obtained.

Corollary

To each symmetric matrix $A = A'$, may be associated an orthogonal matrix Q and a diagonal matrix Λ such that $Q'A\,Q = \Lambda$ or, equivalently, $A = Q\,\Lambda\,Q' = \sum_{i=1}^{n} \lambda_i q_i q_i'$.
In other words, any symmetric matrix is similar to a diagonal matrix.

This corollary may also be written as :

$$A = \sum_{i=1}^{m} \lambda_i Q_i$$

where m is the number of different eigenvalues, Q_i is the projector onto the corresponding invariant subspaces E_i, and may therefore be written as:

$$Q_i = \sum_{1 \leq j \leq n_i} q_{ij} q'_{ij}$$

with $n_i = \dim(E_i) = \mathrm{r}(Q_i)$ and $\{q_{ij} \,:\, 1 \leq j \leq n_i\}$ is an orthonormal basis of E_i. One may also define:

$$Q_i^* = [q_{i1}, q_{i2}, \cdots, q_{i,n_i}] \qquad n \times n_i$$

then:

$$Q_i^* Q_i^{*\prime} = Q_i \qquad Q_i^{*\prime} Q_i^* = I_{(n_i)}$$

Theorem 6.6.2 *(Two Symmetric Matrices)*

If	$A = A' \quad n \times n$
	$B = B' \quad n \times n$ *and* $B > 0$
then	$\exists R$ *regular and* $\Lambda = diag(\lambda_1, ..., \lambda_n)$ $\lambda_i \in I\!R$
	such that $A = R'\Lambda R$
	$B = R'R$
If, furthermore,	$A \geq 0$
then	$\lambda_i \geq 0$
If, furthermore,	$B \geq A \geq 0$
then	$0 \leq \lambda_i \leq 1$

Remark Partitioning R according to the columns : $R = (r_1, ..., r_n)$, the conclusions of Theorem 6.6.2 may be written as follows:

$$A = \sum_{i=1}^{n} \lambda_i r_i r'_i$$

$$B = \sum_{i=1}^{n} r_i r'_i$$

In terms of quadratic forms, one gets:

$$\left.\begin{array}{lllll} x'A\,x & = & y'\Lambda y & = & \sum_{i=1}^{n} \lambda_i y_i^2 \\ x'B\,x & = & y'y & = & \sum_{i=1}^{n} y_i^2 \end{array}\right\} \text{with } y = Rx$$

Theorem 6.6.3 *(Commuting matrices)*

If $A_i = A_i' \qquad n \times n \qquad i = 1, ..., k$

then

$$\left[\begin{array}{ll} \exists Q & n \times n \textit{ such that :} \\ & Q'Q = I_{(n)} \\ & Q'A_iQ = diag(\lambda_{ij}, ..., \lambda_{in}) \\ & \qquad i = 1, ..., k \end{array}\right] \begin{array}{l} \\ \Leftrightarrow A_iA_j = A_jA_i \\ \Leftrightarrow A_iA_j = (A_iA_j)' \\ \\ \end{array}$$

6.7 Projections and Idempotent matrices

Let V be a vector space (on $\mathbb{R}$)

Definition $P : V \to V$ is a **projection**

$$\begin{array}{ll} \text{iff} & \text{(i)} \quad P \text{ is linear} \\ & \text{(ii)} \quad P^2 = P \end{array}$$

Let $E_1 = Im(P) \equiv \{x \in V | \exists y \in V : \ x = Py\}$: sub-vector space of V
$E_0 = Ker(P) = \{x \in V | Px = 0\}$: sub-vector space of V

Theorem 6.7.1 *(Properties of a Projection)*

Let $P : V \to V$*, projection*

then

i) $I - P$ *is a projection*

ii) $P(I - P) = (I - P)P = 0$

iii) $V = E_1 \bigoplus E_0$ *and therefore*$E_1 \bigcap E_0 = \{0\}$

iv) $Im(P) = Ker(I - P) = E_1$ *i.e.* , $x \in E_1 \Longleftrightarrow Px = x$;
$Ker(P) = Im(I - P) = E_0$ *i.e.* , $x \in E_0 \Longleftrightarrow (I - P)x = x$.

If, furthermore, dim $V = n < \infty$
then, $|A - \lambda I| = (-1)^n \lambda^{n-r} (\lambda - 1)^r$
where $r = dimE_1 = r(P), \;\; n - r = dimE_0 = r(I - P)$.

Remarks

1) Part iii) of the theorem shows that any vector $x \in V$ may be uniquely decomposed as follows: $x = x_1 + x_2$ with $x_1 = Px \in E_1$ and $x_2 = (I - P)x \in E_0$. We shall also say that P **projects onto** E_1 **parallely to** E_0.

2) In the finite-dimensional case, typically $V = I\!R^n$,:

(i) the operator P may be identified with a square matrix such that $P^2 = P$ and we shall write:
$E_1 = \mathcal{C}(P) =$ the space generated by the columns of $P = \{x \in I\!R^n | x = Py\}$
$E_0 = \mathcal{N}(P) =$ the nullity space of the matrix $P = \{x \in I\!R^n | Px = 0\} = Ker(P)$.
(ii) By the above theorem : $r(P) = tr(P)$

Let us now consider v, an **inner product** on V, *i.e.* :

$$
\begin{array}{llll}
v: & V \times V \to I\!R & \text{bilinear :} & v(\alpha\, x_1 + \beta\, x_2, y) \; = \; \alpha v(x_1, y) \; + \; \beta v(x_2, y) \\
 & & \text{symmetric :} & v(x, y) \; = \; v(y, x) \\
 & & \text{positive :} & v(x, x) \geq 0 \quad \forall x \\
 & & & v(x, x) = 0 \Leftrightarrow x = 0
\end{array}
$$

Remark

If $V = I\!R^n$, either $v(x, y) = x'y$ or $v(x, y) = x'Ay$ for a given (P.D.S) A matrix (corresponding to a chosen basis) will be considered. A change of the inner product then corresponds to a change of basis.

To v is associated:

- a notion of **norm**, *i.e.* of length of a vector: $||x|| = [v(x,x)]^{1/2}$

- a notion of **distance** between two vectors: $d(x,y) = ||x-y|| = [v(x-y, x-y)]^{1/2}$

- a notion of **transposition** of a linear opérator L : L' is defined by $v(Lx,y) = v(x, L'y) \quad \forall\, x, \forall\, y.$

- a notion of **symmetry** : $L = L'$

- a notion of **angle** among vectors : $cos(x,y) = \dfrac{v(x,y)}{||x||.||y||}$

- a notion of **perpendicularity** : $x \perp y \Leftrightarrow v(x,y) = 0$
  we shall say that: $x \, and \, y$ are mutually **orthogonal** (relatively to v).

The concept of orthogonality may be extended to the subsets of V:

$$\begin{array}{ll} \text{Let} & A, B \subseteq V \\ \text{then} & A \perp B \Leftrightarrow [x \in A \,\text{and}\, y \in B \Rightarrow x \perp y] \\ & A^{\perp} = \{x \in V | y \in A \Rightarrow x \perp y\} \end{array}$$

$A^{\perp}$ is called the **orthogonal complement** of A: this is the s.v.s. of the vectors orthogonal to **all** the vectors of A.

Theorem 6.7.2

$$\begin{array}{ll} \textit{Let} & dimV < \infty \\ & E \;\; a \;\, s.v.s. \;\; of \, V \\ \textit{then} & dimE + dim\, E^{\perp} = dimV \end{array}$$

Remark : Unless mentionned otherwise, we assume in the sequel that $V(x,y) = x'y$.

Definition $P : V \to V$ is an **orthogonal projection** if P is a projection and is symmetric *i.e.* $P = P' = P^2$.

When $dimV < \infty$ and P is represented by a matrix, one speaks of an **idempotent matrix** .

Remark : In these notes, the concept of an idempotent matrix includes therefore the property of symmetry. This is not always the case.

When the projection P is orthogonal, one has: $E_0 = E_1^{\perp}$. Any vector $x \in V$ may then be decomposed into two mutually orthogonal components:

$$x = x_1 + x_2 \text{ with} : x_1 = Px \quad x_2 = (I - P)x \quad x_1 \perp x_2$$

Such a decomposition is unique.

Examples Let $V = \mathbb{R}^2$

$$P = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \text{represents an orthogonal projection}$$

$$P = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \text{represents a non-orthogonal projection}$$

Notation P_A denotes the orthogonal projection onto A, when A is a s.v.s. or onto $\mathcal{C}(A)$ (the space genrated by the columns of A) when A is an $n \times m$ matrix ; this is also the projection onto $Im(A)$ when $A : \mathbb{R}^m \to \mathbb{R}^m$,linear.

Theorem 6.7.3 *(Properties of orthogonal projections)*

Let A and B: s.v.s. of V

1. *then the following three properties are equivalent:*
   (i) $A \perp B (et\, therefore\, A \subseteq B^{\perp} and B \subseteq A^{\perp})$

   (ii) $P_A P_B = P_B P_A = 0$

   (iii) $P_A + P_B = P_{A+B}$ *($A + B$ is also the s.v.s. generated by $A \cup B$)*

2. *then the following three properties are equivalent:*

(i) $B \subset A$	*(i bis)* $A^{\perp} \subset B^{\perp}$
(ii) $P_A P_B = P_B P_A = P_B$	*(ii bis)* $P_{A^{\perp}} P_{B^{\perp}} = P_{B^{\perp}} P_{A^{\perp}} = P_{A^{\perp}}$
(iii) $P_A - P_B = P_{A \cap B^{\perp}}$	*(iii bis)* $P_{B^{\perp}} - P_{A^{\perp}} = P_{A \cap B^{\perp}}$

Besides being symmetric, the idempotent matrices enjoy noticeable properties, derived from the previous results; for convenience, they are gathered below.

Theorem 6.7.4 *(Properties of idempotent matrices)*

Let $A, B, R : n \times n$
A *and* B : *idempotent and* $r(A) = r \leq n$
R :*orthogonal*

then 1) $\phi_A(\lambda) \equiv |A - \lambda I| = (-1)^n \lambda^{n-r} (\lambda - 1)^r$

2) $r(A) = tr(A) =$ *number of eigenvalues equal to 1*

3) $r(A) = n \Rightarrow A = I_{(n)}$

$r(A) < n \Rightarrow A$ *is S.P.D.S.*$\Rightarrow a_{ii} \geq 0 \, \forall i$

$\Rightarrow \exists \, C$ *orthogonal:* $C'AC = \begin{bmatrix} I_{(r)} & 0 \\ 0 & 0 \end{bmatrix}$

4) $a_{ii} = 0 \Rightarrow a_{ij} = a_{ji} = o \quad j = 1, ..., n$

5) $R' A \, R$ *is idempotent*

6) $I - A$ *is idempotent and* $A(I - A) = (I - A)A = 0$
$r(I - A) = n - r$

7) $AB = BA \Rightarrow AB$ *idempotent*

8) $A + B$ *idempotent* $\Leftrightarrow AB = BA = 0$

9)$A - B$ *idempotent* $\Leftrightarrow AB = BA = B$

Theorem 6.7.5 *(Sum of Projections)*
Let $A_i = A_i' \quad (n \times n) \quad i = 1, ..., k$

then *a) Any two of the following properties imply the third one:*

1) $A_i = A_i^2 \quad i = 1, ..., k$
2) $\sum_{i=1}^{k} A_i = (\sum_{i=1}^{k} A_i)^2$
3) $A_i A_j = 0 \quad i \neq j$

b) furthermore, any two of these properties imply:
$r(\sum_{i=1}^{k} A_i) = \sum_{i=1}^{k} r(A_i)$

6.8 Generalized Inverse

Theorem 6.8.1 *(Moore-Penrose generalized Inverse)*

Let $A : n \times m$ *with any rank*

then there exists a ***unique*** $m \times n$ *matrix, denoted* A^+*, such that:*

(i) $AA^+A = A$

(ii) $A^+AA^+ = A^+$

(iii) $AA^+ = (AA^+)'$

(iv) $A^+A = (A^+A)'$

Definition A^+ is called the **"(Moore-Penrose) generalized inverse "** of A

Properties of the generalized inverse :

(i) $(A')^+ = (A^+)'$

(ii) $(A^+)^+ = A$

(iii) $(A^+A)^r = A^+A \quad r \geq 1, integer$

(iv) $(AA^+)^r = AA^+ \quad r \geq 1, integer$

Remarks: 1) In general : $(AB)^+ \neq B^+A^+$
2) The properties (ii) and (iv) show that AA^+ and A^+A are idempotent matrices. With the notation of section 3, we have :
$AA^+ = P_A$: projection onto $\mathcal{C}(A)$
$A^+A = P_{A'}$: projection onto $\mathcal{C}(A')$

Construction of the generalized inverse

Method 1

1) If $A : n \times n$ and $r(A) = n$ then $A^+ = A^{-1}$
2) If $A = diag(a_1, ..., a_n)$
then $A^+ = diag(a_1^+, ..., a_n^+)$ where $a_i^+ \quad = a_i^{-1}$ if $a_1 \neq 0$
$= 0$ if $a_i = 0$
3) If $A = A' \quad n \times n, r(A) = r \leq n$
let $A = Q\Lambda Q' \quad Q'Q = I_{(n)}$ $\quad Q$: eigenvectors of A
$\Lambda = diag(\lambda_1, ..., \lambda_n)$
$|A - \lambda_i I_{(n)}| = 0$
then $A^+ = Q\Lambda^+Q'$
$= Q_1\Lambda_1^{-1}Q_1' \quad Q$: eigenvectors associated to $\lambda_i \neq 0$
and $\Lambda_1 : diag(\lambda_1, ..., \lambda_r) \quad \lambda_i \neq 0$
4) If $A \quad n \times m$ of any rank
then $A^+ = (A'A)^+A'$

Method 2

Let $A = n \times m \; (m \geq n)$, one may always write :
$A = \Gamma(D_\lambda : 0)\Delta'$

where Γ: eigenvectors of AA' $(\Gamma\Gamma' = I_{(n)})$;
Δ: eigenvectors of $A'A$ $(\Delta'\Delta = I_(m))$
$\mathrm{D}_\lambda = diag(\lambda_1, ..., \lambda_n)$
λ_i^2 : eigenvalues of AA'
λ_i positive root of λ_i^2
$0 : n \times (m - n)$

then $A^+ = \Delta[\mathrm{D}_\lambda^+ : 0]\Gamma'$

These two methods do NOT describe efficient numerical algorithms; they give a constructive proof of the existence (but not of the unicity!) of a Moore-Penrose generalized inverse, along with some analytic characterizations.

Generalized Inverse and systems of linear equations

Theorem 6.8.2 *(General Solution of a Linear System)*

There exists a solution to the equation $AXB = C$ iff : $C = AA^+CB^+B$; in such a case, any solution may be written in the form:

$$X = A^+C\,B^+ + Y - A^+A\,Y B\,B^+$$

where Y is an arbitrary matrix (of suitable dimension).

Theorem 6.8.3

Let $A : n \times m \quad r(A) = m$

then any solution of the equation $XA = I_{(m)}$ may be written in the form: $X = (A'MA)^{-1}A'M$ where M is any matrix such that :

i) $M = M' : n \times n$

ii) $r(A'MA) = m$

Sum of Quadratic Forms

Let $Q_i = (x - m_i)' A_i(x - m_i) \quad A_i = A_i' : n \times n \quad i = 1, 2$
$A_* = A_1 + A_2$
$A_*^- =$ any solution of $A_* A_*^- A_* = A_*$
$m_* =$ any solution of $A_* m_* = A_1 m_1 + A_2 m_2$

General case : $Q_1 + Q_2 = (x - m_*)' A_*(x - m_*) + m_1' A_1 m_1 + m_2' A_2 m_2 - m_* A_* m_*$
If $r(A_*) = n$: $Q_1 + Q_2 = (x - m_*)' A_*(x - m_*) + (m_1 - m_2)' A_1 A_*^{-1} A_2 (m_1 - m_2)$
If $r(A_1) = r(A_2) = n \qquad A_1 A_*^{-1} A_2 = (A_1^{-1} + A_2^{-1})^{-1} = A_2 A_*^{-1} A_1$

Remark : In general : $m_*' A_* m_* = (A_1 m_1 + A_2 m_2)' A_*^- (A_1 m_1 = A_2 m_2)$

6.9 Trace, Kronecker Product, Stacked Matrices

Theorem 6.9.1 *(Alternative Definitions of the Trace)*

For any square symmetric matrix, the sum of the elements of the main diagonal is equal to the sum of the eigenvalues (taking multiplicities into account) and define the **trace** *of the matrix, namely :*

$$tr(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i \, r_i$$

where the λ_i's are the different eigenvalues and r_i are their respective multiplicities

Theorem 6.9.2 *(Properties of the trace)*

Let $A : (n \times n) \quad B : (n \times n) \quad C : (n \times r) \quad D : (r \times n) \;\; c \in I\!R$
then

(i) tr(.) is a linear function defined on the square matrices : $tr(A + B) = tr(A) + tr(B)$ *and* $tr(cA) = c\, tr(A)$

(ii) $tr(I_{(n)}) = n$

(iii) $trA = trA'$

(iv) $tr(CD) = tr(DC)$

Definition *Let* $A : (r \times s) \qquad B : (u \times v)$
then *(Kronecker product)*
$A \otimes B = [a_{ij}B] : (ru \times sv)$

Theorem 6.9.3 *(Properties of Kronecker product)*

(i) $(A \otimes B).(C \otimes D) = AC \otimes BD$ *(**p**rovided the products are well defined!)*

(ii) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

(iii) $(A \otimes B)' = A' \otimes B'$

(iv) $(A \otimes B) \otimes C = A \otimes (B \otimes C) = A \otimes B \otimes C$

(v) $(A + B) \otimes (C + D) = (A \otimes C) + (A \otimes D) + (B \otimes C) + (B \otimes D)$

(vi) $Ax_i = \lambda_i x_i$ *and* $By_j = \gamma_j y_j \Rightarrow (A \otimes B)(x_i \otimes y_j) = \lambda_i \gamma_j (x_j \otimes y_j)$
in particular :
$|A - \lambda_i I_{(n)}| = |B - \gamma_j I_{(n)}| = 0 \Rightarrow |A \otimes B - \lambda_i \gamma_j I_{(m.n)}| = 0$

(vii) $|A \otimes B| = |A|^m |B|^n$ *where* $A : n \times n \quad B : m \times m$

(viii) $tr(A \otimes B) = tr(A).tr(B)$

(ix) $a \otimes b' = b' \otimes a = ab'$ *where* $a : n \times 1 \quad b : m \times 1$

(x) $\alpha A \otimes \beta B = \alpha\beta(A \otimes B)$ *where* α *and* $\beta \in \mathbb{R}$

Definition Let $A = [a_1, a_2, ..., a_p] : n \times p$
$a_i \in \mathbb{R}^n$

then (**stacked matrices** - or : stacking operator):

$$A^v = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} : np \times 1$$

Theorem 6.9.4 *(Relationships between trace, stacked matrices and Kronecker product)*

Let $X : (n \times p)$ $\quad Y : (n \times p)$ $\quad P : p \times q$

$Q : (q \times r)$ $\quad D : (p \times p)$ $\quad V : (n \times n)$

$a : (n \times 1)$ $\quad b : (n \times 1)$

then *(i)* $[XPQ]^v = [Q' \otimes X]P^v$

(ii) $tr(XY') = tr[X^v(Y^v)'] = (Y^v)'X^v$

(iii) $tr(X'VXD) = (X^v)'(D \otimes V)X^v$

(iv) $(a \otimes b')^v = b \otimes a$

(v) $a'XP = (X^v)'(P \otimes a) = (X'^v)'(a \otimes P)$

Remark : in general $(A')^v \neq (A^v)'$

Chapter 7

Complements on Linear Regression (Appendix B)

7.1 Linear restrictions in Linear models

7.1.1 Estimating under Linear restrictions

Let a multiple regression model

$$y \;=\; Z\beta \;+\; \epsilon \tag{7.1}$$

with $y : (n \times 1), Z : (n \times k)$. In general, s linear restrictions such as :

$$A\,\beta \;=\; a_0 \qquad \text{with :}\;\; A\;: s \times k \;\; \text{and} \;\; r(A) = s \tag{7.2}$$

are equivalently written as:

$$\beta \;=\; G\,\gamma \,+\, g_0 \qquad \text{with :}\;\; g_0 = \,A'(A\,A')^{-1}a_0 \tag{7.3}$$

where G is any matrix such that:

$$G\;: k \times r,\;\; r(G) \;=\; r,\;\; r+s \;=\; k, \;\text{and}\;\; A\,G \;=\; 0\;\;(s \times r)$$

and γ represents r free coefficients under the s linear restrictions. Therefore, the BLUE estimator of:

$$y \;=\; Z\,\beta \;+\; \epsilon \quad V(\epsilon) \;=\; \sigma^2 I_{(n)} \quad \text{subject to :}\;\; A\,\beta \;=\; a_0$$

may be written as

$$b \;=\; g_0 \;+\; G\,(Z_*'Z_*)^{-1}Z_*'y_*$$

with :

$$y_* = y - Z\,g_0 \qquad Z_* = Z\,G \;\; (n \times r)$$

This estimator amounts to an OLS regression of y_* on Z_*.

7.1.2 Regression with singular residual covariance matrix

Let a multiple regression model with a singular residual covariance matrix:

$$y = Z\,\beta + \epsilon \qquad V(\epsilon) = \sigma^2\,\Omega \qquad r(\Omega) = r < n$$

with $y : (n \times 1), Z : (n \times k), \beta : (k \times 1)$. There exists a matrix T:

$$T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \qquad \text{with : } \;\; T : n \times n, \quad T_1 : r \times n, \quad T_2 : (n-r) \times n$$

such that each of these matrices have full (row) rank and:

$$T\,\Omega\,T' = \begin{bmatrix} I_{(r)} & 0 \\ 0 & 0 \end{bmatrix}$$

Thus the BLUE estimator of β may be obtained through :

$$OLS[T_1\,y \;:\; T_1\,Z\,] \qquad \text{subject to : } \; T_2\,y = T_2\,Z\,b$$

Note that model (7.1) and (7.2) is equivalently written as:

$$\begin{pmatrix} y \\ a_0 \end{pmatrix} = \begin{pmatrix} Z \\ A \end{pmatrix} \beta + \varepsilon_* \qquad V(\varepsilon_*) = \sigma^2 \begin{pmatrix} I_{(n)} & 0 \\ 0 & 0 \end{pmatrix}$$

7.1.3 Testing Linear restrictions

Consider now a normal multiple regression model

$$y = Z\beta + \epsilon \qquad \epsilon \sim N(0, \sigma^2\,I_{(n)}\,)$$

subject to the hypotheses:

$$H_0 : \;\; A\,\beta = 0 \qquad \text{or, equivalently: } \; \beta = G\,\gamma$$

against the complementary alternative:

$$H_1 : \ A\beta \neq 0$$

Define the residual sums of squares, relatively to each hypotheses:

$$S_i^2 \ = \ y' M_i y \qquad i = 0, 1$$

$$M_0 \ = \ I_{(n)} \ - \ Z_* Z_*^+ \qquad \text{with} : \ Z_* \ = \ Z\,G$$

$$M_1 \ = \ I_{(n)} \ - \ ZZ^+$$

Under the null hypothesis, the statistic:

$$F \ = \ \frac{(S_0^2 - S_1^2)/(l_0 - l_1)}{S_1^2)/l_1},$$

where $l_i \ = \ r(M_i)$, is distributed as an F-distribution with $(l_0 - l_1, l_1)$ degrees of freedom.

7.2 Decomposition of a Linear regression

Let a multiple regression model with a partitionned matrix of exogenous variables:

$$y \ = \ Z\beta \ + \ \epsilon \ = \ Z_1\beta_1 \ + \ Z_2\beta_2 + \ \epsilon$$

with $y : (n \times 1), Z : (n \times k), Z_i : (n \times k_i), \beta_i : (k_i \times 1)$ and $k_1 + k_2 = k$, and let

$$b \ = \ (Z'Z)^{-1}Z'y \ = \ \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

be the Ordinary Least Squares estimator of β, also in partitionned form, then

$$b_1 \ = \ (Z_1' M_2 Z_1)^{-1} Z_1' M_2 y \tag{7.4}$$

where:

$$M_2 \ = \ I_{(n)} \ - \ Z_2(Z_2'Z_2)^{-1}Z_2'$$

is the projection of $\mathbb{R}^n$ onto the orthogonal complement of the subspace generated by the columns of Z_2. Thus, for instance, $M_2 y$ is the vector of the residuals of the regression of y on the Z_2. Note that when k is large, this formula requires the inversion of a symmetric matrix $(k_1 \times k_1)$ and the inversion of another symmetric matrix $(k_2 \times k_2)$, rather than the inversion of a symmetric matrix $(k \times k)$

7.3 Equivalence between OLS and GLS

Let a multiple regression model with non -spherical residuals:

$$y = Z\beta + \epsilon \qquad V(\epsilon) = \sigma^2\,\Omega$$

with $y : (n \times 1), Z : (n \times k), \beta : (k \times 1)$. Let us compare the Ordinary Least Squares and the Genereralized Least Squares estimators. When Z has full rank, *i.e.* $r(Z) = k$, we have:

$$b_{OLS} = (Z'Z)^{-1}Z'y \tag{7.5}$$
$$b_{GLS} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y \tag{7.6}$$

More generally, when $r(Z) = r \leq k$, the estimable linear combinations of β may be represented as:

$$\lambda = a'Z\beta \qquad a \in I\!R^n \text{ arbitrary known}$$

i.e. these are the linear combinations the vector of coefficients of which are generated by $\mathcal{C}(Z')$, the row space of Z.

Theorem 7.3.1 *The following conditions are equivalent:*

$$b_{OLS} = b_{GLS} \tag{7.7}$$
$$\mathcal{C}(Z) \quad \textit{is an invariant subspace of } \mathcal{C}() \tag{7.8}$$
$$\textit{there exist } r \quad \textit{eigenvectors of} \Omega \textit{ providing a basis for } \mathcal{C}(Z) \tag{7.9}$$
$$\Omega = I_n + ZCZ' + WDW' \quad \textit{with } C \textit{ and } D \textit{ arbitrary, and } W \textit{ such that} Z'W = 0. \tag{7.10}$$
$$\exists \; T \quad (k \times k) \;\; \textit{such that } \; \Omega Z = ZT \tag{7.11}$$
$$\Omega P = P\Omega \tag{7.12}$$
$$(\Omega P)' = \Omega P \tag{7.13}$$

where P *is the orthogonal projector on* $\mathcal{C}(Z)$*:*

$$P = ZZ^{+}$$

Corollary
1. If furthermore $X = \Omega X$, then $X'X = X'\Omega_{-1}X$, *i.e.* in such a case, the

non -sphericity of the residuals has no impact on the precision of the GLS estimator.
2. If $Z = (\iota\ Z_2)$, the OLS estimator of β_2 is equivalent to the GLS estimator of β_2 for *any* Z_2 if and only if Ω has the form $\Omega = (1\rho)\, I_n + \rho\, \iota\iota'$

Exercise. Check that (7.4) may also be viewed as an OLS estimator on the transformed model:

$$M_2\, y = M_2\, Z\beta + M_2\, \epsilon \qquad V(M_2\, \epsilon) = \sigma^2\, M_2 \tag{7.14}$$

therefore with non-spherical residuals. Check however that the condition for equivalence between OLS and GLS is satisfied in (7.14). ■

7.4 Prediction under non-spherical residuals

Let us start again with a multiple regression model, in a time-series context of T observations, with non -spherical residuals:

$$y = Z\beta + \epsilon \qquad V(\epsilon) = \sigma^2\, \Omega$$

with $y : (T \times 1)$, $Z : (T \times k)$, $\beta : (k \times 1)$. We now want to construct a predictor for the next S $(S \geq 1)$ post-samples values of y, namely $y_F = (y_{T+1}, y_{T+2}, \cdots, y_{T+S})'$, relatively to the corresponding future values of the exogenous variables, specified in a matrix $Z_F : S \times k$ and taking into account the possible covariances between the past and the future values of y, specified in an $S \times T$- matrix $\Gamma_F = [cov(y_{T+i}, y_t)] \quad 1 \leq i \leq S,\ 1 \leq t \leq T$. If we want to limit the attention to simple procedures, we only consider predictors that are linear in the past observations:

$$\hat{y}_F = Ay \qquad A : S \times T$$

where the elements of the matrix A are typically function of Z and of Z_F but not of y. For the same reason of simplicity, we also want an unbiased predictor, in the sense:

$$E(\hat{y}_F \mid Z,\, Z_F,\, \theta) = E(y_F \mid Z,\, Z_F,\, \theta).$$

We finally want the predictor to be "optimal" , in the sense of minimizing, in the restricted class, the variance of any linear combination of the predictor.

Supposing Γ_F and Ω known, we obtain the Best Linear Unbiased Predictor (BLUP) as follows:

$$\hat{y}_F \;=\; Z_F\, b_{GLS} \;+\; \Gamma_F\, \Omega^{-1}\, (y \;-\; Z\, b_{GLS})$$

This predictor may be viewed as the best linear estimator of $E(y_F \mid Z,\, Z_F,\, \theta)$ corrected by the best linear predictor of the *conditional* expectation of the future residual $E(\varepsilon_F \mid \varepsilon \;=\; y - Z\, b_{GLS},\, Z,\, Z_F,\, \theta)$, provided that $(Z, Z_F) \perp\!\!\!\perp (\varepsilon,\, \varepsilon_F)$.

7.5 Spectral Decomposition of the Residual Covariance Matrix

7.5.1 The model

Let us consider a multiple regression model with non-spherical residuals:

$$y \;=\; Z\,\beta \;+\; \epsilon : \qquad (7.15)$$

$$V(\epsilon) \;=\; \Omega \qquad (7.16)$$

with $y : (n \times 1), Z : (n \times k), \beta : (k \times 1)$. Assume Ω non-singular and consider the spectral decomposition of Ω:

$$\Omega \;=\; \sum_{1 \leq i \leq p} \lambda_i Q_i \qquad (7.17)$$

where p is the number of different eigenvalues of Ω and (see section 6.6):

- $\lambda_i > 0$
- $Q_i = Q_i' = Q_i^2$
- $Q_i\, Q_j = 0 \qquad j \neq i$
- $r(Q_i) = tr\, Q_i = r_i \qquad \sum_{1 \leq i \leq p} r_i = n$
- $\sum_{1 \leq i \leq p} Q_i = I_{(n)}$
- $\Omega\, Q_i \;=\; Q_i\, \Omega \;=\; \lambda_i\, Q_i$

Note that

$$\forall r \in \mathbb{R} \;:\quad \Omega^r \;=\; \sum_{1\leq i\leq p} \lambda_i^r Q_i$$

Furthermore, if we define the $np \times n$-matrix Q as follows:

$$Q \;=\; \begin{pmatrix} Q_1 \\ Q_2 \\ \cdots \\ Q_p \end{pmatrix}$$

then:

$$V(Q\epsilon) \;=\; diag(\lambda_i Q_i) \tag{7.18}$$

In this section, we shall always assume that the projectors Q_i are perfectly known and that only the characteristic values λ_i are functions of unknown parameters. We shall consider the estimation of the regression coefficients β and of the eigenvalues $\lambda_i,\ i = 1,\ \cdots,\ p$.

7.5.2 Moment Estimation of the regression coefficients

Once the residuals are non-spherical, i.e. $V(\epsilon)\ =\neq\ \sigma^2\, I_{(n)}$, the BLU estimator of β may be obtained by means of the Generalized Least Squares:

$$\hat{\beta}_{GLS} \;=\; (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y \tag{7.19}$$

As Ω has typically large dimension (that of the sample size), its inversion is often more efficiently operated by making use of its spectral decomposition (7.17), namely:

$$\begin{aligned}\hat{\beta}_{GLS} \;&=\; [Z'(\sum_{1\leq i\leq p} \lambda_i^{-1} Q_i)Z]^{-1}[Z'(\sum_{1\leq i\leq p} \lambda_i^{-1} Q_i)y] \\ &=\; [\sum_{1\leq i\leq p} \lambda_i^{-1} Z'Q_iZ]^{-1}[\sum_{1\leq i\leq p} \lambda_i^{-1} Z'Q_iy] \end{aligned} \tag{7.20}$$

Consider now the OLS estimators of Q_iy on Q_iZ:

$$b_i \;=\; [Z'Q_iZ]^{-1}\, Z'Q_iy \qquad \text{when } \; r_i \geq k \tag{7.21}$$

$$\;=\; \text{any solution of } \; [Z'Q_iZ]\ b_i \;=\; Z'Q_iy \quad \text{, in general} \tag{7.22}$$

and define:

$$W_i = \lambda_i^{-1} Z' Q_i Z \tag{7.23}$$

then:

$$W_i\, b_i = \lambda_i^{-1} Z' Q_i y \tag{7.24}$$

$$\hat{\beta}_{GLS} = [\sum_{1 \le i \le p} W_i]^{-1} [\sum_{1 \le i \le p} W_i b_i]$$

$$= \sum_{1 \le i \le p} W_i^* b_i \qquad \text{where } W_i^* = [\sum_{1 \le i \le p} W_i]^{-1} W_i \tag{7.25}$$

Thus, the GLS estimator may be viewed as a matrix weighted average of the OLS estimators of the transformed models

$$Q_i\, y = Q_i\, Z\beta + Q_i\, \epsilon \tag{7.26}$$

$$V(Q_i\, \epsilon) = \lambda_i\, Q_i \tag{7.27}$$

Exercise. Check that these OLS estimators are equivalent to the GLS estimators in the same models. ■

We have in particular that if:

$$Q_1 = \bar{J}_n = \frac{1}{n}\iota\iota' \qquad r_1 = r(Q_1) = 1 \tag{7.28}$$

then:

$$\epsilon' Q_1\, \epsilon = n\bar{\epsilon}^2 \tag{7.29}$$

$$E(\epsilon' Q_1\, \epsilon) = \lambda_1 \tag{7.30}$$

therefore:

$$E[\bar{\epsilon}^2] = V(\bar{\epsilon}) = \frac{\lambda_1}{n} = \frac{\iota'\Omega\iota}{n^2} \tag{7.31}$$

$$\lambda_1 = \frac{\iota'\Omega\iota}{n} \tag{7.32}$$

Furthermore:

$$Q_1\, y = \iota\bar{y} \qquad (n \times 1) \tag{7.33}$$

$$Q_1\, Z = \iota\bar{z}' \qquad (n \times k) \tag{7.34}$$

$$W_1 = \lambda_1^{-1} n\bar{z}\bar{z}' \qquad (k \times k) \tag{7.35}$$

$$W_1\, b_1 = \lambda_1^{-1} n\bar{z}\bar{y} \qquad (k \times 1) \tag{7.36}$$

where $\bar{z}$ is the k-vector of column averages of Z; therefore:

$$\bar{z}\bar{z}' b_1 \;=\; \bar{z}\bar{y} \tag{7.37}$$

If, furthermore,

$$Z \;=\; [\iota \;\; Z_2] \qquad Z_2 \;:\; n \times (k-1), \tag{7.38}$$

and :

$$b_1 \;=\; \begin{bmatrix} b_{1,1} \\ b_{2,1} \end{bmatrix} \qquad b_{2,1} \;:\; (k-1)\times 1 \qquad \bar{z} \;=\; \begin{bmatrix} 1 \\ \bar{z}_2 \end{bmatrix}$$

then the solution of (7.37) leaves $b_{2,1}$ arbitrary, provided:

$$b_{1,1} \;=\; \bar{y} \;-\; b'_{2,1}\bar{z}_2 \tag{7.39}$$

Define:

$$M_1 \;=\; I_{(n)} \;-\; Q_1.$$

Note that:

$$M_1\, Q_1 \;=\; Q_1\, M_1 \;=\; 0 \tag{7.40}$$

$$M_1\, Q_i \;=\; Q_i\, M_1 \;= Q_i \qquad \forall\, i \neq 1 \tag{7.41}$$

$$M_1\, Z \;=\; (0 \quad M_1\, Z_2) \tag{7.42}$$

Therefore:

$$M_1\, y \;=\; M_1\, Z_2\, \beta_2 \;+\; M_1\, \epsilon\,, \tag{7.43}$$

with

$$V(M_1\, \epsilon) \;=\; \sum_{2\leq i\leq p} \lambda_i\, Q_i \;=\; \Omega - \lambda_1\, Q_1 \;=\; \Omega_* \;\; \text{, say },$$

represents the regression with all the variables taken in deviation from the sample mean, for instance: $M_1\, y \;=\; [y_i \;-\bar{y}]$. Moreover, from (7.41), we also have:

$$Q_i\, M_1\, [y \quad Z] \;=\; Q_i\, [y \quad Z]$$

Therefore, (7.43) also implies:

$$Q_i\, y \;=\; Q_i\, Z_2\, \beta_2 \;+\; Q_i\, \epsilon \qquad \text{with } \; V(Q_i\, \epsilon) \;=\; \lambda_i Q_i \qquad \forall\, i \neq 1 \tag{7.44}$$

If furthermore,

$$\forall i \neq 1 \;:\; r_i \;\geq\; k$$

the BLUE estimators in (7.44) are given by:

$$b_{2,i} \;=\; [Z_2' Q_i Z_2]^{-1}\, Z_2' Q_i y \qquad \forall\, i \neq 1 \tag{7.45}$$

Therefore, under (7.28) and (7.38), b_{GLS} takes the form

$$b_{GLS,1} \;=\; \bar{y} \;-\; b'_{GLS,2}\bar{z}_2 \qquad (7.46)$$

$$\begin{aligned} b_{GLS,2} \quad &= \quad [\sum_{2\leq i\leq p} W_{2,i}]^{-1}[\sum_{2\leq i\leq p} W_{2,i}b_{2,i}] \\ &= \quad \sum_{2\leq i\leq p} W_i^{*}b_{2,i} \end{aligned} \qquad (7.47)$$

where

$$\begin{aligned} W_{2,i}\, b_{2,i} \;&=\; \lambda_i^{-1} Z_2' Q_i y \qquad \forall\, i \neq 1 \\ W_i^{*} \;&=\; [\sum_{2\leq i\leq p} W_{2,i}]^{-1} W_{2,i} \end{aligned}$$

Remark

Remember that the projection matrices Q_i are singular with rank: $r(Q_i) = r_i < n$. Thus, in the transformed model $Q_i y = Q_i Z\beta + Q_i\varepsilon$, the n observations of the n-dimensional vector $Q_i y$ actually represent r_i linearly independent observations only, instead of n. This redundance of observations may be avoided by remarking that there exists an $r_i \times n$-matrix Q_i^{*} such that:

$$(i)\;\; Q_i^{*}\, Q_i^{*'} \;=\; I_{(r_i)} \qquad (ii)\;\; Q_i^{*'}\, Q_i^{*} \;=\; Q_i \qquad (7.48)$$

where the matrix Q_i^{*} is unique up to a pre-multiplication by an arbitrary orthogonal matrix only. Note that (i) means that Q_i^{*} is made of r_i rows of an $n \times n$ (suitably specified) othogonal matrix. We therefore obtain that $GLS(Q_i y : Q_i Z)$ is equivalent to $OLS(Q_i y : Q_i Z)$, by Theorem 7.3.1, which is in turn equivalent to $OLS(Q_i^{*} y : Q_i^{*} Z)$ because of spherical residuals.

Note the following particular cases:

(i) When $Q_i \;=\; \frac{1}{n}\,\iota\iota'$, we obtain: $r(Q_i) \;=\; 1$, $Q_i^{*} \;=\; \frac{1}{\sqrt{n}}\,\iota'$ and $Q_i^{*}\, y \;=\; \sqrt{n}\,\bar{y}$

(ii) When $Q_i \;=\; I_{(n)} \;-\; \frac{1}{n}\,\iota\iota'$, we obtain: $r(Q_i) \;=\; n \;-\; 1$ and Q_i^{*} is an $(n \;-\; 1) \times n$-matrix such that $Q_i^{*}\iota \;=\; 0$, $Q_i^{*}\, Q_i^{*'} \;=\; I_{(r_i)}$ and $Q_i^{*'}\, Q_i^{*} \;=\; I_{(n)} \;-\; \frac{1}{n}\,\iota\iota'$

(iii) When $Q_i \;=\; R \otimes S$ with $Q_i^{*'}\, Q_i^{*} \;=\; Q_i$, $R^{*'} R^{*} \;=\; R$ and $S^{*'} S^{*} \;=\; S$ all defined as above, we obtain: $Q_i^{*} \;=\; R^{*} \otimes S^{*}$, with $Q_i^{*} \,:\, q \times n$, $q = rk(Q_i) \;=\; rs$, $r \;=\; rk(R)$, $s \;=\; rk(S)$.

7.5.3 Moment Estimation of the eigenvalues

Note that the estimators (7.25), or (7.46) and (7.47) are not feasible, because they involve the characteristic roots λ_i that typically depend on unknown parameters. Designing simple estimators of these characteristic roots is the object of this section.

Note first that $Q_i\varepsilon \sim (0,\ \lambda_i\, Q_i)$ implies that:

$$E(\epsilon' Q_i \epsilon) \ = \ \lambda_i\, r_i$$

Thus, *if* ε were observable, an unbiased estimator of λ_i could be:

$$L_i \ = \ \frac{\varepsilon' Q_i\, \varepsilon}{r_i}$$

Moreover, as $Q_i\varepsilon$ and $Q_j\varepsilon$ are uncorrelated, L_i and L_j are independent under a normality assumption. The fact that the vector ε is not observable leads to develop several estimators of the eigenvalues λ_i.

Let us first consider the residuals of the OLS($Q_i\ :\ Q_i Z$), namely:

$$e_{(i)} \ = \ S_{(i)}\varepsilon \ = \ S_{(i)}y \qquad \text{where:} \qquad (7.49)$$
$$S_{(i)} \ = \ Q_i \ - \ Q_i Z(Z'Q_i Z)^{-1} Z'Q_i \qquad (7.50)$$

i.e. $S_{(i)}$ is the projector on the space orthogonal to the space generated by the columns of ZQ_i.

Exercise.
(i)Check that $S_{(i)}\Omega \ = \ \Omega S_{(i)} \ = \ \lambda_i S_{(i)}$
(ii) Check that $tr\, S_{(i)} \ = \ r_i \ - \ k$, *provided* that $r_i \ < \ k \ = \ r(Z'Q_i Z)$ ■

Therefore:

$$\mathbb{E}\,[e'_{(i)} e_{(i)}] \ = \ \mathbb{E}\,[\epsilon' Q_i \epsilon] \ = \ \lambda_i\,(r_i \ - \ k)$$

Thus an unbiased estimator of λ_i is provided by :

$$\hat{\lambda}_i \ = \ \frac{y' S_{(i)} y}{r_i \ - \ k}$$

but the condition $r_i \ < \ k \ = \ r(Z'Q_i Z)$ will typically not hold when r_i is small.

Let us now consider M_i, the projector on the orthogonal complement of the invariant subspaces of Ω:

$$M_i \;=\; I_{(n)} \;-\; Q_i \tag{7.51}$$

Note that:

$$M_i\, Q_j \;=\; Q_j\, M_i \;=\; Q_j \quad j \neq i \qquad M_i\, Q_i \;=\; Q_i\, M_i \;=\; 0 \tag{7.52}$$

$$M_i\, \Omega \;=\; \Omega\, M_i \;=\; \Omega \;-\; \lambda_i\, Q_i \;=\; \sum_{j\neq i} \lambda_j\, Q_j \tag{7.53}$$

Therefore:

$$V(M_i\, \varepsilon) \;=\; \Omega \;-\; \lambda_i\, Q_i \tag{7.54}$$

$$\mathbb{E}\,(\varepsilon' M_i\, \varepsilon) \;=\; tr\, \Omega \;-\; \lambda_i\, r_i \;=\; \sum_{j\neq i} \lambda_j\, r_j \tag{7.55}$$

In the model (7.15) transformed by M_i:

$$M_i\, y \;=\; M_i\, Z\, \beta \;+\; M_i\, \epsilon \tag{7.56}$$

the residual ε may be estimated unbiasedly through the OLS residuals of the transformed regression (7.56):

$$\hat{\varepsilon}_{[i]} \;=\; R_{(i)}\, y \tag{7.57}$$

$$R_{(i)} \;=\; M_i \;-\; M_i Z (Z' M_i Z)^{-1} Z' M_i \tag{7.58}$$

where $R_{(i)}$ is the projector on the space orthogonal to the space generated by the columns of $M_i\, Z$:

$$R^2_{(i)} \;=\; R'_{(i)} \;=\; R_{(i)} \qquad\qquad R_{(i)}\, M_i Z \;=\; 0$$

Therefore:

$$V(R_{(i)}\, y) \;=\; V(R_{(i)}\, \varepsilon) \;=\; R_{(i)}\, \Omega\, R_{(i)} \tag{7.59}$$

$$\begin{aligned} \mathbb{E}\,(y' R_{(i)}\, y) \;&=\; \mathbb{E}\,(\varepsilon' R_{(i)}\, \varepsilon) \;=\; tr\; R_{(i)}\, \Omega \\ &=\; tr\; [I_{(n)} \;-\; M_i Z (Z' M_i Z)^{-1} Z' M_i\;](\sum_{j\neq i} \lambda_j\, Q_j) \\ &=\; \sum_{j\neq i} \lambda_j\; [r_j \;-\; tr\; M_i Z (Z' M_i Z)^{-1} Z'\, Q_j] \end{aligned} \tag{7.60}$$

Consider the case where $p = 2$.

$$\begin{aligned} \mathbb{E}\,(y' R_1\, y) &= \mathbb{E}\,(\varepsilon' R_1\, \varepsilon) \;=\; tr\; R_1\, \Omega \\ &= tr\; [I_{(n)} \,-\, M_1 Z (Z' M_1 Z)^{-1} Z' M_1 \;]\; \lambda_2\, Q_2 \\ &= \lambda_2\; [r_2 \,-\, tr\; M_1 Z (Z' M_1 Z)^{-1} Z'\, Q_2] \\ &= \lambda_2\; [r_2 \,-\, k] \end{aligned} \tag{7.61}$$

because, when $p = 2$, $M_1 = Q_2$, and provided that $r_2 > k = r(Z' M_1 Z)$. Thus an unbiased estimator of λ_2 may be obtained as follows

$$\hat{\lambda}_2 \;=\; \frac{y' R_1\, y}{r_2 \,-\, k} \tag{7.62}$$

and similarly for $\hat{\lambda}_1$.

More generally, for $p > 2$, (7.60) provides a linear system

$$C\, \lambda \;=\; c \qquad\qquad C \;:\; p \times p \tag{7.63}$$

where: $c_i = y' R_{(i)} y$, $c_{ii} = 0$ and $c_{ij} = [r_j \,-\, tr\; M_i Z (Z' M_i Z)^{-1} Z' M_i\, Q_j]$. Now, (7.63) may be solved in λ, easily in models where the λ_i 's are not constrained, and provided that $c_{ij} > 0$.

In several models, the eigenvalues λ_i are constrained by being a linear transformation of a smaller number of underlying parameters, typically components of variances. Let $\alpha \in I\!\!R^q_+$, with $q < p$, represent those variation-free parameters such that:

$$\lambda \;=\; A\, \alpha \qquad A \;:\; p \times q \tag{7.64}$$

A simple, least-squares, solution of (7.63) under (7.64) is given by:

$$\hat{\alpha} \;=\; (A' C' C A)^{-1} A' C' c \qquad\qquad \hat{\lambda} \;=\; A \hat{\alpha} \tag{7.65}$$

7.5.4 Maximum likelihood estimation

Under a normality assumption, the data density of model (7.15)-(7.17), is:

$$\begin{aligned} p(y \mid Z, \theta) \;=\; &(2\,\pi)^{-\frac{1}{2} NT} \;\mid \Omega \mid^{-\frac{1}{2}} \\ &\exp -\tfrac{1}{2}\,(y \,-\, Z\,\beta)'\, \Omega^{-1}\, (y \,-\, Z\,\beta) \end{aligned} \tag{7.66}$$

where $\theta = (\beta', \lambda')'$. Reminding that the determinant of Ω is the product of its characteristic values, taking into account the multiplicities, we may write $L(\beta, \lambda) = -2 \ln p(y \mid Z, \theta)$ (to be minimized) as follows:

$$\begin{aligned} L(\beta, \lambda) &= const. + \sum_{1 \leq i \leq p} r_i \ln \lambda_i \\ &\quad + \sum_{1 \leq i \leq p} \lambda_i^{-1} (y - Z\beta)' Q_i (y - Z\beta) \end{aligned} \tag{7.67}$$

Defining and reparametrizing

$$\eta_i = \lambda_i^{-1} \qquad f_i(\beta) = (y - Z\beta)' Q_i (y - Z\beta), \tag{7.68}$$

we may rewrite (7.66) as follows:

$$L(\beta, \eta) = const. - \sum_{1 \leq i \leq p} r_i \ln \eta_i + \sum_{1 \leq i \leq p} \eta_i f_i(\beta) \tag{7.69}$$

Note that $f_i(\beta)$ is the sum of the OLS square residuals of the regression (7.26)-(7.27). In some cases, there are constraints among the λ_i (or equivalently, the η_i), but the first order conditions of the *unconstrained* minimization provide a simple stepwise optimization:

$$\begin{aligned} \frac{\partial}{\partial \eta_i} L(\beta, \eta) &= -r_i (\eta_i)^{-1} + f_i(\beta) \\ &= 0 \quad \Rightarrow \quad r_i (\eta_i)^{-1} = f_i(\beta) \end{aligned} \tag{7.70}$$

Therefore:

$$\hat{\eta}_i(\beta) = \frac{r_i}{f_i(\beta)} \qquad \hat{\lambda}_i(\beta) = \frac{f_i(\beta)}{r_i} \tag{7.71}$$

Thus the concentrated log-likelihood becomes:

$$\begin{aligned} L_*(\beta) &= L(\beta, \hat{\eta}(\beta)) \\ &= const. \sum_{1 \leq i \leq p} r_i [\ln r_i - \ln f_i(\beta)] + \sum_{1 \leq i \leq p} \frac{r_i}{f_i(\beta)} f_i(\beta) \\ &= const. + \sum_{1 \leq i \leq p} r_i \ln f_i(\beta), \end{aligned} \tag{7.72}$$

the first partial derivatives of which are:

$$\frac{\partial}{\partial \beta_j} L_*(\beta) = \sum_{1 \le i \le p} r_i \frac{1}{f_i(\beta)} [\frac{\partial}{\partial \beta_j} f_i(\beta)] \tag{7.73}$$

$$\frac{d}{d \beta} L_*(\beta) = 2 \sum_{1 \le i \le p} \frac{r_i}{f_i(\beta)} [Z'Q_i Z \beta Z'Q_i y] \tag{7.74}$$

Evaluating the roots of (7.74) amounts to solve a non-linear system. Efficient numerical procedures require to make a best use of the peculiarities of the functions $f_i(\beta)$, i.e. of the particular structure of the projectors Q_i. Furthermore second-order conditions should be checked in order to ensure a global minimum rather than a local one, or a saddle-point, or a maximum!

Remark. In many models, there are restrictions among the eigenvalues, both in the form of inequalities (*i.e.* there is an order among the λ_i's) and in the form of equalities, because the p different λ_i's are functions of less than p variation-free variances.

Chapter 8

Complements of Probability and Statistics (Appendix C)

8.1 Probability: Miscellaneous results

8.1.1 Independence in probability

Marginal Independence

Theorem
Let $X = (Y, Z)$ be a random vector . The following conditions are equivalent and define "*Y and Z are independent in probability*" or "*Y and Z are stochastically independent*, to be denoted as:

$$Y \perp\!\!\!\perp Z$$

$$\begin{aligned}
\mathbb{E}\left[g(Y)\; h(Z)\right] &= \mathbb{E}\left[g(Y)\right] \mathbb{E}\left[h(Z)\right] \\
& \qquad \forall\; g(Y) \text{et } h(Z) \text{ integrable} \\
\mathbb{E}\left[g(Y) \mid Z\right] &= \mathbb{E}\left[g(Y)\right] \\
& \qquad \forall\; g(Y) \text{ integrable} && (8.1) \\
\mathbb{E}\left[h(Z) \mid Y\right] &= \mathbb{E}\left[h(Z)\right] \\
& \qquad \forall\; h(Z) \text{ integrable} && (8.2)
\end{aligned}$$

Conditional Independence

Theorem
Let $X = (W, Y, Z)$ be a random vector . The following conditions are equivalent and define *Y and Z are independent in probability, or: stochastically independent, conditionally on W*", to be denoted as:

$$Y \perp\!\!\!\perp Z \mid W$$

$$\begin{aligned}
\mathbb{E}\left[g(Y)\, h(Z) \mid W\right] &= \mathbb{E}\left[g(Y) \mid W\right] \mathbb{E}\left[h(Z) \mid W\right] \\
&\quad \forall\, g(Y) \text{et } h(Z) \text{ integrable} && (8.3) \\
\mathbb{E}\left[g(Y) \mid Z, W\right] &= \mathbb{E}\left[g(Y) \mid W\right] \\
&\quad \forall\, g(Y) \text{ integrable} && (8.4) \\
\mathbb{E}\left[h(Z) \mid Y, W\right] &= \mathbb{E}\left[h(Z) \mid W\right] \\
&\quad \forall\, h(Z) \text{ integrable} && (8.5) \\
\mathbb{E}\left[\mathbb{E}\left(f(Y, W) \mid W\right) \mid Z\right] &= \mathbb{E}\left[f(Y, W) \mid Z\right] \\
&\quad \forall\, f(Y, W) \text{ integrable} && (8.6)
\end{aligned}$$

Properties

8.1.2 Singular Covariance Matrix

Theorem
Let X be a random vector of $\mathbb{R}^p$, with a vector of mathematical expectation μ, and covariance matrix Σ and support Supp_X Then:

$$Supp_X \subset \{\mu\} + Im(\Sigma)$$

Equivalently:

$$P[X \in \{\mu\} + Im(\Sigma)] = 1$$

If $r(\Sigma) = r < p$, there exists a basis $(a_1, a_2, \cdots, a_{p-r})$ of $\mathcal{N}(\Sigma) = Im(\Sigma)^{\perp}$ such that $V(X'a_j) = 0 \quad j = 1, \cdots, p-r$ ■

Remark
If the distribution of the random vector X is representable through a density, we have:

$$\text{Supp}_X = \{x \in \mathbb{R}^p \mid f_X(x) > 0\}$$

More generally, Supp_X is the smallest set, the closure of which has probability 1.

8.2 Distribution theory

Independence between linear and quadratic forms of normal variates

Theorem
Let

- $X \sim \mathcal{N}_k(O, \Sigma) \qquad r(\Sigma) = k$
- A and B be matrices (of suitable order)

then the following conditions are equivalent:

- $A\Sigma B = 0$
- $X'AX \perp\!\!\!\perp X'BX$ with A and B $k \times k$ SPDS matrices
- $X'AX \perp\!\!\!\perp BX$ with A $k \times k$ SPDS and B $r \times k$ matrices
- $AX \perp\!\!\!\perp BX$ with A $s \times k$ and B $r \times k$ matrices ■

Chi-square distribution

Theorem
Let $X \sim \mathcal{N}_k(O, \Sigma)$ and A be a $k \times k$ SPDS matrix
then the following conditions are equivalent:

- $X'AX \sim \chi^2_{(r)}$
- $\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$
- $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ is a projector

in which case: $r = tr(A\Sigma) = r(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})$ ■

8.3 Asymptotic theory

Different types of convergence
Asymptotic theorems

8.4 Statistical Inference

8.4.1 Notations

In this section, we consider, under suitable regularity conditions, a statistical model, written in terms of densities:

$$[R_X,\, \mathcal{X},\, \{p(x \mid \theta) \;:\; \theta \in \Theta] \qquad \Theta \subset \mathbb{R}^k$$

along with its log-likelihood, its score, its statistical information and its Fisher Information matrix:

$$l_\theta(X) = L(\theta) = \ln p(X \mid \theta) \tag{8.7}$$

$$s_\theta(X) = S(\theta) = \frac{d}{d\,\theta} L(\theta) = [\frac{\partial}{\partial\,\theta_i} \ln p(X \mid \theta)] \tag{8.8}$$

$$j_\theta(X) = J(\theta) = -\frac{d^2}{d\,\theta\, d\,\theta'}\, l_\theta(X) = -[\frac{\partial^2}{\partial\,\theta_i\,\partial\,\theta_j} \ln p(X \mid \theta)] \tag{8.9}$$

$$I_X(\theta) = V(s_\theta(X) \mid \theta) = \mathbb{E}\,(j_\theta(X) \mid \theta) \tag{8.10}$$

$$= -\mathbb{E}\,[\frac{\partial^2}{\partial\,\theta_i\,\partial\,\theta_j} \ln p(X \mid \theta) \mid \theta] \tag{8.11}$$

Under i.i.d. sampling, we have, introducing a subscript n to denote the sample size:

$$L_n(\theta) = \sum_{1 \le k \le n} l_\theta(X_k) \tag{8.12}$$

$$S_n(\theta) = \sum_{1 \le k \le n} s_\theta(X_k) \tag{8.13}$$

$$J_n(\theta) = \sum_{1 \le k \le n} j_\theta(X_k) \tag{8.14}$$

$$I_n(\theta) = V(S_n(\theta)) = n I_{X_1}(\theta) \tag{8.15}$$

8.4.2 Maximun Likelihood Estimation

The maximum likelihood estimator (m.l.e.) of θ is defined as:

$$\hat{\theta} = \arg\sup_{\theta\in\Theta} L(\theta) = \arg\sup_{\theta\in\Theta} p(X \mid \theta) \tag{8.16}$$

Two simple iterative methods to evaluate an m.l.e. Let $\tilde{\theta}_{n,0}$ be an initial estimator.

- Newton-Raphson:

$$\tilde{\theta}_{n,k+1} \;=\; \tilde{\theta}_{n,k} \;+\; [J_n(\tilde{\theta}_{n,k})]^{-1} S_n(\tilde{\theta}_{n,k})$$

- Score

$$\tilde{\theta}_{n,k+1} \;+\; [I_n(\tilde{\theta}_{n,k})]^{-1} S_n(\tilde{\theta}_{n,k})$$

8.4.3 Asymptotic tests

Consider an hypothesis testing,

$$H_0 \;:\; \theta \in \Theta_0 \subset \Theta \qquad \text{against} \qquad H_1 \;:\; \theta \notin \Theta_0$$

where Θ_0 is specified either in the form

$$H_0 \;:\; g(\theta) = 0 \qquad \text{where } \; g \;:\; \Theta \;\longrightarrow\; I\!R^r \;\; (r \leq k)$$

or in the form:

$$H_0 \;:\; \theta = h(\alpha) \qquad \text{where } \; h \;:\; I\!R^p \;\longrightarrow\; \Theta \;\; (p \leq k)$$

i.e. :

$$H_0 \;:\; \Theta_0 \;=\; g^{-1}(0) = Im(h).$$

Let also $\hat{\theta}_0$ be the m.l.e. of θ under H_0 and $\hat{\theta}$ be the unconstrained m.l.e. of θ .

$$\hat{\theta}_0 \;=\; \arg \sup_{\theta \in \Theta_0} p(X \mid \theta) \tag{8.17}$$

$$\hat{\theta} \;=\; \arg \sup_{\theta \in \Theta} p(X \mid \theta) \tag{8.18}$$

Note that:

$$\begin{aligned} \hat{\theta}_0 &= h(\hat{\alpha}_0) \\ \text{where : } \hat{\alpha}_0 &= \arg \sup_{\alpha \in I\!R^p} p(X \mid h(\alpha)) \end{aligned} \tag{8.19}$$

Three standard ways of building a test statistic are the following.

Likelihood Ratio Test

$$L \;=\; -2 \ln \frac{p(X \mid \hat{\theta}_0)}{p(X \mid \hat{\theta})} \tag{8.20}$$

Wald Tests

$$W = (\hat{\theta}_0 - \hat{\theta})' I_X(\hat{\theta})(\hat{\theta}_0 - \hat{\theta}) \tag{8.21}$$

Rao or Lagrange Multiplier Test

$$R = S(\hat{\theta}_0)'[I_X(\hat{\theta}_0)]^{-1} S(\hat{\theta}_0) \tag{8.22}$$

For these three statistics, the critical (or, rejection) region corresponds to large values of the statistics. The level of those tests is therefore the survivor function, (*i.e.* 1- the distribution function), under the null hypothesis, and, under suitable regularity conditions, their asymptotic distribution, under the null hypothesis, is χ^2 with $l - l_0$ degrees of freedom where l, resp l_0, is the (vector space) dimension of Θ, resp.Θ_0.(Note: the vector space dimension of Θ = the dimension of the smallest vector space containing Θ.)

8.4.4 The δ-method

Let $T_n = f_n(X_1, \cdots, X_n)$ be a statistic.

Theorem
If

- $a_n(T_n - b_n) \xrightarrow{d} Y \in \mathbb{R}^p$
- $a_n \uparrow +\infty \qquad b_n \longrightarrow b$
- $g : \mathbb{R}_p \rightarrow \mathbb{R}_q$ continuously differentiable
- $\bigtriangledown = \left[\frac{\partial g_i(y)}{\partial y_j}\right] = \frac{dg}{dy'} \; : \; p \times q$

then:

$$a_n[g(T_n) - g(b_n)] \xrightarrow{d} \bigtriangledown' g(Y)$$

■

In particular, if:

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

then

$$\sqrt{n}[g(T_n) - g(\theta)] \xrightarrow{d} \mathcal{N}(0, \bigtriangledown' \Sigma \bigtriangledown)$$

References

Baltagi, Badi H. (1995), *Econometric Analysis of Panel Data*, New York: Wiley and Sons.

Hausman, J.A. (1978), Specification tests in econometrics, *Econometrica*, **46**, 1251-1271.

Hsiao, C. (1986), *The Analysis of Panel Data*, Cambridge (U.K.): Cambridge University Press.

Little, R.J. and D.Rubin (1987), *Statistical Analysis with Missing Data*, New York: Wiley.

Mátyás, László and Patrick Sevestre (editors) (1996), *The Econometrics of Panel Data*, Dordrecht: Kluwer Academic Publishers.